

Unraveling the Evolution of the Atlantic Cod's (*Gadus morhua* L.) Alternative Immune Strategy

Martin Malmstrøm¹, Sissel Jentoft¹, Tone F. Gregers², Kjetill S. Jakobsen^{1*}

1 CEES (Centre for Ecological and Evolutionary Synthesis), Department of Biosciences, University of Oslo, Oslo, Norway, **2** CIR (Centre for Immune Regulation), Department of Biosciences, University of Oslo, Oslo, Norway

Abstract

Genes encoding the major histocompatibility complex (MHC) have been thought to play a vital role in the adaptive immune system in all vertebrates. The discovery that Atlantic cod (*Gadus morhua*) has lost important components of the MHC II pathway, accompanied by an unusually high number of MHC I genes, shed new light on the evolution and plasticity of the immune system of teleosts as well as in higher vertebrates. The overall aim of this study was to further investigate the highly expanded repertoire of MHC I genes using a cDNA approach to obtain sequence information of both the binding domains and the sorting signaling potential in the cytoplasmic tail. Here we report a novel combination of two endosomal sorting motifs, one tyrosine-based associated with exogenous peptide presentation by cross-presenting MHC I molecules, and one dileucine-based associated with normal MHC II functionality. The two signal motifs were identified in the cytoplasmic tail in a subset of the genes. This indicates that these genes have evolved MHC II-like functionality, allowing a more versatile use of MHC I through cross-presentation. Such an alternative immune strategy may have arisen through adaptive radiation and acquisition of new gene function as a response to changes in the habitat of its ancestral lineage.

Citation: Malmstrøm M, Jentoft S, Gregers TF, Jakobsen KS (2013) Unraveling the Evolution of the Atlantic Cod's (*Gadus morhua* L.) Alternative Immune Strategy. PLoS ONE 8(9): e74004. doi:10.1371/journal.pone.0074004

Editor: Rachel Louise Allen, University of London, St George's, United Kingdom

Received: May 5, 2013; **Accepted:** July 30, 2013; **Published:** September 3, 2013

Copyright: © 2013 Malmstrøm et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Supported by Norwegian Research Council grant# 187940/S10 and University of Oslo, EMBIO. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: k.s.jakobsen@bio.uio.no

Introduction

The antigen presenting class I and II genes of the major histocompatibility complex (MHC) have been identified as crucial components of the adaptive immune system (AIS) in all higher vertebrates, including teleosts [1,2]. Until now, it was generally believed that the MHC genes and their associated immune components have been conserved since their emergence in the jawed fishes, approximately 500 million years BP [1–3]. Sequencing of the Atlantic cod (*Gadus morhua*) genome [4], however, revealed the loss of MHC II as well as the MHC II interacting molecule CD4, required for T-cell activation, and the invariant chain (Ii), facilitating MHC II assembly, transport and peptide loading [5]. As an important part of the AIS, the antigen-presenting molecules MHC I and II help distinguish between self and non-self. Class I molecules normally present endogenously derived pathogens, typically of viral or tumoral origin, while class II molecules present exogenous pathogens such as bacteria and endoparasites. Post-infectious immunological memory, and the acquisition of immunity normally depend on the class II pathway [6]. Malfunction of the MHC II pathway is generally considered to lead to severe immune deficiency or even death.

Another unique feature of the immune system of Atlantic cod is the extreme expansion of MHC I genes. Earlier investigations have shown that Atlantic cod has an extended MHC I repertoire [7,8], compared to other vertebrates. These findings were firmly established by the complete genome sequencing of this species, demonstrating that the Atlantic cod harbors about 100 copies of MHC I in its genome [4], more than twice as many as previously

reported. In evolutionary time, the MHC I gene family has undergone several expansions, and subsequent reductions, especially following speciation and adaptive radiations within phylogenetic lineages [8,9]. These differences are illustrated by the reduced MHC I and II repertoire of early Euteleosts [10], compared to the more advanced Neoteleost, like cichlids [11,12]. The evolutionary arms race against co-evolving pathogens [13], co-evolution with commensal bacteria [14], and in some cases sexual selection [15], have all contributed to a diverse MHC repertoire. The extreme expansion of MHC I genes observed in cod, however, is unique in extant species, as most species retain a few highly conserved (yet polymorphic) “classical” (Ia), and several more divergent “non-classical” (Ib) MHC I genes [16]. It is currently unclear whether this expansion is functionally and evolutionarily linked to the loss of the MHC II pathway.

Furthermore, the immune functionality of Atlantic cod have caused some controversy due to contradicting reports on low to moderate specific antibody response [17,18], and the fact that challenge tests show that Atlantic cod can survive, as well as establish immunity against bacteria [19,20]. These findings indicate that components other than the classical adaptive immune system provide protection. One possible explanation is that the loss of MHC II functionality coincided with changes that allow a more versatile usage of MHC I, indicative through the expansion of this gene complex and the presence of two clades [4]. In mammals, it has been shown that CD8+ T cells can be activated through both the classical MHC I pathway as well as the alternative cross-presentation pathway [21], in which MHC I molecules mimic the

function of class II molecules, presenting exogenous antigens to T-cells [22,23]. In the classical MHC I pathway endogenously derived peptides are loaded onto MHC I within the ER and subsequently presented to CD8+ T-cells at the cell surface (Figure 1a) [24]. MHC II on the other hand is transported from the ER to the endosomal pathway facilitated by endosomal sorting signals within the cytoplasmic tail of the MHC II associated Ii. Here Ii is sequentially degraded and subsequently replaced by peptides derived from exogenous antigens taken up by the cell through endocytosis. Peptide-loaded MHC II is transported to the cell surface for presentation to CD4+ T cells (Figure 1a). In the cross-presentation pathway, exogenously derived peptides are presented in the context of MHC I, as phagocytized bacterial antigenic peptides are loaded onto phagosomal MHC I, which is recycled from the cell surface (Figure 1b) [21,25,26]. The relative importance and functionality of this alternative pathway in mammals is still debated [21,27].

The cross-presentation pathway, like other intracellular transporting pathways, relies on trafficking of molecules facilitated by specific adaptor proteins which recognize and bind to intracellular sorting motifs embedded in the cytoplasmic tail of membrane spanning molecules [28,29]. These conserved motifs act as signals, and in the AIS, they target proteins involved in pathogen recognition and transport them to the endosomes and lysosomes [30]. Dileucine-based and tyrosine-based motifs are the two main classes of sorting signals for endosomal trafficking, important in the

degradation and preparation of extracellular antigen presentation. In humans, both signals have been shown to be involved in cross-presentation via MHC I [30,31], while MHC II trafficking is exclusively facilitated by dileucine signals [32]. The functions of these signaling motifs are highly conserved in all vertebrates, including teleosts, and found in numerous membrane spanning molecules [33]. Both motifs are present in genes involved in antigen presentation in terrestrial vertebrates [30], whereas only dileucine-based signals have been reported in teleost MHC I [16] and II pathways [34].

The rationale for this study was to improve our understanding of the alternative immune system in Atlantic cod, by further characterization of the diverse repertoire of MHC class I genes. Of particular relevance was looking for the presence of sorting signals which would indicate enhanced cross-presentation functionality, thus allowing us to assess whether this pathway could have evolved to play a prominent role in the AIS. Here we report the discovery of a novel combination of sorting motifs in the cytoplasmic tail of MHC I molecules of Atlantic cod, and its proposed role in this alternative immune system.

Results

Expansion of MHC I Loci in Atlantic Cod

In this study we investigated the complete coding regions of the transcribed MHC class I molecules, including the three α -

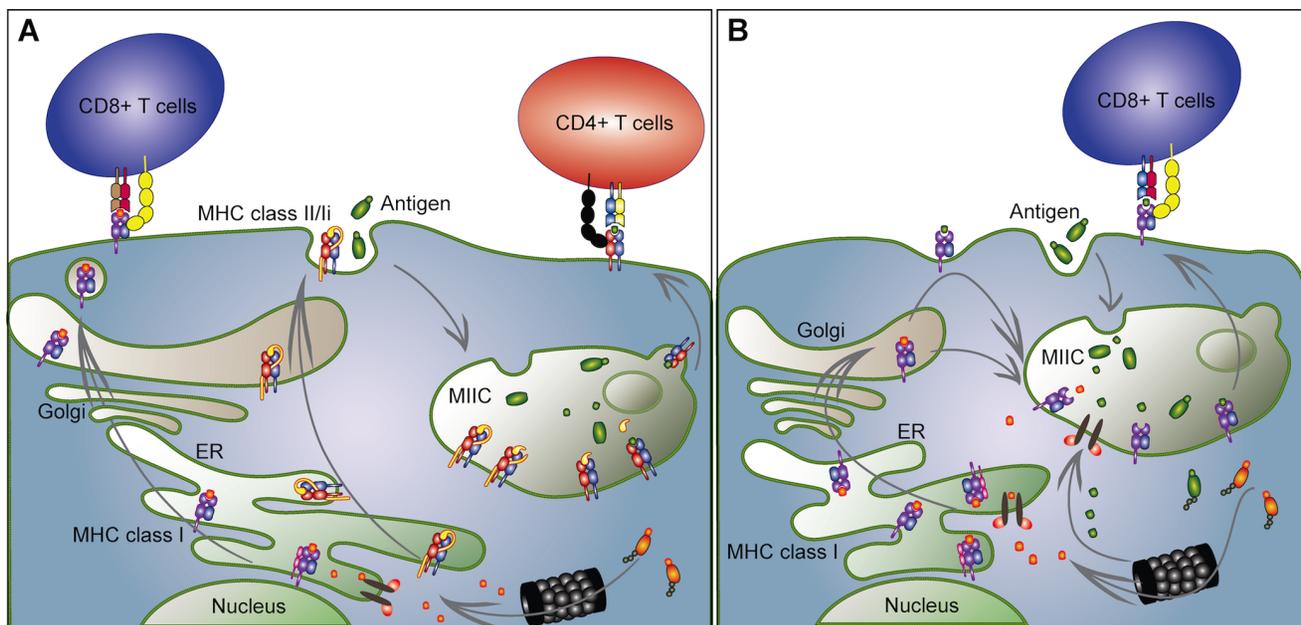


Figure 1. Classical and alternative pathways for antigen presentation. A) *Classical antigen presentation pathways.* MHC class I molecules assemble in the ER together with dedicated chaperones (like tapasin) that retain the MHC class I molecules until peptide binding. Ubiquitinated antigens are degraded by the proteasome, and the resulting peptides are transported via the transporters associated with antigen presentation (TAPs) into the ER lumen. Here the peptides are loaded onto MHC class I, tapasin is released and the peptide-MHC class I complex is transported through the Golgi to the cell surface where they are recognized by specific CD8+ T cells. MHC class II molecules also assemble in the ER with the dedicated chaperone Invariant chain (Ii). Ii mediates trafficking of MHC class II from the ER, through the Golgi, and via the cell surface to the endosomal pathway. Ii is exchanged for degraded exogenous antigenic peptides in specialized MHC class II loading compartments (MIIC). Peptide-loaded MHC class II molecules are released from the endosomal compartment to the cell surface where they are recognized by specific CD4+ T cells (reviewed in [6]). B) *Alternative (Cross-presentation) pathway for exogenous derived peptides by MHC I molecules.* MHC class I molecules carrying signal motifs in the cytoplasmic tail are transported to the endosomal pathway where endocytosed antigens are degraded. Peptides can then be loaded directly in the endosomes in a TAP-independent manner, or the antigens can translocate to the cytosol for proteasomal degradation. The processed antigens can then either be loaded on MHC class I in the ER, or transported back via TAP transporters that have been recruited to the endosomal membrane (reviewed in [35]). Peptide-loaded MHC class I molecules are subsequently released to the cell surface for antigen presentation to CD8+ T cells.

doi:10.1371/journal.pone.0074004.g001

domains, the transmembrane region, and the cytoplasmic tail. Numerous cDNA clones with correct insert length (≈ 1150 bp) were generated from 16 separate PCR reactions, and a total of 192 clones (12 per individual PCR reaction) were selected for Sanger sequencing (see Materials and Methods). Manual curation, including removal of duplicates and sequences likely containing PCR artifacts, reduced the number of unique nucleotide sequences to 143 (Figure 2a). Phylogenetic analysis confirms the previously

observed split of these sequences into two fully supported main clades.

In order to link the information of binding abilities encoded by the $\alpha 1$ and $\alpha 2$ domains to any putative C-terminal signals, we focused our investigation on sequences we could confidently determine not to be chimeric due to PCR artifacts. Only sequences representing identical clones originating from two or more separate PCR reactions were included in the further analysis (see Materials and Methods). 20 sequences fulfilled this criterion (Figure 2b). The selected subset of sequences represents the majority of the basal branches observed in the complete dataset (elongated branches, Figure 2a).

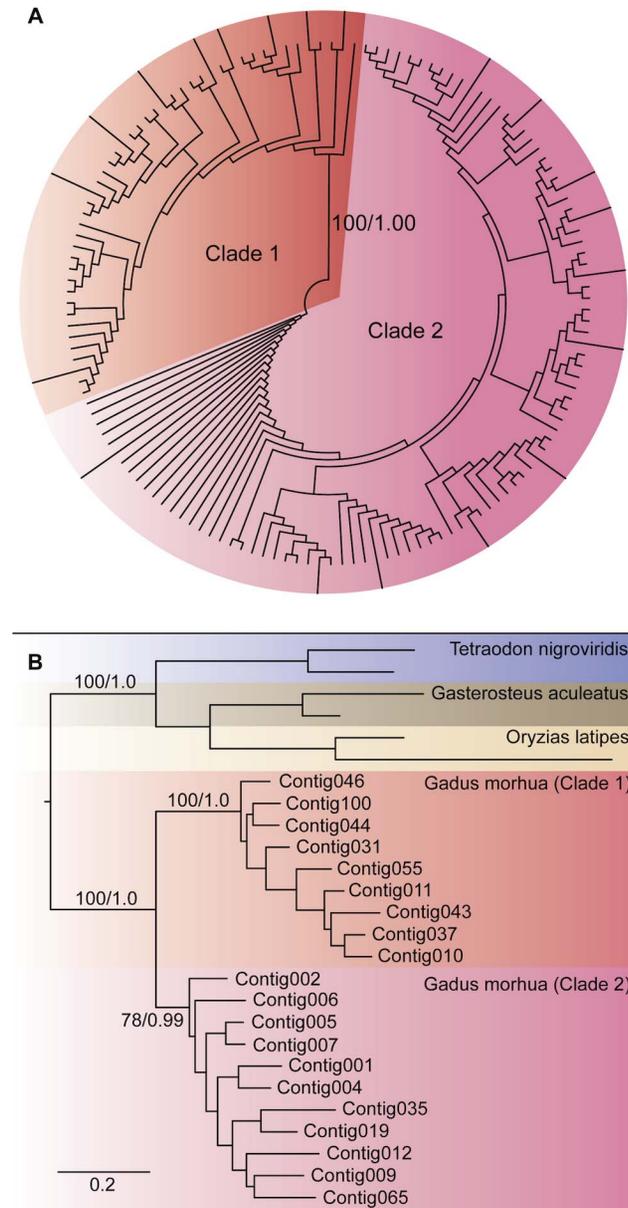


Figure 2. Phylogeny of MHC I diversity in Atlantic cod. A) Unrooted polar cladogram of all unique cDNA sequences of MHC Ia and Ib in Atlantic cod, based on amino acid sequence alignment. Elongated branches illustrate sequences originating from at least two independent PCR reactions. B) Subset of sequences highlighted in a), rooted with additional teleost Ia and Ib sequences from Ensembl. Maximum likelihood (ML) and Bayesian posterior probabilities are shown for the basal branches. Scale bar represents number of amino acid substitutions pr site. doi:10.1371/journal.pone.0074004.g002

Structural Conservation of Sequences

To determine whether the molecules encoded by either clade were atypical in any respect, we investigated the three-dimensional structure predicted by the sequence data. Several conserved features of typical MHC I structure and function were identified. The cysteine bridges in the alpha2 domain (pos. 100 and 164) and alpha3 domain (pos. 200 and 259) as well as the N-glycosylation site (NQT at sites 86 to 88) are completely conserved in all sequences (Figure 3). Other important structural features, e.g. the conserved salt bridges, were also identified (H3-D28, R41-E61, H92-D118, K143-D/E147, D217-R256) in both clades. Further, the acidic domain presumed to be involved in recognition of the T-cell co-receptor CD8 (ELHEQVDPGE at pos. 221 to 230), was also present in all sequences. With the exception of Contig043, which apparently has no cytoplasmic domain, all sequences also contain a transmembrane region and a cytoplasmic tail consistent with typical MHC I structure and function.

Sequence Variation in Binding Domains

In addition to overall structure, polymorphisms within the binding domains are an important trait in typical MHC class I molecules. Using the one-tailed Z-test for positive selection, we revealed a significant excess of non-synonymous mutations in the antigen presenting sites (APS) in both clades (Table 1, Figure 1). As these loci are paralogs from the same individual, signs of positive selection should be interpreted as balancing selection, and hence an evolutionary force promoting a diverse repertoire. Higher d_N to d_S ratio was also observed in non-APS in clade 2. There is no sign of positive selection in the highly conserved and structurally important exon 4 ($\alpha 3$ domain).

Based on the conserved anchoring sites in the binding groove (Figure 3) we assessed whether these molecules could potentially bind peptides. In clade 1, two sequences have sufficient conservation of anchoring sites to be regarded as classical (black branches in top half of Figure 3). Both sequences have eight of the nine anchoring sites conserved, indicating that these genes function as peptide-presenting molecules. The remaining clade 1 sequences are more divergent in their anchoring sites, where only four to six of these sites are conserved, indicating that these transcripts represent non-classical MHC I (Ib) (grey branches in Figure 4). In clade 2 all sequences are highly conserved in their anchoring sites, implying classical (Ia) function for this clade as a whole.

Signaling Motifs in Cytoplasmic Tail

Investigations of the cytoplasmic domain of the selected subset of sequences resulted in the discovery of two putative signal motifs for endosomal trafficking (Figure 5). These signals were only identified in sequences belonging to clade 1. The first signal is a dileucine-based motif (EGQKLA), found in five of the nine sequences in clade 1. The second motif is a tyrosine-based signal

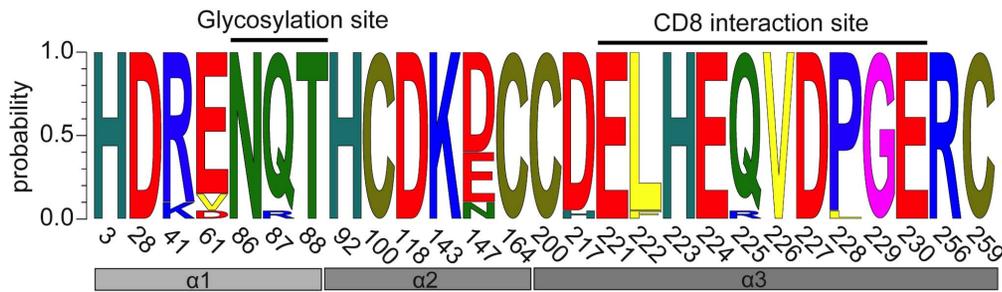


Figure 3. Conserved amino acid sites in typical MHC I molecules. WebLogo presentation of important selected structural and functional sites for subset of MHC I sequences from Atlantic cod. Letter size indicates the probability of the particular amino acids at the given site. Coloring scheme follows standard presentation in MEGA 5.05, reflecting amino acid properties. Numbering is based on consensus sequence, starting at the $\alpha 1$ domain (exon 2).

doi:10.1371/journal.pone.0074004.g003

motif (YQPL) located just two amino acids downstream of the first signal. In Contig055 the second signal contains a point mutation, where the tyrosine (Y) at position 350 has changed to phenylalanine (F), but given the chemical similarity between Y and F, this amino acid change is unlikely to be deleterious for the signal. The degree of nucleotide conservation surrounding the signal motifs suggests that both signals evolved through point mutations rather than by gene recombination mechanisms.

Stop codons causing premature termination were identified in both clades. In clade 1, an insertion leading to a stop codon at position 325 in Contig043 (open triangle in Figure 5) terminates the sequence following the transmembrane region, and thus eliminates the signal motifs during the process of amino acid translation. In clade 2, a point mutation (filled triangle in Figure 5), leads to a truncated cytoplasmic tail in three of the sequences.

Additionally, an extensive comparative analysis of all full-length MHC I coding regions available in the Ensembl Genome Browser from zebrafish, medaka, stickleback, tetraodon and tilapia (see Materials and Methods), revealed no sequences containing both motifs (File S1). A single putative tyrosine based motif was identified in tilapia, and several putative dileucine motifs were identified in a subset of the sequences in all species.

Discussion

Evolution of Novel Signaling Motifs in MHC I – Evidence for an Altered Immune Strategy

Signaling motifs are heavily involved in intracellular transport of immune related molecules. The type of motif determines what adaptor proteins they bind and regulates transportation of the molecule. Normally, one signal is sufficient as the molecules are specialized to follow one particular pathway, and conduct a specific task. Up until now, no molecule has been reported to carry two different signals.

Notably, in Atlantic cod the signal motifs are always found together, and only in one of the MHC I clades (clade 1), implying altered immune function for this clade. The signaling motifs resemble those known from the MHC II pathway [33], and cross-presentation pathway in other vertebrates [30]; this indicates that some of the clade 1 molecules have evolved to function more like class II molecules, as outlined in Figure 1b. This innovation may be an important part of the altered immune strategy that has evolved in Atlantic cod, enabling it to handle exogenous pathogens in absence of the normal MHC II pathway. The two signals will in theory allow the signal-carrying MHC I molecules to follow multiple trafficking pathways into the endosomal compartments and subsequently present extracellular peptides to T-cells [6,35]. Sequences encoding Clade 2 molecules all appear to be classical (Ia) and without any signal motifs, thus they likely maintain the classical function of endogenous peptide presentation to CD8+ T-cells through the constitutive secretory route (see Figure 1a).

In what way the expansion of MHC I genes and the evolution of novel signaling motifs are linked to the loss of MHC II remains to be investigated. So far, two alternative ancestral selection scenarios have been suggested to explain the loss of MHC II in Atlantic cod [36]. One scenario explains the expansion of MHC I as a compensatory mechanism for the loss of the MHC II pathway, while in the other scenario the expansion occurred prior to the loss, rendering the MHC II system obsolete. Large-scale comparative genomics analysis of closely and distant related teleost lineages is needed to disentangle the two scenarios.

In mammalian systems MHC I and II genes are genetically linked [37,38]. In teleosts however, this linkage is broken, as the gene clusters reside on different chromosomes [39–41]. This allows selection to act on each system independently. It follows that “alternative” (in comparison to mammalian) immune strategies are more likely to arise in the teleost lineage than in other vertebrate groups. The extreme expansion of MHC I genes, and the fact that these are divided into two well supported clades, suggests that these genes have been under strong positive

Table 1. Synomus (d_S) and non-synomus (d_N) mutations in functional sites of MHC I molecules.

Sites	no. sequences	Selection	
		$d_N - d_S$	P
Clade 1			
APS	9	6,942	0,000
non-APS	9	0,530	0,298
Exon 4	9	-0,885	1,000
Clade 2			
APS	11	3,993	0,000
non-ABS	11	4,429	0,000
Exon 4	11	0,269	0,616

Identification of antigen presenting sites (APS and non-APS) follows Kaufmann et. al. (1994), with 37 and 148 sites pr. sequence, respectively. Significant values in bold.

doi:10.1371/journal.pone.0074004.t001

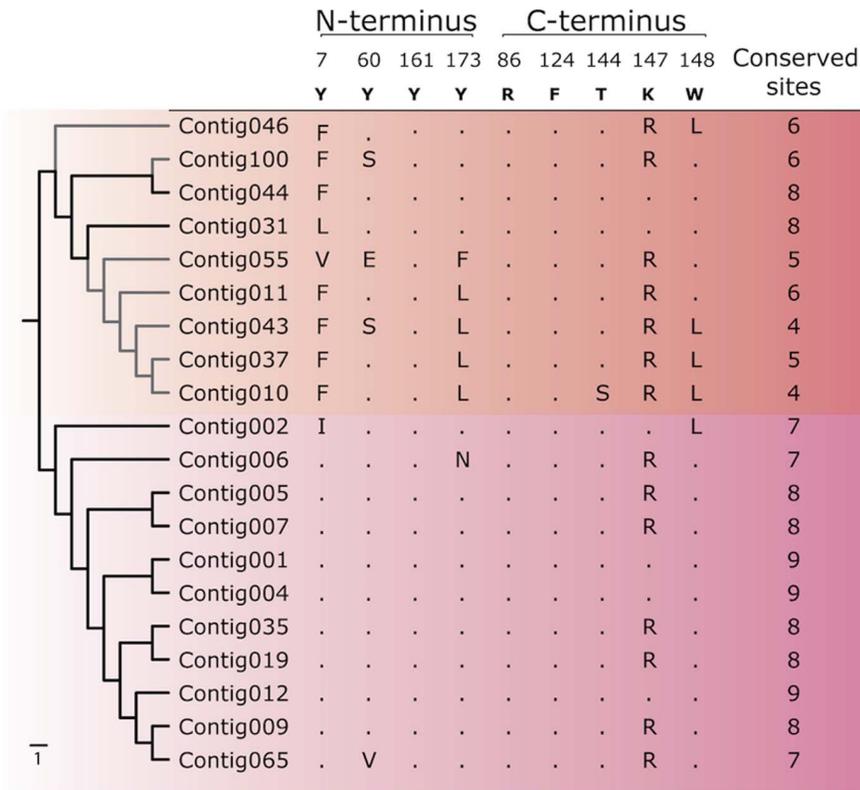


Figure 4. Anchoring sites in MHC I sequences. Amino acids found at conserved anchoring sites are shown for the selected subset of MHC Ia and Ib sequences in Atlantic cod. Conserved teleost amino acids are shown on top. Dots indicate coherence with conserved amino acid, while letters indicate substitute amino acids at each position for each contig. Gray branches represent Ib contigs containing six or fewer conserved sites. Numbering is based on consensus sequence, starting at the $\alpha 1$ domain (exon 2). doi:10.1371/journal.pone.0074004.g004

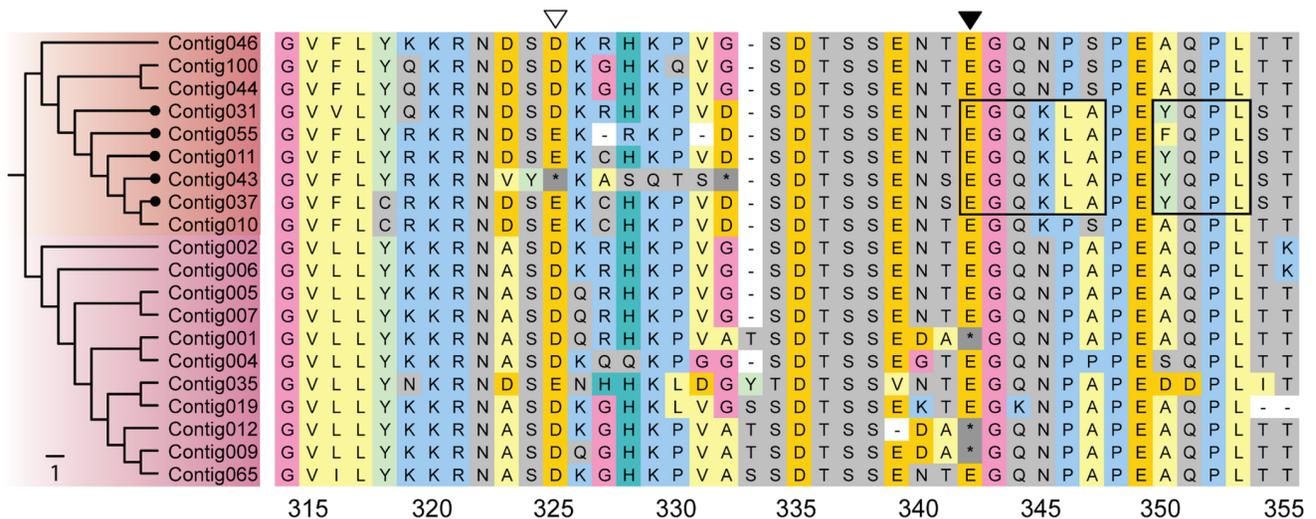


Figure 5. Signal motifs in cytoplasmic tail of MHC I. Amino acid sequences for a manually curated ClustalW nucleotide alignment for a selected subset of MHC Ia and Ib sequences in Atlantic cod. Targeting motifs are boxed (ExxxLA and YxxL), and sequences containing these are indicated with round terminal branches. Open and filled triangles indicate the position of a stop codon (*) in clade 1 and 2 respectively. The “-” represent sequence gaps. The coloring scheme follows standard presentation in MEGA 5.05, reflecting amino acid properties. Numbering of amino acids is based on consensus sequence, starting at the $\alpha 1$ domain (exon 2). doi:10.1371/journal.pone.0074004.g005

(diversifying) selection, and indicates that they have evolved one or several novel functions within the Atlantic cod immune system. Our results suggest that the two clades have experienced different evolutionary pressures; one clade has maintained functionality reflecting 'classical' MHC I, while lack of evolutionary constraint has led to MHC II-like functionality for some representatives of the other clade. As this alternative immune system may be shared with at least some of the other gadoids [4], this system is likely to have evolved millions of years ago.

MHC Class Ia/Ib and MHC I-like Molecules

Both classical (Ia) and non-classical (Ib) MHC I molecules have the same typical appearance and organization, but Ib molecules have usually evolved to serve other immune-related functions such as lipid binding, NK-activation and other immune regulatory functions [42]. Some Ib molecules, however, have been shown to present bacterial antigens [43]. Ia loci are, by definition, highly polymorphic, but as data on population-based polymorphism of specific loci in these highly expanded genes is currently impossible to obtain for Atlantic cod, we have only used conserved anchoring sites in this study as an indication of Ia or Ib function. In this regard we find it valid to question whether the conventional definition of Ia and Ib function is applicable to the unconventional immune system we find in this species, as this definition is based on systems where both classes of MHC molecules are present [42].

The binding abilities are encoded in the groove constituted by the $\alpha 1$ and $\alpha 2$ domains [44]. Most of these antigen-presenting sites are polymorphic, but the nine most *N*- and *C*-terminal amino acids are highly conserved and function to anchor the peptide [45,46]. In the 420my since teleosts diverged from their last common ancestor with mammals [47], the set of conserved mammalian anchoring sites (YYYYYYTKW) is somewhat different in the teleost lineage where the consensus is YYYYRTFKW [48]. In order to present peptides, MHC I molecules should have at least seven of these sites conserved, and thus be coined as classical (Ia) [49]. Interestingly, we find that most, but not all, of the sequences containing the signaling motifs have evolved towards non-classical function, and may no longer have the ability to present peptides. Of course, it should be considered whether the conserved set of anchoring sites for teleosts in general is strictly applicable for Atlantic cod. The prevalent replacement of lysine (K) with arginine (R) at position 147 in both clades seems to be specific for this species. Further analysis is needed to determine whether this arginine should actually be considered to be the most prevalent amino acid at this position. If so, additional clade 1 sequences would be coined classical. The fact that both variants (147 K/R) are found in both clades clearly indicates that the two clades originate from duplication of several genes, and not a single gene duplication event (see Figure 4). Nevertheless, some of the sequences in clade 1 seem to have evolved to serve other immune-related functions as they presumably have lost the ability to present peptides. In this regard, our results on Ia and Ib classification are consistent with findings in other teleosts such as medaka, cichlids, zebrafish, pufferfish, carp [50], rainbow trout [16] and salmon [51], as well as with previous investigations on Atlantic cod [8].

Our findings confirm the conserved structural characteristics of MHC I molecules (Figure 3), show a high degree of variability in the antigen presenting sites of the binding groove (Table1) [52], and reveal an evolutionary pattern in the conserved anchoring sites (Figure 4) for both clades. Collectively these data support the notion that both clades originate from classical MHC I genes and do not represent MHC I-like molecules, or a separation purely of Ia and Ib sequences.

Mutation Pattern in Anchoring Sites Indicate Early Evolution of Signaling Motifs

The mutation pattern of the anchoring sites in the binding groove follows the phylogeny of the complete transcripts to a great extent (see Figure 4). This correlation indicates that non-classical function has evolved several times within clade 1. All sequences in this clade have a mutation in the N-terminus, replacing tyrosine (Y) with a hydrophobic amino acid. This mutation most likely has only minor effects on the peptide binding ability, as the same mutation is also found at site 124(F) for all teleosts, compared to the mammalian counterpart [48]. Following the separation of the two clades our data show a mutation in the C-terminus where tryptophan (W) is replaced by leucine (L), presumably leading to non-classical function of the loci represented by Contig046. As the tryptophan (148 W) reappears in sequences branching off at the more distal nodes in the tree, some loci have retained the conserved amino acids – represented by contigs 044 and 031. This finding suggests that the evolution of the cytoplasmic signaling motifs has occurred prior to the emergence of genes represented by Contig031, indicating that these motifs most likely evolved basally in the Atlantic cod lineage.

Conclusions

We here report the discovery of a novel combination of two sorting motifs that are normally associated with exogenous peptide presentation and cross-presentation by MHC class II molecules and MHC class I molecules, respectively. These findings indicate an altered functionality of MHC class I molecules in Atlantic cod and elucidate new insight into the plasticity and evolution of the vertebrate immune system.

Materials and Methods

Ethical Statement

We always aim to limit the effect of our research on populations and individuals. Whenever possible we collaborate with other sources, such as commercial fisheries or aquaculture farms, where samples can be harvested freely in combination with their normal business. This way, no animals need to be euthanized to serve our scientific purpose alone. The specimen used in this study comes from a wild population and was part of a larger haul of commercially fished individuals intended for human consumption. Following capture the fish were immediately stunned by bleeding following standard procedure by a local fisherman. Sampling in this manner does not fall under any specific legislation in Norway, but it is in accordance with the guidelines set by the 'Norwegian consensus platform for replacement, reduction and refinement of animal experiments' (www.norecopa.no).

Sample Extraction and Purification

Spleen tissue from a single individual of Atlantic cod from the Lofoten area (68°8'48"N 13°36'35"E) in Norway was used. Total mRNA was extracted using 'Dynabeads DIRECT mRNA Isolation Kit' (Life Technologies, Carlsbad, California, U.S.), and gDNA was removed following the 'Qiagen RNeasy MinElute Cleanup' (Quiagen, Venlo, Netherlands) protocol. cDNA was synthesized using random hexamer primers (Roche, Penzberg, Germany) and 'First strand cDNA Synthesis Kit' (Fermentas, Vilnius, Lithuania). A final clean up and concentration was conducted with the 'QIAquick PCR Purification Kit' (Quiagen, Venlo, Netherlands), for a final concentration of 30 ul. All procedures were carried out following the manufacturers' instructions.

Amplification, Cloning and Sequencing

We chose to use cDNA in this experiment; this enabled us to sequence both the 5' and 3' end of the molecules and sequence them in one reaction. This was important as we wanted to investigate the cytoplasmic tail of these molecules, and couple any informational signals there to the upstream parts of the molecule. This approach also excludes any unexpressed pseudo-genes from the dataset. We have not attempted to analyze the total diversity of MHC class I, but rather illustrate the novelties which lay hidden in this diverse repertoire. Due to the extreme expansion and repetitive nature of the MHC I gene, with large regions present in near-identical copies between loci, it is still not possible, even with high throughput sequencing technologies and state-of-the-art bioinformatics, to assemble, classify and determine the genomic structure of all MCH I loci in Atlantic cod.

Universal MHC I primers for Atlantic cod, based on all available data from the Cod Genome Project (GenBank accession numbers JX567622 - JX567728) and other NCBI sequences (AJ132511–132529 and AF414203–AF414220), were designed for exon 1 (5'-CTGCTGTTGRTCTTTGGTCA) and exon 7 (5'-AAYGTGAGAAGMCTCTTCATG). As MHC I sequences are particularly prone to chimeric PCR generated errors [53], we ran 16 independent PCR reactions in parallel. Each PCR reaction of 10 μ l was run with 'BD Advantage 2 Polymerase Mix' (BD Biosciences, San Jose, California, U.S.) under the following conditions: 94°C denaturation for 2 min, then running 25 cycles of 94°C 30 s, 56°C 30 s, 68°C 60 s, and 68°C elongation for 5 min. Following the PCR amplification, 3'-A- overhangs were added using 'Dream-Taq DNA polymerase' (Fermentas, Vilnius, Lithuania), before each pool of amplicons was cleaned up using 'Wizard® SV Gel and PCR Clean-Up System' (Promega, Fitchburg, Wisconsin, U.S.). Cloning was performed independently for amplicons from each of the 16 PCR reactions, using 'TOPO TA-Cloning Kit' in 'One Shot TOP10 Chemically Competent E. coli' by Invitrogen (Life Technologies, Carlsbad, California, U.S.) following manufacturers' instructions. 48 clones originating from each PCR reaction were screened on an agarose gel, of which 12 clones with the right insert size were picked, for a total of 192 clones. These were sequenced with conventional ABI 3730 technology. The 143 sequences included in this study have been submitted to GenBank with submission ID: 1563074.

Sequence Handling and Phylogenetic Analysis

Raw sequence was manually inspected and corrected in Sequencher 5.0.1 (Gene Codes Corporation). Contigs representing unique sequences on a nucleotide level were aligned using ClustalW [54] as implemented in MEGA 5.05 [55] and the alignment was manually curated. Comparative sequences for rooting were downloaded from the Ensembl Genome Browser with accession numbers: ENSTNIT00000000248, ENSTNIT000000003613, ENSGACT000000002570, ENSGACT00000000184, ENSORLT000000021463 and ENSORLT000000008514; these are pufferfish (*Tetraodon nigroviridis*), stickleback (*Gasterosteus aculeatus*), and Medaka (*Oryzias latipes*), respectively.

Tree topology and bootstrapping ($n=100$) for Maximum Likelihood were computed using RAxML HPC-PTHREADS (Version 7.2.6) [56] under the PROTGAMMAIJTTF model, suggested by ProtTest (Version 2.4) [57]. Bayesian posterior probabilities were calculated using MrBayes v3.1.2 [58,59], run with 4 chains and with 5.0 million generations, and were sampled every 1000th generation. Burnin was set to 40000. Site-specific rate

model was set to "variable", and the rate matrix for amino acids set to "fixed (jones)". Parameters for the likelihood model were set to "invgamma", and the model allowed the site-specific rate of change to vary over its evolutionary history using the "covarion" setting.

WebLogo was created using WebLogo 3 [60], and colours adjusted to MEGA 5.05 standard in Adobe Illustrator CS4.

Comparative Analysis of Cytoplasmic Sequences

A total of 151 amino acid-translated transcripts from zebrafish, medaka, stickleback, tetraodon and tilapia were selected for comparative analysis of the cytoplasmic tail. These were detected using the in-built BLAST function of the Ensembl Genome Browser, with ENSDARP00000020667 (*Danio rerio*), ENSORLP00000001303 (*Oryzias latipes*), ENSGACP00000000148 (*Gasterosteus aculeatus*), ENSTNIP000000002995 (*Tetraodon nigroviridis*) and ENSONIP000000006183 (*Oreochromis niloticus*) as queries. All sequences were aligned using ClustalW [54] as implemented in MEGA 5.05 [55] individually for each species. Sequences too divergent to be aligned, or missing larger sections of sequence, including the cytoplasmic domain, were removed. The resulting 72 sequences were then manually inspected and compared to a selection of sequences from Atlantic cod, both with and without signaling motifs (File S1).

Detection of Selection

Comparison of non-synonymous (dN) and synonymous (dS) mutations for detection of selection per site was done with the 'One-tailed Z-test' as implemented in MEGA 5.05 [55]. When the relative rate of dN to dS is equal ($dN - dS = 0$) a site is evolving neutrally. An excess of dN relative to dS ($dN - dS > 0$) is indicative of positive (diversifying/balancing) selection, whereas an the opposite is indicative of purifying (negative) selection. The test report average values of dN-dS for each of the sequence partitions, and sequence sets tested. The P score represents the probability of rejecting the null hypothesis of strict-neutrality ($dN = dS$) in favor of the alternative hypothesis ($dN > dS$). The variance of the difference was estimated using the bootstrap method (1000 replicates). Analyses were conducted using the Nei-Gojobori method [61]. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). A total of 285 positions were analyzed, 37 of which are defined as APS, 148 as non-APS, and 100 as Exon 4.

Supporting Information

File S1 Teleost cytoplasmic tail sequences. Alignment of cytoplasmic tail for the 72 full-length MHC I coding regions available in the Ensembl Genome Browser from zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*) and tilapia (*Oreochromis niloticus*). A subset of Atlantic cod (*Gadus morhua*) sequences, with and without signaling motifs, is included for comparison. All gaps have been removed. (DOCX)

Acknowledgments

We would like to thank Monica H. Solbakken, Bastiaan Star and Anna V. B. Mazzarella for helpful discussions and critical reading of the manuscript. The ABI platform at the Department of Biosciences and the Norwegian Sequencing Centre (NSC; www.sequencing.uio.no) are acknowledged for their help.

Author Contributions

Conceived and designed the experiments: MM SJ TFG KSJ. Performed the experiments: MM. Analyzed the data: MM. Contributed reagents/materials/analysis tools: MM KSJ. Wrote the paper: MM SJ TFG KSJ.

References

- Litman GW, Rast JP, Fugmann SD (2010) The origins of vertebrate adaptive immunity. *Nat Rev Immunol* 10: 543–553.
- Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet* 11: 47–59.
- Schluter SF, Bernstein RM, Marchalonis JJ (1999) ‘Big Bang’ emergence of the combinatorial immune system. *Dev Comp Immunol* 23: 107–111.
- Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, et al. (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477: 207–210.
- Landsverk OJ, Bakke O, Gregers TF (2009) MHC II and the endocytic pathway: regulation by invariant chain. *Scand J Immunol* 70: 184–193.
- Neeffes J, Jongsma MLM, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 11: 823–836.
- Persson A-C, Stet RJM, Pilstrom L (1999) Characterization of MHC class I and β 2-microglobulin sequences in Atlantic cod reveals an unusually high number of expressed class I genes. *Immunogenetics* 50: 49–59.
- Miller KM, Kaukinen KH, Schulze AD (2002) Expansion and contraction of major histocompatibility complex genes: a teleostean example. *Immunogenetics* 53: 941–963.
- Klein J, Sato A, O’huigin C (1998) Evolution by gene duplication in the major histocompatibility complex. *Cytogenetic and Genome Research* 80: 123–127.
- Miller KM, Withler RE (1998) The salmonid class I MHC: limited diversity in a primitive teleost. *Immunol Rev* 166: 279–293.
- Sato A, Klein D, Sultmann H, Figueroa F, O’huigin C, et al. (1997) Class I MHC genes of cichlid fishes: identification, expression, and polymorphism. *Immunogenetics* 46: 63–72.
- Málaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, et al. (1998) Linkage Relationships and Haplotype Polymorphism Among Cichlid Mhc Class II B Loci. *Genetics* 149: 1527–1537.
- Borghans JM, Beltman J, Boer R (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics* 55: 732–739.
- Lee YK, Mazmanian SK (2010) Has the microbiota played a critical Role in the evolution of the adaptive immune system? *Science* 330: 1768–1773.
- Edwards SV, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution* 13: 305–311.
- Dijkstra J, Kiryu I, Yoshiura Y, Kumánovics A, Kohara M, et al. (2006) Polymorphism of two very similar MHC class Ib loci in rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics* 58: 152–167.
- Pilstrom L, Warr GW, Strömberg S (2005) Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation. *Fisheries Science* 71: 961–971.
- Schröder MB, Ellingsen T, Mikkelsen H, Norderhus EA, Lund V (2009) Comparison of antibody responses in Atlantic cod (*Gadus morhua* L.) to *Vibrio anguillarum*, *Aeromonas salmonicida* and *Francisella* sp. *Fish & Shellfish Immunology* 27: 112–119.
- Espelid S, Rødseth OM, Jørgensen TO (1991) Vaccination experiments and studies of the humoral immune responses in cod, *Gadus morhua* L., to four strains of monoclonal-defined *Vibrio anguillarum*. *Journal of Fish Diseases* 14: 185–197.
- Mikkelsen H, Lund V, Larsen R, Seppola M (2011) Vibriosis vaccines based on various sero-subgroups of *Vibrio anguillarum* O2 induce specific protection in Atlantic cod (*Gadus morhua* L.) juveniles. *Fish & Shellfish Immunology* 30: 330–339.
- Amigorena S, Savina A (2010) Intracellular mechanisms of antigen cross presentation in dendritic cells. *Current Opinion in Immunology* 22: 109–117.
- Kovacovics-Bankowski M, Rock KL (1995) A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* 267: 243–246.
- Rock KL (1996) A new foreign policy: MHC class I molecules monitor the outside world. *Immunology Today* 17: 131–137.
- Pamer E, Cresswell P (1998) Mechanisms of MHC class I-restricted antigen processing. *Annu Rev Immunol* 16: 323–358.
- Gromme M, Uytdehaag FG, Janssen H, Calafat J, van Binnendijk RS, et al. (1999) Recycling MHC class I molecules and endosomal peptide loading. *Proc Natl Acad Sci U S A* 96: 10326–10331.
- Zou L, Zhou J, Zhang J, Li J, Liu N, et al. (2009) The GTPase Rab3b/3c-positive recycling vesicles are involved in cross-presentation in dendritic cells. *Proc Natl Acad Sci U S A* 106: 15801–15806.
- Ramachandra L, Simmons D, Harding CV (2009) MHC molecules and microbial antigen processing in phagosomes. *Curr Opin Immunol* 21: 98–104.
- Breifeld PP, Casanova JE, Simister NE, Ross SA, McKinnon WC, et al. (1989) Sorting signals. *Current Opinion in Cell Biology* 1: 617–623.
- Mellman I (1996) Endocytosis and molecular sorting. *Annu Rev Cell Dev Biol* 12: 575–625.
- Lizée G, Basha G, Jefferies WA (2005) Tails of wonder: endocytic-sorting motifs key for exogenous antigen presentation. *Trends in Immunology* 26: 141–149.
- Basha G, Lizée G, Reinicke AT, Scipp RP, Omilusik KD, et al. (2008) MHC Class I Endosomal and Lysosomal Trafficking Coincides with Exogenous Antigen Loading in Dendritic Cells. *PLoS ONE* 3: e3247.
- Bakke O, Nordeng TW (1999) Intracellular traffic to compartments for MHC class II peptide loading: signals for endosomal and polarized sorting. *Immunological Reviews* 172: 171–187.
- Bonifacio JS, Traub LM (2003) Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annual Review in Biochemistry* 72: 395–447.
- Silva DSP, Reis MIR, Nascimento DS, Vale Ad, Pereira PJB, et al. (2007) Sea bass (*Dicentrarchus labrax*) invariant chain and class II major histocompatibility complex: Sequencing and structural analysis using 3D homology modelling. *Molecular Immunology* 44: 3758–3776.
- Joffre OP, Segura E, Savina A, Amigorena S (2012) Cross-presentation by dendritic cells. *Nat Rev Immunol* 12: 557–569.
- Star B, Jentoft S (2012) Why does the immune system of Atlantic cod lack MHC II? *BioEssays* 34: 648–651.
- Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, et al. (2000) Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. *Proc Natl Acad Sci U S A* 97: 4712–4717.
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Reviews* 190: 95–122.
- Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, et al. (1997) Mapping of Mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* 46: 129–134.
- Hansen JD, Strassburger P, Thorgaard GH, Young WP, Du Pasquier L (1999) Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. *J Immunol* 163: 774–786.
- Sato A, Figueroa F, Murray BW, Málaga-Trillo E, Zaleska-Rutczynska Z, et al. (2000) Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics* 51: 108–116.
- Rodgers JR, Cook RG (2005) MHC class Ib molecules bridge innate and acquired immunity. *Nat Rev Immunol* 5: 459–471.
- Pamer EG, Bevan MJ, Lindahl KF (1993) Do nonclassical, class Ib MHC molecules present bacterial antigens to T cells? *Trends in Microbiology* 1: 35–38.
- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, et al. (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329: 506–512.
- Du Pasquier L (1992) Origin and evolution of the vertebrate immune system. *APMIS: acta pathologica, microbiologica, et immunologica Scandinavica* 100: 383–392.
- Kaufman J, Salomonsen J, Flajnik M (1994) Evolutionary conservation of MHC class I and class II molecules—different yet the same. *Seminars in Immunology* 6: 411–424.
- Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56: 683–695.
- Hashimoto K, Okamura K, Yamaguchi H, Ootake M, Nakanishi T, et al. (1999) Conservation and diversification of MHC class I and its related molecules in vertebrates. *Immunological Reviews* 167: 81–100.
- Madden DR (1995) The Three-Dimensional Structure of Peptide-MHC Complexes. *Annual Review of Immunology* 13: 587–622.
- Matsuo MY, Asakawa S, Shimizu N, Kimura H, Nonaka M (2002) Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). *Immunogenetics* 53: 930–940.
- Lukacs MF, Harstad H, Bakke HG, Beetz-Sargent M, McKinnel L, et al. (2010) Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. *BMC Genomics* 11: 154.
- Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167–170.
- Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci – Implications for evolutionary analysis. *Gene* 427: 117–123.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
- Stamatakis A, Ludwig T, Meier H (2005) RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21: 456–463.

57. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105.
58. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science* 294: 2310–2314.
59. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
60. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188–1190.
61. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426.