Evolutionary and functional insight into the teleost immune system — lessons learned from Atlantic cod and other teleosts

BY

MONICA HONGRØ SOLBAKKEN

DISSERTATION PRESENTED FOR THE DEGREE OF

PHILOSOPHIAE DOCTOR

2016

Centre for Ecological and Evolutionary Synthesis

Department of Biosciences

Faculty of Mathematics and Natural Sciences

University of Oslo

“To find yourself, think for yourself.”

- Socrates -

Acknowledgments

The work presented in this thesis was performed at the Centre for Ecological and Evolutionary Synthesis, Department of Biosciences at the University of Oslo. It was funded by The Research Council of Norway and supervised by Professor Kjetill S. Jakobsen, Dr. Sissel Jentoft, Dr. Tone F. Gregers and Professor Oddmund Bakke.

Dear supervisors, I would like to thank you for encouraging me to do a project initially far outside of my comfort zone. I have learned amazingly much exploring aspects of biology that were new to me. Collectively, you have fulfilled every role from enthusiastic motivators to critical challengers and I am very grateful for all that you have done for me. Kjetill and Sissel, I am so happy you stayed for the ride. Your mentorship has influenced me tremendously and I could not have done this without you. Thank you! Tone, thank you for helping me negotiate the difficult transition between classic and comparative immunology and for the support when I needed it. Oddmund, even though the frequency of meetings has been low, I appreciate the enthusiasm and interest in my projects and for the support and help during the final stages of my PhD.

Dear collaborators, both on PhD papers as well as other projects, thank you for the opportunity to collaborate with you, for the materials given, for the analyses provided, for the interpretations delivered, for discussions made and for allowing me to do the same. I find that it has challenged me and broadened my horizon working with you and I have enjoyed every project.

Nils Christian Stenseth, I would like to thank you for running such a great center and for being there when I needed insight from someone outside the bubble. I also would like to thank the CEES administration for smooth operations and Sumera Majid for the open door policy throughout this PhD.

Dear colleagues at CEES, you have provided a great working environment. You have listened, given feedback, educated, supported and encouraged. You are a great bunch of people! Martin, thank you for being my great officemate for all these years — through the good and the bad! Helle and Rebekah, you have made the final leg of this PhD much smoother with your great personalities and scientific insights. It has been great having someone who works more along the lines that I do and who quickly became my friends. Thank you for everything!

My gals: Live, Raimonda, Helene, Kine, Karène, Lene and Benedicte. Sometimes you meet people who you connect with on a different level. You are these people for me. I deeply cherish all of you and hope to have you as my friends for many years to come. Thank you for the space given these last couple of years and for your everlasting support!

Kjære mamma og pappa, stepappa, lillebror, svigermor og svigerfar. Jeg vet at dette har vært en reise som har vært vanskelig å forstå og få innsikt i. Jeg er utrolig takknemlig for at dere har holdt ut nedturene, har gitt meg litt ekstra takhøyde, for rådene dere har gitt, for at dere har støttet meg, at dere har vist interesse og genuint har hatt lyst til å forstå hva jeg egentlig har hold på med. Tusen takk!

Lars Erik, my dearest husband, you are my both my opposite and my equal. Your strength is always there to support me, your smile whenever I need cheering up and your heartfelt excitement whenever things go my way. Words cannot express how much you mean to me!

My dearest son, thank you for the joy you bring me every single day!

Monica H. Solbakken

Oslo, June 2016

Table of Contents

List of papers

Paper I

Evolutionary redesign of the Atlantic cod (Gadus morhua L.) Toll-like receptor repertoire by gene losses and expansions

Monica H. Solbakken, Ole K. Tørresen, Alexander J. Nederbragt, Marit Seppola, Tone F. Gregers, Kjetill S. Jakobsen, Sissel Jentoft

Scientific Reports, 6:25211, DOI: 10.1038/srep25211

Paper II

Unveiling the evolution of the teleost innate immune system

Monica H. Solbakken, Kjetil Lysne Voje, Kjetill Sigurd Jakobsen, Sissel Jentoft

Submitted

Paper III

Successive losses of central immune genes characterize the Gadiformes' alternate immunity

Monica H. Solbakken, Matthew L. Rise, Kjetill S. Jakobsen, Sissel Jentoft.

Submitted

Paper IV

Disentangling the immune response and host-pathogen interactions in Francisella noatunensis infected Atlantic cod

Monica H. Solbakken, Sissel Jentoft, Trond Reitan, Helene Mikkelsen, Tone F. Gregers, Oddmund Bakke, Kjetill S. Jakobsen, Marit Seppola.

Manuscript

Paper V

Whole transcriptome analysis of the Atlantic cod vaccine response reveals no conventional adaptive immunity

Monica H. Solbakken, Sissel Jentoft, Trond Reitan, Helene Mikkelsen, Kjetill S. Jakobsen, Marit Seppola.

Manuscript

Abbreviations

ALR	Absent in melanoma (AIM) like receptor
AMP	Antimicrobial peptide
APC	Antigen presenting cell
B2M	Beta 2-microglobulin
BCR	B-cell receptor (alias immunoglobulin, antibody)
CLR	C-type lectin receptor
DAMP	Damage-associated molecular pattern
DC	Dendritic cell
DNA	Deoxyribonucleic acid
IFN	Interferon
IgX	Immunoglobulin of isotype X (alias BCR, antibody)
IL	Interleukin
ILC	Innate lymphoid cell
LPS	Lipopolysaccharide
MHC	Major Histocompatibility Complex
Mx	Myxovirus-resistance
NAb	Natural antibody
NK	Natural killer
NKT	Natural killer T-cell
NLR	Nucleotide-binding oligomerization domain (NOD) like receptor
PAMP	Pathogen-associated molecular pattern
PRR	Pattern recognition receptor
RLR	Retinoic acid-inducible (RIG) like receptor
RNA	Ribonucleic acid
RNAseq	In this thesis: total mRNA sequencing
TCR	T-cell receptor
TGF	Transforming growth factor
TLR	Toll-like receptors
TNF	Tumor necrosis factor
V(D)J	Variable, diversity and joining type segments

Thesis summary

Teleosts comprise a very diverse group of species where genome sequencing the last decade has revealed great variability regarding their genetic basis for immunity. With emphasis on the innate immune system, this thesis addresses evolutionary and functional aspects of teleost immunity using high-throughput sequencing and bioinformatics analysis tools. The losses and expansions of Toll-like receptors (*TLR*) and loss of Major Histocompatibility Complex class II (*MHCII*) were originally discovered in Atlantic cod. *TLRs* are a gene family of pattern recognition receptors central to the functionality of innate immunity. The first paper in this thesis describes the in-depth characterization of Atlantic cod *TLRs*. The *TLR* repertoire, extreme compared to other teleosts and vertebrates, indicated a correlation with *MHCII* loss. In addition there were signs of diversifying selection within the *TLR* gene expansions suggesting sub- and neofunctionalization of the duplicated genes. Recently, through the use of new genome resources from 66 teleost species and corresponding species phylogeny, the loss of *MHCII* from the entire Gadiformes order was established. Using these genomes and phylogeny, we established the correlation between *MHCII* loss and *TLR* expansion within the order of Gadiformes (cod-like fish species). Moreover, we established correlations between the *TLR* expansion, species maximum depth and species latitudinal distribution — likely proxies for environmental abiotic factors such as temperature. We also investigated another long sought after gene in Atlantic cod — the Myxovirus resistance gene (*Mx*) — a viral infection marker with unknown function and considered a part of the innate immune system. Adding the *Mx* related findings to the teleost phylogeny showed that the changes to the teleost immune system are of a successive nature. The timing of the phylogeny demonstrated that well-described large alterations in past environment such as oceanic oxygen levels, temperature and layout of tectonic plates overlapped with the changes to the teleost immune system — illustrating

the combined effect of host-intrinsic, biotic and abiotic factors on the evolution of the teleost immune system.

In summary, the teleost immune system display a higher degree of diversity compared to other vertebrates. Intriguingly, the functional adaptive immune response of Atlantic cod and other Gadiformes have been found to deviate from that of other investigated teleosts. In contrast, functional studies on Gadiformes, mainly Atlantic cod, demonstrate the presence of well-described innate responses such as inflammation, cellular and humoral defenses. However, the underlying genetic repertoire (loss of *MHCII* and related factors and the extreme repertoire of *TLRs*) means that the well-studied mechanisms leading to immunity in other vertebrates do not apply to Gadiformes. In this thesis we present the first overall description of the transcriptomic mechanisms related to bacterial infection and immersion vaccination. Overall, Atlantic cod — and thus likely most Gadiformes since they lack *MHCII* — paint a transcriptional picture fitting the classic usage of innate defenses with inflammation and recruitment of phagocytic cells. With respect to adaptive immunity, Atlantic cod uses cytotoxic defenses through the presentation of both endogenous and exogenous antigen on MHCI and T-cell independent activation of B-cells for the generation of antibodies and possibly establishment of subsequent memory through both B-cell and T-cell lineages. However, in relation to vaccination, Atlantic cod appears to apply unconventional mechanisms. There were no significant findings of inflammation or cell recruitment combined with a very weak response related to MHCI and antibody production. Thus, alternative mechanisms leading to memory, related to both innate and adaptive immunity, were considered. Here, innate memory through NK-cell lineages or through a metabolism-related epigenetic imprinting was deemed the most likely based on the expression data. However, none of the suggested involved mechanisms in either experiment — adaptive or innate — excludes the other.

Overall, this thesis elaborates on the more recent knowledge regarding the immunological strategy of vertebrates. Most importantly, it shows that teleosts appear to harbor the most immunological diversity within the vertebrate lineage. This is particularly evident within the Gadiformes lineage containing paradigm changing gene losses (*MHCII*) and expansions (*MHCI*, *TLRs*) where immunity is found to be well orchestrated by the Atlantic cod transcriptional investigations presented here. Collectively, the results described in this thesis enables the targeted design of future experimental investigations to further deduce the functional details of mechanisms underlying immunological memory in Atlantic cod and other Gadiformes.

Background

Bony fish (superclass Osteichthyes), where teleost make up the largest infraclass, comprises an exceptionally diverse group with species inhabiting numerous marine and freshwater habitats across the globe. This diversity is also mirrored by their life history strategies, morphological varieties and migratory behavior [1-3]. After years of sequencing non-model vertebrate genomes, it is now evident that the bony fish diversity also encompasses the genetic basis of immunity [4-9]. In 2011, the Atlantic cod (*Gadus morhua*) genome was published, where we discovered unforeseen gene losses and expansions of components central to the vertebrate immune system [10]. These findings spurred several projects of which a new version of the Atlantic cod genome has been generated [11], and characterization of the same gene losses found in Atlantic cod has been performed in 66 new teleost species [12]. Five studies, emphasizing the innate immune system in Atlantic cod, and in teleosts overall, form the foundation of this thesis. Collectively, they address both genetic and functional aspects of the Atlantic cod immune system, as well as characterizing immunological repertoires in teleosts overall within an evolutionary framework.

Vertebrate genome duplications and teleost diversity

The evolution and diversification of the vertebrate lineage coincided with novel genetic innovations, which also affected the genetic repertoire underlying the vertebrate immune system. The increase in vertebrate morphological complexity has been connected to whole genome duplications, and animals generally allocate large genome resources to their repertoire of immune genes. In the vertebrate lineage there are two well-characterized duplication events — the first in the vertebrate ancestor and the second in the transition between jawless and jawed vertebrates [13]. In addition, bony fish have experienced a third event [14, 15]. Many consider these genome duplications in the context of Ohno's hypothesis implying that the genome duplications, by generating excessive

amounts of genes, have permitted extensive innovations by the selection of sub- and neofunctionalization [13].

Evolution of the vertebrate immune system

Evolutionary trajectory

All organisms harbor elements that protect them from pathogens, such as physical barriers and antimicrobial peptides, which are collectively termed innate immunity. Further, and also a part of the innate immune system, all organisms are at some level able to discriminate between self and non-self. This discrimination is focused at maintaining homeostasis, integrity, and survival of the organism by enabling detection of food sources, sexual exchange of genetic material and the ability to separate safe from harmful. Overall, discrimination in eukaryotes is generated through receptors that are paired with phagocytosis. This mechanism provides uptake of nutrients by engulfment of extracellular material simultaneously to functioning as an inducible defense mechanism able to clear pathogens from the immediate vicinity of the eukaryote [16-19]. Moving beyond the unicellular eukaryotic organism, the term immunity develops into a description of a much more complex system. Protecting the host is here strongly influenced by the increased complexity of the organism itself and the presence of commensal bacteria. This requires additional levels of immune recognition, regulation and response compared to the unicellular eukaryotes. To enable a more complex immune system, multicellular organisms display compartmentalization and enhancement of immune functions and the further development of somatic cells into specialized immune cells [16, 17].

In the common ancestor of plants and invertebrates, specialized pattern recognition receptors (PRRs) arose, which further developed into a plethora of PRR diversity, both in terms of gene families but also in terms of function [20-23]. In the case of self-defense, ligand interaction with the PRR induces signaling,

initiating the production of effector molecules such as antimicrobial peptides, suppressors of pathogen replication and phagocyte recruitment [24]. In parallel to the origin of PRRs, new immune cell lineages with killer functions evolved, adding to the repertoire of defense mechanisms seen in the earlier ancestor [20-22].

Overall, the evolutionary origin of vertebrates coincides with the appearance and expansion of some immune gene families, and the contraction or downfall of others. At a functional level, response cascades developed a higher degree of complexity — especially in the form of additional regulators — and immune cell lineages diverged to generate subsets with corresponding specialization and enhancement of function. Nevertheless, the vertebrate immune system displays strong parallels to the invertebrate immune system, such as the presence of PRRs and phagocytic cells. The vertebrate PRRs are considered to be of purely immune-related functionality in contrast to invertebrate PRRs, which are involved in both immunity and development. Moreover, the extreme *PRR* gene repertoire found in invertebrates is much more conservative in the vertebrate lineage [8, 16]. Thus, PRRs illustrate a nice example on how components of the invertebrate immune system have evolved further in vertebrate lineage. Similar findings related to immune-gene families has also been reported for lineages originating later — e.g. bony and cartilaginous fish, amphibians, reptiles, birds and mammals — demonstrating the ever continuous evolution of the immune system [16].

The above mentioned immune defenses are all considered part of an organisms' germline encoded innate immune system. With the appearance of jawless and jawed vertebrates, immune systems capable of adapting their receptors towards a specific pathogen evolved. This has been termed adaptive, or acquired, immunity [25]. Adaptive immunity was first described in jawed vertebrates (here called conventional adaptive immunity). This system consists of a set of receptors of

which some are somatically recombined and undergo random mutation to increase receptor diversity. Also, with this system came the ability to continuously improve receptor affinity throughout the immune response. Related to this, the ability to establish immunological memory by differentiating certain immune cells into long-lived memory cells bearing high affinity receptors for previously encountered pathogen evolved [25]. It was long believed that adaptive immunity was a trait only found in jawed vertebrates, but a functionally analogous system was later discovered in jawless vertebrates. Similarly to jawed vertebrates, adaptive immunity is obtained through receptors and cell lineages, but the genetic components are of a different origin, and they use different underlying mechanisms to increase receptor diversity [26].

The teleost immune system

Most of what we know about vertebrate immunity has been obtained from studies of mammalian species, especially in mice and humans. In this era of high-throughput sequencing, new genome resources from non-model teleost species demonstrate great diversity in immunological strategies deviating from the norm of the mammalian immune system [10, 27-32]. Below, I will briefly present the overall basis of the teleost immune system highlighting differences between mammalian strategies and those found in teleosts studied to date. Throughout, I will use the terms innate (germline encoded components) and adaptive (somatically altered components), which will aid in the presentation of such a complex system. However, this segregation does not reflect a clear boundary and there is significant cross-talk between the two.

Cells of the immune system

Although this thesis will bear a substantial genetic focus a brief presentation of the most common immune cell lineages is in order. The various cell types are of myeloid or lymphoid origin and can generally be ascribed to either innate or

adaptive immunity — the majority being innate (Figure 1). Cells of the adaptive immune system consist of B and T lymphocytes (B- and T-cells) where the T-cells appear in two sub-lineages: the cytotoxic T-cells, which are CD8+ and the helper T-cells, which are CD4+. The CD4+ T-cells can be further divided into subsets such as Th1, Th2, Th17 and Tregs. There is continuous discovery of new cell lineages related to immunity and recently innate lymphoid cells (ILCs), and natural killer T-cells (NKT) were reported, and the repertoire continues to expand [33-35].

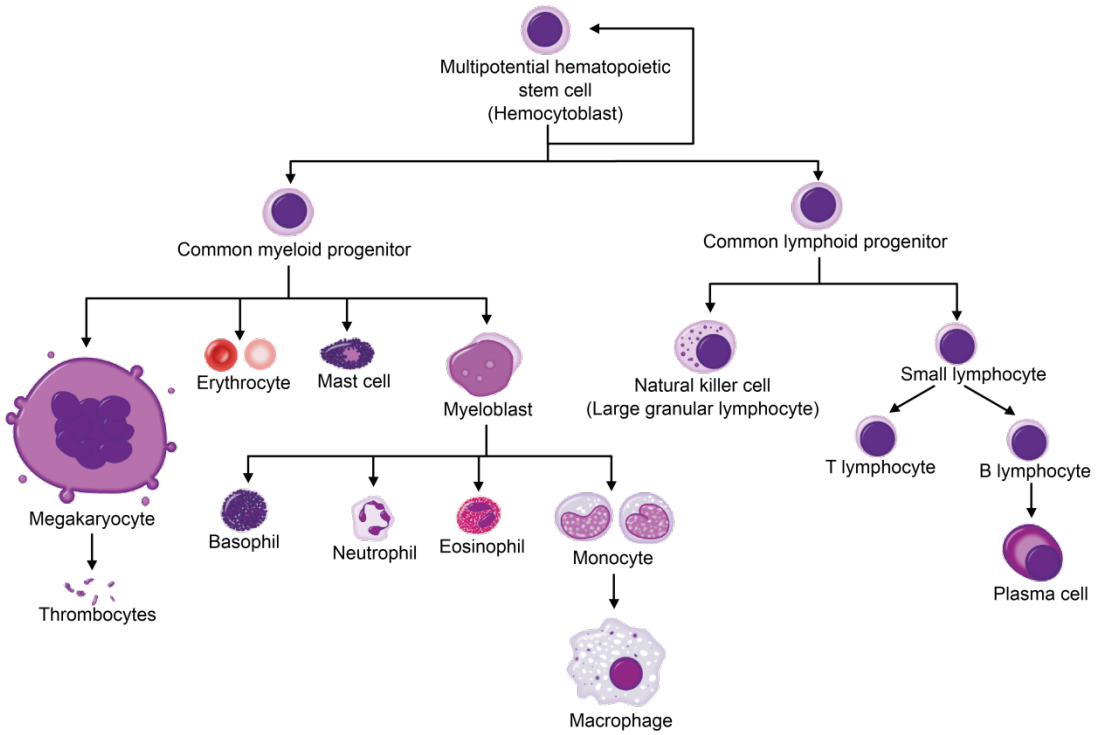


Figure 1 General overview of ontogeny leading to the most common cell lineages of the mammalian immune system showing both the myeloid and lymphoid lineages.

Neutrophils circulate the peripheral vascular system and upon activation they increase the expression of surface receptors to enable their migration to sites of inflammation. Upon reaching their destination they initiate respiratory burst releasing the contents of their granules causing bacterial destruction, but also collateral tissue damage. Macrophages and dendritic cells function as antigen

presenters, also called professional antigen presenting cells (APCs) together with B-cells, to the adaptive immune system (see section about adaptive immunity). They are potent phagocytes, especially macrophages, and the presented antigens are derived from the engulfed material. Furthermore, they also secrete initiators and mediators of inflammation. Platelets are crucial for the initiation of coagulation aiding the healing of wounds. They express several pathogen-detecting receptors and interact with neutrophils, monocytes and lymphocytes. Both eosinophils and erythrocytes are regulators of immunity secreting regulatory signal mediators called cytokines and chemokines (also see cytokines and chemokines). On the lymphoid side NK and NKT cells are cytotoxic effectors secreting perforins and granzymes, but they also display regulatory functions through the release of cytokines [34]. The ILCs are activated by stress signals, microbial compounds and cytokines. They are early effectors of the early immune response and are highly reactive. They have counterparts within the CD4+ T-cell lineages of the adaptive immune system, but do not express the T-cell specific receptors. Instead they act as regulators responding to the same targets as the corresponding T-cells [35]. The CD4+ T-cells are helpers and regulators, which can be separated by their cytokine expression profiles. Th1 aids in the elimination of intracellular pathogen through cell-mediated immunity by activating phagocytes and cytotoxic CD8+ T-cells. Th2 aids with the elimination of extracellular parasites stimulating a humoral (secreted) response from B-cells in addition to a range of regulatory effects. Th17 aids against extracellular bacteria and fungi, but mainly at mucosal surfaces. It is also involved in autoimmunity. Treg is a master regulator of immune responses capable of inducing tolerance towards self, but also to foreign antigens and thus protects against any immunopathology. B-cells mediate humoral adaptive immunity and carry antigen receptors, which upon activation are secreted into the extracellular environment [36].

Teleosts do not have bone marrow (myeloid lineage) and thymus (lymphoid lineage) like mammals, but tissue equivalents have been described. Still, many of the cell lineages connected to immunity in mammals have been found in the teleost immune system displaying similar functionality; neutrophils, macrophages, mast cells, dendritic cells, thrombocytes, B-cells, CD4+ and CD8+ T-cells and Tregs [37-43]. An ongoing debate is the polarization of cellular responses in teleost and if they establish typical Th1 and Th2 responses similar to mammals or if the polarize innate cells like macrophages instead. [44].

Innate immunity

Antimicrobial peptides

Antimicrobial proteins (AMPs) are small peptides found in the plasma and mucus of vertebrates. They display antimicrobial, antifungal and antiparasitic properties, which they apply through various mechanisms such as pathogen lysis and interfering with pathogen DNA replication. The most conserved antimicrobial peptides are defensins, cathelicidins, hepcidin and lysozyme [45, 46]. The antimicrobial peptide repertoire of teleosts consists of about 90 peptides distributed across five classes. These classes are similar to those of mammals, but there are also some teleost-specific lineages such as the piscidins. Overall, the functionality of teleost AMPs resemble the function of mammalian homologs, but the peptides generated by teleost species appear adapted to the unique aquatic environments they inhabit as well as the pathogens encountered [6, 47, 48].

Acute-phase response

Acute-phase proteins are found in plasma and mucus, similar to the AMPs, where they play important roles in the early innate immune response. The main mammalian proteins are C-reactive protein (CRP), serum amyloid A and P (SAA, SAP), haptoglobin (HP), α1-acid glycoprotein (ORM2), α2-macroglobulin (A2M), ceruloplasmin (CP), fibrinogen (FG) and transferrin (TF). Their functionalities

range from pathogen recognition and pathogen clearance through immune cell attraction and regulation of inflammation to preventing collateral damage of reactive oxygen species. Most of the acute phase reactants found in mammals are present in teleosts and shown to share similar functionality [45]. However, some discrepancies are to be expected as exemplified by the suggested analogous CRP-functionality of SAP in some species, and the suggested presence of only one CRP-like gene in teleosts [49, 50]. Also, some differences in reactivity between homologs has been reported [45].

Pattern recognition receptors

The initial discrimination between self and non-self is in vertebrates generally performed by pattern recognition receptors (PRRs). Their overall functionality is to initiate signaling upon ligand interaction, where the ligand is a pathogen- or damage-associated molecular pattern (PAMP and DAMP, respectively), culminating in initiation of inflammation and acute phase response, phagocytosis, recruitment of lymphoid cells and establishment of communication with the adaptive immune system [45, 51]. In mammals there are five major classes of PRRs: the Toll-like receptors (TLRs), the C-type lectin receptors (CLRs), the nucleotide-binding oligomerization domain (NOD) like receptors (NLRs), the retinoic acid-inducible (RIG) like receptors (RLRs) and the absent in melanoma (AIM) like receptors (ALRs). Collectively, they monitor both the intracellular and extracellular environments and function either as membrane associated receptors (TLRs and CLRs) or soluble receptors (NLRs, RLRs and ALRs) [overview of mammalian PRRs presented in 52]. Several classes of PRRs have been characterized in teleosts such as the TLRs, NLRs, RLRs and CLRs. Those proteins that are homologous to a mammalian counterpart generally display similar function and downstream effects. However, for some classes of PRRs the gene repertoire of teleosts is more diverse compared to mammals [48]. One clear example is the TLRs where TLR1-13 has been characterized in mammals contrary

to TLR1-26 reported in teleost (with the exception of TLR6 and TLR20) [53]. Furthermore, some of these display distinct characteristics such as soluble variants, alternate exon-intron structures and dissimilar ligand profiles compared to the mammalian version [48, 54]. Other examples are the additional lineages of NLRs and the diverse repertoire of lectin receptors (CLRs included) found in teleosts [48, 55].

Cytokines and chemokines

The vertebrate immune system uses a range of molecules, small and large, to coordinate its efforts of which two major families are the cytokines and chemokines. Cytokines are small inducible proteins that regulate inflammation, recruit various cell types and promote cell differentiation, maturation and activation. Dependent on their genomic region, target receptors, signaling pathways and function they are divided into classes where the major ones are interferons (IFNs), interleukins (ILs), tumor necrosis factors (TNFs) and transforming growth factors (TGFs) [45]. In mammals there are three IFN classes: type I IFNs, type II and type III. All interferons are mainly involved in antiviral defenses, but type II is also involved in the regulation of Major Histocompatibility complex (MHC) protein expression (see "adaptive immunity"), stimulates phagocytosis, and inhibits cell growth and apoptosis. The interleukins display a range of functions from being potent initiators and regulators of inflammation, supporting differentiation of immune cells and recruiting neutrophils, macrophages and leukocytes. Tumor necrosis factor (TNF), the major member of the TNF family, is involved in leukocyte chemoattraction and macrophage stimulation. Finally, transforming growth factor beta (TGFB) representing the TGF family is highly immunomodulatory through its regulation of inflammation, immunosuppressive effect and induction of tolerance in addition to growth-related functions [45]. Teleosts harbor all major cytokines

families. However, they display additional gene paralogs and there are several lineage-specific expansions of some gene families [56].

Chemokines are, in contrast to cytokines, mainly of chemotactic functionality controlling cell migration and positioning during an immune response, but also throughout development. In addition they facilitate interaction between subsets of immune cells — one interaction being the interface between innate and adaptive immunity. In mammals there are about 50 chemokines, which are subdivided into four groups dependent on the positioning of the initial cysteine residue: XC, CC, CXC and CX3C. The about 20 corresponding cell-surface receptors are transmembrane G-protein coupled proteins displaying variable binding affinity and within ligand group promiscuity[the mammalian chemokine system is reviewed in 57]. In contrast to cytokines, chemokines in teleosts are a diverse group of genes. Many mammalian homologs are present, however, a range of teleost-specific chemokines have been described [the teleost chemokine system is reviewed in 7].

Inflammation and inflammasomes

Inflammation is the overall initial response of the innate immune system upon infection or tissue damage. Key to inflammation are cytokines, which are both pro- and anti-inflammatory. There are several well-known pro-inflammatory cytokines such as IL1, IL12, IL18, IL23 and TNF, but IL1 and TNF are among the more studied examples. IL1 is transcribed from two genes, *IL1A* and *IL1B*, where *IL1A* is constitutively expressed by most cells and responds to tissue damage in contrast to the tightly regulated *IL1B* expression mainly by myeloid immune cells. The initiation of *IL1B* expression is induced by PRRs interacting with ligand [58]. Both IL1A and IL1B are generated as inactive precursors and processing is required to generate IL1 biological activity. This processing is performed by the inflammasome — a cytosolic multiprotein complex, which is assembled upon infection/tissue damage and provides a secondary level of IL1 regulation. The

inflammasome contains a pattern recognition receptor, most often an NLR containing a pyrin or card domain (NLRP or NLRC) and a pro-caspase (caspase 1 or 11, CASP1/11). Upon inflammasome assembly the caspase matures and its proteolytic ability is activated so it can cleave pro-IL1B into its active form: IL1B. The inflammasome is also responsible for cleavingpro-IL18 [59, inflammasomes are reviewed in 60]. In addition to provide the host with active pro-inflammatory cytokines, the inflammasome can induce cell death (apoptosis or pyroptosis) as part of the defense mechanism. Apoptosis destroys the infected host cell whereas pyroptosis releases the DAMPs within the target cell which further act as non-cytokine initiators of inflammation [59, 60]. As inflammation is established IL1B recruits neutrophils and induce the differentiation of T-cells into the Th17 subset. IL18 on the other hand promotes inflammation through initiation of IFNG production inducing Th1 responses demonstrating the differences in response established by the different cytokines [58].

Natural antibodies

Natural antibodies (NAbs) are made by B-cells without any antigenic stimulation and are considered part of the humoral innate immune system in vertebrates. They reside in vertebrate plasma and consist mainly of the IgM isotype (also see "adaptive immune system"). They present restricted variability compared to antibodies generated upon antigenic stimulation, but react against foreign antigens and can, in addition to neutralization, activate clearance and lysis of the pathogen through the complement system (also see "complement") [45, 61]. NAbs are also considered a bridge between innate and adaptive immunity as they have been found to prime the mammalian adaptive immune system by presenting their bound antigens to immune cells lymph nodes [61].

Complement

The complement system consists of a large and complex network of proteins, both soluble and membrane bound, which through a cascade reaction can

respond toward non-self and damaged self. This is a highly regulated process as unwanted complement activation can cause significant collateral damage. Activation of complement results in the clearance and/or lysis of the target together with enhancement of inflammation. However, additional functionality of complement has been suggested as an important mechanism during pregnancy, nervous system development and host-graft interactions [62, 63]. Traditionally, the complement components are presented as the core of the complement cascade and are termed C1 through C9 (C1 is a complex and not a single protein). In mammals, three pathways can activate this cascade: classical, lectin and alternative. The classical pathway is initiated with an antibody-antigen complex (may be a NAb) and the C1 complex, which through cleavage of C2, C3 and C4 produces a C5 convertase. The lectin pathway accomplishes the same, however, with a C1-like complex consisting of a mannose-binding lectin (MBL) or ficolin bound to a carbohydrate and MBL-associated serine proteases (MASPs). The alternative pathway is activated in a slightly different manner with the spontaneously hydration of C3 to (C3(H2O)). This process is tightly regulated and with the help of several positive regulatory factors this pathway generates an alternative C5 convertase. All three activation mechanisms converge at the terminal pathway, which through subsequent cleavage of C5, C6, C7, C8 and C9, creates a membrane attack complex forming a pore in the target membrane. All the various cleavage steps, as well as the membrane attack complex itself, generate or stimulate the release of mediators of inflammation, cell proliferation and cell death through apoptosis. The membrane attack complex can also activate the inflammasome to further enhance overall inflammation [62]. The teleost complement system appears identical to that of mammals displaying all three activation pathways and downstream formation of the membrane attack complex. There are also some striking dissimilarities such as the subcomponents of the C1 complex, which in mammals consist of, among others, C1q, C1r and C1s whereas in teleosts discrimination of C1r and C1s has proven difficult. In teleosts, the

discovery of a gene equally similar to both *C2* and one of the complement regulatory components called *factor B*, indicate that both C2 and factor B functionality are covered by a single gene in teleosts. It also appears that teleosts display fewer regulators of complement activation compared to mammals and they all appear to be soluble factors contrary to the membrane-associated mammalian repertoire. Finally, the most striking difference is the additional isotypes of C3, C4, C5, C7, MBL, factor B and factor I (another regulator of complement activation) described in teleost species generated by gene duplications. The teleost C3 isoforms are among the best studied and display structural differences in catalytic sites, in hemolytic activity and in binding specificity to various targets. Further, the isotypes of the complement components display some tissue-dependent expression patterns as well as being present in additional tissues like head kidney, spleen, intestine, gill, brain and gonads in contrast to mainly serum in mammals. The overall expression of complement indicates a focus of innate defenses in tissues representing internal-external transition zones in teleost. Overall, it has been hypothesized that the complement diversification enables teleosts to enhance their innate immune recognition and the effector functionality of their complement system [48, 64].

The adaptive immune system

The conventional adaptive immune system associated with jawed vertebrates revolves around a complex interaction between host cells, APCs, B- and T-cells and their receptors. It consists of both of cellular and humoral components coupled to cellular cytotoxicity and the generation of antigen-specific antibodies. Both establish long-lasting immunological memory enabling a rapid and specific response in the case of a pathogen reencounter [65-67] In the case of host cells, their antigen-presenting receptors (*MHCI*, see "The MHC complexes") form an immunological synapse with complementary T-cell receptors (TCRs) on CD8+ T-cells. If all co-stimulatory signals are present the T-cell is activated, proliferates

and initiates its cytotoxic effector function killing host cells with matching MHCI-antigen complexes presented [68]. The most common activation of a B-cell is through CD4+ T-cell help. The antigen can be presented by an APC or the B-cell can function as the APC itself. If all co-stimulatory signals are present, the B-cell and T-cell will differentiate and proliferate into several cell subsets consisting of both short-lived and long-lived cells such as Th T-cell subsets, antibody-producing B cells, plasma cells producing large amounts of affinity matured antibodies and long-lived memory B- and T-cells [69] (see figure2 for outline of the MHC-TCR-BCR interaction).

The MHC complexes

The Major Histocompatibility Complexes (MHCs), class I and class II are proteins that generally present antigens from the intracellular or extracellular environment to CD8+ and CD4+ T-cells, respectively. Some interlinking has been observed in the form of cross-presentation where MHCI molecules are loaded with exogenous antigens. Similarly, endogenous antigens can be loaded onto MHCII if they are degraded through autophagy [65].

MHCI is expressed by all nucleated cells, which together with beta-2-microglobulin (B2M) presents antigens mainly generated by proteasomes. The aim of the MHCI pathway is to report any intracellular infection to minimize further infection of neighboring cells. In humans, the MHCI region contains six loci; *HLA-A*, *B*, *C*, *E*, *F* and *G* where *HLA-A*, *B* and *C* are highly polymorphic. These polymorphism generate a range of slightly different antigen binding grooves to enable binding of a broad antigen repertoire [65]. The human MHCII region also contains six loci; *HLA-DR*, *DQ*, *DP*, *DM*, *DOA* and *DOB* where *HLA-DR*, *DQ* and *DP* are highly polymorphic, *HLA-DOA* and *DOB* are less polymorphic and HLA-DM is a chaperone active in the peptide loading process. In contrast to MHCI interacting with B2M, MHCII is a heterodimeric receptor made up by two MHCII peptide chains. Similar to MHCI, the polymorphic

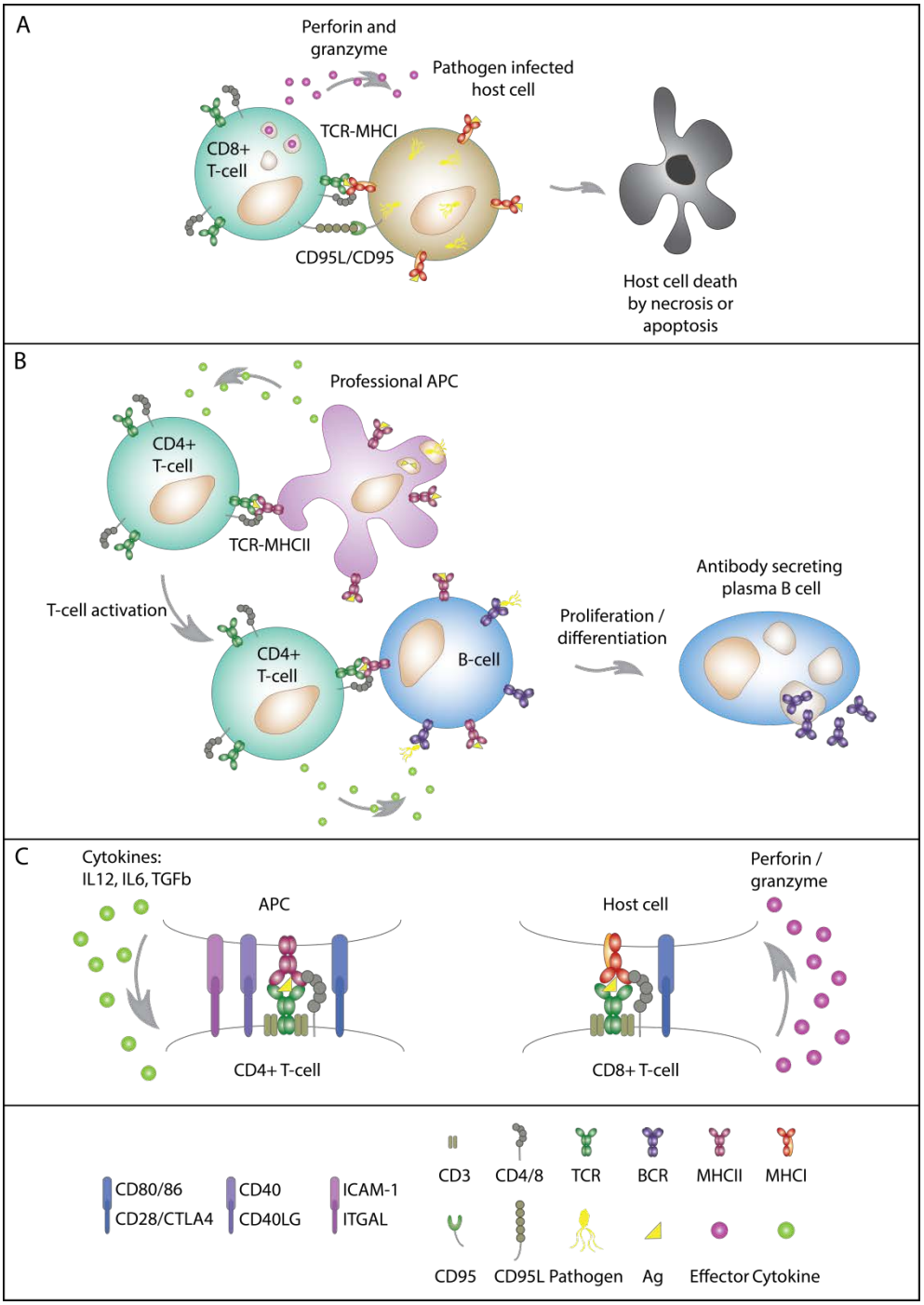


Figure 2 Basic outlines of the cell and protein interactions laying the foundation for conventional adaptive immunity in jawed vertebrates. A) The interaction between host cell displaying foreign antigen on MHCI and T-cell with antigen-complementary TCR. Through CD95 or the release of enzymes the host cell dies. B) One of the ways of activating a B-cell is through CD4+ T-cell help, and the interactions are between MHCII, TCR and BCR. The proliferation and differentiation of the full panel of T- and B-cell subsets is not depicted. C) The immunological synapse formed between APCs/CD4+ T-cells and host cells/CD8+ T-cells, respectively. Only the core set of interacting molecules within the synapses are depicted.

nature of MHCII provides a range of slightly different antigen binding grooves enabling them to present an array of different antigens. MHCII is mainly expressed on cells with APC functionality such as dendritic cells (DCs), macrophages and B-cells, and present peptides obtained through degradation of exogenous material in the endocytic pathway [65, 70].

MHC molecules in teleosts have similar functionality to those of the mammalian system [48]. In teleosts, *MHCI* is found in five different classes: U, Z, S, L and P where class U contains those genes designated as classical MHCI. For *MHCII*, there appears to be two lineages (A and E), possibly three (B) where lineage A contains those appearing to have classical MHCII functionality. Interestingly, the number of genes in the MHCI and MHCII clusters varies greatly within the teleost lineage, which further may be attributed to the unlinked nature of teleost MHC clusters – in contrast to mammals – and their different divergence rates [70, 71].

B-cell and T-cell receptors

The MHC complexes present antigens that are both self and non-self. It is up to the T-cells to determine if the antigen is of a harmful origin. The T-cells display TCRs used to screen the antigen-presenting MHCs. The TCR gene locus in humans consists of a series of disconnected gene segments combined by RAG recombinases in a process called V(D)J-recombination (variable, diversity and joining type segments). Collectively, this can give rise to an astonishing amount of different TCRs. The repertoire is increased further by the insertion of non-encoded random nucleotides in between each joined segment as well as variable exonuclease trimming. Each T-cell will generate two different TCR chains, which are combined into the finished heterodimeric TCR. This TCR is tested for self-reactivity in the human thymus where there is positive selection of those who have weak affinity for the human MHC. Subsequently, there is second round of negative selection removing TCRs with too strong affinity towards MHC. These

two steps are measures taken to minimize potential self-reactivity. Only T-cells with a TCR fulfilling both these demands are released from the thymus [68]. The TCR complex in teleosts is recombined similarly to that of mammals and the loci are overall organized similarly to that of mammals. However, specific regions (CDR loops and connecting regions) within the TCR chains are longer or shorter compared to mammals. For CDR loops this affects how the TCR interacts with its corresponding MHC whereas for connecting regions it affects disulfide bridges. It also appears that some TCRs are expressed in different isoforms, that others have tissue-specific expression patterns and that some teleosts may have additional constant segments in their TCR loci [72].

The B-cell receptor (BCR, immunoglobulin receptor, which in soluble form also are called antibodies) present on B-cells is also encoded in a multi gene segment locus where the segments are combined with V(D)J-recombination. The BCR consists of two heavy chains, usually of the Mu or Delta isotype (IgM, IgD) and two light chains (kappa or lambda). Also, like the TCR, BCRs have to be screened for any self-reactive combinations and clonal deletion is performed in the bone marrow. The diversity of BCRs is even further increased through the process of somatic hypermutation where random mutations are inserted into the regions that will later interact with antigen [69, 73].

Mammals have five BCR isotypes: IgM, IgG, IgE, IgA and IgD. In teleosts there are three identified: IgM, IgD and IgT (also called IgZ) [74]. This change in diversity between teleosts and mammals is in contrast to the more conserved TCR loci in vertebrates [72]. Teleost IgM is similar in structure and function to the mammalian IgM. IgT is likely a functional analog to human IgA by protecting mucosal surfaces together with IgD. Mammals have two light chains, kappa and lambda, whereas teleosts have four (nomenclature varies) L1, L2, L3 and lambda. L1 and L3 have been suggested as kappa homologs and L2 is homologous to the amphibian sigma chain. In addition to having another repertoire of

immunoglobulin classes compared to mammals, teleost genomes also contain some of the largest immunoglobulin loci described with up to several hundred available variable regions available for V(D)J-recombination [74]. Teleosts do not display isotype switch of their immunoglobulins after initiation of adaptive immunity. Also, their overall affinity maturation of immunoglobulins is much lower than in mammals. However, this has been related to the lack of secondary lymphoid organs with similar organization to the mammalian lymph node, which enables efficient cell interaction and immunoglobulin maturation [5, 74]

Discoveries leading to new immunological paradigms

The human immune system has been featured as the endpoint of immune-related evolution displaying the most sophisticated immune system. As stated earlier, genome sequencing is challenging many of the well-established assumptions in immunological research. Below, I will briefly present some of the more groundbreaking findings in immunology.

Innate memory

Innate immunity has for long been considered a generic, or at the most a semi-specific, host defense system. There is now accumulating evidence for an innate memory mechanism, also called trained immunity, where, upon a reencounter with an identical or heterologous pathogen, the innate immune system presents a heightened response compared to a primary pathogen encounter. This has also been observed in connection with vaccination where the vaccine enables the host to develop a more general protection against similar pathogens to that/those the vaccine was developed to protect against [75]. The detailed molecular establishment of innate memory is largely unknown other than it generally involves cells of the innate immune system such as macrophages, NK-cells and innate lymphoid cells. However, epigenetic reprogramming has been suggested as a factor [76].

Adaptability

Conventional adaptive immunity has been assumed to be the only adaptive immune system found in vertebrates and then only in those with jaws. Other organisms have been assumed heavily dependent on their innate immune systems. Expanding the term “adaptability” and moving away from gene homology has revealed a plethora of adapting immune systems — even in prokaryotes with their CRISPR/cas system (clustered regularly interspaced short palindromic repeats CRISPR and CRISPR-associated cas) [18, 77, 78]. Similar findings have also been seen in invertebrates and protochordates where RNA processing generates thousands of variants of DSCAM (Down syndrome cell adhesion molecule) in Drosophila [79] and the VCBP (variable region-containing chitin-binding protein) in protochordates maintaining the commensal microbiota in the intestine similar to the adaptive immune system in mammals [80]. However, the most striking unconventional adaptive immune system has been found in jawless vertebrates showing analogous function to jawed vertebrate lymphoid cells and BCRs/TCRs. The genetic basis for developing antibody secreting B-cell lineages is present in jawless vertebrate genomes. There is also genetic basis for the development of T-cells to a certain stage of maturation where MHC is required — a genetic locus not found in jawless vertebrates [81]. Instead they have variable lymphocyte receptors (VLRs) consisting of leucine-rich repeats (LRRs). The "adaptability" is achieved by combining variable LRRs resulting in a diverse VLR repertoire analogous to the antigen receptor (BCR/TCR) repertoire in jawed vertebrates. [82-85].

It has been assumed that all jawed vertebrates have the MHC-TCR-BCR system of adaptive immunity and that alteration, or lack of such, within this lineage hardly is compatible with life. In 2013, when the coelacanth (lobe-finned fish) genome project was published, the loss of IgM — otherwise found in all investigated vertebrates — was documented. In contrast, it was found to have

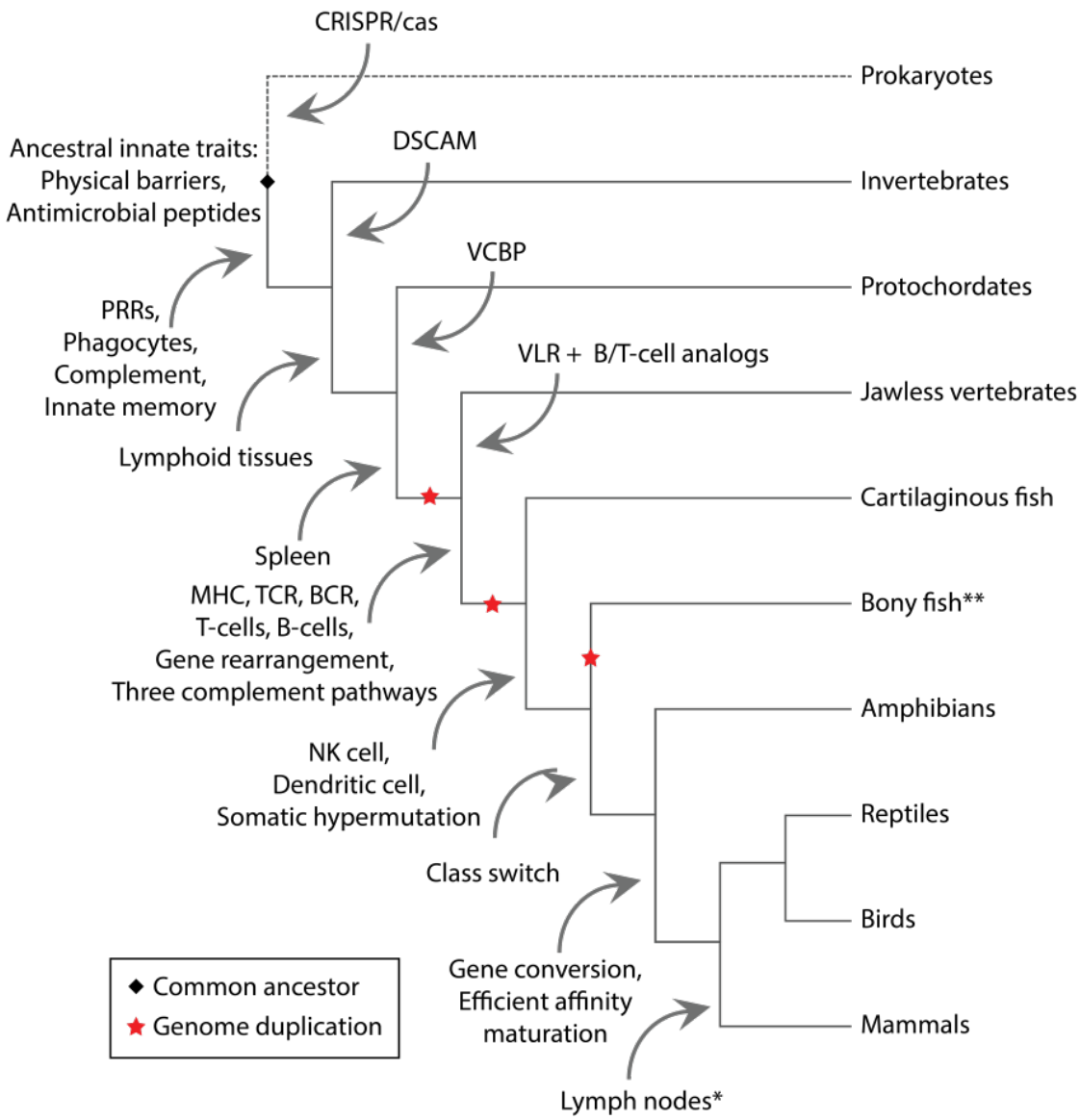


Figure 3 The evolution of innate and adaptive immunity with emphasis in the vertebrate lineage including unconventional adaptive mechanisms reported. Traits are mapped upon a simple phylogenetic cartoon together with the commonly accepted vertebrate genome duplications. The various immunological traits depicted are described in [16, 45, 78-80, 86]. *Reptiles and birds display the primordial form of lymph nodes. **The order of Gadiformes has lost *MHCII* and related factors [12].

immunoglobulins similar to those of cartilaginous fish [4]. Moreover, the genome sequencing of elephant shark discovered the lack of *CD4* and *MHCII* with no CD4 interaction domain. In addition, a range of cytokine receptor ligands were found to be missing [4]. As already mentioned, sequencing of the Atlantic cod genome revealed the first evidence of a jawed vertebrate lacking the MHCII

pathway [10] confirming the hypothesized loss of *MHCII* made by Pilström in 2005 [87]. The ability of jawed vertebrate species to survive without *MHCII* has further been supported in a distantly related species — the pipefish (*Syngnathus typhle*) — by a functional loss of *MHCII*, i.e. not detected in the transcriptome [88]. Recently, Malmstrøm et al. found that the loss of *MHCII* is a trait common for the entire Gadiformes (cod-like fish) lineage completely refuting the jawed vertebrate-conventional adaptive immunity connection [12].

Combining the classic knowledge with the newer information provided the last decade or so through genome sequencing the new status quo indicates innate immunity in all living organisms, but possibly a form of adaptive immunity as well (Figure 3). It further demonstrates an increased need of investigating non-model species, both genetically and functionally, to thoroughly reveal how the vertebrate immune system is organized and how it applies its plethora of defense mechanisms.

Aims

The overarching aim of this thesis was to obtain a better understanding of the immune system in Atlantic cod and other Gadiformes — especially due to the loss of *MHCII*. By taking advantage of state of the art high-throughput sequencing technology, high performance computing clusters and bioinformatical tools this thesis addresses descriptive, evolutionary and functional aspects of the Atlantic cod immune system emphasizing the innate defenses. We aimed at fully characterizing the PRR families in Atlantic cod through a comparative approach to infer function, search for signs of ongoing selection and to compare the Atlantic cod *PRR* repertoire to those of other reference genome sequenced teleosts (Paper I). Considering the unconventional immunological strategy of Atlantic cod we aimed at comparing Atlantic cod to other teleost. With emphasis on the cod-like lineage of Gadiformes, genome information of 66 new teleost species and corresponding species phylogeny enabled the characterization of teleost *PRR* repertoires. Furthermore, we aimed at correlating these repertoires towards the loss of *MHCII* and abiotic factors such as species depth and species latitudinal distribution — presented in both a contemporary and more ancient setting (Paper II). In paper III we wanted to investigate the evolutionary pattern of long sought-after innate immune gene in Atlantic cod using the 66 teleost genomes. In paper IV, by whole transcriptome sequencing (RNAseq) of Atlantic cod individuals infected with a bacterial pathogen, we aimed at characterizing the overall innate response as well as the transition into adaptive immunity. In paper V, using the same methodological strategy as in paper IV, we aimed at characterizing the response of Atlantic cod towards an immersion vaccine. This vaccine has earlier been shown to establish increased resistance towards the pathogen and thus we aimed at uncovering the underlying transcriptional mechanism establishing immunological memory.

Summaries of paper I-V

Paper I

Evolutionary redesign of the Atlantic cod (*Gadus morhua* L.) Toll-like receptor repertoire by gene losses and expansions

Scientific Reports 6, Article number: 25211 (2016) doi:10.1038/srep25211

Paper I is a continued analysis of the findings reported with the genome sequencing of Atlantic cod [10]. Here, we further characterized the *TLR* gene repertoire in Atlantic cod and presented our findings in a comparative and evolutionary setting. We established loss of *TLR1/6*, *TLR2* and *TLR5* in parallel with gene expansion of *TLR7*, *TLR8*, *TLR9*, *TLR22* and *TLR25* using local gene synteny analyses. Upon closer investigation of the expanded *TLRs* we found several sites under diversifying selection. Mapping these sites onto modeled protein structures of Atlantic cod *TLRs* demonstrate a distribution of sites concentrated in the ecto-domain in regions assumed to be involved in TLR dimerization and/or TLR-ligand interaction. These findings indicate that Atlantic cod increases its detectable ligand repertoire through neo- and subfunctionalization. Using RNAseq we looked at the gene expression patterns of all Atlantic cod *TLRs*. We found patterns indicative of both tissue-specific and developmental stage- specific expression. Finally, by using the mammalian TLR signaling pathway as a reference, we characterized all homologous pathway genes in Atlantic cod. We found it likely that the Atlantic cod TLR repertoire uses a similar pathway for downstream TLR signaling and thus have comparable functionality. This was further supported by the presence of conserved endolysosomal sorting signals across all investigated species in TLRs in need of processing to become functional Looking at our findings in a broader perspective, using a comprehensive vertebrate *TLR* phylogeny with representatives from mammals, birds, reptiles, amphibians, teleosts, non-teleosts and jawless

vertebrates, we found that the Atlantic cod *TLR* repertoire is extreme with respect to gene losses and expansions. We also describe a shift in *TLR* repertoires in the evolutionary transition from aquatic (teleost, non-teleost and jawless vertebrates) to terrestrial (amphibians, reptiles, birds and mammals) lifestyle. This change is evident in that different members of each *TLR*-family is used by different classes of species and that, contrary to earlier assumed, some *TLR*-families are not represented in all species classes. Collectively, our findings in this paper provide insight into the function and evolution of *TLRs* in Atlantic cod, but also the evolutionary history of vertebrate innate immunity.

Paper II

Unveiling the evolution of the teleost innate immune system

Manuscript submitted

Recently, a study by Malmstrøm et al. demonstrated that all Gadiformes have lost *MHCII* similar to that of Atlantic cod using genome sequences from 66 new teleost species in addition to 11 available reference genomes [12]. Moreover, in paper I we found that the teleost innate immune system has demonstrated a different set of *Toll-like receptors (TLR)* compared to other vertebrates and that Atlantic cod harbors an extreme variant of the teleost TLR repertoire. In paper II we characterize the *TLR* repertoire of the 76 teleosts used in Malmstrøm et al. aiming at revealing the underlying selective mechanisms driving the variety of immunological strategies observed in teleosts and why they arose (Figure4). We also wanted to investigate the possible link between the loss of *MHCII*, past environmental conditions and the genetic architecture of the innate immune system. In paper II, we show that the teleost *TLR* repertoire contains an array of lineage-specific losses and expansions, with the Gadiformes lineage as an extreme outlier. Interestingly, within the Gadiformes we discovered expansions of *TLR* genes to be correlated with the loss of *MHCII,* whereas *TLR* copy number

variation correlated with species latitudinal distribution in teleosts overall. This suggests that there is a strong on-going selection of the innate immune system linked to specific environmental factors. Furthermore, timing of the lineage-specific losses overlaps with well-described changes in paleoclimate and continental drift, and hence unveils past adaptive signatures driving the genetic change within the teleost immune system. Our study reveals a remarkable evolutionary flexibility of teleost innate immunity, which has played an essential role in the survival and radiation of the teleost lineage.

Paper III

Successive losses of central immune genes characterize the Gadiformes' alternate immunity

Manuscript submitted

Studies on the mammalian immune system are what dominate our understanding of the vertebrate immune system. Genome sequencing of non-model vertebrates has revealed genetic diversity that surpasses mammalian diversification. Teleosts in particular have been found to harbor gene families not found in mammals [45], but more importantly some teleosts have lost immune genes earlier assumed to be required for vertebrate survival [12]. In paper III, we show that genes central to the innate mammalian immune system are lost from the immune gene repertoire of teleosts predating the loss of key adaptive components in codfishes (Figure4). In detail demonstrate that the innate Myxovirus resistance gene (*Mx*) is lost from the ancestor of Gadiformes and the closely related *Stylephorus chordatus*, thus predating the loss of *Major Histocompatibility Complex class II* in Gadiformes. Although the functional implication of *Mx* loss is still unknown, we demonstrate that this loss is one of several ancient events appearing in successive order throughout the evolution of teleost immunity. In particular, we find that the loss of *Toll-like receptor 5*predates

the loss of *Mx* involving the entire Paracanthopterygii lineage. Using a time-calibrated phylogeny we show that these losses overlap with major paleoclimatic and geological events indicating adaptive losses promoting survival and speciation in environments where maintaining these genes was less favorable. From a paleoclimatic and geographic viewpoint these dramatic immunological changes suggest that major events in earth's history were important catalysts in shaping the teleost immune system. We conclude that the observed gene losses are linked to historic environmental changes causing scenarios where maintaining these genes was less favorable.

Paper IV

Disentangling the immune response and host-pathogen interactions in *Francisella noatunensis* infected Atlantic cod

Manuscript

The immune gene repertoire of Atlantic cod deviates from other genome sequenced teleosts as well as vertebrates. So far, no experimental immunological studies have been able to fully deduce its functionality. In this study we, by full transcriptome profiling, investigate the overall immune response of Atlantic cod, but also the host-pathogen interaction in Atlantic cod infected with *Francisella noatunensis*. This pathogen is a gram-negative facultative intracellular bacterium, mainly infecting macrophages, causing the severe disease francisellosis in wild and farmed fish species worldwide. We discovered that Atlantic cod displays an overall classic initiation of immunity with inflammation, acute phase response and cell recruitment. Further, we found that *Francisella noatunensis* alters the immune response in Atlantic cod similar to that seen in other teleosts, but also similar to the mammalian equivalent tularemia. In Atlantic cod the affected pathways involve iron homeostasis, phagosome and autophagosome formation, oxidative burst and apoptosis. Looking closer at the transition between innate

and adaptive immunity we found an extensive up-regulation of *MHCI*. Our results indicate that they are likely to present endogenous as well as exogenous antigen with corresponding cytotoxic cellular responses. Finally, our results indicate T-cell independent B-cell activation with the help of TLRs and possibly also with help from neutrophils and NK-cells. Collectively, this study provides further insight into the host gene expression patterns underlying francisellosis and novel functional insight into the orchestration of the Atlantic cod immune response.

Paper V

Whole transcriptome analysis of the Atlantic cod vaccine response reveals no conventional adaptive immunity

Manuscript

Genome sequencing demonstrated that Atlantic cod lacks *MHCII*, which is central for presenting antigen to the adaptive immune system. In functional studies, Atlantic cod appears to establish an adaptive response towards pathogens and protection post vaccination indicative of adaptive mechanisms. Here we investigate the immunological response of Atlantic cod using whole transcriptome sequencing characterizing the transcriptional response towards a *Vibrio anguillarum* vaccine. We used siblings from an Atlantic cod breeding stock found to be highly susceptible towards vibriosis and where vaccination gave rise to increased pathogen resistance. In-depth gene expression analysis at 2, 4, 21 and 42 days post vaccination was conducted. We found that the innate responses are more or less absent and found few differentially expressed genes related to conventional adaptive immunity. However, there is a strong response from non-immune related pathways involving muscle and neuron development as well as from range of metabolic pathways. These findings are in line with earlier reports demonstrating changes in muscle growth and increased neuron development

post vaccination. Moreover, the up-regulation of metabolism-related pathways demonstrates a shift towards glycolysis, which has in earlier studies been linked to the development of innate memory. The lack of a clear transcriptomic component combined with other functional studies demonstrating significant memory responses in Atlantic cod indicate the usage of an unknown mechanism establishing immunological memory. Likely candidates are CD8+ memory T-cells, memory B-cells activated through T-cell independent mechanisms, innate memory induced through NK-cells or shift in metabolic strategy maintaining epigenetic changes.

Discussion

Drivers of teleost immune system evolution – host interactions

The progressive changes seen in the immune system during the evolution of the vertebrate lineage can easily be linked to the vertebrate genome duplication events (Figure 4). However, within an ecosystem setting, all is interconnected and the likelihood of genome duplications being the sole evolutionary driver is small. More than 150 years ago, Darwin addressed how evolution through natural selection, as responses to change in biotic and abiotic environments, influences the biodiversity on a geological time-scale [89]. Today, we know that natural selection in response to environment is a key driver in genomic diversification. Host immunity is readily affected by factors such as nutrient availability, temperature, pathogen load and diversity, other host-intrinsic fitness-related systems and intra-species co-evolution of genes. Thus, if these factors affect genetic components providing a change in fitness for a host subpopulation they will enable adaptation of the immune system [90]. It is within this framework I will present the findings in paper I-III and touch upon paper IV and V.

Host intrinsic factors: the *MHC-TLR* interaction

Interacting partners in immunity, be it innate versus adaptive immunity or members of the same gene families, affect the evolution of each other as they rapidly co-evolve in their fight against pathogen [25, 91, 92]. In paper II we demonstrated the correlation between *TLR* expansions and the loss of *MHCII*. This provides a nice example of host intrinsic factors that likely affect the continued evolution of the host immune system. This interaction is also evident for the innate and adaptive immune system as a whole. In the vertebrate lineage there is decreased diversity related to the innate immune system appearing in reverse-parallel to an increasing level of regulation and nuances in the adaptive immune system [25, 91, 92].

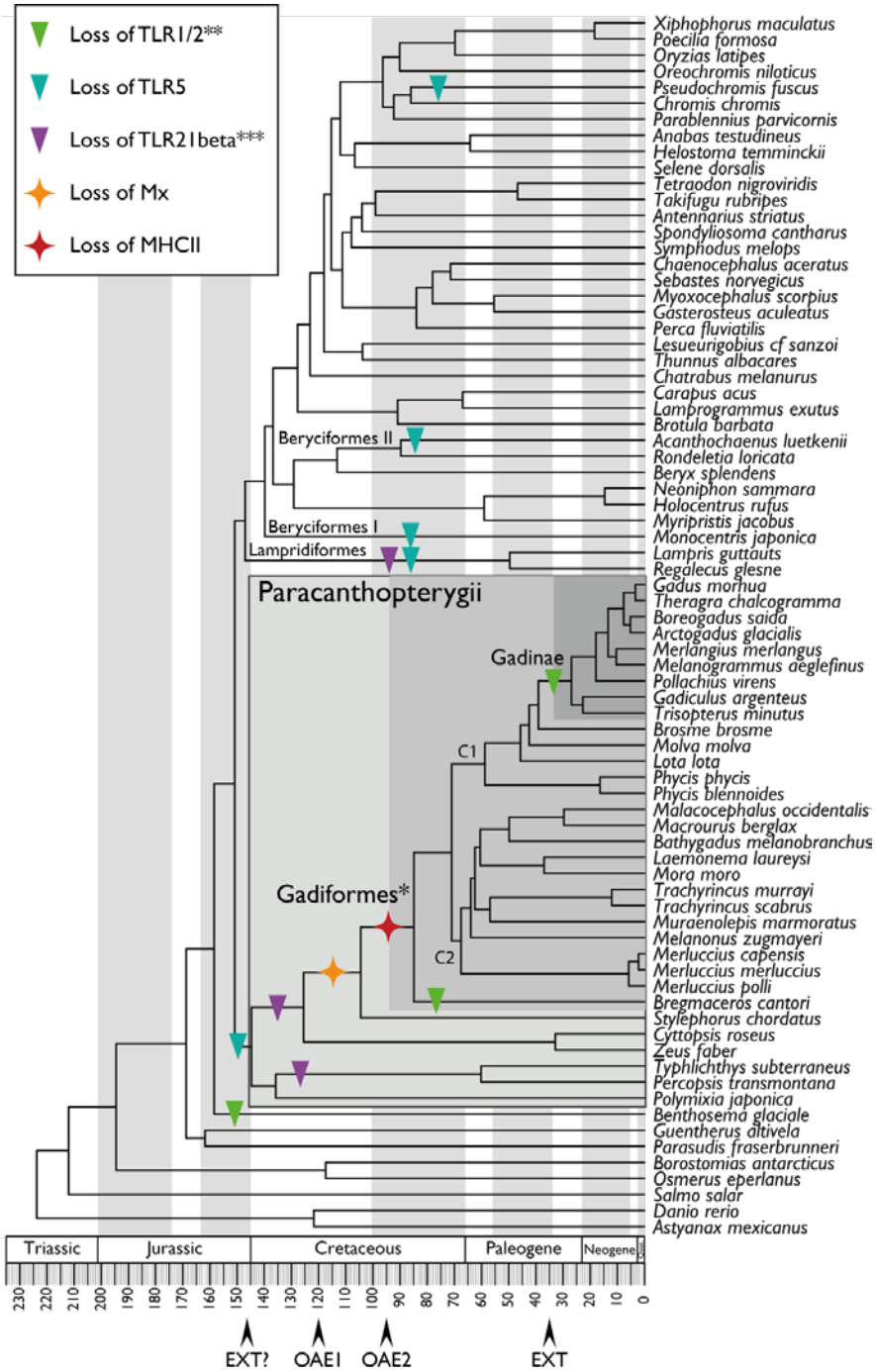


Figure 4 The loss of *MHCII*, *TLRs* (paper II) and *Mx* (paper III) mapped onto the timed phylogeny generated by Malmstrøm et al [12]. OAE: global oceanic anoxia events 1 and 2 [93]. EXT: extinction event with cooling and high eustatic sea levels [94]. EXT?: likely extinction event [95-99].* The Gadiformes lineage display significant expansion of *TLRs* 7, 8, 9, 22 and 25 with high, but variable, copy numbers compared to the smaller expansions outside Gadiformes. **For two species there is loss of either *TLR1* or *TLR2* not depicted here. ***There are some species-specific losses of *TLR21beta* outside the Paracanthopterygiian lineage not depicted here (see paper II).

Host-biotic interactions: the commensal and the pathogen

The appearance of commensal microbiota has likely contributed greatly to the evolution of the vertebrate immune system [16, 17, 80]. With the evolution of more complex body plans in the earliest animals the prokaryotes evolved the ability to inhabit anatomical niches creating the commensals. In return for being a favorable niche the host obtains help with digesting complex carbohydrates and is provided with essential nutrients [100]. Further, commensals actively compete with pathogens for this niche [101]. However, for the immune system the commensals present a tricky problem. They are an enormous antigenic burden capable of activating the immune system, they may return to a pathogenic state and they carry the same molecular markers as pathogenic strains. In jawed vertebrates the adaptive immune system is able to discriminate between commensal and pathogen by reading the cues provided by the host microbiota and thus is able to initiate proper responses. It is this interaction that likely has generated the development of additional regulatory circuits in vertebrates able to induce ignorance and tolerance as well as immune responses leading to pathogen clearance [80, 81, 101, 102].

In terms of the host-pathogen interaction this directly affects the evolution of immunity due to the continuous race between host and pathogen adaptation. The host aims at resisting, detecting and removing any pathogens to promote its own survival. In contrast, the pathogen aims for efficient host entry, avoidance of detection and within-host replication. This result in a never-ending evolutionary struggle termed the host-pathogen arms race — a dynamic interaction readily described by the Red Queen evolutionary hypothesis [16, 103, 104]. Host pathogen antagonist genes are often under positive selection possibly enabling the development of new mechanisms for pathogen detection or clearance [103, 104]. This is evident within several of the Atlantic cod *TLR* expansions, which displays signs of positive (diversifying) selection (Paper I). For the pathogen

there are similar selective pressures, but on the development of new mechanisms for host entry, survival and replication [103, 104]. However, the pathogen, due to shorter generation times and higher mutation rates, adapt quicker. Contrary, the host, with less genome size restrictions can utilize mechanisms like gene/genome duplications to provide new genetic material promoting faster adaptation of pathogen antagonist genes [103, 104]. For the vertebrate lineage the genome duplications have likely provided new genetic material promoting host adaptation. For bony fish in particular it likely reflect the extreme diversity of this lineage both with respect to life history strategies, that they inhabit the entire range of freshwater and marine habitats, but also the diversity revealed in their immune gene repertoires (Paper II and III) [1-3]. Finally, whereas the pathogen has a certain host or hosts to adapt towards, the host has to handle a range of possible pathogens either individually or as co-infections. As they likely utilize different approaches to attempt host entry and survival the host has to maintain a plethora of defense mechanisms to protect itself [90]. Thus, the host from a population point of view often displays large genetic diversity within the pathogen antagonist genes also reflected by large individual differences in the immune response [103, 104]. This could also be one of the reasons for the large gene expansions observed in Atlantic cod and Gadiformes. In addition we observe considerable variation in gene expression levels between the treated replicates in our RNAseq investigations (Paper IV and V, supplementary information) indicative of larger inter-individual variation. Diversity on both the genetic and functional (gene expression) level contributes to maintaining overall population pathogen resistance through balancing selection [103, 104].

Host-abiotic interactions: gene losses, climate and geography

Upholding the immune system comes at a cost because the host has to spend energy to maintain the genetic components as well as risking collateral damage to use it. In addition, the immune system is directly related to fitness and usage

will directly affect other fitness-related system such as reproduction [90]. We found in paper II and III that the loss of *MHCII* (first presented by Malmstrøm et al [12]), *Mx* and certain *TLRs* overlaps with global oceanic anoxia events, abrupt changes in temperature and other likely extinction events as well as changes in the layout of tectonic plates (Figure 4, Paper II and III). For *MHCII* and *Mx* the losses overlapped with global oceanic anoxia events. As the energetic costs increases in such environments [105] our findings support gene loss due to too large energetic demands for maintaining these systems [106]. For the losses of *TLRs* it is likely caused by changes in temperature combined with new available habitats (Paper II). Large environmental changes can alter the functionality of the immune system with respect to disease resistance. On a molecular level heat-stress has been correlated with reduced effect of phagocytosis, oxidative capacity and antibody synthesis. On a systemic level, environmental deviation affects the disease transmission due to changes in available pathogen vectors, effects of free-living pathogen stages as well as pathogen and host density [107]. Occasions of dramatic environmental changes provide the opportunity for an organism to diversify and geologically expand or in the worst case, may lead to extinction. Such extreme changes in environment tend to occur on rather short time scales, and thus it directly influences a population's possibility to adapt decided by the physiological tolerance of the individuals [107]. Furthermore, by geographical reallocation, the host will be exposed to new pathogens with no co-evolutionary past i.e. making the host highly susceptible to infection. Likely it will lead to decrease in overall host survival and maybe extinction unless the immune system is able to adapt [107].

Overall, our findings in paper I-III demonstrate signs of both biotic and abiotic interactions and beautifully illustrate the complexity of the system as we still have not clearly elucidated all underlying mechanisms driving the evolution of teleost immunity.

How does the immune system of Atlantic cod operate?

Paper I-III consisted mainly of genomic information analyzed with bioinformatical tools describing gene repertoires and presenting them within an evolutionary framework. In paper IV and V we applied RNAseq adding the first level of functional information onto this system. Below, I will present our overall conclusions on how the Atlantic cod, and likely Gadiformes, immune system is orchestrated. I will also present some alternative mechanisms, which were not well supported in each of the papers alone, but when viewed collectively present a somewhat clearer picture.

Overall, teleosts display more diversity related to their innate responses compared to mammals. This is evident when observing the increased diversity within innate gene families and in addition teleosts also harbor genes with no clear mammalian homologs adding to the diversity [6, 7, 45, 47, 48, 54-56, 64]. One can argue that Atlantic cod and other cod-like fish have an even stronger presence of innate immunity compared to teleosts harboring genetic components of both innate and conventional adaptive immunity. In support of this are the findings in paper I and II demonstrating the extreme repertoire of *TLRs* in species lacking *MHCII*. Functionally, in paper IV, we observe significant up-regulation of factors recruiting phagocytic cells and up-regulation of *PRRs*. In an evolutionary setting this indicates a more ancestral strategy similar to non-vertebrate species in Atlantic cod and likely in all Gadiformes. Furthermore, we also observe in paper IV a considerable contribution from the MHCI system, both classically as well as cross-presentation, with likely cytotoxic responses. This appears to be supported by T-cell independent activation of B-cells in response towards infection. We found no signs of conventional adaptive immunity in either paper IV or V which was expected due to the lack of *MHCII* and *CD4*. This further affects the possibility of CD4+ T-cell polarization. There is ample evidence for polarization of CD4+ T-cell subsets in mammals where the most common subsets

are Th1, Th2 and Th17. Their differentiation is driven by cytokines and transcription factors such as interleukins, the STAT family of transcription factors, TBET, GATA3 and more [36] of which we, in our data, find little response. The polarization of T-cells in teleosts overall has been debated, and there has been proposed polarization of macrophages instead creating inflammatory and tolerogenic M1 and M2 populations [44]. We find more factors related to this system (IFNG, MHCI, STATs, JUN, pro-inflammatory cytokines and more). However, this observation could be biased as the system is affected by the intracellular lifestyle of the pathogen preferring macrophages as their cellular host [108].

In paper V, where we describe the vaccination experiment, there was an overall lack of response from the immune system. Previous investigations have demonstrated protection post vaccination in this very system [109]. There could be ongoing differential expression at time-points that we did not sample throughout our experiment, but the overall absent response indicates otherwise. Thus, we considered alternative mechanisms that could potentially explain the increase in protection post vaccination in Atlantic cod. Contrary to paper IV, there was no significant evidence of T-cell independent B-cell activation, but it should still not be completely disregarded. However, it implicates immunological strategies outside of the conventional framework. Assuming very subtle responses not detected by our analyses, long-lived plasma (B) cells can potentially be generated without the help of T-cells. In mammals, these display much shorter life-time compared to plasma cells generated with T-cell help. In addition, they also appear to be generated outside of the germinal centers [110]. The latter supports a T-cell independent mechanism potentially being present in teleosts as teleosts do not generate germinal centers [111], but even more in Atlantic cod and Gadiformes due to the lack of *CD4* [10, 12].

Looking at the overall data in paper V, we suggested innate memory as the most likely mechanism where cells of the innate immune system become trained to rapidly respond to pathogens reencountered [76]. We found little recruitment of innate cells, likely due to the lack of inflammation. However, innate memory can be generated through alternative mechanisms. We found some support of a metabolism-related / epigenetic mechanism correlated with a shift from oxidative phosphorylation to glycolysis and the mTOR signaling pathway [75, 112, 113]. Moreover, innate memory can also be conducted through cells of lymphoid origin like NK-cells, innate lymphoid cells (ILCs) and T-cells [76]. Much like B- and T-cells, NK-cells express germ-line encoded receptors able to detect pathogen. In mice, these cells behave similar to B- and T-cells upon infection where they become activated, expand and then contract to generate long-lived memory cells [114]. NK-cell receptors and cell surface markers are poorly described in teleosts, but we still observed some signs of NK-cells in paper V together with apoptosis and mitophagy then likely responsible for NK-cell contraction upon establishment of the memory NK cell lineage [114]. CD8+ T-cells also generate memory lineages both in an antigen-dependent and antigen-independent manner. The former is more classical whereas the latter is induced by MHC-TCR interactions where MHC is presenting self antigen or the induction is facilitated by a certain combination of cytokines. After priming, expansion and massive contraction the remaining CD8+ T-cell subset display memory-like traits where they upon pathogen reencounter differentiate into effector cells conferring some protection in mice. Upon pathogen reencounter these cells migrated to the spleen where they interact with monocytes and neutrophils t o promote a greater immune response [115, 116].

The difficulty with these innate memory systems, collectively, is the diverse set of cytokines observed involved in mammals, i.e. cytokines that are not found in teleosts or because of sequence divergence are hard to annotate correctly. In

addition, there are cytokines reported found in teleosts, not found in mammals nor functionally tested, which may be involved in these innate memory mechanisms instead [7, 56]. Thus, we lack information about involved cytokines in these potential memory processes in Atlantic cod which restricts us to a gentle speculation about the presence and importance of these mechanisms.

In summary, we demonstrated that the immune system of Atlantic cod respond with a classic inflammation with acute phase proteins, PRRs, phagocytes and complement. Further, we observed a great contribution of MHCI, both endogenous and exogenous peptide presentation, and T-cell independent B-cell activation. In relation to establishment of memory the mechanisms are more unclear. However, a memory mechanism is present based on vaccine trials reporting protection and likely mechanisms are CD8+ T-cell memory, T-cell independent B-cell memory, innate memory in relation to the NK-cell lineage or innate memory via a metabolic shift and possibly epigenetic imprinting.

Concluding remarks and future perspectives

Using genomic and transcriptomic resources from Atlantic cod and teleosts species, this thesis demonstrates the combined effect of host-intrinsic, biotic and abiotic factors on the evolutionary development of teleost immunity and how great changes in past times have affected the immune system in contemporary teleost species. Further, we have also obtained insight into the functional orchestration of Atlantic cod immunity and by proxy the Gadiformes. However, further studies are needed to reveal the range of functional mechanisms and increase the level of details regarding evolutionary processes affecting this system. The 66 newly generated teleost genomes used in paper II and III have proven extremely useful for an improved understanding of the evolution of the teleost adaptive (see also Malmstrøm et al.) and innate (this thesis) immune system. Although these genomes were produced with a low-coverage approach,

they extensively cover the "gene space". Based on the findings in paper II and III, it should be possible to select small number of evolutionary interesting species and add sequencing (Illumina, PacBio) to generate reference-level genomes. This will enable a deeper and more complete characterization of the *TLR* family similarly to the investigations of Atlantic cod. More complete genomes will facilitate determination of possible pseudogenes, signs of selection, verification of proposed gene losses and comparative synteny. Also, the *TLR* and *MHC* complexes are not sole contributors to immunity and the 66 genomes provide a great opportunity to fully investigate the overall diversity of immune genes in the teleost lineage.

The functional studies on Atlantic cod revealed both classic and unconventional strategies. However, transcriptome analysis is not equal to end-point function. Still, some additional transcriptomic analyses on both vaccination and infection experiments with other pathogens, alternative vaccination strategies and different time-points should provide a better overview of immune mechanisms in Atlantic cod. This would in turn enable better design of end-point function studies both *in vitro* and *in vivo* needed to verify the proposed functions presented in this thesis. Furthermore, functional studies should also be conducted in other Gadiformes species to determine inter-species variations and correlate this with the differences in genetic repertoires.

Collectively, the findings in this thesis illustrates how we need to move away from interpreting teleosts within the mammalian framework and even the assumption that all teleosts maintain the same immunological strategy. A comparative approach where we assume that each species may apply its own unique strategy will likely reveal a plethora of interesting immunological mechanisms.

Acknowledgements

I would like to thank Helle T. Baalsrud, Rebekah Oomen and Ole K. Tørresen for reading and commenting on papers and thesis.

References

1. Faircloth, B.C., et al., *A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs).* PLoS One, 2013. **8**(6): p. e65923.
2. Sallan, L.C., *Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity.* Biol Rev Camb Philos Soc, 2014. **89**(4): p. 950-71.
3. Volff, J.N., *Genome evolution and biodiversity in teleost fish.* Heredity (Edinb), 2005. **94**(3): p. 280-94.
4. Buonocore, F. and M. Gerdol, *Alternative adaptive immunity strategies: coelacanth, cod and shark immunity.* Mol Immunol, 2015.
5. Bengten, E. and M. Wilson, *Antibody Repertoires in Fish.* Results Probl Cell Differ, 2015. **57**: p. 193-234.
6. Katzenback, B.A., *Antimicrobial Peptides as Mediators of Innate Immunity in Teleosts.* Biology (Basel), 2015. **4**(4): p. 607-39.
7. Bird, S. and C. Tafalla, *Teleost Chemokines and Their Receptors.* Biology (Basel), 2015. **4**(4): p. 756-84.
8. Palti, Y., *Toll-like receptors in bony fish: from genomics to function.* Dev Comp Immunol, 2011. **35**(12): p. 1263-72.
9. Zou, J. and C.J. Secombes, *Teleost fish interferons and their role in immunity.* Dev Comp Immunol, 2011. **35**(12): p. 1376-87.
10. Star, B., et al., *The genome sequence of Atlantic cod reveals a unique immune system.* Nature, 2011. **477**(7363): p. 207-10.
11. Tørresen, O.K., et al., *An improved genome assembly uncovers a profilic tandem repeat structure in Atlantic cod.* Submitted.
12. Malmstrøm, M., et al., *Evolution of the immune system influences speciation rates in teleost fishes.* Nat Genet, 2016.
13. McCurley, N., et al., *Immune related genes underpin the evolution of adaptive immunity in jawless vertebrates.* Curr Genomics, 2012. **13**(2): p. 86-94.
14. Kassahn, K.S., et al., *Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates.* Genome Res, 2009. **19**(8): p. 1404-18.
15. Glasauer, S.M. and S.C. Neuhauss, *Whole-genome duplication in teleost fishes and its evolutionary consequences.* Mol Genet Genomics, 2014. **289**(6): p. 1045-60.
16. Buchmann, K., *Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals.* Front Immunol, 2014. **5**: p. 459.
17. Dzik, J.M., *The ancestry and cumulative evolution of immune reactions.* Acta Biochim Pol, 2010. **57**(4): p. 443-66.
18. Abedon, S.T., *Bacterial 'immunity' against bacteriophages.* Bacteriophage, 2012. **2**(1): p. 50-54.
19. Hammami, R., et al., *Anti-infective properties of bacteriocins: an update.* Cell Mol Life Sci, 2013. **70**(16): p. 2947-67.
20. Mills, C.D., et al., *Sequential Immune Responses: The Weapons of Immunity.* J Innate Immun, 2015. **7**(5): p. 443-9.
21. Buckley, K.M. and J.P. Rast, *Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes.* Dev Comp Immunol, 2015. **49**(1): p. 179-89.
22. Satake, H. and T. Sekiguchi, *Toll-like receptors of deuterostome invertebrates.* Frontiers in Immunology, 2012. **3**: p. 34.
23. Buckley, K.M. and J.P. Rast, *Dynamic evolution of toll-like receptor multigene families in echinoderms.* Frontiers in Immunology, 2012. **3**: p. 136.
24. Kingsolver, M.B., Z. Huang, and R.W. Hardy, *Insect antiviral innate immunity: pathways, effectors, and connections.* J Mol Biol, 2013. **425**(24): p. 4921-36.
25. Ward, A.E. and B.M. Rosenthal, *Evolutionary responses of innate immunity to adaptive immunity.* Infect Genet Evol, 2014. **21**: p. 492-6.
26. Boehm, T., et al., *VLR-based adaptive immunity.* Annu Rev Immunol, 2012. **30**: p. 203-20.

27. Jones, F.C., et al., *The genomic basis of adaptive evolution in threespine sticklebacks.* Nature, 2012. **484**(7392): p. 55-61.
28. Aparicio, S., et al., *Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.* Science, 2002. **297**(5585): p. 1301-10.
29. Brawand, D., et al., *The genomic substrate for adaptive radiation in African cichlid fish.* Nature, 2014. **513**(7518): p. 375-81.
30. McGaugh, S.E., et al., *The cavefish genome reveals candidate genes for eye loss.* Nat Commun, 2014. **5**: p. 5307.
31. Mitani, H., et al., *The medaka genome: why we need multiple fish models in vertebrate functional genomics.* Genome Dyn, 2006. **2**: p. 165-82.
32. Schartl, M., et al., *The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits.* Nat Genet, 2013. **45**(5): p. 567-72.
33. Eisenstein, E.M. and C.B. Williams, *The T(reg)/Th17 cell balance: a new paradigm for autoimmunity.* Pediatr Res, 2009. **65**(5 Pt 2): p. 26r-31r.
34. Misra, D.P. and V. Agarwal, *Innate immune cells in the pathogenesis of primary systemic vasculitis.* Rheumatol Int, 2016. **36**(2): p. 169-82.
35. Eberl, G., et al., *Innate lymphoid cells. Innate lymphoid cells: a new paradigm in immunology.* Science, 2015. **348**(6237): p. aaa6566.
36. Luckheeram, R.V., et al., *CD4(+)T cells: differentiation and functions.* Clin Dev Immunol, 2012. **2012**: p. 925135.
37. Hodgkinson, J.W., L. Grayfer, and M. Belosevic, *Biology of Bony Fish Macrophages.* Biology (Basel), 2015. **4**(4): p. 881-906.
38. Fink, I.R., et al., *Immune-relevant thrombocytes of common carp undergo parasite-induced nitric oxide-mediated apoptosis.* Dev Comp Immunol, 2015. **50**(2): p. 146-54.
39. Shao, T., et al., *Characterization of surface phenotypic molecules of teleost dendritic cells.* Dev Comp Immunol, 2015. **49**(1): p. 38-43.
40. Sfacteria, A., M. Brines, and U. Blank, *The mast cell plays a central role in the immune system of teleost fish.* Mol Immunol, 2015. **63**(1): p. 3-8.
41. Bassity, E. and T.G. Clark, *Functional identification of dendritic cells in the teleost model, rainbow trout (Oncorhynchus mykiss).* PLoS One, 2012. **7**(3): p. e33196.
42. Rombout, J.H., et al., *Teleost intestinal immunology.* Fish Shellfish Immunol, 2011. **31**(5): p. 616-26.
43. Nakanishi, T., Y. Shibasaki, and Y. Matsuura, *T Cells in Fish.* Biology (Basel), 2015. **4**(4): p. 640-63.
44. Wiegertjes, G.F., et al., *Polarization of immune responses in fish: The 'macrophages first' point of view.* Mol Immunol, 2016. **69**: p. 146-56.
45. Riera Romo, M., D. Perez-Martinez, and C. Castillo Ferrer, *Innate immunity in vertebrates: an overview.* Immunology, 2016.
46. Lombardi, L., et al., *Insights into the antimicrobial properties of hepcidins: advantages and drawbacks as potential therapeutic agents.* Molecules, 2015. **20**(4): p. 6319-41.
47. Masso-Silva, J.A. and G. Diamond, *Antimicrobial peptides from fish.* Pharmaceuticals (Basel), 2014. **7**(3): p. 265-310.
48. Zhu, L.Y., et al., *Advances in research of fish immune-relevant genes: a comparative overview of innate and adaptive immunity in teleosts.* Dev Comp Immunol, 2013. **39**(1-2): p. 39-62.
49. Cray, C., J. Zaias, and N.H. Altman, *Acute phase response in animals: a review.* Comp Med, 2009. **59**(6): p. 517-26.
50. Gisladottir, B., et al., *Isolation of two C-reactive protein homologues from cod (Gadus morhua L.) serum.* Fish Shellfish Immunol, 2009. **26**(2): p. 210-9.
51. Yin, Q., et al., *Structural biology of innate immunity.* Annu Rev Immunol, 2015. **33**: p. 393-416.
52. Brubaker, S.W., et al., *Innate immune pattern recognition: a cell biological perspective.* Annu Rev Immunol, 2015. **33**: p. 257-90.
53. Quiniou, S.M., P. Boudinot, and E. Bengten, *Comprehensive survey and genomic characterization of Toll-like receptors (TLRs) in channel catfish, Ictalurus punctatus: identification of novel fish TLRs.* Immunogenetics, 2013. **65**(7): p. 511-30.
54. Boltana, S., et al., *PAMPs, PRRs and the genomics of gram negative bacterial recognition in fish.* Dev Comp Immunol, 2011. **35**(12): p. 1195-203.
55. Vasta, G.R., et al., *Structural and functional diversity of the lectin repertoire in teleost fish: relevance to innate and adaptive immunity.* Dev Comp Immunol, 2011. **35**(12): p. 1388-99.
56. Zou, J. and C.J. Secombes, *The Function of Fish Cytokines.* Biology (Basel), 2016. **5**(2).
57. Sokol, C.L. and A.D. Luster, *The chemokine system in innate immunity.* Cold Spring Harb Perspect Biol, 2015. **7**(5).

58. Keyel, P.A., *How is inflammation initiated? Individual influences of IL-1, IL-18 and HMGB1.* Cytokine, 2014. **69**(1): p. 136-45.
59. Said-Sadier, N. and D.M. Ojcius, *Alarmins, inflammasomes and immunity.* Biomed J, 2012. **35**(6): p. 437-49.
60. de Vasconcelos, N.M., N. Van Opdenbosch, and M. Lamkanfi, *Inflammasomes as polyvalent cell death platforms.* Cell Mol Life Sci, 2016.
61. Panda, S. and J.L. Ding, *Natural antibodies bridge innate and adaptive immunity.* J Immunol, 2015. **194**(1): p. 13-20.
62. Triantafilou, M., et al., *Complementing the inflammasome.* Immunology, 2016. **147**(2): p. 152-64.
63. Meri, S., *Self-nonself discrimination by the complement system.* FEBS Lett, 2016. **590**(15): p. 2418-34.
64. Nakao, M., et al., *The complement system in teleost fish: progress of post-homolog-hunting researches.* Dev Comp Immunol, 2011. **35**(12): p. 1296-308.
65. Neefjes, J., et al., *Towards a systems understanding of MHC class I and MHC class II antigen presentation.* Nat Rev Immunol, 2011. **11**(12): p. 823-836.
66. Jameson, S.C. and D. Masopust, *Diversity in T cell memory: an embarrassment of riches.* Immunity, 2009. **31**(6): p. 859-71.
67. Kurosaki, T., K. Kometani, and W. Ise, *Memory B cells.* Nat Rev Immunol, 2015. **15**(3): p. 149-59.
68. Clambey, E.T., et al., *Molecules in medicine mini review: the alphabeta T cell receptor.* J Mol Med (Berl), 2014. **92**(7): p. 735-41.
69. Maddaly, R., et al., *Receptors and signaling mechanisms for B-lymphocyte activation, proliferation and differentiation–insights from both in vivo and in vitro approaches.* FEBS Lett, 2010. **584**(24): p. 4883-94.
70. Kelley, J., L. Walter, and J. Trowsdale, *Comparative genomics of major histocompatibility complexes.* Immunogenetics, 2005. **56**(10): p. 683-95.
71. Grimholt, U., *MHC and Evolution in Teleosts.* Biology (Basel), 2016. **5**(1).
72. Castro, R., et al., *T cell diversity and TcR repertoires in teleost fish.* Fish Shellfish Immunol, 2011. **31**(5): p. 644-54.
73. Herzog, S. and H. Jumaa, *Self-recognition and clonal selection: autoreactivity drives the generation of B cells.* Curr Opin Immunol, 2012. **24**(2): p. 166-72.
74. Fillatreau, S., et al., *The astonishing diversity of Ig classes and B cell repertoires in teleost fish.* Front Immunol, 2013. **4**: p. 28.
75. Netea, M.G., et al., *Trained immunity: A program of innate immune memory in health and disease.* Science, 2016. **352**(6284): p. aaf1098.
76. Netea, M.G., et al., *Innate immune memory: a paradigm shift in understanding host defense.* Nat Immunol, 2015. **16**(7): p. 675-9.
77. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-400.
78. Barrangou, R., *The roles of CRISPR-Cas systems in adaptive immunity and beyond.* Curr Opin Immunol, 2015. **32**: p. 36-41.
79. Schmucker, D., et al., *Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity.* Cell, 2000. **101**(6): p. 671-84.
80. Dishaw, L.J., et al., *Gut immunity in a protochordate involves a secreted immunoglobulin-type mediator binding host chitin and bacteria.* Nat Commun, 2016. **7**: p. 10617.
81. Boehm, T. and J.B. Swann, *Origin and evolution of adaptive immunity.* Annu Rev Anim Biosci, 2014. **2**: p. 259-83.
82. Oisi, Y., et al., *Craniofacial development of hagfishes and the evolution of vertebrates.* Nature, 2013. **493**(7431): p. 175-80.
83. Pancer, Z., et al., *Variable lymphocyte receptors in hagfish.* Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9224-9.
84. Li, J., et al., *Definition of a third VLR gene in hagfish.* Proc Natl Acad Sci U S A, 2013. **110**(37): p. 15013-8.
85. Kasahara, M. and Y. Sutoh, *Two forms of adaptive immunity in vertebrates: similarities and differences.* Adv Immunol, 2014. **122**: p. 59-90.
86. Hirano, M., *Evolution of vertebrate adaptive immunity: immune cells and tissues, and AID/APOBEC cytidine deaminases.* Bioessays, 2015. **37**(8): p. 877-87.
87. Pilstrom, L., G. Warr, and S. Stromberg, *Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation.* Fisheries Science, 2005(71): p. 961-971.
88. Haase, D., et al., *Absence of major histocompatibility complex class II mediated immunity in pipefish, Syngnathus typhle: evidence from deep transcriptome sequencing.* Biol Lett, 2013. **9**(2): p. 20130044.
89. Darwin, C., *On the origin of species.* 1871, New York :: D. Appleton and Co.

90. Schulenburg, H., et al., *Introduction. Ecological immunology.* Philos Trans R Soc Lond B Biol Sci, 2009. **364**(1513): p. 3-14.
91. Parham, P., et al., *Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules.* Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1590): p. 800-11.
92. Hofmann, J., et al., *B-cells need a proper house, whereas T-cells are happy in a cave: the dependence of lymphocytes on secondary lymphoid tissues during evolution.* Trends Immunol, 2010. **31**(4): p. 144-53.
93. Wilson, P.A. and R.D. Norris, *Warm tropical ocean surface and global anoxia during the mid-Cretaceous period.* Nature, 2001. **412**(6845): p. 425-9.
94. Goldner, A., N. Herold, and M. Huber, *Antarctic glaciation caused ocean circulation changes at the Eocene-Oligocene transition.* Nature, 2014. **511**(7511): p. 574-7.
95. Benson, R.B., et al., *Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates.* Proc Biol Sci, 2010. **277**(1683): p. 829-34.
96. Benson, R.B. and P.S. Druckenmiller, *Faunal turnover of marine tetrapods during the Jurassic-Cretaceous transition.* Biol Rev Camb Philos Soc, 2014. **89**(1): p. 1-23.
97. Bambach, R.K., *Phanerozoic biodiversity mass extinctions.* Annual Review of Earth and Planetary Sciences, 2006. **34**: p. 127-155.
98. Alroy, J., *The shifting balance of diversity among major marine animal groups.* Science, 2010. **329**(5996): p. 1191-4.
99. Cavin, L., *Diversity of Mesozoic semionotiform fishes and the origin of gars (Lepisosteidae).* Naturwissenschaften, 2010. **97**(12): p. 1035-40.
100. Gomez, D., J.O. Sunyer, and I. Salinas, *The mucosal immune system of fish: the evolution of tolerating commensals while fighting pathogens.* Fish Shellfish Immunol, 2013. **35**(6): p. 1729-39.
101. Abt, M.C. and D. Artis, *The dynamic influence of commensal bacteria on the immune response to pathogens.* Curr Opin Microbiol, 2013. **16**(1): p. 4-9.
102. Lee, Y.K. and S.K. Mazmanian, *Has the microbiota played a critical role in the evolution of the adaptive immune system?* Science, 2010. **330**(6012): p. 1768-73.
103. Duggal, N.K. and M. Emerman, *Evolutionary conflicts between viruses and restriction factors shape immunity.* Nat Rev Immunol, 2012. **12**(10): p. 687-95.
104. Jack, R.S., *Evolution of Immunity and Pathogens.* Results Probl Cell Differ, 2015. **57**: p. 1-20.
105. Wu, R.S., *Hypoxia: from molecular responses to ecosystem responses.* Mar Pollut Bull, 2002. **45**(1-12): p. 35-45.
106. Star, B. and S. Jentoft, *Why does the immune system of Atlantic cod lack MHC II?* Bioessays, 2012. **34**(8): p. 648-51.
107. Martinez, J. and S. Merino, *Host-parasite interactions under extreme climatic conditions.* Current Zoology, 2011. **57**(3): p. 390-405.
108. Bakkemo, K.R., et al., *Intracellular localisation and innate immune responses following Francisella noatunensis infection of Atlantic cod (Gadus morhua) macrophages.* Fish Shellfish Immunol, 2011. **31**(6): p. 993-1004.
109. Mikkelsen, H. and M. Seppola, *Response to vaccination of Atlantic cod (Gadus morhua L.) progenies from families with different estimated family breeding values for vibriosis resistance.* Fish Shellfish Immunol, 2013. **34**(1): p. 387-92.
110. Bortnick, A. and D. Allman, *What is and what should always have been: long-lived plasma cells induced by T cell-independent antigens.* J Immunol, 2013. **190**(12): p. 5913-8.
111. Sunyer, J.O., *Fishing for mammalian paradigms in the teleost immune system.* Nat Immunol, 2013. **14**(4): p. 320-6.
112. Gardiner, C.M. and K.H. Mills, *The cells that mediate innate immune memory and their functional significance in inflammatory and infectious diseases.* Semin Immunol, 2016.
113. Cheng, S.C., et al., *mTOR- and HIF-1alpha-mediated aerobic glycolysis as metabolic basis for trained immunity.* Science, 2014. **345**(6204): p. 1250684.
114. O'Sullivan, T.E., J.C. Sun, and L.L. Lanier, *Natural Killer Cell Memory.* Immunity, 2015. **43**(4): p. 634-45.
115. Lauvau, G. and S.M. Soudja, *Mechanisms of Memory T Cell Activation and Effective Immunity.* Adv Exp Med Biol, 2015. **850**: p. 73-80.
116. Jameson, S.C., Y.J. Lee, and K.A. Hogquist, *Innate memory T cells.* Adv Immunol, 2015. **126**: p. 173-213.

Paper I

SCIENTIFIC REPORTS

OPEN

Evolutionary redesign of the Atlantic cod (*Gadus morhua* L.) Toll-like receptor repertoire by gene losses and expansions

Received: 20 January 2016
Accepted: 07 April 2016
Published: 29 April 2016

Monica H. Solbakken[1], Ole K. Tørresen[1], Alexander J. Nederbragt[1,5], Marit Seppola[2], Tone F. Gregers[3], Kjetill S. Jakobsen[1] & Sissel Jentoft[1,4]

Genome sequencing of the teleost Atlantic cod demonstrated loss of the Major Histocompatibility Complex (MHC) class II, an extreme gene expansion of MHC class I and gene expansions and losses in the innate pattern recognition receptor (PRR) family of Toll-like receptors (TLR). In a comparative genomic setting, using an improved version of the genome, we characterize *PRRs* in Atlantic cod with emphasis on *TLRs* demonstrating the loss of *TLR1/6*, *TLR2* and *TLR5* and expansion of *TLR7*, *TLR8*, *TLR9*, *TLR22* and *TLR25*. We find that Atlantic cod *TLR* expansions are strongly influenced by diversifying selection likely to increase the detectable ligand repertoire through neo- and subfunctionalization. Using RNAseq we find that Atlantic cod *TLRs* display likely tissue or developmental stage-specific expression patterns. In a broader perspective, a comprehensive vertebrate *TLR* phylogeny reveals that the Atlantic cod *TLR* repertoire is extreme with regards to losses and expansions compared to other teleosts. In addition we identify a substantial shift in *TLR* repertoires following the evolutionary transition from an aquatic vertebrate (fish) to a terrestrial (tetrapod) life style. Collectively, our findings provide new insight into the function and evolution of *TLRs* in Atlantic cod as well as the evolutionary history of vertebrate innate immunity.

Functional understanding of teleost immunity and its diversity is still in its infancy. Homologs of both mammalian innate and adaptive immune genes have been detected in teleost genomes, however, teleosts display greater genetic diversity as well as some functional discrepancies - for examples see references[1–3]. Central to innate immunity are pattern recognition receptors (PRRs) that detect pathogen associated molecular patterns (PAMPs) and initiate various features of the host's immune system - see[4] and references therein. One of the largest PRR families is the Toll-like receptors (TLRs). Upon ligand interaction, TLRs initiate the production of cytokines, anti-viral components and co-stimulatory molecules via the TLR signalling pathway - see[5] and references therein. The diversity of TLR repertoires among multicellular organisms is substantial. The invertebrate TLR repertoire spans from several hundred genes in the sea urchin (*Strongylocentrotus purpuratus*) to only two genes in the ascidian *Ciona intestinalis*[6]. This is in stark contrast to the less extensive vertebrate repertoire that generally display between 10–13 TLR genes - overview in[7–9].

Currently, there are ~20 known vertebrate *TLRs* (*TLR1-26*, the annotation used for individual genomes varies) where mammals display *TLR1-13* in contrast to fish which also display *TLR14–26*. Vertebrate *TLRs* form six families; *TLR1*, *TLR3*, *TLR4*, *TLR5*, *TLR7* and *TLR11* and individual species generally harbours at least one member from each family[8]. However, some exceptions are known such as the lack of *TLR11 -family* representatives in mammals. Teleosts display greater genetic diversity of *TLRs* but functional studies on mammalian *TLR* homologs overall report identical protein function - see[7,8].

In contrast to the genetic diversity found within the innate immune system the adaptive immune system is shown to display an intra-genetic polymorphic nature, i.e. to enable adaptation of the immune response towards

[1]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway. [2]Department of Medical Biology, The Arctic University of Norway, Tromsø, Norway. [3]Department of Biosciences, University of Oslo, Oslo, Norway. [4]Department of Natural Sciences, University of Agder, Kristiansand, Norway.[5]Research Group for Biomedical Informatics, Department of Informatics, University of Oslo, Oslo, Norway. Correspondence and requests for materials should be addressed to M.H.S. (email: m.h.solbakken@ibv.uio.no)

specific targets[10]. Large structural or functional alterations affecting acquired immunity have been perceived as less likely. During the last decade, however, several alternative immune strategies have been identified in vertebrate species - for details see[1,11,12]. Atlantic cod (*Gadus morhua*) is a particularly interesting case as genome sequencing revealed complete loss of the *MHC-II* pathway accompanied by an extreme gene expansion of *MHC-I* and gene losses and expansions within the *TLRs*[13–15]. By taking advantage of a new and substantially improved genome assembly combined with large scale genomic analyses we here perform a deep characterization of the major innate immune gene families in Atlantic cod, with emphasis on *TLRs*. Our phylogenetic analysis shows that the gene losses and expansions in Atlantic cod are extreme compared to other vertebrate lineages, including other teleosts. Comparative gene syntenies firmly establish the loss of *TLR1/6*, *TLR2* and *TLR5* and expansion of *TLR7*, *TLR8*, *TLR9*, *TLR22* and *TLR25*. Further, we are also able to more accurately determine *TLR* copy number, characterize *TLRs* not found in the earlier version of the genome and perform multiple selection analyses. We detect varying numbers of sites under diversifying selection within the *TLR* expansions most likely increasing the detectable ligand repertoire through neo- and subfunctionalization. Protein structure modelling and phylogenetic analysis suggest that *TLR* losses do not reduce the available genetic toolkit to detect pathogens. Furthermore, our transcriptome profiling of Atlantic cod *TLRs* show a likely tissue specific paralog usage. Finally, a comprehensive vertebrate *TLR* phylogeny demonstrates that there is a shift in *TLR* repertoires following the transition from aquatic to terrestrial life styles mirroring different selective pressures in the two environments.

Results

Atlantic cod PRR gene families – the deviating TLRs. We have investigated all major PRR gene families in Atlantic cod using the new and improved genome assembly (for details see method section "Genome assembly"). The *TLR* repertoire in Atlantic cod is clearly different compared to the other investigated teleosts and vertebrates. Within the collectin, pentraxin, retinoic acid-inducible (RIG) 1-like and nucleotide-binding oligomerization domain (NOD)-like families no clear differences were found – except for two genes: Atlantic cod has no evident homolog of *NOD2* and *AIM2* (Supplementary Tables 1–3). We have therefore focused on the *TLR* repertoire in the following investigations.

Gene syntenies verify TLR gene losses and expansions. We performed gene synteny analyses on all genomic regions in the assembly containing complete *TLRs* in Atlantic cod against the genomes of medaka (*Oryzias latipes*), fugu (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), zebrafish (*Danio rerio*) and stickleback (*Gasterosteus aculeatus*). We found conserved gene organization up- and downstream of *TLR1/6*, *TLR2* and *TLR5* proving their absence from the Atlantic cod genome. Comparatively, each species contained some genomic reshuffling and additional open reading frames – particularly prominent in zebrafish (Fig. 1). We find that *TLR7*, *TLR8*, *TLR9*, *TLR22* and *TLR25* are expanded in Atlantic cod and that the gene copies display both tandem and non-tandem organization in numerous contigs (Fig. 2). The *TLR8* and *TLR22* expansions are the most numerous with twelve copies each. The three *TLR7* copies are interspersed among the twelve *TLR8* copies. They are present in three different contigs where two have partial gene synteny compared to the other investigated teleosts (Fig. 2). Again, zebrafish display the most deviating local genomic architecture (Fig. 2). The five copies of *TLR9* are tandemly organized on a single contig that display general conserved synteny with the other species, however with some minor gene shuffling (Fig. 2). The twelve copies of *TLR22* are found in eight contigs. Three of these contigs have tandem organization of the *TLR22* copies, but most contigs are short and only contain a single gene. In only two contigs could synteny with flanking genes be determined (Fig. 2). The *TLR22* synteny also reveals that zebrafish has lost *TLR22*. This species also harbours a local inversion involving four genes downstream of the predicted *TLR22* region and display several additional open reading frames upstream compared to the other investigated species (Fig. 2). Finally, *TLR25* consists of seven copies in Atlantic cod found in three contigs. Two of the contigs demonstrate partial synteny and contigs with several *TLR25* copies display tandem organization. Medaka was the only other species containing *TLR25* and no local synteny directly downstream of the *TLR25* genomic region was evident for this species (Fig. 2). The single copy Atlantic cod *TLRs*, *TLR3*, *TLR14*, *TLR21* and *TLR23* were also located to genomic regions displaying conserved local synteny compared to the other investigated species (data not shown).

***TLR* expression patterns using RNAseq.** To investigate *TLR* expression patterns in Atlantic cod we performed RNAseq using the spleen/head kidney of healthy juvenile cod where the resulting reads were mapped towards all full-length *TLRs* found in the new Atlantic cod genome assembly. Most of the 43 full-length *TLRs* had detectable expression levels; however, four *TLRs* (two *TLR8* and two *TLR25*) had very low to no detectable expression. For the remaining *TLRs*, substantial variation in expression levels was observed (Fig. 3). The four genes with the lowest expression levels also displayed poor sequence quality resulting in protein translations containing frameshifts and stop codons possibly indicating pseudogenes. This was also the case for an additional six *TLRs*. In total 10 full-length *TLR* genes were excluded from further analysis (Supplementary Table 4).

Endolysosomal sorting signals in Atlantic cod. We compared known endolysosomal sorting signals from mammalian *TLRs* in the transmembrane, linker and cytosolic region against the corresponding regions of Atlantic cod *TLRs*. We found that the sorting signal in *TLR3* and *TLR9* were well conserved across all investigated species with the exception of *TLR3* in lamprey (Fig. 4A). We also searched for similar signals in the remaining *TLRs*: *TLR7*, *TLR8*, *TLR14*, *TLR21*, *TLR23* and *TLR25*. For *TLR25* a putative sorting signal was found (Fig. 4B), but for the other *TLRs* no clear conserved signalling motifs could be discerned (data not shown).

Protein structure modelling and diversifying selection. We modelled the 3D protein structure of all full-length *TLRs* in Atlantic cod (excluding those in Supplementary Table 4) onto the mammalian TLR5 structure (Fig. 5, Supplementary Figs 1, 2 and 3) as the overall structure of the TLR protein is central to TLR function. All

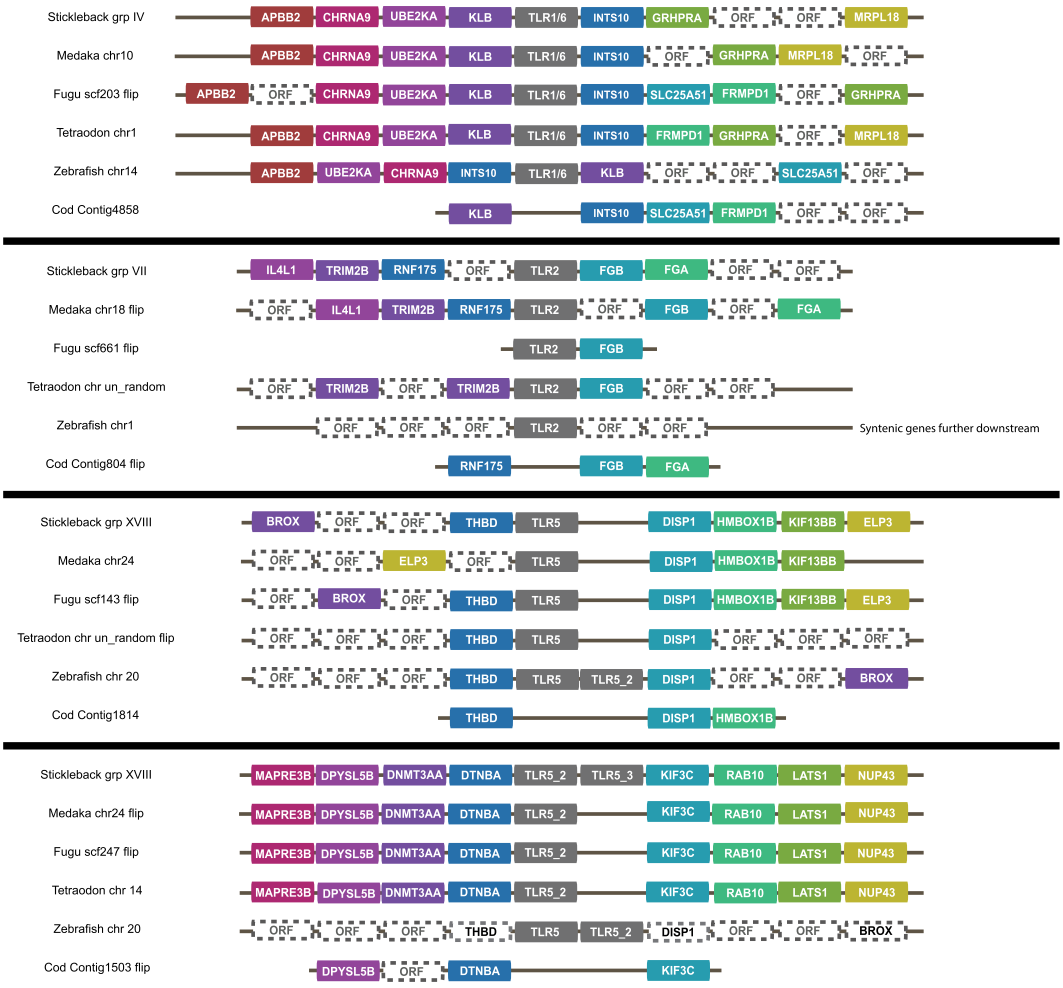


Figure 1. Gene synteny comparison of genomic regions in Atlantic cod towards genomic regions in stickleback, medaka, fugu, tetraodon and zebrafish containing *TLRs* not found in Atlantic cod (*TLR1/6*, *TLR2* and *TLR5*). Genes with colored boxes were found in several of the investigated species whereas white boxes designated ORF represents open reading frames which are species-specific and without certain annotation. Some genomic regions have been drawn in reversed order for visual purposes – designated "flip". For *TLR1/6* synteny is well conserved upstream of the *TLR* where zebrafish show a local inversion. Downstream of *TLR1/6* several genes are syntenic, but the gene order varies between species and there are some species - specific open reading frames. Atlantic cod has one contig that display syntenic genes towards the other species demonstrating the loss of *TLR1/6* from its genome. For *TLR2* synteny is less conserved, however, several common genes are found. *TLR2* in zebrafish is not located to the same genomic region as in the other fish; however, the syntenic genes are located further downstream on zebrafish chromosome 1. The fugu scaffold containing *TLR2* is short and only contains one additional annotation. Atlantic cod displays three syntenic genes, but no *TLR2*, demonstrating the loss of this gene. There were two genomic regions containing *TLR5* in the investigated species. The first *TLR5* region displays limited synteny upstream but more conserved synteny downstream of *TLR5*. Zebrafish has its two *TLR5* genes tandemly organized and also seems to have a local inversion compared to the other fish. Synteny is well conserved in the second *TLR5* region with the exception of zebrafish. Atlantic cod has one additional open reading frame compared to the other species. The syntenic genes in both putative *TLR5* regions in Atlantic cod demonstrate the loss of *TLR5* from its genome.

modelled *TLRs* conformed to the overall TLR structure with a solenoid ecto-domain, transmembrane domain, linker and Toll/interleukin-1 receptor (TIR) domain. *TLR3*, *TLR7*, *TLR8*, *TLR9*, *TLR21*, *TLR22* and *TLR23* displayed a longer solenoid ecto-domain structure (Fig. 5, Supplementary Figs 1 and 2). *TLR14* and *TLR25* demonstrated a somewhat shorter structure with loops modelled in their ecto-domains - more similar to the structure of other plasma membrane TLRs in mammals (Supplementary Figs 2 and 3).

The expanded Atlantic cod *TLRs*, with the exception of *TLR7* due to low copy number, were analyzed for sites under selection using three phylogeny-guided methods; SLAC, FEL and REL (see methods for details and Table 1). *TLR22* appears to have the most sites under diversifying selection and *TLR25* the least. Sites common between two or more selection analyses were mapped onto one of the modelled protein structures for each of the *TLR8*, *TLR9*, *TLR22* and *TLR25* gene expansions demonstrating that the sites are mainly located to loops interspersed between the leucine-rich repeat elements in the *TLRs* ecto-domains (Fig. 5A–D).

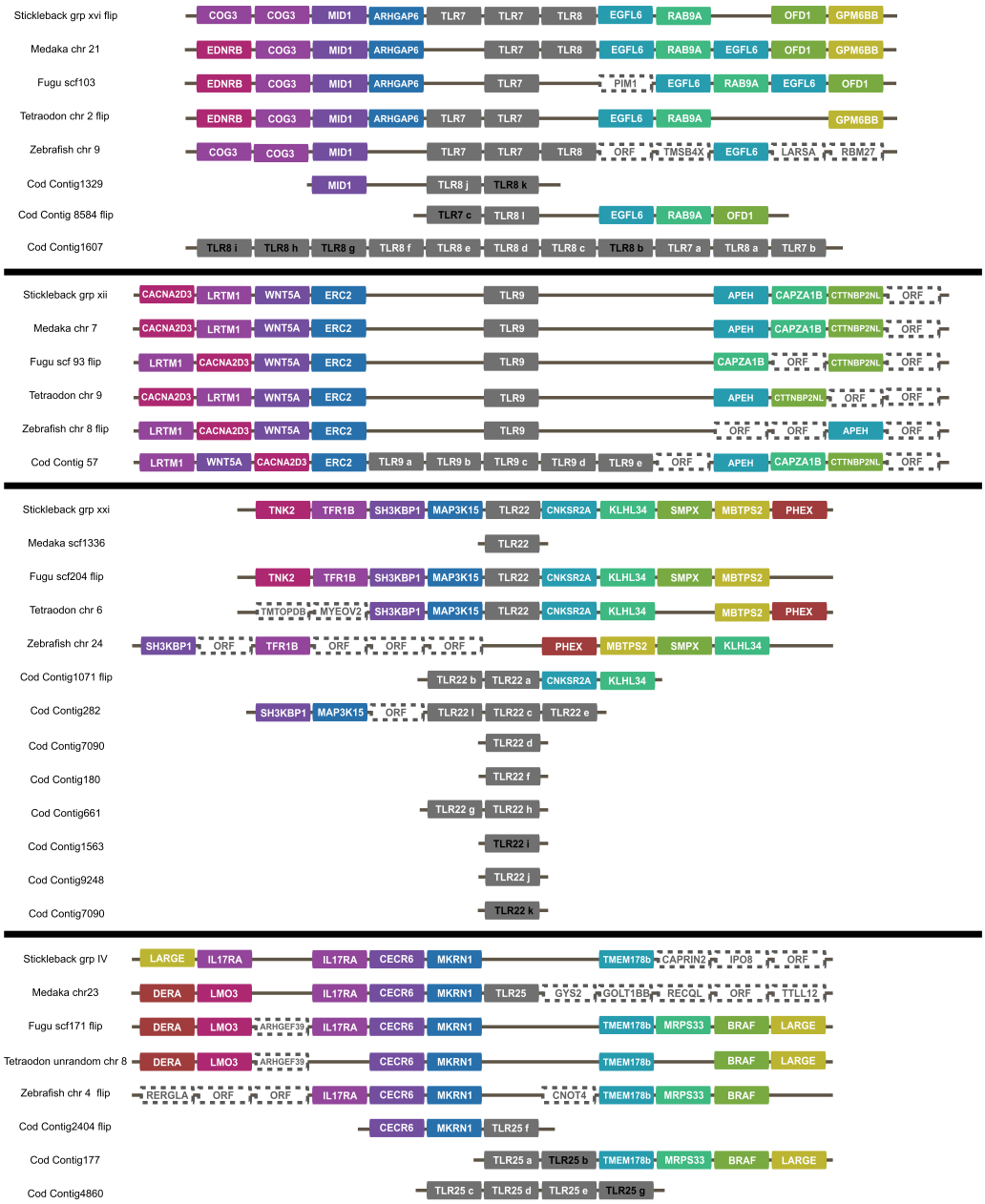


Figure 2. Gene synteny comparison of genomic regions in Atlantic cod towards genomic regions in stickleback, medaka, fugu, tetraodon and zebrafish containing *TLR7*, *TLR8*, *TLR9*, *TLR22* and *TLR25*. Genes with colored boxes were found in several of the investigated species whereas white boxes designated ORF represents open reading frames which are species-specific without certain annotation. Some genomic regions have been drawn in reversed order for visual purposes – designated "flip". *TLRs* in Atlantic cod removed from further analyses due to lacking expression and/or poor sequence quality listed in Supplementary Table S1 4 are written in black. *TLR7* and *TLR8* are located to the same genomic regions in the investigated fish species. Gene synteny is well conserved, however, zebrafish displays additional open reading frames of which some have proper annotation. Stickleback, tetraodon and zebrafish have two *TLR7* whereas fugu and tetraodon lacks *TLR8*. Atlantic cod has three contigs containing both *TLR7* and *TLR8* copies interspersed. Two of these contigs have partial synteny towards the other fish species. *TLR9* is also located to genomic regions with conserved synteny. Zebrafish displays less synteny downstream of its *TLR9*. Atlantic cod has five *TLR9* copies tandemly organized on a single contig with well conserved synteny. Also *TLR22* is located to a genomic region with relatively conserved synteny among the fish species. Medaka *TLR22* is present on a scaffold with no other annotated genes present. No *TLR22* was found in zebrafish and this species has a local inversion in the predicted *TLR22* region. Atlantic cod has eight contigs with *TLR22* gene copies present where two display partial synteny and tandem organization of the *TLR22* copies. The remaining contigs are short and contains only that single gene. The predicted *TLR25* regions have relatively well conserved synteny; however, synteny is absent downstream of medaka *TLR25* and somewhat disturbed downstream in stickleback and upstream in zebrafish. *TLR25* was only found in medaka and Atlantic cod. Atlantic cod *TLR25* copies are present on three contigs of which two have partial synteny. Contigs with several *TLR25* copies display tandem organization.

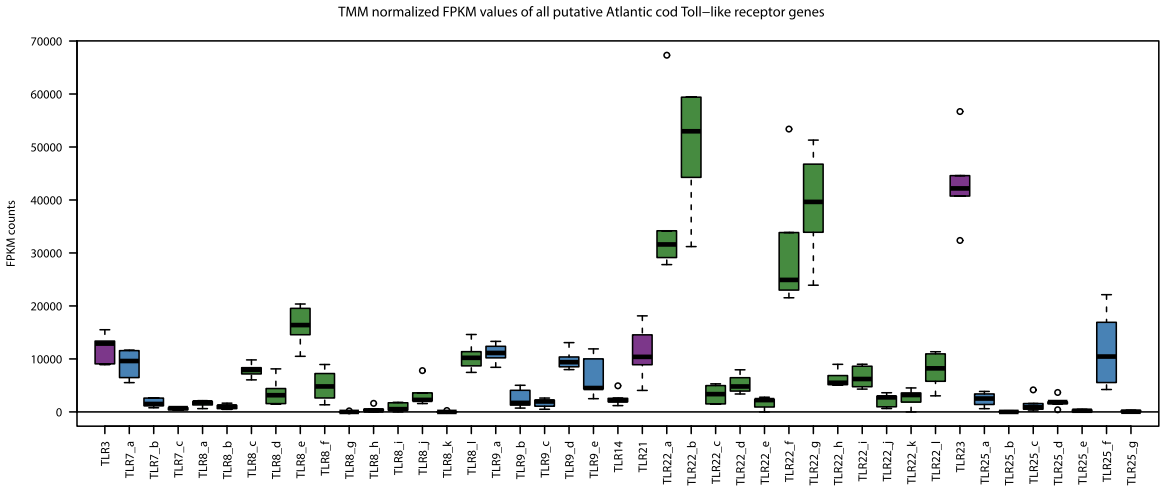


Figure 3. Transcriptome profiling of all Atlantic cod *TLRs*. Adapter and quality trimmed 100 bp paired-end Illumina RNAseq reads derived from the head kidney/spleen of six healthy juvenile cod were mapped towards an index of all full-length *TLRs* in Atlantic cod (S1 Table 2). The raw counts were converted to TMM normalized FPKM values and are displayed here as a box plot with average, standard deviation and outliers. The boxes have been colored for visualization purposes only. Some paralogs of *TLR7*, *TLR8* and *TLR25* have very low expression counts and the remaining *TLR* expansions display highly variable expression levels.

The TLR signalling pathway is intact in Atlantic cod. Using the mammalian TLR signalling network we searched for homologous genes in the new version of the Atlantic cod genome assembly (Supplementary Table 5). All components of the TLR signalling pathway were detected with the exception of TLR4 associated co-factors and some downstream T-cell/B-cell co-stimulatory molecules which were difficult to confirm due to distant sequence homology (Fig. 6). One downstream cytokine, interleukin-8 (*IL8*) showed substantial gene expansion: eight copies in total of which six were assembled to full-length (Supplementary Table 6). The translated sequences were subjected to a maximum likelihood (ML) protein sequence phylogenetic analysis together with *IL8* from fugu, tetraodon, tilapia, stickleback, medaka and human. The phylogeny grouped Atlantic cod *IL8's* in two clades (Supplementary Fig. 4). Transcriptome profiling of *IL8* (identical to that performed on Atlantic cod *TLRs*) did not resolve the paralogs sufficiently and thus the expression pattern of each clade or individual paralogs could not be further addressed (data not shown).

TLR annotation and vertebrate repertoires. We performed a multi-*TLR*, multi-species phylogenetic analysis using the translated sequence of the transmembrane, linker and TIR-domain regions of all *TLR* genes in selected vertebrate species with a main emphasis on teleosts (Supplementary Tables 2–4). The phylogeny resolved all six major *TLR* families, however, the *TLR11* and *TLR5* families display weaker support than the remaining families likely connected to the placement of *TLR21*, *TLR26* and *TLR13* (Fig. 7). Atlantic cod was the only species not harbouring any *TLRs* phylogenetically grouping within the *TLR1/6* and the *TLR2* clades of the *TLR1*-family. However, *TLR14* and *TLR25* are well supported within the *TLR1*-family clade. *TLR14* was not found in chicken and human. *TLR13* was present in the anole lizard (*Anolis carolinensis*), xenopus (*Xenopus tropicalis*) and coelacanth (*Latimeria chalumnae*). *TLR25* and *TLR26* were both sparsely found among the investigated fish species. Humans were the only species not displaying any members of the *TLR11*-family. The *TLR5*-family was not represented in either Atlantic cod or lamprey and the *TLR4*-family was only found in zebrafish, chicken (*Gallus gallus*), anole lizard and humans. Furthermore, the phylogeny demonstrates that the *TLR* gene expansions in Atlantic cod are rather extreme compared to the relatively few duplicates, triplicates and a single quadruplet expansion (xenopus *TLR14*) seen in the other species. No expansions were found within the human *TLR* repertoire (Fig. 7, Table 2).

Discussion

Signs of compensatory mechanisms for lost TLRs. Our *TLR* phylogeny indicates that Atlantic cod is the only known species lacking *TLR1/6* and *TLR2* which is confirmed by gene synteny analysis (Figs 1 and 7). These *TLRs*, members of the *TLR1*-family, are known to recognize peptidoglycan/lipoproteins at the plasma membrane. Roach *et al.*[8] have demonstrated a convincing link between phylogenetic relationships and function within vertebrate TLR families. Our *TLR* phylogeny suggests that Atlantic cod has other representatives within the *TLR1*-family – *TLR14* and *TLR25* – and thus any reduced ability to detect peptidoglycan/lipoprotein by TLRs could be alleviated (Fig. 7). Our phylogeny and synteny analyses also describe the loss of *TLR5* in Atlantic cod, a plasma membrane associated TLR detecting flagellin[7,8]. However, no compensatory mechanism similar to that of the *TLR1*-family was found as no other Atlantic cod *TLR* was placed within the *TLR5*-family (Figs 1 and 7). However, due to overlapping ligand profiles flagellin detection is likely covered by other PRR families in this species - see[16].

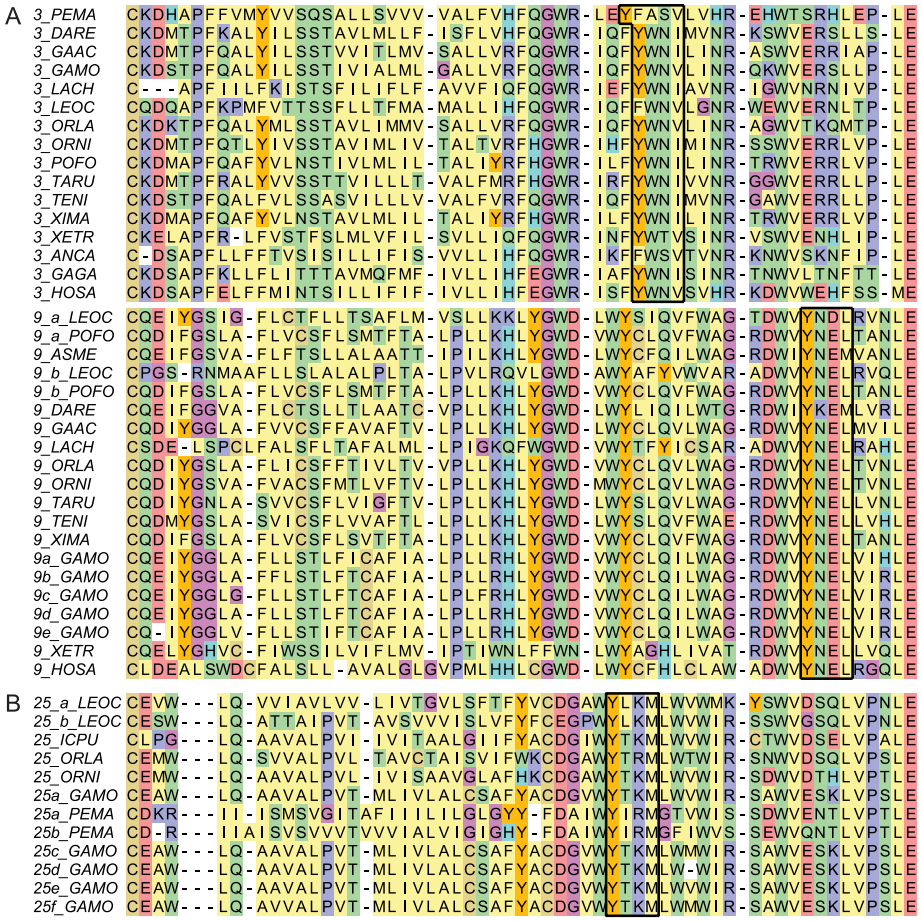


Figure 4. Edited amino acid alignments of the linker and transmembrane region of *TLR3*, *TLR9* and *TLR25* displaying known or putative tyrosine-containing endolysosomal sorting signals. (**A**) The known *TLR3* endolysosomal sorting signal is well conserved across species (black box) with the exception of *TLR3* in lamprey which has a phenylalanine in the tyrosine position and a tyrosine in the position before. For *TLR9* the signal is conserved in all species (black box). (**B**) For *TLR25* we propose an endolysosomal sorting signal in the linker region conserved across all species investigated that contain *TLR25*.

Functional assessment of *TLRs* through comparative analyses. With the aim of inferring function on Atlantic cod *TLRs* we performed several comparative analyses based on sequence homology which we interpreted using established links between function and phylogenetic relationships, protein structure and sorting signals. For *TLR3*, *TLR7*, *TLR8* and *TLR9* our findings support earlier functional reports demonstrating nucleic acid ligands and intracellular localization identical to their mammalian counterparts (Figs 2,4A,5A,5B and 7 and Supplementary Fig. 2)[17]. There are limited functional studies on non-mammalian TLRs (TLR11–26) of which TLR14–26 are present in teleosts. For TLR14 and TLR25 functional studies have so far not fully resolved ligand specificity. However, interesting results include transcriptional up-regulation of *TLR14* after exposure to viable gram negative bacteria[18] and transcriptional up-regulation of *TLR25* in response to parasites[19]. We propose a *TLR1*-family-like function for *TLR14* and *TLR25* implying plasma membrane localization and peptidoglycan or lipopolysaccharide-like ligands. This is further supported by protein structure modelling resolving shorter disrupted solenoid structures (Supplementary Figs 2 and 3) – structures correlated with plasma membrane localization and non-nucleic acid ligands[7,20], Furthermore, the presence of an intact TLR signalling pathway (Fig. 6) also supports the proposed function of TLR14 and TLR25. Otherwise one would expect a concurrent loss of adaptor proteins and co-factors specific for plasma membrane associated TLR proteins – in line with the observed loss of all TLR4-associated adapters in species lacking *TLR4*[21]. Lastly, our analysis revealed a putative endolysosomal sorting signal in *TLR25* similar to that of mammalian TLR3 and TLR9 (Fig. 4B)[22–25]. For TLR21 reports suggest that it is an intracellular TLR with a nucleic acid ligand[26,27]. No firm conclusion can be drawn for TLR22; there are several incongruent reports indicating a cell surface location with a nucleic acid ligand as well as transcriptional response towards several non-nucleic acid stimulants like peptidoglycan and lipopolysaccharide[28–32]. The function of TLR23 is also not established[29]. *TLR21*, *TLR22* and *TLR23* all belong to the *TLR11*-family (Fig. 7) and display the longer solenoid structures indicative of intracellular localization and nucleic acid ligands (Supplementary Figs 1 and 2). Considering that the rodent-specific TLR11 and TLR12 of the *TLR11*-family is shown to have

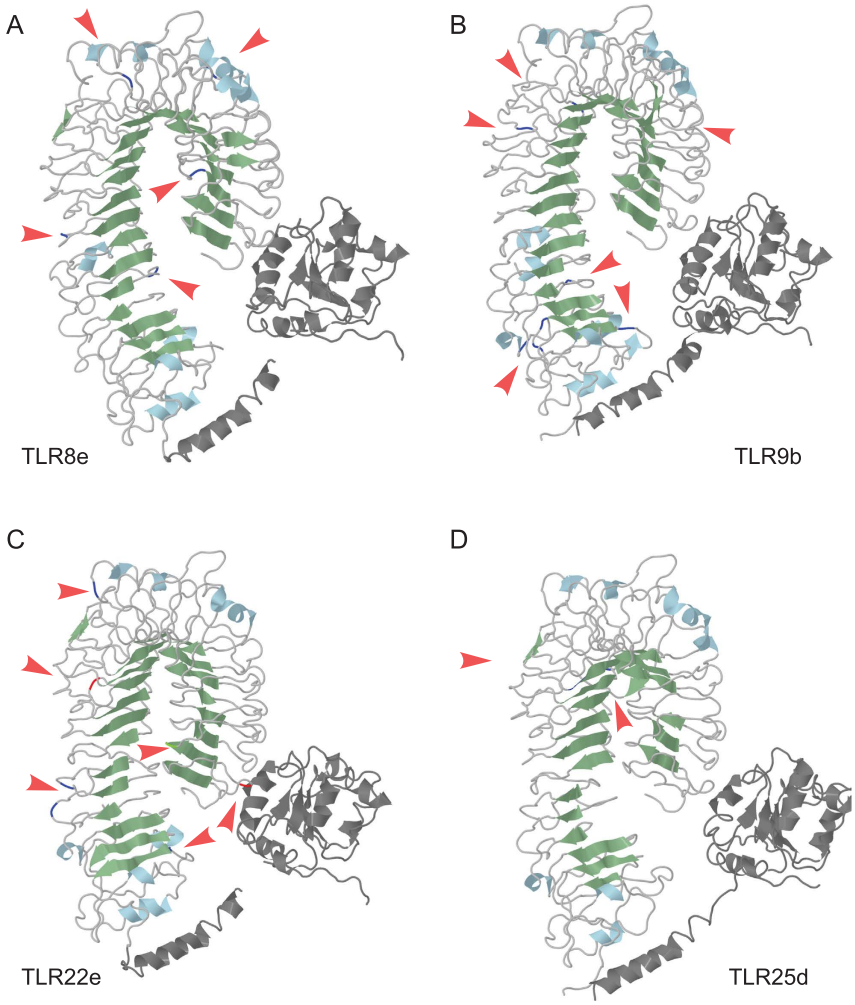


Figure 5. Sites under diversifying selection mapped onto the protein modeled structures of one paralog from each of the gene expansions *TLR8*, *TLR9*, *TLR22* and *TLR25* in Atlantic cod. The transmembrane, linker and TIR domain is colored dark grey whereas the ecto-domain is colored light grey with its sheets in pale green and helices in light blue. Sheets overlap with leucine-rich repeats in the ecto-domain. Arrows pointing at bright blue/bright red/bright green represents sites under diversifying selection as reported in Table 1. (**A**) Five sites (blue) mapped onto the modeled structure of *TLR8e*.The five sites are located both within and on the surface of the ecto-domain. (**B**) Eight sites (blue) mapped onto *TLR9b*. The sites are mainly located to two clusters in the ecto-domain with one cluster right at the border towards the transmembrane domain and one cluster in the middle of the ecto-domain. The sites are located both within and on the surface of the structure. (**C**) One, three and four sites (green, red and blue, respectively) are mapped onto *TLR22e*. With the exception of one site at the tip of the ecto-domain, the sites are located to the first half of the ecto-domain, mainly on the outer surface of the ecto-domain surface. (**D**) Two sites (blue) mapped onto *TLR25d* located to the middle and within the ecto-domain.

endosomal localization and that computational data supports a nucleic acid ligand for TLR22, our findings suggest that this whole family of TLRs do have nucleic acid ligands and most like intracellular localization[28,33–35].

Functional implications of lost and expanded *TLRs*. We detected diversifying selection among paralogs within the expanded Atlantic cod *TLRs*: *TLR8*, *TLR9*, *TLR22* and *TLR25* (Table 1). *TLR9* and *TLR22* stand out with the highest number of sites reported. Upon PAMP recognition, TLRs form TLR-homodimer:ligand complexes[36]. Vertebrates can further expand their detectable ligand repertoire by forming heterodimers within or between TLR families as have been demonstrated for TLR1/2, TLR2/6, TLR11/12 and TLR4/6[37–41]. The number of sites under diversifying selection in the ecto-domain of *TLR9* and *TLR22* suggests that the Atlantic cod's innate immune strategy partly involves an increase in its detectable ligand repertoire relative to other investigated fish species through "heterodimerization" between paralogs or possibly heterodimerization of paralogs with other TLRs. For *TLR8* and *TLR25*, the number of sites detected was much lower and somewhat inconsistent between the different methods (Table 1) suggesting that increased detectable ligand repertoire is not the main force maintaining these two gene expansions. We investigated the possibility of increased gene dosage by performing a

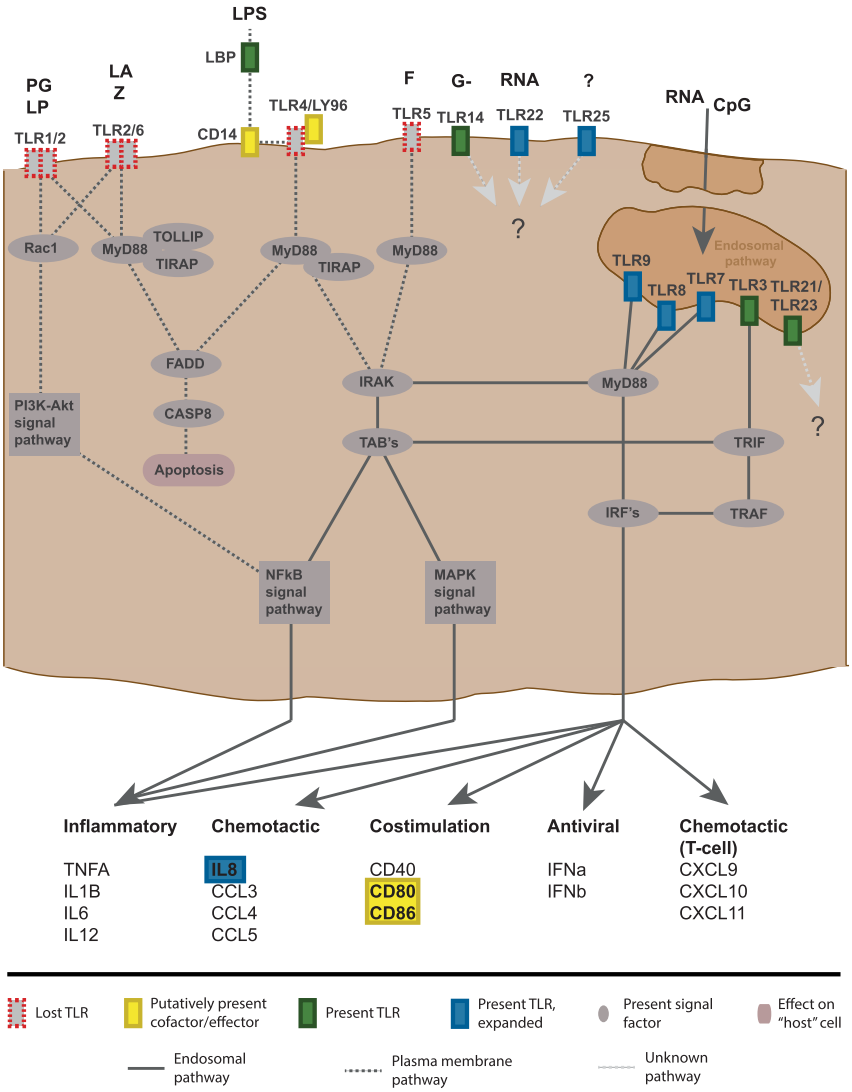
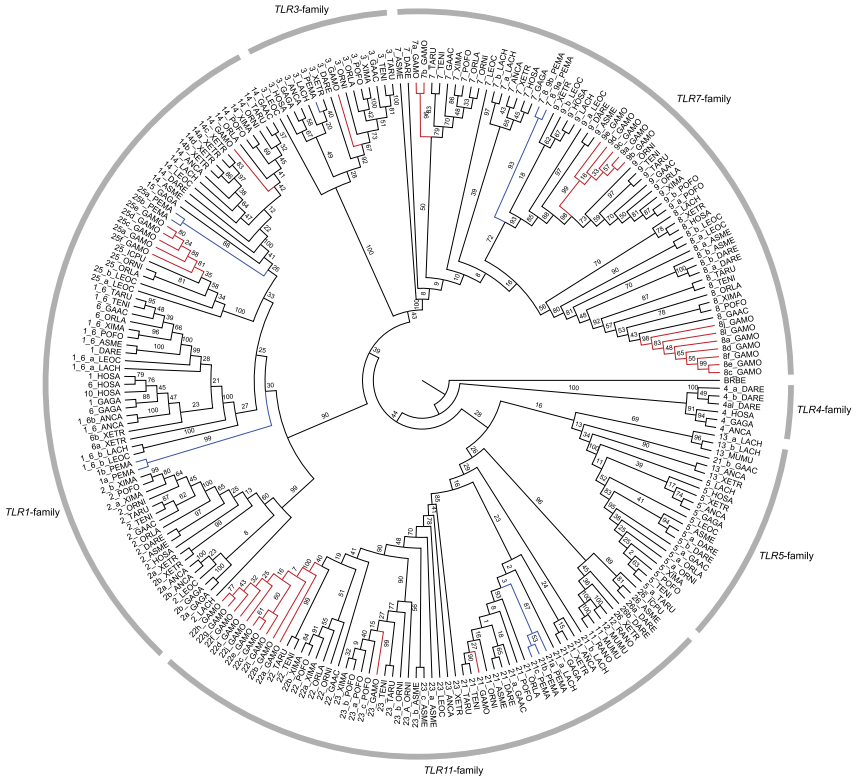


Figure 6. The mammalian TLR signaling pathway as depicted in KEGG condensed and presented to fit the proposed situation in Atlantic cod. Ligands are: PG – peptidoglycan (gram positive bacteria), LP – lipoprotein, LA – lipoarabinomannan, Z – zymosan (yeast), LPS – lipopolysaccharide (gram negative bacteria), G- – gram negative bacteria, F – flagellin, CpG – umethylated CpG DNA from bacteria. *TLR1/6*, *TLR2*, *TLR4* and *TLR5* are not found in Atlantic cod (also see Figs. 1 and 7). The presence of CD14, LY96 and CD80/86 was difficult to determine and are thus marked as putative. *TLR14*, *TLR21*, *TLR22*, *TLR23* and *TLR25* have unknown signaling pathways, but are drawn at their most likely affiliated membranes with the exception of *TLR22* drawn at the plasma membrane due to incongruent reports.

Analysis	TLR8	TLR9	TLR22	TLR25
SLAC	0	0	3	0
FEL	5	9	27	2
REL	0	44	7	0
Common sites	0	8*	1/3/4**	0

Table 1. Sites under diversifying selection as reported by SLAC, FEL and REL analyses. *Sites reported that are common between FEL and REL. **Sites reported that are common between all, SLAC and FEL or FEL and REL respectively.

transcriptome profiling of all *TLRs* expressed in the spleen/head kidney of healthy juvenile Atlantic cod. Here we found no evident need of increased gene dosage, however, it suggests more tissue-specific *TLR* and *TLR* paralog usage (Fig. 3). This is supported by *TLR* expression analyses by Sundaram *et al.*[29] in Atlantic cod (including *TLR22* paralogs) and by different expression levels of *TLRs* in various tissues in zebrafish and chicken[30,42].



BRBE	Amphioxus	*Branchiostoma belcheri*
ANCA	Anole lizzard	*Anolis carolinensis*
ASME	Cave fish	*Astyanax mexicanus*
DARE	Zebrafish	*Danio rerio*
GAAC	Stickleback	*Gasterosteus aculeatus*
GAGA	Chicken	*Gallus gallus*
GAMO	Atlantic cod ●	*Gadus morhua*
HOSA	Human	*Homo sapiens*
ICPU	Channel catfish	*Ictalurus punctatus*
LACH	Coelacanth	*Latimeria chalumnae*
LEOC	Spotted gar	*Lepisosteus oculatus*
MUMU	Mouse	*Mus musculus*
ORLA	Medaka	*Oryzias latipes*
ORNI	Tilapia	*Oreochromis niloticus*
PEMA	Lamprey ●	*Petromyzon marinus*
POFO	Amazon molly	*Poecilia formosa*
RANO	Rat	*Rattus norvegicus*
TARU	Fugu	*Takifugu rubripes*
TENI	Tetraodon	*Tetraodon nigroviridis*
XETR	Xenopus	*Xenopus tropicalis*
XIMA	Platyfish	*Xiphophorus maculatus*

Figure 7. A ML-phylogeny made from the transmembrane, linker and TIR-domains from all full length *TLRs* found in all investigated vertebrate species listen in S1 Table 3 displayed with bootstrap values (see also Table 2). An Amphioxus *TLR* gene was used as the root. Atlantic cod genes are marked in red and lamprey in blue. The six major *TLR* families are marked with grey bars with corresponding family name. The Atlantic cod expansions are extreme compared to other teleost. Xenopus contains the largest expansion in addition to Atlantic cod with 4 copies of *TLR14*. Humans do not have representatives from the *TLR11*-family. Atlantic cod and lamprey do not have *TLR5*-family members. Atlantic cod is the only species without *TLR1/6* and *TLR2*. Some TLRs are only found in some species such as *TLR4*, *TLR10*, *TLR13*, *TLR15*, *TLR25* and *TLR26*. The resolution of the *TLR5*- and *TLR11*-families is somewhat poor compared to the other families due to the placement of *TLR13*, *TLR21* and *TLR26*.

Teleost *TLR* repertoires are more diverse compared to other vertebrates. Our phylogenetic analysis of vertebrate *TLRs* revealed substantial variation in *TLR* repertoires. All investigated fish species, except zebrafish, lack representatives of the *TLR4*-family, *TLR5* is not found in lamprey and Atlantic cod and *TLR22* is lost in zebrafish (Figs 2 and 7 and Table 2). In contrast, certain *TLRs* are only present in a few species independent of phylogenetic relationships – i.e. *TLR13, TLR23*, *TLR25* and *TLR26*. With regard to the gene expansions observed, duplications seems to be more frequent within teleosts and less frequently occurring in other vertebrate lineages (Fig. 7 and Table 2). This pattern may be connected to the teleost genome duplication event where a causal connection between gene/genome duplication and subsequent neofunctionalization of paralogs has been established in contrast to the usual reciprocal loss of gene duplicates[43]. This is also in line with the sites under diversifying selection detected in the Atlantic cod *TLR* expansions (Table 1). Our data also demonstrate that *TLR14* is lost from birds and humans and that humans lack the entire *TLR11*-family. Notably, the *TLR* diversity and phylogeny suggest that life history strategies involving aquatic life stages require a different array of *TLR11*-family members and additional *TLRs* from the *TLR1*-family (Fig. 7 and Table 2). Thus, the transition from

an aquatic to a terrestrial lifestyle is associated with a shift in TLR repertoires – a shift that likely is linked to a highly different selection pressure on *TLRs* in the two environments.

The birth-and-death of *TLRs*. Multigene families connected to the immune system tend to follow a birth-and-death (BD) evolutionary model promoting diversification that manifests as general phylogenetic inter-specific gene clustering patterns, the presence of pseudogenes and gene losses[44,45]. Furthermore, gene expansions subjected to BD evolution and strong purifying selection undergo functional differentiation of the paralogs via sub- or neofunctionalization[44]. *TLRs* in general and especially their TIR-domains and leucine-rich repeat elements are known to be under strong purifying selection[46–48]. Our vertebrate *TLR* phylogeny demonstrates that gene losses and expansions are common in most lineages. However, the pattern is less pronounced in non-teleost lineages. Among teleosts, Atlantic cod shows the most pronounced loss and expansion pattern (Fig. 7 and Table 2). The BD model further supports our finding that sites under diversifying selection within *TLR8* and *TLR22* (and possibly *TLR9* and *TLR25*) in Atlantic cod (Table 1) likely increase the detectable ligand repertoire in this species. Finally, the extreme case of Atlantic cod compared to other teleosts indicates that its *TLR* repertoire is associated with the loss of *MHC-II*, i.e. that the loss of such a major adaptive immune system component has boosted evolutionary innovation through interlinked gene losses and expansions leading to high complexity and greater relative dependence on the innate immune system in this species.

Materials and Methods

Genome assembly. The genome assembly used in this study is one of four assemblies used to produce a new release of the Atlantic cod genome (Tørresen & Nederbragt *et al.* in prep). In short, overlapping sequencing reads from Illumina (180 bp insert size, 100 nt read length) were merged with FLASH using default options[49]. Meryl and merTrim were used to count and correct the reads, both programs from the Celera Assembler package 8.1[50]. 454 reads used in Star *et al.*[13] were converted from .sff files with sffToCA (also from Celera Assembler package) and corrected with merTrim, before trimmed with overlap based trimming (OBT, Celera Assembler program). Celera Assembler 8.2 alpha was used to trim subreads of PacBio sequencing reads. 20x of the merged Illumina 180 bp insert size reads, all paired 454 reads and the trimmed PacBio reads were used in an assembly with the Celera Assembler. The resulting genome assembly had some gaps closed with PBJelly[51] and was polished by Pilon[52]. Details are available upon request and later in Tørresen & Nederbragt *et al.* (in prep).

Genome mining for PRRs. We searched for PRR genes representing the major PRR families known in mammals listed in Supplementary Table 1 collected from Ensembl and UniProt[53,54]. The search was performed using TBLASTN from the BLAST+ suite with an e-value cut-off of 1e−1[55]. The low e-value was used to capture distant sequence homologs. Homologous relationships are described in Supplementary Table 1.

Selection of full-length TLR genes for further analyses. Annotated *TLR* sequences from selected species in Ensembl and GenBank covering all known *TLR* genes (listed in Supplementary Table 2) were compared towards the Atlantic cod genome using TBLASTN from the BLAST+ suite with an e-value cut-off of 1e−10 and otherwise default parameters[53,55,56]. All putative contigs containing *TLRs* were loaded into MEGA5[57] where regions of interest in each scaffold were extracted. Only full-length *TLRs* containing a complete ecto-domain, transmembrane domain, linker and complete TIR-domain were evaluated further. We performed RNAseq to evaluate expression levels as some of the full-length *TLR*s extracted contained several insertions and deletions making poor translated protein sequences. All extracted full-length *TLRs* were used to make an Atlantic cod *TLR* index. The quality and adapter trimmed RNAseq sequences from six healthy juvenile Atlantic cod (see RNAseq method section) were mapped towards this database and raw counts extracted using the RSEM/Bowtie wrapper included in Trinity v2.0.6[58]. These raw counts were normalized using the included edgeR scripts in Trinity to obtain TMM normalized FPKM counts[59]. *TLRs* with large amounts of insertions/deletions, either alone or in combination with low read counts, were excluded from further analysis as the accuracy of the translate protein sequences was questionable (Supplementary Table 4). Count matrix is available in the GitHub repository (https://github.com/uio-cels/Solbakken_TLRs).

Fish and totalRNA isolation for RNA sequencing. Total RNA was isolated from the head kidney/spleen of six healthy juvenile Atlantic cod. These fish originate from the Norwegian cod breeding program and were reported to be healthy without any history of diseases. The use of live Atlantic cod was approved by the National Animal Research authority in Norway (FOTS id 1147) and all methods were in accordance with the approved guidelines. The fish were transported at approx. 2 g to 100 L tanks at the Aquaculture Research Station (Tromsø, Norway) for grow-out in seawater of 3.4% salinity at 10 °C, 24 hour light and fed *ad libitum* with commercial feed (BioMar, Norway). The rates of water inflow were adjusted to an oxygen saturation of 90–100% in the outlet water. The tissue was stored on RNAlater (Life Technologies) and total RNA was isolated using Trizol (Life Technologies) according to protocol but using half the amount of tissue per volume Trizol recommended by the manufacturer. The complete laboratory protocol is available in the GitHub repository (https://github.com/uio-cels/Solbakken_TLRs). Sequencing libraries were produced according to the IlluminaTruSeq protocol (Illumina, Inc., San Diego, CA). Illumina HiSeq2000 100 bp paired-end sequencing services were provided by the Norwegian Sequencing Centre (http://www. sequencing.uio.no). Sequences were trimmed for adapters using Cutadapt v1.0 and trimmed on quality using Sickle using known Illumina adapter sequences, a Q threshold of 20 and otherwise default parameters[60,61].

Synteny analyses. The Ensembl[53] genome browser v78 (unless otherwise stated) was used to chart annotated open reading frames around *TLRs* annotated in the selected fish species. Protein sequences from these genes

	TLR1	TLR2	TLR6	TLR10	TLR14	TLR15	TLR25	TLR3	TLR4	TLR5	TLR13	TLR7	TLR8	TLR9	TLR21	TLR22	TLR23	TLR26
Homo sapiens	x	x	x	x				x	x	x		x	x	x				
Gallus gallus	x	x^2	x			x		x	x	x		x			x			
Anolis carolinensis	x^2	x^2	?		x	?		x	x	x	x	x			x		x	
Xenopus tropicalis	x^2	x^2	x		x^4			x		x	x	x	x	x	x		x	x
Gadus morhua					x		x^5(7)	x				x^2(3)	x^7(12)	x^5	x	x^8(12)	x	
Oreochromis niloticus	Frag.	x	?		x		x	x		x^2		x	x	x	x	x	x^2	
Poecilia formosa	x	x	?		x			x		x		x	x	x^2	x	x	x^3	
Takifugu rubripes	x	x	?		x			x		x^2		x	x	x	x	x	x	
Tetraodon nigroviridis	x	x	?		x			x		x		x	x	x	x	x	x	
Xiphophorus maculatus	x	x^2	?		x			x		x		x	x	x	x	x^2	x	
Astyanax mexicanus	x	x			x			x		x		x	x^2	x	x		x^3	x
Lepisosteus oculatus	x^2	x	?		x		x^2	x		x		x	x	x^2			x	
Gasterosteus aculeatus	x	x	?		x			x		x^3		x	x	x	x^2	x		
Oryzias latipes	x	x	?		x		x	x		x^2		x	x	x	x	x		
Danio rerio	x	x			x			x	x^3[1]	x^2		x	x^2	x	x			x^2
Latimeria chalumnae	x^2	x	?		x			x		x	x^2	x^2	x	x	x^2			
Petromyzon marinus	x^2		?		x		x^2	x				x^2			x^3			

Table 2. Overview of the full length *TLRs* found in all investigated species. Caret (^): the number of copies for a given gene if expanded. For *Gadus morhua* the number presented within () includes the genes excluded from further analyses given in S1 Table 4. For *TLR1* and *TLR6* – if homology could not be determined with confidence the copy was assigned to *TLR1* and a? designation given for *TLR6*. [1]TLR4 in zebrafish does not have homologous function to mammalian TLR4 (see reference Sepulcre, *et al.* 2009).

were downloaded and used in a TBLASTN[55] towards the Atlantic cod genome together with *TLR* representatives with an e-value cut-off of 1e−10. If a certain *TLR* was not annotated in one or several of the selected fish genomes in Ensembl we used the Ensembl BLAST tool with protein queries towards nucleic acid resources (TBLASTN) with default parameters to find the genomic region of interest. Some genome regions were reverse complemented for figure. drawing purposes and this is noted in the respective figures (Figs 1 and 2).

Endolysosomal sorting signals. Characterized *TLR* sorting signals were obtained from the literature[22,23]. Protein sequence was obtained for all *TLR3* and all *TLR9* genes investigated in this study (Supplementary Table 2). These were aligned with default settings using MEGA5 and ClustalW (Fig. 4A)[57]. We also searched for similar tyrosine based signals in the linker region of the remaining Atlantic cod *TLRs* (*TLR7*, *TLR8*, *TLR14*, *TLR21*, *TLR22*, *TLR23* and *TLR25*) (Fig. 4B).

TLR signalling pathway. The mammalian TLR signalling pathway available through the KEGG database[62] was used as a basis for mapping the pathway components in the Atlantic cod genome. The connected UniProt sequences for each pathway component were used in a TBLASTN search together with annotated homologs from fish species available at Ensembl or UniProt (Supplementary Table 5) towards the Atlantic cod genome with an e-value cut-off of 1e−1[53–55]. The low e-value was used due to distant homology of sequences between fish and mammals. Genes that were difficult to verify are highlighted in Fig. 6.

Protein structure prediction. Translated Atlantic cod *TLR* sequences were submitted to the Phyre2 structure prediction server for modelling[63]. All sequences were modelled against TLR5. All TLRs from *Homo sapiens* (human), *Petromyzon marinus* (lamprey), *Anolis carolinensis* (lizard) and *Oreochromis niloticus* (tilapia) were also submitted to Phyre2 and modelled onto the human TLR5 crystal structure (Fold library id: c3j0aA). The structures were coloured for visualization purposes using Jmol[64], differentiating between loops, sheets and helices

as well as the transmembrane, linker and TIR-domain (Supplementary Table 2 and Supplementary Figs 1–3). All Atlantic cod PDB files are available in the GitHub repository (https://github.com/uio-cels/Solbakken_TLRs).

Selection analyses. The expanded Atlantic cod *TLR*s with three or more full-length copies (*TLR8*, *TLR9*, *TLR22* and *TLR25*) were analyzed using Datamonkey[65]. Nucleotide sequences were imported into MEGA5 for alignment using default ClustalW parameters. The alignment was then manually edited to ensure proper translation to amino acids. A maximum likelihood phylogeny was made using partial deletion, a Jukes-Cantor model of sequence evolution with gamma distributed rate heterogeneity[57]. The resulting phylogeny was submitted together with the nucleotide alignment to Datamonkey. For each *TLR* expansion a model test was first run. The proposed best model was used before running selection analyses with the SLAC, FEL and REL methods. These are codon based maximum likelihood methods estimating rates of nonsynonymous and synonymous changes at each site in an alignment to identify sites under positive or negative selection. These tests are originally designed to be run on interspecies alignments. Here, since the tests are run on intraspecies paralogs, we argue that the sites reported to be under positive selection actually are under diversifying selection. The term diversifying selection is thus used throughout this report. Fixed effects likelihood model (FEL) estimates the ratio of nonsynonymous to synonymous substitution rates for each site in a sequence alignment with fixed estimates for branch lengths and substitution rate bias parameters. Random effects likelihood model (REL) allows rate variation in both nonsynonymous and synonymous rates and a general underlying nucleotide substitution model. Single-likelihood ancestor counting (SLAC) model weights the nucleotide substitution biases which are estimated from the data and allow ambiguous codons in the data. Sites reported to be under diversifying selection in two or more tests are highlighted in one of the protein structure models made for each of the *TLR8*, *TLR9*, *TLR22* and *TLR25* expansions. In cases where only one test has reported sites it is noted in the Fig. legend (Table 1 and Fig. 5). Phylogenies and alignments are available in the GitHub repository (https://github.com/uio-cels/Solbakken_TLRs).

Vertebrate *TLR* phylogeny. Full-length protein sequences were not alignable due to large variations in the ecto-domain of the *TLR*s. Thus, the transmembrane region, linker and TIR-domain were used as basis for phylogenetic analysis after alignment and minor curation of the data using MEGA5[57]. PROTTEST[66] was used for substitution model optimalization with the Bayesian Information Criterion (BIC) model selection criterion and testing all seven models available. PROTTEST suggested the JTT+I+G+F as the best substitution model. A maximum likelihood tree was produced using Randomized Axelerated Maximum Likelihood (RAxML) HPC-PTHREADS version 7. 2. 6 with the PROTCATJTT model[67]. The rapid bootstrap/search for the best tree simultaneously option was used and the analysis was run with 500 bootstraps. The resulting phylogeny was used as the basis for the final *TLR* annotations of all sequences used and described in this study (Supplementary Table 2). The tree was imported into FigTree v1.4[68] for cladogram transformation and then edited in Adobe Illustrator for improved Fig. visualization (Fig. 7). The alignment is available in the GitHub repository (https://github.com/uio-cels/Solbakken_TLRs).

References

1. Buonocore, F. & Gerdol, M. Alternative adaptive immunity strategies: coelacanth, cod and shark immunity. *Mol Immunol*, doi: 10.1016/j.molimm.2015.09.003 (2015).
2. Zhu, L. Y., Nie, L., Zhu, G., Xiang, L. X. & Shao, J. Z. Advances in research of fish immune-relevant genes: a comparative overview of innate and adaptive immunity in teleosts. *Dev Comp Immunol* **39,** 39–62, doi: 10.1016/j.dci.2012.04.001 (2013).
3. Magnadottir, B. Innate immunity of fish (overview). *Fish Shellfish Immunol* **20,** 137–151, doi: 10.1016/j.fsi.2004.09.006 (2006).
4. Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140,** 805–820, doi: 10.1016/j.cell.2010.01.022 (2010).
5. Kawasaki, T. & Kawai, T. Toll-like receptor signaling pathways. *Frontiers in immunology* **5,** 461, doi: 10.3389/fimmu.2014.00461 (2014).
6. Satake, H. & Sekiguchi, T. Toll-like receptors of deuterostome invertebrates. *Frontiers in immunology* **3,** 34, doi: 10.3389/fimmu.2012.00034 (2012).
7. Palti, Y. Toll-like receptors in bony fish: from genomics to function. *Dev Comp Immunol* **35,** 1263–1272, doi: 10.1016/j.dci.2011.03.006 (2011).
8. Roach, J. C. *et al.* The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci USA* **102,** 9577–9582, doi: 10.1073/pnas.0502272102 (2005).
9. Rauta, P. R., Samanta, M., Dash, H. R., Nayak, B. & Das, S. Toll-like receptors (TLRs) in aquatic animals: signaling pathways, expressions and immune responses. *Immunol Lett* **158,** 14–24, doi: 10.1016/j.imlet.2013.11.013 (2014).
10. Boehm, T. & Swann, J. B. Origin and evolution of adaptive immunity. *Annual review of animal biosciences* **2,** 259–283, doi: 10.1146/annurev-animal-022513-114201 (2014).
11. Boehm, T. *et al.* VLR-based adaptive immunity. *Annu Rev Immunol* **30,** 203–220, doi: 10.1146/annurev-immunol-020711-075038 (2012).
12. Haase, D. *et al.* Absence of major histocompatibility complex class II mediated immunity in pipefish, Syngnathus typhle: evidence from deep transcriptome sequencing. *Biology letters* **9,** 20130044, doi: 10.1098/rsbl.2013.0044 (2013).
13. Star, B. *et al.* The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477,** 207–210, doi: 10.1038/nature10342 (2011).
14. Malmstrom, M., Jentoft, S., Gregers, T. F. & Jakobsen, K. S. Unraveling the evolution of the Atlantic cod's (Gadus morhua L.) alternative immune strategy. *PLoS One* **8,** e74004, doi: 10.1371/journal.pone.0074004 (2013).
15. Star, B. & Jentoft, S. Why does the immune system of Atlantic cod lack MHC II? *Bioessays* **34,** 648–651, doi: 10.1002/bies.201200005 (2012).
16. Saleh, M. The machinery of Nod-like receptors: refining the paths to immunity and cell death. *Immunological Reviews* **243,** 235–246, doi: 10.1111/j.1600-065X.2011.01045.x (2011).
17. Zhang, J. *et al.* Toll-like receptor recognition of bacteria in fish: ligand specificity and signal pathways. *Fish Shellfish Immunol* **41,** 380–388, doi: 10.1016/j.fsi.2014.09.022 (2014).
18. Hwang, S. D., Kondo, H., Hirono, I. & Aoki, T. Molecular cloning and characterization of Toll-like receptor 14 in Japanese flounder, Paralichthys olivaceus. *Fish Shellfish Immunol* **30,** 425–429, doi: 10.1016/j.fsi.2010.08.005 (2011).
19. Zhao, F. *et al.* Expression profiles of toll-like receptors in channel catfish (Ictalurus punctatus) after infection with Ichthyophthirius multifiliis. *Fish Shellfish Immunol* **35,** 993–997, doi: 10.1016/j.fsi.2013.05.023 (2013).

20. Kang, J. Y. & Lee, J. O. Structural biology of the Toll-like receptor family. *Annual review of biochemistry* **80,** 917–941, doi: 10.1146/annurev-biochem-052909-141507 (2011).
21. Boltana, S., Roher, N., Goetz, F. W. & Mackenzie, S. A. PAMPs, PRRs and the genomics of gram negative bacterial recognition in fish. *Dev Comp Immunol* **35,** 1195–1203, doi: 10.1016/j.dci.2011.02.010 (2011).
22. Nishiya, T., Kajita, E., Miwa, S. & Defranco, A. L. TLR3 and TLR7 are targeted to the same intracellular compartments by distinct regulatory elements. *The Journal of biological chemistry* **280,** 37107–37117, doi: 10.1074/jbc.M504951200 (2005).
23. Leifer, C. A. *et al.* Cytoplasmic targeting motifs control localization of toll-like receptor 9. *The Journal of biological chemistry* **281,** 35585–35592, doi: 10.1074/jbc.M607511200 (2006).
24. Qi, R., Singh, D. & Kao, C. C. Proteolytic processing regulates Toll-like receptor 3 stability and endosomal localization. *The Journal of biological chemistry* **287,** 32617–32629, doi: 10.1074/jbc.M112.387803 (2012).
25. Ewald, S. E. *et al.* The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. *Nature* **456,** 658–662, doi: 10.1038/nature07405 (2008).
26. Brownlie, R. *et al.* Chicken TLR21 acts as a functional homologue to mammalian TLR9 in the recognition of CpG oligodeoxynucleotides. *Mol Immunol* **46,** 3163–3170, doi: 10.1016/j.molimm.2009.06.002 (2009).
27. Keestra, A. M., de Zoete, M. R., Bouwman, L. I. & van Putten, J. P. Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. *J Immunol* **185,** 460–467, doi: 10.4049/jimmunol.0901921 (2010).
28. Matsuo, A. *et al.* Teleost TLR22 recognizes RNA duplex to induce IFN and protect cells from birnaviruses. *J Immunol* **181,** 3474–3485 (2008).
29. Sundaram, A. Y., Consuegra, S., Kiron, V. & Fernandes, J. M. Positive selection pressure within teleost Toll-like receptors tlr21 and tlr22 subfamilies and their response to temperature stress and microbial components in zebrafish. *Molecular biology reports* **39,** 8965–8975, doi: 10.1007/s11033-012-1765-y (2012).
30. Sundaram, A. Y., Kiron, V., Dopazo, J. & Fernandes, J. M. Diversification of the expanded teleost-specific toll-like receptor family in Atlantic cod, Gadus morhua. *BMC evolutionary biology* **12,** 256, doi: 10.1186/1471-2148-12-256 (2012).
31. Reyes-Becerril, M. *et al.* Molecular cloning and comparative responses of Toll-like receptor 22 following ligands stimulation and parasitic infection in yellowtail (Seriola lalandi). *Fish Shellfish Immunol* **46,** 323–333, doi: 10.1016/j.fsi.2015.06.020 (2015).
32. Salazar, C., Haussmann, D., Kausel, G. & Figueroa, J. Molecular cloning of Salmo salar Toll-like receptors (TLR1, TLR22, TLR5M and TLR5S) and expression analysis in SHK-1 cells during Piscirickettsia salmonis infection. *J Fish Dis*, doi: 10.1111/jfd.12354 (2015).
33. Pifer, R., Benson, A., Sturge, C. R. & Yarovinsky, F. UNC93B1 is essential for TLR11 activation and IL-12-dependent host resistance to Toxoplasma gondii. *The Journal of biological chemistry* **286,** 3307–3314, doi: 10.1074/jbc.M110.171025 (2011).
34. Koblansky, A. A. *et al.* Recognition of profilin by Toll-like receptor 12 is critical for host resistance to Toxoplasma gondii. *Immunity* **38,** 119–130, doi: 10.1016/j.immuni.2012.09.016 (2013).
35. Sahoo, B. R. *et al.* Understanding the distinguishable structural and functional features in zebrafish TLR3 and TLR22, and their binding modes with fish dsRNA viruses: an exploratory structural model analysis. *Amino acids* **47,** 381–400, doi: 10.1007/s00726-014-1872-2 (2015).
36. Botos, I., Segal, D. M. & Davies, D. R. The structural biology of Toll-like receptors. *Structure (London, England: 1993)* **19,** 447–459, doi: 10.1016/j.str.2011.02.004 (2011).
37. Farhat, K. *et al.* Heterodimerization of TLR2 with TLR1 or TLR6 expands the ligand spectrum but does not lead to differential signaling. *Journal of leukocyte biology* **83,** 692–701, doi: 10.1189/jlb.0807586 (2008).
38. Raetz, M. *et al.* Cooperation of TLR12 and TLR11 in the IRF8-dependent IL-12 response to Toxoplasma gondii profilin. *J Immunol* **191,** 4818–4827, doi: 10.4049/jimmunol.1301301 (2013).
39. Andrade, W. A. *et al.* Combined action of nucleic acid-sensing Toll-like receptors and TLR11/TLR12 heterodimers imparts resistance to Toxoplasma gondii in mice. *Cell host & microbe* **13,** 42–53, doi: 10.1016/j.chom.2012.12.003 (2013).
40. Stewart, C. R. *et al.* CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. *Nat Immunol* **11,** 155–161, doi: 10.1038/ni.1836 (2010).
41. Jin, M. S. *et al.* Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130,** 1071–1082, doi: 10.1016/j.cell.2007.09.008 (2007).
42. Kannaki, T. R., Reddy, M. R., Verma, P. C. & Shanmugam, M. Differential Toll-like receptor (TLR) mRNA expression patterns during chicken embryological development. *Animal biotechnology* **26,** 130–135, doi: 10.1080/10495398.2014.939658 (2015).
43. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet* **10,** 725–732, doi: 10.1038/nrg2600 (2009).
44. Eirin-Lopez, J. M., Rebordinos, L., Rooney, A. P. & Rozas, J. The birth-and-death evolution of multigene families revisited. *Genome dynamics* **7,** 170–196, doi: 10.1159/000337119 (2012).
45. Nei, M. & Rooney, A. P. Concerted and birth-and-death evolution of multigene families. *Annual review of genetics* **39,** 121–152, doi: 10.1146/annurev.genet.39.073003.112240 (2005).
46. Areal, H., Abrantes, J. & Esteves, P. J. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC evolutionary biology* **11,** 368, doi: 10.1186/1471-2148-11-368 (2011).
47. Barreiro, L. B. *et al.* Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS genetics* **5,** e1000562, doi: 10.1371/journal.pgen.1000562 (2009).
48. Mikami, T., Miyashita, H., Takatsuka, S., Kuroki, Y. & Matsushima, N. Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene* **503,** 235–243, doi: 10.1016/j.gene.2012.04.007 (2012).
49. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)* **27,** 2957–2963, doi: 10.1093/bioinformatics/btr507 (2011).
50. Miller, J. R. *et al.* Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics (Oxford, England)* **24,** 2818–2824, doi: 10.1093/bioinformatics/btn548 (2008).
51. English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* **7,** e47768, doi: 10.1371/journal.pone.0047768 (2012).
52. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9,** e112963, doi: 10.1371/journal.pone.0112963 (2014).
53. Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res* **43,** D662–669, doi: 10.1093/nar/gku1010 (2015).
54. UniProt. UniProt: a hub for protein information. *Nucleic Acids Res* **43,** D204–212, doi: 10.1093/nar/gku989 (2015).
55. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10,** 421, doi: 10.1186/1471-2105-10-421 (2009).
56. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res* **41,** D36–42, doi: 10.1093/nar/gks1195 (2013).
57. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28,** 2731–2739, doi: 10.1093/molbev/msr121 (2011).
58. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29,** 644–652, doi: 10.1038/nbt.1883 (2011).
59. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26,** 139–140, doi: 10.1093/bioinformatics/btp616 (2010).
60. Sickle - a windowed adaptive trimming tool for FASTQ files using quality.

61. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Bioinformatics in Action* **17,** 10–12, doi: citeulike-article-id:11851772 (2012).
62. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28,** 27–30 (2000).
63. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10,** 845–858, doi: 10.1038/nprot.2015.053 (2015).
64. *Jmol: an open-source Java viewer for chemical structures in 3D*, http://www.jmol.org/.
65. Delport, W., Poon, A. F., Frost, S. D. & Kosakovsky Pond, S. L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics (Oxford, England)* **26,** 2455–2457, doi: 10.1093/bioinformatics/btq429 (2010).
66. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)* **21,** 2104–2105, doi: 10.1093/bioinformatics/bti263 (2005).
67. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* **22,** 2688–2690, doi: 10.1093/bioinformatics/btl446 (2006).
68. FigTree http://tree.bio.ed.ac.uk/software/figtree/ (2015).

Acknowledgements

This work was supported by The Research Council of Norway (Grant number 199806/S40 to KSJ). The genome assembly was made using the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/. The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities. We would like to thank Dr. Simon MacKenzie for valuable comments and discussions.

Author Contributions

M.H.S. performed all analyses. O.K.T. and A.J.N. made and provided the new version of the Atlantic cod genome. MS provided material for RNAseq. M.H.S., T.F.G., S.J. and K.S.J. interpreted the results. M.H.S. wrote the main text and prepared all figures with the assistance of T.F.G., S.J. and K.S.J. All authors contributed to review of the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Solbakken, M. H. *et al.* Evolutionary redesign of the Atlantic cod (*Gadus morhua* L.) Toll-like receptor repertoire by gene losses and expansions. *Sci. Rep.* **6**, 25211; doi: 10.1038/srep25211 (2016).

Paper II

Unveiling the evolution of the teleost innate immune system

Monica Hongrø Solbakken[1*], Kjetil Lysne Voje[1], Kjetill Sigurd Jakobsen[1], Sissel Jentoft[1,2*]

[1] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

[2] Department of Natural Sciences, University of Agder, Kristiansand, Norway

*Corresponding authors:

Monica Hongrø Solbakken. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. m.h.solbakken@ibv.uio.no, +47 22 84 41 64

Sissel Jentoft. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. sissel.jentoft@ibv.uio.no, +47 22 85 72 39

Genome sequencing efforts of non-model organisms have provided new insight into the extreme diversity of the teleost lineage including evidence for several alternate immunological strategies. The discoveries of the genetic loss of the Major Histocompatibility (*MHC*) class II pathway in Atlantic cod (*Gadus morhua*) as well as the functional loss in the more distant *Syngnathus typhle* [1, 2], show that *MHCII* is not crucial for the defence against pathogens and survival in some fish species. These findings are further supported in a recent study by Malmstrøm et al., which demonstrated that the loss of *MHCII* is shared by the entire Gadiformes lineage [3]. Moreover, characterization of the teleost innate immune system has demonstrated a different set of *Toll-like receptors (TLR)* compared to other vertebrates [4-6]. However, the underlying selective mechanisms driving the variety of immunological strategies observed and why they arose are poorly understood. Using genome assemblies from 66 teleost species our aim was to characterise teleost *TLRs* with emphasis on the Gadiformes lineage and thereby investigate the possible link between the loss of *MHCII*, past environmental conditions and the genetic architecture of the innate immune system. We show that the teleost *TLR* repertoire contain an array of lineage-specific losses and expansions, with the Gadiformes lineage as an extreme outlier. Interestingly, within the Gadiformes we discovered expansions of *TLR* genes to be correlated with the loss of *MHCII*, whereas *TLR* copy number variation correlated with latitudinal distribution in teleosts overall. In contrast, a minor correlation was found towards depth for *TLR9* and *TLR22*. This suggests that there is a strong on-going selection of the innate immune system linked to specific environmental factors. Furthermore, timing of the lineage-specific losses overlaps with well-described changes in paleoclimate and continental drift, and hence unveils past adaptive signatures driving the genetic change within the teleost immune system. Our study reveals a remarkable evolutionary flexibility of teleost innate immunity, which has played an essential role in the survival and radiation of the teleost lineage.

Mapping all the identified teleost *TLRs* — extracted from the 66 genome assemblies – onto the phylogeny of Malmstrøm et al. demonstrates the presence of comprehensive *TLR* repertoires in all investigated teleosts (Figure 1) similar to that found in other vertebrates [4, 6, 7]. However, most notable was the observation of three lineage-specific gene losses, several lineage-specific gene expansions and a substantial number of recorded species-specific variants (Figure 1). Specifically, *TLR1/2* are lost from the Gadinae (40-16 mya) in addition to being completely or partially lost in *Bregmaceros cantori*, *Benthosema glaciale*, *Stylephorus chordatus and Guentherus altivela*. *TLR5* is lost from the entire Paracanthopterygii superorder and the order Lampridiformes (175-130 mya) in addition to *Monocentris japonica*, *Acanthochaenus luetkenii* and *Pseudochromis fuscus*. Further, we discovered a new *TLR*, annotated as *TLR21beta* based on sequence homology, which is also absent in all Paracanthopterygiian species, with the exception of *Polymixia japonica,* and Lampridiformes. However, in contrast to *TLR5*, the presence of *TLR21beta* does not follow any clear phylogenetic pattern outside Paracanthopterygii/Lampridiformes (Figure 1). The Gadinae is the only clade consistent with the recently reported alternative *TLR* repertoire in Atlantic cod [1, 7] due to the prominent gene losses of *TLR1/2*. The more ancient loss of *TLR5* (Figure 1) supports previous discoveries of *TLR5* loss in the Atlantic cod [1, 7] as well as in the superorder Paracanthopterygii and the order Lampridiformes (Solbakken et al., paper III in this thesis). Evident lineage-specific gene losses, here demonstrated by *TLR1/2*, *TLR5* and *TLR21b* (Figure 1), have been previously suggested to be the result of adaptation to changes in species' habitat [8]. This was also suggested for *TLR5* and Myxovirus resistance (*Mx*) gene investigated by Solbakken et al. where these losses correlated with well-described past changes in climate (Solbakken et al., paper III in this thesis).

Three *TLRs* are found in all species; *TLR3*, *TLR14* and *TLR21*, the latter with the exception of *Benthosema glaciale*. Within the Gadiformes we find gene expansions for

69 *TLR7-9*, *TLR22*, *TLR23* and *TLR25*, especially within the C1 clade (see Figure 1).
70 Outside the Gadiformes the presence of *TLR25* displays no obvious phylogenetic
71 pattern. This is in contrast to *TLR7-9* which are present in all species with the
72 exception of a single *TLR8* loss in *Guentherus altivela*. *TLR22* and *TLR23* are found in
73 all Gadiformes except in *Bregmaceros cantori* and show a substantial degree of gene
74 expansion within the Gadiformes lineage, particularly for *TLR22*. Outside the
75 Gadiformes, the expansion of *TLR22* is less pronounced whereas, in contrast, *TLR23*
76 is frequently expanded. Otherwise, among the non-Gadiformes species, *TLR7-9*,
77 *TLR22*, *TLR23* and *TLR25* have non-structured phylogenetic patterns of presence and
78 gene loss as well as gene expansions (Figure 1, Supplementary table 1). Finally, there
79 are two rare teleost *TLRs*, i.e. – *TLR4* and *TLR26*. *TLR4* is found in the
80 Holocentriformes and in 3 out of 4 Beryciformes species in addition to *Danio rerio*,
81 *Polymixia japonica* and *Guentherus altivela*. *TLR26* is mainly found in species basal to
82 the Gadiformes and in two Beryciformes: *Rondeletia loricata* and *Beryx splendens*
83 (Figure 1, Supplementary table 1). Overall, vertebrate and teleost genome
84 duplications may explain some of the teleost *TLR* repertoire variation demonstrated
85 here with respect to gene expansions. However, the extreme numbers seen for some
86 of the *TLR* expansions within the Gadiformes indicate that these genes have
87 undergone additional lineage-specific duplication events — a phenomenon also seen
88 for other genes in teleost species [9]. Gene duplicates preserved after a duplication
89 event commonly undergo neo- or subfunctionalization [10 and references therein]. In
90 Atlantic cod, we have previously demonstrated that the *TLR* expansions and their
91 paralogs show signs of diversifying selection. For some expansions, this was
92 indicative of neofunctionalization due to high numbers of sites under selection in
93 likely dimerization and ligand-interacting regions. For other expansions it was more
94 indicative of subfunctionalization due to fewer sites under selection combined with
95 tissue-specific expression patterns [7 and references therein]. Therefore, our findings

96 suggest that neo-/subfunctionalization is the main mechanism resulting in the large
97 gene copy number of *TLRs* present within the Gadiformes.

98 Associations between specific *TLR* expansions, species latitudinal distributions,
99 species maximum depth as well as the absence of *MHCII* — specific for the
100 Gadiformes lineage (Figure 1) — were further investigated using Stochastic Linear
101 Ornstein-Uhlenbeck Models for Comparative Hypotheses (SLOUCH) [11]. Models
102 using the specified latitudinal categories as predictor variables showed that latitude
103 explained 19-32 % of the *TLR* copy number variation for *TLR8*, *TLR9*, *TLR22* and
104 *TLR25* whereas species maximum depth explained 4-10 % of the variation seen in
105 *TLR9* and *TLR22* (Supplementary information). Especially northern latitudinal
106 categories were found to be associated with higher copy numbers in *TLR8*, *TLR22*
107 and *TLR25*, while increased copy numbers in *TLR9* were associated with more
108 tropical latitudes, particularly in the equatorial region (Table 1, Supplementary table
109 1). However, for *TLR23* there was no indication that the copy number has evolved as
110 a consequence of changes in latitude (Table 1). Moreover, within the Gadiformes
111 lineage we found strong support for scenarios where *TLR8*, *TLR9*, *TLR22* and *TLR25*
112 have evolved additional gene copies with the loss of *MHCII* explaining between 14-
113 27 % of the copy number variation (Table 2). The explained variation in copy
114 numbers was 3-6 % larger (compared to latitude alone) and 3-16 % larger (compared
115 to *MHCII* loss alone) when we ran models where copy numbers of*TLR8*, *TLR22* and
116 *TLR25* evolved towards optima jointly defined by latitudinal categories and
117 presence/absence of *MHCII*. This indicates that both latitude and loss of *MHCII* have
118 contributed to the expansion of these *TLRs.* However, we were not able to
119 distinguish the relative contribution of *MHCII* and latitude, respectively. This is
120 contrary to the striking result obtained for *TLR9* where the combination of latitude

121 and loss of *MHCII* explained 50 % of the copy number variation – compared to 20 %
122 and 22 % for latitude and *MHCII* loss separately (Table 2).

123 Extreme northern or southern distributions, here given by latitudinal coordinates,
124 are proxy indicators for temperature as these regions are cooler but also have
125 undergone a larger degree of paleoclimatic changes compared to the more tropical
126 regions [12]. The latitudinal species richness gradient, however, reverse-complement
127 this climatic pattern by showing a larger number of species in tropical areas likely
128 reflecting the more stable paleoclimatic conditions in this region [12 and references
129 therein]. Thus, for the species inhabiting the northern hemisphere, the observed
130 expansions for*TLR8*, *TLR22* and *TLR25* indicate selection towards higher copy
131 number optima. This could be explained by different pathogen loads or pathogen
132 community compositions connected to highly variable paleoclimatic arctic
133 environment. On the other hand, the expansion discovered for *TLR9* showed higher
134 optimal copy number in tropical regions (Table 1) most likely driven by the specific
135 biotic or abiotic factors encountered in the tropics. Collectively, our findings indicate
136 that, for the Gadiformes, both the loss of *MHCII* as well as paleogeographic
137 distribution reflecting the environments these species have inhabited through time,
138 have been vital drivers in particular for the expansion of *TLR9* but also for *TLR8*,
139 *TLR22* and *TLR25*.

140 The dated phylogeny shows that the successive alterations to the teleost immune
141 system occurred in periods with substantial paleoclimatic fluctuations as well as
142 oceanic changes due to continental drift. Such events are often associated with
143 periods of extinction followed by population diversification and subsequent
144 speciation enabling the invasion of new niches [13, 14]. Our data suggests that the
145 overall loss of *TLR5* and *TLR21beta* (175-130 mya) overlap the Jurassic-Cretaceous (J-
146 K) boundary (Figure 1). Although this transition between geological periods does

147 not harbour a well-defined extinction event, there is accumulating evidence of both
148 coinciding species extinctions and radiations at the J-K boundary [15-19]. The loss of
149 *TLR5* and *TLR21beta* therefore may have occurred during adaptation to new habitats
150 such as the expanding Central Atlantic Ocean by the ancestors of Paracanthopterygii
151 and Lampridiformes [20]. The increase in marine fish species family richness
152 overlapping the J-K boundary, indirectly derived from fossil data [21], also implies
153 that these gene losses promoted new adaptations and species radiations among the
154 ancestral teleosts.

155 Within the Gadiformes clade we find that the loss of *MHCII* coincides with the
156 overall gene expansion patterns of *TLR7*, *TLR8*, *TLR9*, *TLR22*, *TLR23* and *TLR25*,
157 spanning a total interval 110-64 mya. This further overlaps with the early-late
158 Cretaceous transition which includes one of the late Cretaceous global anoxia events
159 (95 mya). This anoxic environment, although likely allowing a small degree of
160 specialized adaptation, generally deprived the deep seas of species [22, 23]. Anoxic
161 conditions led to higher extinction rates during this time period [24-27], fitting with
162 the metabolic cost scenario proposed to promote the loss of *MHCII* [28]. In this
163 scenario the benefits of maintaining the MHCII system in some environments could
164 not compensate for the metabolic cost of expressing it. Coinciding with the anoxic
165 event is the further northward opening of the Central Atlantic Ocean [20] and the
166 propagation of the South Atlantic Ocean to meet the Central Atlantic Ocean [29-31].
167 The stress imposed by global ocean anoxia therefore appears simultaneously with
168 the appearance of new habitats. Further, this time period is associated with a
169 decrease in bony fish family richness, indirectly derived from fossil data [21],
170 indicating that these secondary changes to the Gadiformes immune system may
171 have had slightly more adverse effects here compared to the initial ones occurring at
172 the J-K boundary. However, this likely had a positive effect supporting species

173 survival and radiation in the long term. The more recent loss of *TLR1/2* from the
174 Gadinae subfamily (40 – 16 mya) is likely a temperature-driven adaptation caused
175 by an abrupt cooling of global climate and loss of habitat due to the drastic decrease
176 of eustatic sea levels ~ 34 mya [27, 32, 33] overlapping with the opening of the North
177 Atlantic Ocean between Greenland and Norway [20].

178 Overall, our findings reveal unprecedented variability within the teleost innate
179 immune system and particularly within the Gadiformes. The successive nature of
180 these changes to the ancestral teleost immune system combined with the extensive
181 evolvability of the innate immune system described here have likely contributed to
182 the overall survival and successful radiation of this lineage.

183 Materials and methods

184 Sequencing and assembly summary

185 The 66 teleost genomes and species phylogeny were generated by Malmstrøm et al.
186 [3] In short DNA was isolated from 66 teleost species and subjected to Illumina
187 HiSeq sequencing (2 x 150 bp paired-end reads) which after trimming resulted in an
188 overall coverage between 9 and 34X. The genomes were assembled using the Celera
189 Assembler. For the phylogenetic reconstruction 9 reference fish species were added
190 from Ensembl together with *Salmo salar*. An alignment of 71,418 bp was used as
191 input for phylogenetic reconstruction with the Bayesian software BEAST [34].

192 The phylogeny was made using the Bayesian software BEAST combined with fossil
193 time-calibration. Teleost *TLR* characterization was performed using BLAST against
194 all 76 teleost species depicted in the phylogeny. Conserved Toll/interleukin-1
195 receptor (TIR) domain protein sequences from *TLRs* annotated in all Ensembl fish
196 species and *TLRs* described by Solbakken et al. [7] representing all known *TLRs* to
197 date were used as queries.

198 Note: all timings derived from the phylogeny presented in this study are branch
199 range times to illustrate the uncertainty underlying the fossil calibration performed
200 by Malmstrøm et al. [3] and thus are longer than the branches depicted in the
201 phylogeny (Figure 1).

202 **Gene searches**

203 Protein query TIR domain sequences from Atlantic cod [7],all fish genomes available
204 at Ensembl [35] and channel catfish [36], collectively representing all known
205 vertebrate *TLR* genes to date, were used for TBLASTN searches towards the 66 fish
206 genomes supplied by Malmstrøm et al. *TLR* copy numbers for the Ensembl species
207 were taken from Solbakken et al [7].The NCBI BLAST tool was used to search the
208 *Salmo salar* genome (ICSASG_v2, GCA_000233375.4) with default settings using the
209 same query sequences. TBLASTN from Blast+ 2.2.26 [37] was used with an e-value
210 cut-off at 1e-10 and in some cases also lower to capture the largest gene expansions.
211 The number of detected TIR domains was counted for each *TLR* gene. Due to the
212 fragmented nature of some of the genomes conservative estimates of copy numbers
213 have been added to Supplementary table 1. These copy numbers form the
214 foundation for the *TLR* repertoires depicted in Figure 1.

215 Note on gene annotation: *TLR* gene annotation varies greatly between species. In this
216 study the following annotations are used (similar to that of Solbakken et al. [7]):
217 *TLR1*, *TLR1/6* (in cases where annotation has not been provided and phylogeny
218 cannot determine stronger homology towards *TLR1* or *TLR6*), *TLR2*, *TLR3*, *TLR4*,
219 *TLR5*, *TLR6*, *TLR7*, *TLR8*, *TLR9*, *TLR10*, *TLR11*, *TLR12*, *TLR13*, *TLR14*, *TLR15*, *TLR16*,
220 *TLR18* is by phylogeny determined to be *TLR14*, *TLR15*, *TLR16*, *TLR19* is by
221 phylogeny determined to be *TLR26*, *TLR20* is by phylogeny determined to be *TLR26*,
222 *TLR21*, *TLR22*, *TLR23*, *TLR25* and *TLR26*.

223 ***TLR, MHC,* latitude and depth correlations using SLOUCH**

224 For genes displaying more than 4 different gene copy numbers (*TLR8*, *TLR9*, *TLR22*,
225 *TLR23*, *TLR25*) we ran SLOUCH — Stochastic Linear Ornstein-Uhlenbeck Models for
226 Comparative Hypotheses. This is a phylogenetic comparative method designed to
227 study adaptive evolution of a trait along a phylogeny currently implemented in the
228 R program SLOUCH [11, 38, 39]. The output of models analysed in SLOUCH can be
229 summarized by a regression, which includes information on whether the analysed
230 traits are evolving towards the estimated optima, how fast (or slow) this evolution is,
231 and how much of the trait variation that is explained by evolution towards these
232 optima. We used SLOUCH to test whether *TLR* copy numbers have evolved towards
233 optima that are influenced by the species' latitudinal distribution (values obtained
234 from Fishbase.org [40]), species maximum depth (values obtained from Fishbase.org
235 [40]) and evolutionary losses of the *MHCII* complex. We defined 6 latitudinal
236 categories based on latitude 75, 50, 25, 0 (equator), -25 and -50. If a species'
237 latitudinal distribution includes or crosses one of these it was assigned to that
238 respective category (multiple assignments are possible). Some species were not
239 included in any of the categories due failure to cross the defined latitudes or where
240 data on depth was unavailable and thus were excluded from the phylogeny
241 resulting in a reduced tree.

242 The model of evolution in SLOUCH is based on an Ornstein–Uhlenbeck process and
243 assumes that a trait (e.g. gene copy number) has a tendency to evolve towards a
244 'primary' optimum Θ. We assume that average copy number in a lineage can take
245 any non-negative real number (i.e., intraspecies variation in copy numbers exist). A
246 primary optimum is defined as the average optimal state that species will reach in a
247 given environment when ancestral constraints have disappeared [38], at a rate
248 proportional to a parameter α. As an example, in some of our analyses, we

249 investigated whether species sharing the same latitudinal distribution have a
250 tendency to evolve similar copy numbers for a given *TLR* locus. Lag in adaptation
251 towards primary optima is quantified by a half-life parameter, $t_{1/2} = \ln(2)/\alpha$, which
252 can be interpreted as the average time it takes a species to evolve half the distance
253 from the ancestral (copy number) state towards the predicted optimal (copy number)
254 state. For example, a half-life of zero signifies immediate adaptation of the trait to
255 any change in the optimum for every lineage present in the phylogeny. A half-life
256 above zero indicates adaptation is not immediate, with the amount of constrained
257 evolution increasing with an increasing half-life. The model of evolution used in
258 SLOUCH also includes a stochastic component with standard deviation σ, which can
259 be interpreted as evolutionary changes in the trait (e.g. copy numbers) due to
260 unmeasured selective forces and genetic drift. This component of the model is
261 reported as $v_y = \sigma^2/2\alpha$, and can be interpreted as the expected residual variance when
262 adaptation and stochastic changes have come to an equilibrium.

263 Our latitudinal categories, maximum depth and evolutionary losses of *MHCII*
264 represent 'niches' and the model estimates one primary optimum for each niche
265 included in any particular model. The different states of niches (e.g. presence and
266 absence of *MCHII*) are known for all extant species in our phylogeny, but are
267 unobserved for internal branches in the tree. We therefore mapped a separate state
268 called *ancestral* to all internal nodes in the phylogeny to avoid having to infer
269 uncertain primary optima. The method uses generalized least squares for estimation
270 of the regression parameters (i.e., the influence of the predictor on the primary
271 optimum) and maximum likelihood for estimation of α and σ^2 in an iterative
272 procedure. For a full description of the model implemented in SLOUCH, see Hansen
273 et al. 2008. All analyses were performed in R version 3.0 [39].

274 We used SLOUCH to estimate the phylogenetic effect in the data. A phylogenetic

275 effect means that some part of the variation in the trait is explained by shared
276 ancestry (i.e. phylogeny), which means closely related species tend to have more
277 similar trait values compared to more distantly related species. The phylogenetic
278 effect can be estimated in SLOUCH by running a model without any predictor
279 variables (i.e. no latitudinal categorical variables). The half-life parameter in such a
280 model will represent an estimate for how important shared history is in explaining
281 the distribution of trait means on the phylogeny: A half-life of zero means the trait
282 data is not phylogenetically structured, while a half-life > 0 indicates that there exists
283 an influence of phylogeny on the data. A phylogenetic effect can be due to slowness
284 of adaptation, adaptation towards phylogenetically structured optima, or a
285 combination of both. To investigate which of these scenarios we find support for, we
286 contrasted the phylogenetic effect model with a model run with predictor variables
287 (e.g. latitudinal distribution) using AICc. A better (lower) AICc value for a model
288 including predictor variables indicate evidence for a scenario where the traits in our
289 models are evolving towards optima that are shared by species across niches (e.g.
290 the same latitudinal section).

291 **References**

292 1. Star, B., et al., *The genome sequence of Atlantic cod reveals a unique immune system.* Nature,
293 2011. **477**(7363): p. 207-10.
294 2. Haase, D., et al., *Absence of major histocompatibility complex class II mediated immunity in*
295 *pipefish, Syngnathus typhle: evidence from deep transcriptome sequencing.* Biol Lett, 2013.
296 **9**(2): p. 20130044.
297 3. Malmstrøm, M., et al., *Evolution of the immune system influences speciation rates in teleost*
298 *fishes.* Nat Genet, In press.
299 4. Palti, Y., *Toll-like receptors in bony fish: from genomics to function.* Dev Comp Immunol, 2011.
300 **35**(12): p. 1263-72.
301 5. Rebl, A., T. Goldammer, and H.M. Seyfert, *Toll-like receptor signaling in bony fish.* Vet
302 Immunol Immunopathol, 2010. **134**(3-4): p. 139-50.

6. Roach, J.C., et al., *The evolution of vertebrate Toll-like receptors.* Proc Natl Acad Sci U S A, 2005. **102**(27): p. 9577-82.
7. Solbakken, M.H., et al., *Evolutionary redesign of the Atlantic cod (Gadus morhua L.) Toll-like receptor repertoire by gene losses and expansions.* Sci Rep, 2016. **6**: p. 25211.
8. Olson, M.V., *When less is more: gene loss as an engine of evolutionary change.* Am J Hum Genet, 1999. **64**(1): p. 18-23.
9. Yang, L., et al., *Genome-wide identification, characterization, and expression analysis of lineage-specific genes within zebrafish.* BMC Genomics, 2013. **14**: p. 65.
10. Kassahn, K.S., et al., *Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates.* Genome Res, 2009. **19**(8): p. 1404-18.
11. Hansen, T.F., J. Pienaar, and S.H. Orzack, *A comparative method for studying adaptation to a randomly evolving environment.* Evolution, 2008. **62**(8): p. 1965-77.
12. Lawson, A.M. and J.T. Weir, *Latitudinal gradients in climatic-niche evolution accelerate trait evolution at high latitudes.* Ecol Lett, 2014. **17**(11): p. 1427-36.
13. Wellborn, G.A. and R.B. Langerhans, *Ecological opportunity and the adaptive diversification of lineages.* Ecol Evol, 2015. **5**(1): p. 176-95.
14. Simoes, M., et al., *The Evolving Theory of Evolutionary Radiations.* Trends Ecol Evol, 2016. **31**(1): p. 27-34.
15. Benson, R.B., et al., *Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates.* Proc Biol Sci, 2010. **277**(1683): p. 829-34.
16. Benson, R.B. and P.S. Druckenmiller, *Faunal turnover of marine tetrapods during the Jurassic-Cretaceous transition.* Biol Rev Camb Philos Soc, 2014. **89**(1): p. 1-23.
17. Bambach, R.K., *Phanerozoic biodiversity mass extinctions.* Annual Review of Earth and Planetary Sciences, 2006. **34**: p. 127-155.
18. Alroy, J., *The shifting balance of diversity among major marine animal groups.* Science, 2010. **329**(5996): p. 1191-4.
19. Cavin, L., *Diversity of Mesozoic semionotiform fishes and the origin of gars (Lepisosteidae).* Naturwissenschaften, 2010. **97**(12): p. 1035-40.
20. Melankholina, E.N. and N.M. Sushchevskaya, *Development of passive volcanic margins of the Central Atlantic and initial opening of ocean.* Geotectonics, 2015. **49**(1): p. 75-92.

21. Guinot, G. and L. Cavin, *'Fish' (Actinopterygii and Elasmobranchii) diversification patterns through deep time.* Biol Rev Camb Philos Soc, 2015.
22. Priede, I.G. and R. Froese, *Colonization of the deep sea by fishes.* Journal of Fish Biology, 2013. **83**(6): p. 1528-1550.
23. Rogers, A.D., *The role of the oceanic oxygen minima in generating biodiversity in the deep sea.* Deep-Sea Research Part Ii-Topical Studies in Oceanography, 2000. **47**(1-2): p. 119-148.
24. Takashima, R., et al., *Greenhouse World and the Mesozoic Ocean.* Oceanography, 2006. **19**.
25. Wilson, P.A. and R.D. Norris, *Warm tropical ocean surface and global anoxia during the mid-Cretaceous period.* Nature, 2001. **412**(6845): p. 425-9.
26. Sinninghe Damsté, J.S., et al., *A CO2 decrease-driven cooling and increased latitudinal temperature gradient during the mid-Cretaceous Oceanic Anoxic Event 2.* Earth and Planetary Science Letters, 2010. **293**(1–2): p. 97-103.
27. Harnik, P.G., et al., *Extinctions in ancient and modern seas.* Trends in Ecology & Evolution, 2012. **27**(11): p. 608-617.
28. Star, B. and S. Jentoft, *Why does the immune system of Atlantic cod lack MHC II?* Bioessays, 2012. **34**(8): p. 648-51.
29. Granot, R. and J. Dyment, *The Cretaceous opening of the South Atlantic Ocean.* Earth and Planetary Science Letters, 2015. **414**: p. 156-163.
30. Voigt, S., et al., *Tectonically restricted deep-ocean circulation at the end of the Cretaceous greenhouse.* Earth and Planetary Science Letters, 2013. **369**: p. 169-177.
31. Murphy, D.P. and D.J. Thomas, *The evolution of Late Cretaceous deep-ocean circulation in the Atlantic basins: Neodymium isotope evidence from South Atlantic drill sites for tectonic controls.* Geochemistry Geophysics Geosystems, 2013. **14**(12): p. 5323-5340.
32. Liu, Z., et al., *Global cooling during the eocene-oligocene climate transition.* Science, 2009. **323**(5918): p. 1187-90.
33. Goldner, A., N. Herold, and M. Huber, *Antarctic glaciation caused ocean circulation changes at the Eocene-Oligocene transition.* Nature, 2014. **511**(7511): p. 574-7.
34. Drummond, A.J., et al., *Bayesian phylogenetics with BEAUti and the BEAST 1.7.* Mol Biol Evol, 2012. **29**(8): p. 1969-73.
35. Cunningham, F., et al., *Ensembl 2015.* Nucleic Acids Res, 2015. **43**(Database issue): p. D662-9.
36. Benson, D.A., et al., *GenBank.* Nucleic Acids Res, 2013. **41**(Database issue): p. D36-42.

365 37. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p.
366 421.
367 38. Hansen, T.F., *Stabilizing selection and the comparative analysis of adaptation.* Evolution,
368 1997. **51**(5): p. 1341-1351.
369 39. R-Core-Team. *R: A Language and Environment for Statistical Computing*. 2015; Available
370 from: http://www.R-project.org.
371 40. Froese, R. and D. Pauly. 2015; Available from: www.fishbase.org.

372 Acknowledgement

373 This work was supported by The Research Council of Norway (Grant numbers
374 222378/F20 and 199806/S40 to KSJ/SJ).Some of the teleost genomes were assembled
375 using the Abel Cluster, owned by the University of Oslo and the Norwegian
376 metacenter for High Performance Computing (NOTUR), and operated by the
377 Department for Research Computing at USIT, the University of Oslo IT-department
378 http://www.hpc.uio.no/. The authors would like to thank Dr. Mark Ravinet for
379 valuable comments.

380 Author information

381 Affiliations

382 MHS, KLV, KSJ and SJ are affiliated with the Department of Biosciences, Centre for
383 Ecological and Evolutionary Synthesis (CEES), University of Oslo, Oslo, Norway. SJ
384 is also affiliated with the Department of Natural Sciences, University of Agder,
385 Kristiansand, Norway.

386 Author contributions

387 MHS performed BLAST searches. KLV performed all SLOUCH analyses. SJ and KSJ
388 provided the overall evolutionary context. MHS made all figures/tables and wrote
389 the overall text with significant aid of SJ and KSJ. KLV wrote all sections related to

SLOUCH. All authors contributed with comments, edits and proofreading of the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Corresponding author

Correspondence and requests for materials should be addressed to MHS: m.h.solbakken@ibv.uio.no and SJ: sissel.jentoft@ibv.uio.no.

398 ## Figures and tables

399 **Table 1 Phylogenetic comparative analyses of the evolution of *TLR* copy numbers in**
400 **relation to species latitudinal distributions using SLOUCH.** For each model, we show the
401 phylogenetically corrected r^2, and the AICc score. Lower AICc scores indicate a better model.
402 Detailed output from each model is given in supplementary information. The model called
403 "phylogeny" does not include any explanatory variables and is given as a reference point for
404 comparison to models with predictor variables.

	TLR8		***TLR9***		***TLR22***		***TLR23***		***TLR25***	
Category	**AICc**	**r2**	**AICc**	**r2**	**AICc**	**r2**	**AICc**	**r2**	**AICc**	**r2**
Phylogeny	266.41	0.00	243.91	0.00	430.27	0.00	307.65	0.00	241.36	0.00
Group 75 latitude	260.29	18.32	239.07	18.91	**418.86**	**24.63**	311.72	0.96	**226.61**	**32.26**
Group 50 latitude	**259.75**	**19.02**	240.67	15.49	427.26	13.70	310.88	2.30	232.46	21.96
Group 25 latitude	259.98	18.72	240.34	17.24	429.86	8.86	**307.22**	**7.91**	233.31	20.89
Group 0 latitude	259.90	20.13	**238.24**	**19.99**	427.05	13.99	311.27	1.69	232.56	21.84
Group -25 latitude	260.06	18.63	239.78	16.69	429.34	9.62	309.78	4.00	233.38	20.80
Group -50 latitude	260.31	16.35	240.16	16.18	429.62	9.21	311.54	1.24	233.45	20.71

405

406

407 **Table 2 Phylogenetic comparative analyses of the evolution of *TLR* copy numbers in**
408 **relation to species latitudinal distributions and *MHCII* status using SLOUCH.** For each
409 model, we show the phylogenetically corrected r^2, and the AICc score. Lower AICc scores
410 indicate a better model. Detailed output from each model is given in supplementary
411 information. The model called "phylogeny" in Table 1 does not include any explanatory
412 variables and is given as a reference point for comparison to models with predictor variables.

	TLR8		*TLR9*		*TLR22*		*TLR23*		*TLR25*	
Category	**AICc**	**r2**	**AICc**	**r2**	**AICc**	**r2**	**AICc**	**r2**	**AICc**	**r2**
Group MHCII	259.43	19.44	239.01	22.41	427.99	14.53	328.13	2.65	231.30	26.94
Group MHCII + Group 75 lat.	264.37	19.52	240.39	31.32	**420.41**	**30.23**	315.34	3.21	**228.98**	**35.08**
Group MHCII + Group 50 lat.	**262.31**	**22.16**	243.11	25.72	431.07	17.15	314.69	4.23	235.60	27.76
Group MHCII + Group 25 lat.	263.76	20.32	239.13	32.69	431.69	16.32	**311.21**	**9.41**	234.82	28.67
Group MHCII + Group 0 lat.	**261.00**	**27.11**	**228.48**	**53.53**	430.16	18.36	314.90	3.92	233.63	30.02
Group MHCII + Group -25 lat.	263.37	20.82	230.06	52.33	432.07	15.80	313.49	6.01	234.84	28.64
Group MHCII + Group -50 lat.	264.39	19.50	240.15	29.17	432.19	15.63	314.92	3.81	235.60	27.76

413

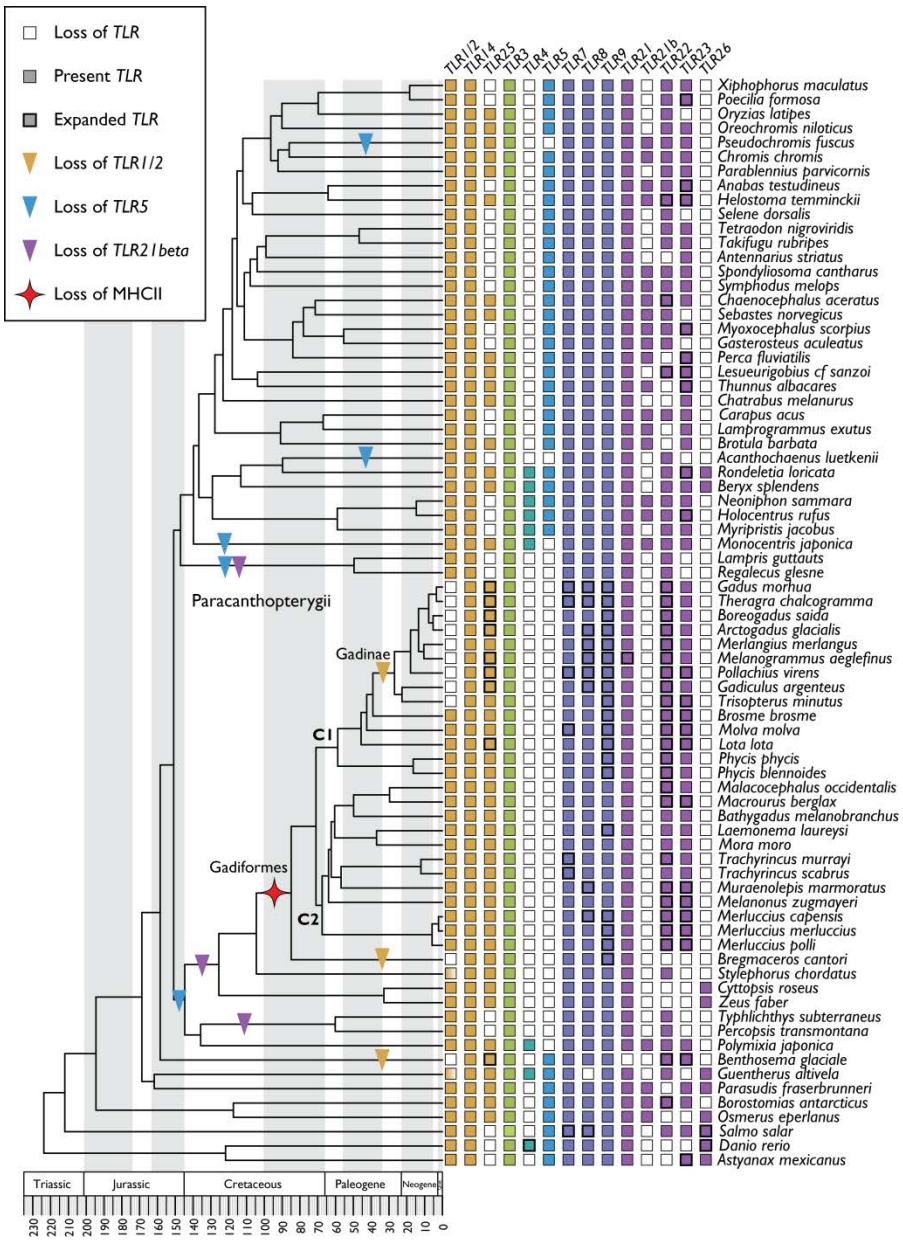


Figure 1 The *TLR* repertoires of 76 teleosts mapped onto a time-calibrated species phylogeny. All *TLRs* characterized in the new 66 teleost genomes as well as in 10 reference teleosts genomes (Ensembl and GenBank) mapped onto a species phylogeny generated by Malmstrøm et al. The phylogeny demonstrates the loss of *MHCII* 110-64 mya (branch range time, bright red star) reported by Malmstrøm et al. Lineage-specific *TLR* losses are marked by arrows (ochre for *TLR1/2*, blue for *TLR5* and violet for *TLR21beta*). The individual species' repertoires are depicted with boxes

422 where the coloration represents the six major *TLR* families: *TLR1-family* (orange),
423 *TLR3-family* (green), *TLR4-family* (turquoise), *TLR5-family* (blue), *TLR7-family* (indigo)
424 and *TLR11-family* (violet). Filled boxes indicate presence, empty boxes indicate
425 absence, boxes with bold borders indicate gene expansions with 3 copies or more
426 (see Supplementary table 1) and for *TLR1/2* a gradient-filled box indicates the
427 presences of either *TLR1* or *TLR2*.

428

429 **Supplementary table 1 Overview of *TLR* copy number and northern and southern latitude boundaries collected from**

430 **Fishbase.**

Latin Name	TLR 1/6	TLR 2	TLR 3	TLR 4	TLR 5	TLR 7	TLR 8	TLR 9	TLR 14	TLR 21	TLR 21_beta	TLR 22	TLR 23	TLR 25	TLR 26	North.	South.	Average
Arctogadus glacilis	0	0	1	0	0	1	3	7	1	2	0	22	1	4	0	87	69	78
Boreogadus saida	0	0	1	0	0	1	1	6	1	1	0	16	1	10	0	87	52	69.5
Trisopterus minutus	0	0	1	0	0	2	1	8	1	2	0	6	3	1	0	66	28	47
Pollachius virens	0	0	1	0	0	3	5	5	1	1	0	15	3	7	0	77	33	55
Melanogrammus aeglefinus	0	0	1	0	0	1	12	5	1	3	0	6	2	4	0	79	35	57
Merlangius merlangus	0	0	1	0	0	2	7	6	1	1	0	11	1	2	0	72	35	53.5
Theragra chalcogramma	0	0	1	0	0	3	7	5	1	1	0	27	1	4	0	68	34	51
Gadiculus argentus	0	0	1	0	0	1	9	10	1	2	0	7	2	5	0	74	24	49
Phycis phycis	1	1	1	0	0	1	1	5	1	2	0	8	2	1	0	45	13	29
Molva molva	1	1	1	0	0	3	1	5	1	2	0	10	3	1	0	75	35	55
Lota lota	1	1	1	0	0	1	1	5	1	2	0	5	3	3	0	78	40	59

Brosme brosme	1	1	1	0	0	1	1	11	1	2	0	9	4	1	0	83	37	60
Merluccius merluccius	1	1	1	0	0	2	1	4	1	2	0	4	5	1	0	76	18	47
Merluccius capensis	1	1	1	0	0	1	4	5	1	2	0	6	4	2	0	-11	-37	-24
Merluccius polli	1	1	1	0	0	1	1	3	1	2	0	5	6	1	0	29	-19	5
Melanonus zugmayeri	1	1	1	0	0	1	1	1	1	1	0	13	6	1	0	60	-49	5.5
Macrourus berglax	1	1	1	0	0	1	2	1	1	1	0	16	3	1	0	82	37	59.5
Malacocephalus occidentalis	1	1	1	0	0	2	1	1	1	1	0	8	2	1	0	43	-37	3
Bathygadus melanobranchus	1	1	1	0	0	1	1	1	1	1	0	2	2	1	0	53	-34	9.5
Muraenolepis marmoratus	1	1	1	0	0	1	3	1	1	1	0	4	5	1	0	-44	-56	-50
Bregmaceros cantori	0	0	1	0	0	1	1	3	1	1	0	0	0	1	0			
Mora moro	1	1	1	0	0	1	1	1	1	1	0	1	1	1	0	64	-51	6.5
Laemonema laureysi	1	1	1	0	0	1	1	4	1	1	0	1	1	1	0	8	-8	0
Polymixia japonica	1	1	1	1	0	1	1	1	1	1	1	2	2	1	0	40	6	23
Percopsis transmontana	1	1	1	0	0	1	1	1	1	1	0	1	0	0	0	44	43	43.5

Typhlichthys subterraneus	1	1	1	0	0	1	1	1	1	1	0	1	0	0	0	39	34	36.5
Zeus faber	1	1	1	0	0	1	1	1	1	1	0	0	0	1	1	75	-49	13
Cyttopsis roseus	1	1	1	0	0	1	1	1	1	1	0	0	0	1	1			
Lamprogrammus exutus	1	1	1	0	1	1	1	1	1	1	1	0	2	0	0	12	-23	-5.5
Brotula barbata	1	1	1	0	1	1	1	1	1	1	1	0	2	1	0	30	-14	8
Carapus acus	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	42	-15	13.5
Myripristis jacobus	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	37	-23	7
Holocentrus rufus	1	1	1	1	1	1	1	1	1	1	1	1	3	0	0	33		33
Trachyrincus scabrus	1	1	1	0	0	7	1	1	1	1	0	1	2	1	0	55	-27	14
Chatrabus melanurus	1	1	1	0	1	1	1	1	1	1	0	0	1	1	0	-35		-35
Parasudis fraserbrunneri	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	21		21
Regalecus glesne	1	1	1	0	0	1	1	1	1	1	0	1	0	0	0	72	-52	10
Lampris guttatus	1	1	1	0	0	1	1	1	1	1	0	1	0	0	0	70	-45	12.5
Guentherus altivela	0	1	1	1	1	1	0	1	1	1	0	1	0	1	1			

Antennarius striatus	1	1	1	0	1	1	1	1	1	1	0	0	1	0	0	43	-50	-3.5
Osmerus eperlanus	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	70	43	56.5
Perca fluviatilis	1	1	1	0	1	1	1	1	1	1	1	0	17	1	0	74	38	56
Sebastes norvegicus	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	79	38	58.5
Chaenocephalus aceratus	1	1	1	0	1	1	1	1	1	1	1	3	1	1	0	-53	-65	-59
Borostomias antarcticus	1	1	1	0	1	1	1	1	1	1	1	4	1	1	0	66	-66	0
Benthosema glaciale	0	0	1	0	1	1	1	2	1	0	0	49	8	4	0	81	11	46
Rondeletia loricata	1	1	1	1	1	1	1	1	1	1	0	2	3	1	1	67	-42	12.5
Beryx splendens	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	45	-43	1
Neoniphon sammara	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	30	-30	0
Monocentris japonica	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0			
Acanthochaenus luetkenii	1	1	1	0	0	1	1	1	1	2	0	1	0	0	0	40	-57	-8.5
Stylephorus chordatus	1	0	1	0	0	1	1	2	1	1	0	1	0	1	0	45	-37	4
Spondyliosoma cantharus	1	1	1	0	1	1	1	1	1	1	1	1	2	0	0	63	-20	21.5

Thunnus albacares	1	1	1	0	1	1	1	1	1	1	1	0	7	1	0	52	-45	3.5
Helostoma temminckii	1	1	1	0	1	1	1	1	1	1	1	3	14	1	0	16	-6	5
Anabas testudineus	1	1	1	0	1	1	1	1	1	1	1	2	3	0	0	28	-10	9
Selene dorsalis	1	1	1	0	1	1	1	1	1	1		2	0	0	0	39	-28	5.5
Chromis chromis	1	1	1	0	1	1	1	1	1	1	1	2	2	1	0	46	-12	17
Parablennius parvicornis	1	1	1	0	1	1	1	1	1	1	0	1	1	1	0	36	-6	15
Symphodus melops	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	63	28	45.5
Pseudochromis fuscus	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	26	-24	1
Myoxocephalus scorpius	1	1	1	0	1	1	1	1	2	1	1	2	4	0	0	80	40	60
Trachyrincus murrayi	1	1	1	0	0	3	1	1	1	1	0	3	2	1	0			
Phycis blennoides	1	1	1	0	0	1	1	7	1	2	0	5	1	1	0	71	20	45.5
Lesueurigobius cf sanzoi	1	1	1	0	1	1	1	1	1	1	0	3	7	2	0	42	-21	10.5
Gadus morhua	0	0	1	0	0	4	8	5	1	1	0	12	1	6	0	83	35	59
Astyanax mexicanus	1	1	1	0	1	1	2	1	1	1	0	0	3	0	1	36	24	30

Danio rerio	1	1	1	3	2	1	2	1	1	1	0	0	0	0	4	33	8	20.5
Gasterosteus aculeatus	1	1	1	0	1	1	1	1	1	2	1	1	0	0	0	71	26	48.5
Oreochromis niloticus	0	1	1	0	1	1	1	1	1	1	0	1	2	1	0	32	10	21
Oryzias latipes	1	1	1	0	1	1	1	1	1	1	0	1	0	1	0	55	10	32.5
Poecilia formosa	1	1	1	0	1	1	1	2	1	1	0	1	3	0	0	27	25	26
Takifugu rubripes	1	1	1	0	1	1	1	1	1	1	0	1	1	0	0	46	21	33.5
Tetraodon nigroviridis	1	1	1	0	1	1	1	1	1	1	0	1	1	0	0			
Xiphophorus maculatus	1	2	1	0	1	1	1	1	1	1	0	2	1	0	0	23	17	20
Salmo salar	2	2	2	0	2	4	4	2	2	2	0	2	2	0	4	72	37	54.5

BEST ESTIMATES & MODEL FIT – TLR8 group 50
MODEL PARAMETERS

	Estimate
Rate of adaptation	16.9060288
Phylogenetic half-life	0.0410000
Phylogenetic correction factor	0.9408495
Stationary variance	3.3300000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	6.0220382	1.2377241
NO	0.9751567	0.3641890
YES	1.3916402	0.3719837

MODEL FIT

	Value
Support	-124.34093
AIC	258.68187
AICc	259.75330
SIC	269.31754
r squared	19.02160
SST	76.67101
SSE	62.08696

> model.fit(ancestor, indata2$time, seq(0.001,0.1, 0.01), seq(3.3, 3.7, 0.01), response= indata2$TLR8, me.response= NULL, fixed.fact=Group_50, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR9 group 0
MODEL PARAMETERS

	Estimate
Rate of adaptation	4.3052620
Phylogenetic half-life	0.1610000
Phylogenetic correction factor	0.7708612
Stationary variance	2.6900000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	4.90548398	0.9785545
NO	1.03000046	0.5615206
YES	-0.02680546	0.4932277

MODEL FIT

	Value
Support	-113.58672
AIC	237.17343
AICc	238.24486
SIC	247.80910
r squared	19.98517
SST	77.39867
SSE	61.93042

> model.fit(ancestor, indata2$time, seq(0.001,0.24, 0.01), seq(2.4, 3, 0.01), response= indata2$TLR9, me.response= NULL, fixed.fact=Group_0, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR22 group 75

MODEL PARAMETERS

	Estimate
Rate of adaptation	17.3286795
Phylogenetic half-life	0.0400000
Phylogenetic correction factor	0.9422922
Stationary variance	43.3500000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	14.896336	4.502874
NO	1.840849	1.025818
YES	10.526348	2.283833

MODEL FIT

	Value
Support	-203.89235
AIC	417.78470
AICc	418.85613
SIC	428.42038
r squared	24.63373
SST	82.27239
SSE	62.00563

> model.fit(ancestor, indata2$time, seq(0.01,0.07, 0.01), seq(43.1, 43.6, 0.01), response= indata2$TLR22, me.response= NULL, fixed.fact=Group_75, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR25 group 75

MODEL PARAMETERS

	Estimate
Rate of adaptation	23.1049060
Phylogenetic half-life	0.0300000
Phylogenetic correction factor	0.9567191
Stationary variance	1.9300000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	5.2370077	1.0103873
NO	0.6168722	0.2122440
YES	2.1608119	0.4634191

MODEL FIT

	Value
Support	-107.77088
AIC	225.54177
AICc	226.61319
SIC	236.17744
r squared	32.25999
SST	91.51601
SSE	61.99296

> model.fit(ancestor, indata2$time, seq(0.01,0.08, 0.01), seq(1.8, 2, 0.01), response= indata2$TLR25, me.response= NULL, fixed.fact=Group_75, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR8 group 50 + MHCII

MODEL PARAMETERS

	Estimate
Rate of adaptation	16.9060288
Phylogenetic half-life	0.0410000
Phylogenetic correction factor	0.9408495
Stationary variance	3.2100000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	5.8463339	1.2807175
NoNO	0.4190979	0.9921280
NoYES	2.0202578	0.5889784
YesNO	1.0768274	0.3831597
YesYES	0.9711968	0.4654550

MODEL FIT

	Value
Support	-123.11637
AIC	260.23274
AICc	262.30681
SIC	275.12268
r squared	22.15806
SST	79.53722
SSE	61.91331

> model.fit(ancestor, indata3$time, seq(0.001,0.1, 0.01), seq(3, 3.6, 0.01), response= indata3$TLR8, me.response= NULL, fixed.fact=Group_50_MHCII , fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR8 group 0 + MHCII

MODEL PARAMETERS

	Estimate
Rate of adaptation	22.3595865
Phylogenetic half-life	0.0310000
Phylogenetic correction factor	0.9552765
Stationary variance	3.1000000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	5.5081865	1.3379683
NoNO	2.7748094	0.6246435
NoYES	0.7863605	0.7210838
YesNO	1.1396624	0.4709201
YesYES	0.9948441	0.3675523

MODEL FIT

	Value
Support	-122.46304
AIC	258.92607
AICc	261.00015
SIC	273.81601
r squared	27.11397
SST	85.15327
SSE	62.06484

> model.fit(ancestor, indata3$time, seq(0.001,0.1, 0.01), seq(3, 3.6, 0.01), response= indata3$TLR8, me.response= NULL, fixed.fact=Group_0_MHCII , fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR9 group 0 + MHCII

MODEL PARAMETERS

	Estimate
Rate of adaptation	8.5573726
Phylogenetic half-life	0.0810000
Phylogenetic correction factor	0.8831642
Stationary variance	1.9400000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	5.2355351	0.9269600
NoNO	5.3595659	0.7952621
NoYES	0.6606602	0.7439025
YesNO	0.9759444	0.3906908
YesYES	0.8283597	0.3103422

MODEL FIT

	Value
Support	-106.20238
AIC	226.40476
AICc	228.47883
SIC	241.29470
r squared	53.52633
SST	133.68217
SSE	62.12702

> model.fit(ancestor, indata3$time, seq(0.001,0.24, 0.01), seq(1.5, 2.5, 0.01), response= indata3$TLR9, me.response= NULL, fixed.fact=Group_0_MHCII , fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR22 group 75 + MHCII

MODEL PARAMETERS

	Estimate
Rate of adaptation	23.1049060
Phylogenetic half-life	0.0300000
Phylogenetic correction factor	0.9567191
Stationary variance	40.5000000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	15.032341	4.794099
NoNO	5.026109	2.023606
NoYES	9.151258	2.856095
YesNO	1.109897	1.108735
YesYES	13.008162	3.184843

MODEL FIT

	Value
Support	-202.16960
AIC	418.33921
AICc	420.41328
SIC	433.22915
r squared	30.23004
SST	88.97227
SSE	62.07592

```
> model.fit(ancestor, indata3$time, seq(0.01,0.06, 0.01), seq(38, 40.5, 0.01), response= indata3$TLR22, me.response= NULL, fixed.fact=Group_75_MHCII , fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)
```

BEST ESTIMATES & MODEL FIT – TLR25 group 75 + MHCII

MODEL PARAMETERS

	Estimate
Rate of adaptation	23.1049060
Phylogenetic half-life	0.0300000
Phylogenetic correction factor	0.9567191
Stationary variance	1.8500000

PRIMARY OPTIMA

UNKNOWN mapped to the root of the tree and includes the coefficent for the ancestral state (Ya)

	Estimates	Std.error
UNKNOWN	4.8155471	1.0246259
NoNO	1.0017827	0.4324982
NoYES	2.7388718	0.6104231
YesNO	0.5119824	0.2369661
YesYES	1.4982955	0.6806854

MODEL FIT

	Value
Support	-106.45249
AIC	226.90499
AICc	228.97906
SIC	241.79493
r squared	35.08048
SST	95.47346
SSE	61.98091

```
> model.fit(ancestor, indata3$time,  seq(0.01,0.08, 0.01), seq(1, 4, 0.01),  response= indata3$TLR25, me.response= NULL, fixed.fact=Group_75_MHCII , fixed.cov= NULL, me.fixed.cov= NULL,  mecov.fixed.cov=NULL,  random.cov=  NULL,  me.random.cov=NULL, mecov.random.cov=NULL,  intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)
```

BEST ESTIMATES & MODEL FIT – TLR9 phylogenetic effect

MODEL PARAMETERS

	Estimate
Rate of adaptation	5.7762265
Phylogenetic half-life	0.1200000
Phylogenetic correction factor	0.6305712
Stationary variance	6.3700000

PRIMARY OPTIMA

	Estimate	Std.error
Theta_global	2.324661	0.7098559

MODEL FIT

	Value
Support	-1.183494e+02
AIC	2.426988e+02
AICc	2.430988e+02
SIC	2.491755e+02
r squared	-2.221451e-14
SST	6.397104e+01
SSE	6.397104e+01

model.fit(ancestor, time, seq(0,0.5, 0.01), seq(5, 10, 0.01), response= TLR9, me.response= NULL, fixed.fact=NULL, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR9 maximum depth

MODEL PARAMETERS

	Estimate
Rate of adaptation	4.3321699
Phylogenetic half-life	0.1600000
Phylogenetic correction factor	0.5517876
Stationary variance	6.1900000

	Maxdepth
Predictor theta	7.176028e+02
Predictor variance	2.127460e+07

PRIMARY OPTIMA

Evolutionary regression

	Estimate	Std. Error
Intercept	2.55726	0.76584
Maxdepth	-0.00052	0.00020

Optimal regression

	Estimate	Std. Error
K	2.40257	0.75533
Maxdepth	-0.00093	0.00036

Bias-corr. regression parameters

K	2.5572620431
Maxdepth	-0.0005178823

Decomposition of K assuming Ya = Xa to get the optimal regression intercept Bo

[1] 2.70278

(Use this as the intercept when plotting the regression line)

MODEL FIT

Support	-114.709044
AIC	237.418088

AICc 238.096054

SIC 246.053621

r squared 9.512911

SST 72.640560

SSE 65.730328

model.fit(ancestor, time, seq(0.05,0.3, 0.01), seq(6.1, 6.3, 0.01), response= TLR9, me.response= NULL, fixed.fact=NULL, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= Maxdepth, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR22 phylogenetic effect

MODEL PARAMETERS

	Estimate
Rate of adaptation	17.3286795
Phylogenetic half-life	0.0400000
Phylogenetic correction factor	0.8526366
Stationary variance	70.6000000

PRIMARY OPTIMA

	Estimate	Std.error
Theta_global	4.499373	1.638639

MODEL FIT

	Value
Support	-214.89433
AIC	435.78867
AICc	436.18867
SIC	442.26532
r squared	0.00000
SST	64.02347
SSE	64.02347

model.fit(ancestor, time, seq(0,0.1, 0.01), seq(60, 100, 0.1), response= TLR22, me.response= NULL, fixed.fact=NULL, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= NULL, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

BEST ESTIMATES & MODEL FIT – TLR22 maximum depth

MODEL PARAMETERS

	Estimate
Rate of adaptation	13.8629436
Phylogenetic half-life	0.0500000
Phylogenetic correction factor	0.8200624
Stationary variance	72.1100000

	Maxdepth
Predictor theta	7.176028e+02
Predictor variance	2.127460e+07

PRIMARY OPTIMA

Evolutionary regression

	Estimate	Std. Error
Intercept	2.89181	2.06756
Maxdepth	0.00176	0.00100

Optimal regression

	Estimate	Std. Error
K	2.96982	2.04058
Maxdepth	0.00217	0.00122

Bias-corr. regression parameters

K	2.891805411
Maxdepth	0.001762484

Decomposition of K assuming Ya = Xa to get the optimal regression intercept Bo

[1] 2.689926

(Use this as the intercept when plotting the regression line)

MODEL FIT

	Value
Support	-213.507480
AIC	435.014959
AICc	435.692926
SIC	443.650492
r squared	4.743322
SST	67.049714
SSE	63.869330

model.fit(ancestor, time, seq(0.03,0.06, 0.01), seq(72, 73, 0.01), response= TLR22, me.response= NULL, fixed.fact=NULL, fixed.cov= NULL, me.fixed.cov= NULL, mecov.fixed.cov=NULL, random.cov= Maxdepth, me.random.cov=NULL, mecov.random.cov=NULL, intercept="root", ultrametric=TRUE, support=NULL, convergence=NULL)

Paper III

Successive losses of central immune genes characterize the Gadiformes' alternate immunity

Monica H. Solbakken[1*], Matthew L. Rise[2], Kjetill S. Jakobsen[1], Sissel Jentoft[1,3*]

[1] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

[2] Department of Ocean Sciences, Memorial University of Newfoundland, Canada

[3] Department of Natural Sciences, University of Agder, Kristiansand, Norway

*Corresponding authors:

Monica Hongrø Solbakken. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. m.h.solbakken@ibv.uio.no, +47 22 84 41 64

Sissel Jentoft. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. sissel.jentoft@ibv.uio.no, +47 22 85 72 39

Abstract

Great genetic variability among teleost immunomes with gene losses and expansions of central adaptive and innate components has been discovered through genome sequencing over the last few years. Here, we demonstrate that the innate Myxovirus resistance gene (*Mx*) is lost from the ancestor of Gadiformes and the closely related *Stylephorus chordatus*, thus predating the loss of Major Histocompatibility Complex class II (*MHCII*) in Gadiformes. Although the functional implication of *Mx* loss is still unknown, we demonstrate that this loss is one of several ancient events appearing in successive order throughout the evolution of teleost immunity. In particular, we find that the loss of *Toll-like receptor 5* predates the loss of *Mx* involving the entire Paracanthopterygii lineage. Using a time-calibrated phylogeny we show that these losses overlap with major paleoclimatic and geological events indicating adaptive losses promoting survival and speciation in environments where maintaining these genes was less favourable.

Background

Comprehensive characterization of immune gene repertoires has, over the last decade, provided the scientific community with new discoveries that have challenged our perception of the evolution of vertebrate immunity. The detection of variable lymphocyte receptors in jawless vertebrates reveals an alternative adaptive immune system, lack of Major Histocompatibility (*MHC*) class II in Atlantic cod (*Gadus morhua*) and possibly in pipefish (*Syngnathus typhle*) indicate that classic adaptive immunity is more flexible than initially believed. Further, the discovery of different repertoires of central innate immunity genes reflects great plasticity in the vertebrate innate immune system [1-6]. Recently, Malmstrøm et al. demonstrated that the loss of central adaptive immunity components found in Atlantic cod [1] is a common immunological trait in the Gadiformes lineage [7]. They show that the *MHCII* pathway was lost

44 approximately 105 mya (million years ago) in the common ancestor of
45 Gadiformes. This was followed by an independent event resulting in the
46 expansion of *MHCI*. Moreover, in Atlantic cod, additional gene losses and
47 expansions within the central innate gene family of *Toll-like receptors (TLRs)* have
48 been reported [1]. This *TLR* repertoire has been found to be extreme compared to
49 other teleosts [8]. In this study we take advantage of the genome resources and
50 phylogeny generated by Malmstrøm et al. to further elucidate the evolutionary
51 origin of the immunological strategy common to Gadiformes and to infer our
52 findings in a broader paleontological perspective.

53 **Results and discussion**

54 **An ancient loss of *Mx***

55 We here show that the innate myxovirus resistance (*Mx*) gene is lost from the
56 Gadiformes and *Stylephorus chordatus*, and this predates the loss of *MHCII* (Figure
57 1). Further, we find that the copy number in teleost genomes harbouring *Mx*
58 overall lies between 1 and 3 with the exception of 7 in *Danio rerio* (SI table 1). For
59 15 of the 38 non-reference teleost genomes containing *Mx* partial synteny was
60 possible to obtain, and all are sharing the same *Mx* containing genomic region (SI
61 table 1). This partial synteny was then compared to the *Mx* genomic regions in
62 the fish reference genomes available as well as a selected number of vertebrates
63 [9]; with the exception of *Latimeria chalumnae* in which *Mx* could not be found. All
64 teleosts investigated with the exception of *Danio rerio* and *Astyanax mexicanus*
65 share local gene synteny. In *Danio rerio* we find 7 copies of *Mx* which are
66 distributed among four clusters in the genome (Table 1) where one of them shares
67 synteny with the *Mx* region in *Astyanax mexicanus.* Moreover, we find that
68 *Lepisosteus oculatus* share synteny with another of the identified *Mx* regions in
69 *Danio rerio*. *Petromyzon marinus*' single *Mx* is located on a short scaffold without
70 any similarity to the other species investigated. The *Mx* regions of *Homo sapiens*,
71 *Mus musculus*, *Gallus gallus*, *Anolis carolinensis and Xenopus tropicalis* share synteny.

However, these *Mx* regions are dissimilar to the *Mx* regions found in the investigated teleosts (Table 1). The synteny patterns demonstrated are likely related to the vertebrate genome duplications where different *Mx* genomic regions have been preserved while superfluous genetic material has been discarded throughout evolution [10]. Interestingly, we further find that the loss of *TLR5* reported in Atlantic cod [1] predates the loss of *Mx* as it affects the entire Paracanthopterygii and Lampridiformes lineages with a few additional species. Using the time calibrated phylogeny made by Malmstrøm et al. we were able to date the loss of *TLR5* to 151-147 mya (Figure 1).

The role of Mx in teleost immunity

Although the specific function of Mx is still unknown the diverse nature of its targets and responses between species indicate that *Mx* is under to strong selection and thus is important in vertebrate innate immunity. From mammals we know that *Mx* gene products are interferon-inducible dynamin-like large GTPases that block the early steps of virus replication [11]. Furthermore, Mx shows broad antiviral activity and the gene is usually present in two copies. However, the known diversity of antiviral targets and responses related to Mx does not correspond to the apparent copy number stability [12 and references therein]. *Mx* has been studied in various fish species like Atlantic salmon (*Salmo salar*), Atlantic halibut (*Hippoglossus hippoglossus*), gilthead seabream (*Sparus aurata*) and European Eel (*Anguilla anguilla*), and in these species showing similar function to mammalian Mx confirming a diverse range of Mx targets and responses also in fish [13-16]. In gilthead seabream the three variants of *Mx* responds to both RNA and DNA viruses from different families *in vitro*. However, this species' response towards DNA viruses cannot be replicated in other fish species [15 and references therein]. Strong diversifying selection combined with lineage-specific exchanges between paralogs conserving key enzymatic and structural characteristics, as well as acquiring new antiviral specificities, have

100 been proposed as the underlying mechanisms [12 and references therein]. A
101 single study reports Mx in Atlantic cod using a cross-reactive polyclonal antibody
102 generated against Atlantic salmon Mx [17]. Conversely in this study, we have
103 demonstrated a loss of *Mx* in Atlantic cod as well as the Gadiformes and
104 *Stylephorus chordatus* (Figure 1). The loss of *Mx* shown is in accordance with the
105 proposed lineage-specific adaptation of *Mx* [15 and references therein], whereas a
106 loss instead of diversifying selection of *Mx* as seen in other species. In a recent
107 publication, Braun et al. (2015) reported on the discovery of an evolutionary loss
108 of function of Mx for toothed whales, where it was suggested that
109 pseudogenization of *Mx* hinder entry of virus particles into host cells, i.e.
110 protecting the ancestral toothed whale species against harmful virus outbreaks
111 [18]. Cumulatively, these findings fit the scenario that lineage-specific gene loss
112 events are adaptive responses towards changes in a species' environment [19].

113 **Loss of *Mx* – a putative precursor to the loss of *MHCII***

114 Here, combined with findings reported in the literature [1, 7], we find a
115 succession of immune-relevant gene losses throughout the evolution of the teleost
116 immune system: *TLR5* 151-147 mya, *Mx* 126-104 mya and *MHCII* 105-85 mya. The
117 loss of *TLR5* in the late Jurassic is encompassing the Paracanthopterygii
118 superorder together with the Lampridiformes and a few other species. The loss of
119 *Mx* in Gadiformes and *Stylephorus chordatus* appears in the early Cretaceous
120 followed by the loss of *MHCII* in Gadiformes during the transition from the early
121 to the late Cretaceous. Viewing the successive gene losses in light of changes in
122 paleontological climate, oceanography and major extinctions we see that the loss
123 of *TLR5* is close to the Jurassic-Cretaceous (J-K) boundary. There is accumulating
124 evidence of both species extinctions and radiations coinciding with this transition
125 together with an ongoing debate about average global temperatures in the same
126 period [20-26]. This is further supported by the fact that periods of extinctions are
127 often followed by population diversification and subsequent species radiation

enabling the invasion of new habitats [27, 28]. Habitat wise, the formation of the central Atlantic Ocean in the early Jurassic continued with a subsequent northward expansion in the Early Cretaceous [29]. Thus, if there were large changes in climate, or possibly an unknown larger extinction event, the loss of *TLR5* may be associated with adaptation of new species possibly towards new habitats within the opening Atlantic Ocean.

The loss of *Mx* is close to the early/late Cretaceous boundary and also overlapping one of the global anoxia events within this period approximately 120 mya. The loss of *MHCII* is also close to the boundary but spanning a second global anoxia event approximately 95 mya [30, 31]. Coinciding with these two anoxia events was the continued opening northward of the Central Atlantic Ocean expanding the North Atlantic Ocean further and the formation of a gateway between the South Atlantic Ocean and the Central Atlantic Ocean [29, 32]. The metabolically taxing anoxic environments, even though some adaptation likely was possible, resulted in the deep seas being depleted of fish [33, 34]. This is supported by higher extinction rates in the same period [35, 36]. The anoxic scenario fits with one of several mechanisms proposed to promote loss of *MHCII* – metabolic cost [37] – but could also be connected to post extinction speciation in which new species invade habitats where maintaining *MHCII* and *Mx* was less favourable.

Our findings can be further compared to the level of bony fish species family richness, diversification and extinction rates through evolutionary history. Bony fish species family richness gradually increased from Jurassic to modern time. However, there is a shift from increasing to decreasing richness with the J-K transition following the *TLR5* loss event combined with a small increase in extinction rate [38]. The loss of *Mx* and the global anoxia event ~120 mya are associated with a small increase in extinction rate but otherwise overall higher and stable species richness levels compared to the J-K transition. The loss of

156 *MHCII* spanning the second global anoxia event ~95 mya coincides with a large
157 drop in species richness combined with an increase in extinction rate and a large
158 increase in species diversification rate. As the losses of *TLR5*, *Mx* and *MHCII* are
159 clearly lineage specific and likely responses towards changes in species' habitats
160 [19] the loss of *TLR5* can be seen as an adaptation to events in the J-K transition
161 that led to extinctions promoting survival and speciation in the subsequent early
162 Cretaceous which is characterized by an increase in species richness and
163 diversification rates [38]. The loss of *Mx* spanning a global anoxia event ~120 mya
164 does not overlap with any large changes in species richness, extinction or
165 speciation rates. However, after this event there is an increase in species richness
166 and speciation rate and thus *Mx* loss can be viewed as a beneficial adaptation in
167 the anoxic environment leading to subsequent increased speciation [38]. The loss
168 of *MHCII* spanning the second global anoxia event ~95 mya presents a different
169 pattern than *TLR5* and *Mx*. Here there is an overlap between the gene loss and
170 large drops in species richness and origination rates [38]. This indicates that the
171 loss *MHCII* had more adverse effects than the loss of *TLR5* and *Mx*, however, still
172 over time promoting speciation within the Gadiformes lineage [7].

173 Even though the functional implication of *TLR5*, *Mx* and *MHCII* loss on the
174 teleost immune system remains unclear our data indicates that the J-K transition
175 harbours events central to shaping the teleost immune system initiated by the
176 loss of *TLR5*. Further, the loss of *Mx* directly outside of the Gadiformes lineage
177 indicates that this loss might have been a catalyst for the subsequent loss of
178 *MHCII*. This combined with the increased metabolic cost to maintain the *MHCII*
179 system in an anoxic environment likely lead to the alternate immune system seen
180 in Gadiformes today.

181 **Materials and methods**

182 The generation of teleost sequences, assemblies and time-calibrated phylogeny is
183 described in detail in Malmstrøm et al. [7] and briefly in Supporting information.

Query *Mx* and *TLR5* protein sequences were obtained from Ensembl v.82 (SI table 2 and 3) [9]. The NCBI BLAST tool was used to search the *Salmo salar* genome (ICSASG_v2, GCA_000233375.4) with default settings. All *Mx*/*TLR5* sequences were used as queries in a BLAST+ v. 2.2.26 a TBLASTN search against the non-reference teleost unitigs with e-value 1e-10 and outformat 6 with the 'sseq' option added [39]. The reported targets for *Mx* were aligned against queries using MEGA5 to eliminate hits from other GTPase genes and to establish *Mx* copy number [40]. To establish synteny protein sequences from genes flanking *Mx* in Ensembl vertebrate genomes (SI table 2 and 3) were used in TBLASTN searches as described above where partial synteny was obtained for 15 of 38 non-reference teleosts harbouring *Mx*.

Acknowledgements

This work was supported by The Research Council of Norway (Grant number 222378/F20 to KSJ/SJ). Some of the teleost genomes were assembled using the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/.

References

1. Star, B., et al., *The genome sequence of Atlantic cod reveals a unique immune system.* Nature, 2011. **477**(7363): p. 207-10.

2. Haase, D., et al., *Absence of major histocompatibility complex class II mediated immunity in pipefish, Syngnathus typhle: evidence from deep transcriptome sequencing.* Biol Lett, 2013. **9**(2): p. 20130044.

3. Buonocore, F. and M. Gerdol, *Alternative adaptive immunity strategies: coelacanth, cod and shark immunity.* Mol Immunol, 2015.

4. Boehm, T., et al., *VLR-based adaptive immunity.* Annu Rev Immunol, 2012. **30**: p. 203-20.

5. Pancer, Z., et al., *Variable lymphocyte receptors in hagfish.* Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9224-9.

214 6. Han, B.W., et al., *Antigen recognition by variable lymphocyte receptors.* Science,
215 2008. **321**(5897): p. 1834-7.

216 7. Malmstrøm, M., et al., *Evolution of the immune system influences speciation*
217 *rates in teleost fishes.* Nat Genet, In press.

218 8. Solbakken, M.H., et al., *Evolutionary redesign of the Atlantic cod (Gadus morhua*
219 *L.) Toll-like receptor repertoire by gene losses and expansions.* Sci Rep, 2016. **6**: p.
220 25211.

221 9. Cunningham, F., et al., *Ensembl 2015.* Nucleic Acids Res, 2015. **43**(Database issue):
222 p. D662-9.

223 10. Glasauer, S.M. and S.C. Neuhauss, *Whole-genome duplication in teleost fishes*
224 *and its evolutionary consequences.* Mol Genet Genomics, 2014. **289**(6): p. 1045-
225 60.

226 11. Haller, O., et al., *Mx GTPases: dynamin-like antiviral machines of innate immunity.*
227 Trends Microbiol, 2015. **23**(3): p. 154-63.

228 12. Mitchell, P.S., et al., *Evolutionary Analyses Suggest a Function of MxB Immunity*
229 *Proteins Beyond Lentivirus Restriction.* PLoS Pathog, 2015. **11**(12): p. e1005304.

230 13. Bergan, V. and B. Robertsen, *Characterization of Atlantic halibut (Hippoglossus*
231 *hippoglossus) Mx protein expression.* Dev Comp Immunol, 2004. **28**(10): p. 1037-
232 47.

233 14. Das, B.K., A.E. Ellis, and B. Collet, *Induction and persistence of Mx protein in*
234 *tissues, blood and plasma of Atlantic salmon parr, Salmo salar, injected with poly*
235 *I:C.* Fish Shellfish Immunol, 2009. **26**(1): p. 40-8.

236 15. Fernandez-Trujillo, M.A., et al., *Mx1, Mx2 and Mx3 proteins from the gilthead*
237 *seabream (Sparus aurata) show in vitro antiviral activity against RNA and DNA*
238 *viruses.* Mol Immunol, 2013. **56**(4): p. 630-6.

239 16. Huang, B., W.S. Huang, and P. Nie, *Characterization of four Mx isoforms in the*
240 *European eel, Anguilla anguilla.* Fish Shellfish Immunol, 2013. **35**(3): p. 1048-54.

241 17. Das, B.K., et al., *Induction of Mx protein in Atlantic cod with poly I:C: immuno-*
242 *cross reactive studies of antibodies to Atlantic salmon Mx with Atlantic cod.* Fish
243 Shellfish Immunol, 2008. **25**(3): p. 321-4.

244 18. Braun, B.A., et al., *Mx1 and Mx2 key antiviral proteins are surprisingly lost in*
245 *toothed whales.* Proc Natl Acad Sci U S A, 2015. **112**(26): p. 8036-40.

246 19. Olson, M.V., *When less is more: gene loss as an engine of evolutionary change.*
247 Am J Hum Genet, 1999. **64**(1): p. 18-23.

20. Benson, R.B., et al., *Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates.* Proc Biol Sci, 2010. **277**(1683): p. 829-34.

21. Benson, R.B. and P.S. Druckenmiller, *Faunal turnover of marine tetrapods during the Jurassic-Cretaceous transition.* Biol Rev Camb Philos Soc, 2014. **89**(1): p. 1-23.

22. Bambach, R.K., *Phanerozoic biodiversity mass extinctions.* Annual Review of Earth and Planetary Sciences, 2006. **34**: p. 127-155.

23. Alroy, J., *The shifting balance of diversity among major marine animal groups.* Science, 2010. **329**(5996): p. 1191-4.

24. Cavin, L., *Diversity of Mesozoic semionotiform fishes and the origin of gars (Lepisosteidae).* Naturwissenschaften, 2010. **97**(12): p. 1035-40.

25. Korte, C., et al., *Jurassic climate mode governed by ocean gateway.* Nat Commun, 2015. **6**: p. 10015.

26. Price, G.D., et al., *Isotopic evidence for long term warmth in the Mesozoic.* Sci Rep, 2013. **3**: p. 1438.

27. Wellborn, G.A. and R.B. Langerhans, *Ecological opportunity and the adaptive diversification of lineages.* Ecol Evol, 2015. **5**(1): p. 176-95.

28. Simoes, M., et al., *The Evolving Theory of Evolutionary Radiations.* Trends Ecol Evol, 2016. **31**(1): p. 27-34.

29. Melankholina, E.N. and N.M. Sushchevskaya, *Development of passive volcanic margins of the Central Atlantic and initial opening of ocean.* Geotectonics, 2015. **49**(1): p. 75-92.

30. Wilson, P.A. and R.D. Norris, *Warm tropical ocean surface and global anoxia during the mid-Cretaceous period.* Nature, 2001. **412**(6845): p. 425-9.

31. Sinninghe Damsté, J.S., et al., *A CO2 decrease-driven cooling and increased latitudinal temperature gradient during the mid-Cretaceous Oceanic Anoxic Event 2.* Earth and Planetary Science Letters, 2010. **293**(1–2): p. 97-103.

32. Granot, R. and J. Dyment, *The Cretaceous opening of the South Atlantic Ocean.* Earth and Planetary Science Letters, 2015. **414**: p. 156-163.

33. Priede, I.G. and R. Froese, *Colonization of the deep sea by fishes.* Journal of Fish Biology, 2013. **83**(6): p. 1528-1550.

34. Rogers, A.D., *The role of the oceanic oxygen minima in generating biodiversity in the deep sea.* Deep-Sea Research Part Ii-Topical Studies in Oceanography, 2000. **47**(1-2): p. 119-148.

282 35. Takashima, R., et al., *Greenhouse World and the Mesozoic Ocean.* Oceanography,
283 2006. **19**.

284 36. Harnik, P.G., et al., *Extinctions in ancient and modern seas.* Trends in Ecology &
285 Evolution, 2012. **27**(11): p. 608-617.

286 37. Star, B. and S. Jentoft, *Why does the immune system of Atlantic cod lack MHC II?*
287 Bioessays, 2012. **34**(8): p. 648-51.

288 38. Guinot, G. and L. Cavin, *'Fish' (Actinopterygii and Elasmobranchii) diversification*
289 *patterns through deep time.* Biol Rev Camb Philos Soc, 2015.

290 39. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics,
291 2009. **10**: p. 421.

292 40. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using*
293 *maximum likelihood, evolutionary distance, and maximum parsimony methods.*
294 Mol Biol Evol, 2011. **28**(10): p. 2731-9.

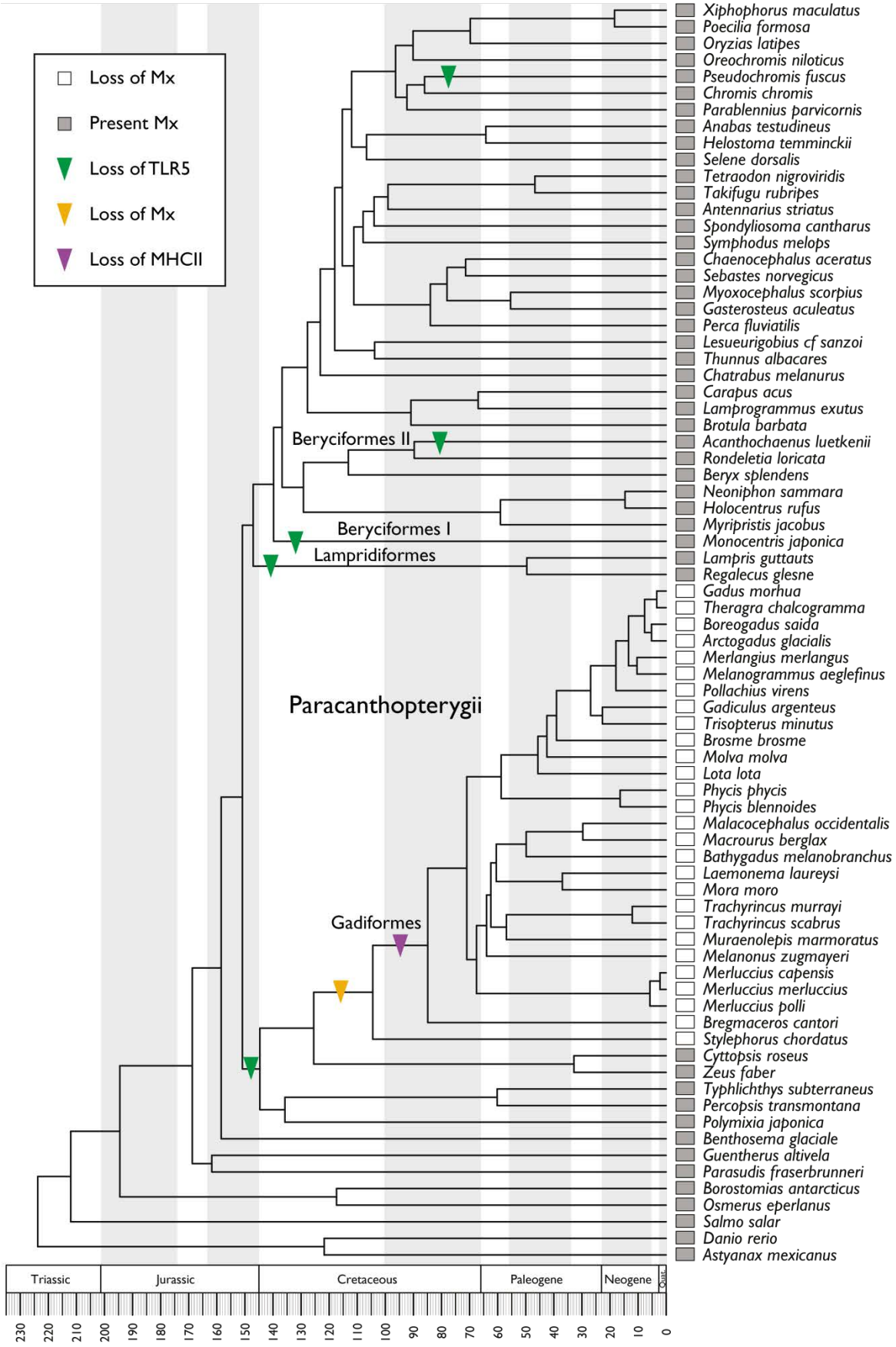
295

296 ## Tables and figures

297 **Table 1** The genomic region containing *Mx* and *Mx* copy numbers for all
298 investigated species. * Synteny analysis was only possible for some non-reference
299 teleost species and then in a partial manner displaying a single flanking region. **
300 The two *Mx* regions found in *Mus musculus* are located directly adjacent to each
301 other. *Mx* is lost from the genome of *Latimeria chalumnae*. ORF indicates open
302 reading frames without annotation in reference species.

Species	**Flanking gene**	**Mx region**	**Flanking gene**
Homo sapiens	FAM3B	2 x Mx	TMPRSS2
Mus musculus region # 2 **	Gm9242	1 x Mx	TMPRSS2
Mus musculus region # 1	BACE2	1 x Mx	FAM3B
Gallus gallus	FAM3B	1 x Mx	TMPRSS2
Anolis carolinensis	FAM3B	2 x Mx	Mx2
Xenopus tropicalis	FAM3B	1 x Mx	TMPRSS2
Latimeria chalumnae		No Mx	
Non-Ensembl phylogeny species w/ Mx *	THOC7	1-3 x Mx	SYNPR
Xiphophorus maculatus	THOC7	1 x Mx	SYNPR
Poecilia formosa	THOC7	1 x Mx	SYNPR
Oryzias latipes	THOC7	1 x Mx	SYNPR
Oreochromis niloticus	THOC7	2 x Mx	SYNPR
Tetraodon nigroviridis	THOC7	1 x Mx	IP6K2A
Takifugu rubripes	THOC7	1 x Mx	IP6K2A
Gasterosteus aculeatus	THOC7	2 x Mx	SYNPR
Salmo salar	THOC7	2 x Mx	SYNPR
Danio rerio region # 1	EFNB2B	2 x Mx	PCNP
Danio rerio region # 2	ORF	2 x Mx	HPX
Danio rerio region # 3	ORF	2 x Mx	ORF
Danio rerio region # 4	ABCG1	1 x Mx	PGM2L1
Astyanax mexicanus	EFNB2B	1 x Mx	PCNP
Lepisosteus oculatus	ORF	3 x Mx	HPX
Petromyzon marinus	End of scf	1 x Mx	GLRA3

303


Loss of Mx
Present Mx
Loss of TLR5
Loss of Mx
Loss of MHCII
Beryciformes II
Beryciformes I
Lampridiformes
Paracanthopterygii
Gadiformes
Xiphophorus maculatus
Poecilia formosa
Oryzias latipes
Oreochromis niloticus
Pseudochromis fuscus
Chromis chromis
Parablennius parvicornis
Anabas testudineus
Helostoma temminckii
Selene dorsalis
Tetraodon nigroviridis
Takifugu rubripes
Antennarius striatus
Spondyliosoma cantharus
Symphodus melops
Chaenocephalus aceratus
Sebastes norvegicus
Myoxocephalus scorpius
Gasterosteus aculeatus
Perca fluviatilis
Lesueurigobius cf sanzoi
Thunnus albacares
Chatrabus melanurus
Carapus acus
Lamprogrammus exutus
Brotula barbata
Acanthochaenus luetkenii
Rondeletia loricata
Beryx splendens
Neoniphon sammara
Holocentrus rufus
Myripristis jacobus
Monocentris japonica
Lampris guttauts
Regalecus glesne
Gadus morhua
Theragra chalcogramma
Boreogadus saida
Arctogadus glacialis
Merlangius merlangus
Melanogrammus aeglefinus
Pollachius virens
Gadiculus argenteus
Trisopterus minutus
Brosme brosme
Molva molva
Lota lota
Phycis phycis
Phycis blennoides
Malacocephalus occidentalis
Macrourus berglax
Bathygadus melanobranchus
Laemonema laureysi
Mora moro
Trachyrincus murrayi
Trachyrincus scabrus
Muraenolepis marmoratus
Melanonus zugmayeri
Merluccius capensis
Merluccius merluccius
Merluccius polli
Bregmaceros cantori
Stylephorus chordatus
Cyttopsis roseus
Zeus faber
Typhlichthys subterraneus
Percopsis transmontana
Polymixia japonica
Benthosema glaciale
Guentherus altivela
Parasudis fraserbrunneri
Borostomias antarcticus
Osmerus eperlanus
Salmo salar
Danio rerio
Astyanax mexicanus
Triassic
Jurassic
Cretaceous
Paleogene
Neogene
Quat.
230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0

304

305 **Figure 1** Phylogenetic distribution of *Mx* genes in 76 teleost species. *Mx* is mapped onto a
306 teleost phylogeny generated by Malmstrøm et al[7]. The presence of *Mx* is marked by
307 grey boxes. The loss of *Mx* is marked by an orange arrow. The losses of *MHCII* and *TLR5*
308 are marked by purple and green arrows, respectively. The absence of *Mx* is a
309 characteristic of the Gadiformes and *Stylephorus chordatus* and thus predates the loss of
310 *MHCII* from the Gadiformes. The absence of *TLR5* affects the entire Paracanthopterygii
311 superorder together with the Lampridiformes, two species representative from the
312 Beryciformes and *Pseudochromis fuscus*. The loss of *Mx* occurs between 126-104 mya, the
313 loss of *MHCII* 105-85 mya and the loss of *TLR5* 151-147 mya.

314

Paper IV

Disentangling the immune response and host-pathogen interactions in *Francisella noatunensis* infected Atlantic cod

Monica Hongrø Solbakken*[1], Sissel Jentoft*[1,2], Trond Reitan[1], Helene Mikkelsen[3], Tone F. Gregers[4], Oddmund Bakke[5], Kjetill S. Jakobsen[1], Marit Seppola[6]

[1]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

[2]Department of Natural Sciences, University of Agder, Kristiansand, Norway

[3]The Northern Norway Regional Health Authority, Tromsø, Norway

[4]Department of Biosciences, University of Oslo, Oslo, Norway

[5]Department of Biosciences, Centre for Immune Regulation, University of Oslo, Oslo, Norway

[6]Department of Medical Biology, The Arctic University of Norway, Tromsø, Norway

*Corresponding authors:

Monica Hongrø Solbakken. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. m.h.solbakken@ibv.uio.no, +47 22 84 41 64

Sissel Jentoft. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. sissel.jentoft@ibv.uio.no, +47 22 85 72 39

The immune gene repertoire of Atlantic cod is shown to deviate from that of genome sequenced teleosts as well as other vertebrates. So far, no experimental immunological studies have been able to fully unravel its functionality. By global transcriptome profiling, we here investigate the immune response and host-pathogen interaction of Atlantic cod juveniles infected with *Francisella noatunensis* — a pathogen causing the severe disease francisellosis in wild and farmed fish species worldwide. We show that Atlantic cod displays an overall classic initiation of immunity with inflammation, acute phase response and cell recruitment. Related to adaptive immunity we find an extensive up-regulation of Major Histocompatibility Complex class I (MHCI). These are likely to present endogenous as well as exogenous antigens with corresponding cytotoxic cellular responses. Our results indicate T-cell independent B-cell activation with the help of Toll-like receptors and possibly also with help from neutrophils and Natural Killer cells. Further, we find that *F. noatunensis* alters the immune response in Atlantic cod similar to that seen in other fish but also similar to the mammalian equivalent tularemia. This is evident from the effects on pathways in iron homeostasis, phagosome and autophagosome formation, oxidative burst and apoptosis. Collectively, we have obtained further insight into the gene expression mechanism underlying francisellosis. Moreover, our results provide novel insight into the orchestration of the Atlantic cod immune response indicating that Atlantic cod have a phagocyte-dominated initial defense, employs MHCI – both classically and through cross-presentation — and generates antibodies through direct B-cell activation without the conventional help from T-cells or NKT-cells.

Introduction

The Atlantic cod's (*Gadus morhua* L.) unconventional immunity compared to other teleost species, was revealed through genome sequencing showing the loss of the

47 Major Histocompatibility Complex (MHC) class II pathway, gene expansion of
48 *MHCI* and gene losses and expansion within the family of Toll-like receptors (*TLRs*)
49 [1, 2]. Although additional studies have further investigated these large gene
50 expansions and gene losses, and hypothesized on functional outcomes [3-5], no
51 overarching functional examination of this particular immune system, or its
52 interactions with pathogen, has been conducted. To elucidate the orchestration of the
53 Atlantic cod immune response within a host-pathogen interaction framework we
54 chose a common disease affecting wild and farmed fish species worldwide —
55 francisellosis [6-10].

56 In fish, francisellosis is a systemic granulomatous inflammatory disease
57 characterized by granulomas in visceral organs such as spleen and head-kidney. It is
58 caused by the gram negative facultative intracellular bacterium *Francisella*
59 *noatunensis*. Currently, there is no vaccine available and treatments with
60 antimicrobial compounds have been reported with highly variable effects [11, 12].
61 Most of the knowledge gained of this disease comes from studies of the mammalian
62 counterpart tularemia which is most often caused by *F. tularensis* [13-15]. However,
63 in recent years characterization of the mechanisms underlying fish-specific infections
64 with *F. noatunensis* subspecies have been conducted and demonstrate several
65 similarities to the mechanisms described in mammals. In both fish and mammals,
66 *Francisella spp.* resides within phagocytic cells – mainly macrophages [16-19]. It likely
67 enters through phagocytosis involving surface receptors such as mannose- and
68 complement receptors [13-15]. *Francisella spp.* is demonstrated to delay apoptosis,
69 hampering the final stages of phagosome maturing into phagolysosomes, inhibiting
70 the defense mechanism oxidative burst and preventing autophagy. Dysregulation of
71 the immune response caused by *Francisella spp.* in mammals leads to excessive
72 amounts of inflammatory cytokines and recruitment of large amounts of neutrophils.

73 Furthermore, most of the well-described immune evasion strategies of *Francisella spp.*
74 are shown to affect both the innate immune system as well as the initiation of
75 adaptive immunity — linked to its intracellular lifestyle within professional antigen
76 presentation cells [13-15, 17, 19, 20]. The immune evasion is mediated through
77 interference with interferon gamma (*IFNG*) signaling: i.e. *Francisella* induces the
78 expression of anti-inflammatory cytokines and inhibits the expression of pro-
79 inflammatory cytokines by targeting *IFNG* receptors and preventing activation of
80 downstream transcription factors. In mammals *Francisella* triggers the degradation of
81 *MHCII* through ubiquitination restricting presentation of antigen on the cell surface
82 but this does not prevent a robust antibody production consisting of both
83 immunoglobulin gamma (IgG2) and immunoglobulin mu (IgM) [13, 14].
84 Additionally, *Francisella spp*. skews the development of the adaptive immune
85 response towards a more tolerogenic setting which again results in reduced
86 activation of immune cells [14]. In comparison, the effect of *Francisella noatunensis* on
87 the adaptive immune system of fish is poorly characterized beyond demonstrating
88 an increase of antibody expression that likely consists of IgM [2, 21].

89 Here, we in-depth characterize the immune response and the host-pathogen
90 interaction in *F. noatunensis* infected Atlantic cod juveniles using global
91 transcriptome profiling. Overall, Atlantic cod displays classic inflammation, acute
92 phase response and recruitment of immune cells. Furthermore, the effect of
93 *Francisella* on the innate immune system, more specifically delay of apoptosis, delay
94 of phagosome maturation, inhibition of oxidative burst and autophagy are likely
95 explanations for many of the differential gene expression patterns observed.
96 However, we also demonstrate significant changes in gene expression providing
97 insight into the defense mechanisms of Atlantic cod such as MHCI cross-

98 presentation, T-cell independent B-cell activation and likely a neutrophil-dependent
99 response towards francisellosis.

100 **Results**

101 In this study we have chosen a multifaceted approach to detect differentially
102 expressed genes. It consists of both *de novo* and reference-genome based
103 transcriptomics (Trinity[22] and Tuxedo[23], respectively) combined with R-
104 packages EdgeR[24] and CummeRbund [23] for final differential expression analysis
105 and result presentation. In addition, due to our experimental setup we have applied
106 a custom analysis script clustering genes by their expression pattern over time (for
107 details see methods section). Overall, we found that the three approaches detected
108 similar trends but with somewhat different sensitivities. This was especially
109 prominent in relation to annotation where we found that Trinity readily detected
110 immune genes whereas Cufflinks had improved resolution for non-immune genes
111 (Table 1). Below we present our findings focusing on the output from the Tuxedo
112 pipeline and supplement with findings from the other two approaches. This is to
113 capture genes that may not have corresponding gene models in the reference
114 genome [25] due to filtering of smaller genome contigs, thus immune genes located
115 to these contigs are only found using the Trinity approach.

116 The TopHat-Cufflinks-CummeRbund (TCC) pipeline reported 90 differentially
117 expressed genes 6 hrs. post infection compared to control. Gene ontology (GO) terms
118 associated with the annotated genes found indicate up-regulation of systems
119 involved in muscle functionality but also increased inhibition of nuclear factor
120 kappa-light-chain-enhancer of activated B cells (NFkB) transcription factor activity
121 and increased up-regulation of granulocyte chemotaxis (Table 2). On the individual
122 gene level we found increased inflammation through the up-regulation of
123 interleukin 1B (*IL1B*) and inflammasome components caspase 1 (*CASP1*) and

124 Nucleotide-Binding Oligomerization Domain, Leucine Rich Repeat and CARD
125 Domain Containing 3 (*NLRC3*). This is counteracted by the up-regulation of anti-
126 inflammatory *IL10*. Simultaneously there is up-regulation of the neutrophil
127 attractant *CXCL8* (Table 3).

128 At 2 days post infection 878 differentially expressed genes were identified. Their
129 corresponding annotations and related GO:terms demonstrated a major contribution
130 from genes related to antigen processing and presentation of antigens by MHCI.
131 There was also an overall response to cytokines and response to oxygen-containing
132 compounds in addition to an up-regulation of genes involved in apoptosis, iron
133 homeostasis and ribosome biogenesis (Table 1). The down-regulated genes
134 displayed less prominent trends with GO:terms mainly related to cell-substrate
135 junction assembly and triglyceride metabolism (Table 2). Looking closer at the
136 individual genes there is a continued up-regulation of pro- and anti-inflammatory
137 cytokines (*IL1B*, *IL10*, Transforming Growth Factor beta (*TGFB*), neutrophil
138 attractant *CXCL8* and monocyte attractant *CCL2* at this stage. However, the acute-
139 phase reactants became more prominent such as transferrin (*TF*), Fibrinogen (*FG*),
140 ceruloplasmin (*CP*) together with the antimicrobials hepcidin (*HAMP*), *IL4L1* and
141 lysozyme (*LYG2*) as well as pattern recognition receptors (*PRRs*) from the Toll-like
142 family, C-type lectin family and NOD-like family. There is also evidence of
143 apoptosis through caspases with up-regulation of *CASP3*, *CASP6* and *CASP7*,
144 granzyme B (*GZMB*) and *BAX*. Finally, there are signs of increased phagosome
145 activity (Table 3).

146 At 4 days post infection 1231 differentially expressed genes were identified and this
147 it the time-point with the highest GO:term diversity. There is an overall increased
148 response to organic and chemical stimulus combined with response to cytokines.
149 Further, there is extensive regulation of apoptosis and some regulation of single-

150 organism transport and ferric iron transport. In contrast, there is down-regulation of
151 collagen catabolic process and fructose metabolic process (Table 2). With respect to
152 the individual immune genes, we at this time-point, observed the initiation of
153 interferon gamma (IFNG), complement, the continued inflammation, the continued
154 increased level of apoptosis, increased level of MHCI, antimicrobial peptides and
155 acute-phase reactants. The differentially expressed transcripts indicate a continued
156 effect on phagosomes. However, the production of reactive oxygen species declined
157 at day 4 post infection (Figure 1. Table3).

158 At 7 days post infection 1130 differentially expressed genes were identified. Here,
159 we observed a continued expression of MHCI but the GO:terms indicated a more
160 prominent possibility of cross-presentation than the earlier time-points. The
161 response towards wounding, viral entry into host cell and cell-cell adhesion declines
162 (Table 2), which is also reflected at the individual gene level with decline of early
163 innate defenses such as inflammation, acute-phase reactants and complement. With
164 respect to phagosomes there is a decline in expression compared to day 4 (Figure 2).

165 The GO:terms derived from annotated genes clustered with our custom script
166 displayed similar trends to that of the pair-wise differential gene expression analyses.
167 Up-regulated transcripts over time were heavily influenced by the presentation of
168 antigen on MHCI but also metabolic processes and transmembrane transport. Genes
169 demonstrating an internal maximum (quadratic, positive) expression pattern were
170 connected to a range of systems such as negative regulation of intracellular signal
171 transduction, response to lipopolysaccharide, positive regulation of apoptosis and
172 cytokine signaling. Genes related to wound healing (among others), were decreasing
173 over time, whereas genes with an internal minimum (quadratic negative) expression
174 pattern were related to a range of metabolic processes. Finally, the freestyle pattern

175 (alternating trends over time) genes were related to positive regulation of ubiquitin
176 protein transferase activity, response to unfolded protein and more (Table 4).

177 Discussion

178 By global transcriptome profiling we have obtained a more systemic overview of the
179 innate defense mechanisms, the host-pathogen interactions as well as the transition
180 into the unconventional adaptive immune mechanisms in *F: noatunensis* infected
181 Atlantic cod. Collectively, we find strong resemblances to the immune response of
182 mammals with up-regulation of inflammation and acute-phase reactants including
183 complement, cytokines and chemokines, antimicrobial peptides and PRRs (Table 2
184 and 3)[26, 27].

185 A prominent inflammatory response

186 Earlier reports have shown that *Francisella spp.* suppresses pro-inflammatory
187 cytokines and increases anti-inflammatory cytokines to dampen cell-mediated
188 immune responses in mammals [14]. In this study we observe transcriptomic up-
189 regulation of pro-inflammatory cytokines *IL1B* and *IL12B* (subunit for both *IL12A*
190 and *IL23A* of which none are found) whereas no significant differential expression of
191 *TNF* and *IL6* was found (Table 3). Furthermore, it has been proposed that the level of
192 suppression is only required to be at a level where it keeps inflammasomes from
193 being activated [14]. Our data does not support this observation as *IL1B* is up-
194 regulated from a very early stage and signs of inflammasome up-regulation is seen
195 throughout the experiment (Table 3). Lastly, *Francisella* has been observed to initiate
196 production of anti-inflammatory *IL10* in mammals [14], however, we observe up-
197 regulation of both *IL10* as well as the anti-inflammatory *TGFb* (Table 3). *Francisella* is
198 also known to induce the expression of several antimicrobial peptides [14] and in
199 line we this we find up-regulation of *HAMP*, *IL4L1* and *LYG2* (Table 3). However,
200 antimicrobial peptides does not efficiently hinder host entry by *Francisella* as it has

evolved defense mechanisms including changing its cell surface charge counteracting cationic antimicrobial peptides and expressing multidrug efflux pumps in addition to that its intracellular lifestyle efficiently protects it from host defenses [14].

Tightly interwoven with the inflammation and acute-phase response is the iron homeostasis – a key nutrient both for the host and for the pathogen as iron ions are part of important enzymes and redox reactions. The up-regulation of *FTH1*, *CP*, *TF*, *F3*, *HAMP* (Table 3) observed in our dataset, indicate that the iron homeostasis is affected. These genes are all thought to be involved in sequestering the iron from the pool available to the pathogen during infection. Most of the iron available for use by pathogens is located within host cells but mostly sequestered by iron-containing enzymes and iron storage proteins such as ferritin. Upon infection host cells decrease the influx of iron into cells by down-regulating transferrin receptors as well as increasing expression of ferritin to sequester as much iron as possible both intracellularly and extracellularly. However, in the case of *HAMP*, this up-regulation prevents efflux of iron from the host cell by HAMP binding to ferroportin. Thus, this otherwise protective mechanism ends up providing an iron source for the pathogen due to *Francisella's* intracellular lifestyle [14].

Signs of extensive neutrophil recruitment

There is pronounced up-regulation of hepoxilin-metabolism related genes such as arachidonate 12-lipoxygenase (*ALOX12*) and arachidonate 15-lipoxygenase type B (*ALOX15B*) throughout the timeline of this study (Table 3). These genes are commonly involved in fatty acid metabolism maintaining skin and mucus membranes in mammals, but there has also been described a function for these genes in relation to inflammation and recruitment of neutrophils across endothelial cell layers in mammals [28]. The lipooxygenases generate hepoxilin which establishes a

227 gradient for neutrophil migration [28]. This interaction is in our data well
228 represented with up-regulation of *ALOX12* and *ALOX15B*. Another well-known
229 neutrophil attractant, *CXCL8* (alias interleukin 8, *IL8*) [29], which in Atlantic cod
230 exists in 8 copies [4], is also up-regulated. Collectively, the prominent expression of
231 *CXCL8* and hepoxilin-related genes indicate that Atlantic cod commits a neutrophil
232 defense upon infection with *Francisella*. Such a strategy would also correspond to the
233 high titers of neutrophils in Atlantic cod blood [30]. Furthermore, we also find up-
234 regulation of *CCL2* recruiting monocytes as well as some NK-cell markers (*NCAM1*
235 and *ITGAL*) indicative of mature NK-cell activity at day 2 and day 4. However, these
236 latter NK-cell markers are also found on other cell populations (Table 3) [31].

237 **Delay of apoptosis**

238 Cell death is a well-known defense mechanism for the handling of intracellular
239 pathogens as well as a mechanism enabling proper clearance of immune cells such
240 as neutrophils, i.e. minimize tissue damage and release of toxic compounds. It is
241 dependent on detection through PRRs such as TLRs, NLRs and NK-cell receptors.
242 Depending on the down-stream signaling pathway the end results is either cell
243 death or pyroptosis [32, 33]. The former involves death receptors and caspases 3, 8
244 and 9 (*CASP3*, *8* and *9*) leading to permeabilized cell membranes. The latter is
245 dependent on the inflammasome and *CASP1* and releases large amounts of
246 pyrogens and inflammatory cytokines through lysis of host cells [34]. Studies have
247 found that various *Francisella* strains initiate both apoptosis and pyroptosis in
248 mammalian cells. Our results demonstrate a stronger *CASP3* response supported by
249 the pro-apoptopic gene *BAX*, *CASP6* and *CASP7* indicating that in our system cell
250 death by apoptosis is more prevalent. *Francisella* strains have also been shown to
251 inhibit the initiation of apoptosis in mammalian neutrophil cells where the natural
252 onset of apoptosis begins within 12 hrs and is effective by 24 hrs. In contrast,

253 *Francisella* infected neutrophils displayed onset of apoptosis beyond 48 hrs [33]. In
254 our data the initiation of apoptosis by *CASP3* is seen at day 2, continues at day 4 and
255 is further support by *BAX* at 7 days post infection (Table 3).This delay in apoptosis,
256 promoting pathogen survival, likely increases the life span of central immune cells
257 which further may be responsible for the dysregulated immune response forming
258 granulomas in *Francisella* infected organisms [33].

259 **Pathogen detection and communication with the adaptive immune system**

260 The ability to detect a pathogen upon host entry plays an important role for the
261 overall orchestration and outcome of the immune response as well as the
262 establishment of communication with the adaptive immune system [35, 36]. The
263 various families of PRRs are located throughout the cell and respond to a range of
264 pathogen-associated and damage-associated molecular patterns (PAMPs and
265 DAMPs, respectively) [35, 36]. This *F. noatunensis* infection is dominated by the up-
266 regulation cell-surface located TLRs and mannose-receptors — in particular *TLR2/6*
267 (annotated as *TLR25* by Solbakken et al. [4]), and *MRC1* (alias *CD206*) (Table 3).There
268 was no evident increased expression of PRR transcripts associated with the cytosol
269 or intracellular membranes, which correlates with the suggested inhibition of
270 intracellular PRR signaling by *Francisella spp.*in mammals [14]. The function of *TLR25*
271 has been implicated in the detection of surface structures derived from bacteria due
272 to the gene's phylogenetic relationship to *TLR1/2/6* [4] and the response pattern
273 demonstrated here further supports this. *MRC1* are receptors found to be involved
274 in phagocytosis and enabling presentation of antigen on MHCII [37]. However, as
275 Atlantic cod lacks the *MHCII* pathway [1], it its more likely that the up-regulation of
276 *MRC1* expression found is related to phagocytosis. As *Francisella* strains infect and
277 replicate within phagocytic cells like macrophages and neutrophils they have
278 evolved to avoid mechanisms leading to their clearance. Three of these mechanisms

279 are delay of phagosome maturation, inhibiting the production of reactive oxygen
280 and nitrogen species with subsequent oxidative burst aimed at clearing
281 phagocytosed material and prevention of autophagy [13-15].Final phagosome
282 maturation would result in an environment prematurely killing *Francisella* [13-15].
283 Here, we find that the phagosome pathway is affected with up-regulation of *RAB7*
284 and tubulin suggesting that there is no delay in phagosome maturation. *Francisella*
285 also needs the further acidification of the phagosome to be able to escape into the
286 cytosol [13-15]. In our data this is supported by the up-regulation of *vacuolar ATPases*
287 (Figure 1, Table 3) suggesting a fine-tuned balance for the pathogen between
288 immune evasion and immune responses promoting its life cycle. It has further been
289 found that *Francisella* strains inhibit the oxidative burst mechanism in various ways –
290 also in Atlantic cod [16, 38, 39]. We found an overall down-regulation of neutrophil
291 cytosolic factor 1 (*NCF1* alias *p47phox*), a part of the NADPH activating complex
292 enabling production of reactive oxygen species indicating a protective environment
293 (Table 3). In the event of an unsuccessful formation of mature phagosomes as a
294 defense mechanism autophagy can be initiated to clear pathogens from the
295 intracellular environment. The avoidance mechanism used by *Francisella* preventing
296 autophagy is not clearly understood and it is suggested that certain sugar moieties
297 surrounding *Francisella* strains protects against recognition in mammalian cells [40,
298 41, and references therein]. We do not find convincing significant differential
299 expression of transcripts related to autophagy (Table 4) suggesting that *Francisella*
300 successfully has inhibited this self-defense mechanism in Atlantic cod.

301 Interferon gamma (*IFNG*) is a key regulator in the transition from innate to adaptive
302 immunity. Its signaling, even though being delayed by *Francisella* infection,
303 overcomes the inhibitory effects of the bacterium and thus can facilitate clearance by
304 increased nitric oxide production, induction of authophagy as well as by increasing

305 antigen presentation on MHCI and II. Mechanistically, *Francisella* will down-regulate
306 *IFNGR1*, the required *IFNGR1* transcription factor *STAT1* and increase the
307 expression of *SOCS3* which is a negative inhibitor of IFNG signaling [14]. In support
308 of this, in our data we observe an *IFNG* response at day 4 post infection and up-
309 regulation of *SOCS3* at day 2 and day 4, but a contrasting up-regulation of *STAT1* at
310 day 2 and day 4 (Supplementary table 1,2).

311 *Francisella spp.* has been shown to actively degrade MHCII through ubiquitination in
312 mammalian macrophages [42]. Due to the fact that Atlantic cod lacks *MHCII*, the
313 observed increased ubiquitination in our data is most likely related to the
314 degradation of other proteins. On the other hand Atlantic cod has a large gene
315 expansion of *MHCI* [1] where ubiquitinated material potentially can be presented.
316 Additionally, some of these *MCHI* genes carry signal peptides indicative of
317 specialized use in cross-presentation of exogenous antigen [3]. Moreover, support for
318 an active cross-presentation pathway is provided by our data and the corresponding
319 GO:term analysis (Table 2). MHCI may therefore play a central part in fighting this
320 particular pathogen. The increased phagosome activity, where MHCI can be loaded
321 for cross-presentation within the endosomal pathway, is further supporting this [43].
322 However, the functionality of the cross-presentation pathway in Atlantic cod needs
323 to be experimentally validated.

324 **The antibody response in Atlantic cod**

325 The antibody response of an organism can be initiated with or without T-cell help,
326 where the commonly described mechanism is the interaction between a antigen
327 presenting cell, a CD4+ T-cell and a B-cell within a germinal center culminating in
328 the production of antibodies [44]. Since Atlantic cod lacks CD4 [1] there will be no
329 conventional T-cell help, or help from other CD4+ cell lineages such as NKT-cells
330 [44]. However, there are T-cell /NKT-cell help-independent mechanisms usually

331 initiated through myeloid cells or directly with the B-cell itself if the antigen can
332 provide a sufficiently strong signal upon interacting with the B-cell receptors (BCR)
333 [44]. In line with this, our transcriptome analysis reveals no up-regulation of genes
334 involved in the conventional T-cell dependent or the more elaborate T-cell
335 independent mechanisms. Simpler systems such as direct B-cell stimulation with
336 additional signals from surface TLRs or neutrophils is more likely [44]. This is
337 supported in our data by the up-regulation of surface-located *TLRs* and significant
338 recruitment of neutrophils and also monocytes. Furthermore, we observe a response
339 towards lipopolysaccharide (LPS) in the GO:term analyses — indicating the presence
340 of antigens (such as LPS) able to initiate T-cell independent B-cell activation — and
341 thus most likely responsible for the slight up-regulation of immunoglobulins in our
342 data (Table 2, 3, 4). Finally, functional studies on Atlantic cod adaptive responses
343 have established the presence of a memory mechanism [45-47] and the
344 aforementioned direct stimulation of B-cell together with TLR signals are able to
345 establish memory contrary to the other mechanisms [44].

346 **Conclusions**

347 We find that Atlantic cod display an overall classic innate immune response. We also
348 find that this particular host-pathogen interaction results in trends similar to other
349 host-pathogen interactions described in mammals and fish as seen for different
350 members of the *Francisella* genus. Lastly, we observe that Atlantic cod, for this
351 particular infection, uses MHCI, both classically and through cross-presentation to
352 handle Francisella combined with direct stimulation of B-cell without the
353 conventional help from T-cells or NKT-cells. To further deduce the underlying
354 mechanisms, future experiments should extend beyond the sampled time-points in
355 our study, which should provide further insight into the adaptive responses. Also,

356 additional experiments should aim at using an extracellular and/or gram positive
357 pathogen to elucidate differences in host response patterns.

358 **Methods**

359 Please see GitHub repository: for details.

360 **Fish and experiment setup**

361 Atlantic cod juveniles (n=66) from the Norwegian cod breeding program
362 (www.nofima.no) were transported at approx. 2 g to 100 L tanks at the Aquaculture
363 Research Station (Tromsø, Norway) for grow-out in seawater of 3.4 % salinity at 10
364 ºC, 24 hour light and fed *ad libitum* with commercial feed (BioMar, Norway). The
365 rates of water inflow were adjusted to an oxygen saturation of 90-100 % in the outlet
366 water. The fish were reported to be healthy without any history of diseases and the
367 experiment was approved by the National Animal Research authority in Norway.
368 The fish were distributed in two circular, centrally drained, fiberglass tanks (250 L)
369 with 30 fish in each tank (density <20 kg/dm^3). The use of live Atlantic cod was
370 approved by the National Animal Research authority in Norway (FOTS id 1147) and
371 all methods were in accordance with the approved guidelines.

372 *Francisella noatuensis* subsp. *noatuensis* NCIMB 14265 isolate used for challenge was
373 originally isolated from diseased Atlantic cod (*Gadus morhua*) in Norway, and was
374 provided by Dr. Duncan Colquhoun at the National Veterinary Institute Oslo,
375 Norway [48, 49]. The bacteria were cultivated at 21 ºC for 7-10 days on CHAB agar:
376 heart infusion broth (Merck) pH 6.8 ± 0.2, supplemented with cysteine 0.1 % (Merck,
377 Germany), haemoglobin 2 % (Oxoid, England), glucose 1 %, agar 1.5 % and 5 %
378 human blood concentrate. The bacteria were stored in glycerol cultures at -80 ºC.
379 Pure colonies were inoculated in Bacto heart infusion broth (Becton and Dickson,
380 USA) pH 7, supplemented with cysteine 0.07 %, $FeCl_3$ 2 mM and glucose 1 %, and

381 incubated with agitation at 21 ºC for 24-30 hours before being used in the challenge
382 study. CHAB plates were used for determination of colony forming units (cfu) of
383 challenge dose and re-isolation of *F. noatunensis* from challenged fish.

384 The fish were acclimated to 15 ºC and starved 24 h before injection. Prior to intra-
385 peritoneal (ip) injection the fish (approx. 25 g) were anaesthetised with Metacainum
386 (50 mg/l, Norsk Medisinaldepot), and injected with 100 µl of either *F. noatunensis* (5 x
387 10^7 cfu per fish) or 0.9 % NaCl (control). When sampled fish were rapidly killed by
388 cranial concussion and blood was removed by bleeding the fish from the *vena*
389 *caudalis*. Head kidney and spleen from 6 individuals were sampled at 6 hours, 2, 4
390 and 7 days post challenge from both the treated and untreated groups (n = 48). Head
391 kidney and spleen were aseptically removed and transferred to RNA-Later (Ambion)
392 and kept at 4 °C overnight before being stored at -80 °C. No mortality was recorded
393 in any of the tanks.

394 **RNA isolation, library preparation and sequencing**

395 The samples and controls were subjected to the TRIzol reagent (Invitrogen) RNA
396 isolation protocol. 30 mg of tissue was homogenized using sterile pistils in sterile 1.5
397 ml tubes (VWR) in 300 µl TRIzol reagent (Invitrogen). 60 µl of Chloroform (VWR
398 International) and subsequently 150 µl of isopropanol (Sigma Aldrich) were added
399 to the homogenate. Otherwise the TRIzol (Invitrogen) protocol was followed.

400 Some of the samples were taken from totalRNA to messengerRNA (mRNA) before
401 library preparation. mRNA isolation was performed using the Dynabeads® mRNA
402 direct kit (Life technologies) according to the manufacturers recommendations
403 (noted in sample overview in the GitHub repository). All RNA isolates (totalRNA or
404 mRNA) were quality controlled using an Agilent 2100 Bioanalyzer (BioRad) before
405 library preparation.

406 All libraries were prepared using the TruSeq™ RNA low-throughput (LT) protocol
407 (Illumina). Most samples were total RNA. mRNA samples were included before the
408 fragmentation step. All samples were fragmented for 4 minutes to obtain the size
409 distribution desired according to the TruSeq protocol. A library overview is
410 available the Github repository-

411 All libraries were sequenced 100bp paired-end (PE) at the Norwegian Sequencing
412 Centre on the Illumina HiSeq 2000 (www.sequencing.uio.no). Obtained sequences
413 were cleaned for adapters using Cutadapt version 1.0 [50]. Low quality regions were
414 trimmed using Sickle with a 40 bp minimum remaining sequence length, a Sanger
415 quality threshold of 20 and no 5' end trimming [51]. Results were quality controlled
416 using FastQC version 0.9.2 to ensure improvement compared to raw data [52].

417 **Reference-genome based approach using Tuxedo**

418 The second version of the Atlantic cod genome [25] was used as reference for a
419 Topphat/Cufflinks pipeline according to the workflow described in [23]. Mapping of
420 samples towards the reference-genome GFF3 file was performed with Tophat v2.0.14
421 with default settings. Sample-specific transcriptomes were generated with Cufflinks
422 v2.1.1. Cuffmerge was used to concatenate all the individual transcriptomes.
423 Differential expression analysis was performed with Cuffdiff in a pair-wise manner
424 between treated and control for each time-point. The output from Cuffdiff was
425 further analyzed using CummeRbund v2.8.2 in R v3.1.3 for presentation purposes
426 [53, 54].

427 **Reference-genome-guided approach using Trinity**

428 Two RNAseq studies provided reads for the transcriptome assembly used here – the
429 reads derived from the *Francisella* challenge described above and the reads derived
430 from the vibriosis vaccination study with the same number of samples (Solbakken et

431 al., paper V in this thesis). In total, the 96 libraries (48 from each experiment)
432 provided on average 20.51 million trimmed read-pairs resulting in 1 969.31 million
433 reads in total.

434 We applied the Trinity transcriptome assembler v 2.0.6 using the genome-guided
435 option with the second version of the Atlantic cod genome [25]. The genome was
436 indexed using Bowtie1 (v1.0.0) and then mapped using Tophat (v2.0.9) and sorted
437 with Samtools (v0.1.19). The built-in normalization step of Trinity was applied
438 reducing the trimmed read dataset to approximately 45 million read pairs [22, 55].
439 The following parameters were changed for the Trinity run: genome-guided, max
440 intron 10 000, max memory 150 Gb, bflyHeapSpaceMax 10G, bflyCPU 12 and CPU
441 10.

442 The assembly was evaluated with the built-in trinity_stats.pl and
443 align_and_estimate_abundance.pl — the latter with RSEM estimation method and
444 bowtie aligner. The abundance estimation output was further used to filter the
445 assembly on transcript level with FPKM = 1 using filter_fasta_by_rsem_values.pl.
446 This resulted in 44 543 transcripts with an overall contig N50 of 2 568 bp, median
447 contig length of 1 132 bp and a total of ~73.3 million assembled bases. Based on the
448 longest open reading frames (ORFs) the transcript dataset was reduced to 32 934
449 "genes" with an overall contig N50 of 2 490 bp and median contig length of 1 014 bp.

450 Overall annotation was performed using Trinotate v. 2.0.1 following all mandatory
451 steps with default parameters on the non-filtered assembly and transferred to the
452 filtered assembly transcripts. The annotation of genes specifically discussed in this
453 study have been verified through reciprocal BLAST by extracting the longest isoform
454 of the gene in question and subjecting it to a BLASTX towards all UniProt entries
455 using the UniProt BLAST tool with default settings [56].

456 **Sample mapping, read count extraction**

457 The trimmed reads from all Vibrio-related samples were mapped against the filtered
458 Trinity assembly to simplify interpretation of results using the built-in
459 align_and_estimate_abundance script in Trinity with the RSEM estimation method
460 and bowtie aligner, before extracting raw counts using
461 abundance_estimates_to_matrix.pl again with RSEM as the estimation method.

462 **Error distributions and differential expression analyses**

463 Most RNAseq analysis packages assume that such data follows a negative binomial
464 distribution of variability. We tested this assumption using a custom script testing
465 the fit of the Poisson distribution, the negative binomial distribution and the zero-
466 inflated negative binomial distribution (using the pscl package in R, script available
467 in the GitHub repository). About 90 % of all genes were classified as having negative
468 binomial distribution and thus, in all cases, the negative binomial distribution was
469 used for all down-stream analyses.

470 For the reference-genome based analysis CuffDiff performed the differential
471 expression analysis with default parameteres. For the Trinity-generated read-counts,
472 differential expression analysis was performed using the R-package edgeR
473 specifying the following contrasts: 6 hrs *Francisella* versus control, 2 day *Francisella*
474 versus control, 4 day *Francisella* versus control, and 7 day *Francisella* versus control,
475 and otherwise default settings.

476 **Custom script approach for gene expression pattern clustering**

477 We wanted to further characterize the behavior of the dataset outside of what the
478 most common RNAseq differential expression analysis packages could provide in
479 terms of the "genes'" being dependent on time and/or treatment (most analysis
480 packages provide pair-wise analysis options or time-series with a time 0 —not time-

481 series analysis with control samples for each time-point). Expression patterns were
482 to be classified into categories: increasing expression over time, decreasing
483 expression over time, expression pattern with an internal maximum (quadratic,
484 positive), expression pattern with an internal minimum (quadratic, negative) and
485 freestyle expression pattern (alternating trends over time) — in both control-
486 dependent and independent manners. In addition two categories named no
487 control/no time dependency and control dependency were added. Note that if a
488 quadratic effect was found but with minima/maxima outside the data material, it
489 would be classified as either increasing or decreasing, depending on the estimated
490 quadratic effect.) This categorization was performed with a set of regression models;
491 no time dependency, linear time dependency, quadratic time dependency, factorial
492 time dependency, pure treatment effect (no time dependency), treatment combined
493 with linear time (interaction), treatment combined with quadratic time (interaction)
494 and treatment combined with factorial time (interaction). Estimated regression
495 coefficients were then used for determining in which time dependency category each
496 gene expression was to be classified.

497 **Evaluation of RNAseq experiment**

498 The overall quality evaluation of the samples revealed similar trends for dispersion
499 and good clustering of treated and control samples with the exception of the 6 hrs
500 time-point which displayed some overlap in sample clustering (Supplementary
501 figures 1,2). The primary differential expression analysis (cutoff p=0.05) reported in
502 total 3 329 differentially expressed genes (DEGs) (Table 1).

503 **GO and gene network analyses**

504 The reported differentially expressed genes from the primary analyses were
505 analyzed in Cytoscape [57] using the plugin ClueGO [58]. ClueGO was run with

506 default settings selecting biological and immunological related systems and a p-
507 value cutoff of 0.05 unless otherwise stated.

508 ## References

509 1. Star, B., et al., *The genome sequence of Atlantic cod reveals a unique immune system.* Nature,
510 2011. **477**(7363): p. 207-10.
511 2. Zhu, L.Y., et al., *Advances in research of fish immune-relevant genes: a comparative overview*
512 *of innate and adaptive immunity in teleosts.* Dev Comp Immunol, 2013. **39**(1-2): p. 39-62.
513 3. Malmstrom, M., et al., *Unraveling the evolution of the Atlantic cod's (Gadus morhua L.)*
514 *alternative immune strategy.* PLoS One, 2013. **8**(9): p. e74004.
515 4. Solbakken, M.H., et al., *Evolutionary redesign of the Atlantic cod (Gadus morhua L.) Toll-like*
516 *receptor repertoire by gene losses and expansions.* Sci Rep, 2016. **6**: p. 25211.
517 5. Malmstrøm, M., et al., *Evolution of the immune system influences speciation rates in teleost*
518 *fishes.* Nat Genet, In press.
519 6. Birkbeck, T.H., S.W. Feist, and D.W. Verner-Jeffreys, *Francisella infections in fish and shellfish.*
520 J Fish Dis, 2011. **34**(3): p. 173-87.
521 7. Colquhoun, D.J. and S. Duodu, *Francisella infections in farmed and wild aquatic organisms.*
522 Vet Res, 2011. **42**: p. 47.
523 8. Soto, E., et al., *Identification of Francisella noatunensis in novel host species French grunt*
524 *(Haemulon flavolineatum) and Caesar grunt (Haemulon carbonarium).* J Zoo Wildl Med, 2014.
525 **45**(3): p. 727-31.
526 9. Leal, C.A., G.C. Tavares, and H.C. Figueiredo, *Outbreaks and genetic diversity of Francisella*
527 *noatunensis subsp orientalis isolated from farm-raised Nile tilapia (Oreochromis niloticus) in*
528 *Brazil.* Genet Mol Res, 2014. **13**(3): p. 5704-12.
529 10. Soto, E., et al., *Prevalence of Francisella noatunensis subsp. orientalis in cultured tilapia on*
530 *the island of Oahu, Hawaii.* J Aquat Anim Health, 2013. **25**(2): p. 104-9.
531 11. Isachsen, C.H., et al., *Antimicrobial susceptibility of Francisella noatunensis subsp.*
532 *noatunensis strains isolated from Atlantic cod Gadus morhua in Norway.* Dis Aquat Organ,
533 2012. **98**(1): p. 57-62.
534 12. Soto, E., et al., *Efficacy of florfenicol for control of mortality associated with Francisella*
535 *noatunensis subsp. orientalis in Nile tilapia, Oreochromis niloticus (L.).* J Fish Dis, 2013. **36**(4):
536 p. 411-8.

537 13. Steiner, D.J., Y. Furuya, and D.W. Metzger, *Host-pathogen interactions and immune evasion*
538 *strategies in Francisella tularensis pathogenicity.* Infect Drug Resist, 2014. **7**: p. 239-51.
539 14. Jones, C.L., et al., *Subversion of host recognition and defense systems by Francisella spp.*
540 Microbiol Mol Biol Rev, 2012. **76**(2): p. 383-404.
541 15. Asare, R. and Y.A. Kwaik, *Exploitation of host cell biology and evasion of immunity by*
542 *francisella tularensis.* Front Microbiol, 2010. **1**: p. 145.
543 16. Vestvik, N., et al., *Francisella noatunensis subsp. noatunensis replicates within Atlantic cod*
544 *(Gadus morhua L.) leucocytes and inhibits respiratory burst activity.* Fish Shellfish Immunol,
545 2013. **35**(3): p. 725-33.
546 17. Bakkemo, K.R., et al., *Intracellular localisation and innate immune responses following*
547 *Francisella noatunensis infection of Atlantic cod (Gadus morhua) macrophages.* Fish Shellfish
548 Immunol, 2011. **31**(6): p. 993-1004.
549 18. Furevik, A., et al., *The intracellular lifestyle of Francisella noatunensis in Atlantic cod (Gadus*
550 *morhua L.) leucocytes.* Fish Shellfish Immunol, 2011. **30**(2): p. 488-94.
551 19. Brudal, E., et al., *Establishment of three Francisella infections in zebrafish embryos at*
552 *different temperatures.* Infect Immun, 2014. **82**(6): p. 2180-94.
553 20. Vojtech, L.N., et al., *Host immune response and acute disease in a zebrafish model of*
554 *Francisella pathogenesis.* Infect Immun, 2009. **77**(2): p. 914-25.
555 21. Schrøder, M.B., et al., *Comparison of antibody responses in Atlantic cod (Gadus morhua L.) to*
556 *Vibrio anguillarum, Aeromonas salmonicida and Francisella sp.* Fish Shellfish Immunol, 2009.
557 **27**(2): p. 112-9.
558 22. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-seq using the Trinity*
559 *platform for reference generation and analysis.* Nat Protoc, 2013. **8**(8): p. 1494-512.
560 23. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq*
561 *experiments with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.
562 24. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for*
563 *differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p.
564 139-40.
565 25. Tørresen, O.K., et al., *An improved genome assembly uncovers a profilic tandem repeat*
566 *structure in Atlantic cod.* Submitted.
567 26. Polepalle, T., et al., *Acute Phase Proteins and Their Role in Periodontitis: A Review.* J Clin
568 Diagn Res, 2015. **9**(11): p. Ze01-5.

569 27. Riera Romo, M., D. Perez-Martinez, and C. Castillo Ferrer, *Innate immunity in vertebrates: an*
570 *overview.* Immunology, 2016.
571 28. Szabady, R.L. and B.A. McCormick, *Control of neutrophil inflammation at mucosal surfaces by*
572 *secreted epithelial products.* Front Immunol, 2013. **4**: p. 220.
573 29. Havixbeck, J.J. and D.R. Barreda, *Neutrophil Development, Migration, and Function in Teleost*
574 *Fish.* Biology (Basel), 2015. **4**(4): p. 715-34.
575 30. Ronneseth, A., H.I. Wergeland, and E.F. Pettersen, *Neutrophils and B-cells in Atlantic cod*
576 *(Gadus morhua L.).* Fish Shellfish Immunol, 2007. **23**(3): p. 493-503.
577 31. Montaldo, E., et al., *Human NK cell receptors/markers: a tool to analyze NK cell development,*
578 *subsets and function.* Cytometry A, 2013. **83**(8): p. 702-13.
579 32. Storek, K.M. and D.M. Monack, *Bacterial recognition pathways that lead to inflammasome*
580 *activation.* Immunol Rev, 2015. **265**(1): p. 112-29.
581 33. Schwartz, J.T., et al., *Francisella tularensis inhibits the intrinsic and extrinsic pathways to*
582 *delay constitutive apoptosis and prolong human neutrophil lifespan.* J Immunol, 2012. **188**(7):
583 p. 3351-63.
584 34. Lai, X.H., et al., *Macrophage cell death upon intracellular bacterial infection.* Macrophage
585 (Houst), 2015. **2**: p. e779.
586 35. Drickamer, K. and M.E. Taylor, *Recent insights into structures and functions of C-type lectins*
587 *in the immune system.* Curr Opin Struct Biol, 2015. **34**: p. 26-34.
588 36. Kawasaki, T. and T. Kawai, *Toll-like receptor signaling pathways.* Front Immunol, 2014. **5**: p.
589 461.
590 37. Gazi, U. and L. Martinez-Pomares, *Influence of the mannose receptor in host immune*
591 *responses.* Immunobiology, 2009. **214**(7): p. 554-61.
592 38. McCaffrey, R.L. and L.A. Allen, *Francisella tularensis LVS evades killing by human neutrophils*
593 *via inhibition of the respiratory burst and phagosome escape.* J Leukoc Biol, 2006. **80**(6): p.
594 1224-30.
595 39. McCaffrey, R.L., et al., *Multiple mechanisms of NADPH oxidase inhibition by type A and type*
596 *B Francisella tularensis.* J Leukoc Biol, 2010. **88**(4): p. 791-805.
597 40. Rabadi, S.M., et al., *Antioxidant Defenses of Francisella tularensis Modulate Macrophage*
598 *Function and Production of Proinflammatory Cytokines.* J Biol Chem, 2016. **291**(10): p. 5009-
599 21.

600 41. Case, E.D., et al., *The Francisella O-antigen mediates survival in the macrophage cytosol via*
601 *autophagy avoidance.* Cell Microbiol, 2014. **16**(6): p. 862-77.
602 42. Wilson, J.E., B. Katkere, and J.R. Drake, *Francisella tularensis induces ubiquitin-dependent*
603 *major histocompatibility complex class II degradation in activated macrophages.* Infect
604 Immun, 2009. **77**(11): p. 4953-65.
605 43. Neefjes, J., et al., *Towards a systems understanding of MHC class I and MHC class II antigen*
606 *presentation.* Nat Rev Immunol, 2011. **11**(12): p. 823-836.
607 44. Vinuesa, C.G. and P.P. Chang, *Innate B cell helpers reveal novel types of antibody responses.*
608 Nat Immunol, 2013. **14**(2): p. 119-26.
609 45. Mikkelsen, H. and M. Seppola, *Response to vaccination of Atlantic cod (Gadus morhua L.)*
610 *progenies from families with different estimated family breeding values for vibriosis*
611 *resistance.* Fish Shellfish Immunol, 2013. **34**(1): p. 387-92.
612 46. Mikkelsen, H., et al., *Vibriosis vaccines based on various sero-subgroups of Vibrio*
613 *anguillarum O2 induce specific protection in Atlantic cod (Gadus morhua L.) juveniles.* Fish
614 Shellfish Immunol, 2011. **30**(1): p. 330-339.
615 47. Gudmundsdottir, S., et al., *Specific and natural antibody response of cod juveniles vaccinated*
616 *against Vibrio anguillarum.* Fish Shellfish Immunol, 2009. **26**(4): p. 619-24.
617 48. Mikalsen, J., et al., *Francisella philomiragia subsp. noatunensis subsp. nov., isolated from*
618 *farmed Atlantic cod (Gadus morhua L.).* Int J Syst Evol Microbiol, 2007. **57**(Pt 9): p. 1960-5.
619 49. Olsen, A.B., et al., *A novel systemic granulomatous inflammatory disease in farmed Atlantic*
620 *cod, Gadus morhua L., associated with a bacterium belonging to the genus Francisella.* J Fish
621 Dis, 2006. **29**(5): p. 307-11.
622 50. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.*
623 Bioinformatics in Action, 2012. **17**(1): p. 10-12.
624 51. Joshi, N., *Sickle - a windowed adaptive trimming tool for FASTQ files using quality*.
625 52. Andrews, S. *The FastQC project*. 2011; Available from:
626 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
627 53. R-Core-Team. *R: A Language and Environment for Statistical Computing*. 2015; Available
628 from: http://www.R-project.org.
629 54. Goff, L., C. Trapnell, and D. Kelley, *CummeRbund: Analysis, exploration, manipulation, and*
630 *visualization of Cufflinks high-throughput sequencing ata*. 2013.

631 55. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a*
632 *reference genome.* Nat Biotechnol, 2011. **29**(7): p. 644-52.
633 56. UniProt, *UniProt: a hub for protein information.* Nucleic Acids Res, 2015. **43**(Database issue):
634 p. D204-12.
635 57. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular*
636 *interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
637 58. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene*
638 *ontology and pathway annotation networks.* Bioinformatics, 2009. **25**(8): p. 1091-3.
639 59. Tanabe, M. and M. Kanehisa, *Using the KEGG database resource.* Curr Protoc Bioinformatics,
640 2012. **Chapter 1**: p. Unit1 12.

641

642 ## Acknowledgements

643 This work was supported by The Research Council of Norway (Grant number
644 199806/S40 to KSJ. The reference-genome assembly was made using the Abel Cluster,
645 owned by the University of Oslo and the Norwegian metacenter for High
646 Performance Computing (NOTUR), and operated by the Department for Research
647 Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/.
648 The sequencing service was provided by the Norwegian Sequencing Centre
649 (www.sequencing.uio.no), a national technology platform hosted by the University
650 of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs
651 of the Research Council of Norway and the Southeastern Regional Health
652 Authorities.

653 ## Additional information

654 This manuscript has a GitHub repository providing all data.

655 ## Competing financial interests

656 The authors declare no competing financial interests.

Corresponding author

Monica Hongrø Solbakken, m.h.solbakken@ibv.uio.no and Sissel Jentoft; sissel.jentoft@ibv.uio.no

661 ## Tables and figures

662 **Table 1** Genes reported as significantly different from control, with or without
663 corresponding annotation, for all analysis approaches applied (CuffDiff, EdgeR, Custom
664 scripts). For the custom script output, only genes reported with expression patterns in a
665 control dependent manner are depicted. Rows corresponding to EdgeR are derived from the
666 de novo transcriptome whereas those derived from CummeRbund are derived from the
667 reference-genome gene models.

Method	Time-point or pattern	No of "genes"	No of annotated "genes"
EdgeR	6hrs up	26	10
	2day up	294	142
	4day up	294	134
	7day up	134	67
	6hrs down	5	1
	2day down	48	22
	4day down	179	104
	7day down	181	98
Custom script	Increase	485	213
	Internal max	181	87
	Decrease	2688	1385
	Internal min	975	594
	Freestyle	895	608
CummeRbund	6hrs up	64	33
	2day up	679	504
	4day up	751	5
	7day up	529	373
	6hrs down	26	5
	2day down	199	145
	4day down	480	341
	7day down	601	424
Custom script	Increase	588	458
	Internal max	576	370
	Decrease	859	543
	Internal min	1075	669
	Freestyle	1246	817

668

669 **Table 2** All major GO:term clusters reported by ClueGO in Cytoscape (biological and
670 immunological processes) with a p-value cutoff = 0.05. Gene annotations dervied from
671 CummeRbund only.* have a p-value cutoff = 0.0001.

Time	GO:terms
6 hrs up	Muscle filament sliding
	Negative regulation of NK-kappaB transcription factor activity
	Regulation of granulocyte chemotaxis
6 hrs down	No significant GO:terms reported
2 days up	Antigen processing and presentation ofpeptide antigen via MHC class I
	Response to cytokine
	Response to oxygen-containing compound
	Ribosome biogenesis
	Ferric iron transport
	Establishment of protein localization
2 days down	Cell-substrate adherens junction assembly
	Triglyceride catabolic process
	Trabecula formation
	Regulation of epithelial to mesenchymal transition
	Autophagosome assembly
	Cellular response to ketone
4 days up*	Cellular response to organic substance
	Response to organic substance
	Response to cytokine
	Cellular response to chemical stimulus
	Regulation of apoptopic process

	Single-organism transport
	Extracellular matrix organization
	Ferric iron transport
	Anatomical structure morphogenesis
	Single organism cell adhesion
	Regulation of cell proliferation
4 days down	Collagen catabolic process
	Single-organism carbohydrate catabolic process
	Negative regulation of membrane potential
	Protein trimerization
7 days up*	Antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent
	Hydrogen ion transmembrane transport
	Amino acid activation
	Response to unfolded protein
7 days down	Regulation of response to wounding
	Response to wounding
	Viral entry in host cell
	Single organismal cell-cell adhesion
	Extracellular matrix organization
	Cell activation
	Negative regulation of wound healing
	Hemopoiesis
	Positive regulation of secretion by cell

672

673 **Table 3** Key genes involved in immunological processes. For each gene the results derived
674 from Cufflinks/CummeRbund are depicted. Results from Trinity/EdgeR are presented if
675 there were no significant findings reported by Cufflinks. * gene has been manually
676 annotated. NS = not significant

Gene	Up	Down	Pattern	R package
Cytokines, chemokines and inflammasome related				
ASC			Freestyle	edgeR
CASP1	2,4,7			CummeRbund
CD40	2			CummeRbund
CXCL8s	6,2,4,7			CummeRbund
IFNg*	4			edgeR
IL10	6,2,4,7			edgeR
IL12B	2,4			CummeRbund
IL1B	6,2,4,7			CummeRbund
IL6*			NS	
MCP-1 (CCL2)	2,4,7			CummeRbund
TGFb (TGFB3)	4,7			CummeRbund
TNF*			NS	
Pattern recognition				
MRC1	2	4,7		CummeRbund
NLRC3 variant	6			CummeRbund
NLRC3 variant		6		CummeRbund
NLRC3 variant	2			CummeRbund
NLRP12	4			CummeRbund
TLR13 (TLR23)		4,7		CummeRbund
TLR22			Int. min	CummeRbund
TLR21			Int.min	CummeRbund
TLR2/6 TLR25	2,4,7			edgeR
Complement				
C1Q (L2/TNF3)		4		CummeRbund
C3	4	7		CummeRbund
C4	4			CummeRbund
C7	4			CummeRbund
C8G			Freestyle	edgeR
Antimicrobials, acute-phase, iron homeostasis				
CP	2			CummeRbund
CRP		4		CummeRbund

F3			Int.max	CummeRbund
FGB	6	7		CummeRbund
FGG	2,4	7		
FTH1	7	4		CummeRbund
HAMP	2,4,7			CummeRbund
IL4L1 (LAO)	2,4,7			CummeRbund
LYG2	2,4,7			CummeRbund
PTX3	2,4	7		edgeR
SERPINE1	4			CummeRbund
SERPINA1		7		CummeRbund
SLC40A1		2,4,7		CummeRbund
TF	2,4			CummeRbund
Apoptosis				
BAX	2,7			CummeRbund
CASP3	2,4,7			CummeRbund
CASP6	2,7	4		CummeRbund
CASP7	2,4,7			CummeRbund
FASLG			Decrease	CummeRbund/edgeR
GZMa		4,7		CummeRbund
GZMB	2,4,7			CummeRbund
MHC, TCR, BCR and related				
CD8A*			Decrease	edgeR
HLA-A	2			CummeRbund
IGKC		4,7		CummeRbund
IGLC6	2	7		CummeRbund
TRBC2			Decrease	CummeRbund
T-cell subsets and functions				
BCL6			Decrease	edgeR
CCR7			Decrease	edgeR
CXCR5*			Decrease	CummeRbund
GZMA		4,7		CummeRbund
GZMB	2,4,7			CummeRbund
PRF1	4			CummeRbund
NK-cell markers				
ITGAL			Int.max	CummeRbund
CD132 (IL2RG)			Decrease	edgeR
CD244		4		CummeRbund/edgeR
IL7R (CD127)		7		edgeR
NCAM1(CD56)	2,4			CummeRbund

677

678 **Table 4 All major GO:term clusters reported by ClueGO in Cytoscape (biological and**
679 **immunological processes) with a p-value cutoff = 0.05unless * which is p=0,0001.**

Pattern	GO:terms (Custom scripts)
Increasing ctr dependent*	Antigen processing and presentation of peptide antigen via MHC class I
	Cellular amino acid metabolic process
	Energy coupled proton transmembrane transport against electrochemical gradient
	Organonitrogen compound metabolic process
	Mitochondrial transmembrane transport
Internal maximum ctr dependent	Negative regulation of intracellular signal transduction
	Response to organic cyclic compound
	Response to lipopolysaccharide
	Positive regulation of apoptopic process
	Negative regulation of cell proliferation
	Cytokine-mediated signaling pathway
	Response to cytokine
	Single organismal cell-cell adhesion
	Extracellular matrix organization
	Response to unfolded protein
	Type I interferon signaling pathway
	Negative regulation of apoptosis
	Heterotypic cell-cell adhesion
	Positive regulation of secretion
	Regulation of sequence-specific DNA bindnig transcription factor activity

	Cell junction organization
	Negative regulation of cell cycle
	Positive regulation of cell proliferation
	Endoderm formation
Decreasing ctr dependent	Wound healing
	Negative regulation of membrane potential
	Vesicle-mediated transport
	Fructose metabolic process
Internal minimum ctr dependent	Diterpenoid metabolic process
	Cellular modified amino acid biosynthetic process
	Arachidonic acid metabolic process
	Cellular aldehyde metabolic process
	Protein trimerization
	Glutathione metabolic process
Freestyle ctr dependent	Positive regulation of ubiquitin-protein transferase activity
	Response to unfolded protein
	Substantia nigra development
	ATP synthesis coupled electron transport
	Copper ion transport
	Chaperone-mediated protein folding

680

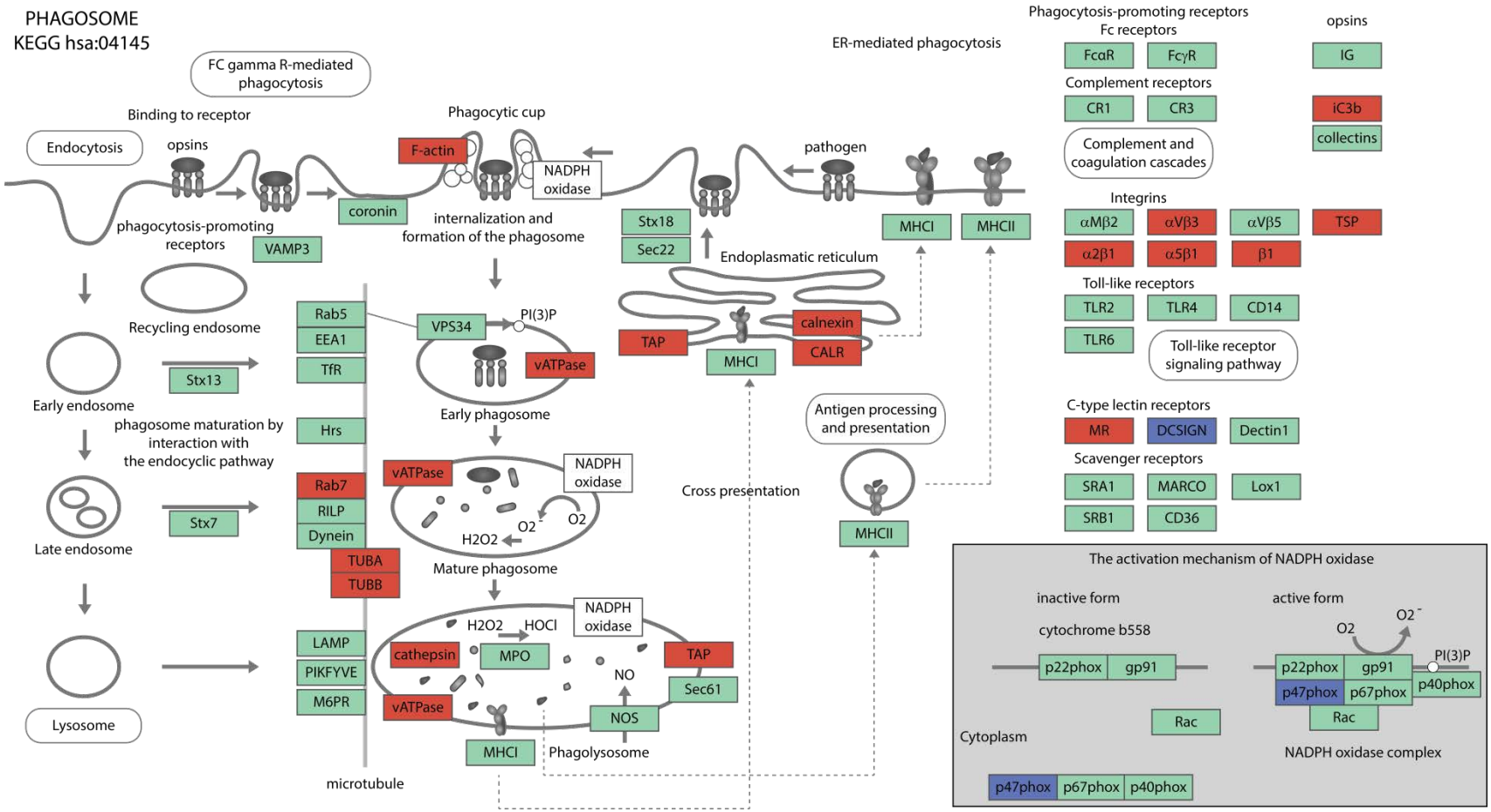


Figure 1 Mapping of genes reported as significantly differentially expressed at day 4 post infection in the TCC pipeline. Red denoted genes are up-regulated and blue denoted genes are down-regulated. Green denoted genes where expression data have not been provided. The pathway drawn after the phagosome maturation and related processes pathway has:04145 obtained from KEGG [59].

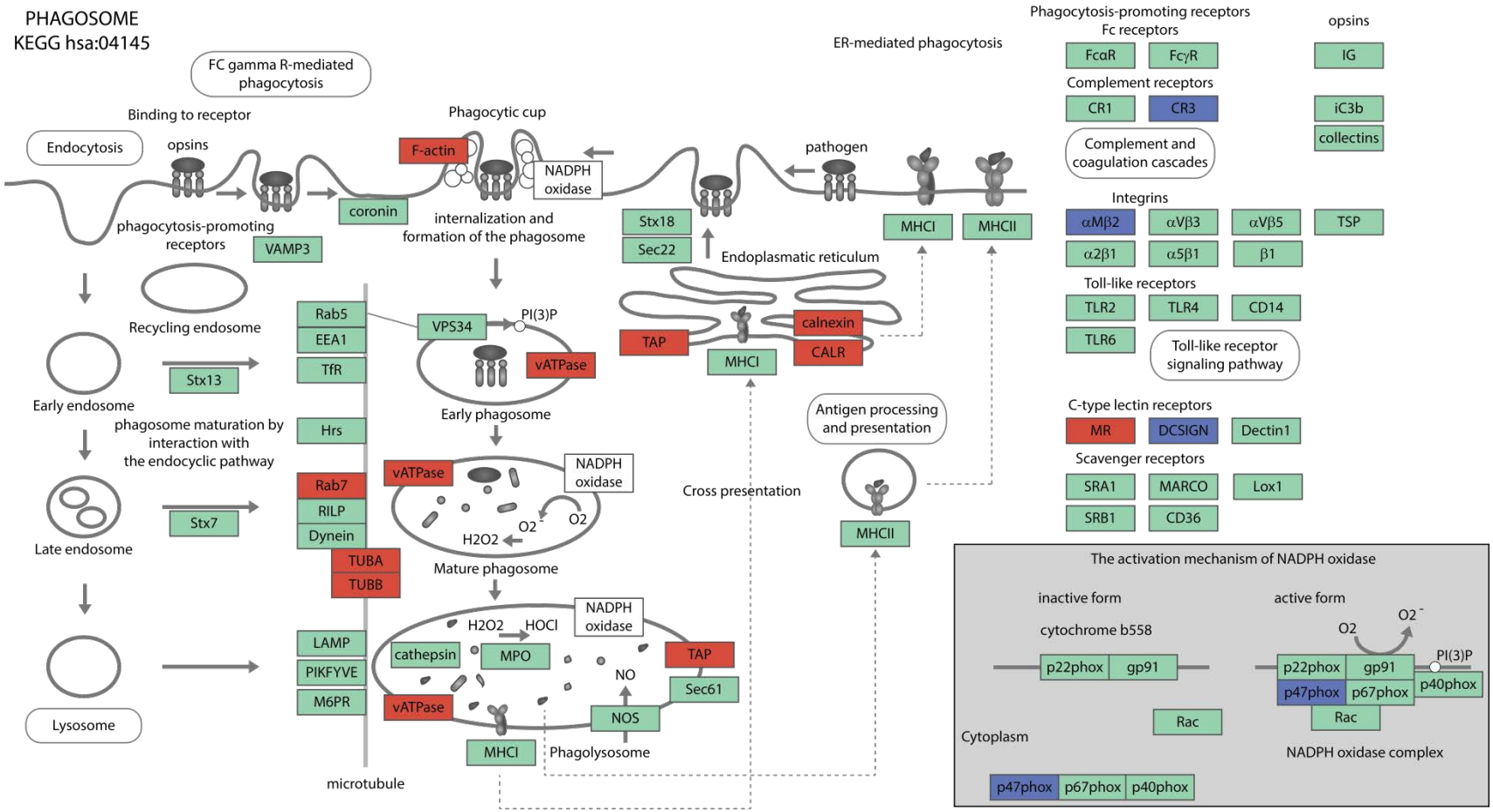
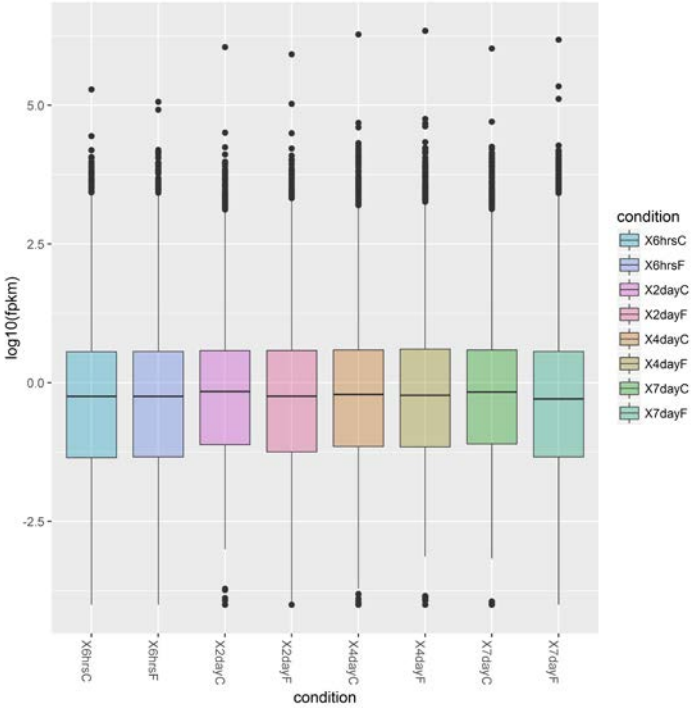
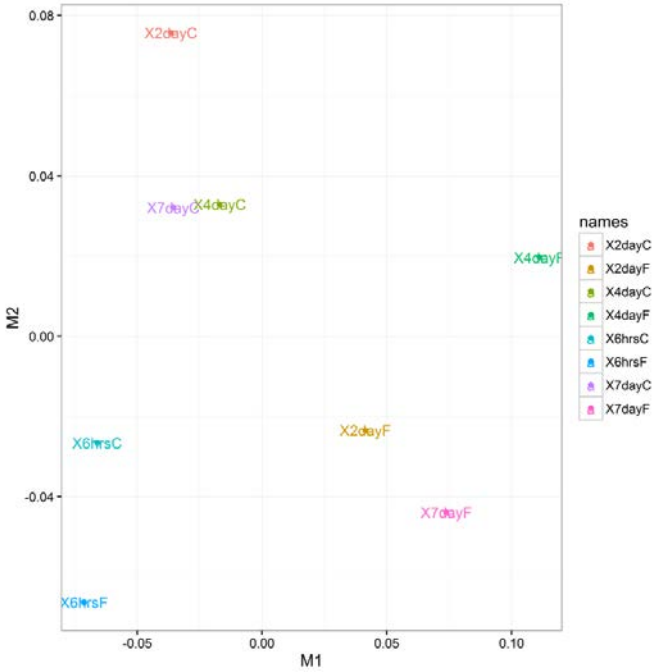


Figure 2 Mapping of genes reported as significantly differentially expressed at day 7 post infection in the TCC pipeline. Red denoted genes are up-regulated and blue denoted genes are down-regulated. Green denoted genes where expression data have not been provided. The pathway drawn after the phagosome maturation and related processes pathway has:04145 obtained from KEGG [59].



Supplementary figure 1 Box plot displaying the overall expression concatenated for all biological replicates for each sample. Note: one sample in 6hrsC (6hrsC_1) that deviates from the others with respect to coverage of lowly expressed genes (data not shown). Made using CummeRbund.



Supplementary figure 2 MDS plot of all samples with concatenated biological replicates Made using CummeRbund.

Supplementary table 1 All annotated genes reported with significant differential expression in the Tophat-Cufflinks-CuffDiff pipeline.

6hrs annotation	6hrs log2 fold change	2day annotation	2 day log2 fold change	4day annotation	4 day log2 fold change	7 day annotation	7 day log2 fold change
ACTA1	3.41305	ABCB9	2.03448	AADAC	-2.72479	AADAC	-3.306
ACTA1	2.8163	ABCB9	2.40409	ABCA4	-1.82778	ABCA4	-2.0101
CASP3	2.59719	ABCE1	1.16637	ABCB10	1.13727	ABCB5	-1.22552
CASQ1	3.22836	ABCE1	1.35214	ABCB9	2.91493	ABCB9	1.94737
CASQ2	3.57993	ABCF2	1.44709	ABCF2	1.16283	ABCB9	1.86518
CH25H	2.21089	ABHD6/ PXK	1.3773	ABHD5	1.82026	ABHD12	-1.46979
CKM	3.6438	ABTB2	1.6676	ABHD6-B/ PXK	1.57327	ABTB1	-1.14954
CKM	2.8749	ADGRG3	1.5375	ABTB1	-1.30812	ABTB2	1.66809
CLEC3B	1.67234	ADM2	2.14682	ABTB2	2.37031	ACE2	3.20857
CMKLR1	2.76076	AG2	1.93823	ACO2	1.00309	ACE2	3.44271
CXCL8	5.1393	AHSA1	1.20903	ACSS1	-2.47772	ACER2	-1.15872
CXCL8	5.41006	AIMP2	1.14016	ACTB	1.55172	ACKR4	2.55288
FGG	6.22989	AK1	1.07804	ADAMTS1	1.19403	ACSS1	-1.75585
GALAXIN	1.88955	AKAP12	1.48309	ADAP1	-1.26231	ACTA1	-2.18038
GGH	2.1797	ALCAM	1.4009	ADCK3	-2.32161	ACTB	0.999559
GNA12	3.68133	ALDH1A2	1.25612	ADCY6	-1.12459	ADAM19	-1.6254
IL1B	6.8073	ALOX12B	5.48423	ADCYAP1	-1.70528	ADAM8	-1.987
IRGC	1.28134	ALOX15B	5.78055	ADGRG3	2.41697	ADCK3	-1.15158
MUSTN1	3.47056	ALOX15B	6.3862	ADM	1.69147	ADCY6	-1.22691
MYH4	4.69311	ALPK1	2.34762	ADM2	2.34496	ADD2	-1.11599
MYL1	3.81345	ALYREF-B	0.869499	ADM2	1.85171	ADGRG3	1.21548
NFKBIA	1.11386	ANGPT1	1.56703	ADSS	1.14043	ADH1	-1.49647
NLRC3	3.6088	ANXA13	1.53648	AGA	0.908851	ADM2	0.948789

PSBP1	4.18608
PVALB	3.29265
PVALB	4.17757
RPL9	1.38924
RPLP2	1.30582
SEPP1	4.03052
SPIRE1	2.05643
TF	1.99552
TNNC2	2.76917
TPM1	3.64055
HSPA1A	-2.05883
HSPA1A	-3.65978
NLRC3	-2.45861
PSG1	-2.36292
SAMHD1	-3.07839

APBB3	2.38099
AQP3	2.46097
ARL4C	2.49458
ARL8BA	1.01082
ASS1	1.35782
ATAD1	1.5842
ATF3	1.33621
ATP2A2	0.971239
ATP6AP1	1.16736
ATP6V0A2	1.29108
ATP6V0B	1.77696
ATP6V0C	1.20708
ATP6V0D1	1.05505
ATP6V1C1A	1.59352
ATP6V1E1	1.47308
ATP6V1G1	1.0091
BAG2	1.53105
BAX	1.15773
BCL2L14	1.07481
BPI	2.67314
BTG3	1.6532
BTG4	1.16546
BYSL	1.34
C21ORF33	1.48444
CALR	1.13297
CALR	1.65841

AGL	-1.21448
AGPAT9L	-1.52998
AHCYL2	1.00538
AIMP2	1.30028
AKAP12	2.63769
AKAP2	1.0509
AKT2	1.13346
ALCAM	2.39826
ALDH1A2	1.99307
ALOX12B	5.81733
ALOX12B	5.12489
ALOX15B	5.18723
ALOX5AP	-1.14794
ALPK1	2.93558
ALPL	-1.6186
AMIGO3	-3.49462
AMY1	-0.996126
ANGPT1	1.74149
ANGPTL2	-2.11867
ANGPTL4	2.66414
ANGPTL4	1.80153
ANPEP	-1.98757
ANPEP	-1.9177
ANXA13	-2.43732
APBB3	2.82982
APLP2	1.35674

ADRB2	-1.66885
ADSS	0.937509
AGT	-4.10138
AHSA1	0.985169
AHSG	-Inf
AIMP1	1.49705
AIMP2	1.54777
ALAS1	-1.39393
ALDH9A1	-1.10294
ALOX15B	4.47648
ALOX5AP	-1.63696
ALPK1	2.65313
ALYREF-B	0.891951
AMY1	-3.01419
ANGPT1	1.11266
ANGPTL2	-2.23635
ANK1	-1.78554
ANO1	-Inf
ANPEP	-1.51558
ANTXR1	-1.40643
APBB3	2.17917
APCS	-1.39618
APOEB	-1.65423
APOH	-2.42005
APOL6	2.07487
APOLIPOPROTEIN A-I-2	-5.95985

CALR	1.61604	AQP8	-0.95827	AQP3	-2.54797
CALU	2.06789	ARHGAP12	1.12989	ARFIP2	1.77703
CALUB	1.24011	ARHGAP21	2.17216	ARHGEF12	-1.13257
CANX	1.29492	ARHGAP21	2.55651	ARL13B	-1.8919
CARS	1.25675	ARHGEF12	-0.870185	ARL3	0.901343
CASP1	1.70532	ARHGEF19	-4.92328	ARL4C	1.46921
CASP1	1.92232	ARL13B	-1.18921	ARL8BA	1.43018
CASP3	1.77274	ARL4C	1.87718	ARMC3	-1.65608
CASP3	1.09483	ARL9	-1.41352	ARPC1A	-1.08382
CASP3	2.56298	ARMC3	-1.87685	ARPC1A	-1.21707
CASP3	1.64384	ARPC1A	-1.20951	ART1	-2.15747
CASP6	1.10906	ARPC1A	-1.37171	ASNS	1.43529
CASP7	0.918064	ARRDC3	1.13511	ASS1	1.46236
CCDC137	1.18242	ASIC1	-1.6087	ATF2	1.03952
CCDC25	1.16857	ASNS	1.67053	ATP5I	1.11748
CCDC43	1.47827	ATAD1B	1.33422	ATP5J	1.37311
CCDC79	2.29396	ATAD3	1.49709	ATP5J2	1.14533
CCDC86	1.2168	ATF3	1.73425	ATP6V0A2	1.35529
CCL2	1.11869	ATG4D	1.07696	ATP6V0B	1.26944
CCNY	1.10789	ATP2A2	1.28832	ATP6V0C	1.12659
CD209	3.20281	ATP6V0A2	1.71776	ATP6V0D1	1.15166
CD276	1.18801	ATP6V0B	1.20681	ATP6V1A	1.73101
CD276	1.33869	ATP6V0D1	1.59553	ATP6V1B2	1.71519
CD38	2.1704	ATP6V1A	2.06328	ATP6V1C1A	1.36696
CD40	1.32472	ATP6V1B2	1.64002	ATP6V1E1	2.0109

CDA	1.59777	ATP6V1C1A	1.26218	ATP6V1F	1.48409
CDH1	1.29972	ATP6V1E1	1.58755	ATP6V1G1	1.77358
CDK2AP1	1.2713	ATP6V1G1	1.36859	ATP6V1H	1.11203
CDO1	1.53344	ATP6V1H	1.38145	ATP8B2	-2.71929
CDR2	2.27017	ATP8B2	-2.66697	ATXN2L	-1.41419
CEBPB	1.11671	ATPIF1	-1.26833	B3GALT2	-1.92575
CFP	2.22655	ATXN2L	-1.35283	B3GNT7	-1.35131
CH25H	3.67372	B3GNT7	-1.43796	BACH2	-1.62075
CHCHD4	1.47999	BAG2	1.02632	BACH2	-1.54916
CILP2	2.19264	BANF1	2.12831	BANF1	1.53953
CLCN7	1.18345	BCL2L14	1.07954	BANP	-1.72597
CLDN1	4.34	BHLHE40	1.63431	BAX	1.57848
CLDN5	1.66915	BHLHE40	1.44009	BCL11A	-1.33683
CLIC2	1.16522	BNIP3	-2.64211	BCL11B	-1.22606
CMLKR1	2.91056	BNIPL	-2.25953	BCO1	-2.97643
CMLKR1	5.82942	BPI	2.1007	BHLHB3	3.41629
CNDP2	2.03486	BRD4	1.10979	BLNK	-1.21562
CNDP2	1.26236	BTG4	1.03677	BNIPL	-1.54227
CNFN	1.59781	C1GALT1B	1.40079	BOLA2	0.964128
CNN2	1.031	C1QL2	-3.65572	BTG3	1.05158
CNPY1	2.24334	C1QTNF3	-2.51227	C10ORF11	1.5877
COPZ1	0.946954	C21ORF33	1.22208	C21ORF33	1.34299
COX5A	1.38034	C2CD4CC2CD4 FAMILY	-1.75421	C2CD4CC2CD4 FAMILY	-3.30912
CP	2.19792	C3	2.98253	C5AR1	-1.71746
CREBBP	1.31515	C4	1.6701	CA6	-2.50775
CRELD2	1.68866	C7	2.00274	CADM4	1.06193

CREM	3.80324	CABP1	-1.73062	CALCOCO1	-1.48338
CSRNP1	1.89016	CABP4	-1.17843	CALHM3	-3.51041
CSRP1	2.25358	CALHM3	-2.77614	CALR	1.36871
CTSS	1.03661	CALR	1.22863	CALR	1.59742
CXCL8	7.01863	CALR	0.989608	CALUA	1.38619
CXCL8	7.01539	CALUA	1.13434	CAMK1	-1.18905
CXCL8	1.97958	CALUB	1.0588	CANX	1.15317
CXCR2	1.33297	CAMK1	-0.933037	CARS	1.09107
CYB5R2	1.15828	CANX	1.28983	CASP1	2.06565
CYC	1.95522	CARNS1	-1.44685	CASP1	2.42994
CYC-B	1.36687	CASP1	1.1865	CASP1	1.69915
CYLD	0.96173	CASP3	1.7789	CASP3	1.48255
CYR61	1.83062	CASP3	2.45425	CASP3	0.962432
DCUN1D5	1.84761	CASP3	2.09539	CASP3	1.57878
DDIT4L	1.8848	CASP3/OSBP	1.23888	CASP6	0.95444
DHX35	0.989009	CASP6	-1.36515	CASP6	1.13069
DHX58	1.56554	CASP7	1.14653	CASP7	1.0854
DLC1	1.05476	CBS	2.37715	CAST	-2.1115
DMBT1	1.79243	CBX1	-0.927916	CAV2	-1.6197
DNAJB11	1.82166	CBX4	-1.31733	CBX4	-1.70258
DOCK9	1.67988	CCDC79	2.75338	CCDC115	1.29502
DPP3	1.29587	CCL2	1.92013	CCDC136	-1.23921
DPP3	2.50574	CCND1	0.909611	CCDC3	-2.16041
DRAM1	2.76998	CCNY	0.980886	CCDC43	1.29353
DTX3L	1.47398	CCR2	-1.84569	CCDC79	2.06085
DUSP1	2.34908	CD101	0.976349	CCDC80	2.22899
DUSP16	1.105	CD209	-1.78012	CCL2	1.45418

DUSP5	2.32425	CD209E	2.26136	CCNA2	1.47258
EBNA1BP2	1.42938	CD209E	-3.07293	CCNB2	1.25759
EBP	1.32241	CD22	-1.89277	CCND1	1.00703
EFHD2	0.985837	CD22	-3.27566	CCPG1	-1.18996
EHD1	2.3776	CD22	-1.96429	CCR2	-1.3683
EHD1	0.983211	CD22	-2.29895	CCR2	-3.08196
EIF1AX	1.07485	CD276	1.66042	CD209	-1.24723
EIF3J	0.903696	CD276	1.58725	CD209E	2.50143
EIF4E	1.52421	CD38	2.52709	CD209E	-3.48673
EIF4G2	1.5228	CD48	1.30933	CD22	-2.26821
EMC9	1.20086	CD79B	-1.19847	CD22	-2.06409
ENDOU	2.32943	CD81	-1.39545	CD22	-1.90815
EPD1	2.31237	CD9	1.30232	CD2AP	-1.49791
ERAP1	1.16964	CD97/ COLGALT 1	1.0987	CD48	1.14607
EREG	3.65017	CDC16	-1.02721	CD79B	-1.81626
EXOC3L2	1.71269	CDH1	1.58561	CD9	1.02001
EXOSC5	1.00234	CDH2	1.55921	CDC20	1.03959
EXOSC9	1.0425	CDK2AP1	1.15452	CDH1	-1.81834
FAAP24	1.1793	CDO1	2.05491	CEACAM2	-1.38961
FAM129B	1.24679	CDV3	0.925927	CEBPE	-2.83075
FAM136A	1.07479	CEBPB	0.932655	CECR1A	-1.23449
FAM49A	1.43255	CEBPE	-2.60075	CELA2A	-3.20173
FCF1	0.971091	CEP131	1.41306	CEP131	1.16534
FGB	1.5544	CES2/ KIAA0513	-1.1546	CFI	-1.52019
FGG	3.40587	CFP	2.26685	CFP	1.65372
FGL2	2.66704	CH25H	3.32368	CHAC1	1.34968

FITM2	1.38445	CHST3	2.23074	CHT1	-1.06378
FKBP11	1.31037	CIART	1.56815	CIPC	-1.97477
FKBP4	1.0239	CILP2	2.70679	CKM	-2.06444
FLVCR2	1.10491	CLCC1	1.11913	CLCN7	1.42389
FNDC4	2.11974	CLCN7	1.10185	CLDN1	-1.6282
FRMD4B	1.85712	CLDN1	4.4537	CLDN4	-1.23756
FZD1	2.25169	CLDN5	1.33538	CLEC10A	-1.69442
GAS7	1.75121	CLEC10A	-2.31041	CLEC17A	-1.80744
GATC	1.86959	CLEC3A	3.79385	CLEC3A	3.26495
GCSH	1.10767	CLEC3B	-2.19979	CLEC4F	-1.04514
GGA3	1.35846	CLMN	-1.68829	CLEC4M	-3.32083
GGH	7.0852	CLU	-1.0876	CLU	-1.66703
GINS4	1.18668	CMLKR1	2.4998	CMLKR1	1.36778
GJA1	1.94422	CMLKR1	4.40032	CMLKR1	3.24579
GJA3	2.07687	CMTM7	-1.32288	CMTM7	-1.62035
GLOD4	1.21226	CNDP2	1.16494	CNDP2	1.68412
GNG12	2.79089	CNFN	1.70377	CNDP2	1.76476
GNPDA1	0.975793	CNN2	1.13862	CNFN	1.38765
GNPNAT1	1.42142	CNR2	-2.2387	CNN2	1.0255
GRIK1	1.81044	COL10A1	-1.64566	CNPY1	1.57713
GRN	1.95564	COL12A1	2.40889	CNR2	-3.23112
GRPEL1	1.01963	COL16A1	-2.79047	COL17A1	-2.89906
GRWD1	1.28596	COL17A1	-2.75501	COL19A1	-2.39333
GYG1	1.17024	COL1A1	-1.90135	COMMD2	1.01612
GZMB	2.21645	COL1A2	-1.95626	COX17	0.997525
HAMP1	4.13502	COL5A1	-1.09039	COX5A	0.996239
HARS	1.0826	COL6A3	-1.24311	COX6B1	-1.86907
HBEGF	1.2683	COL8A2	-1.5694	COX7B	1.03555

HCAR2	1.58036	COLGALT1	-1.18517	COX8A	1.00647
HCAR3	1.14535	COX4I2	-1.57358	CPZ	-1.59871
HEPHL1	1.24374	CRABP2	2.87085	CRABP2	3.37068
HERC3	1.31662	CRABP2	-3.24573	CREG2	-1.10196
HKDC1	0.932155	CRBN	1.10868	CRELD2	1.37483
HLA-A	1.12686	CREBBP	1.97201	CREM	1.50395
HMGCS1	2.29596	CREG2	-0.979178	CRYSTALLIN J1C	-1.16699
HMOX	1.18387	CRELD2	1.24303	CSRNP2	-1.47199
HSC71	1.38933	CREM	3.2048	CSRP1	2.15804
HSP90AB3P	1.83859	CRIM1	1.7237	CTSK	-1.07913
HSP90B1	1.39025	CRP	-2.00794	CTSZ	1.1088
HSP90B1	2.28935	CRY1	-1.59223	CXCL8	4.85244
HSPA4	1.62092	CRYSTALLIN J1C	-1.19957	CXCL8	6.31639
HSPA5	2.24191	CSGALNACT1	-1.1075	CXCL8	3.00865
HSPA9	0.91758	CSRNP1	1.68343	CXCR1	3.14337
HSPE1	1.46284	CSRP1	2.01537	CXCR2	-2.59154
HTRA1	1.29016	CST3	-1.33746	CXCR4	-1.61203
IDH3A	1.18959	CTGF	1.77964	CYB5R2	1.38435
IER5L	1.34785	CTNNA1	1.12783	CYC	1.55146
IFI27L2	2.84604	CTSH	-1.00947	CYP21A2	1.24523
IFI44	1.5611	CTSS	2.00458	CYP2A13	-1.45675
IFI44	1.59159	CTSZ	1.46532	CYP2D15	-2.83001
IFI44L	1.68075	CUL9	1.22957	CYP4F22	-2.43969
IFITM10	1.14573	CXCL1	2.10998	CYR61	-1.89012

IGLC6	2.01856	CXCL8	4.95609	DAPK3	-1.91802
IGSF6	2.22381	CXCL8	6.36972	DBN1	1.58859
IL10RB	1.26089	CXCL8	2.07766	DCSTAMP	-1.64266
IL12B	2.43777	CXCR1	2.39938	DDC	-1.35801
IL12RB2	0.891295	CXCR4	-1.44589	DDIT4L	2.4588
IL1B	8.34125	CYLD	1.10115	DEF6	-0.932726
IL20RB	2.70442	CYP26A1	-3.9891	DFNA5	-1.77741
IL22RA2	2.78435	CYP2A13	-2.47028	DGAT1	-1.96371
IL4I1	Inf	CYP4F22/ CYP4F3	-2.11163	DGKB	-1.23608
IL4I1	4.06771	CYR61	1.56899	DGKQ	-2.25541
IMPA1	1.63687	DAPK2	-1.1614	DHRS13	1.20282
IPCEF1	0.986887	DBN1	1.89861	DLL4	-1.86707
IRF2	2.81842	DCK	-0.928682	DMPK	-2.47731
IRF4	1.34757	DDIT4L	2.53021	DNAJB11	1.22
IRGC	3.54819	DDR1	2.19743	DNASE1L3	-0.963737
IRGC	6.07169	DEF6/ PPARD	-1.00325	DOCK9	-1.57999
IRGC	4.97881	DENND4A	2.40655	DOK2	-1.35599
IRGC	2.06174	DFNA5	-0.958085	DPP3	1.7432
IRGC	2.51192	DGAT1	-1.19499	DPP3	2.28869
IRGC	2.04764	DGAT1	-2.23889	DPP9	1.13081
ITGB1BP1	1.22631	DGKB	-1.30222	DRAM1	2.14382
JUNB	2.12059	DGKQ	-2.99653	DSCAM	-1.61098
KARS	1.76712	DHX58	1.71959	DSP	0.936983
KIF27	1.32086	DIP2A	1.16887	DTNBP1	-1.82055
KLF5	2.31509	DIRC2	2.04578	DTX3	-1.30345
KLHL40	1.43182	DLC	1.39081	DUSP16	1.11454
KNOP1	1.00356	DMBT1	1.48018	DUSP2	-1.70338

LDHA	2.2242	DNASE1L3	-1.18002	DUSP22B	-2.07371
LECT2	8.47876	DOCK9	2.22718	EBF1	-1.5282
LIN52	1.02476	DOCK9	-1.10595	EBNA1BP2	1.31854
LIPH	2.08018	DOK2	-1.30668	EEF1D	1.01766
LMAN1	0.906555	DOPEY2	2.93485	EEF1E1	1.30022
LMOD1	0.952924	DOPEY2	3.00299	EEPD1	-1.68631
LRAT	1.63462	DPP3	1.55668	EHD1	1.83182
LRPAP1	0.875178	DPP3	3.76075	EIF1AX	1.08494
LRRC15	2.43908	DPYSL3	-1.66776	EIF4E	1.28241
LYAR	1.00794	DRAM1	2.42705	EIF4E3	-1.44676
LYG2	3.03606	DSCAM	-2.94301	EIF4G1	1.03003
LYG2	3.29112	DTNBP1	-1.66256	ENDOUC	-1.43397
MANF	2.26101	DTX3	-1.15013	EPD1	2.10043
MAPKAPK 2	0.977533	DUSP1	2.27949	EPOR	-1.07852
MARCH5	1.20587	DUSP16	1.86197	EPS15L1	-1.45315
MARCKSL 1	3.32345	DUSP5	2.38493	EPX	-2.98484
MBD2	1.41493	DUSP6	0.962863	EPX	-1.29063
MECR	2.10407	E2F2	-1.30482	ERAP1	1.24096
MED24	2.25639	ECE1	1.10521	EREG	2.21353
MERTK	2.87335	EEPD1	-1.76	ERGIC2	-2.27755
MESDC2	0.899414	EGR1	1.83864	ETS1	-0.932715
MGEA5	1.10231	EGR1	1.61596	F5	-2.49038
MIMI R795	1.66049	EHD1	2.0975	FAAH	0.953541
MINA	0.963609	EHD1	1.91141	FABP2	-3.24141
MLXIPL	1.96359	EIF4E3	-1.16699	FAM129B	1.15542
MMCM6	1.85479	EIF4G1	1.33404	FAM136A	1.0435
MMP13	4.07869	EIF4G2	1.31006	FAM13A	-1.48881

MMP14	1.26611	ELF3	1.18201	FAM214B	-1.45446
MMP9	2.82889	ELMO1	-1.23488	FAM65C	-1.15604
MOB3C	1.08029	ELMO3	1.1344	FBP1	-3.2362
MPHOSPH 10	1.19349	ELMOD3	1.56881	FBXO32	-1.93138
MPHOSPH 6	1.10034	ENDOU	3.99933	FECH	1.11055
MPV17L2	1.70417	ENDOUC	-1.8017	FGA	-4.32233
MPZL2	1.06264	EPB41	-1.17043	FGB	-2.68863
MR1	1.34951	EPD1	2.17227	FGD5	-1.40536
MRC1	3.19281	EPHX2	-3.05308	FGG	-9.02601
MRPL12	1.06398	EPOR	-0.953677	FGL2	0.986276
MRPS17	0.966217	EPX	-3.42989	FKBP11	1.20399
MSMB	2.27486	EPX	-1.14867	FKBP4	1.2257
MST1R	1.88217	ERAP1	1.12787	FOXP4	-2.38741
MYC	1.33002	EREG	3.38216	FRMD4B	1.96115
MYH4	3.01961	ERGIC1	1.08559	FRRS1	7.29721
MYO1C	8.60528	ETF1	0.887772	FUT9	-3.64618
MYO1D	1.11902	ETNPPL	1.45958	FXYD6	-1.41663
MYOF	1.33307	ETS2	1.41961	FZD1	1.73072
NAMPT	2.2554	ETV5	1.2828	G3BP1	1.00615
NAMPT	1.92775	EXOC3L2	1.47106	G6PD	-1.29115
NAT1	1.14658	F10	-1.46855	GADL1	-1.0568
NCAM1	1.74978	F5	-2.12409	GARS	1.38922
NCL	1.0399	FABP2	-4.09852	GAS1	-1.13639
NEOVERR UCOTOXI N SUBUNIT	3.41483	FABP3	-1.0577	GATC	0.977826

ALPHA					
NEU3	1.13589	FABP3	-3.05448	GBGT1	-1.46366
NFAT5	1.9452	FABP6	1.71706	GBP	-1.04737
NFKBIA	1.7156	FAM129B	1.28702	GCSH	1.27528
NFKBIA	2.10239	FAM173A	-1.25181	GFI1	-2.35172
NFKBIE	1.40948	FAM212A	-2.74671	GGH	6.27686
NHP2	1.26853	FAM49A	1.29619	GGT5	-1.34785
NIFK	0.947151	FAM65C	-1.07911	GLDN	2.37489
NLRC3	1.19813	FBP1	-3.70137	GLIPR2	1.23929
NOLC1	1.02863	FBXO21	1.31201	GLUL	-1.31047
NOP10	1.08442	FEZ1	1.43867	GNG12	2.23776
NPM1	0.971397	FGB	2.77503	GNG8	-2.57857
NR13	1.50107	FGD5	-1.53675	GNMT	1.58554
NRBP1	1.94547	FGL2	2.92667	GNPDA1	1.28598
NRN1	2.24579	FILIP1L	2.5503	GPD1	-1.58197
NRP1	1.28041	FKBP5	2.05191	GPR149	-2.14542
NUBP2	1.20192	FLVCR2	1.55414	GRAMD1C	1.58699
NUDC	1.46749	FMO5	-1.16567	GREM1	-3.92425
NUS1	1.03446	FNDC3B	1.49037	GRID2IP	-3.84907
OGFR	1.20468	FOXO3	-1.02438	GRIK1	1.59641
OGFRL1	1.71785	FREM1	-1.94761	GRN	3.00292
OLFM4	1.13479	FRMD4B	2.77891	GRPEL1	1.00129
OLFM4	1.36503	FTH1	-1.58415	GSN	-1.25005
PA2G4	0.931502	FZD1	2.71179	GSTK1	-0.973533
PARP14	1.42269	FZD4	-1.18814	GZMA	-2.38402
PDE4B	1.62038	G6PD	-1.24228	GZMB	3.58924
PDIA3	1.2597	GADL1	-1.56444	H1F0	-1.8355
PDIA6	1.5664	GANAB	1.27403	H2AFX	-0.952576

PEA15	1.31541	GARS	1.04135	HABP2	-Inf
PFKFB3	2.42497	GAS7	1.12876	HAMP1	2.17445
PHACTR1	1.51093	GATC	2.31544	HARS	1.24861
PHC2	1.5607	GBGT1	-2.83569	HCAR2	-1.27059
PHEX	2.20695	GCH1	3.11576	HCAR3	1.17225
PI4K2A	1.56797	GCNT4/ ANKRD31	-1.39786	HCEA	-3.31574
PID1	1.73054	GDI1	1.16074	HEBP2	-4.22525
PIP5K1C	1.19659	GEM	2.27709	HEPHL1	1.4509
PIR	1.46641	GFI1	-1.90777	HES7	-1.69045
PKP3	1.77779	GGH	6.53864	HHLA2	-1.374
PLA2G16	1.92542	GGT5	-2.20774	HHLA2	-1.21033
PLAU	1.33031	GJA1	2.25944	HHLA2	-1.17359
PLCD4	1.37725	GLDN	2.59343	HIST1H1E	-1.72295
PMP22	1.36703	GLIPR2	-3.46517	HLF	-1.62127
PNP	1.6532	GLS	1.71441	HMGCS1	1.9215
POLR1C	1.00815	GNB4	1.08916	HMOX	1.47161
POLR2H	1.19944	GNG12	2.0143	HNRNPH1	1.04636
POLR2I	1.12163	GNG8	-2.15042	HPD	-1.76503
POMP	1.21725	GPD1	-1.34921	HPDL	1.41715
PPA1	1.16905	GPD1	-2.05407	HPGD	-1.18781
PPAN	0.912769	GPI	-1.0181	HPX	-Inf
PPIA	0.951838	GPR182	-1.12419	HSP90AB3 P	1.91957
PPIB	1.30741	GPR4	1.62922	HSP90B1	1.5908
PPP5C	0.970648	GRAMD1C	1.50597	HSPA4	1.55917
PRDX1	1.00564	GRID2IP	-3.34766	HSPA5	1.60248
PRMT1	1.07725	GRIK1	2.40318	HSPA9	0.941054
PRPF31	0.915663	GRK5	1.08648	HSPE1	1.8597

PSMA1	1.38096	GSTK1	-1.51092	IAAA	1.98787
PSMA2	1.42977	GUCA1A	-2.45957	IER5L	1.15442
PSMA4	1.35225	GZMA	-2.43722	IFI27L2A	2.92923
PSMA5	1.44575	GZMB	3.98524	IFI44	2.71278
PSMA6	1.35893	H2AFX	-1.47944	IG HEAVY CHAIN V-III REGION NIE	-1.29824
PSMA6	0.907927	HACD4	-1.58982	IG KAPPA CHAIN V-IV REGION B17	-2.15575
PSMA7	1.29571	HADHA	1.02274	IGF1R	-1.33268
PSMB3	1.12501	HAMP1	3.07343	IGF2	-1.62512
PSMB7	1.13092	HAS2	2.97131	IGFALS	-Inf
PSMC1	0.904215	HAVCR1	1.99077	IGFBP3	-1.50476
PSMC3	0.916548	HCAR2	1.37568	IGFBP5	-0.990875
PSMC4	1.0147	HCEA	-3.2383	IGHV1-61	-1.35299
PSMC6	0.874256	HEBP2	-3.86454	IGKC	-1.94898
PSMD12	1.14172	HEPHL1	1.78258	IGKC	-1.13047
PSMD13	1.01873	HERC3	1.8287	IGKC	-1.1566
PSMD14	1.18036	HGFAC	2.89932	IGKC	-1.58392
PSMD6	1.17781	HIVEP1	1.3091	IGKC	-1.55971
PSMD8	1.0592	HK2	0.973576	IGKC	-1.39672
PSME1	1.13287	HKDC1	1.15466	IGKV3 REGION VH	-1.61986
PSME2	1.24973	HLF	-1.88665	IGKV3 REGION VH	-1.15987

PTAFR	2.50445	HLX	-0.972863	IGLC6	-1.24378
PTGES	1.70969	HMGB1	-1.26415	IGLC6	-2.29893
PTGR1	0.952508	HMGCS1	1.8201	IGLC6	-3.47882
PUMILIO DOMAIN-CONTAINING PROTEIN KIAA0020 HOMOLOG	0.864285	HMOX	1.77692	IGLC6	-1.50808
PVALB	1.84733	HNRNPH1	1.16012	IL17RA	-1.3465
PWP2	1.40492	HOMER3	1.43813	IL1B	5.54297
QPCT	4.21091	HPGD	-1.79663	IL1RL1	-1.49668
RAB29	1.81565	HSC71	1.13787	IL20RB	2.19586
RAB39B	1.91271	HSD17B7	-1.80631	IL34	-2.91508
RAB8B	0.937452	HSP70	2.59073	IL4I1	3.67418
RABL6	1.57215	HSP90AB3P	1.27251	IL6R	-1.64532
RAN	0.983912	HSP90B1	2.25617	IMPA1	-1.18405
RAP1GAP	1.1726	HSPA1A	-3.91244	IRF2	1.93498
RASGEF1BA	3.01541	HSPA4	1.49677	IRF4	1.03053
RBCK1	1.55626	HSPA5	2.05904	IRGC	2.53727
RBM44	2.4419	HSPA9	1.19838	IRGC	1.7314
RDH10	2.5805	HTRA1A	1.18585	IRGC	1.81079
RDX	1.15448	IAAA	2.11122	IRGC	1.62152
REXO2	1.09847	ICT1	-1.60872	ISYNA1-A	-1.49087
RGCC	1.56053	ID1	-1.11896	ITGA6	-1.69295
RGS1	3.14662	ID4	-1.38007	ITGAE	-1.51198

RGS1	2.03421	IER5L	1.2054	ITGAM	-1.11369
RHO	1.98362	IFI27L2A	3.6931	ITGAX	-1.08727
RHOB	1.63966	IFI44	1.35406	ITGB6	-1.05968
RIPK2	1.29737	IFI44L	1.51954	ITM2B	-1.08076
RNF114	1.30638	IFITM10	1.78711	JUNB	1.38145
RNF186	1.73353	IG KAPPA CHAIN V-VI REGION NQ2-6.1	-1.91492	KARS	2.18934
RNF213	1.61968	IGFBP1	3.65359	KBP	-3.84027
RNF4	1.09753	IGFBP3	-1.59333	KCNG1	-1.57853
RPL9	1.309	IGKV4-1	-2.51807	KCNH1	1.0232
RRBP1	1.6587	IL10RB	1.02269	KCNK1	-1.19138
RRP15	1.06452	IL12B	2.6522	KCNT1	-3.6512
RRP9	1.11032	IL18R1	2.29768	KCTD7	-1.45755
RSL1D1	1.03832	IL1B	6.95879	KDR	-1.53411
SAPCD2	3.88038	IL20RB	3.74933	KIAA1161	-2.08688
SDC4	1.03091	IL22RA2	3.6213	KIF27	1.12232
SEC61A1	1.27024	IL34	-2.55534	KIRREL	-2.29924
SEC61B	1.5531	IL4I1	4.41663	KLF11	-1.69845
SENP8	1.15627	IL6R	-1.73254	KLF13	-1.30854
SFRP2	2.23702	IMPA1	2.57209	KLF2	-1.16546
SGK1	1.54107	IMPDH1B	1.25772	KLF5	2.56027
SGK1	1.40698	IRF2	2.44713	KLF6	-1.01035
SIGLEC1	1.21984	IRF4	1.34273	KLHL24	-1.94088
SIGLEC1	2.89255	IRGC	1.41884	KPNA2	1.32595
SIN3A	1.03971	IRGC	2.2375	KRT13	-3.44668
SKAP2	1.19472	IRGC	2.5498	KRT13	-3.97555
SLC12A9	1.70674	IRGC	2.63688	KRT13	-3.5952

SLC22A2	2.05509	IST1	0.945941	KRT18	-1.6566
SLC25A28	1.99241	ITGA5	1.29291	L1CAM	-1.30904
SLC25A32	1.17924	ITGA6	-1.2374	LDHA	1.91857
SLC27A4	0.946131	ITGB1	1.04449	LDHBA	-1.8617
SLC29A2	1.01869	ITGB3	1.63441	LECT2	5.29398
SLC2A6	3.73075	ITIH3	-1.24611	LFNG	-1.39198
SLC35B2	1.30978	IVNS1ABPA	1.64692	LGALS1	1.71904
SLC38A3	1.21564	JUNB	2.39481	LGALS1	-1.40467
SLC39A6	2.9078	JUND	1.69645	LIPH	1.56537
SLC43A2	1.07474	KALRN	1.24265	LMBRD1	-1.08308
SLC51A	1.82543	KARS	2.73735	LONRF1	-1.70918
SLC5A7	1.65394	KCNK1	-1.18513	LPAR6	-1.37112
SLC5A8	1.17948	KCNT1	-2.10542	LPL	-1.12749
SMYD5	0.968128	KCTD7	-2.11069	LPXN	-1.48394
SNRPA1	0.998221	KDR	-1.35994	LRMP	-1.50233
SNRPD3	0.907949	KHK	-1.46202	LRRC15	1.31804
SNX10B	2.09809	KIAA1324L	-1.81779	LRRC17	-2.15769
SOCS1	3.95469	KIF27/GKAP1	0.930555	LTA4H	-2.62496
SOCS3	2.65769	KLF11	-1.28301	LYG2	3.88411
SOCS3	2.09103	KLF5	3.26797	LYG2	3.38492
SPAG1	1.46914	KLF9	1.72331	MAG	-1.73873
SPIRE1	2.77481	KLHL26	1.82529	MAMDC2	-1.24169
SPTLC2	2.20329	KRT13	-6.49792	MANF	1.22235
SSR2	1.19105	KRT13	-6.04266	MAPK1	0.89351
SSR3	0.981698	KRT13	-5.68714	MAPK11	-1.16034
STAR	1.23597	KRT13	-4.98534	MAPK13	-0.975688
STAT1	1.8868	L1CAM	-1.00891	MAPKAPK	0.897896

				2	
STAT3	0.989627	L-ASPARAGINASE	-2.11408	MARC1	-1.20387
STEAP2	2.9033	L-ASPARAGINASE	-3.39791	MARC1	-1.14502
STEAP4	1.73003	LCP1	1.05973	MARCKSL1	1.35881
STEAP4	2.32432	LDHA	2.03535	MAST CELL PROTEASE 3	-1.78917
STEAP4	2.82917	LDHBA	-1.30047	MCTP1	-1.67972
STEAP4	2.45904	LDLR-A	4.38495	MED9	1.0733
STK38L	1.40162	LECT2	6.79705	MERTK	2.4267
STX5	1.22559	LGALS1	-2.34292	MFSD1	1.23213
SULT2B1	1.45787	LGALS1	-1.83557	MFSD8	2.15945
SWT1	0.972561	LIFR	1.51067	MID1IP1B	-1.30648
TAP1	1.54563	LIPE	1.47825	MID1IP1L	-1.76722
TAPBPL	1.604	LIPH	3.28846	MIEN1	-1.97433
TBC1D2	2.46561	LMO4-B	-3.24113	MINOS1	1.16424
TCAF	1.57168	LMOD1	0.973048	MIOX	-1.37409
TDH	1.03729	LOXL2A	1.92999	MKNK2	-1.24372
TF	5.42779	LPAR4	-1.28514	MLEC	0.931604
TGFB1I1	1.07916	LPIN3	1.54345	MLXIPL	1.26496
TGL2	1.5259	LPL	-1.38676	MMCM6	1.64623
TGM1	1.3908	LPXN	-1.22734	MMP13	2.8931
TIMM13-A	1.28584	LRMP	-1.06839	MMP2	-1.15584
TIMM44	1.34005	LRRC15	3.03352	MMP25	-2.08623

TIMP2	1.94959	LRRC17	-2.05476	MMP9	1.64771
TM4SF1	1.13628	LTB4R	-2.279	MPHOSPH6	1.00868
TMA7	0.986889	LTBP1	1.35035	MPPE1	-1.54166
TMCO1	0.97326	LYG2	4.76707	MPV17L2	1.30313
TMED2	1.29791	LYG2	1.28001	MRC1	1.7547
TMEM147	1.02294	LYG2	3.21863	MRC1	-1.59971
TMEM208	1.25714	MAMDC2	-2.97986	MRC1	-1.8153
TNFAIP3	2.12941	MANF	1.16013	MRPL53	1.44363
TNFRSF11B	3.62057	MANSC1	1.25491	MRPS16	0.946955
TNFRSF4	2.68863	MAP1B	1.32218	MRPS17	1.28742
TNIP1	0.982296	MAP3K15	-2.45331	MRPS33	0.98133
TOMM22	1.23913	MAP3K8	1.58891	MSMB	1.7175
TOP3B	3.23198	MAPK1	1.01482	MSN	-1.8463
TPH2	4.26761	MAPKAPK2	1.20388	MTSS1L	-2.22957
TPH2	4.08705	MAPRE1	0.944348	MXRA8	-1.12405
TRHDE	2.13947	MARCKSL1	3.18065	MYDGF	1.21498
TRIB2	1.67947	MBD2	1.59027	MYG1	1.07415
TRIM62	1.00719	MBP	-1.45092	MYH4	-4.77998
TRMT11	1.19345	MCF2L2	1.4317	MYH4	-4.58812
TRMT112	0.985911	MCTP1	-1.1208	MYOF	1.46169
TSR2	1.49604	MDK-B	-1.52199	NAMPT	1.62758
TUBB	1.08071	MECR	1.38201	NAT1	1.06359
TUFT1	1.45388	MEGF6	-3.55481	NCAM2	-1.69704
TUSC5	2.10456	MERTK	3.2523	NCF1	-1.24906
TXNL1	0.870311	METTL7A	-0.924637	NDUFA3	0.944974

TYMP	2.2825	MFSD1	1.31331	NEK2	1.0899
TYMP	1.89652	MFSD8	2.49863	NETO2	-2.16347
TYPE-4 ICE-STRUCTURING PROTEIN AFP4	1.72608	MGAT4B	-2.19032	NEU3	1.06934
UBA5	1.47042	MGP	7.45343	NEU3	-1.1937
UNCHARACTERIZED PROTEIN C10ORF88 HOMOLOG	5.15735	MGST1	-1.14454	NFAT5	1.27386
UNCHARACTERIZED PROTEIN C10ORF88 HOMOLOG	2.97764	MICALL2	1.5298	NFKBIA	1.31657
UPP1	0.9589	MICALL2	-1.59344	NFKBIA	0.914506
USP2	1.42052	MID1IP1B	-1.64557	NGDN	1.17781
UTP11L	1.02278	MID1IP1L	-2.17742	NLRC3	-1.29937
UTP15	1.01232	MIEN1	-2.63977	NLRC3	-1.80346
UTP3	1.15071	MIER2	1.36714	NLRC3	-1.36141
VCAM1	2.98604	MINPP1	-1.08862	NLRC3	-2.37403
VCAM1	2.07974	MMP13	3.61596	NLRC3	-2.67358

VCAN	1.2567	MMP14	1.52475	NLRC3	-2.28652
VDAC1	1.11743	MMP14	1.43102	NLRP12	-1.64436
VDAC2	2.36956	MMP2	-1.25131	NLRP12	-1.18164
VERRUCOTOXIN SUBUNIT BETA	2.1485	MMP25	-1.91862	NLRX1	-1.45841
VPS11	1.41258	MMP9	2.55254	NODAL	-2.54956
VRK1	1.31287	MOB3C	0.992699	NOLC1	1.34078
VWA5A	1.42126	MPV17L2	1.38429	NPM1	0.98472
WDFY2	1.33989	MPZL2	1.02556	NPTN	-1.23121
WDR77	1.08073	MR1	1.19496	NPTX1	-2.68749
WDR83OS	1.20921	MRC1	-2.03017	NR1D2	-2.1257
XMRK	2.13725	MRC1	-1.42061	NRBP1	1.45706
YES1	2.11721	MSMB	3.57131	NRIP2	-1.16966
YRDC	1.03536	MSN	-2.36954	NRN1	3.52404
ZFP36L3	1.00131	MST1R	2.53028	NT5C2	-1.51742
AADAC	-1.40591	MT	1.29697	NTRK2	-3.0352
ABTB1	-1.34959	MUSTN1	-2.89733	NUBP2	1.66535
ACSS1	-1.45147	MUTYH	-1.16803	NUDC	1.21154
ADCK3	-1.28694	MYADM	1.91397	OAZ1	-1.38015
ADCY6	-0.994807	MYC	0.9979	OGFRL1	1.55052
AMIGO3	-2.5487	MYO1C	Inf	OLFML2A	-1.39904
APOE	-1.06538	MYO1D	1.79093	OXSR1	1.88655
ARHGAP35	-1.83158	MYO9B	1.95902	P2RY1	1.80973
ARMC3	-1.02964	MYOF	1.42621	P2RY1	-1.08489
ART1	-1.44156	NAMPT	2.39691	PAIP2B	-1.52107
ATG2A	-1.66806	NAMPT	1.85653	PAPLN	-2.40906

ATP8B2	-1.32575	NCAM1	2.08274	PAQR5A	-1.02542
ATPIF1	-0.902218	NCF1	-0.970639	PAX5	-1.47261
BNIPL	-1.13536	NETO2	-1.95274	PCBD1	1.18678
C2CD4CC2CD4 FAMILY	-1.5757	NEU3	-1.83988	PCMTD1	-1.22774
CABP1	-1.88489	NEURL3	3.22098	PCMTD1	-1.17403
CABP4	-1.23586	NFAT5	2.66632	PDE4DIP	-1.5014
CALCOCO1	-1.20362	NFKBIA	2.21234	PDIA3	1.09747
CALHM3	-1.70731	NFKBIA	2.10217	PDIA6	1.19171
CAMK2G	-1.6376	NFKBIE	1.52264	PDLIM3	-1.59163
CD209	-1.74046	NLRC3	1.05034	PFDN1	1.13477
CES2	-0.910585	NLRC3	-1.16091	PFKFB1	-2.06685
CLCN2	-1.49861	NLRC3	-1.2761	PFKFB3	1.67633
CMTM7	-0.997117	NLRC3	-3.02474	PGK1	0.895771
COLGALT1	-1.12691	NLRC3	-1.28492	PHACTR1	1.55621
DAAM1	-1.53316	NLRC3	-1.46883	PHC2	2.27923
DAB2	-1.06093	NLRP12	1.70344	PHF5A	1.30372
DGAT1	-1.86944	NLRP12	-1.33375	PHPT1	1.31388
DGKB	-1.29394	NMRK2	-1.18655	PI4K2A	1.25273
DGKQ	-2.24146	NMT1	1.10511	PID1	1.34921
DOK2	-1.2091	NODAL	-4.24713	PIK3IP1	-1.13869
DSCAM	-2.19958	NOLC1	1.87448	PIR	1.69349
EIF4E3	-1.11283	NPHS1	-1.57756	PKNOX2	-1.12727
EML4	-0.9719	NPM1	1.32509	PLA2G16	1.96322
EXOC3L1	-1.74929	NPTN	-1.09832	PLA2G16	3.01566
F5	-1.39677	NPTX1	-1.63237	PLA2G4C	-4.99735

FABP3	-1.78681	NR13	2.24875	PLCD4	1.14048
FAM214B	-1.13596	NR1H4	-1.67889	PLEC	-1.45353
FBN2	-1.92373	NRBP1	2.23803	PLLP	-0.949272
FBP1	-1.62512	NRN1	1.93616	PLXNB2	1.4282
FBXO32	-1.08416	NRP1A	2.46215	PMP22	1.3206
FZD4	-1.25001	NT5C2	-1.49597	PNP	2.39145
GADL1	-1.06846	NT5C2	-1.06225	POLR1D	1.24631
GGT5	-1.92938	NUAK1	2.3235	POLR2H	0.977498
GLIPR2	-1.87466	NUBP2	1.10741	POLR2I	1.06565
GM2A	-0.975785	NUDT4	-1.51931	POMP	1.43162
GRB10	-1.22317	OGFRL1	2.75871	POSTN	-1.67156
GREM1	-2.2377	OLFML2A	-1.83586	PPA1	1.31902
GRID2IP	-2.71729	P2RY1	1.92801	PPDPFA	-1.59411
HACD4	-1.03569	P3H2	1.92261	PPIA	1.10866
HBP1	-1.22524	PARP12	1.31225	PPIB	1.22666
HCEA	-1.41929	PARVB	-1.61253	PPP1R3A	-1.75467
HPGD	-0.99599	PCOLCE	1.73701	PPP1R3CB	-1.16311
HSPA1A	-3.18858	PDE4B	2.16903	PPP1R3D	-2.53273
ID4	-1.04775	PDE6D	-1.59597	PPP2R2B	1.50111
IL6R	-0.953091	PDGFRA	1.24113	PRDX1	1.08148
KCNG1	-1.13021	PDGFRB/ CSF1R2	1.53636	PREP	1.17666
KCNT1	-1.66915	PDIA3	1.3071	PRMT5	1.26877
KCTD7	-1.49442	PDIA6	1.03577	PROSC	1.15059
KDR	-1.35829	PDLIM3	-2.07025	PRR5	-1.16827
KLF11	-1.88708	PEA15	1.67646	PSAP	-1.31948
KLF13	-1.16519	PELI2	1.83708	PSBP1	-7.58379
KLHL22	-1.07942	PFKFB1	-2.12095	PSMA1	1.26768
KLHL24	-1.58693	PFKFB2	-2.02628	PSMA2	1.60969

KLHL28	-1.14919	PFKFB3	2.94171	PSMA4	1.50064
KMT2A	-1.01701	PGAP2	0.985648	PSMA5	1.48736
KRT13	-3.18243	PGM1	-1.40517	PSMA6	1.47734
KRT13	-2.84886	PHACTR1	2.69662	PSMA6	1.02075
KRT13	-2.66962	PHC2	2.60604	PSMB10	1.40575
KRT18	-1.18828	PHEX	2.54531	PSMB2	1.65149
LGALS4	-1.17624	PHYHD1	-1.23175	PSMB3	1.17852
LMTK2	-0.922135	PID1	1.71415	PSMB4	1.41643
LONRF1	-1.57868	PIK3CB	1.12307	PSMB5	0.946396
LPL	-0.874423	PIK3IP1	-0.951577	PSMC4	1.03045
LRRK1	-1.58959	PIM1	1.06115	PSMC6	0.957961
MAMDC2	-2.02114	PIR	2.4593	PSMD14	1.15631
MAP3K15	-2.86945	PITPNB	-1.75496	PSMD8	1.03968
MAPK11	-0.874734	PKP3	1.77876	PSME1	1.45995
MEGF6	-1.42579	PLA2G16	2.73341	PSME2	1.38602
MGAT4B	-2.05698	PLA2G16	2.44266	PSPH	1.55899
MKL1	-1.42122	PLA2G16	3.03412	PTAFR	1.50651
MSN	-1.10121	PLA2R1	1.59817	PTGER4	-1.72535
MTSS1L	-1.42371	PLAU	1.932	PTGR1	1.04886
MYO18A	-1.31535	PLCD3A	1.51866	PTN	-1.8422
NLRC3	-3.5783	PLCD4	1.75015	PTPN7	-1.0313
NLRC3	-3.00162	PLEC	-1.44565	PTRF	-1.53225
NLRC3	-2.75499	PLEKHA5	1.24384	PVALB	-2.12354
NLRC3	-1.91038	PLLP	-1.43697	PWP2	1.24253
NLRC3	-2.87453	PLVAP	-1.78239	PYCARD	1.05257
NODAL	-2.23944	PLXNA1	1.74696	PYGL	-2.34634
NPHS1	-1.49431	PLXNB2	1.31433	RAB11B	-0.937799
NPL	-1.25083	PMP22	1.16471	RAB11FIP4A	-2.07694

NPTX1	-1.43462	PMP22	-1.74428	RAB29	2.12346
NR1H4	-1.96475	PNP	1.40225	RAB32	1.79455
NUPR1	-1.82344	POLC	-1.64729	RAB39B	1.58522
PANK4	-1.11476	PPDPFB	-0.983514	RAB3D	-1.73993
PAQR5	-1.15646	PPP1R14B	1.48334	RAB40B	-1.25199
PCMTD1	-1.32705	PRDM1	2.47081	RAB7A	1.07937
PCMTD1	-1.32271	PREP	1.15565	RAG1	-3.25684
PLCL2	-1.29989	PRF1	1.33133	RAG2	-2.61308
PLVAP	-1.00137	PROM1A	-1.31512	RAMP1	-2.23586
PTPN12	-1.64178	PRR18	-1.47943	RAP1GAP	1.18989
PTPN7	-0.913739	PRR5L	-1.51622	RARS	1.54128
PYGL	-1.52213	PSMA1	0.938979	RASA3	-1.46888
RAB11FIP2	-1.32862	PSMB10	1.26445	RASGEF1BA	2.20483
RAB40B	-1.37921	PSMD2	1.12775	RASGRP2	-1.07205
RAMP1	-1.31535	PSME1	1.196	RBM38	-1.18645
RASA3	-1.24351	PTAFR	1.62776	RBP4A	-2.3107
RBP4	-1.92237	PTGES	1.97786	RDX	1.36719
RNF123	-1.64356	PTN	-1.97777	REC8	-2.95972
RPS6KA5	-1.92313	PTP4A1	1.36132	REXO2	1.0182
SCARA3	-1.50498	PTPN13	1.50136	RGS1	1.51661
SESN1	-1.99679	PTPRB	-2.02044	RGS21	-2.76889
SLC16A5	-0.961779	PTPRO	2.00881	RGS9	-1.95908
SLC40A1	-2.30396	PTRF	-1.29616	RHAG	-1.02663
SLC43A2	-1.51871	PVALB	-1.3065	RHOAC	1.25849
SLC43A3	-1.63513	PWP2	1.71847	RHOQ	1.23356
SLC44A5	-1.28912	PYGL	-2.63935	RNASEK-A	1.25107
SLC6A12	-4.70938	PYGM	-2.36972	RPA3	1.14998
SLC6A6	-0.964563	QPCT	4.91594	RPL18A	-2.00261

SLC6A6	-1.23807	QSOX2	1.2867	RRBP1	1.82116
SLC9A5	-1.52571	RAB11A/ INCENP-A	1.01103	S100A6	-1.80419
SMPD2	-1.23543	RAB11B	-1.2113	S1PR4	-1.65392
STX19	-1.74402	RAB11FIP4 A	-1.72174	SAAL1	1.04163
SYNPO	-1.18186	RAB23	1.06079	SAMD9	-2.79528
TCP11L2	-1.25953	RAB29	1.70546	SAT1	-1.4674
THBS1	-1.21039	RAB32	2.09942	SBDS	1.00768
TJP2	-1.13994	RAB39B	1.69532	SCARA3	-2.73286
TKT	-1.09568	RAB40B	-1.56447	SDS	-1.46802
TMEM131	-1.38929	RAB7A	1.23841	SEPP1A	-1.73512
TMEM230	-1.04823	RAB8B	1.2818	SEPP1B	-5.86588
TMEM86A	-2.65531	RABL6	1.60287	SEPT9	0.902032
TNK2	-2.24587	RAD23B	1.206	SEPT9	-1.09737
TP53INP1	-1.20326	RAG2	-2.89088	SERPINA1	-4.51642
TP53INP1	-0.936195	RAI14	1.3131	SESN1	-1.48427
TRPV1	-2.39455	RAPH1	-1.80497	SFRP2	1.0947
TWF1	-2.09271	RARS/ WWC1	1.34695	SFXN5	-0.97651
UBE2H	-0.899732	RASA3	-1.20475	SGK1	2.50192
ULK2	-1.70166	RASGEF1B A	2.49565	SGK3	0.942628
UNKL	-0.939057	RBMS1	1.82423	SH2D1A	-1.27374
VIM	-1.80694	RBP4A	-4.8235	SH3BGRL3	0.903625
VWA7	-1.92633	RDH10A	2.86288	SHMT2	0.967911
VWA7	-2.0362	RDH12	2.16903	SHTN1	2.67527
ZFP36L1	-1.31219	RDX	1.4748	SIGLEC1	1.46141
ZMYND8	-0.988908	REC8	-2.68387	SIGLEC14	-2.47352

RERGL	2.46136	SIGLEC5	-1.9397
RFESD	-1.05149	SKAP2	1.01369
RGCC	1.58223	SLC12A9	1.98459
RGL1	2.2047	SLC16A5	-2.2629
RGS1	1.53277	SLC16A7	-1.87153
RGS1	2.31101	SLC22A2	3.54699
RGS5	2.23965	SLC22A5	-1.46526
RHAG	-0.949559	SLC23A1	-1.83674
RHBDF1	1.25992	SLC25A16	-1.11482
RHOAC/ PPM1H	1.03464	SLC25A28	1.62043
RHPN1	2.57317	SLC27A4	1.22152
RIPK2	1.3326	SLC28A3	-1.6961
RNF114	1.345	SLC2A1	-0.94423
RNF186	2.75959	SLC2A6	2.84134
RNF213	2.14793	SLC2A9	-1.54306
RNF213	1.45314	SLC35A5	-1.0083
RRBP1	2.07296	SLC38A3	1.53104
RRM2	-2.40195	SLC40A1	-0.984834
RRP12	1.08367	SLC43A2	-2.01494
RRP1B	1.13656	SLC43A3	-4.17531
S100A6	-1.79001	SLC44A2	-1.26414
SAMD4A	2.32863	SLC44A2	-1.15617
SAPCD2	2.82493	SLC44A5B	-1.53281
SAT1	-1.41486	SLC51A	-2.03408
SBDS	1.09123	SLC5A12	-1.04219
SCARA3	-2.19861	SLC5A7	2.33443
SCCPDH	-1.54684	SLC6A13	-2.31814
SCN4B	-2.86999	SLC6A6	-1.16589

SDC4	1.60406	SLC6A6	-0.952902
SDHA	1.02198	SLC8A2	-1.38148
SDR16C5	-1.09455	SLC9A5	-2.3687
SEC31A	1.17485	SMP	-1.7174
SEC61A	1.1004	SMPDL3A	1.46707
SELH	-1.20198	SNRPB2	1.06221
SEMA3D	1.06007	SNRPD3	1.1274
SERBP1	-1.04862	SNRPF	1.09022
SERPINE1	2.42848	SNX10B	1.67648
SESN1	-1.97982	SOCS1	2.43926
SFRP2	2.42298	SOCS7	-2.47924
SGK1	1.35513	SORD	-0.995171
SGK1	1.32515	SORT1	-1.88681
SGK3	1.27401	SOX4	-1.61293
SH3BGRL2	3.28994	SOX6	-1.31968
SH3GL3	0.963002	SPA17	-1.76578
SI	1.97174	SPIDR	-1.67061
SI	2.341	SPIRE1	1.83096
SIGLEC1	1.83297	SPOCK3	-1.60858
SIGLEC1	1.32297	SPR	1.49028
SIGLEC1	2.77347	SPTLC2	1.47735
SIGLEC14	-1.03943	SRPX	-1.18358
SIGLEC5	-1.65271	SRPX2	-1.46954
SIL1	-1.24057	SRSF10	-2.07084
SIM1	1.65654	ST3GAL1	-2.27091
SIN3A	1.78826	ST3GAL1	-1.62857
SIRT5	1.04053	ST6GAL2	-1.11943
SLC12A3	-4.17981	STAT1	1.17816

SLC12A9	2.13677	STIP1	1.26639
SLC13A3	1.14693	STMN1	-1.23286
SLC16A5	-1.84964	STOML2	1.23184
SLC25A16	-1.35622	STX19	-1.45876
SLC25A28	1.79262	STX5	0.901775
SLC25A53	-1.98869	SUCNR1	-3.81
SLC27A4	1.05132	SUSD3	-1.78158
SLC27A6	1.87583	SWAP70	-1.67121
SLC28A2	-2.3002	TAP1	1.02551
SLC29A1	-1.88149	TAS1R1	-1.17062
SLC2A1	1.82395	TCAF	-1.92767
SLC2A6	3.11759	TCP11L2	-1.58993
SLC34A1	-1.70284	TGFB1I1	0.998794
SLC35A5	-1.18954	TGFB3	1.84938
SLC35B2	1.36645	TGM2	-1.20674
SLC38A2	1.78539	THAP11	-2.22061
SLC38A3	1.36037	THBD	-2.89812
SLC39A6	3.16676	TIMM13-A	1.29138
SLC39A9-A	-2.17028	TIMM44	1.24513
SLC40A1	-1.62423	TIMP3	-1.2229
SLC43A2	1.37403	TLR13	-1.86451
SLC43A2	-1.47504	TMEM183	1.37554
SLC43A3	-2.0079	TMEM229 A	-1.102
SLC43A3	-3.18915	TMEM25	-1.55014
SLC44A2	-1.40335	TMEM86A	-2.30281
SLC44A5B	-1.8613	TMPRSS2	1.97784
SLC5A7	2.36186	TMPRSS2	-1.33747
SLC5A8	1.91832	TNFAIP2	1.59116

SLC6A13	-1.14301	TNFAIP3	1.14795
SLC6A14	-3.74282	TNFRSF11B	1.54891
SLC6A8	2.23848	TNFRSF4	1.49063
SLC7A6OS	1.18363	TOMM22	1.46877
SLC8A2	-1.63304	TOMM34	1.19042
SLC9A3R2	-1.15464	TOMM40	1.09745
SLCO3A1	1.44201	TOP3B	1.2459
SMARCC2	-4.30909	TOPORS	1.10273
SMIM5	-1.46416	TP53INP1	-1.1448
SMOC1	1.79136	TRANSMEMBRANE PROTEIN C9ORF91 HOMOLOG	1.23658
SMP	-0.97383	TRIM16	-2.82531
SMTNL2	-1.40288	TRMT10A	1.24205
SNX10B	1.55503	TRMT112	0.93739
SNX22	-1.36352	TRP53INP1	-1.61174
SNX27	1.27092	TRPC2	-2.51356
SOCS1	4.46988	TRPC4AP	-1.27799
SOCS3	2.82032	TRPV1	-4.67068
SOCS3	2.65583	TSPAN33	1.33168
SOX11-B	-1.47971	TSPAN7	-1.2472
SPA17	-1.91725	TTF2	1.2717
SPAG1	1.44643	TTPA	-2.61198
SPARC	-1.01213	TTYH2	1.43567
SPIRE1	2.94014	TUBA1C	0.9819

SPRY3	2.13341	TUBB	0.933406
SPTLC2	2.4284	TWF1	-2.45577
SQLE	-1.41266	TXNDC17	0.963951
SRPX	-1.7652	TXNIP	-1.50984
SRPX2	-1.25663	TXNL4B	1.00081
ST6GAL2	-1.377	TYMP	2.33982
ST6GALNAC2	-1.51689	TYMP	2.59016
STAR	1.42786	UBE2D4	1.06211
STAT1	1.67876	UBE2H	-0.970789
STAT3	1.40172	UCK2A	-1.15559
STEAP4	2.68654	ULK2	-1.53058
STEAP4	3.21939	UNC13D	-1.40727
STEAP4	3.38331	UNC93B1	1.09772
STEAP4	2.14507	UQCRH	1.43522
STMN1	-1.63138	USP2	1.2883
STONUSTOXIN SUBUNIT ALPHA	2.90355	USP2	1.18186
STX19	-2.15752	UTP11L	0.938296
STX5	1.39523	VAMP8	-0.860901
SUCNR1	-3.14881	VAV2	-1.27091
SULT1ST1	-1.03083	VCAM1	1.10586
SULT2B1	2.03584	VDAC2	0.862274
SULT2B1	3.72964	VDAC2	2.30991
SUSD3	-1.38307	VERRUCOTOXIN SUBUNIT	-2.87749

			BETA
SWAP70	-1.13936	VIM	-3.41101
SWT1	1.74487	VPS11	0.970752
TAL1	-0.964692	VTN	-5.70894
TAP1	1.87269	VWA5A	1.21191
TAPBPL	1.56281	VWA7	-1.78665
TARS	1.03249	WBSCR27	1.32864
TAS1R1	-1.21888	WDFY2	1.99108
TBC1D2	2.51592	XPA	-1.88608
TCAF	2.61117	YES1	1.10992
TCEA1	-1.07557	YPEL5	-1.19406
TDH	1.24431	ZAN	-1.09084
TDRD7B	1.49406	ZBTB7A	1.22627
TEKT1	1.91241	ZEB2	-1.3194
TEKT2	1.60697	ZFP36L1	-1.1224
TF	5.97514	ZFP36L2	-1.05046
TFG	1.43306	ZGC	-3.24866
TFPI2	-2.417	ZNF513	-1.66768
TGFB1I1	1.57882	ZNF831	-2.55146
TGFB3	1.93966	SLC26A6	-0.979583
TGM1	1.74015	SLC28A3	-1.6961
TGM2	-1.86858	SLC2A1	-0.94423
THAP11	-2.18651	SLC2A9	-1.54306
THBS1	3.55603	SLC35A5	-1.0083
TIMM44	1.03756	SLC40A1	-0.984834
TIMP2	2.22498	SLC43A2	-2.01494
TIMP4	-4.43585	SLC43A3	-4.17531
TJP2	-1.29363	SLC44A2	-1.26414

TKT	-0.99358	SLC44A2	-1.15617
TLR13	-1.21567	SLC44A5B	-1.53281
TM7SF3	1.11008	SLC51A	-2.03408
TM9SF2	-1.38164	SLC5A12	-1.04219
TMEM150C	2.26866	SLC6A13	-2.31814
TMEM183	1.07803	SLC6A6	-1.16589
TMEM230	-1.10252	SLC6A6	-0.952902
TMEM25	-1.20315	SLC8A2	-1.38148
TMEM86A	-2.21323	SLC9A5	-2.3687
TMEM87A	1.13695	SMIM5	-0.997489
TMPRSS2	1.94791	SMP	-1.7174
TNFAIP2	1.16071	SOCS7	-2.47924
TNFAIP3	2.27242	SORD	-0.995171
TNFRSF11B	5.96732	SORT1	-1.88681
TNFRSF4	3.32576	SOX4	-1.61293
TNIK	1.06654	SOX6	-1.31968
TNIP1	1.44713	SPA17	-1.76578
TOMM22	1.37753	SPIDR	-1.67061
TOP3B	3.57976	SPOCK3	-1.60858
TPH2	4.01098	SRPX	-1.18358
TPH2	3.71837	SRPX2	-1.46954
TRAF3IP3/DNASE1L3	-1.38189	SRSF10	-2.07084
TRANSMEMBRANE PROTEIN C9ORF91 HOMOLO	1.22499	ST3GAL1	-2.27091

G			
TRHDE	2.40309	ST3GAL1	-1.62857
TRIB2	2.44248	ST6GAL2	-1.11943
TRIM62	0.929034	STMN1	-1.23286
TRIP10	-0.996205	STX19	-1.45876
TRP53INP1	-1.2567	SUCNR1	-3.81
TRPV1	-3.34313	SUN1	-0.963406
TSHB	4.23088	SUSD3	-1.78158
TSPAN1	1.01567	SWAP70	-1.67121
TSPAN13	-1.48561	TAS1R1	-1.17062
TSPAN33	2.09555	TCAF	-1.92767
TTPA	-2.24288	TCP11L2	-1.58993
TTPA	-3.01731	TGM2	-1.20674
TTYH2	2.02517	THAP11	-2.22061
TUBA1C	0.905931	THBD	-2.89812
TUBB	0.960218	TIMP3	-1.2229
TUSC5	1.84507	TKT	-0.888602
TWF1	-2.86677	TLR13	-1.86451
TYMP	2.10564	TMEM229A	-1.102
TYMP	1.87496	TMEM25	-1.55014
TYPE-4 ICE-STRUCTURING PROTEIN AFP4	3.65842	TMEM86A	-2.30281
UBA1	1.4723	TMPRSS2	-1.33747
UBE2D2	1.26721	TP53INP1	-1.1448

UBE2D2	1.03723	TRIM16	-2.82531
UBE2K	1.04132	TRP53INP1	-1.61174
UBE2L3	1.29717	TRPC2	-2.51356
UBQLN4	0.96148	TRPC4AP	-1.27799
UNCHARACTERIZED PROTEIN C10ORF88 HOMOLOG	2.28492	TRPV1	-4.67068
URGCP	2.71064	TSPAN7	-1.2472
USF1	-0.958026	TTPA	-2.61198
UST	1.38994	TWF1	-2.45577
VCAM1	2.97433	TXNIP	-1.50984
VCP	1.14035	UBE2H	-0.970789
VDAC2	1.9274	UCK2A	-1.15559
VERRUCOTOXIN SUBUNIT BETA	2.35025	ULK2	-1.53058
VIM	-5.135	UNC13D	-1.40727
VPS11	1.3917	VAMP8	-0.860901
VPS53	-1.54623	VAV2	-1.27091
VWA5A	1.5969	VERRUCOTOXIN SUBUNIT BETA	-2.87749
VWA7	-2.25223	VIM	-3.41101
VWA7	-1.57274	VPS37C	-1.34572

WAP	1.12838
WDFY2	1.7919
XMRK	2.33741
XPA	-1.68296
YES1	2.32117
YWLC	2.43647
ZBTB34	-0.962449
ZBTB7A	1.77295
ZDHHC2	-1.9962
ZFP14	-0.994444
ZFP36L3	1.20463
ZNF827	-1.13204
ZNFX1	1.35411
SCARA3	-2.19861
SCCPDH	-1.54684
SCN4B	-2.86999
SDR16C5	-1.09455
SELH	-1.20198
SEMA6D	-1.75697
SERBP1	-1.04862
SESN1	-1.97982
SIGLEC14	-1.03943
SIGLEC5	-1.65271
SIL1	-1.24057
SLC12A3	-4.17981
SLC16A5	-1.84964
SLC25A16	-1.35622
SLC25A53	-1.98869

VPS53	-1.20674
VTN	-5.70894
VWA7	-1.78665
XPA	-1.88608
YPEL3	-0.931113
YPEL5	-1.19406
ZAN	-1.09084
ZBTB16A	-2.09911
ZEB2	-1.3194
ZFP36L1	-1.1224
ZFP36L2	-1.05046
ZGC	-3.24866
ZNF513	-1.66768
ZNF831	-2.55146

SLC26A6	-1.04367
SLC28A2	-2.3002
SLC29A1	-1.88149
SLC34A1	-1.70284
SLC35A5	-1.18954
SLC39A9-A	-2.17028
SLC40A1	-1.62423
SLC43A2	-1.47504
SLC43A3	-2.0079
SLC43A3	-3.18915
SLC44A2	-1.40335
SLC44A5B	-1.8613
SLC6A13	-1.14301
SLC6A14	-3.74282
SLC8A2	-1.63304
SLC9A3R2	-1.15464
SMARCC2	-4.30909
SMIM5	-1.46416
SMP	-0.97383
SMTNL2	-1.40288
SNX22	-1.36352
SOX11-B	-1.47971
SPA17	-1.91725
SPARC	-1.01213
SQLE	-1.41266
SRPX	-1.7652
SRPX2	-1.25663
ST6GAL2	-1.377

ST6GALNAC2	-1.51689
ST8SIA6	-1.469
STMN1	-1.63138
STX19	-2.15752
SUCNR1	-3.14881
SULT1ST1	-1.03083
SUSD3	-1.38307
SWAP70	-1.13936
TAL1	-0.964692
TAS1R1	-1.21888
TCAIM	-1.11677
TCEA1	-1.07557
TFPI2	-2.417
TGM2	-1.86858
THAP11	-2.18651
TIMP4	-4.43585
TJP2	-1.29363
TKT	-0.99358
TLR13	-1.21567
TM9SF2	-1.38164
TMEM230	-1.10252
TMEM25	-1.20315
TMEM86A	-2.21323
TRAF3IP3/DNASE1L3	-1.38189
TRIP10	-0.996205
TRP53INP1	-1.2567
TRPV1	-3.34313

TSPAN13	-1.48561
TTPA	-2.24288
TTPA	-3.01731
TUBA2	-2.87469
TWF1	-2.86677
USF1	-0.958026
VIM	-5.135
VPS53	-1.54623
VWA7	-2.25223
VWA7	-1.57274
XPA	-1.68296
ZBTB34	-0.962449
ZDHHC2	-1.9962
ZFP14	-0.994444
ZNF827	-1.13204

Paper V

Whole transcriptome analysis of the Atlantic cod vaccine response reveals no conventional adaptive immunity

Monica Hongrø Solbakken*[1], Sissel Jentoft*[1,2], Trond Reitan[1], Helene Mikkelsen[3], Kjetill S. Jakobsen[1], Marit Seppola[4]

[1]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

[2]Department of Natural Sciences, University of Agder, Kristiansand, Norway

[3]The Northern Norway Regional Health Authority, Tromsø, Norway

[4]Department of Medical Biology, The Arctic University of Norway, Tromsø, Norway

*Corresponding authors:

Monica Hongrø Solbakken. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. m.h.solbakken@ibv.uio.no, +47 22 84 41 64

Sissel Jentoft. Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066, Blindern, 0316 Oslo, Norway. sissel.jentoft@ibv.uio.no, +47 22 85 72 39

Abstract

Genome sequencing demonstrated that Atlantic cod lacks the Major Histocompatibility Complex class II (*MHCII*), which is key in presenting antigen to the adaptive immune system and thus for establishment of conventional immunological memory. Here, we investigate the immunological response of Atlantic cod using whole transcriptome sequencing during the time-course after vaccination with *Vibrio anguillarum*. The experiment was conducted using siblings from an Atlantic cod family found to be highly susceptible towards vibriosis, in which vaccination has demonstrated improved pathogen resistance. In-depth gene expression analysis at 2, 4, 21 and 42 days post vaccination were conducted. Weak initiation of innate defenses was detected. With respect to genes involved in conventional adaptive immunity, we observed sparse significant differential expression. Intriguingly, the panel of differentially expressed genes was dominated by up-regulated muscle, neuron and metabolism-related pathways. These findings are in line with earlier reports demonstrating changes in muscle growth and increased neuron development post vaccination. Moreover, the up-regulation of metabolism-related pathways demonstrates a shift towards glycolysis, which has in earlier studies been linked to the development of innate memory. Collectively, we find that there is a lack of a clear immunological transcriptomic response related to this vibriosis vaccine. In the light of functional studies demonstrating significant memory in Atlantic cod post vaccination, this indicates the presence of an unknown adaptive mechanism responsible for the establishment memory in Atlantic cod. Likely candidates are CD8+ memory T-cells, memory B-cells activated through T-cell independent mechanisms, innate memory induced through NK-cells or shift in metabolic strategy maintaining epigenetic changes.

Introduction

Teleosts are among the oldest, most diverse and numerous vertebrate infraclasses [1]. Recently, it has become evident that the genetic and phenotypic diversification of this lineage also has manifested through a diverse genetic repertoire related to immunity [2, 3]. Studies investigating the functional outcome of teleost immune responses report classic patterns of gene expression for species harboring genes related to conventional adaptive immunity such as Major Histocompatibility Complex class II (*MHCII*), *CD4*, T-cell and B-cell receptors (*TCR* and *BCR*, respectively). This is exemplified by zebrafishes, which upon vaccination elicit both innate and adaptive responses. More specifically, initial detection is likely performed by pattern recognition receptors (PRRs) with a subsequent onset of inflammation characterized by interleukin 1 beta (IL1B) and interleukin 8 (IL8, CXCL8) and antimicrobial defenses. The classic transition into adaptive immunity is mediated through cytokine signaling leading to the up-regulation of MHCI and MHCII with their corresponding CD8 and CD4 T-cell subsets where the latter generates antibody production by B-cells [4 and references therein]. However, for some teleost species, such as Atlantic cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*), studies collectively describe inconsistent response patterns related to the specificity of the generated antibodies and conclusions as to whether these species have conventional adaptive immunity [5-13]. These studies do, at a gene expression level, describe overall inflammation, antimicrobial peptides, acute phase proteins as well as some interferon-related genes and cell-mediated cytotoxicity [5-7, 14]. However, there are so far no thorough investigations addressing global gene expression of the whole array of genes related to the genetic mechanisms underlying conventional adaptive immunity in these two species. Recently, the large group of cod-like fishes (i.e the order Gadiformes) was found to lack *MHCII*, *CD4* and invariant chain (*Ii*) similarly to Atlantic cod [15, 16]. This explains the lack of reports describing a more or less

71 conventional immune response in codfish. Accordingly, a more detailed and
72 overarching investigation of the Gadiformes' unconventional immune strategy as
73 well as functionality are truly needed. Here, we present a global transcriptome
74 analysis of an Atlantic cod vaccine response. The target disease, vibriosis, occurs in a
75 range of teleost species both under farmed and wild conditions. Left untreated the
76 infection causes fatal hemorrhagic septicemic disease [17]. The fish used in this
77 experiment were family material from the national breeding program of Atlantic cod.
78 The chosen siblings were individuals shown to be highly susceptible to vibriosis and
79 additionally proven to have an increased resistance against the pathogen post
80 vaccination [7]. The vaccine consisted of gram negative *Vibrio anguillarum* (*Listonella*
81 *anguillarum*) bacterin (i.e. suspension of attenuated bacteria). Despite the fact that
82 Atlantic cod show increased resistance post vaccination [7], we find that the
83 transcriptome analyses show barely any inflammatory and innate defense responses.
84 In contrast, highly profound transcriptome responses involving diffuse muscle- and
85 neuron development as well as metabolic pathways were evident. Thus, we find no
86 transcriptomic evidence of conventional adaptive immunity indicative of
87 unconventional mechanisms leading to the adaptive phenotype observed in Atlantic
88 cod.

89 **Results**

90 In this study we have chosen a multifaceted approach to detect differentially
91 expressed genes. It consists of both *de novo* and reference genome based
92 transcriptomics (Trinity [18] and Tuxedo [19], respectively) combined with R-
93 packages EdgeR [20] and CummeRbund [19] for final differential expression analysis
94 and result visualization. In addition, due to our experimental setup with pair-wise
95 controls (mock vaccinated fish) at all time-points, we have applied a custom analysis
96 clustering genes by their expression pattern over time (for details see methods

97 section). Overall, the various approaches applied to this dataset gave somewhat
98 different results but the use of a new reference genome [21] clearly provided better
99 insight as the number of detected differentially expressed genes was increased and
100 the relative amount of annotated genes was improved (Table 1). Below we present
101 our findings focusing on the output from the Tuxedo pipeline and supplement with
102 findings from the other two approaches. This is to capture genes that may not have
103 corresponding gene models in the reference genome [21] due to filtering of smaller
104 genome contigs, thus immune genes located to these contigs are only found using
105 the Trinity approach.

106 The TopHat-Cufflinks-CummeRbund (TCC) pipeline reported 72 differentially
107 expressed genes 2 days post vaccination. The gene ontology (GO) terms of the up-
108 regulated genes detected were reported to be involved in muscle contraction and
109 cardiac muscle cell development. At this stage, there was mainly down-regulation of
110 fibrinolysis but also some platelet degranulation (Table 2). On day 4 there were 132
111 reported differentially expressed genes of which most were up-regulated. The up-
112 regulated genes supported a continued effect on muscle-related systems but also a
113 strong positive effect on metabolism-related pathways such as carbohydrate
114 catabolic and pyruvate metabolic processes (Table 2). The 21 day post vaccination
115 comparison showed 291 differentially expressed genes. The muscle and metabolism-
116 related systems were still up-regulated, but an extensive network of transcripts
117 related to neuron development, vesicle generation and movement as well as
118 membrane potentials appeared. In parallel, genes related to endoderm development
119 and endocrine pancreatic systems were down-regulated (Table 2). At 42 days post
120 vaccination there were 376 differentially expressed genes. Transcripts associated
121 with metabolism-related pathways were still up-regulated. However, neuron-related
122 pathways and intracellular transport appeared as more prominent features.

123 Simultaneously, there was an overall down-regulation of transcripts associated with
124 muscle-related processes and blood vessel development as well as down-regulation
125 of virus entry-associated transcripts (alias endocytosis) (Table 2).

126 The custom clustering of genes according to expression patterns over time in a
127 control dependent manner detected 600 and 180 significant genes for Trinity and
128 Cufflinks, respectively (Table 1). The corresponding GO:terms for genes with an
129 overall increased expression pattern suggests extensive effects on amino acid and
130 iron ion transport. There was an up-regulation of transcripts involved in immune-
131 related phagocytosis, movement of protein to the endoplasmatic reticulum and
132 regulation of mRNA processing (Table 3). In contrast, a range of systems such as
133 ribonucleoprotein complex biogenesis, positive regulation of calcium ion transport
134 into cytosol, neuron projection extension, melanin biosynthetic process and signal
135 transduction involved in cell cycle checkpoint were down-regulated (Table 3). Lastly,
136 genes with a "freestyle" expression pattern (alternating trends over time) were
137 related to ATP synthesis, cobalamin (B-vitamin) synthesis, ribosome assembly, RNA
138 splicing, posttranscriptional regulation of expression, non-coding RNA processing
139 and more (Table 3).

140 Among the immune-related genes there are only a few found to be significantly
141 differentially expressed or with a clear expression pattern over time in a control
142 dependent manner. Overall, they are related to a range of immunological systems
143 (Table 4). Furthermore, genes related to to the same or co-dependent systems do not
144 appear to be differentially expressed within overlapping time frames. Intriguingly,
145 there is little or no sign of up-regulation of transcripts involved in inflammation.
146 Further, there is an initial down-regulation of acute phase proteins (day 2) followed
147 by up-regulation combined with an increase in genes related to iron metabolism
148 (day 4 and onward). There is up-regulation of heat shock proteins throughout the

149 experiment. Further, we found some differential expression among genes related to
150 complement, autophagy, pattern recognition, phagosome development and
151 apoptosis. But we also detected down-regulation of chemotactic factors at day 42,
152 and overall down-regulation of proteasomal subunits. The only genes detected that
153 were directly related to adaptive immunity were immunoglobulin light chains and
154 beta 2-microglobulin (*B2M*), which displayed increasing expression patterns over
155 time (Table 4).

156 ## Discussion

157 **No clear evidence for conventional adaptive immunity**

158 Vaccine responses in teleosts, by both immersion and injection techniques, towards
159 various pathogens have been characterized mainly by using quantitative PCR
160 methods or estimating antibody titers. There is some disparity in gene expression
161 response, which generally appears to be connected to vaccination method
162 (immersion or injection technique), and vaccine content and adjuvant use [4-10, 22-
163 25]. Further, the conclusions drawn on functional investigations of the antibody
164 response vary from a heterogenic repertoire of natural antibodies, through
165 pathogen-lipopolysaccharide-specific antibodies, to specific antibodies able to
166 discriminate between pathogen serogroups [7-13]. Immersion vaccination towards
167 vibriosis in Atlantic cod has been shown to induce expression of *IL1B*, *IL10* and
168 *IL12B* (the latter is a subunit of both *IL12* and *IL23*) together with hepcidin (*HAMP*).
169 Induction of interferon gamma (*IFNG*) and lysozyme (*LYG2*) varies somewhat and
170 appers to be correlated with vaccine content [7, 8], whereas injection-based strategies
171 have induced *LYG2*, transferrin (*TF*), interferon regulatory factor 1 (*IRF1*), granzyme
172 and apolipoproteins [5, 6]. In this study, we find that innate defenses in Atlantic cod
173 are only marginally affected by this immersion vaccination against *V. anguillarum*
174 with no significant up-regulation of *IL1B*, *IL10* and *IL12B*. However, an up-

175 regulation of inflammasome-related components *NLRP3* and *CASP1* early in the
176 time-series indicate that a minor response is present (Table 4). This is concordant
177 with the weak differential expression of immune genes previously reported using
178 qPCR on this system [7]. We further find that transcripts from chemotactic factors,
179 such as *CXCL1*, *CXCL8*, *CXCL12* and *CCL2*, are down-regulated at day 42 or display
180 an overall decreasing expression pattern (Table 4). We also detect an overall up-
181 regulation of a single antibacterial transcript *LYG2* together with acute-phase
182 reactants *TF*, *FGB*, *HPX*, apolipoproteins and *HSPA1A* (Table 4). In teleosts
183 harboring a conventional adaptive immune system, vaccination affects the
184 expression of *MHCI* and *MHCII* as well as the corresponding co-receptors on T-cells
185 (*CD4* and *CD8*) together with an increase of immunoglobulin expression [4 and
186 references therein]. We find an overall up-regulation of *B2M*, immunoglobulin light
187 chains, but notably, no corresponding up-regulation of annotated *MHCI* or T-cell co-
188 receptors. In addition, we found that several proteasomal subunits are down-
189 regulated (Table 4). The observed subtle significant changes to the phagosomal
190 pathway (Table 4) could indicate a small amount of cross-presentation on MHCI [26],
191 which correlate with the reported MHCI gene expansion and presence of endosomal
192 sorting signals in Atlantic cod likely related to alternative loading of antigen on
193 MHCI [27]. We further observed the up-regulation of genes related to cytotoxic and
194 lysosome functions such as perforin (*PRF1*) and cathepcin (*CTSC*) (Table 4)
195 indicating some increase in phagocytosis and cellular cytotoxicity defense
196 mechanisms [26, 28]. CD8+ T-cells are capable of establishing memory [29]. However,
197 we do not detect any of the cytokines involved in differentiation of the CD8+
198 memory lineages. Considering that Atlantic cod does not have *CD4* [15] and that
199 fishes do not form germinal centers [30], T-cell independent activation of B-cells
200 could explain the increase in immunoglobulin expression [31]. However, there was
201 no indication of B-cell activation, with or without T-cell help beyond the up-

202 regulation of immunoglobulin light chains (Table 4). Collectively, we find no clear
203 transcriptomic signals indicating the establishment of memory using conventional
204 pathways in Atlantic cod suggesting that other alternative strategies exist.

205 **Vaccination affects muscle and neuron development**

206 Our results reveal that the immune system in Atlantic cod is not heavily influenced
207 by the vaccination as seen by the minute number of significantly differentially
208 expressed immune genes. However, there are major responses from other non-
209 immune system related pathways. We observed large changes in muscle- and
210 neuron-related pathways (Figure 1). Similar findings have been reported in other
211 teleost studies where gene regulation post vaccination has been investigated using
212 microarrays. These studies reported that the contribution by the immune system was
213 overshadowed by other pathways such as metabolism [23, 24, 32]. Vaccination has
214 been shown to have an effect on growth rate in salmon as well as the development of
215 the heart muscle, which is in concordance with our findings [33, 34]. However, none
216 of the earlier reports have revealed such a dominant presence of muscle and neuron-
217 related pathways as seen in this study. Moreover, we find an additional up-
218 regulation of genes related to neuronal development — especially within the two
219 last time-points (Table 2). This is in concordance with findings shown in mice; when
220 immunized early in life they tend to develop more complex structures in certain
221 areas of the brain characterized by more neuron interactions. Comparable
222 discoveries have been reported in humans where there is an unknown mechanism
223 leading to a neuronal protective effect post immunization [35, 36]. Taken together
224 our global transcriptome profiling did not reveal any strong effect on gene
225 expression of immune genes, but in contrast showed up-regulation of metabolism,
226 muscle and neuron-developmental related pathways.

227 **Shift in metabolic strategy — indicative of epigenetic imprinting and innate**
228 **memory?**
229 The profound changes in the metabolic strategy of Atlantic cod upon vaccination
230 (Figure 1, Table 1-4), concordant with transcriptomic studies on other fish species
231 post vaccination demonstrating similar effects regarding up-regulated metabolic
232 pathways [23, 24, 32]. However, in contrast to Atlantic cod, these species (all
233 possessing *MHCII*) also display clear responses from the immune system, both with
234 respect to innate and adaptive defenses. Our observed change in metabolism might
235 be associated with the alterations in muscle and neuronal development, but it could
236 also be a characteristic indicative of the immunological paradigm called innate or
237 trained memory [37]. The underlying mechanisms are yet to be fully elucidated, in
238 both invertebrates and vertebrates. In the invertebrate *Artemia*, a recent study
239 demonstrates the interaction between of both genetic and epigenetic factors to form
240 innate memory where up-regulation of *HSP70* (equivalent to *HSPA1A* in mammals)
241 and *HMGB1* (high mobility group box 1) — genes involved in pathogenic stress
242 responses — were implicated [38]. In our data, we also find up-regulation of
243 *HSPA1A*, but no significant differential expression of *HMGB1/2* (Table 4). In
244 mammals innate memory has been attributed to several cell lineages of the immune
245 system such as monocytes, macrophages, natural-killer (NK) cells and innate
246 lymphoid cells. There is also a more prominent role for PRRs combined with an
247 elaborate orchestration of cytokines not involved in the development of conventional
248 adaptive memory through the B-cell/T-cell interaction [39, 40]. In this regard, we
249 observe some up-regulation of NK-cell related markers, active down-regulation of
250 neutrophil and monocyte recruitment and down-regulation of a single interleukin
251 IL34 (Table 4) indicating that a potential innate memory mechanism in Atlantic cod
252 could be related to NK-cells. However, we find no candidate cytokines, which likely
253 is related to the greater cytokine diversity in fish compared to mammals, making

254 annotation of cytokines problematic [2, 41, 42]. There is also an epigenetic factor in
255 mammalian innate memory initiated through various transcription factors enabling
256 epigenetic reprogramming [39, 40] where a shift in metabolism — specifically from
257 oxidative phosphorylation to glycolysis via the MTOR signaling pathway — has
258 been suggested to provide metabolites responsible for maintaining epigenetic
259 changes [37, 39, 40]. Our data support a shift towards glycolysis with up-regulation
260 of a range of genes, including *mTOR*, suggesting the presence of innate memory
261 mechanisms in Atlantic cod.

262 To summarize, our findings do not strongly support any conventional adaptive
263 mechanism in the vaccine response against vibriosis at a gene expression level in
264 Atlantic cod. Although this is predicted due to the lack of *MHCII* and *CD4*, this has
265 — to our knowledge — not previously been shown. Further, we have demonstrated
266 that the innate response associated with vaccination, such as inflammation, is more
267 or less absent. However, we show that Atlantic cod seems to have a strong post-
268 vaccination response involving muscle and neuron development as well as a range
269 of metabolic pathways. The lack of a clear immunological transcriptomic response
270 shown in this study — taken together with other functional studies demonstrating
271 significant memory in Atlantic cod post vaccination — indicates the presence of an
272 unknown mechanism responsible for the establishment of innate or another
273 unconventional immunological pathway conveying memory. Likely candidates are
274 memory-CD8+ T-cells, memory B-cells activated through T-cell independent
275 mechanisms and innate memory induced through NK-cells or shift in metabolic
276 strategy maintaining epigenetic changes.

277 **Methods**

278 Please see GitHub repository: for details.

Fish and experiment setup

The fish selected for this study were part of a larger vaccine investigation setup published in 2013 [7]. The fish used in this experiment were family material from the national breeding program of Atlantic cod (www.nofima.no) where the individuals were siblings derived from a family highly susceptible towards vibriosis. The fish were transported (at approx. 1.6 g) to the Aquaculture Research Station (Tromsø, Norway) for grow-out in seawater of 3.4 % salinity at 10 °C, 24 h light and fed with commercial feed (BioMar, Norway). The rates of water inflow were adjusted to an oxygen saturation of 90–100 % in the outlet water. We selected a family with low resistance towards Vibriosis for further investigations. See Mikkelsen et al. for further information [7]. The fish were reported to be healthy without any history of diseases and the experiment was approved by the National Animal Research authority in Norway. All methods were in accordance with the approved guidelines.

The original experiment had the following setup: The experimental dip vaccine produced by PHARMAQ AS (Norway) contained bacterin of *V. anguillarum* serotype O2b isolate 4299. Cod (approx. 2.5 g) were dip vaccinated by immersion for 30 seconds in diluted vaccine (1:10 in seawater), according to the manufacturer's instruction. Controls were mock vaccinated by dipping in sea water without vaccine. The fish, 10 vaccinated and 10 control groups, were distributed in 20 parallel, circular, centrally drained, fiberglass tanks (100 L) with approx. 100-125 fish in each tank (density <20 kg/dm3). Fish for pre-challenge (n= 72) were left untreated and kept in a separate tank. One week prior to challenge the fish were anaesthetized with Metacainum (Norsk Medisinaldepot, Norway) (70 mg L-1) and marked at the operculum Visible Implant Fluorescent Elastomer (Northwest Marine Technology Inc. US) before being distributed in 4 x 500 L tanks (2 tanks with coastal cod and North East Arctic cod families, respectively) with 80 fish from each family (40

305 vaccinated and 40 controls) in each tank (Table 1). Samples used for RNA was
306 collected at 2 days, 4 days, 21 days and 42 days post vaccination. Six individuals
307 were collected from each vaccinated and control group (n = 48). Head kidney and
308 spleen were aseptically removed and transferred to RNA-Later (Ambion), and kept
309 at 4 °C overnight before being stored at -80 °C.

310 **RNA isolation, library preparation and sequencing**

311 The Vibrio vaccinated treated and control sampled tissues were homogenized in 1x
312 lysis buffer using MagNA Lyser Green Beads and the MagNa Lyser Instrument
313 (Roche Diagnostics). Total RNA was purified using an ABI Prism 6100 Nucleic Acid
314 Prep Station (Applied Biosystems) with the recommended on-column DNase
315 treatment.

316 All RNA isolates (totalRNA or mRNA) were quality controlled using Agilent 2100
317 Bioanalyzer (BioRad) before library preparation.

318 Libraries for RNAseq were prepared using the TruSeq™ RNA low-throughput (LT)
319 protocol (Illumina). All samples were fragmented for 4 minutes to obtain the size
320 distribution desired according to the TruSeq protocol. A library overview is
321 available in the Github repository.

322 All libraries were sequenced 100bp paired-end (PE) at the Norwegian Sequencing
323 Centre on the Illumina HiSeq 2000 (www.sequencing.uio.no).

324 The obtained sequences were trimmed using Sickle with a 40 bp minimum
325 remaining sequence length, a Sanger quality threshold of 20 and no 5′ end trimming
326 [43]. Remaining sequencing adapters were removed using Cutadapt v1.0 [44].
327 Results were quality controlled using FastQC v0.9.2 to ensure improvement
328 compared to raw data [45].

329 **Reference-genome based approach using Tuxedo**

330 The second version of the Atlantic cod genome [21] was used as reference for a
331 Topphat/Cufflinks pipeline according to the workflow described in [19]. Mapping of
332 samples towards the reference-genome GFF3 file was performed with Tophat v2.0.14
333 with default settings. Sample-specific transcriptomes were generated with Cufflinks
334 v2.1.1. Cuffmerge was used to concatenate all the individual transcriptomes.
335 Differential expression analysis was performed with Cuffdiff in a pair-wise manner
336 between treated and control for each time-point. The output from Cuffdiff was
337 further analyzed using CummeRbund v2.8.2 in R v3.1.3 for presentation purposes
338 [46, 47].

339 **Reference-genome-guided approach using Trinity**

340 Two RNAseq studies provided reads for the transcriptome assembly used here – the
341 reads derived from the *Vibrio* vaccination described above and the reads derived
342 from a *Francisella* challenge study with the same number of samples (Solbakken et al.,
343 paper IV in this thesis). In total, the 96 libraries (48 from each experiment) provided
344 on average 20.51 million trimmed read-pairs resulting in 1 969.31 million reads in
345 total.

346 We applied the Trinity transcriptome assembler v 2.0.6 using the genome-guided
347 option with the second version of the Atlantic cod genome [21].The genome was
348 indexed using Bowtie1 (v1.0.0) and then mapped using Tophat (v2.0.9) and sorted
349 with Samtools (v0.1.19). The built-in normalization step of Trinity was applied
350 reducing the trimmed read dataset to approximately 45 million read pairs [18, 48].
351 The following parameters were changed for the Trinity run: genome-guided, max
352 intron 10 000, max memory 150 Gb, bflyHeapSpaceMax 10G, bflyCPU 12 and CPU
353 10.

354 The assembly was evaluated with the built-in trinity_stats.pl and
355 align_and_estimate_abundance.pl — the latter with RSEM estimation method and
356 bowtie aligner. The abundance estimation output was further used to filter the
357 assembly on transcript level with FPKM = 1 using filter_fasta_by_rsem_values.pl.
358 This resulted in 44 543 transcripts with an overall contig N50 of 2 568 bp, median
359 contig length of 1 132 bp and a total of ~73.3 million assembled bases. Based on the
360 longest open reading frames (ORFs) the transcript dataset was reduced to 32 934
361 "genes" with an overall contig N50 of 2 490 bp and median contig length of 1 014 bp.

362 Overall annotation was performed using Trinotate v2.0.1 following all mandatory
363 steps with default parameters on the non-filtered assembly and transferred to the
364 filtered assembly transcripts. The annotation of genes specifically discussed in this
365 study have been verified through reciprocal BLAST by extracting the longest isoform
366 of the gene in question and subjecting it to a BLASTX towards all UniProt entries
367 using the UniProt BLAST tool with default settings [49].

368 **Sample mapping, read count extraction**

369 The trimmed reads from all Vibrio-related samples were mapped against the filtered
370 Trinity assembly to simplify interpretation of results using the built-in
371 align_and_estimate_abundance script in Trinity with RSEM estimation method and
372 bowtie aligner, before extracting raw counts using
373 abundance_estimates_to_matrix.pl again with RSEM as the estimation method.

374 **Error distributions and differential expression analyses**

375 Most RNAseq analysis packages assume that such data follows a negative binomial
376 distribution of variability. We tested this assumption using a custom script testing
377 the fit of the Poisson distribution, the negative binomial distribution and the zero-
378 inflated negative binomial distribution (using the pscl package in R, script available

379 in the GitHub repository). About 90 % of all genes were classified as having negative
380 binomial distribution and thus, in all cases, the negative binomial distribution was
381 used for all down-stream analyses.

382 For the reference-genome based analysis CuffDiff performed the differential
383 expression analysis with default parameteres. For the Trinity-generated read-counts,
384 differential expression analysis was performed using the R-package edgeR
385 specifying the following contrasts: 2 day vaccinated vs control, 4 day vaccinated
386 versus control, 21 day vaccinated versus control, and 41 day vaccinated versus
387 control, and otherwise default settings.

388 **Custom script approach for gene expression pattern clustering**

389 We wanted to further characterize the behavior of the dataset outside of what the
390 most common RNAseq differential expression analysis packages could provide in
391 terms of the "genes'" being dependent on time and/or treatment (most analysis
392 packages provide pair-wise analysis options or time-series with a time 0 — not time-
393 series analysis with control samples for each time-point). Expression patterns were
394 to be classified into categories: increasing expression over time, decreasing
395 expression over time, expression pattern with an internal maximum (quadratic,
396 positive), expression pattern with an internal minimum (quadratic, negative) and
397 freestyle expression pattern (alternating trends over time) — in both control-
398 dependent and independent manners. In addition, two categories named no
399 control/no time dependency and control dependency were added. Note that if a
400 quadratic effect was found but with minima/maxima outside the data material, it
401 would be classified as either increasing or decreasing, depending on the estimated
402 quadratic effect). This categorization was performed with a set of regression models;
403 no time dependency, linear time dependency, quadratic time dependency, factorial
404 time dependency, pure treatment effect (no time dependency), treatment combined

405 with linear time (interaction), treatment combined with quadratic time (interaction)
406 and treatment combined with factorial time (interaction). Estimated regression
407 coefficients were then used for determining in which time dependency category each
408 gene expression was to be classified.

409 **Evaluation of RNAseq experiment**

410 The overall quality evaluation of the samples revealed similar trends for dispersion
411 and biological variance between samples. However, the MDS clustering revealed no
412 clear clustering of samples. (Supplementary figures 1,2). The primary differential
413 expression analysis (cutoff p=0.05) reported in total 871 differentially expressed
414 genes (DEGs) (Table 1).

415 **GO and gene network analyses**

416 The reported differentially expressed genes from the primary analyses were
417 analyzed in Cytoscape [50] using the plugin ClueGO [51]. ClueGO was run with
418 default settings selecting biological and immunological related systems and a p-
419 value cutoff of 0.05 unless otherwise stated.

420 ## References

421 1. Eschmeyer, W.N. and R. Fricke, *CATALOG OF FISHES: GENERA, SPECIES, REFERENCES.* .
422 2015.
423 2. Zhu, L.Y., et al., *Advances in research of fish immune-relevant genes: a comparative*
424 *overview of innate and adaptive immunity in teleosts.* Dev Comp Immunol, 2013.
425 **39**(1-2): p. 39-62.
426 3. Solbakken, M.H., et al., *Evolutionary redesign of the Atlantic cod (Gadus morhua L.)*
427 *Toll-like receptor repertoire by gene losses and expansions.* Sci Rep, 2016. **6**: p. 25211.
428 4. Liu, X., et al., *Profiling immune response in zebrafish intestine, skin, spleen and kidney*
429 *bath-vaccinated with a live attenuated Vibrio anguillarum vaccine.* Fish Shellfish
430 Immunol, 2015. **45**(2): p. 342-5.

431 5. Caipang, C.M., M.F. Brinchmann, and V. Kiron, *Profiling gene expression in the spleen*
432 *of Atlantic cod, Gadus morhua upon vaccination with Vibrio anguillarum antigen.*
433 Comp Biochem Physiol B Biochem Mol Biol, 2009. **153**(3): p. 261-7.
434 6. Caipang, C.M., et al., *Intraperitoneal vaccination of Atlantic cod, Gadus morhua with*
435 *heat-killed Listonella anguillarum enhances serum antibacterial activity and*
436 *expression of immune response genes.* Fish Shellfish Immunol, 2008. **24**(3): p. 314-22.
437 7. Mikkelsen, H. and M. Seppola, *Response to vaccination of Atlantic cod (Gadus*
438 *morhua L.) progenies from families with different estimated family breeding values*
439 *for vibriosis resistance.* Fish Shellfish Immunol, 2013. **34**(1): p. 387-92.
440 8. Mikkelsen, H., et al., *Vibriosis vaccines based on various sero-subgroups of Vibrio*
441 *anguillarum O2 induce specific protection in Atlantic cod (Gadus morhua L.) juveniles.*
442 Fish Shellfish Immunol, 2011. **30**(1): p. 330-339.
443 9. Corripio-Miyar, Y., et al., *Vaccination experiments in the gadoid haddock,*
444 *Melanogrammus aeglefinus L., against the bacterial pathogen Vibrio anguillarum.*
445 Vet Immunol Immunopathol, 2007. **118**(1-2): p. 147-53.
446 10. Gudmundsdottir, S., et al., *Specific and natural antibody response of cod juveniles*
447 *vaccinated against Vibrio anguillarum.* Fish Shellfish Immunol, 2009. **26**(4): p. 619-24.
448 11. Schrøder, M.B., et al., *Comparison of antibody responses in Atlantic cod (Gadus*
449 *morhua L.) to Vibrio anguillarum, Aeromonas salmonicida and Francisella sp.* Fish
450 Shellfish Immunol, 2009. **27**(2): p. 112-9.
451 12. Lund, V., S. Bordal, and M.B. Schroder, *Specificity and durability of antibody*
452 *responses in Atlantic cod (Gadus morhua L.) immunised with Vibrio anguillarum O2b.*
453 Fish Shellfish Immunol, 2007. **23**(4): p. 906-10.
454 13. Magnadottir, B., et al., *Natural antibodies of cod (Gadus morhua L.): specificity,*
455 *activity and affinity.* Comp Biochem Physiol B Biochem Mol Biol, 2009. **154**(3): p.
456 309-16.
457 14. Ellingsen, T., et al., *Francisella noatunensis in Atlantic cod (Gadus morhua L.);*
458 *waterborne transmission and immune responses.* Fish Shellfish Immunol, 2011. **31**(2):
459 p. 326-33.

460 15. Star, B., et al., *The genome sequence of Atlantic cod reveals a unique immune system.*
461 Nature, 2011. **477**(7363): p. 207-10.
462 16. Malmstrøm, M., et al., *Evolution of the immune system influences speciation rates in*
463 *teleost fishes.* Nat Genet, In press.
464 17. Frans, I., et al., *Vibrio anguillarum as a fish pathogen: virulence factors, diagnosis and*
465 *prevention.* J Fish Dis, 2011. **34**(9): p. 643-61.
466 18. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-seq using the*
467 *Trinity platform for reference generation and analysis.* Nat Protoc, 2013. **8**(8): p.
468 1494-512.
469 19. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq*
470 *experiments with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.
471 20. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for*
472 *differential expression analysis of digital gene expression data.* Bioinformatics, 2010.
473 **26**(1): p. 139-40.
474 21. Tørresen, O.K., et al., *An improved genome assembly uncovers a profilic tandem*
475 *repeat structure in Atlantic cod.* Submitted.
476 22. Chettri, J.K., et al., *Comparative evaluation of infection methods and environmental*
477 *factors on challenge success: Aeromonas salmonicida infection in vaccinated rainbow*
478 *trout.* Fish Shellfish Immunol, 2015. **44**(2): p. 485-95.
479 23. Zhang, H., et al., *Transcriptome profiling reveals Th17-like immune responses induced*
480 *in zebrafish bath-vaccinated with a live attenuated Vibrio anguillarum.* PLoS One,
481 2013. **8**(9): p. e73871.
482 24. Pridgeon, J.W., et al., *Global gene expression in channel catfish after vaccination with*
483 *an attenuated Edwardsiella ictaluri.* Fish Shellfish Immunol, 2012. **32**(4): p. 524-33.
484 25. Schrøder, M.B., et al., *Early vaccination and protection of Atlantic cod (Gadus*
485 *morhua L.) juveniles against classical vibriosis.* Aquaculture, 2006. **254**(1-4): p. 46-53.
486 26. Neefjes, J., et al., *Towards a systems understanding of MHC class I and MHC class II*
487 *antigen presentation.* Nat Rev Immunol, 2011. **11**(12): p. 823-836.

488 27. Malmstrom, M., et al., *Unraveling the evolution of the Atlantic cod's (Gadus morhua*
489 *L.) alternative immune strategy.* PLoS One, 2013. **8**(9): p. e74004.
490 28. Trapani, J.A. and M.J. Smyth, *Functional significance of the perforin/granzyme cell*
491 *death pathway.* Nat Rev Immunol, 2002. **2**(10): p. 735-47.
492 29. Jameson, S.C. and D. Masopust, *Diversity in T cell memory: an embarrassment of*
493 *riches.* Immunity, 2009. **31**(6): p. 859-71.
494 30. Magor, B.G., *Antibody Affinity Maturation in Fishes-Our Current Understanding.*
495 Biology (Basel), 2015. **4**(3): p. 512-24.
496 31. Vinuesa, C.G. and P.P. Chang, *Innate B cell helpers reveal novel types of antibody*
497 *responses.* Nat Immunol, 2013. **14**(2): p. 119-26.
498 32. Jiang, J., et al., *Differential transcriptomic response in the spleen and head kidney*
499 *following vaccination and infection of Asian seabass with Streptococcus iniae.* PLoS
500 One, 2014. **9**(7): p. e99128.
501 33. Fraser, T.W., et al., *Vaccination and triploidy increase relative heart weight in farmed*
502 *Atlantic salmon, Salmo salar L.* J Fish Dis, 2015. **38**(2): p. 151-60.
503 34. Berg, A., et al., *Time of vaccination influences development of adhesions, growth and*
504 *spinal deformities in Atlantic salmon Salmo salar.* Dis Aquat Organ, 2006. **69**(2-3): p.
505 239-48.
506 35. Li, Q., et al., *Neonatal vaccination with bacille Calmette-Guerin promotes the*
507 *dendritic development of hippocampal neurons.* Hum Vaccin Immunother, 2016.
508 **12**(1): p. 140-9.
509 36. Fissolo, N., et al., *Treatment with MOG-DNA vaccines induces CD4+CD25+FoxP3+*
510 *regulatory T cells and up-regulates genes with neuroprotective functions in*
511 *experimental autoimmune encephalomyelitis.* J Neuroinflammation, 2012. **9**: p. 139.
512 37. Cheng, S.C., et al., *mTOR- and HIF-1alpha-mediated aerobic glycolysis as metabolic*
513 *basis for trained immunity.* Science, 2014. **345**(6204): p. 1250684.
514 38. Norouzitallab, P., et al., *Probing the phenomenon of trained immunity in*
515 *invertebrates during a transgenerational study, using brine shrimp Artemia as a*
516 *model system.* Sci Rep, 2016. **6**: p. 21166.

517 39. Netea, M.G., et al., *Trained immunity: A program of innate immune memory in*
518 *health and disease.* Science, 2016. **352**(6284): p. aaf1098.
519 40. Gardiner, C.M. and K.H. Mills, *The cells that mediate innate immune memory and*
520 *their functional significance in inflammatory and infectious diseases.* Semin Immunol,
521 2016.
522 41. Buchmann, K., *Evolution of Innate Immunity: Clues from Invertebrates via Fish to*
523 *Mammals.* Front Immunol, 2014. **5**: p. 459.
524 42. Bird, S. and C. Tafalla, *Teleost Chemokines and Their Receptors.* Biology (Basel), 2015.
525 **4**(4): p. 756-84.
526 43. Joshi, N., *Sickle - a windowed adaptive trimming tool for FASTQ files using quality*.
527 44. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing*
528 *reads.* Bioinformatics in Action, 2012. **17**(1): p. 10-12.
529 45. Andrews, S. *The FastQC project*. 2011; Available from:
530 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
531 46. R-Core-Team. *R: A Language and Environment for Statistical Computing*. 2015;
532 Available from: http://www.R-project.org.
533 47. Goff, L., C. Trapnell, and D. Kelley, *CummeRbund: Analysis, exploration, manipulation,*
534 *and visualization of Cufflinks high-throughput sequencing ata*. 2013.
535 48. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data*
536 *without a reference genome.* Nat Biotechnol, 2011. **29**(7): p. 644-52.
537 49. UniProt, *UniProt: a hub for protein information.* Nucleic Acids Res, 2015.
538 **43**(Database issue): p. D204-12.
539 50. Shannon, P., et al., *Cytoscape: a software environment for integrated models of*
540 *biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
541 51. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene*
542 *ontology and pathway annotation networks.* Bioinformatics, 2009. **25**(8): p. 1091-3.

543

544

Acknowledgements

This work was supported by The Research Council of Norway (Grant number 199806/S40 to KSJ). The reference genome assembly was made using the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/. The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities.

Additional information

This manuscript has a GitHub repository providing all data.

Competing financial interests

The authors declare no competing financial interests.

Corresponding author

Monica Hongrø Solbakken, m.h.solbakken@ibv.uio.no

Sissel Jentoft; sissel.jentoft@ibv.uio.no

Tables and figures

Table 1 Genes reported as significantly different from control, with or without corresponding annotation, for all analysis approaches applied. For the custom script output only genes reported with expression patterns in a control dependent manner are depicted. Rows corresponding to EdgeR are derived from the de novo transcriptome whereas those derived from CummeRbund are derived from the reference genome gene models. Note that the Trinity pipeline resulted in more genes than the reference genome approach which affects the numbers presented here.

Method	Time-point or pattern	No of "genes"	No of annotated "genes"
EdgeR	2day up	37	1
	4day up	0	0
	21day up	75	20
	42day up	1	0
	2day down	3	0
	4day down	7	5
	21day down	9	4
	42day down	0	0
Custom script	Increase time	600	128
	Internal max time	121	44
	Decrease time	545	232
	Internal min time	900	254
	Freestyle time	1360	780
CummeRbund	2day up	53	35
	4day up	124	109
	21day up	217	174
	42day up	117	93
	2day down	19	13
	4day down	8	3
	21day down	74	52
	42day down	259	209
Custom script	Increase time	180	121
	Internal max time	36	25
	Decrease time	261	215
	Internal min time	62	39
	Freestyle time	176	92

573 **Table 2** All major GO:term clusters reported by ClueGO in Cytoscape with a p-value cutoff
574 = 0.05 from genes clustered according to expression patterns using our custom script.

Time	GO:terms
2 days up	Striated muscle contraction
	Cardiac cell development
	Muscle filament sliding
2 days down	Platelet degranualtion
	Fibrinolysis
4 days up	Pyruvate metabolic process
	Muscle filament sliding
	Carbohydrate catabolic process
	Phosphatidylcholine metabolic process
	Negative regulation of actin filament polymerization
	Regulation of ATPase activity
	Response to magnesium ion
4 days down	No significant terms
21 days up	Synaptic vesicle transport
	Regulation of neuron projection development
	NADH regeneration
	Regulation of membrane potential
	Actin-mediated cell contraction
	Intermediate filament cytoskeleton organization
	Negative regulation of neuron apoptopic process
	Adult behaviour
	Neuron maturation

	Negative regulation of neuron apoptopic process
	Serine phosphorylation of STAT3 protein
	Establishment of protein localization to plasma membrane
	Phosphatidylcholine metabolic process
	Neuron cell-cell adhesion
	Azole transport
	Negative regulation of microtubule polymerization
	Membrane biogenesis
21 days down	Endoderm cell differentiation
	Endocrine pancreas development
42 days up	Regulation of neurotransmitter levels
	Cytoskeleton-dependent intracellular transport
	Hexose catabolic process
	Negative regulation of microtubule polymerization
	Neural nucleus development
	Paranodal junction assembly
	Regulation of potassium ion transmembrane transport
	Actin-mediated cell contraction
	Golgi to plasma membrane transport
42 days down	Muscle contraction
	Regulation of calcium ion transport
	Regulation of muscle contraction
	Viral entry into host

	Blood vessel morphogenesis
	Positive regulation of locomotion
	Glomerular filtration
	Mesenchyme development
	Glycosaminoglycan biosynthetic process
	Vasculogenesis
	Negative regulation of behavior
	Vascular endothelial growth factor signaling pathway
	Regulationof cell-substrate adhesion
	Negative regulation of phosphatase activity
	Regulation of stress fiber assembly

575

576 **Table 3** All major GO:terms as well as pathways from KEGG and WIKI pathway databases
577 reported by ClueGO in Cytoscape with a p-value cutoff = 0.05.

Pattern	**GO:terms (Custom scripts)**
Increase over time ctr dependent	Amino acid transport
	Iron ion transport
	SRP-dependent co-translational protein targeting to membrane
	Immune response-regulating cell surface receptor signaling pathway involved in phagocytosis
	Regulation of mRNA processing
	Negative regulation of mitochondrion organization
	Filopodium assembly

	Regulation of membrane depolarization
	Regulation of mitophagy
Internal maximum ctr dependent	No significant terms
Decrease over time ctr dependent	Ribonucleoprotein complex biogenesis
	Positive regulation of calcium ion transport into cytosol
	Cytoskeleton-dependent intracellular transport
	'de novo' posttranslational protein folding
	Cell-substrate adherence junction assembly
	Regulation of mRNA splicing, via spliceosome
	Regulation of cellular amine metabolic process
	Pigment metabolic process
	RNA localization
Internal minimum ctr dependent	No significant terms
Freestyle ctr dependent	ATP synthesis coupled electron transport
	Cobalamin metabolic process
	Ribosome assembly

578

579

580 **Table 4** Overview of immune genes with corresponding significant expression pattern.
581 Results from cufflinks supplemented with edgeR or custom scripts where Cufflinks data are
582 missing.*genes have been manually annotated. **there are several gene copies.

Gene	Up	Down	Pattern	R package
Pro/anti-inflammatory				
IL1B			NS	
IL10			NS	
IL12B			NS	
IL6*			NS	
TGFb		42		CummeRbund
TNF*			NS	
Complement				
C1QBP			Freestyle	edgeR/CummeRbund
C1QL4	21			
CD93		42		
C3			Freestyle	edgeR
C6			Freestyle	CummeRbund
CD59			Increase	CummeRbund
Pattern recognition				
CLEC2G			Decrease	CummeRbund
CLEC6A			Freestyle	edgeR
DHX58	48			CummeRbund
NLRC3**		2,21		CummeRbund
NLRC3**			Various patterns	edgeR
NLRP12		4		edgeR
NLRP3			Increase	CummeRbund
Acute phase				
APOB		2		
APOD	4	21		CummeRbund
APOEB	4	21		CummeRbund
APOH	42			CummeRbund
Apolipoproteins	21,42	2		CummeRbund
F2			Decrease	CummeRbund
FG (A, B, G)			Increase	CummeRbund
HPX	21,42			CummeRbund
PLG		2		CummeRbund
TF			Increase	CummeRbund
Interferon				
IFNg*			NS	
STAT1			Increase	CummeRbund

Cell recruitment				
CXCL1		42		CummeRbund
CXCL12		42		CummeRbund
CXCL8		42		CummeRbund
MCP-1 (CCL2)			Decrease	edgeR
Apoptosis				
CASP1			Freestyle	CummeRbund
CASP6		21		CummeRbund
CASP8			Internal min	CummeRbund
CTSC			Increase	CummeRbund
Animicrobials				
IL4L1 (LAO)		42		CummeRbund
LYG2			Increase	edgeR
Phagosome and intracellular transport				
ACTB			Internal max	CummeRbund
ATG10			Freestyle	edgeR
ATP6V0C			Increase	CummeRbund
ATP6V0E2			Increase	CummeRbund
CALR			Freestyle	edgeR
EEA1			Internal min	edgeR
ITGA5			Increase	CummeRbund
ITGAM (CD11B)			Freestyle	edgeR
RAB34		21		CummeRbund
SEC61A1			Decrease	CummeRbund
SEC61G			Freestyle	CummeRbund
THBS1		42		CummeRbund
TUBA1C	21,42			CummeRbund
TUBA4			Decrease	CummeRbund
TUBB	21			CummeRbund
TUBB2A	21,42			CummeRbund
TUBB2B			Decrease	CummeRbund
Immunoglobulins				
IGKC			Increase	CummeRbund
IGKC			Internal min	CummeRbund
IGLC6	2			CummeRbund
MHCI related				
B2M			Increase	CummeRbund
Cytotoxic and T-cell related				
BCL2		42		CummeRbund
CD80			Freestyle	edgeR
MCL1			Internal min	CummeRbund
PRF1	4,21	42		CummeRbund
RANKL		42		CummeRbund

(TNFSF11)				
NK-cell markers				
B3GAT1 (CD57)			Internal max	CummeRbund
BNIP3L			Increase	CummeRbund
KIT (CD117)			Increase	edgeR
NCAM1 (CD56)		42		CummeRbund
TRAIL (TNFSF10)		42		CummeRbund
Heat shock				
HSPA1A (HSP70)	2,4	42		CummeRbund
Interferon regulatory factors				
IFI27			Decrease	CummeRbund
IFI44	21,42			CummeRbund
IRF2			Freestyle	edgeR
IRF3			Internal min	edgeR
IRF4			Internal min	CummeRbund
Metabolism				
HIF1N			Increase time	edgeR
MTOR			Freestyle	edgeR
Macrophage				
ATF4			Freestyle	edgeR
ATF4			Internal min	edgeR
Interleukin				
IL34		42		CummeRbund
Proteasome				
PSMB1			Decrease	CummeRbund
PSMB6			Decrease	CummeRbund
PSMC6			Decrease	CummeRbund
PSMD11			Decrease	CummeRbund
PSMD6			Decrease	CummeRbund
PSMD7			Decrease	CummeRbund
PSMD8			Decrease	edgeR

583

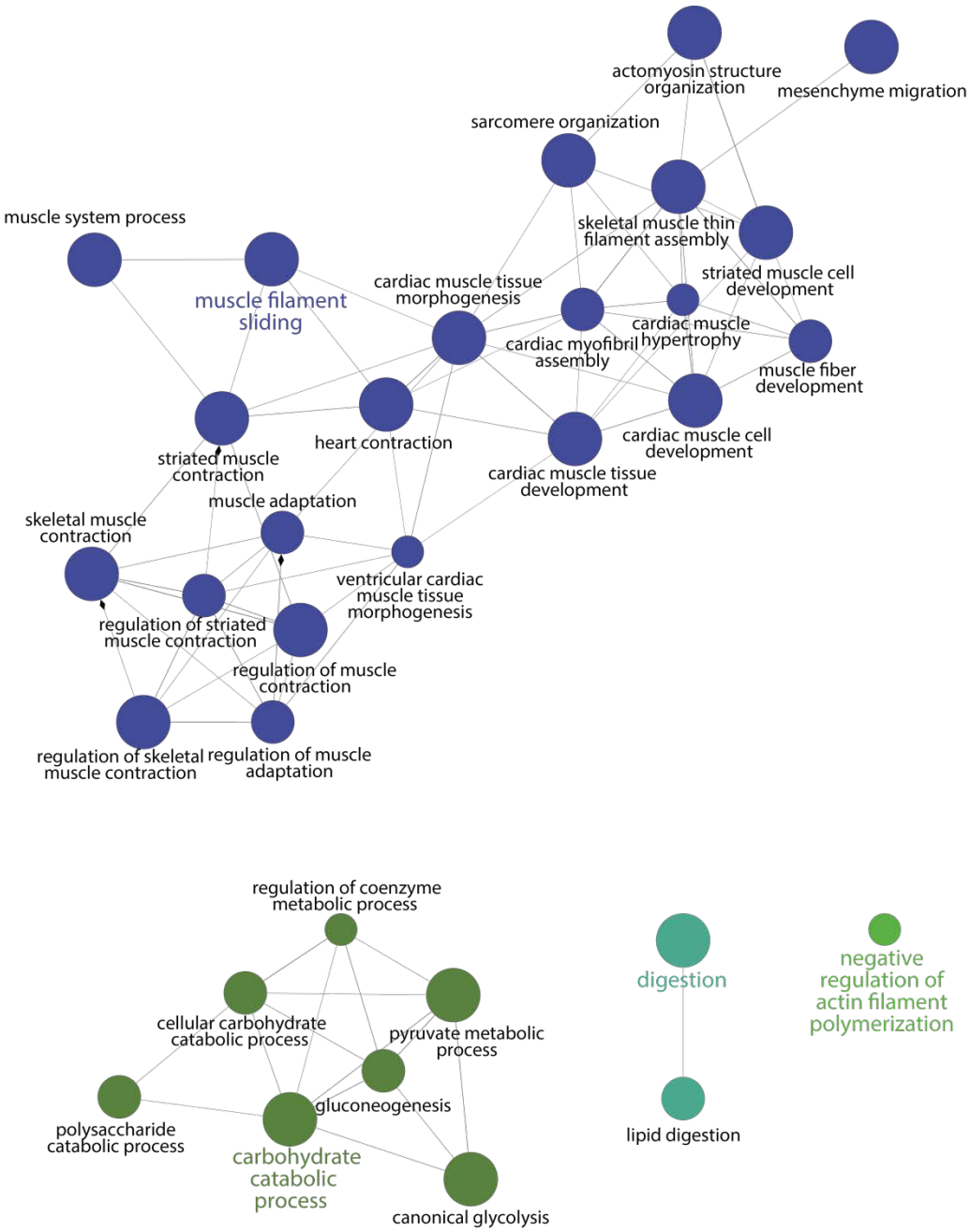
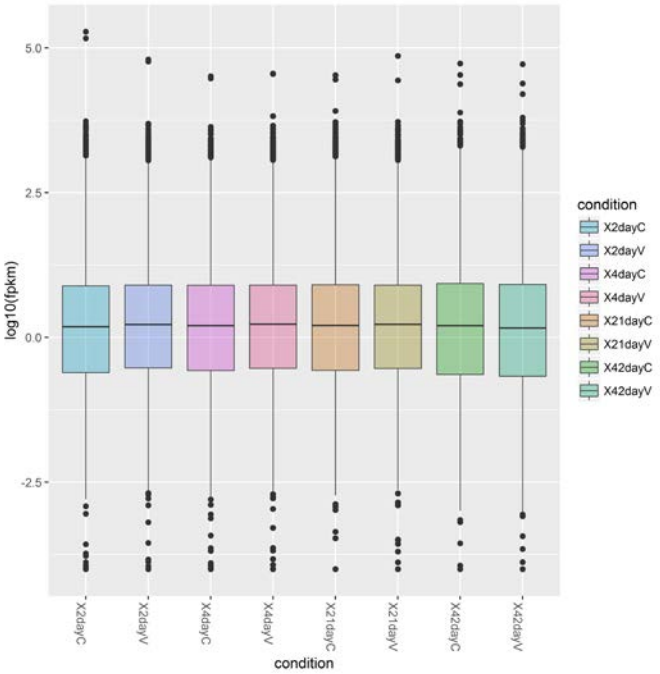
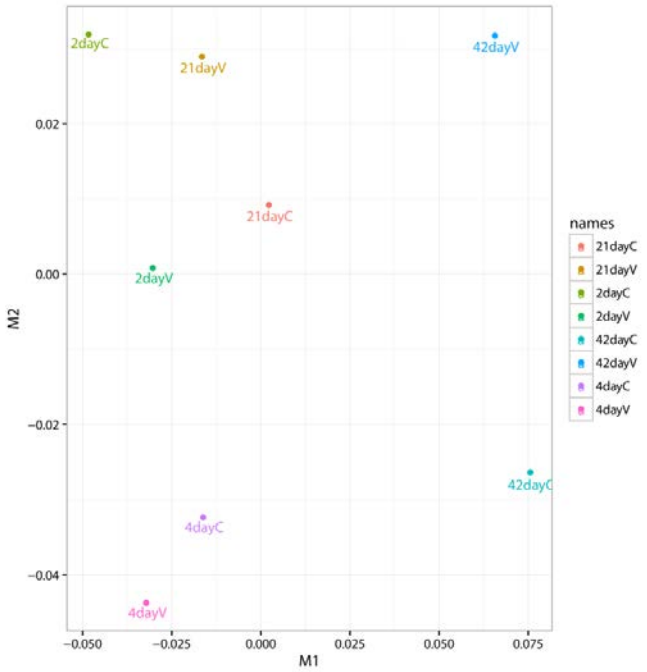


Figure 1 GO:term clustering derived from all up-regulated genes at day 4 post vaccination significantly different from control reported by the TCC pipeline. The network was generated using ClueGO in Cytoscape with p-value cufoff = 0.05 and focusing on biological and immunological systems. In addition fusion of related GO:terms was applied to improve readability and thus the reported terms deviate some from table 2.



591

592 **Supplementary figure 1** Box plot displaying the overall expression concatenated for all
593 biological replicates for each sample. Made using CummeRbund.



594

595 **Supplementary figure 2** MDS plot of all samples with concatenated biological replicates
596 Made using CummeRbund.