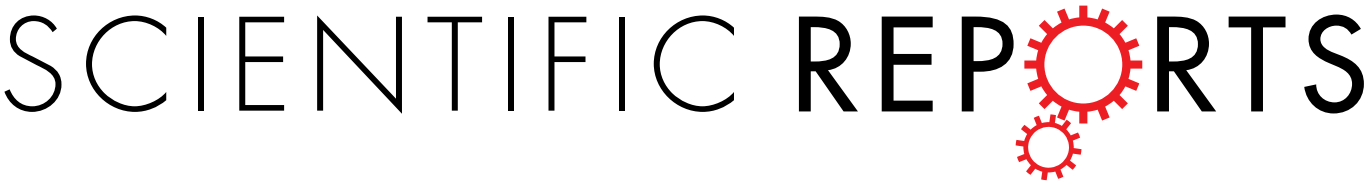

OPEN

Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod

Received: 21 December 2015
Accepted: 02 March 2016
Published: 17 March 2016

Paul R. Berg[1], Bastiaan Star[1], Christophe Pampoulie[2], Marte Sodeland[3,4], Julia M. I. Barth[1], Halvor Knutsen[1,3,4], Kjetill S. Jakobsen[1] & Sissel Jentoft[1,4]

Identification of genome-wide patterns of divergence provides insight on how genomes are influenced by selection and can reveal the potential for local adaptation in spatially structured populations. In Atlantic cod – historically a major marine resource – Northeast-Arctic- and Norwegian coastal cod are recognized by fundamental differences in migratory and non-migratory behavior, respectively. However, the genomic architecture underlying such behavioral ecotypes is unclear. Here, we have analyzed more than 8.000 polymorphic SNPs distributed throughout all 23 linkage groups and show that loci putatively under selection are localized within three distinct genomic regions, each of several megabases long, covering approximately 4% of the Atlantic cod genome. These regions likely represent genomic inversions. The frequency of these distinct regions differ markedly between the ecotypes, spawning in the vicinity of each other, which contrasts with the low level of divergence in the rest of the genome. The observed patterns strongly suggest that these chromosomal rearrangements are instrumental in local adaptation and separation of Atlantic cod populations, leaving footprints of large genomic regions under selection. Our findings demonstrate the power of using genomic information in further understanding the population dynamics and defining management units in one of the world's most economically important marine resources.

Genomic differentiation between populations can display complex patterns, involving selection of numerous genome wide loci[1–4] and challenge the understanding of the different evolutionary processes that shape such genomic signatures[5,6]. The key lies in understanding the genetic architecture of adaptive divergence and the balance between divergent selection and homogenizing gene flow. Genome-wide single nucleotide polymorphism (SNP) analyses as well as large-scale sequencing of natural populations address this challenge by identifying areas of the genome involved in diversification[1,2,7,8], and sometimes also the underlying candidate genes involved in population divergence[4,9,10]. In some cases, genomic islands of divergence[5,6] – linked loci within genomic regions under selection – have been observed, whereby elevated levels of divergence between individuals or populations expand over extensive regions[4,11]. Such patterns can emerge via divergence hitchhiking[12] or by other factors that reduce recombination across the genome, such as chromosomal rearrangements and thereby maintaining polymorphism in complex traits[13]. Chromosomal inversion polymorphism may play a key role in the process of local adaptation if it captures several locally adapted alleles since the inversion suppresses meiotic recombination in heterozygous individuals, thereby avoiding the association of adapted/maladapted allele combinations[13]. The identification of a limited number of differentiated genomic regions, in combination with little or no genetic differentiation in other parts of the genome, presumed not to be under selection, is usually interpreted as a sign of ecological differentiation through local adaptation in the compared populations or species[12,14]. Indeed, if genetic structuring is low in the presumably neutrally evolving part of the genome, divergent regions are most likely of functional importance[6,15,16]. Nevertheless, in most species, and in marine species in particular, the potential for such adaptation and the underlying genetic architecture remains unclear.

[1]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, N-0316 Oslo, Norway. [2]Marine Research Institute, Skúlagata 4, 101 Reykjavik, Iceland. [3]Institute of Marine Research, Flødevigen, N-4817 His, Norway. [4]Department of Natural Sciences, University of Agder, N-4604 Kristiansand, Norway. Correspondence and requests for materials should be addressed to P.R.B. (email: p.r.berg@ibv.uio.no)

Atlantic cod is one of the most studied and exploited marine species in the world. Despite this fact, the degree of population structuring[17] and the potential for local adaptation[18] remains debated. In 1933, two distinct groups of Atlantic cod were described by Rollefsen[19], based on growth zones and patterns of otoliths in what is now known as Northeast Arctic cod (NEAC) and Norwegian coastal cod (NCC). Since then, a controversy has existed on whether NEAC and NCC are genetically distinct populations. After more than 80 years of controversy, these issues are still not fully resolved[20], even though genetic markers under selection, such as hemoglobin[21] and *Pan* I[22] display significant frequency differences between NEAC and NCC. The NEAC is characterized by long distance migrations from the spawning grounds along the Norwegian coast to feeding areas in the Barents Sea. The main spawning grounds are off the Lofoten islands[23] and after spawning, the majority of eggs and larvae drift along the coast into the nursery area in the Barents Sea. In contrast to NEAC, NCC inhabits coastal- and fjord areas along the Norwegian coast, perform relatively short coastal migrations[24] and spawn along most of the Norwegian coast[25], including the Lofoten area[26]. To date, it is uncertain if the NCC is a self recruited population or if it is a stock recruited in part also by vagrant NEAC individuals. Several mechanisms have been proposed to hinder the potential for hybridization between NEAC and NCC like lekking spawning behavior[27], depth/temperature preferences at spawning[26], drift trajectories of offspring[28] and settling depth for juveniles[29]. The clearly observed phenotypic diversity between ecotypes of migrating and non-migrating Atlantic cod populations that nonetheless spawn in the vicinity of each other, offers an excellent opportunity to identify the potential for local adaptation and investigate its genomic architecture, when both natural selection and gene flow are potentially high, in a major marine resource.

Population divergence of Atlantic cod populations in northern Norway have so far been described by a small to moderate number of genetic markers[20,30] or by pooled population data[31], which limits inference of the genomic architecture underlying local adaptation as well as the level of neutral divergence. Nonetheless, distinct regions of elevated divergence have been detected by comparing Atlantic cod populations in other parts of its geographical range[4,30]. Moreover, it has been suggested that these regions consist of genomic rearrangements like chromosomal inversions[11]. We here aim at elucidating the genomic distinctions between migratory cod (NEAC) and costal cod (NCC) using the available SNP chip resource featuring more than 8.000 SNPs distributed throughout the genome. An additional comparison with a more remote population from the North Sea enables us to compare the population structure of NEAC and NCC that spawn in the vicinity of each other to a population that spawns at a distinctly different location and hence better quantify the genomic differences within and between different areas. We use population genetic theory and two different outlier approaches to identify SNPs and thus, genomic regions under selection. Distinct genomic regions separating the two populations were demonstrated and chromosomal rearrangement patterns were further investigated, using different approaches based on linkage disequilibrium (LD) and haplotype tagging. Finally, we discuss the mechanisms driving the observed patterns of local adaptation and separation in Atlantic cod and in marine fish species in general.

Results

A total of 8.168 SNPs were analyzed in 141 individuals of Atlantic cod (Fig. 1, Table 1). The SNPs were distributed over all 23 linkage groups[32,33] (LGs) with an average distance between SNPs of 94.000 bp (based on a genome size of 830 Mb)[33]. Out of these SNPs, a total of 5.205 SNPs were located within 5.000 bp of 4.247 Ensembl annotated genes.

Outlier detection and identification of genomic regions under selection. The BAYESCAN analyses identified 336 SNPs (4.1%) as candidates for divergent selection in the NEAC/NCC comparison ($q < 0.01$), while FDIST2, implemented in LOSITAN, identified 479 outlier SNPs (5.9%, $q < 0.01$, Supplementary Table S1). All SNPs identified as outliers by BAYESCAN were also identified by LOSITAN, resulting in a final outlier dataset of 336 SNPs (Fig. 2a, Supplementary Table S2). LG1, 2 and 7 have the highest number of outliers: 146, 35, 154 respectively while a single outlier SNP was detected in LG4 (Supplementary Table S2). Out of these SNPs, 244 loci were located in or within 5 Kb of a known gene, of which 134 were located in exons and 114 were non-synonymous substitutions causing amino acid changes. Notably, a single outlier SNP in a gene with unknown function (ENSGMOG 00000011194) was detected in LG4 in the final outlier dataset in addition to a limited number of outlier SNPs in 16 other LGs (Supplementary Table S1) that were only detected as outliers using LOSITAN. These are all single outliers, not representing any larger outlier blocks and not residing within linked regions of the genome, however indicating that a few smaller areas of the genome also could play a role in the genomic diversification of the NEAC and NCC populations. The identified outlier pattern between the NEAC and the NS population (Fig. 2b) resembles the NEAC/NCC comparison, except that outlier signals are generally stronger and 109 additional outlier SNPs were also detected within LG12 (Supplementary Table S1). Nine SNPs were candidates for selection in the comparison between NCC and NS (Fig. 2c, Supplementary Table S1).

SNPs were categorized as outlier SNPs or as neutral SNPs based on the comparison between NEAC and NCC populations. The final outlier dataset of 336 SNPs are represented by 86 tag-SNPs while the neutral dataset of 7.702 SNPs are represented by 7.384 tag-SNPs (details on tag-SNP selection are given in Materials and Methods).

Population genetic structuring. F_{ST} values of the outlier SNPs ($F_{ST} = 0.35053$) were orders of magnitude larger than those of the neutral SNPs ($F_{ST} = 0.00123$) between the NEAC and NCC populations, and all F_{ST} values were significantly different from 0 (Table 2). Moreover, the elevated F_{ST} values predominantly occur in LG1, 2 and 7, (Fig. 3a, Supplementary Table S3). We observed slightly larger F_{ST} differences when comparing the NEAC population to the more geographically distant NS population (Table 2), predominantly through elevated F_{ST} values in LG12 (Fig. 3b, Supplementary Table S3). When comparing the NCC and NS populations, F_{ST} values are generally low (Table 2) with small but distinct F_{ST} elevations in LG2 and 12 (Fig. 3c, Supplementary Table S3).

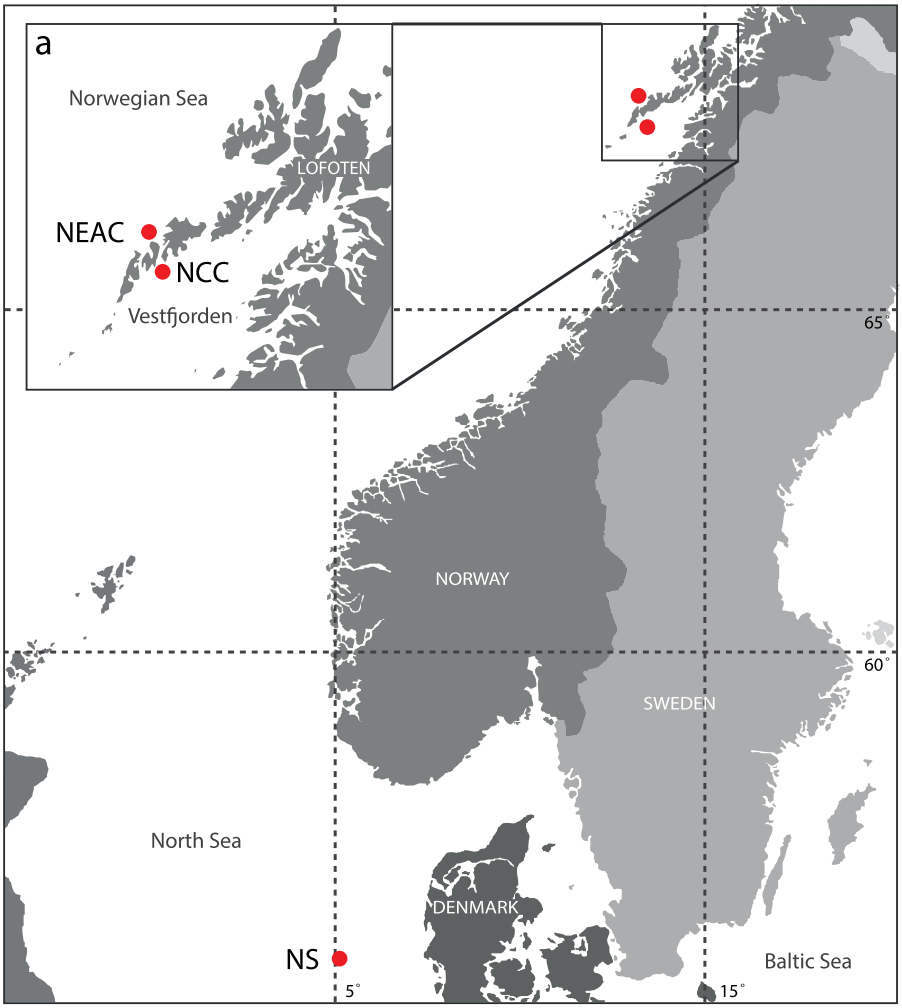

Figure 1. Map of sampling locations for the three Atlantic cod populations used in this study. Red dots indicate the position where the samples were collected. The intersect (**a**) show a detailed view of the Lofoten area. NEAC = Northeast Arctic cod, NCC = Norwegian coastal cod, NS = North Sea cod. See Table 1 for a detailed description of the samples. The map was modified from http://www.graphic-flash-sources.com/europe-free-vector-map/ using Adobe Illustrator CS5.

Based on all SNPs, no private alleles were detected in any of the populations (Supplementary Table S3), although 52 SNPs that were candidates for selection were fixed or close to fixation in the NEAC population (allele frequency >0.95, Supplementary Table S3). Distinctly different patterns of heterozygosity within the different LGs were detected among the populations (Fig. 4, Supplementary Table S3), which correspond well with the areas of high F_{ST} values (Fig. 3) as well as the identified outlier regions (Fig. 2). The exact tests for deviation from Hardy-Weinberg equilibrium (HWE) shows that only one SNP was out of HWE ($q < 0.05$; Supplementary Table S3), indicating no presence of a Wahlund effect. The number of polymorphic loci, observed- and expected heterozygosity (H_o and H_e) were similar in all populations (Table 1).

Population assignment tests using all 8.168 SNPs correctly assigned 136 of the 141 individuals (96.5%) to their presumed source populations (Supplementary Table S4). The miss assignment is due to one individual in the NEAC sample collection assigning as NCC and four individuals in the NCC sample collection assigning as NEAC. The same assignment pattern is obtained using the neutral SNPs (Supplementary Table S4).

Bayesian cluster analyses as implemented in STRUCTURE support a separation between all three populations using the full dataset and the outlier dataset, while little separation between the NEAC and the NCC populations was detected using the neutral dataset (Supplementary Fig. S1). The DAPC analysis, using all SNPs, confirms this population structure (Supplementary Fig. S2a). Moreover, the structure in these data is driven by the three regions within LG1, 2 and 7, according to the DAPC loading plots (Supplementary Fig. S2b).

Linkage disequilibrium patterns and chromosomal rearrangements. A substantial number of SNPs in high LD are detected within LG1, 2, 7 and 12 (Fig. 5a–d, Supplementary Table S5) and the LD pattern is distinctly different between the NEAC and the two other populations in LG1 and 12 (Fig. 5a,d, Supplementary Fig. S3). The LD analyses also reveal eight smaller regions of high LD (Table 3, Supplementary Fig. S3). By using the *R* package inveRsion, the linked regions in LG1, 2, 7 and 12 were suggested as inversions (Supplementary

Sampling ID	Sampling time	Lat.	Long.	Condition	Sample size	Ind. call no. >0.95	Avg. call rate	# Poly-morphic loci	*Ho* (s.d.)	*He* (s.d.)
Northeast Arctic cod (NEAC)	Mar 2011	N68.19	E13.30	Adults, spawn.	51	51	0.983	8113	0.354 (0.149)	0.353 (0.139)
Norwegian coastal cod (NCC)	Jun/Jul 2011	N68.04	E13.41	Adults/juv.	48	48	0.995	8121	0.364 (0.140)	0.369 (0.128)
North Sea cod (NS)	Mar 2002	N55.60	E05.85	Adults, spawn.	42	42	0.982	7953	0.367 (0.144)	0.365 (0.133)

Table 1. Atlantic cod samples included in this study and basic population genetic parameters. Estimates of observed (H_o) and expected heterozygosity (H_e) were calculated using ARLEQUIN[54]. s.d. = standard deviation, Latitude and longitude values are given in degrees and minutes. For sample details on each of the individuals, see Supplementary Table S4.

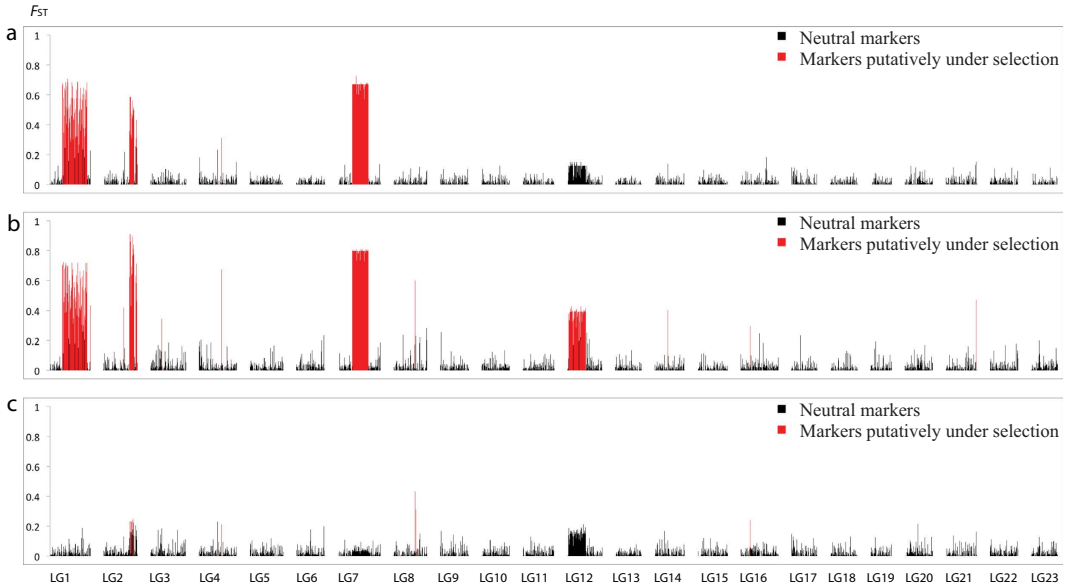

Figure 2. Locus specific F_{ST} values for the pairwise population comparisons between Northeast-Arctic, Norwegian coastal and North Sea cod. (**a**) The observed F_{ST} pattern indicates three distinct regions of the genome with elevated F_{ST} values in the NEAC-NCC comparison. (**b**) In the comparison between NEAC and the geographically more distant NS population, four distinct regions with elevated F_{ST} values are observed. (**c**) Only smaller areas of the genome shows elevated F_{ST} values in the NCC-NS comparison even though they are collected from geographically distant locations. SNPs are ordered according to linkage group and position within linkage groups. SNPs that are identified as putatively under selection are in red color. NEAC = Northeast Arctic cod, NCC = Norwegian coastal cod, NS = North Sea cod.

F-statistics			
	Full dataset (8.168 SNPs)	Neutral dataset (7.384 tag-SNPs)	Outlier dataset (86 tag-SNPs)
NEAC/ NCC	0.04246	0.00123	0.35053
NEAC/ NS	0.06237	0.00861	0.37502
NCC/ NS	0.00839	0.00519	0.01410

Table 2. Pairwise F_{ST} values among Atlantic cod populations, using full-, neutral- and outlier datasets. NEAC = Northeast Arctic cod. NCC = Norwegian coastal cod (Lofoten). NS = North Sea cod. All F_{ST} values are significant values (p-values < 0.0000. 10.000 permutations used) calculated in ARLEQUIN[54].

Fig. S4) and the identified breakpoints correspond well with the identified boundaries for the blocks in high LD (Table 3). Different genotypic combinations of the linked alleles at LG1, 2, 7 and 12, contribute to the observed population divergence (Table 3, Fig. 6), which support the hypothesis that these regions are chromosomal rearrangements. By defining the identified LD areas as regions of interest in the InvClust package, also the smaller regions of high LD are suggested as inversions (Supplementary Fig. S4) but they do not show a population based allele distribution between NEAC and NCC (Table 3, Supplementary Fig. S5).

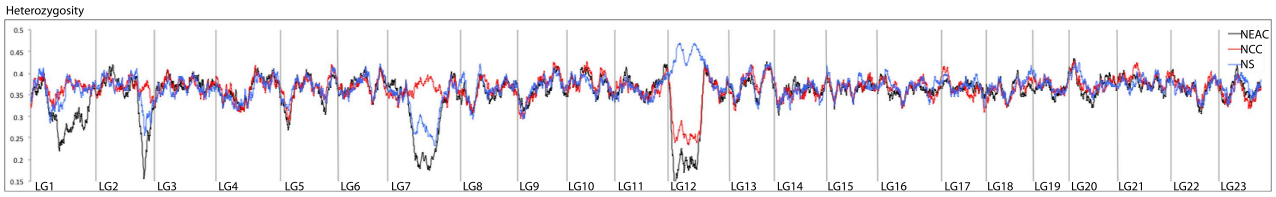

Figure 3. Heterozygosity level across all linkage groups in the Northeast-Arctic, Norwegian coastal and North Sea cod. SNPs are ordered according to linkage group and position within linkage groups. The observed heterozygosity pattern shows four distinct regions of the genome with distinctly different heterozygosity values. NEAC = Northeast Arctic cod, NCC = Norwegian coastal cod, NS = North Sea cod.

The identified LD block on LG1 covers at least 18.5 Mb containing 785 genes (Supplementary Table S6) and shows a distinctly different LD pattern in the NEAC and NCC populations (Fig. 5a). Two smaller and separate LD blocks towards the end of LG2 were identified that cover approximately 5 Mb containing 189 genes (Supplementary Table S6). The identified LD block on LG7 covers at least 9.5 Mb and 297 genes (Supplementary Table S6). In these two latter LGs, the LD pattern is similar in both populations (Fig. 5b,c). The SNPs under selection in all of these there LGs, fall within the identified regions of high LD. Combined, the outlier regions in LG1, 2 and 7 cover more than 33Mb (≈4% of the genome) and contain more than 1.200 genes (Supplementary Table S6). In LG12, we observe a pattern where the identified LD block covers at least 12.5 Mb and the LD pattern is distinctly different in the NEAC population relative to the other two populations (Fig. 5d). Whereas, no significant outliers between NEAC and NCC were detected in LG12, outliers spanning the entire linked region were detected between NEAC and the physically more distant NS population.

Relatively high values of LD ($r^2 > 0.3$) occur inter-chromosomally between SNPs on different LGs (Supplementary Table S5). Notably, the outlier SNPs in the LD blocks in LG1, 2 and 7 has r^2 values between 0.3 and 0.65 (Supplementary Fig. S6), and the linkage disequilibrium between the divergent blocks in LG1 and LG7 is significant (Fishers' exact test; $p = 0.0199$).

Discussion

Here we provide new insight into the evolution of distinct ecotypes of Atlantic cod with different life history strategies. We identify a set of three divergent regions between NEAC and NCC that combined span approximately 4% of the genome. The sizes of these regions, in combination with strong linkage patterns, distinct F_{ST} pattern and a population specific distribution, suggest that these genomic islands of divergence are the results of chromosomal rearrangements. This divergence is in contrast to the remaining parts of the genome, which is characterized by low levels of genomic divergence. Overall, the data indicate a key role for several chromosomal rearrangements in protecting adaptive loci from recombination[13] and hence facilitating adaptive genomic divergence.

Genomic islands of divergence have previously been reported in Atlantic cod[4,30,34–36] and have been discussed in the context of divergence hitchhiking[4,35]. Genomic islands of divergence can also be caused by factors that reduce recombination across the genome such as chromosomal rearrangements where distinct LD blocks will contain the entire rearrangement. Chromosomal rearrangements reduce the rate of crossing over by several orders of magnitudes[37] and may affect large genomic areas[38,39]. As a result, chromosomal rearrangements allow genomic islands of divergence to be larger than in collinear regions[14,40]. Our observations of three distinct and large genomic islands in the NEAC/NCC comparison and the additional large island in the NEAC/NS comparison are in high LD with each other throughout the entire block (Supplementary Table S5). This, in combination with population-specific distribution of the haplotype blocks, suggests that these regions are chromosomal rearrangements (Table 3), possibly large inversions, containing outlier SNPs nested within the regions. A plausible explanation for the divergence in the rearranged loci between the ecotypes, considering the low level of divergence in the remaining parts of the genome is that loci within the rearrangements are indicative of strong adaptive divergence. This is in line with Sodeland *et al.*[11], where rearrangements in LG2, 7 and 12 are identified within coastal and offshore samples of Atlantic cod on the Norwegian Skagerrak coast. Given that the observed rearrangements are inversions, sets of genes involved in local adaptation are either captured and protected from recombination within the inversion[13] or the position of the inversion breakpoints are giving an evolutionary advantage by changing reading frames or expression patterns[41]. To discriminate between these two explanations require further studies of the inversion breakpoints in multiple populations, using either a much denser SNP chip or a whole genome sequencing approach. Either way, the haplotypes detected here, likely provide selective advantage and are protected by the inversion. The fact that, even though there are population specific variation, all inversion haplotypes are in Hardy-Weinberg equilibrium with no heterozygote deficiencies indicate that there are no genomic disadvantages associated with the heterozygote variant.

Reduced recombination rates at chromosomal centromeres could also explain increased differentiation in localized parts of the genome[42], but does not explain the observed difference in genomic divergence within these regions (Fig. 3, Table 3). An alternative explanation for the genomic islands of divergence is divergence hitchhiking. In such a scenario, LD is expected to gradually decrease with distance from the target of selection[43] and as a consequence of recombination events, islands of divergence are expected to be relatively small. Large islands of divergence, as the ones observed here, can potentially be observed if there are several targets of selection within the genomic island[12,44] in addition to sequential buildup of divergence around the targets of selection[45]. However, as LD seems to be persistently high throughout the genomic islands of divergence, this seems to be a less plausible explanation for our results.

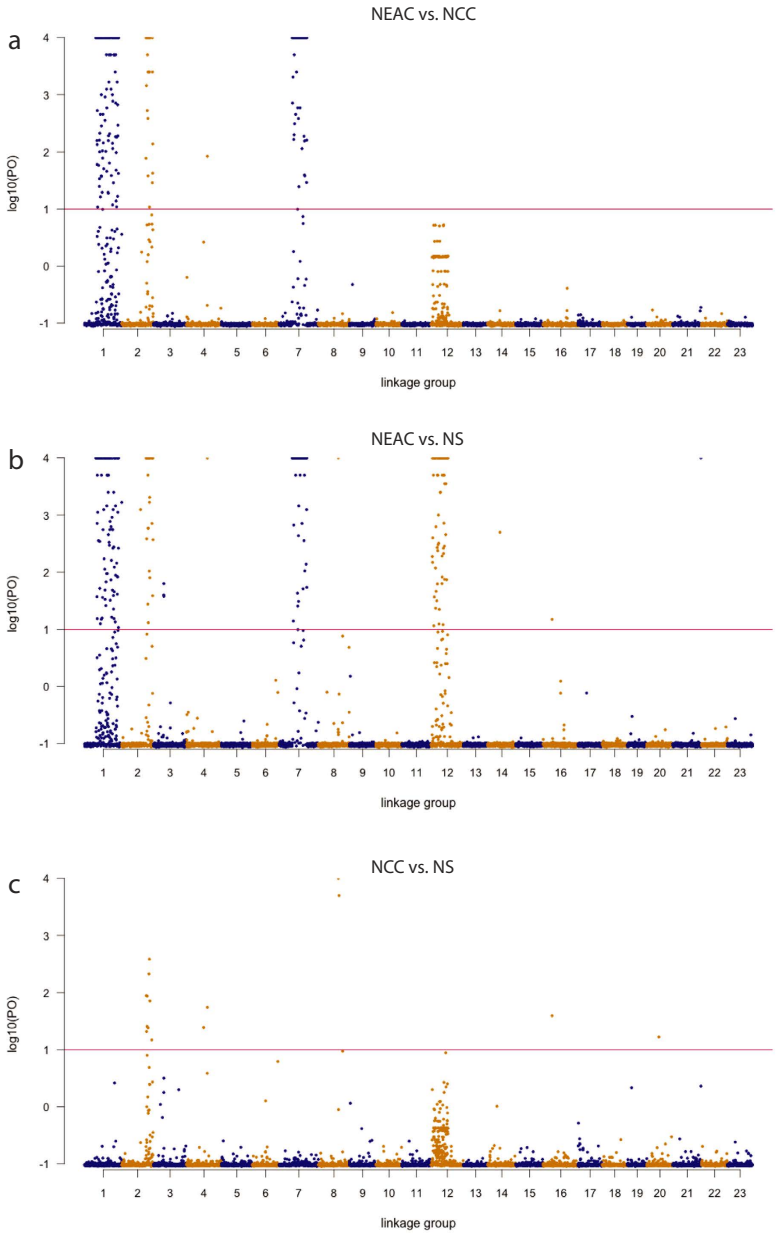

Figure 4. Manhattan plot of pairwise outlier analyses based on median log_{10} (PO) from 10 replicate runs of BAYESCAN. (**a**) The observed outlier pattern between NEAC and NCC indicate that the outliers are clustered within three distinct genomic areas. Only one additional outlier is detected in LG4. (**b**) In the comparison between NEAC and the geographically more distant NS population, an additional outlier area is observed in LG12. (**c**) In the NCC-NS comparison, the outlier area in LG2 is observed, but with lower log_{10} (PO) values. SNPs are plotted according to linkage group and position within the linkage groups along the X-axis. The red line at 1 indicates 'strong association' according to Jeffreys[73]. NEAC = Northeast Arctic cod, NCC = Norwegian coastal cod, NS = North Sea cod. For visualization purpose, maximum log10(PO) values are set to 4 and all underlying values are found in Supplementary Table S1.

The divergent and potentially inverted region in LG1 (Fig. 2a) shows distinct population based frequency distribution (Table 3, Fig. 6). The region covers at least 31 cM according to the linkage map by Hubert *et al.*[32] and corresponds well with the 23 cM genomic region associated with a migratory ecotype, defined by Hemmer-Hansen *et al.*[30]. Interestingly, this region was not detected as significantly divergent between eastern and western Atlantic cod populations by Bradbury *et al.*[36], indicating that the selecting agents on this genomic area may be similar on both sides of the Atlantic Ocean or that the inversion event happened before the split between East- and West Atlantic cod populations, approximately 100.000 years ago[46]. Further research on the inversion status of West Atlantic cod samples is needed to unravel this phenomenon. However, the divergent area is tightly linked (Fig. 5a) and displays low heterozygosity (Fig. 3) and low nucleotide diversity (Table 3) in the NEAC population but not so

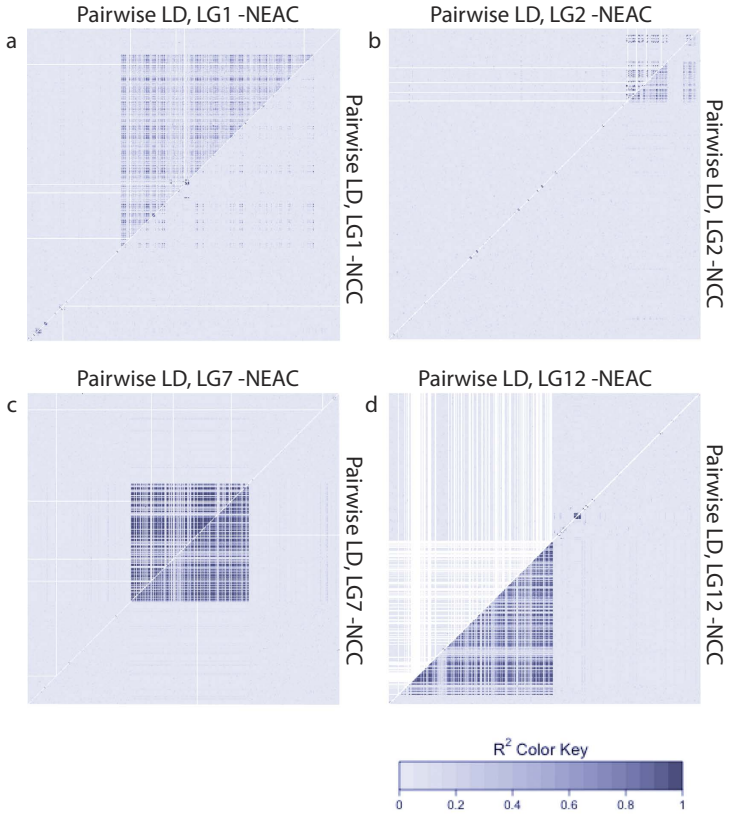

Figure 5. Linkage disequilibrium in linkage group 1, 2, 7 and 12. Pair-wise LD among loci, measured as r^2, are estimated within the Northeast-Arctic cod (above the diagonal) and Norwegian coastal cod (below the diagonal) populations. (**a, d**) A distinctly different LD pattern is observed between NEAC and NCC in LG1 and 12. (**b, c**) The LD pattern is similar within LG2 and 7 between NEAC and NCC. Corresponding patterns for all other linkage groups and for North Sea cod are shown in Supplementary Fig. S3 while the underlying LD measurements are found in Supplementary Table S5. NEAC = Northeast Arctic cod, NCC = Norwegian coastal cod.

in the NCC population, suggesting that the NEAC population is dominated by an inversion haplotype of a more recent origin than the other two populations, which presumably are dominated by a collinear haplotype (Table 3). Alternatively, a scenario of recent selective sweep may have caused the reduced heterozygosity and nucleotide diversity within the inversion haplotype in the NEAC population. In LG2 and 7, the divergent regions, which also shows distinct population differentiation (Table 3, Fig. 6), coincide with the regions detected by Bradbury *et al.*[36] and Hemmer-Hansen *et al.*[30] on both sides of the Atlantic Ocean and by Berg *et al.*[4] in the Baltic Sea. These regions have been associated with variation in temperature[36] and/or salinity and oxygen level[4]. The fact that similar regions are found to be divergent in a wide range of populations across the Atlantic Ocean, indicates that these regions could be predating the split between West- and East Atlantic cod populations and that they potentially play a role in local adaptation in multiple ecological settings. Interestingly, balancing selection has recently been described in the *Ckma* gene[47], which is nested within the divergent region in LG7. The authors suggest that selection could be acting on a larger part of the genome that is locked together by structural variation, which is consistent with our observations. In LG12, the identified divergent region that has been shown to be associated with temperature in two previous studies[4,36], are present on both sides of the Atlantic Ocean and have recently been used to discriminate between fjord and coastal samples of Atlantic cod on the Norwegian Skagerrak coast[11]. Even though this divergent region differentiates between all investigated populations (Table 3, Fig. 6), it is not identified as an outlier region in the NEAC/NCC comparison (Fig. 2) and is likely not responsible for the differentiation between migrating and non-migrating ecotypes *per se*. Interestingly, low heterozygosity was detected in both NEAC and NCC (Fig. 3) but low nucleotide diversity (Table 3) was only detected in NEAC. At the same time, increased heterozygosity and nucleotide diversity was observed in the NS population (Fig. 3, Table 3) indicating that the NEAC and NCC variant could be derived from the NS variant as it only captures a fraction of the genetic variability. In addition to the rearrangements in LG1, 2, 7 and 12, smaller regions of high LD were also suggested as rearrangements in other parts of the genome (Table 3, Supplementary Fig. S5). Although these areas are not under selection and do not contribute to population divergence between NEAC and NCC, their presence shows that such rearrangements can persist in the genome (and may be targets of selection in other populations). Moreover, the detection of these regions shows that it is not the selection *per se* that allows their identification, but rather their specific genomic properties.

LG	LD area* (SNP no.)	Rearrangement frequencies									Genotypic differentiation			Nucleotide diversity (π)				Identified breakpoint
		NEAC			NCC			NS			NEAC/ NCC	NEAC/ NS	NCC/ NS	NEAC	NCC	NS	All	
		AA	AB	BB	AA	AB	BB	AA	AB	BB								
1	137–417	0.02	0.49	0.49	0.85	0.17	0.00	0.98	0.02	0.00	**0.000**	**0.000**	0.139	0.24	0.35	0.33	0.39	134–417
2	751–837	0.96	0.04	0.00	0.19	0.44	0.37	0.00	0.14	0.86	**0.000**	**0.000**	**0.000**	0.21	0.36	0.25	0.37	749–835
4	1391–1413	0.16	0.51	0.33	0.29	0.50	0.21	0.36	0.50	0.14	0.170	0.012	0.626	0.31	0.31	0.26	0.28	
7	2539–2719	0.00	0.14	0.86	0.50	0.42	0.08	0.75	0.25	0.00	**0.000**	**0.000**	0.022	0.19	0.34	0.22	0.43	2537–2720
10	3660–3690	0.27	0.51	0.22	0.23	0.56	0.21	0.30	0.43	0.27	0.965	0.837	1.000	0.41	0.40	0.41	0.39	
11	3890–3909	0.49	0.37	0.14	0.33	0.48	0.19	0.43	0.32	0.25	0.246	0.359	1.000	0.40	0.40	0.38	0.39	
12	4250–4462	0.98	0.02	0.00	0.77	0.17	0.06	0.36	0.48	0.16	**0.004**	**0.000**	**0.003**	0.17	0.31	0.43	0.34	4248–4444
17	6047–6069	0.63	0.33	0.04	0.37	0.46	0.17	0.70	0.25	0.05	0.015	0.667	**0.007**	0.40	0.43	0.39	0.41	
19	6639–6650	0.53	0.29	0.18	0.46	0.37	0.17	0.25	0.55	0.20	0.920	0.095	0.213	0.39	0.39	0.41	0.40	
20	6877–6890	0.63	0.35	0.02	0.50	0.40	0.10	0.45	0.48	0.07	0.190	0.112	1.000	0.39	0.39	0.35	0.34	
21	7203–7222	0.29	0.55	0.16	0.33	0.46	0.21	0.39	0.36	0.25	1.000	0.864	1.000	0.38	0.41	0.40	0.39	
23	7991–8005	0.43	0.45	0.12	0.46	0.44	0.10	0.48	0.43	0.09	0.920	0.667	1.000	0.41	0.42	0.39	0.40	

Table 3. Chromosomal rearrangements in Atlantic cod, their contribution to population genetic structure and nucleotide diversity within these regions. NEAC = Northeast Arctic cod, NCC = Norwegian coastal cod, NS = North Sea cod. AA = frequency of the least common rearrangement variant in the total material, BB = frequency of the most common rearrangement variant in the total material. All rearrangements are detected with InvClust[62]. Rearrangements in LG1, 2, 7 and 12 are also detected with InveRsion[63]. Identified breakpoints are from InveRsion, where left (min) and right (max) values are used. As InvClust does not detect breakpoints, values are missing for rearrangements that are only identified with InvClust. Genotypic differentiation is given in *q*-values. Significant values ($q < 0.01$) are written in bold text. Nucleotide diversity (π) is calculated in DnaSP[58]. MDS plots for all rearrangements are given in Supplementary Fig. S4. *Identified LD area covering more than 10 SNPs. SNP no. corresponds to the SNP order in Supplementary Table S1.

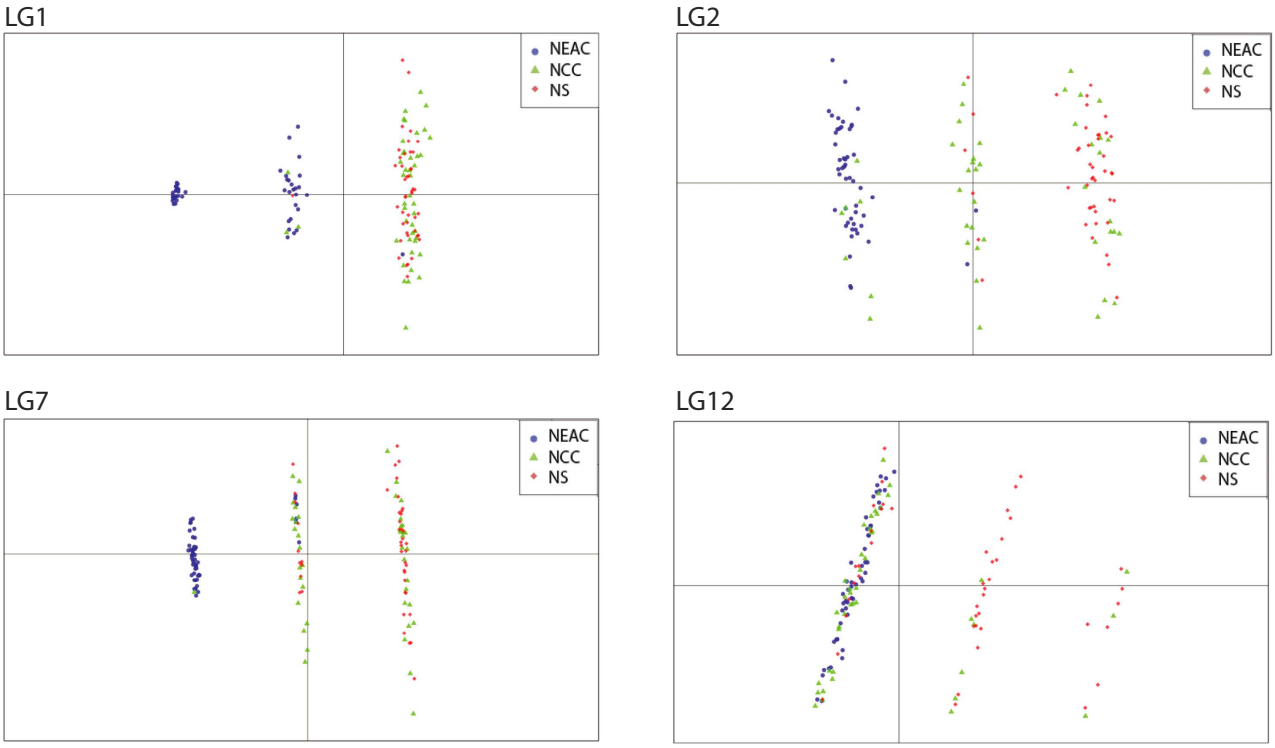

Figure 6. The population structuring within the rearranged regions in LG1, 2, 7 and 12 in Atlantic cod. The first two prinsipal components obtained from PCA of the NEAC, NCC and NS populations, using markers within the rearranged regions in the respective LGs. Each dot represents an individual and the left and right hand clusters represents the homozygotes while the middle cluster represents the heterozygotes. Corresponding patterns for all other linkage groups are shown in Supplementary Fig. S5. NEAC = Northeast Arctic cod, NCC = Norwegian coastal cod, NS = North Sea cod.

As both neutral and selective forces shape the genetic makeup among populations, it is important to disentangle these effects. Even though the majority of the investigated SNPs shows low levels of genetic differentiation (Fig. 3a), large F_{ST} differences (Table 2) and genomic regions under selection (Fig. 2a) are observed between the NEAC and the NCC populations, suggesting adaptive genomic divergence. F_{ST} values based on the unlinked neutral SNP set indicate that there is a low but significant neutral genetic differentiation between NEAC and NCC (Table 2), confirming that these are truly biological distinct populations. The finding of significant neutral divergence between NEAC and NCC supports the 'historical isolation hypothesis' rather then the 'divergent selection hypothesis' (both described in[20]), where historic differentiation between NEAC and NCC, presumably in allopatry, allows for these two populations to occupy the same spawning habitats without interbreeding.

Importantly, the well-studied pantophysin gene (*Pan* I) which have been used to determine individuals as either stationary or migratory[22,48] is only one out of approximately 785 target genes in the outlier region on LG1, but could be used as a proxy for the region under selection as a whole. The estimated frequency, based on the inversion status of the entire divergent region (Table 3), corresponds well with the allele frequency of the investigated *Pan* I locus (Supplementary Table S1) and are somewhat higher than, but still in line with frequencies of the *Pan* I locus in the initial study by Fevolden and Pogson[22]. Since this divergent area in LG1 appears to be inherited as one large rearranged region, it is not necessarily indicating that the *Pan* I locus is under selection. Other genes within the same region have also been described as likely targets of selection, such as rhodopsin[49] and further research is needed to unravel the actual targets of selection.

A natural next step to unravel the history of these and potentially other chromosomal rearrangements is to investigate if the exact locations of inversion breakpoints are shared across a diverse group of Atlantic cod populations. Such analyses would require dense SNP coverage or preferably a whole genome sequencing approach. Our study corroborate by and extent previous work on marine populations, where natural selection shapes the population structure on short spatial scales, despite the high dispersal capacity of these marine organisms[50–53]. The findings reported here show that the majority of the Atlantic cod genome shows little genetic differentiation between NEAC and NCC, apart from 3 distinct genomic regions that are likely the results of chromosomal inversions, maintained under strong diversifying selection. Hence, any linked genetic markers within the respective areas could be used as proxies for the inverted regions as a whole in population based studies.

Materials and Methods

Sample collection, DNA extraction and genotyping. We collected 99 Atlantic cod specimens near the Lofoten Islands at different sampling times (Table 1). Samples collected during spawning time at spawning grounds are here defined as the NEAC population that migrate from the Barents Sea to spawn in the Lofoten area. The samples collected in late June/early July after the spawning NEAC has left the area, are here defined as the resident NCC population. For comparative reasons, a physically more distant population, spawning in a different area, consisting of 42 individuals from the North Sea was included (Fig. 1, Table 1). All fish samples in this study were harvested for human consumption, from which small tissue samples were collected (post mortem). Sampling in this manner does not fall under any specific legislation in Norway, and is in accordance with the guidelines set by the 'Norwegian consensus platform for replacement, reduction and refinement of animal experiments' (www.norecopa.no).

DNA was extracted from ethanol stored muscle tissue, using the E.Z.N.A Tissue DNA kit (Omega Bio-Tek, Norcross, GA, USA) and normalized to 100 ng/μl measured on a NanoDrop DN1000 instrument (Thermo Scientific Inc.), accepting only DNA extractions with a 260/280 ratio >1.8 and a 260/230 ratio >2.0. All individual samples were genotyped using a 12 K Illumina SNP-chip. The 12 K SNP-chip was designed, based on re-sequencing and alignment to the Atlantic cod reference genome[33], of 8 globally collected individuals where all three populations in this study were represented. Out of the 10.913 SNPs on the final SNP-chip, 8.164 SNPs were polymorphic in the NEAC/NCC populations, had a call rate $>95\%$ and showed Mendelian inheritance in a set of >2000 family individuals (data not shown). Genotype clustering and pedigree check was performed using the Genome Studio 2011.1 software from Illumina, where each individual SNP locus was manually inspected and clusters were adjusted if appropriate. In addition, 4 SNPs (one hemoglobin SNP, two Rhodopsin SNPs and one *Pan* I SNP) were genotyped on the MassARRAY (Sequenom Inc.), resulting in a final number of 8.168 SNPs with a minor allele frequency (MAF) >0.05 in any population. Out of these, 602 SNPs were close to selected candidate genes, 1.470 SNPs were non-synonymous coding SNPs while the remaining 5.857 were SNPs randomly distributed throughout the 23 different LGs (the source and selection criteria of each SNP is listed in Supplementary Table S1). The nomenclature of LGs in this paper follows Hubert *et al.*[32] while the order of the SNPs are based on preliminary linkage data as in Berg *et al.*[4].

Population genetics, linkage disequilibrium and rearrangement patterns. Within each population, estimates of observed (H_o) and expected heterozygosity (H_e) were calculated in ARLEQUIN 3.5.1.3[54]. Departure from Hardy-Weinberg Equilibrium (HWE) was tested locus by locus in each population using the exact test in ARLEQUIN with 100.000 iterations and a Markov Chain of 1.000.000. Correction for multiple testing were done by computing the q-value for each locus, using a q-value of 0.05 as a threshold for significance, using the QVALUE package[55] in *R*[56]. Allele frequencies were calculated for all SNPs in all three populations using ARLEQUIN.

SNPs were categorized as outliers or as neutral, based on the outlier analyses in the NEAC and the NCC material (see next section). To avoid bias in the datasets, neutral and outlier SNPs in LD with each other ($r^2 > 0.5$) were represented by tag SNPs, selected using PLINK v1.07[57] and used in the F_{ST} and STRUCTURE analyses. Locus specific F_{ST} values and weighted average F_{ST} values between the three populations were calculated for the full-, the neutral- and the outlier datasets, using ARLEQUIN. For all comparisons, 10.000 permutations were used. We

calculated the nucleotide diversity (π) within all populations and in all populations combined, using a sliding windows approach with a 50-SNP window and 10 SNPs per iteration in DnaSP 5.10[58].

We used the program STRUCTURE v2.3.4[59] to identify major genetic clusters in the dataset, using the correlated allele frequency and admixture model to best reflect the most likely pattern of population connectivity. We performed 10 independent runs for each value of *K*, with $K = 1–4$ (burn in of 10.000 MCMC iterations followed by 100.000 MCMC iterations). Visualization and evaluation of the best *K*-value for the individual STRUCTURE runs was performed, using CLUMPAK[60]. We performed discriminant analysis of principal components (DAPC, using all 8.168 SNPs in the *R* package ADEGENET[61]. DAPC is a method that relies on data transformation using principal component analysis (PCA) prior to the discriminant analysis (DA) step, ensuring that variables in the DA analysis are uncorrelated. Further, loadingplots from DAPC was used to identify the main SNPs that are driving the genetic divergence among the three populations.

We evaluated the presence of linkage disequilibrium (LD) in all three populations separately using all 8.168 SNPs, reporting both inter- and intra- chromosomal LD, quantified with the r^2 estimate, using PLINK. Different approaches were used to investigate rearrangement patterns in the data. We used the *R* package inveRsion[62], which is based on LD differences across inversion breakpoint, to detect and locate large inverted genomic regions and to identify the inversion status of each individual, using block size = 3, min. allele = 0.1 and thbic = 0. By defining regions of interest, based on the LD analyses, we also used the *R* package invClust[63], which is based on haplotype tagging and dimensionality reduction analysis, to assess if the identified LD regions in the Atlantic cod genome were likely to be inversions, visualized as a three-cluster pattern in the first component in a multidimensional scaling (MDS) analysis. This method was also used to identify the inversion status of all individuals within each inverted region. Finally, a method described by Ma[64], where PCA is performed locally within the identified inversions, was used to confirm and visualize the distribution of individuals to their inversion status, based on their population of origin.

Assignment testing and outlier detection. Assignment of all individuals to their presumed source populations was estimated using the Bayesian assignment method in GENECLASS2[65], employing the 'leave-one-out' procedure. Based on the assignment tests, using all 8.168 SNP markers, a dataset for the outlier analyses was defined. This dataset contains only individuals from the source populations of NEAC and NCC (where 5 miss-assigned individuals were excluded) in addition to 42 NS samples, hence containing 136 individuals.

Two independent methods were used to identify candidate loci under selection in all population pairs. First, we used a Bayesian regression approach implemented in BAYESCAN v2.1[66] which, measures the discordance between global and population-specific allele frequencies, based on F_{ST} coefficients. To control for variation in the Bayes Factor (BF) distribution caused by randomness in each run of BAYESCAN, the median value of 10 independent runs were calculated for each SNP. We carried out the simulations, using stringent criteria, assuming selection to be 10%. The false discovery rate (FDR) was set to 0.01. We report both the median $\log_{10}$ value of the posterior odds (PO) as well as the *q*-value for each SNP in all pairwise comparisons. Second, we used the FDIST2 approach implemented in the software LOSITAN[67], where comparisons are made of F_{ST} values in relation to heterozygosity of individual loci, based on a neutral distribution. We carried out the simulations, under the Infinite Allele Method (IAM) with 1.000.000 simulations, a confidence interval of 0.99 with a FDR of 0.01; using the "neutral" mean F_{ST} option and forcing mean F_{ST} option. Based on the probabilities calculated in LOSITAN, *q*-values were calculated and reported for all pairwise comparisons, using the QVALUE package in *R* (using a *q*-value of 0.01 as a threshold for significance). To corroborate our choice of cutoff value, and to be able to use a consistent cutoff value in both the BAYESCAN and the LOSITAN analyses, we used both *q*-values AND cutoff values based on Log_{10} (PO) in determining the significance level in both outlier tests. The consistency of the different approaches for outlier detection and the strength of the identified outlier loci, strongly suggest that the majority of the identified loci and their associated genomic regions are subject to divergent selection. Nevertheless, some outlier loci are only detected by a single approach. This variation can either be caused by different underlying assumptions and detection rates of BAYESCAN and LOSITAN or slightly different cutoff criteria used[68]. It has been suggested that outlier tests may have high false positive rates due to the effects of population demography and bottleneck effect[69]. Even though this is less likely in our case due to presumably large population sizes and shallow neutral population structuring, we performed outlier analyses between pairs of populations, partly omitting the methodological weakness of population structure in the datasets[70].

SNP annotation. All 8.168 SNPs used, were mapped to the published Atlantic cod genome (ATLCOD1C)[33], for which Ensembl annotation is available, in the same way as in Berg *et al.*[4]. SNPs located either within a gene or located within a 5000 bp region up or downstream of a gene were identified using BEDclosest[71] with the option -t "first" and -d to determine distance. Further, protein transcripts of Ensembl genes that were associated with the location of the SNPs through this approach were annotated with BLAST2GO[72] using public database b2g_sep13. Protein transcripts were aligned to the refseq_protein data using the BlastP algorithm in BLAST2GO, allowing a maximum of 20 hits with a minimum e-value of 1E-3. Apart from setting the evidence code weight of IEA (electronic annotation evidence) to 1, default weights were used. Annotation was augmented using the Annex function in BLAST2GO. All SNPs are referred to by their ss# or rs# available in dbSNP (www.ncbi.nlm.nih.gov/SNP/). All raw data are provided in Supplementary File S1 in PLINK format, where the .ped file contains all genotypes for all individuals and all 8168 SNP markers and the .map file contains all SNP names.

References

1. Hohenlohe, P. A. *et al.* Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *Plos Genet* **6,** e1000862 (2010).
2. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484,** 55–61 (2012).

3. Nadeau, N. J. *et al.* Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Phil Trans R Soc B* **367,** 343–353 (2012).
4. Berg, P. R. *et al.* Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (*Gadus morhua* L.). *Genome Biol Evol* **7,** 1644–1663 (2015).
5. Wu, C.-I. The genic view of the process of speciation. *J. Evol. Biol.* **14,** 851–865 (2001).
6. Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. Divergent selection and heterogeneous genomic divergence. *Mol Ecol* **18,** 375–402 (2009).
7. Ellegren, H. *et al.* The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491,** 756–760 (2012).
8. Nosil, P. & Feder, J. L. Genomic divergence during speciation: causes and consequences. *Phil Trans R Soc B* **367,** 332–342 (2012).
9. Guo, B., Defaveri, J., Sotelo, G., Nair, A. & Merilä, J. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol* **13,** 19–19 (2014).
10. Turner, T. L. *et al.* Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nat Genet* **42,** 260–263 (2010).
11. Sodeland, M. *et al.* Islands of divergence' in the Atlantic cod genome are projections of polymorphic chromosomal rearrangements. *Genome Biol Evol* GBE–151112 1–17 doi: 10.1093/gbe/evv (2016).
12. Via, S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Phil Trans R Soc B* **367,** 451–460 (2012).
13. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173,** 419–434 (2006).
14. Feder, J. L., Gejji, R., Yeaman, S. & Nosil, P. Establishment of new mutations under divergence and genome hitchhiking. *Phil Trans Soc Lond B* **367,** 461–474 (2012).
15. Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M. & Hill, W. G. Measures of human population structure show heterogeneity among genomic regions. *Genome Research* **15,** 1468–1476 (2005).
16. Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. & De Meester, L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol Ecol* **22,** 5983–5999 (2013).
17. Reiss, H., Hoarau, G., Dickey-Collas, M. & Wolff, W. J. Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries* **10,** 361–395 (2009).
18. Hutchings, J. A. *et al.* Genetic Variation in Life-History Reaction Norms in a Marine Fish. *Proc R Soc B* **274,** 1693–1699 (2007).
19. Rollefsen, G. The otoliths of the cod. *FiskDir Skr HavUnders* **IV,** 1–14 (1933).
20. Nordeide, J. T., Johansen, S. D., Jørgensen, T. E., Karlsen, B. O. & Moum, T. Population connectivity among migratory and stationary cod *Gadus morhua* in the Northeast Atlantic—A review of 80 years of study. *Mar Ecol Prog Ser* **435,** 269–283 (2011).
21. Moller, D. Genetic differences begween cod groups in the Lofoten area. *Nature* **212,** 824 (1966).
22. Fevolden, S.-E. & Pogson, G. H. Genetic divergence at the synaptophysin (*Syp* I) locus among Norwegian coastal and north-east Arctic populations of Atlantic cod. *J Fish Biology* **51,** 895–908 (1997).
23. Brander, K. Spawning and life history information for North Atlantic cod stocks. *ICES Coop Res Rep* **274,** 1–152 (2005).
24. Hylen, A. Coastal Cod and Skrei in the Lofoten Area. *FiskDir Skr HavUnders* **13,** 27–42 (1964).
25. Jakobsen, T. Coastal cod in northern Norway. *Fish Res* **5,** 223–234 (1987).
26. Nordeide, J. Coastal cod and north-east Arctic cod-Do they mingle at the spawning grounds in Lofoten? *Sarsia* **83,** 373–379 (1998).
27. Nordeide, J. T. & Folstad, I. Is cod lekking or a promiscuous group spawner? *Fish and Fisheries* **1,** 90–93 (2000).
28. Vikebø, F., Jørgensen, C., Kristiansen, T. & Fiksen, Ø. Drift, growth, and survival of larval Northeast Arctic cod with simple rules of behaviour. *Mar Ecol Prog Ser* **347,** 207–219 (2007).
29. Fevolden, S.-E., Westgaard, J. I., Pedersen, T. & Præbel, K. Settling-depth vs. genotype and size vs. genotype correlations at the *Pan* I locus in 0-group Atlantic cod *Gadus morhua*. *Mar Ecol Prog Ser* **468,** 267–278 (2012).
30. Hemmer-Hansen, J. *et al.* A genomic island linked to ecotype divergence in Atlantic cod. *Mol Ecol* **22,** 2653–2667 (2013).
31. Karlsen, B. O. *et al.* Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Mol Ecol* **22,** 5098–5111 (2013).
32. Hubert, S., Higgins, B., Borza, T. & Bowman, S. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genomics* **11,** 191 (2010).
33. Star, B. *et al.* The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477,** 207–210 (2011).
34. Bradbury, I. R. *et al.* Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evol Appl* **6,** 450–461 (2013).
35. Bradbury, I. R. *et al.* Long distance linkage disequilibrium and limited hybridization suggest cryptic speciation in Atlantic cod. *Plos One* **9,** e106380 (2014).
36. Bradbury, I. R. *et al.* Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proc R Soc B* **277,** 3725–3734 (2010).
37. Feder, J. L., Nosil, P. & Flaxman, S. M. Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. *Front. Genet.* **5,** 295 (2014).
38. Noor, M. A., Grams, K. L., Bertucci, L. A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species. *PNAS* **98,** 12084–12088 (2001).
39. Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16,** 351–358 (2001).
40. Flaxman, S. M., Feder, J. L. & Nosil, P. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution* **67,** 2577–2591 (2013).
41. Puig, M., Cáceres, M. & Ruiz, A. Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA. *PNAS* **101,** 9013–9018 (2004).
42. Roesti, M., Hendry, A. P., Salzburger, W. & Berner, D. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol Ecol* **21,** 2852–2862 (2012).
43. Kim, Y. & Stephan, W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160,** 765–777 (2002).
44. Via, S. Natural selection in action during speciation. *Proc Natl Acad Sci USA* **106** Suppl 1, 9939–9946 (2009).
45. Feder, J. L., Egan, S. P. & Nosil, P. The genomics of speciation-with-gene-flow. *Trends Genet* **28,** 342–350 (2012).
46. Bigg, G. R. *et al.* Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. *Proc R Soc B* **275,** 163–172 (2008).
47. Árnason, E. & Halldórsdóttir, K. Nucleotide variation and balancing selection at the *Ckma* gene in Atlantic cod: analysis with multiple merger coalescent models. *PeerJ* **3,** e786–32 (2015).
48. Pogson, G. H. & Fevolden, S.-E. Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (*Pan*I) locus. *Mol Ecol* **12,** 63–74 (2003).
49. Pampoulie, C. *et al.* Rhodopsin gene polymorphism associated with divergent light environments in Atlantic cod. *Behav Genet* **45,** 236–244 (2015).
50. Knutsen, H., Jorde, P. E., André, C. & Stenseth, N. C. Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Mol Ecol* **12,** 385–394 (2003).
51. Defaveri, J., Shikano, T., Shimada, Y. & Merilä, J. High degree of genetic differentiation in marine three-spined sticklebacks (*Gasterosteus aculeatus*). *Mol Ecol* **22,** 4811–4828 (2013).

52. Lamichhaney, S. *et al.* Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proc Natl Acad Sci USA*. **109,** 19345–19350 (2012).
53. Corander, J., Majander, K. K., Cheng, L. & Merilä, J. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Mol Ecol* **22,** 2931–2940 (2013).
54. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* **10,** 564–567 (2010).
55. Storey, J. D. A direct approach to false discovery rates. *J R Statist Soc B* **64,** 479–498 (2002).
56. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2012). at http://www.R-project.org/.
57. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81,** 559–575 (2006).
58. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–1452 (2009).
59. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
60. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Res* **15,** 1179–1191 (2015).
61. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27,** 3070–3071 (2011).
62. Cáceres, A., Sindi, S. S., Raphael, B. J., Cáceres, M. & González, J. R. Identification of polymorphic inversions from genotypes. *BMC Bioinformatics* **13,** 28 (2012).
63. Cáceres, A. & González, J. R. Following the footprints of polymorphic inversions on SNP data: from detection to association tests. *Nucleic Acids Research* **43,** e53–e53 (2015).
64. Ma, J. & Amos, C. I. Investigation of inversion polymorphisms in the human genome using principal components analysis. *Plos One* **7,** e40224 (2012).
65. Piry, S. *et al.* GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered* **95,** 536–539 (2004).
66. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180,** 977–993 (2008).
67. Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A. & Luikart, G. LOSITAN: a workbench to detect molecular adaptation based on a *FST*-outlier method. *BMC Bioinformatics* **9,** 323 (2008).
68. Lotterhos, K. E. & Whitlock, M. C. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol* **23,** 2178–2192 (2014).
69. Narum, S. R. & Hess, J. E. Comparison of *F*ST outlier tests for SNP loci under selection. *Mol Ecol Res* **11** Suppl 1, 184–194 (2011).
70. Vitalis, R., Dawson, K. & Boursot, P. Interpretation of variation across marker loci as evidence of selection. *Genetics* **158,** 1811–1823 (2001).
71. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
72. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–3676 (2005).
73. Jeffreys, H. *Theory of probability 3rd ed.* (Oxford University Press Ed. New-York, 1961).

Acknowledgements

We thank the skipper on the fishing vessel Iversen Jr, Børge Iversen, for invaluable help in getting local knowledge on the fish stocks as well as for hosting us on his boat, providing cod samples. A special thanks also goes to Martin Malmstrøm for valuable fishing effort and to Andrew Snowdon at Nic Haug AS. Thanks also to Matthew P. Kent, Sigbjørn Lien and Mariann Arnyasi at Norwegian University of Life Sciences, CIGENE for SNP genotyping and access to a preliminary linkage map. Initial sequencing for SNP identification was provided by the Norwegian Sequencing Centre, University of Oslo. This work is part of the Cod SNP Consortium (CSC) activities – a collaboration between CIGENE, CEES, IMR and Nofima. Funding was provided by the Research Council of Norway to KSJ (grant numbers 199806 and 187940) and Interreg (MarGen) Öresund-Kattegat-Skagerrak.

Author Contributions

S.J., K.S.J., B.S. and P.R.B. conceived and designed the entire study. P.R.B. drafted the manuscript and analyzed the data. K.S.J., S.J. and B.S. coordinated the study and supervised the research. C.P., M.S. J.M.I.B. and H.K. contributed with valuable advice to the lead author. All authors contributed to the interpretation of the results, the writing of the manuscript and approved on the final 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: Berg, P. R. *et al.* Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci. Rep.* **6**, 23246; doi: 10.1038/srep23246 (2016).