The genome sequence of Atlantic cod reveals a unique immune system

Atlantic cod (Gadus morhua) is a large, cold-adapted teleost that sustains long-standing commercial fisheries and incipient aquaculture1,2. Here we present the genome sequence of Atlantic cod, showing evidence for complex thermal adaptations in its haemoglobin gene cluster and an unusual immune architecture compared to other sequenced vertebrates. The genome assembly was obtained exclusively by 454 sequencing of shotgun and paired-end libraries, and automated annotation identified 22,154 genes. The major histocompatibility complex (MHC) II is a conserved feature of the adaptive immune system of jawed vertebrates3,4, but we show that Atlantic cod has lost the genes for MHC II, CD4 and invariant chain (II) that are essential for the function of this pathway. Nevertheless, Atlantic cod is not exceptionally susceptible to disease under natural conditions5. We find a highly expanded number of MHC I genes and a unique composition of its Toll-like receptor (TLR) families. This indicates how the Atlantic cod immune system has evolved compensatory mechanisms in both adaptive and innate immunity in the absence of MHC II. These observations affect fundamental assumptions about the evolution of the adaptive immune system and its components in vertebrates.

We sequenced the genome of a heterozygous male Atlantic cod (NEAC_001, Supplementary Notes 1 and 2), applying a whole-genome shotgun approach to 40× coverage (estimated genome size of 830 megabases (Mb), Supplementary Note 4 and Supplementary Fig. 2) using 454 technology (Supplementary Note 3). Two programs (Newbler and Celera), Supplementary Notes 5 and 6) produced assemblies with short contigs, yet with scaffolds of comparable size to those of Sanger-sequenced teleost genomes (Supplementary Note 10 and Supplementary Fig. 8). Although fragmentation due to short tandem repeats is difficult to address (Supplementary Note 7), we resolved numerous gaps attributable to heterozygosity (Supplementary Note 8).

The assemblies differ in scaffold and contig length (Table 1), although their scaffolds align to a large extent (Supplementary Note 9 and Supplementary Fig. 9). We obtained about one million single nucleotide polymorphisms (SNPs) by mapping 454 and Illumina reads from the sequenced individual to the Newbler assembly (Supplementary Note 11). Both assemblies cover more than 98% of the reads from an extensive transcriptome data set, indicating that the proteome is well represented (Supplementary Note 13). The assemblies are consistent with four independently assembled bacterial artificial chromosome (BAC) insert clones (Supplementary Note 14 and Supplementary Fig. 9), and with the expected insert size of paired BAC-end reads (Supplementary Note 15 and Supplementary Fig. 10).

A standard annotation approach based on protein evidence was complemented by a whole-genome alignment of the Atlantic cod with the stickleback (Gasterosteus aculeatus), after repeat-masking 25.4% of the Newbler assembly (Supplementary Note 16 and Supplementary Table 6). In this way, 17,920 out of 20,787 protein-coding stickleback genes were mapped onto reorganized scaffolds (Supplementary Note 17). Additional protein-coding genes, pseudogenes and non-coding RNAs were annotated using the standard Ensembl pipeline. These approaches resulted in a final gene set of 22,154 genes (Supplementary Table 7). Comparative analysis of gene ontology classes indicates that the major functional pathways are represented in the annotated gene set (Supplementary Note 18 and Supplementary Fig. 11). We anchored 332 Mb of the Newbler assembly to 23 linkage groups of an existing Atlantic cod linkage map using 924 SNPs (Supplementary Note 19 and Supplementary Table 8). These linkage groups have distinct orthology to chromosomes of other teleosts, on the basis of the number of co-occurring genes, showing that the whole-genome shotgun assembly reflects the expected chromosomal anatomy (Fig. 1, Supplementary Note 20 and Supplementary Table 9).

**Table 1 | Assembly statistics**

<table>
<thead>
<tr>
<th></th>
<th>Newbler</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Bases (Mb)</td>
<td>N50L (bp)</td>
<td>N50 (bp)</td>
<td>ML (bp)</td>
</tr>
<tr>
<td>Contigs%</td>
<td>284,239</td>
<td>536</td>
<td>2,778</td>
<td>50,231</td>
<td>76,504</td>
</tr>
<tr>
<td>Scaffolds</td>
<td>6,467</td>
<td>611</td>
<td>687,709</td>
<td>218</td>
<td>4,999,318</td>
</tr>
<tr>
<td>Entire assembly</td>
<td>157,887</td>
<td>753</td>
<td>459,495</td>
<td>344</td>
<td>4,999,318</td>
</tr>
<tr>
<td>Celera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contigs%</td>
<td>135,024</td>
<td>555</td>
<td>7,128</td>
<td>19,938</td>
<td>117,463</td>
</tr>
<tr>
<td>Scaffolds</td>
<td>3,832</td>
<td>608</td>
<td>488,312</td>
<td>373</td>
<td>2,810,583</td>
</tr>
<tr>
<td>Entire assembly</td>
<td>17,039</td>
<td>629</td>
<td>469,840</td>
<td>395</td>
<td>2,810,583</td>
</tr>
</tbody>
</table>

1Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, PO Box 1066, Blindern, N-0316 Oslo, Norway. 2Department of Molecular Biosciences, Centre for Immune Regulation, University of Oslo, Blindern, N-0316 Oslo, Norway. 3Bioinformatics Core Facility, Institute for Medical Informatics, Oslo University Hospital, Montebello, N-0310 Oslo, Norway. 4Department of Informatics, University of Bergen, N-5020 Bergen, Norway. 5Department of Natural Sciences and Technology, Hedmark University College, P.O. Box 4010, Bedriftsentret, N-2306 Hamar, Norway. 6Department of Animal and Aquacultural Sciences, University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway. 7Department of Biology, Centre for Geobiology, University of Bergen, N-5020 Bergen, Norway. 8Computational Biology Unit, Uni Computing, Uni Research AS, N-5020 Bergen, Norway. 9Illumina Life Sciences, 15 Commercial Street, Branford, Connecticut 06405, USA. 10Welcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. 11Nofima Marine, P.O. Box 5010, N-1430 Ås, Norway. 12Department of Marine Biology, Faculty of Marine Sciences, University of Bergen, N-5020 Bergen, Norway. 13Centre for Integrative Genetics, Norwegian University of Life Sciences, PO Box 5003, 1432 Ås, Norway. 14Faculty of Biosciences and Aquaculture, University of Nordland, N-8049 Bodø, Norway. 15Department of Pharmaceutical Biosciences, School of Pharmacy, University of Oslo, P.O. Box 1068, Blindern, N-0316 Oslo, Norway. 16Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, D-14195 Berlin-Dahlem, Germany. 17Institute for Basic Sciences and Aquatic Medicine, School of Veterinary Sciences, N-0033 Oslo, Norway. 18Department of Medical Biology, Faculty of Health Sciences, University of Tromsø, N-9037 Tromsø, Norway. 19Department of Biology, PO Box 7803, University of Bergen, N-5020 Bergen, Norway.

©2011 Macmillan Publishers Limited. All rights reserved
Well-studied haemoglobin polymorphisms in Atlantic cod are indicative of functional molecular adaptation to thermal variation\(^{10-12}\). The genome contains nine α- and β-globin genes that are organized in two unlinked clusters, \(β\)-\(ξ\)1–\(β\)-\(α\)4 and \(β\)-\(β\)-\(x\)2–\(α\)-\(β\)2 (refs \(13, 14\)). We discovered an indel polymorphism of 73 base pairs (bp) in the intergenic promoter region of the \(α\)-\(β\) globin pair (Fig. 2a and Supplementary Note 21). This promoter polymorphism occurs in highly significant linkage disequilibrium with two known polymorphic sites in the \(β\)1 gene, the Val55Met and Ala62Lys substitutions\(^1\), in eight Atlantic cod populations (Supplementary Note 22 and Supplementary Fig. 12). In fact, in the three most northern Atlantic populations and in both Baltic populations, the cod \(β\)1-globin gene predominantly occurs as a single homoygous genotype consisting of the long promoter and the Val55–Ala62 allele (Supplementary Table 10). By placing the two promoter variants in front of a luciferase reporter gene and transfecting the constructs into salmon kidney cells (Supplementary Note 23), we found that temperature and promoter type have a significant interaction effect (generalized linear model, \(F_2, 36 = 7.85, P = 0.007\), Fig. 2b) and that the long promoter has twofold higher transcriptional activity compared to the short promoter at 15°C and 20°C. Increased globin synthesis of the Val55–Ala62 allele would compensate for its lower oxygen affinity\(^{10,11}\) at high temperatures. Thus, the promoter polymorphism provides a molecular compensatory mechanism that helps to maintain the total oxygen-carrying capacity\(^{15}\). The tight linkage between the two types of polymorphism provides a compelling example of the coevolution of structural and regulatory adaptation, and highlights the relationship between temperature and functional molecular variation in the haemoglobin system\(^{16}\).

The Atlantic cod immune system has unusual properties that set it apart from that of other teleosts: high levels of IgM\(^1\), a minimal antibody response after pathogen exposure\(^{5, 17, 18}\) and abundant phagocytic neutrophils in the peripheral blood\(^{19, 20}\). Despite speculation, the exact causes for these differences remain unknown\(^1\). We found that most genes involved in the vertebrate immune response are present in Atlantic cod (Supplementary Note 24, Supplementary Fig. 13 and Supplementary Table 11). Nevertheless, we did not find genes for the MHCII isoforms, their assembly and trafficking chaperone \(\Gamma\)\(^2\) and the MHCII-interacting protein \(\Gamma\), which is essential for helper T-cell activation. By comparing a comprehensive set of vertebrate MHCII, CD4 and II sequences to the genome assemblies and all unassembled 454 and Illumina sequencing reads (a data set of about 49.5 gigabases), we detected a truncated pseudogene for CD4 (Supplementary Note 25), which is located in a region of conserved synteny (Supplementary Note 27 and Supplementary Fig. 18). No traces of MHCII and II were found in syntenous regions (Supplementary Note 27 and Supplementary Figs 16, 17, 19 and 20) and quantitative PCR (qPCR) targeting a conserved domain in MHCII did not amplify the target sequence (Supplementary Note 26 and Supplementary Fig. 15). The absence of MHCII and II, and the pseudogenic nature of CD4, show that Atlantic cod has lost the function of the classical pathway for adaptive immunity against bacterial and parasitic infections. Nevertheless, Atlantic cod deals adequately with its prevailing pathogen load in its natural ecological settings\(^1\). Previous transcriptional (complementary DNA) studies in Atlantic cod have indicated an expansion of the number of MHC I loci\(^{22, 23}\) and quantitative PCR (qPCR) targeting a conserved domain in MHCII did not amplify the target sequence (Supplementary Note 27 and Supplementary Fig. 18). No traces of MHCII and II were found in syntenous regions (Supplementary Note 27 and Supplementary Figs 16, 17, 19 and 20) and quantitative PCR (qPCR) targeting a conserved domain in MHCII did not amplify the target sequence (Supplementary Note 26 and Supplementary Fig. 15). The absence of MHCII and II, and the pseudogenic nature of CD4, show that Atlantic cod has lost the function of the classical pathway for adaptive immunity against bacterial and parasitic infections. Nevertheless, Atlantic cod deals adequately with its prevailing pathogen load in its natural ecological settings\(^1\). Previous transcriptional (complementary DNA) studies in Atlantic cod have indicated an expansion of the number of MHC I loci\(^{22, 23}\) and quantitative PCR (qPCR) targeting a conserved domain in MHCII did not amplify the target sequence (Supplementary Note 27 and Supplementary Fig. 18). No traces of MHCII and II were found in syntenous regions (Supplementary Note 27 and Supplementary Figs 16, 17, 19 and 20) and quantitative PCR (qPCR) targeting a conserved domain in MHCII did not amplify the target sequence (Supplementary Note 26 and Supplementary Fig. 15). The absence of MHCII and II, and the pseudogenic nature of CD4, show that Atlantic cod has lost the function of the classical pathway for adaptive immunity against bacterial and parasitic infections. Nevertheless, Atlantic cod deals adequately with its prevailing pathogen load in its natural ecological settings\(^1\). Previous transcriptional (complementary DNA) studies in Atlantic cod have indicated an expansion of the number of MHC I loci\(^{22, 23}\) and quantitative PCR (qPCR) targeting a conserved domain in MHCII did not amplify the target sequence (Supplementary Note 27 and Supplementary Fig. 18). No traces of MHCII and II were found in syntenous regions (Supplementary Note 27 and Supplementary Figs 16, 17, 19 and 20) and quantitative PCR (qPCR) targeting a conserved domain in MHCII did not amplify the target sequence (Supplementary Note 26 and Supplementary Fig. 15). The absence of MHCII and II, and the pseudogenic nature of CD4, show that Atlantic cod has lost the function of the classical pathway for adaptive immunity against bacterial and parasitic infections. Nevertheless, Atlantic cod deals adequately with its prevailing pathogen load in its natural ecological settings\(^1\).
pathway—can initiate immune responses in mammals. The cross-presentation pathway represents a structural and cellular modification of the MHC-I machinery that allows activation of CD8⁺ T cells upon bacterial infection. The cytokine gene profile of Atlantic cod (Supplementary Table 11) supports the possibility of generating different subsets of CD8⁺ T cells that either provide direct protection or regulate other immune cells, and thus compensate for the loss of CD4⁺ T cells.

In addition to the MHC-I expansion, we found an unusual composition of the highly conserved TLR families that have a fundamental role in the innate immune response and the initial detection of pathogens. Teleost TLR-encoding genes occur in well-supported phylogenetic clusters, most of which share functional properties with mammalian orthologues, although some are fish-specific. The Atlantic cod TLR genes form monophyletic groups within the known teleost functional groups (Fig. 4, Supplementary Note 30 and Supplementary Fig. 22). Genes for several TLRs that recognize bacterial surface antigens (TLR1, TLR2 and TLR5) are, however, absent, leaving only the teleost-specific TLR4 and TLR18 as members of the TLR1 family in Atlantic cod. Moreover, several families of TLRs that recognize nucleic acids (TLR7, TLR8, TLR9 and TLR22) have markedly expanded, resulting in the highest number of TLRs found in a teleost so far. This TLR repertoire indicates that the Atlantic cod immune system relies relatively heavily on nucleic-acid-detecting TLRs to recognize bacterial pathogens. Notably, the gene expansion of TLR9 coincides with an expansion of interleukin-8 genes (IL-8, Supplementary Table 11). IL-8 is an important chemokine in the innate immune response and is directly induced by TLR9 in human neutrophils. The corresponding expansions of IL-8 and TLR9 indicate that this signalling cascade is particularly important in Atlantic cod.

The loss of MHC-II function and lack of a CD4⁺ T-cell response represent a fundamental change in how the adaptive immune system is initiated and regulated in Atlantic cod. The marked expansion of MHC-I genes and unusual TLR composition signify a shift of its immune system in handling microbial pathogens. An expanded MHC-I repertoire in the presence of a non-polymorphic MHC-II is found in an evolutionarily-distant vertebrate, the axolotl (Ambystoma mexicanum). These observations indicate that anomalous immune systems (possibly analogous to that of Atlantic cod) have evolved independently. Additionally, we did not recover evidence for expressed MHC-II, CD4 and li in the transcriptomes of three other gadoids, indicating that the unusual immune system is a derived characteristic of the gadoid lineage (Supplementary Tables 18 and 19).

We have provided the first annotated genome of a species that supports extensive fisheries and is on the verge of becoming an important aquaculture species. This work provides a major foundation for addressing key issues related to the management of natural Atlantic cod populations, such as the concept of fisheries-induced evolution, which dictates that selective harvesting can change the evolutionary trajectory of major life-history traits of natural populations. Moreover, our novel findings regarding the immune system will allow for more targeted vaccine development, aiding disease management and the process of domestication of Atlantic cod. These findings...
change fundamental assumptions regarding the evolution of the vertebrate immune system.

METHODS SUMMARY

Detailed methods on the sequencing and assembly of data from genomic and transcriptomic origins; annotation, synteny analyses, transfection experiments, bioinformatic analyses and phylogenetic analyses presented in this manuscript are described in the Supplementary Information.

Received 11 February; accepted 28 June 2011.

Published online 10 August 2011.