MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 2130–2143 doi: 10.1111/mec.13592

FROM THE COVER

Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod

TINA GRACELINE KIRUBAKARAN,* HARALD GROVE,* MATTHEW P. KENT,* SIMEN R. SANDVE,* MATTHEW BARANSKI,† TORFINN NOME,* MARIA CRISTINA DE ROSA,‡ BENEDETTA RIGHINO,‡ TORILD JOHANSEN,§ HÅKON OTTERÅ,§ ANNA SONESSON,† SIGBJØRN LIEN* and ØIVIND ANDERSEN*†

*Centre for Integrative Genetics (CIGENE), Department of Animal and Aquaculture Sciences (IHA), Norwegian University of Life Sciences (NMBU), PO Box 5003, Ås N-1430, Norway, †Nofima, PO Box 5010, Ås N-1430, Norway, ‡Institute of Chemistry of Molecular Recognition – CNR and Institute of Biochemistry and Clinical Biochemistry, Catholic University of Rome, Rome 00168, Italy, §Institute of Marine Research, PO Box 6404, Tromsø N-9294, Norway

Abstract

Atlantic cod is composed of multiple migratory and stationary populations widely distributed in the North Atlantic Ocean. The Northeast Arctic cod (NEAC) population in the Barents Sea undertakes annual spawning migrations to the northern Norwegian coast. Although spawning occurs sympatrically with the stationary Norwegian coastal cod (NCC), phenotypic and genetic differences between NEAC and NCC are maintained. In this study, we resolve the enigma by revealing the mechanisms underlying these differences. Extended linkage disequilibrium (LD) and population divergence were demonstrated in a 17.4-Mb region on linkage group 1 (LG1) based on genotypes of 494 SNPs from 192 parents of farmed families of NEAC, NCC or NEACxNCC crosses. Linkage analyses revealed two adjacent inversions within this region that repress meiotic recombination in NEACxNCC crosses. We identified a NEAC-specific haplotype consisting of 186 SNPs that was fixed in NEAC sampled from the Barents Sea, but segregating under Hardy–Weinberg equilibrium in eight NCC stocks. Comparative genomic analyses determine the NEAC configuration of the inversions to be the derived state and date it to ~1.6–2.0 Mya. The haplotype block harbours 763 genes, including candidates regulating swim bladder pressure, haem synthesis and skeletal muscle organization conferring adaptation to long-distance migrations and vertical movements down to large depths. Our results suggest that the migratory ecotype experiences strong directional selection for the two adjacent inversions on LG1. Despite interbreeding between NEAC and NCC, the inversions are maintaining genetic differentiation, and we hypothesize the co-occurrence of multiple adaptive alleles forming a 'supergene' in the NEAC population.

Keywords: chromosomal inversion, gene flow, local adaptation, recombination, supergene, swim bladder

Received 30 November 2015; revision received 8 February 2016; accepted 17 February 2016

Introduction

Adaptive divergence among populations and ultimately speciation is one of the most central subjects in evolutionary biology. Marine fishes are typically associated with shallow population structure and are highly dispersed in various ecological habitats, thus providing excellent models for studying interactions between the homogenizing effect of gene flow and the diversifying effect of selection (Hauser & Carvalho 2008; Nielsen *et al*. 2009). While reduced genetic exchange between

Correspondence: Øivind Andersen, Fax: +47 77629100; E-mail: oivind.andersen@nofima.no

sympatric cryptic species has mainly been explained by various reproductive barriers (Palumbi 1994; Monteiro *et al.* 2012; Shen *et al.* 2015), the adaptive significance of chromosome rearrangements by suppressing recombination between heterozygotes has been evidenced in many plant and animal studies, including coral reef fish species and in marine and freshwater ecotypes of three-spined stickleback (*Gasterosteus aculeatus*) (Jones *et al.* 2012; Martinez *et al.* 2015).

Atlantic cod are widely distributed on the continental shelves and banks on both sides of the North Atlantic Ocean and represent the main demersal fish resource in these regions. The success of this highly exploited fish seems to be related to the different life history strategies of the multiple migratory and stationary populations, but careful management is required as several stocks have been dramatically reduced as a result of overfishing, climate change and pollution (Myers *et al.* 1997; Christensen *et al.* 2003; Mackenzie *et al.* 2004; Robichaud & Rose 2004). Cod fishery dates back to the tenth century A.D. when Vikings used dried Skrei (Old Norse *skríða* means wandering) as a source of nutrition and currency along the European trade routes. Today, Skrei are synonymous with the large Northeast Arctic cod (NEAC) population, which feeds in the Barents Sea and near Svalbard, but the adults undertake annual long-distance migrations to and from the spawning banks along the coast of North Norway, mainly offshore the Lofoten Archipelago (Bergstad & Dragesun 1987; Sundby & Nakken 2008; Ottersen *et al.* 2014). During foraging and spawning migrations, NEAC perform vertical movements down to depths of about 500 m with frequent descending and ascending swimming spanning up to 250 m (Godø & Michalsen 2000; Stensholt 2001). In contrast, the stationary Norwegian coastal cod (NCC) live in shallow coastal waters and fjords throughout the year and generally migrate only short distances at depths down to about 100 m (Hobson *et al.* 2007; Michalsen *et al.* 2014). The vertical divergence between NEAC and NCC is apparent at the 0-group stage when juveniles settle in deep and shallow water, respectively, in northern Norwegian fjords (Løken *et al.* 1994; Fevolden *et al.* 2012). In Iceland, similar ecotypes are represented by the frontal (migratory) and coastal (nonmigratory) populations, which exploit different habitats at depths of 200–600 m and <200 m, respectively (Pálsson & Thorsteinsson 2003; Pampoulie *et al.* 2008; Grabowski *et al.* 2011).

The population structure of NEAC and NCC and the interactions between gene flow and natural selection have been a controversial subject since the 1930s. Although NEAC and NCC occur sympatrically on local spawning grounds with potentially high levels of gene flow, phenotypic and genetic differences are maintained between the two populations (Rollefsen 1933, 1954; Jakobsen 1987; Løken & Pedersen 1996; Nordeide & Pettersen 1998; Nordeide *et al.* 2011). Whereas phenotypic traits such as otolith morphology and vertebrae numbers are likely influences by temperature and hydrostatic pressure at the nursery grounds, strongly divergent allele frequencies have been found in several candidate genes for adaptation to different ecosystems, including pantophysin (*Pan* I), haemoglobin and rhodopsin (Møller 1966, 1968; Fevolden & Pogson 1997; Pogson 2001; Andersen *et al.* 2015; Pampoulie *et al.* 2015). Studying the genetic diversity in Atlantic cod along the Norwegian coast, Møller (1966, 1968, 1969) concluded that NEAC and NCC form two genetically separated populations or sibling species. The extreme divergence between NEAC and NCC at the *Pan* I locus has later been explained by differences in breeding structure, as selection alone would be insufficient to cause the observed levels of genetic differentiation (Fevolden & Sarvas 2001; Sarvas & Fevolden 2005; Westgaard & Fevolden 2007). The Icelandic frontal and coastal ecotypes were consistently proposed to be cryptic species within the Atlantic cod complex that have adapted to environmental factors in shallow and deep waters (Halldórsdóttir & Árnason 2016). Differences in courtship, spawning behaviour or spawning depths could hinder interbreeding between the populations or siblings (Hutchings *et al.* 1999; Nordeide & Folstad 2000; Grabowski *et al.* 2011). However, analysis of the mitochondrial genome revealed no reproductive isolation between NEAC and NCC (Árnason & Pálsson 1996; Karlsen *et al.* 2014), supporting the hypothesis that there is genetic communication between all cod stocks in the North Atlantic (Mork & Sundnes 1985). Accordingly, the genetic differences between NEAC and NCC were suggested to be produced by contemporary selection acting on settling cohorts (Williams 1975; Mork & Sundnes 1985).

The genetic differentiation of NEAC and NCC was recently shown to be uniquely associated with a large genomic region potentially harbouring hundreds of genes on linkage group 1 (LG1) (Hemmer-Hansen *et al.* 2013; Karlsen *et al.* 2013). The underlying mechanism was not unravelled in these studies, but a combination of divergence hitchhiking and chromosomal rearrangement was proposed to be responsible for the genomic island of divergence. We have finally resolved the enigma behind the strong genetic divergence between migratory NEAC and stationary NCC by the identification of a double inversion on LG1 that repress recombination within heterozygotes preventing introgression between cosegregating haplotypes.

Materials and methods

Fish material and DNA extraction

Wild cod were collected from 14 locations ranging from the Irish Sea in the south to the Barents Sea in the north (see Fig. 2). On average, 48 samples were collected from each location. To obtain a representational sampling of NEAC, we collected cod from two locations in the Barents Sea. Farmed cod were sampled from 88 families of the National cod breeding programme maintained by Nofima in Tromsø, Norway, and from eight families of the CODBIOBANK at the Institute of Marine Research in Bergen, Norway.

Totally, 104 cod from the National cod breeding programme were selected for sequencing comprising 50 fish of NEAC origin, 11 fish of NCC origin and 43 fish offspring of NEAC × NCC crosses. The sequenced fish belonged to year classes 2005 (P) and 2006 (F1) and represented the second generation of cod produced in captivity. The original broodstock in the base population were sampled from different geographical areas along the Norwegian coast and were assigned to the NCC and NEAC populations based on sampling locations and the *Pan* I^A and I^B alleles (Fevolden & Pogson 1997; Bangera *et al.* 2013). The Greenland cod (*Gadus macrocephalus ogac*) used to date the inversion was sampled at the Uummannaq Island, Northwest Greenland.

DNA was extracted using either a DNeasy kit from Qiagen (Hilden, Germany) according to the manufacturer's instructions or a high salt precipitation method (http://www.liv.ac.uk/~kempsj/IsolationofDNA.pdf). DNA quality was assessed by electrophoresis on 1% agarose gel to estimate the proportion of high molecular weight (HMW) DNA, and low-quality samples with negligible levels of HMW DNA were excluded from analysis. DNA concentration was assessed fluorometrically using Qubit technologies (Thermo Fisher Scientific, Carlsbad, USA).

Genotyping

Farmed (n = 2951) and wild fish (n = 959) were genotyped for 10 913 SNPs using an Illumina custom Infinium II SNP array (M. Kent, T. G. Kirubakaran, P. R. Berg, M. Baranski, G, Dahle, K. S. Jakobsen, S. Jentoft, T. Johansen, L. Nederbragdt, T, Nome, B. Star & S. Lien, in prep) according to the manufacturer's instructions (Illumina, San Diego, USA). Poorly performing samples displaying call rates below 0.9 were excluded from the analysis. Genotype data was preprocessed by removing low (<0.05) MAF (minor allele frequency) SNPs, and Mendelian errors were set to missing and imputed along with any other failed genotypes using BEAGLE v4 (Browning & Browning 2007). Wild populations were phased using SHAPEIT v2 (https://mathgen.stats.ox.ac.uk/shapeit), and the family material was accurately phased using linkage information. Phased data for 192 parents were used to estimate linkage disequilibrium (LD) between SNPs using HAPLOVIEW 4.2 (Barrett *et al.* 2005). All NEAC samples from the Barents Sea were homozygous for a haplotype consisting of 186 SNPs from the SNP array (see Table S1, Supporting information), and the wild fish were assigned to NEAC, NCC or a cross using this NEAC haplotype.

Linkage mapping and inversion detection

The construction of linkage maps for cod using 12K SNP array is described in detail elsewhere (H. Grove, T. G. Kirubakaran, M. Kent, M. Baranski, T. Nome, S. Sandve, P. R. Berg, M. Sodeland, G. Dahle, A. Sonesson, T. Johansen, Ø. Andersen & S. Lien, in prep). The analyses are mainly based on the CRIMAP linkage program (Green *et al.* 1990) and begin with performing two-point linkage in CRIMAP to sort SNPs into linkage groups. In the present study, we used 494 SNPs mapping to LG1 for the construction of separate multipoint linkage maps for pure NEAC, pure NCC and NEAC × NCC crosses. The sorting of family material for these analyses was determined by haplotyping parents using the 186 SNP set described above.

SNPs on the cod 12K SNP array were carefully chosen to tag as many contigs as possible (M. Kent, T. G. Kirubakaran, P. R. Berg, M. Baranski, G. Dahle, K. S. Jakobsen, S. Jentoft, T. Johansen, L. Nederbragdt, T, Nome, B. Star & S. Lien, in prep) and are well distributed along the linkage groups, thereby also forming a good foundation for building a chromosome sequence for LG1. Scaffolds from two draft assemblies, containing at least one SNP from the linkage map, were selected and used for the construction of the chromosome files. Erroneous scaffolds containing SNPs from more than one LG were broken between conflicting SNP positions. Overlapping scaffolds were identified by comparing SNPs mapping to both assemblies and were merged using coordinates from alignment with LASTZ (Harris 2007), resulting in a total of 40 scaffolds that were used to build the final chromosome sequence. Subsequently, linkage maps were then updated to take into account the more precise SNP order given by individual scaffolds. Finally, all scaffolds were oriented, ordered and concatenated into a new chromosome sequence based on information from the linkage map. The size of the final chromosome sequence for LG1 was 29 521 491 bp.

Sequencing and variant detection

Genomic DNA from the 104 breeding programme fish was prepared for sequencing using the Truseq Library

prep kit from Illumina (Illumina, San Diego, USA). Paired-end sequencing (2 × 100 nts) was carried out using an Illumina HiSeq 2000 instrument to generate approximately 10× coverage of the genome for each sample. Reads were processed using default parameters in TRIMMOMATIC version 0.32 (Bolger *et al.* 2014) before being aligned to the unmasked reference genome based on the NCC map described above using BOWTIE2 version 2.2.3 (Langmead & Salzberg 2012). Within-sample variant detection was performed using GATK HAPLOTYPECALLER version 2.8-1-g932cd3a (McKenna *et al.* 2010). SNPEFF version 4.0e (Cingolani *et al.* 2012) was used to annotate allelic variants. Individual variant calls with a quality score of <20 were excluded from further analysis, as were INDELs and genotypes with read depths below 6 or above 27. Also, variants not detected in >70% of the samples were removed across all samples.

Genomic DNA from a single Greenland cod was prepared for sequencing using a Nextera XT library preparation kit generating a library with an average size of 650 bp. Sequencing was performed using a MiSeq platform with V3 kit chemistry to generate 2 × 301-nt paired-end reads. A total of 18.7 M reads generated 11.2-Gb sequence data. Reads were mapped and variants detected as described above.

Pairwise LD (measured as r^2) for the whole linkage group was calculated based on 48 sequenced NEAC samples using PLINK v1.9 (https://www.cog-genomics.org/plink2) with MAF > 0.1 and HWE > 0.001.

Gene annotation

Gene models were built from multiple data sources including (i) approximately 3 million transcriptome reads (http://www.ncbi.nlm.nih.gov/sra?term=SRP013269) obtained from liver, egg, brain, head, kidney, hindgut, gonad and spleen, generated using GS-FLX 454 Titanium platform (Roche, Switzerland), (ii) ESTs from NCBI (n = 257 218), (iii) predicted RNAs (n = 1541, http://www.codgenome.no/data/ATLCOD1_ANN/) and (iv) approximately 35 million short read mRNA sequences from whole NEAC larvae at 12 and 35 days posthatching (Johnsen & Andersen 2012). To enable model building, short reads were mapped to the reference genome sequence using STAR (version 2.3.1z12) (Dobin *et al.* 2013), while long 454 transcriptome reads were mapped using GMAP (version 2014-07-28) (Thomas & Watanabe 2005) with '–no-chimeras' parameter in addition to default parameters. CUFFLINKS (version 2.2.1) (Trapnell *et al.* 2010) with '–multi-read-correct' parameter in addition to the default parameter assembled the aligned RNA-Seq reads and transcriptome reads into transcripts. Transcript models from RNA-Seq and 454 transcriptome were merged using cuffmerge.

Open reading frame (ORF) prediction was carried out using TransDecoder (http://transdecoder.github.io/) (Haas *et al.* 2013) using the pfamA and pfamB databases for homology searches (–search_pfam) and a minimum length of 30 amino acids for ORFs without pfam support (-m 30). In addition to the pfam homology evidence, we also performed BLASTP (evalue<1e-10) for all predicted proteins against zebrafish (*Danio rerio*) (v9.75) and three-spined stickleback (BROADS1.75) annotations downloaded from Ensembl. Only gene models with support from at least one type of homology search (pfam or BLASTP) were kept.

In total, we mapped 35 million RNA-Seq reads and 3.3 million 454 transcriptome sequences to the whole genome and used this to annotate LG1. A total of 2323 transcripts were left after merging transcript models using cuffmerge. Functional annotations of the transcripts were carried out using BLASTX against the SWISSPROT database. Results from TransDecoder and homology support filtering of putative protein-coding loci are shown in Table S2 (Supporting information).

Origin and dating of inversions

To determine whether NEAC or NCC represents the ancestral state of the inversions, we aligned LG1 sequences representing possible arrangements of the inversions with Northern pike (*Esox lucius*) and stickleback using LASTZ in gap-free mode requiring ≥75% identity and match-count filtering of 100 (Harris 2007).

Hierarchical clustering of the wild stocks was estimated based on genotypes from the SNP array using the R package SNPrelate (Zheng *et al.* 2012). Four linkage groups (LG1, LG2, LG7 and LG12) were excluded from this analysis due to the presence of extended LD (own data; Bradbury *et al.* 2010; Hemmer-Hansen *et al.* 2013). Reads generated from whole-genome sequencing of a single Greenland cod were compared with NEAC and NCC variant calls to identify a set of fixed sequence differences (FSD; single nucleotides fixed within populations) along LG1. FSD counts were then used to calculate pairwise differences among Greenland cod, NEAC and NCC. Under the assumption of a constant clock, we then estimated the NEAC-NCC divergence age relative to their divergence from Greenland cod by calculating the ratio between NEAC-NCC FSD distance and the mean FSD distance between Greenland cod-NEAC and Greenland cod-NCC (i.e. $FSD_{NEAC\text{-}NCC}/FSD_{mean\ (Greenland\ cod\text{-}NEAC,\ Greenland\ cod\text{-}NCC)}$) as shown in Fig. S1 (Supporting information).

Protein modelling

The carbonic anhydrase isoform 6 (*ca6*) was identified as a candidate gene within the double inversion because of

its key role in reducing pH levels in the gas gland of the swim bladder. Homology modelling was performed with the MODELLER software (Sali & Blundell 1993) to build the three-dimensional structure of the NEAC and NCC variants of Ca6 based on the crystal structure of human Ca6 as template (PDB code 3FE4, Pilka *et al.* 2012). The sequences were aligned using CLUSTALW, and identities between targets and template of 58% (NEAC) and 56% (NCC) allowed using the standard MODELLER protocol implemented in DISCOVERYSTUDIO v4.5 (Biovia). We ascertained that no other protein with a known related structure displayed a greater sequence similarity. The best of 50 models according to the PDF (Probability Density Function) score included in MODELLER was selected. The structures were inspected with PROCHECK (Laskowski *et al.* 1993) for inappropriate stereochemistry. Ramachandran maps of NEAC and NCC models revealed that they contained 91.7% of non-Gly-non-Pro residues in most favoured, 7.8% in additional allowed, 0.5% in generously allowed and 0.0% in disallowed regions. These models were further validated for their structure quality by Verify 3D available at http://services.mbi.ucla.edu/, and 95% of the residues of the modelled proteins showed a satisfactory 3D-1D score (>0.2). DISCOVERYSTUDIO v4.5 (Biovia) software was used to visualize the generated models.

Results

Linkage map and LD calculations

A genetic map describing 23 linkage groups in Atlantic cod (H. Grove, T. G. Kirubakaran M. Kent, M. Baranski, T. Nome, S. Sandve, P. R. Berg, M. Sodeland, G. Dahle, A. Sonesson, T. Johansen, Ø. Andersen & S. Lien, in prep.) was constructed by genotyping a large family material of 2739 individuals using a 12K SNP array (Kent *et al.* in prep). The map constructed for LG1 contained 494 SNPs (Fig. 1a; Table S1, Supporting information) and was used to integrate, order and orientate scaffolds from two draft cod assemblies into a cohesive chromosome sequence comprising 29.52 Mb. Accurately phased genotypes from 192 parents were used to estimate LD between SNPs and revealed a distinct block of extended LD from 10 to 27 Mb (Fig. 1c), embracing the *Pan* I locus located at 17.4 Mb. The parents were of known origin and classed as pure NEAC, pure NCC or NEAC × NCC crosses based on Pan I genotyping. Analyses of pure NEAC cod, captured from two locations in the Barents Sea, identified a single haplotype of 186 nonconsecutive SNPs that were homozygous in all individuals (Table S1, Supporting information). All

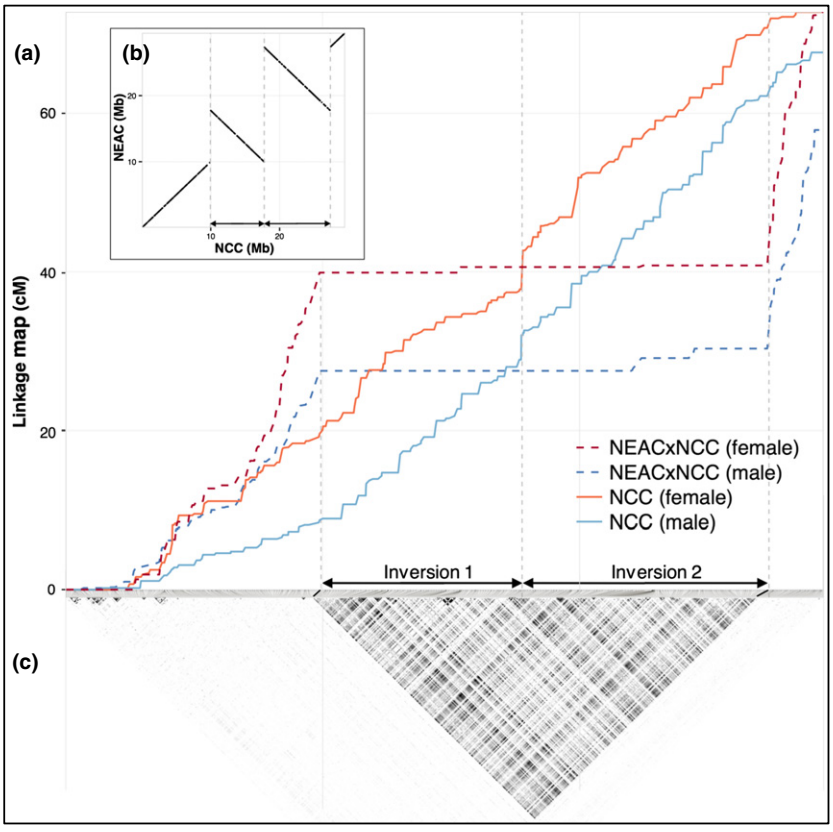

Fig. 1 (a) Linkage map for LG1 created separately for pure NCC and NEAC × NCC crosses. (b) Whole chromosome alignment between the NCC and NEAC sequence. (c) Pairwise LD calculated in 192 parents from the linkage families. Two large inversions inhibit recombination in NEACxNCC crosses corresponding to a region of extended LD on LG1.

NEAC × NCC crosses had one copy of this haplotype, while the NEAC haplotype was completely absent in pure NCC samples.

Because this distinct haplotype in NEAC indicated a substantial differentiation between NEAC and NCC cod, we constructed linkage maps separately for pure NEAC, pure NCC and NEAC × NCC crosses (Table S1, Supporting information). Pure NEAC and NCC showed typical recombination rates between SNPs along the length of LG1, but comparing the linkage maps disclosed a different SNP order within the block with extended LD. NEAC × NCC crosses displayed almost complete repression of recombination within this block, but showed elevated recombination outside the block (Fig. 1b). NEAC and NCC linkage maps were used to order and orient scaffolds to create specific assemblies of LG1 for these two ecotypes of Atlantic cod. Alignment of these sequences revealed the presence of two adjacent inversions of 9.55 and 7.82 Mb (Fig. 1a). Additional evidence for two inversions, in contrast to a single inversion, was found in the LD pattern of 48 whole-genome sequenced NEAC samples (Fig. S2, Supporting information). High LD was found between polymorphisms at 18 and 28 Mb in the NCC version of the assembly. In contrast, the proposed NEAC orientation of the inversions rearranges these two regions to be located close together.

Geographical distribution of NEAC haplotype

To validate whether the NEAC haplotype precisely identified the differing genotypes of migratory and stationary cod ecotypes, we analysed 48 adult cod captured in the Barents Sea and representing pure NEAC based on *Pan* I genotyping. All samples were homozygous for the 186 SNPs within the haplotype block, which endorses its utility as a tool to classify cod as NEAC, NCC or crosses. To explore the distribution of the NEAC haplotype, we tested adult individuals from 14 different localities across the Northeast Atlantic Ocean. In sharp contrast to the fixation in the two locations in the Barents Sea, frequencies of the NEAC haplotype were low or nonexistent in the southern stocks and in the White Sea, while intermediate frequencies were found among samples collected along the Norwegian coast north from Bergen (Porsanger, Senja, Verrabotn, Borgund) (Fig. 2a). The NEAC haplotype was in HWE in all the stocks examined. These results contrasts with the hierarchical clustering analysis performed on SNPs in genomic locations outside regions with suspected inversions causing extended LD on LG1, LG2, LG7 and LG12 (Fig. 2b). The extremely long terminal branches and the absence of a clusters formed by the NEAC fish (Fig. 2b) suggest a complete lack of genetic structure between Barents Sea NEAC and the NCC stocks outside these chromosomal regions.

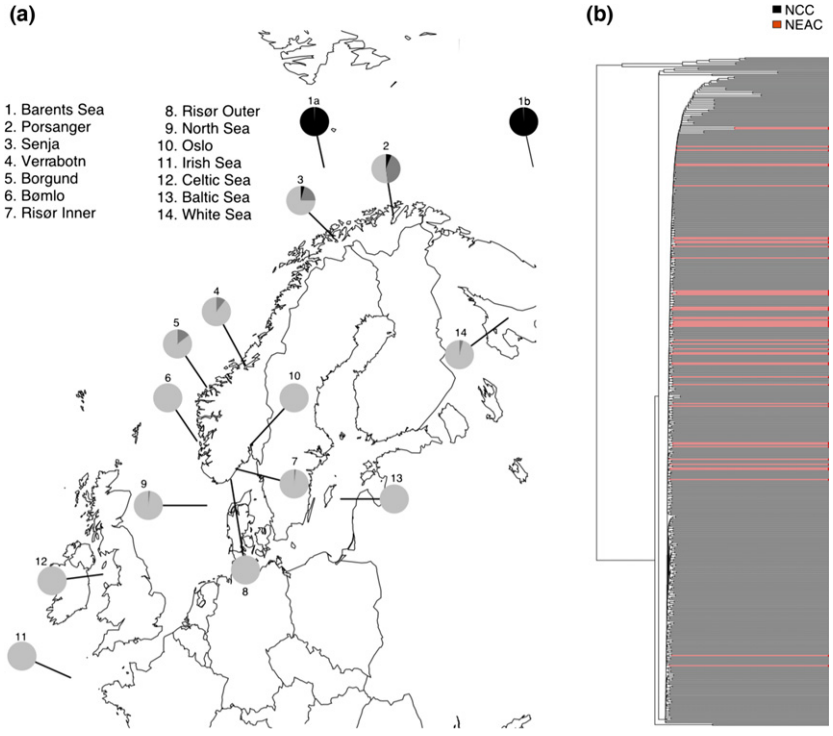

Fig. 2 Genomic divergence between NEAC and NCC. (a) Proportion of fish containing two (black), one (mid-grey) or no (light grey) copies of the NEAC haplotype in different Northeast Atlantic stocks. (b) Hierarchical clustering of SNP variation excluding genomic regions with suspected inversions due to large LD blocks (LG1, LG2, LG7, LG12). NEAC and NCC were represented by red and black tips, respectively. The genetic distance was calculated as identity by state across 7238 SNP loci.

Origin and age of inversions

To determine whether NEAC or NCC represents the ancestral state, we aligned LG1 sequences representing the arrangements of the inversions with Northern pike and stickleback to possibly identify conserved synteny blocks spanning breakage points defining the inversions. This analysis revealed a large block in pike spanning the break points flanking inversion 1, and a smaller block in stickleback bridging the two inversions (Fig. 3, Fig. S3, Supporting information). Taken together, these results suggest that NCC represents the ancestral state of the inverted structure.

The relative SNP density between NEAC and NCC across LG1 was calculated using whole-genome resequencing data from samples classified on the basis of the NEAC haplotype. Analysis of homozygous NEAC (n = 50), homozygous NCC (n = 11) or NEAC × NCC crosses (n = 43) revealed 540 685 SNPs with an average sequencing coverage of 17x. Relative heterozygosity expressed as the number of SNPs per 100 Kb in NEAC divided by the number in NCC revealed a dramatically reduced SNP density in NEAC samples within the LD block (Fig. 3). In contrast, the diversity outside the block was comparable for NEAC and NCC samples and to the rest of the genome. This result suggests a bottleneck event specific to the NEAC haplotype region on LG1 and supports the conclusion that NCC represents the ancestral state of the inverted structure.

The NEAC-NCC divergence relative to their divergence from Greenland cod were estimated to 0.57 and 0.13 within and outside the LG1 inversions, respectively. Assuming a divergence age of 3.5 million years between Greenland cod and Atlantic cod (Carr *et al.* 1999; Coulson *et al.* 2006), the inversion is estimated to be ~2 million years old (3.5*0.57 = 1.99). Although SNP data revealed no apparent genetic population structure between NEAC and NCC outside the inversion (Fig. 2a), we find 1553 FSD counts on LG1 outside the inversion. These divergent FSD sites are likely caused by a sample bias within NEAC and NCC fish because they represent a narrow genetic pool of interrelated individuals from a breeding programme rather than being a true random sample from both populations. Taking this background bias in FSD into account, the normalized Greenland cod–Atlantic cod divergence within the inversion would be ~1.6 million years (3.5*(0.57-0.13) = 1.57).

Candidate genes for adaptation to migratory behaviour

We annotated the LG1 sequence to search for genes involved in the adaptive divergence between the migratory and stationary ecotypes of Atlantic cod. The annotation resulted in the prediction of 1262 gene models for the whole chromosome, whereof 763 genes were located within the 17.4-Mb region containing the two inversions (357 and 406 genes, respectively). Variant detection within the same region revealed 19 206 SNPs that were fixed or very close to fixation for alternative alleles in NEAC and NCC and heterozygous in NEAC × NCC crosses, and included 849 plausible functional variants in 321 genes presenting good hits in the SWISSPROT database (Table S4, Supporting information). The corresponding protein variants displaying several amino acid substitutions included key enzymes in swim bladder function and haem synthesis together with important factors involved in muscle organization and behaviour (Fig. 3). Carbonic anhydrase catalyses the reversible conversion of carbon dioxide and water to bicarbonate and protons of importance for blood acidification and gas secretion into the swim bladder. The predicted NEAC and NCC variants of the secretory carbonic anhydrase isoform 6 (Ca6) differ at five positions, and the replacement of the highly conserved Gln196 with the novel His residue was shown by 3D modelling to reduce the interactions at the dimeric surface (Fig. 4, Table S5, Supporting information). We therefore predict reduced enzyme activity of the NCC variant as dimeric assembly of the enzyme confers an advantage for efficient CO_2 hydration in a variable extracellular milieu, such as the strong pH fluctuations in the gas gland (Pelster 2004; Pilka *et al.* 2012). We identified four additional genes within the double inversion involved in gas secretion by regulating glucose uptake and the production of acid metabolites. Glut1a facilitates glucose transport across cell membrane and is highly expressed in the gas gland cells of Atlantic cod (Hall *et al.* 2014). Two amino acid changes in the Glut1a protein and SNPs in the upstream region might have functional and regulatory effects. We also noted many SNPs in the genes coding for the three enzymes enolase 1 (Eno1), muscle-type phosphofructokinase (Pfkm) and glucose-6-phosphate dehydrogenase (G6pd) catalysing the anaerobic conversion of glucose to the acidic metabolites lactate and CO_2.

The inversions also contained candidate genes associated with the strenuous migratory behaviour, including two *alas* genes, and we found two amino acids changed in the predicted erythroid-specific Alas2 of crucial importance for haemoglobin production. Whereas the polymorphic *hb-1* gene is not located on LG1 (Borza *et al.* 2010), two amino acid changes in the rhesus type B glycoprotein (Rhbg) are probably responsible for the different blood type allele frequencies in NEAC and NCC (Møller 1966). The two populations also differ at four positions in the muscle protein leiomodin (Lmod3) essential for the organization of the sarcomeric thin filaments in skeletal muscle, and precise regulation of the

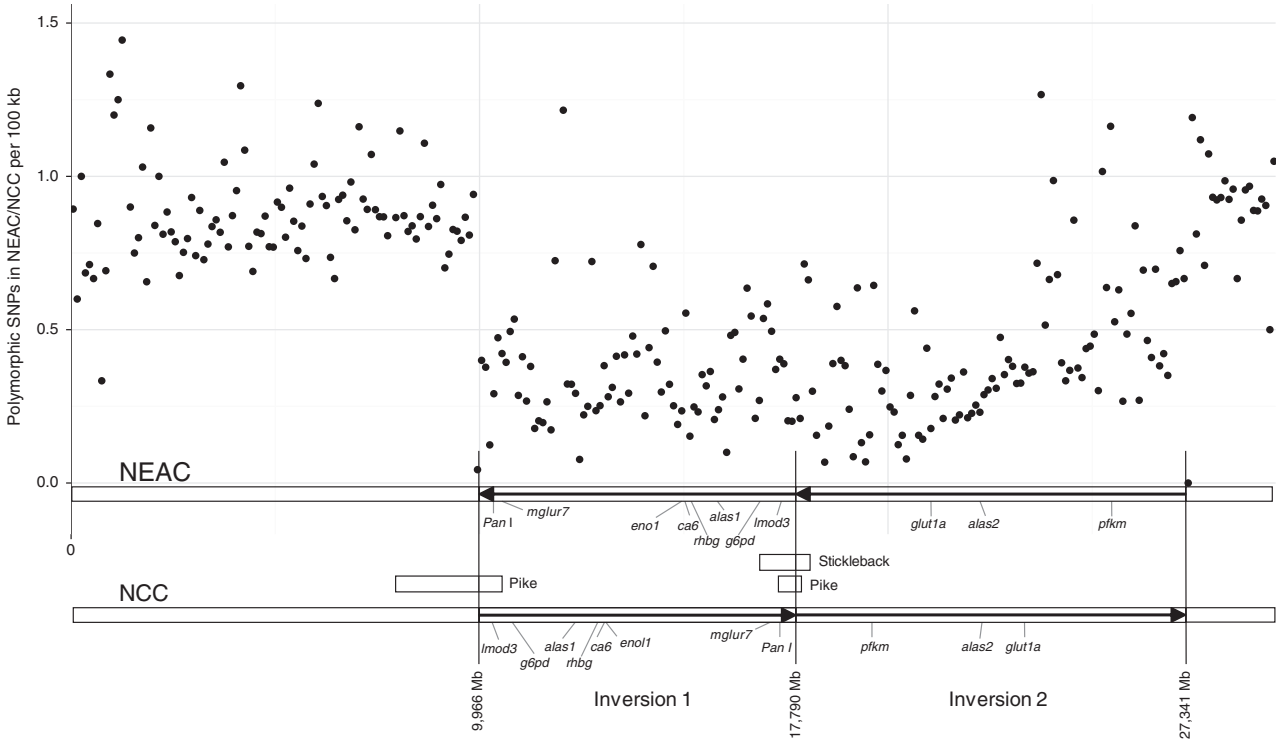

Fig. 3 Graphical representation of two adjacent inversions on LG1 present in NEAC and NCC. The upper part shows the relative difference in heterozygosity, measured as the number of polymorphisms per 100 kb in NEAC divided by the corresponding values in NCC. Conserved synteny blocks bridging inversion breakage points 1 and 2 suggest that NCC is holding the ancestral state of the inversions. Putative adaptive genes within the inversions are indicated.

filament length is crucial for optimal force generation during muscle contraction (Nworu *et al.* 2014; Yuen *et al.* 2014). Intriguingly, the double inversion contains the metabotropic glutamate receptor genes *mglur4* and *mglur7*, which are broadly expressed in the zebrafish brain, including olfactory bulb and retina (Haug *et al.* 2013). mGlur7 plays an important role in hippocampus-dependent spatial learning in mice (Goddyn *et al.* 2015), while the NEAC and NCC variants of the cod receptor were found to differ at three positions in a highly flexible region (data not shown).

Discussion

Population differentiation and adaptive evolution of Atlantic cod have been associated with four discrete islands of genomic divergence located on different chromosomes (Bradbury *et al.* 2010; Hemmer-Hansen *et al.* 2013; Karlsen *et al.* 2013). However, the fragmented nature of the current cod genome assembly (Star *et al.* 2011) has largely restricted our ability to identify genes associated with selection as well as our ability to reveal the mechanisms responsible for the observed patterns. To overcome these constraints, we constructed a dense linkage map and integrated it with draft genome assemblies to produce a cohesive chromosome sequence for Atlantic cod LG1. Separate linkage maps were constructed for pure NEAC, pure NCC and NEAC × NCC crosses in order to study differences in recombination patterns and potentially highlight rearrangements distinguishing the two ecotypes. These analyses revealed two adjacent inversions of 9.55 and 7.82 Mb (Fig. 1b), which clearly differentiate NEAC from NCC, as well as revealing a mechanism resulting in almost complete suppression of homologous recombination in individuals heterozygous for the inversions (Fig. 1a). Unlike the theoretic model of gene flow for simple inversions in *Drosophila* (Navarro *et al.* 1997), chromosomal polymorphisms involving more than one inversion prevent double crossovers and suppress recombination across the entire length of the rearrangements (Munte′ *et al.* 2005; Dyer *et al.* 2007; Huynh *et al.* 2011).

Chromosomal inversions have been associated with adaptive phenotypes in various plants and animals, including migratory species displaying high gene flow between the diverging populations (Rieseberg 2001; Hoffmann *et al.* 2004; Hoffmann & Rieseberg 2008). Polymorphic wing colour mimicry in butterflies is maintained by chromosomal rearrangements in the *Papilio* genus and in *Heliconius numata* (Joron *et al.* 2011;

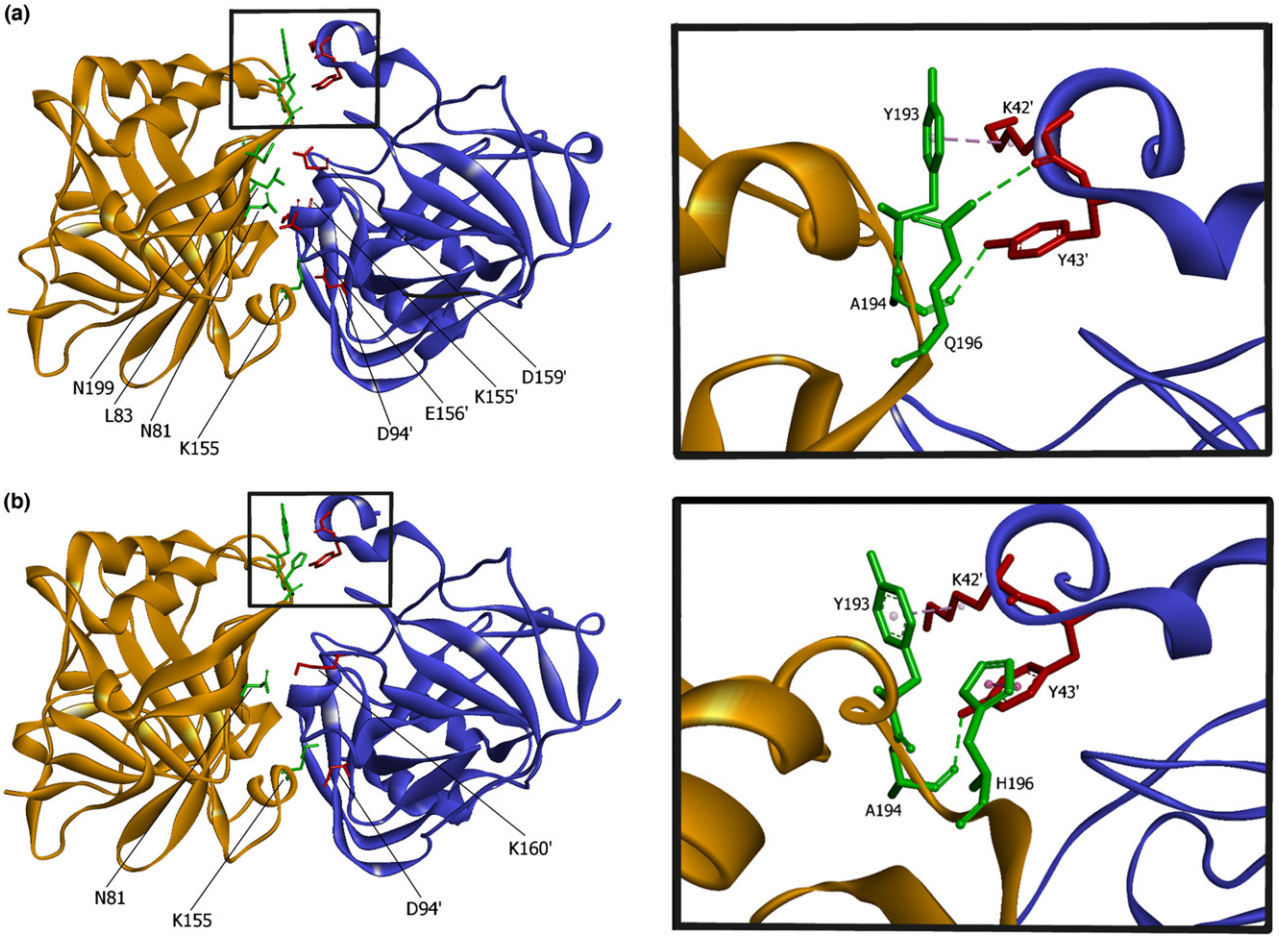

Fig. 4 Ribbon plots of the modelled carbonic anhydrase (Ca6) dimer interface in a) NEAC and b) NCC. The monomer subunits and key interacting residues (Table S5, Supporting information) are given in different colours. The enlarged sections show the dimeric interactions of Gln (Q) and His (H) at position 196.

Nishikawa *et al.* 2015), and at least two inversions spanning about 100 Mb in white-throated sparrow (*Zonotrichia albicollis*) were shown to harbour genes associated with territorial song and plumage (Thomas *et al.* 2008; Huynh *et al.* 2011; Zinzow-Kramer *et al.* 2015). Similarly, alternative reproductive strategies in the ruff (*Philomachus pugnax*) were recently found to be controlled by polymorphic genes affecting sex hormone levels and plumage within an inversion that occurred about 3.8 million years ago (Küpper *et al.* 2015; Lamichhaney *et al.* 2015). The repeated evolution of distinct marine and freshwater ecotypes of three-spined stickleback involves three chromosomal inversions, and alternative orientations of the voltage-gated potassium channel gene *kcnh4* might generate marine- and freshwater-specific isoforms (Jones *et al.* 2012). In rainbow trout (*Oncorhynchus mykiss*), different life history strategies of anadromous (steelhead) and resident ecotypes were recently shown to be associated with multiple loci with strong LD, suggesting the presence of an inversion suppressing recombination (Pearse *et al.* 2014).

The absence of genetic differentiation between NCC and NEAC populations outside the inversions on LG1 (Fig. 2b) supports previous conclusions of high levels of gene flow between migratory and stationary cod ecotypes in the North Atlantic Ocean (Hemmer-Hansen *et al.* 2013; Karlsen *et al.* 2013). However, the fact that the inversion is homozygous in NEAC, but polymorphic and under HWE in NCC populations, suggests that it is under strong directional selection in the migratory ecotype, while confers little or no fitness effects in the stationary ecotype. The absence of the NEAC haplotype in southern populations of NCC is interesting and raises the question whether this chromosomal rearrangement confers negative fitness effects in more southern marine ecosystems, for example due to suboptimal performance at higher water temperatures (Sarvas & Fevolden 2005). However, long-distance LD has been

reported between LG1 and LG2, LG7 and LG12 (Bradbury *et al.* 2013; Halldórsdóttir & Árnason 2016), making it possible that genes conferring the lack of fitness in warmer waters could be located elsewhere.

One of the obvious divergent adaptive challenges between NEAC and NCC populations is the adaptation to high hydrostatic pressure at large depths. The foraging and spawning migrations of NEAC involve vertical movements at depths of 200–400 m along stable thermal paths (Stensholt 2001), while stationary NCC fish exploit much shallower habitats (Hobson *et al.* 2007; Michalsen *et al.* 2014). This is supported by behavioural differences between juvenile NEAC and NCC settling at different depths, whereas the pelagic eggs have similar buoyancy (Løken *et al.* 1994; Fevolden *et al.* 2012; Jung *et al.* 2012). The swim bladder is a crucial organ by maintaining neutral buoyancy that allows fish to stay at their current depth without expending much energy swimming (Fänge 1953; Pelster 2004). Hence, impairment of the swim bladder function was assumed to significantly threaten the success of the spawning migration in the European eel (*Anguilla anguilla*) (Pelster 2014). Frequent descents and ascents lead to negative buoyancy, because gas secretion from the gas gland lags behind gas resorption in the swim bladder (Harden Jones & Scholes 1985; Godø & Michalsen 2000). This effect is amplified at greater depths, and the migratory NEAC should therefore benefit from enhanced gas secretion by increased blood acidification in the gas gland. The important role played by carbonic anhydrase in swim bladder function was demonstrated by inhibiting the enzyme activity in the gas gland that resulted in significantly reduced proton production and gas secretion (Fänge 1953; Skinazi 1953; Pelster 1995; Wurtz *et al.* 1999). While reduced carbonic anhydrase activity predicted for the NCC variant might not be critical for fish inhabiting shallow coastal water, the ability to maintain buoyancy is probably crucial for NEAC during frequent vertical movements to large depths. The energetic costs associated with the strenuous migrations may be further reduced by increased oxygen delivery and enhanced muscular capacity involving a suite of adaptive alleles identified within the inversions.

Gene flow between populations with divergent adaptive challenges can result in large fitness costs when recombination disrupts coinheritance of advantageous genetic variants. A genetic architecture that enforces strong LD between coselected gene variants would therefore be highly favourable under extensive gene flow from divergent populations. Such 'supergenes' have been shown to maintain population-specific adaptations in various organisms and are often caused by larger chromosome rearrangements (Joron *et al.* 2011; Thompson & Jiggins 2014; Twyford & Friedman 2015). We therefore hypothesize that the inversions on LG01 act as a supergene to efficiently maintain coinheritance of several highly favourable genetic variants, which over time have generated the island of genomic divergence observed between migratory and stationary ecotypes of cod.

Similar to NEAC and NCC, Icelandic migratory and stationary cod populations inhabiting different depths show genetic differentiation at the same genomic region as found in the Norwegian cod populations (Pampoulie *et al.* 2008, 2015; Grabowski *et al.* 2011). Additionally, offshore Atlantic cod from Davis Strait clustered with presumably NEAC sampled in the Barents Sea by SNP genotyping trans-Atlantic populations lacking NCC (Bradbury *et al.* 2013). Together, this suggests a common ancestry for the migratory cod populations and supports an old origin of the inversion polymorphism on LG1 associated with divergent behavioural adaptations in Atlantic cod. We estimated that the inversion arose about 1.6–2 million years ago during Pleistocene when glacial barriers and lowered sea level greatly influenced the abundance and distribution of marine species. This epoch probably represented the most important vicariance event in the evolution of Arctic fishes (Mecklenburg *et al.* 2011; Owens 2015). Atlantic cod survived in glacial refugia, but also moved southwards to ice-free regions during the glacial periods (Bigg *et al.* 2008; Kettle *et al.* 2011). We propose that beneficial alleles were captured within the two inversions that occurred in an isolated refugial population and later became fixed. During interglacial periods, local adapted individuals may have dispersed in the Arctic region and are today represented by the large migratory cod populations exploiting the high seasonal productivity in the most northerly environments on both sides of North Atlantic (Robichaud & Rose 2004). Adaptation to the polar environment might have been refined in the hybrid species walleye pollock (*Gadus chalcogrammus*) between Atlantic cod and polar cod (*Boreogadus saida*) that trespasses the ecology of its parents (Halldórsdóttir & Árnason 2016).

In conclusion, we reveal a major difference in the genomic architecture of the migratory NEAC and stationary NCC ecotypes by documenting two adjacent inversions spanning 17.4 Mb on LG1 that effectively block recombination in individuals heterozygous for the inversions. Despite clear signs of interbreeding, this lack of recombination has caused a supergene comprising adaptive alleles related to the migratory ecotype to be preserved without dilution from the stationary ecotype. The status of the two ecotypes differs substantially today that will probably increase in future. The NEAC population has increased greatly in recent years and the spawning stock in Barents Sea is historic high. In contrast, dramatic low levels in the NCC stocks and recruitment have resulted in restrictions on commercial cod fishing in several

coastal regions. As a consequence of increased water temperatures, Atlantic cod is expected to spread northwards and occupy larger areas of Barents Sea, while southern stocks will probably decline or disappear within year 2100 (Drinkwater 2005). Knowledge about the breeding structure of NEAC and NCC and the functional roles played by the differentiating genes might be important for the fisheries management and for predicting the response of the ecotypes to future climate change.

Acknowledgements

We are grateful to Kim Præbel for providing Greenland cod sample. Tom Cross, Phil McGinnity, Richard Fitzgerald, Halvor Knutsen and Roman Wenne are acknowledged for providing Atlantic cod population samples from the Irish Sea, Risør, North Sea and Baltic Sea. Genotypes for LG01 were generated as a part of the Cod SNP Consortium (CSC), a collaboration between CIGENE, CEES, IMR and Nofima. Mariann Arnyasi is acknowledged for genotyping and filtering of SNP data. Ole Kristian Tørresen and Alexander J. Nederbragt are acknowledged for providing draft genome assemblies for Atlantic cod. Funding was provided by the Norwegian University of Life Sciences to cover the salary of T.G.K. and from the Research Council of Norway (Project Nos.: 207680/E40 and 22734/O30).

References

Andersen O, Johnsen H, De Rosa MC *et al.* (2015) Evolutionary history and adaptive significance of the polymorphic *Pan* I in migratory and stationary populations of Atlantic cod (*Gadus morhua*). *Marine Genomics*, **22**, 45–54.

Árnason E, Pálsson S (1996) Mitochondrial cytochrome b DNA sequence variation of Atlantic cod *Gadus morhua*, from Norway. *Molecular Ecology*, **5**, 715–724.

Bangera R, Odegard J, Nielsen HM, Gjoen HM, Mortensen A (2013) Genetic analysis of vibriosis and viral nervous necrosis resistance in Atlantic cod (*Gadus morhua* L.) using a cure model. *Journal of Animal Science*, **91**, 3574–3582.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Bergstad OAJT, Dragesun O (1987) Life history and ecology of the gadoid resources of the Barents Sea. *Fisheries Research*, **5**, 119–161.

Bigg GR, Cunningham CW, Ottersen G *et al.* (2008) Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **275**, 163–172.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Borza T, Higgins B, Simpson G, Bowman S (2010) Integrating the markers *Pan* I and Haemoglobin with the genetic linkage map of Atlantic cod (*Gadus morhua*). *Genetics*, **157**, 317–330.

Bradbury IR, Hubert S, Higgins B *et al.* (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **277**, 3725–3734.

Bradbury IR, Hubert S, Higgins B *et al.* (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in a exploited marine fish. *Evolutionary Applications*, **6**, 450–461.

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, **81**, 1084–1097.

Carr SMKDS, Pepin P, Crutcher DC (1999) Molecular systematics of gadid fishes: implications for the biogeographic origins of Pacific species. *Canadian Journal of Zoology*, **77**, 19–26.

Christensen V, Guenette S, Heymans JJ *et al.* (2003) Hundred-year decline of North Atlantic predatory fishes. *Fish and Fisheries*, **4**, 1–24.

Cingolani P, Platts A, le Wang L *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.

Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome*, **49**, 1115–1130.

Dobin A, Davis CA, Schlesinger F *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Drinkwater KF (2005) The response of Atlantic cod (*Gadus morhua*) to future climate change. *ICES Journal of Marine Science*, **62**, 1327–1337.

Dyer KA, Charlesworth B, Jaenike J (2007) Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 1587–1592.

Fänge R (1953) The mechanisms of gas transport in the euphysoclist swimbladder. *Acta Physiologica Scandinavica. Supplementum*, **30**, 1–133.

Fevolden SE, Pogson GH (1997) Genetic divergence at the synaptophysin (Syp I) locus among Norwegian coastal and north-east Arctic populations of Atlantic cod. *Journal of Fish Biology*, **51**, 895–908.

Fevolden SE, Sarvas T (2001) *Distinct genetic divergence between cod (Gadus morhua) in fjords and cod in offshore waters in northern Norway.* ICES CM 2001/L:04.

Fevolden SE, Westgaard JI, Pedersen T, Praebel K (2012) Settling-depth vs. genotype and size vs. genotype correlations at the Pan I locus in 0-group Atlantic cod *Gadus morhua*. *Marine Ecology Progress Series*, **468**, 267–278.

Goddyn H, Callaerts-Vegh Z, D'Hooge R (2015) Functional dissociation of group III metabotropic glutamate receptors revealed by direct comparison between the behavioral profiles of knockout mouse lines. *International Journal of Neuropsychopharmacology*, **18**, pyv053, doi:10.1093/ijnp/pyv053.

Godø OR, Michalsen K (2000) Migratory behaviour of north-east Arctic cod, studied by use of data storage tags. *Fisheries Research*, **48**, 127–140.

Grabowski TB, Thorsteinsson V, McAdam BJ, Marteinsdottir G (2011) Evidence of segregated spawning in a single marine fish stock: sympatric divergence of ecotypes in Icelandic cod? *PLoS ONE*, **6**, e17528.

Green P, Falls K (1990) *Documentation for CRI-MAP version 2.4.* Washington University School of Medicine, St. Louis, Missouri.

Haas BJ, Papanicolaou A, Yassour M *et al.* (2013) *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*, **8**, 1494–1512.

Hall JR, Clow KA, Short CE, Driedzic WR (2014) Transcript levels of class I GLUTs within individual tissues and the direct relationship between GLUT1 expression and glucose metabolism in Atlantic cod (*Gadus morhua*). *Journal of Comparative Physiology B*, **184**, 483–496.

Halldórsdóttir K, Árnason E (2016) *Whole-genome sequencing uncovers cryptic and hybrid species among Atlantic and Pacific cod-fish. bioRxiv.* preprint first posted online Dec. 20, 2015; doi: http://dx.doi.org/10.1101/034926.

Harden Jones FR, Scholes P (1985) Gas secretion and resorption in the swimbladder of the cod *Gadus morhua*. *Journal of Comparative Physiology B*, **155**, 319–331.

Harris RS (2007) *Improved pairwise alignment of genomic DNA*. Ph.D. Thesis, The Pennsylvania State University.

Haug MF, Gesemann M, Mueller T, Neuhauss SC (2013) Phylogeny and expression divergence of metabotropic glutamate receptor genes in the brain of zebrafish (*Danio rerio*). *The Journal of Comparative Neurology*, **521**, 1533–1560.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333–362.

Hemmer-Hansen J, Nielsen EE, Therkildsen NO *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, **22**, 2653–2667.

Hobson VJ, Righton D, Metcalfe JD, Hays GC (2007) Vertical movements of North Sea cod. *Marine Ecology Progress Series*, **347**, 101–110.

Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, **39**, 21–42.

Hoffmann AA, Sgrò CM, Weeks AR (2004) Chromosomal inversion polymorphism and adaptation. *Trends in Ecology and Evolution*, **19**, 482–488.

Hutchings JA, Bishop TD, McGregor-Shaw CR (1999) Spawning behaviour of Atlantic cod, *Gadus morhua*: evidence of mate competition and mate choice in a broadcast spawner. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 97–107.

Huynh LY, Maney DL, Thomas JW (2011) Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). *Heredity*, **106**, 537–546.

Jakobsen T (1987) Coastal cod in northern Norway. *Fisheries Research*, **5**, 223–234.

Johnsen H, Andersen Ø (2012) Sex dimorphic expression of five dmrt genes identified in the Atlantic cod genome. The fish-specific *dmrt2b* diverged from *dmrt2a* before the fish whole-genome duplication. *Gene*, **505**, 221–232.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Joron M, Frezal L, Jones RT *et al.* (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–206.

Jung KM, Folkvord A, Kjesbu OS *et al.* (2012) Egg buoyancy variability in local populations of Atlantic cod (*Gadus morhua*). *Marine Biology*, **159**, 1969–1980.

Karlsen BO, Klingan K, Emblem A *et al.* (2013) Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, **22**, 5098–5111.

Karlsen BO, Emblem A, Jorgensen TE *et al.* (2014) Mitogenome sequence variation in migratory and stationary ecotypes of North-east Atlantic cod. *Marine Genomics*, **15**, 103–108.

Kettle AJ, Morales Muñiz A, Roselló-Izquierdo E, Heinrich D, Vollestad A (2011) Refugia of marine fish in the northeast Atlantic during the last glacial maximum: concordant assessment from archaeozoology and paleotemperature reconstructions. *Climate of the Past*, **7**, 181–201.

Küpper C, Stocks M, Risse JE (2015) A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, **48**, 79–83.

Lamichhaney C, Fan G, Widemo F *et al.* (2015) Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, **48**, 84–88.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *Journal of Molecular Biology*, **231**, 1049–1067.

Løken S, Pedersen T (1996) Effect of parent type and temperature on vertebrae number in juvenile cod, *Gadus morhua* (L), in Northern Norway. *Sarsia*, **80**, 293–298.

Løken S, Pedersen T, Berg E (1994) Vertebrae numbers as an indicator for the recruitment mechanism of coastal cod of northern Norway. *Cod and Climate Change – Proceedings of a Symposium*, **198**, 510–519.

Mackenzie BR, Almesjo L, Hansson S (2004) Fish, fishing, and pollutant reduction in the Baltic Sea. *Environmental Science & Technology*, **38**, 1970–1976.

Martinez PA, Zuranoa JP, Amadoa TF *et al.* (2015) Chromosomal diversity in tropical reef fishes is related to body size and depth range. *Molecular Phylogenetics and Evolution*, **93**, 1–4.

McKenna A, Hanna M, Banks E *et al.* (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Mecklenburg N, Garcia-Lopez R, Puelles E, Sotelo C, Martinez S (2011) Cerebellar oligodendroglial cells have a mesencephalic origin. *Glia*, **59**, 1946–1957.

Michalsen K, Johansen T, Subbey S, Beck A (2014) Linking tagging technology and molecular genetics to gain insight in the spatial dynamics of two stocks of cod in Northeast Atlantic waters. *ICES Journal of Marine Science*, **71**, 1417–1432.

Møller D (1966) Genetic differences between cod groups in the Lofoten area. *Nature*, **212**, 824

Møller D (1968) Genetic diversity in spawning cod along the Norwegian coast. *Hereditas*, **60**, 1–32.

Møller D (1969) The relationship between arctic and coastal cod in their immature stages illustrated by frequencies of genetic characters. *Fiskeridirektoratets Skrifter. Serie Havundersøkelser*, **15**, 220–223.

Monteiro CA, Serrão EA, Pearson GA (2012) Prezygotic barriers to hybridization in marine broadcast spawners: reproductive timing and mating system variation. *PLoS ONE*, **7**, e35978.

Mork J, Sundnes G (1985) 0-group cod (*Gadus morhua*) in captivity: different survival of certain genotypes. *Helgoländer Meeresuntersuchungen*, **39**, 63–70.

Munte′ A, Rozas J, Aguade M, Segarra C (2005) Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. *Genetics*, **169**, 1573–1581.

Myers RA, Hutchings JA, Barrowman NJ (1997) Why do fish stocks collapse? The example of cod in Atlantic Canada. *Ecological Applications*, **7**, 91–106.

Navarro A, Betran E, Barbadilla A, Ruiz A (1997) Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics*, **146**, 695–709.

Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.

Nishikawa H, Iijima T, Kajitani R *et al.* (2015) A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nature Genetics*, **47**, 405–409.

Nordeide JT, Folstad I (2000) Is cod lekking or a promiscuous group spawner? *Fish and Fisheries*, **1**, 90–93.

Nordeide JT, Pettersen IH (1998) Haemoglobin frequencies and vertebral numbers of cod (*Gadus morhua* L.) off northern Norway – test of a population structure hypothesis. *ICES Journal of Marine Science*, **55**, 134–140.

Nordeide J, Johansen S, Jørgensen T, Karlsen BMoum T (2011) Population connectivity among migratory and stationary cod *Gadus morhua* in the Northeast Atlantic – a review of 80 years of study. *Marine Ecology Progress Series*, **435**, 269–283.

Nworu C, Kraft R, Schnurr D, Gregorio CKrieg P (2014) Leiomodin 3 and tropomodulin 4 have overlapping functions during skeletal myofibrillogenesis. *Journal of Cell Science*, **128**, 239–250.

Ottersen G, Bogstad B, Yaragina N *et al.* (2014) A review of early life history dynamics of Barents Sea cod (*Gadus morhua*). *ICES Journal of Marine Science*, **71**, 2064–2087.

Owens HL (2015) Evolution of codfishes (Teleostei: *Gadinae*) in geographical and ecological space: evidence that physiological limits drove diversification of subarctic fishes. *Journal of Biogeography*, **42**, 1091–1102.

Pálsson OK, Thorsteinsson V (2003) Migration patterns, ambient temperature, and growth of Icelandic cod (*Gadus morhua*): evidence from storage tag data. *Canadian Journal of Fisheries and Aquatic Sciences*, **60**, 1409–1423.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547–572.

Pampoulie C, Jakobsdottir KB, Marteinsdottir G, Thorsteinsson V (2008) Are vertical behaviour patterns related to the pantophysin locus in the Atlantic cod (*Gadus morhua* L.)? *Behavior Genetics*, **38**, 76–81.

Pampoulie C, Skirnisdottir S, Star B *et al.* (2015) Rhodopsin gene polymorphism associated with divergent light environments in Atlantic cod. *Behavior Genetics*, **45**, 236–244.

Pearse DE, Miller MR, Abadia-Cardoso A, Garza JC (2014) Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proceedings of the Royal Society B*, **281**, 20140012.

Pelster B (1995) Mechanisms of acid release in isolated gas gland cells of the European eel Anguilla anguilla. *American Journal of Physiology*, **269**, R793–R799.

Pelster B (2004) pH regulation and swimbladder function in fish. *Respiratory Physiology & Neurobiology*, **144**, 179–190.

Pelster B (2014) Swimbladder function and the spawning migration of the European eel *Anguilla anguilla*. *Frontiers in Physiology*, **5**, 486.

Pilka ES, Kochan G, Oppermann U, Yue WW (2012) Crystal structure of the secretory isozyme of mammalian carbonic anhydrases CA VI: implications for biological assembly and inhibitor development. *Biochemical and Biophysical Research Communications*, **419**, 485–489.

Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin (*Pan* I) locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics*, **157**, 317–330.

Rieseberg L (2001) Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, **16**, 351–358.

Robichaud D, Rose G (2004) Migratory behaviour and range in Atlantic cod: inference from a century of tagging. *Fish and Fisheries*, **5**, 185–214.

Rollefsen G (1933) The otoliths of the cod. *Fiskeridirektoratets Skrifter. Serie Havundersøkelser*, **4**, 1–4.

Rollefsen G (1954) Observations on cod and cod fisheries of Lofoten. *Rapp. P.-v Rapports et procès-verbaux des réunions/ Conseil permanent international pour l'exploration de la mer*, **136**, 40–47.

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, **234**, 779–815.

Sarvas TH, Fevolden SE (2005) Pantophysin (*Pan* I) locus divergence between inshore v. offshore and northern v. southern populations of Atlantic cod in the north-east Atlantic. *Journal of Fish Biology*, **67**, 444–469.

Shen KN, Chang CW, Durand JD (2015) Spawning segregation and philopatry are major prezygotic barriers in sympatric cryptic *Mugil cephalus* species. *Comptes Rendus Biologies*, **338**, 803–811.

Skinazi L (1953) Carbonic anhydrase in two closely related teleosts; inhibition of the secretion of gas from the air bladder of perch by sulfonamides. *Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales*, **147**, 295–299.

Star B, Nederbragt AJ, Jentoft S *et al.* (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, **477**, 207–210.

Stensholt BK (2001) Cod migration patterns in relation to temperature: analysis of storage tag data. *ICES Journal of Marine Science*, **58**, 770–793.

Sundby S, Nakken O (2008) Spatial shifts in spawning habitats of Arcto-Norwegian cod related to multidecadal climate oscillations and climate change. *ICES Journal of Marine Science*, **65**, 953–962.

Thomas DW, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, **21**, 1859–1875.

Thomas JW, Caceres M, Lowman JJ *et al.* (2008) The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics*, **179**, 1455–1468.

Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. *Heredity*, **113**, 1–8.

Trapnell C, Williams BA, Pertea G *et al.* (2010) Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology*, **28**, 511–515.

Twyford A, Friedman J (2015) Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution*, **69**, 1476–1486.

Westgaard JI, Fevolden SE (2007) Atlantic cod (*Gadus morhua* L.) in inner and outer coastal zones of northern Norway display divergent genetic signature at non-neutral loci. *Fisheries Research*, **85**, 306–315.

Williams GC (1975) *Sex and Evolution*. Princeton University Press, Princeton, New Jersey.

Wurtz J, Salvenmoser W, Pelster B (1999) Localization of carbonic anhydrase in swimbladder of European eel (*Anguilla anguilla*) and perch (*Perca fluviatilis*). *Acta Physiologica Scandinavica*, **165**, 219–224.

Yuen M, Sandaradura SA, Dowling JJ *et al.* (2014) Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *Journal of Clinical Investigations*, **124**, 4693–4708.

Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.

Zinzow-Kramer WM, Horton BM, McKee CD *et al.* (2015) Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes, Brain and Behavior*, **14**, 641–654.

S.L. and Ø.A. designed the study with input from T.G.K. H.G and M.P.K. T.G.K, S.L., H.G., S.R.S and T.N. analysed the data. Ø.A., T.G.K. and S.L. examined candidate genes. T.J., H.O. and M.B. provided samples from family material and wild populations. M.B. and A.S. provided sequence data from the National cod breeding programme. M.C.D.R and B.R. modelled the protein variants. Ø.A., T.G.K., H.G, S.R.S., M.P.K. and S.L. wrote the manuscript with contributions from all authors.

Data accessibility

Resequencing data from 104 farmed cod were submitted to the European Nucleotide Archive and can be retrieved under Accession no. PRJEB12803. Genotype data from SNP array are available in Dryad (http://datadryad.org/). All SNPs are referred to by their ss# or rs# available in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Linkage maps generated from 96 families of farmed cod. Maps and map distances were calculated separately for NEAC, NCC and NEAC × NCC crosses, and also split between males and females.

Table S2 Predicted function of open reading frames were found with TransDecoder and homology search using BLASTP against zebrafish and stickleback protein databases.

Table S3 Reads from a Greenland cod and a NCC where aligned to the NEAC reference. Fixed sequence differences were counted for LG1, both inside the inversions (top right) and outside (bottom left).

Table S4 Genes with nonsynonymous SNPs within the inversions. Number of individuals with reference or alternative alleles from NEAC, NCC and NEAC × NCC cross are indicated together with GI number and accession IDs.

Table S5 Interdimeric contacts in carbonic anhydrase (Ca6) of NEAC (Gln196) and NCC (His196). Protein contacts (within 4.5 Å) in the interfaces between A- and B-monomers of the homology models are reported.

Fig. S1 The double inversion was dated by comparing estimates of sequence divergence between NEAC-NCC within the inversion with sequence divergence between Greenland cod and Atlantic Cod under the assumption that this latter species divergence occurred 3.5 million years ago (Carr *et al.* 1999; Coulson *et al.* 2006). The level of sequence divergence was measured in units of fixed sequence differences (FSD). The ratio between NEAC-NCC FSD distance (red line) and the mean FSD distance between Greenland cod-NEAC (blue line) and Greenland cod-NCC (green line) was calculated and then multiplied with the Greenland cod- Atlantic Cod divergence age (3.5 Mya) to get the absolute age of NEAC-NCC divergence.

Fig. S2 Pairwise LD for 48 NEAC, measured as r^2, between all SNPs (MAF>0.1) detected by resequencing on LG1. Left figure is the NCC map, while right figure is the NEAC map. Only values above $r^2 = 0.7$ are shown. Circle indicates a region within the second inversion being in high LD with a region at the end of the chromosome. The NEAC map minimizes the distance between these two regions.

Fig. S3 Comparative map using whole chromosome alignment between the NCC version of LG1 and stickleback LGXIII (a) and pike LG12 (b).