ORIGINAL ARTICLE

# Genomic analysis reveals neutral and adaptive patterns that challenge the current management regime for East Atlantic cod *Gadus morhua L*

Torild Johansen[1] | François Besnier[2] | María Quintela[2] | Per Erik Jorde[3] | Kevin A. Glover[2,4] | Jon-Ivar Westgaard[1] | Geir Dahle[2,4] | Sigbjørn Lien[5] | Matthew P. Kent[5]

[1]Institute of Marine Research (IMR), Tromsø, Norway

[2]Institute of Marine Research (IMR), Bergen, Norway

[3]Institute of Marine Research (IMR), His, Norway

[4]Department of Biology, University of Bergen, Bergen, Norway

[5]Department of Animal and Aquaculture Sciences, Faculty of Biosciences, Centre for Integrative Genetics, Norwegian University of Life Sciences, Ås, Norway

**Correspondence**
Torild Johansen, Institute of Marine Research (IMR), Post box 6404, N-9019 Tromsø, Norway.
Email: torild.johansen@hi.no

## Abstract

Challenging long-held perceptions of fish management units can help to protect vulnerable stocks. When a fishery consisting of multiple genetic stocks is managed as a single unit, overexploitation and depletion of minor genetic units can occur. Atlantic cod (*Gadus morhua*) is an economically and ecologically important marine species across the North Atlantic. The application of new genomic resources, including SNP arrays, allows us to detect and explore novel structure within specific cod management units. In Norwegian waters, coastal cod (i.e. those not undertaking extensive migrations) are divided into two arbitrary management units defined by ICES: one between 62° and 70°N (Norwegian coastal cod; NCC) and one between 58° and 62°N (Norwegian coastal south; NCS). Together, these capture a fishery area of >25,000 km$^2$ containing many spawning grounds. To assess whether these geographic units correctly represent genetic stocks, we analysed spawning cod of NCC and NCS for more than 8,000 SNPs along with samples of Russian White Sea cod, north-east Arctic cod (NEAC: the largest Atlantic stock), and outgroup samples representing the Irish and Faroe Sea's. Our analyses revealed large differences in spatial patterns of genetic differentiation across the genome and revealed a complex biological structure within NCC and NCS. Haplotype maps from four chromosome sets show regional specific SNP indicating a complex genetic structure. The current management plan dividing the coastal cod into only two management units does not accurately reflect the genetic units and needs to be revised. Coastal cod in Norway, while highly heterogenous, is also genetically distinct from neighbouring stocks in the north (NEAC), west (Faroe Island) and the south. The White Sea cod are highly divergent from other cod, possibly yielding support to the earlier notion of subspecies rank.

**KEYWORDS**

chromosomes, haplotype maps, inversions, managements, White Sea cod

## 1 | INTRODUCTION

In marine fisheries, stocks or management units are often defined by national borders or economic zones that do not always reflect the true biological or genetic units (Kerr et al., 2016; Reiss, Hoarau, Dickey-Collas, & Wolff, 2009; Saha et al., 2015). Depletion and overexploitation of vulnerable stocks can occur where a fishery consisting of multiple genetic units or stocks is managed as a single unit (Allendorf, England, Luikart, Ritchie, & Ryman, 2008; Hauser, Adcock, Smith, Ramírez, & Carvalho, 2002; Ruzzante, Taggart, & Cook, 1999). Eroding a vulnerable stock may reduce its overall genetic variation and, in turn, impair the overexploited stocks potential to adapt to environmental change (Hauser et al., 2002). New genomic tools allow for a more accurate assessment of population genetic structure and have been successfully applied to redefine independent stocks and management units within fisheries (Hauser, Waples, & Carvalho, 2008; Reiss et al., 2009).

Atlantic cod (*Gadus morhua* L) is an economically important demersal fish distributed across most of the North Atlantic Ocean (Brander, 1995) where it occurs in both offshore and in coastal areas. Cod from the offshore and coastal areas are typically managed as a single stock, or the coastal cod are simply a subunit relative to the offshore stock component (Berg & Albert, 2003; Bradbury et al., 2013; ICES, 2019; Johansen et al., 2017; Kerr, Cadrin, & Kovach, 2014). However, genetic differences have been detected between offshore

and coastal cod in several regions, including the Gulf of Maine (Kerr et al., 2014; Kovach, Breton, Berlinsky, Maceda, & Wirgin, 2010), Eastern Canada (Bradbury et al., 2013; Ruzzante et al., 1999), Iceland (Berg et al., 2016; Pampoulie et al., 2012), Greenland (Pampoulie et al., 2011), the Faroe Islands (Nielsen, Hemmer-Hansen, et al., 2009) and Norway (Wennevik, Jørstad, Dahle, & Fevolden, 2008). Spies et al. (2018) and Kerr et al. (2016) have highlighted the beneficial effects of taking genetic structure into account when independently estimating the biomass of such stock components.

Norwegian offshore and coastal components of cod have been studied for decades (Nordeide, Johansen, Jørgensen, Karlsen, & Moum, 2011). The offshore and highly migratory Barents Sea cod (also called north-east Arctic cod or north-east Atlantic cod, abbreviated to NEAC) is the world's largest remaining cod stock (Garrod & Schumacher, 1994; Yaragina, Aglen, & Sokolov, 2011). NEAC spawn in Norwegian coastal waters during the spring, from Møre (62°N) in the south to the Russian border in the north (Olsen et al., 2010). Eggs and larvae resulting from NEAC spawning drift northwards to the Barents Sea where they grow and feed until maturity (>7 years), when they migrate back to the Norwegian coast to spawn (Brander, 1995; Johansen et al., 2017; Olsen et al., 2004). In contrast to long-migrating NEAC, Norwegian coastal cod are relatively stationary and are found on the Norwegian shelf and coastline and within fjords, throughout the year (Jakobsen, 1987; Michalsen, Johansen, Subbey, & Beck, 2014).
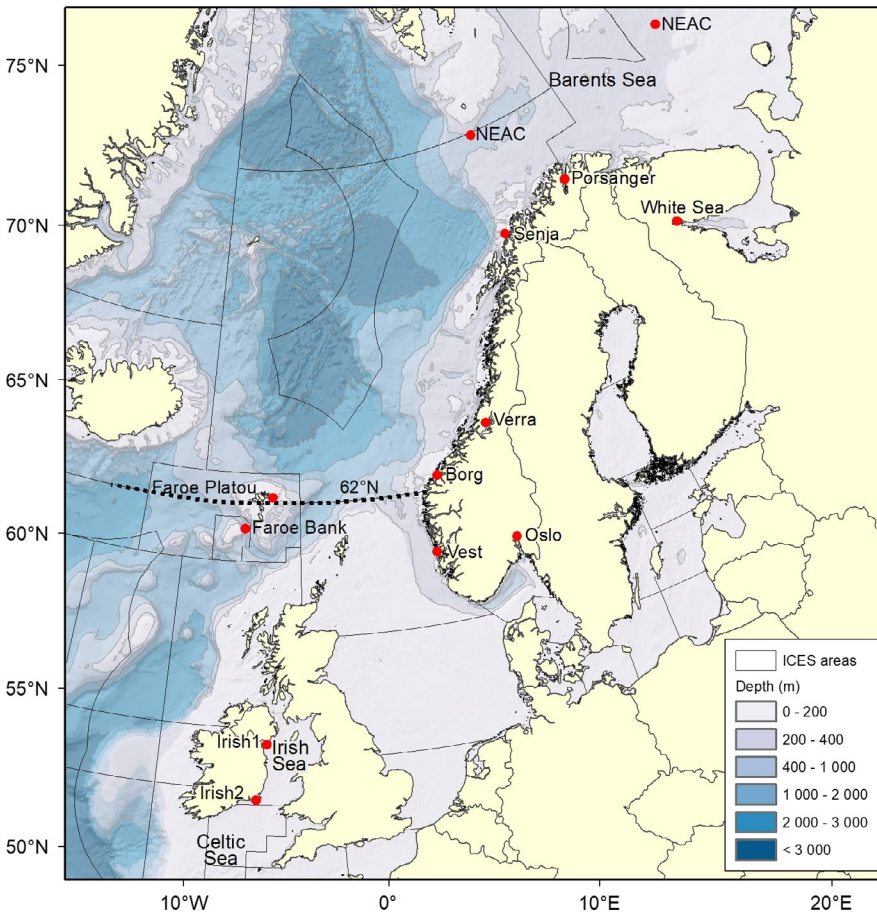


**FIGURE 1** Study area including sampling sites

**TABLE 1** Samples information: *N* corresponds to the initial number of individuals screened, whereas Nf corresponds to the numbers of individuals used for analysis once NEAC (based on otolith reading), and badly amplified samples were removed

| Management areas | Site | N | Nf | Sex (% F) | Age range (year) | Maturity stage (1–5) | Spawning fish |
|---|---|---|---|---|---|---|---|
| Barents Sea | NEAC | 48 | 47 | 0.4 | 1–8 | 1–5 | 0.13 |
| White Sea | White Sea | 48 | 47 | 0.73 | 2–4 | 2–5 | 1.00 |
| NCC (north of 62°N) | Porsangerfjord | 48 | 39 | 0.23 | 5–11 | 2–4 | 1.00 |
| | Senja | 48 | 42 | 0.42 | 1–13 | 1–2 | 0.13 |
| | Verrabotn | 48 | 37 | 0.13 | 3–9 | 2–4 | 1.00 |
| | Borgundfjord | 48 | 45 | 0.77 | 4–10 | 2–4 | 1.00 |
| NSC (south of 62°N) | Vest | 48 | 48 | 0.69 | 2–9 | 1–3 | 0.92 |
| | Oslo | 48 | 44 | 0.56 | 2–12 | 1–5 | 0.79 |
| Outgroups | Faroe Bank | 40 | 40 | NA | NA | NA | NA |
| | Faroe Plateau | 29 | 29 | – | – | – | – |
| | Irish 1 | 48 | 20 | – | – | – | – |
| | Irish 2 | 48 | 48 | – | – | – | – |

*Note:* Sex ratio, age and maturity (in years) are indicated as well as the proportion of spawning fish.

Norwegian coastal cod are currently managed as two separate units: north (Norwegian coastal cod: NCC) and south (Norwegian coastal south: NCS) of 62°N (Figure 1). The management of NCS is linked to the assessment of the North Sea cod (ICES, 2012). NCC have historically been exposed to a higher harvest pressure than NCS, and approximately 70% of the NCC harvest occurs during the spring spawning season (Johansen et al., 2017). Both NCC and NCS fisheries are presently regarded to be outside sustainable harvest limits and therefore in need of a thorough management revision (ICES, 2012, 2018). For both NCC and NCS, spawning typically occurs within fjords and outer skerries, and the cod captured here have traditionally been described as Fjord or Bank cod (Nordeide et al., 2011).

Population genetics of coastal cod in Norwegian waters have been investigated, both in the northern (NCC: Skarstein, Westgaard, & Fevolden, 2007) and in the southern regions (NCS: Knutsen, Jorde, André, & Stenseth, 2003; Knutsen et al., 2011), as well as along the entire Norwegian coastline (Dahle et al., 2018). The latter study, including an analysis of >4,000 cod from 55 spawning sites, revealed a pattern of genetic isolation by distance along the Norwegian coastline without any clearly defined population boundaries. Furthermore, introgression with NEAC, which followed a decreasing N-S latitudinal pattern to almost none south of 62°N, contributed to the overall pattern in genetic structure for NCC. In the southern part of the NCS range, that is Norwegian Skagerrak, several studies have shown low but statistically significant differences between fjord populations (Jorde, Knutsen H, Espeland, & Stenseth, 2007; Knutsen et al., 2003; Sodeland et al., 2016). The observed level of genetic divergence in this region is consistent with data from tagging studies showing low levels of dispersal (Knutsen et al., 2011) and geographic fine-scale variation in life-history traits of coastal cod (Olsen et al., 2004).

A failure to correctly define genetic stocks is illustrated by the historical and current annual quota agreement formulated by Norway

and Russia for coastal cod, where NCC are added to the 5–10 times larger quota of NEAC (ICES, 2019; Jakobsen, 1987). From the mid-1970s until 2003, an expected annual catch of 25, 000 tonnes for Norwegian and Russian NCC was set within this combined NEAC/NCC quota. ICES provides management advice for NEAC and NCC in the Barents Sea and Norwegian Russian management areas, and the total quota has been driven primarily by the status of the NEAC stock, thus leading to an inherent risk of overexploiting the smaller NCC stock. Due to the decline of NCC, ICES advised a zero catch of NCC for the years 2004–2011 supplemented by a recovery plan (ICES, 2015, 2019).

The genomic resources available for Atlantic cod have rapidly expanded in recent years, for example, by the development of genome assemblies (Star et al., 2011; Tørresen et al., 2017) and a SNP array (Berg et al., 2016). These tools have permitted, for example, distinguishing between migratory and nonmigratory ecotypes throughout the species range (Barney, Munkholm, Walt, & Palumbi, 2017; Berg et al., 2017; Berg et al., 2016; Bradbury et al., 2013; Kirubakaran et al., 2016). Between these two ecotypes, distinct islands of divergence have been observed in linkage groups (LGs) 1, 2, 7 and 12 (Berg et al., 2016; Sodeland et al., 2016). The island of divergence on LG1 coincides with a double inversion (Berg et al., 2017; Kirubakaran et al., 2016). Inversions on LG12 seem to differentiate between NCS and the North Sea cod in the south of Norway (Barth et al., 2017; Sodeland et al., 2016), whereas in the Gulf of Maine, islands of divergence on LG2, LG7 and LG12 allow to identify the three spawning units in this region potentially linked with local adaptation (Barney et al., 2017).

Despite recent advances in our understanding of population structure in Atlantic cod, many questions still remain and can be addressed with genomic data to help advise management of cod populations. We used the recently developed SNP array to analyse population genetic structure in coastal cod from multiple locations
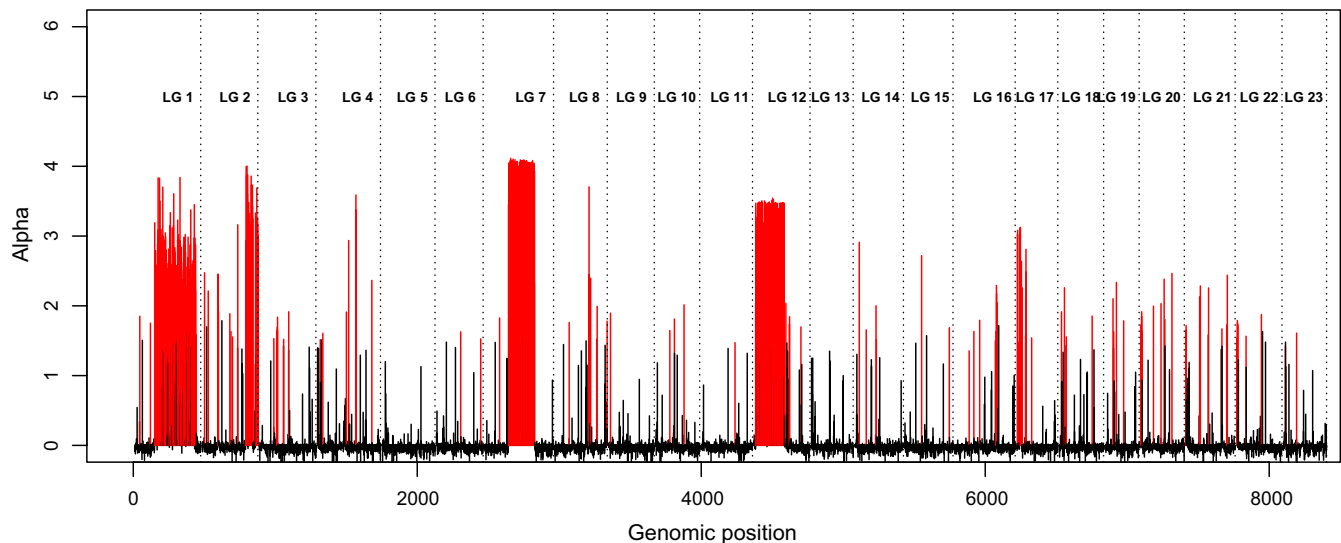
**FIGURE 2** Outlier scan. SNPs are ordered according to their positions on each of the 23 LGs. The estimated alpha coefficient indicates the strength and direction of selection, being positive for diversifying and negative for balancing. The alpha reported on the y-axis: dark lines for $p > .001$ and red lines for $p < .001$

of coastal cod in Norway presently managed as two stocks north and south of 62°N. In addition, samples from White Sea, Ireland and Faroe Islands were included to provide a wider framework for interpreting genomic variation within and among cod stock components.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection, DNA extraction and genotyping

All the Norwegian and White Sea cod included in the present study (Table 1) were sampled as a part of a large sampling programme funded by the Institute of Marine Research in Norway (IMR) by gill-nets or by longline. Cod representing two Irish samples (landings north-east and south of Ireland) and the Faroe Islands (two samples; landing from Faroe Plateau and from offshore Faroe Bank cod) were taken from commercial trawlers. Gill tissue was stored in 96% ethanol prior to DNA extraction. Genomic DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germany). DNA quality and quantity were assessed by agarose gel electrophoresis and Qubit fluorometry (Thermo Fisher Scientific, USA). Taken together, twelve geographic locations were represented by 549 individuals (Figure 1; Table 1).

Otoliths from the Norwegian samples were read to determine the age of the fish and to classify them as NEAC or Norwegian coastal cod (NCC/NCS). Otolith types 1 and 2 depict coastal cod, whereas 4 and 5 are assigned to NEAC, according to (Rollefsen 1933) and Berg and Albert (2003). Both the NCC and NCS show similar otolith patterns. Individual cod with otolith category NEAC were removed from the NCC samples to exclude any migratory NEAC (this meant removing 16 individuals in total from the northern samples).

All samples were genotyped using a custom Illumina SNP array containing assays for 10,913 SNPs (Berg et al., 2016) according to the manufacturer's instructions (Illumina, San Diego, USA). Individuals displaying a call rate below 0.9 were excluded from analyses as were nonpolymorphic SNPs or SNP with a call rate <0.95, leaving a final data set of 486 fish analysed with 8,174 genome-distributed SNPs.

### 2.2 | Outlier detection

In large marine populations, most of the genetic markers might be uninformative about demographic structure (Ward, Woodwark, & Skibinski, 1994), so loci carrying signature of local divergence are useful to outline management units for fisheries management (Russello, Kirk, Frazer, & Askey, 2012). To stratify the SNP data set into loci non-deviating from neutrality and candidates to selection, outlier analyses were conducted using three analytical approaches: BayeScan (Foll & Gaggiotti, 2008), LOSITAN (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008) and PCAdapt (Luu, Bazin, & Blum, 2017). BayeScan v.2.1 (Foll & Gaggiotti, 2008) was used with default parameters, and the $\log^{10}$(BF)>0.5 criterion, that is "substantial" evidence for selection according to Jeffreys (1961), was chosen to define non-neutral markers. In LOSITAN, a neutral distribution of $F_{ST}$ with 1,000,000 iterations was simulated, with forced mean $F_{ST}$ at a significance level of 0.05 under an infinite allele model. It has been suggested that outlier tests may produce high false-positive rates because of population demography and bottlenecks (see, e.g., Lotterhos & Whitlock 2014; Narum & Hess 2011; de Villemereuil & Gaggiotti 2015). A way to circumvent this problem is to conduct the analyses between pairs of populations, since this partly overlooks the methodological weakness of population structure/demographic processes (Vitalis, Dawson, & Boursot, 2001). Hence, a first analytical step consisted of pooling all the samples to detect SNPs under selection across all geographic regions. Secondly,

outlier scans were individually performed for each pair of samples to determine whether the selection pressure changed across locations or whether it reflected any geographic pattern. Markers were regarded as neutral if they were categorized as such in all pairs of samples, and they were considered as outliers if they were found to be under positive selection in, at least, one pair of samples. SNPs under balancing selection, identified by negative alpha value both from BayeScan and from LOSITAN graphic output, were found in extremely low numbers and discarded from this study. The computationally demanding pairwise tests were performed only with BayeScan.

Consensus genome scan of outlier detection revealed large portions of LG1, LG2, LG7 and LG12 to be under likely positive selection (see Figure 2 and Figure S1a,b). Therefore, the data set was subsequently stratified into six sub data sets for population structure analyses (two neutral and four non-neutral ones). Two sets of putative neutral markers were defined: one containing loci outside LG1, LG2, LG7 and LG12, which was named "Neutrals-A" ($n = 5,854$ SNPs), and a second one containing SNPs on LG1, LG2, LG7 and LG12, but outside the regions of these chromosomes that were deemed to be under positive selection, which was called "Neutrals-B" ($n = 1,344$). A further four sets of candidate markers to positive selection were defined within the selected areas on LG1 (LG1S, $n = 281$), LG2 (LG2S; $n = 75$), LG7 (LG7S; $n = 185$) and LG12 (LG12S; $n = 200$), respectively.

## 2.3 | Genetic clustering

STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) was run separately on the six subsets of data, as defined above, under a model assuming admixture and correlated allele frequencies but without using population information. For each data set, ten runs, with a burn-in period consisting of 100,000 replications and a run length of 1,000,000 Markov chain Monte Carlo (MCMC) iterations, were performed for $K = 1$ to $K = 8$ clusters. STRUCTURE runs were automatized with the program ParallelStructure (Besnier & Glover 2013) to reduce computational time. Output was analysed using two approaches: the ad hoc summary statistic $\Delta K$ from Evanno, Regnaut, and Goudet (2005), which detects the uppermost hierarchical level of structure in the data, and using the four statistics (MedMed, MedMean, MaxMed and MaxMean) implemented in StructureSelector (Li & Liu, 2018). The latter has been described as more accurate than the previously used methods to determine the best-fit number of clusters, for both even and uneven sampling data. Finally, runs for the selected Ks were averaged with CLUMPP v.1.1.1 (Jakobsson & Rosenberg 2007) using the FullSearch algorithm and the G' pairwise matrix similarity statistic and were graphically displayed using bar plots.

Genetic clustering was investigated with all six data sets by discriminant analysis of principal components (DAPC) (Jombart, Devillard, & Balloux, 2010), in which the coordinates of all the individuals were calculated on a set of discriminant axes aiming to maximize the variance between region and to minimize the variance within regions. Plotting all the individuals on a two-dimensional space consisting of the two first discriminant axes provided a graphical representation of the genetic distance and relative positions of the samples. The DAPC was performed in R (R Core Team, 2018), using the adegenet package (Jombart, 2008; Jombart & Ahmed, 2011).

## 2.4 | Haplotype reconstruction

The large regions under positive selection in LG1, LG2, LG7 and LG12 were characterized by high levels of linkage disequilibrium. Using a 15-SNP sliding window, extended haplotypes for each LG were reconstructed separately for the twelve sampling locations using the software PHASE v2.1 (Stephens, Smith, & Donnelly, 2004). Some of those haplotypes were represented in high frequency in all, or most of, the sampling sites. The frequency of the haplotypes was reported in all sampling sites for any sequence of 15-SNP haplotype that reached the minimum frequency of 10% in at least one sampling site. The frequency of each haplotype per site was plotted on a map obtained from the R packages "maps" (Becker & Wilks, 2018) and "mapplots" (Gerritsen, 2019).

## 2.5 | Genetic distances and Mantel test

Pairwise genetic distances ($F_{ST}$) between sampling sites were computed separately for each of the six subsets of SNPs defined above using Weir and Cockerham (1984) unbiased estimator implemented in the software Arlequin v.3.5.1.2 (Excoffier, Laval, and Schneider, 2005). Statistical significance for the null hypothesis of no genetic divergence was assessed by 10,000 permutations.

The relationship between genetic and geographic distances was explored by testing whether the genetic data fitted a (linear) pattern of isolation by distance (IBD). A two-tailed Mantel (1967) test was conducted between the matrices of genetic distance (estimated as pairwise $F_{ST}$) and geographic distance (defined as the shortest marine path between sites and measured using Google Earth). The analyses were conducted using PaSSaGE v.2 (Rosenberg & Anderson, 2011), and 10,000 permutations were used to calculate the significance of the correlations.

The slopes of IBD tests allow to obtain qualitative estimates of mean dispersal distances using the theoretical model elaborated by Kinlan and Gaines (2003) and based on Palumbi (2003); i.e. dispersal distance $= 0.0016(\text{IBD slope})^{-1.0001}$. The average dispersal distance obtained for markers under positive selection was compared with the distance for neutral loci to assess whether gene flow might be influenced by any form of environmental pressure.

## 2.6 | Detection of loci associated with environmental variables

The identification of candidate SNPs putatively involved in local adaptation was addressed with two complementary approaches.
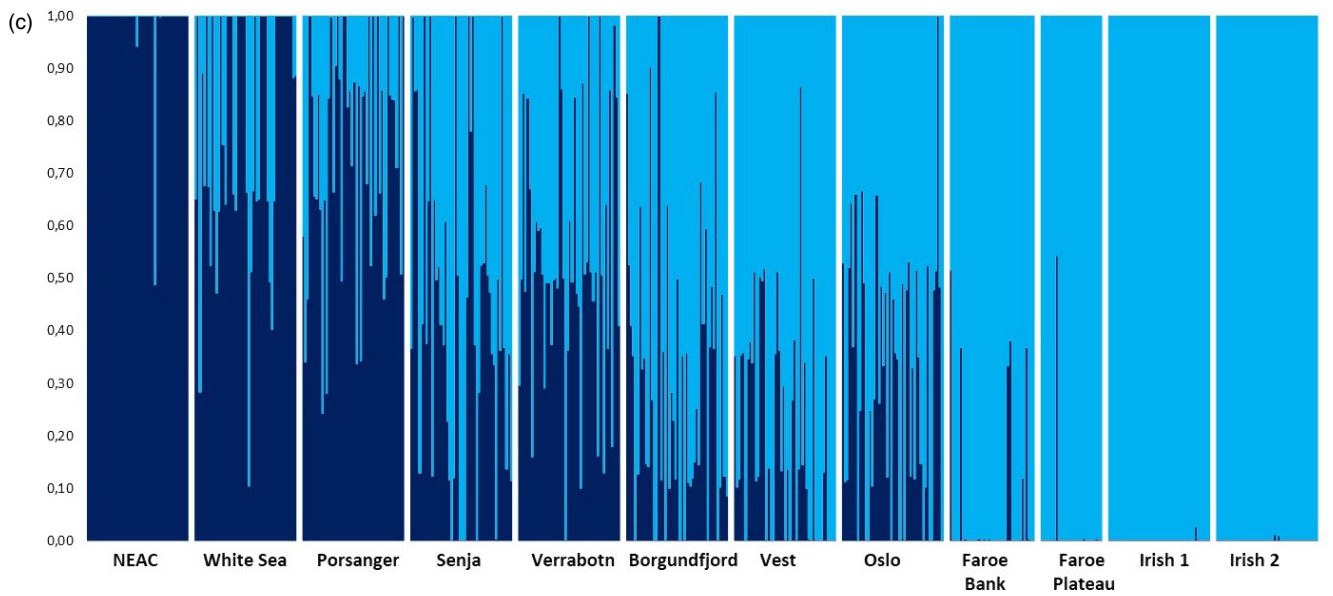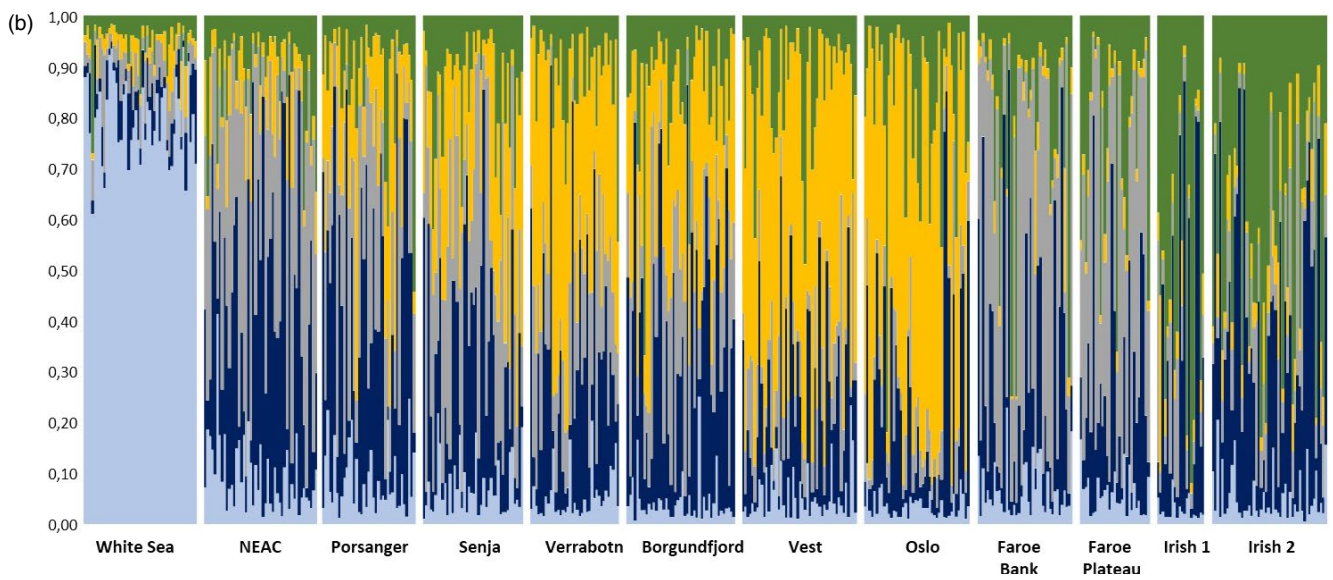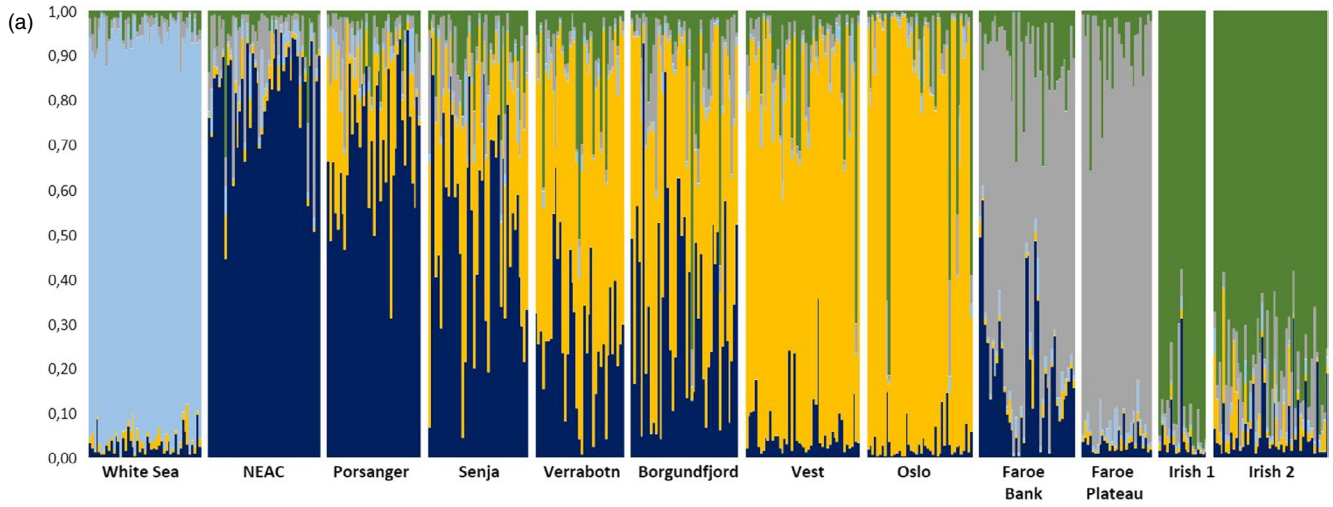
**FIGURE 3** Bayesian clustering of the twelve samples with: (a) 5,854 neutral SNPs (Neutrals-A), (b) 1,433 neutral SNPs within the four LG groups (Neutrals-B), and (c) the pools of SNPs identified as under positive selection on LG1, LG2, LG7 and LG12 when including NEAC. Inferred ancestry of individuals was calculated after averaging ten STRUCTURE runs with CLUMPP; see Table S1 for Evanno test and StructureSelector results

First, outlier analysis was used to flag loci most strongly associated with the observed population structure, which was located within LG1, LG2, LG7 and LG12 (see above). Then, LFMM, "latent factor mixed model" (Frichot, Schoville, Bouchard, & François, 2013), was used to identify loci showing unusual associations with environmental variables compared to the genetic background. LFMM accounts for the underlying population structure by introducing latent factors while simultaneously estimating random effects driven by isolation by distance and population history. The environmental factor tested was temperature, measured at 50 m depth, in the months of March and July as they are assumed to be important for spawning and juvenile growth, respectively. Hence, associations between genetic variation of loci belonging to LG1, LG2, LG7 and LG12 and temperature were assessed while controlling for neutral genetic structure with (random) latent factors. Ten runs of LFMM were conducted using 1,000 sweeps for burn-in and 10,000 additional sweeps. The number of latent factors was set at $K = 4$ according to STRUCTURE as suggested by Frichot et al. (2013). As the variation among runs was extremely low, only the first one was kept. Significance was chosen after Bonferroni correction for multiple tests.

## 3 | RESULTS

### 3.1 | Assessing non-neutral genomic regions of divergence

The initial scan for outlier detection performed across all of the 23 linkage groups (LGs) in all of the samples (8,174 SNPs) revealed that LG1, LG2, LG7 and LG12 contained relatively large genomic regions deviating from neutrality, as indicated by a large number of SNPs scoring a high probability for being under positive selection (Figure 2 and Figures S1a,b and S2). Conversely, other genomic regions showed less evidence of selection.

### 3.2 | Genetic clustering

Evanno test revealed a major division at $K = 2$ (Table S1) for both sets of neutral markers. The highest differentiation was observed between the White Sea and the remaining samples at $K = 2$ (Figure S3a). Conversely, five genetic groups were identified by StructureSelector (Table S1) for all neutral markers. In the Neutrals-A data set, the
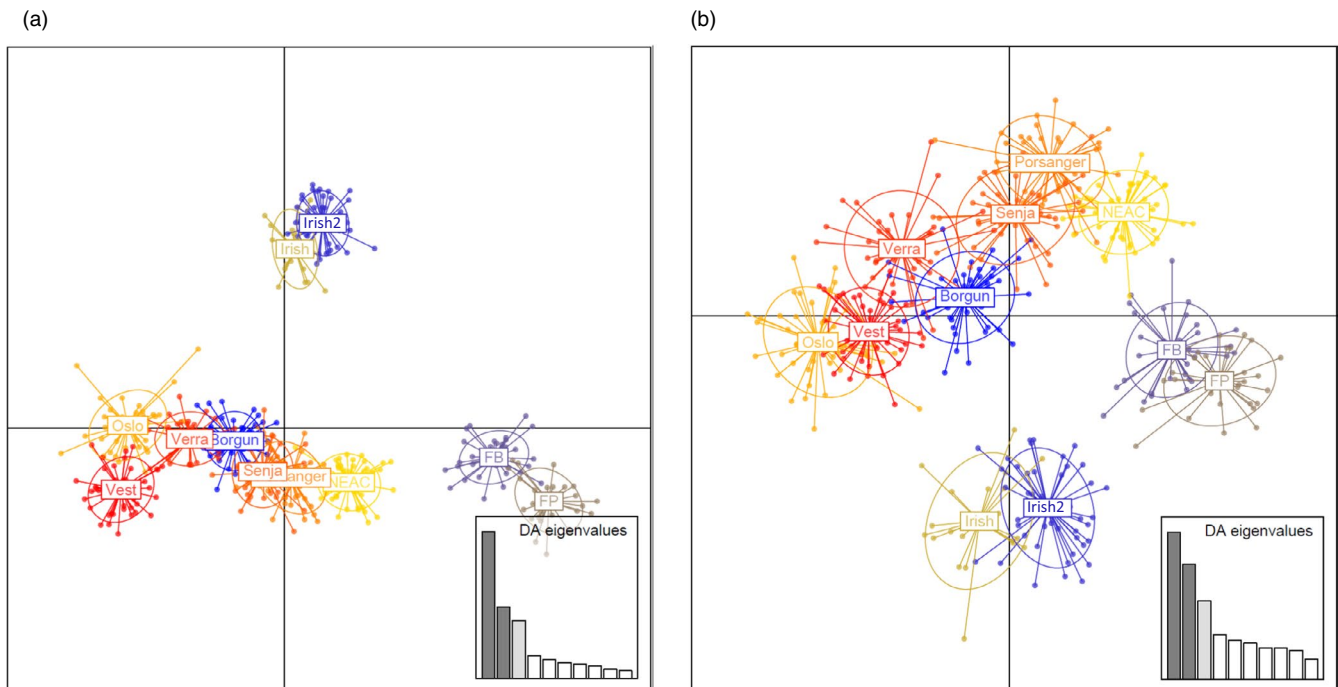


**FIGURE 4** Discriminant analyses of principal components (DAPC) of the samples from eleven sites (i.e. excluding the White Sea) based on neutral markers: (a) 5,854 SNPs (Neutrals-A) and (b) 1,344 SNPs (Neutrals-B). Plots containing the White Sea can be found in Figure S5a-b in Supplementary Information

genetic clustering pattern somehow revealed a geographic underlying pattern (Figure 3a); that is White Sea, Vest-Oslo, Faroe samples, Irish samples and a gradient from NEAC to Borgundfjord in the remaining NCC samples. The clustering pattern was more diffuse in the data set using the neutral SNPs located within the LG, the White Sea being the only distinct group (Figure 3b). The candidate SNPs to positive selection located within LG1, LG2, LG7 and LG12 combined revealed a gradient in which the NCC samples in the northernmost region of Norway were genetically closer to the NEAC (Figure 3c, more details in Figure S4a–k). By removing the NEAC from the plot, the White Sea sample stands out as the most deviating sample (plot not shown).

The DAPC outcome for the sets A and B of neutral markers (Figure 4a-b: including the White Sea; Figure S5a,b) agreed with STRUCTURE, with samples from different countries clustering separately while a gradient of genetic similarity was revealed from north to south in Norwegian coastal cod. However, the patterns revealed by the SNPs under selection varied among the four LGs (Figure 5a-d). For instance, within LG1, the axis corresponding to the first component (PC1) defined three pools: NEAC, White Sea and the remaining samples (Figure 5a), whereas for LG12 (Figure 5d), the three pools distinctly separated by PC1 were as follows: White Sea, the Irish samples and the rest. While the LG2 (Figure 5b) and LG7
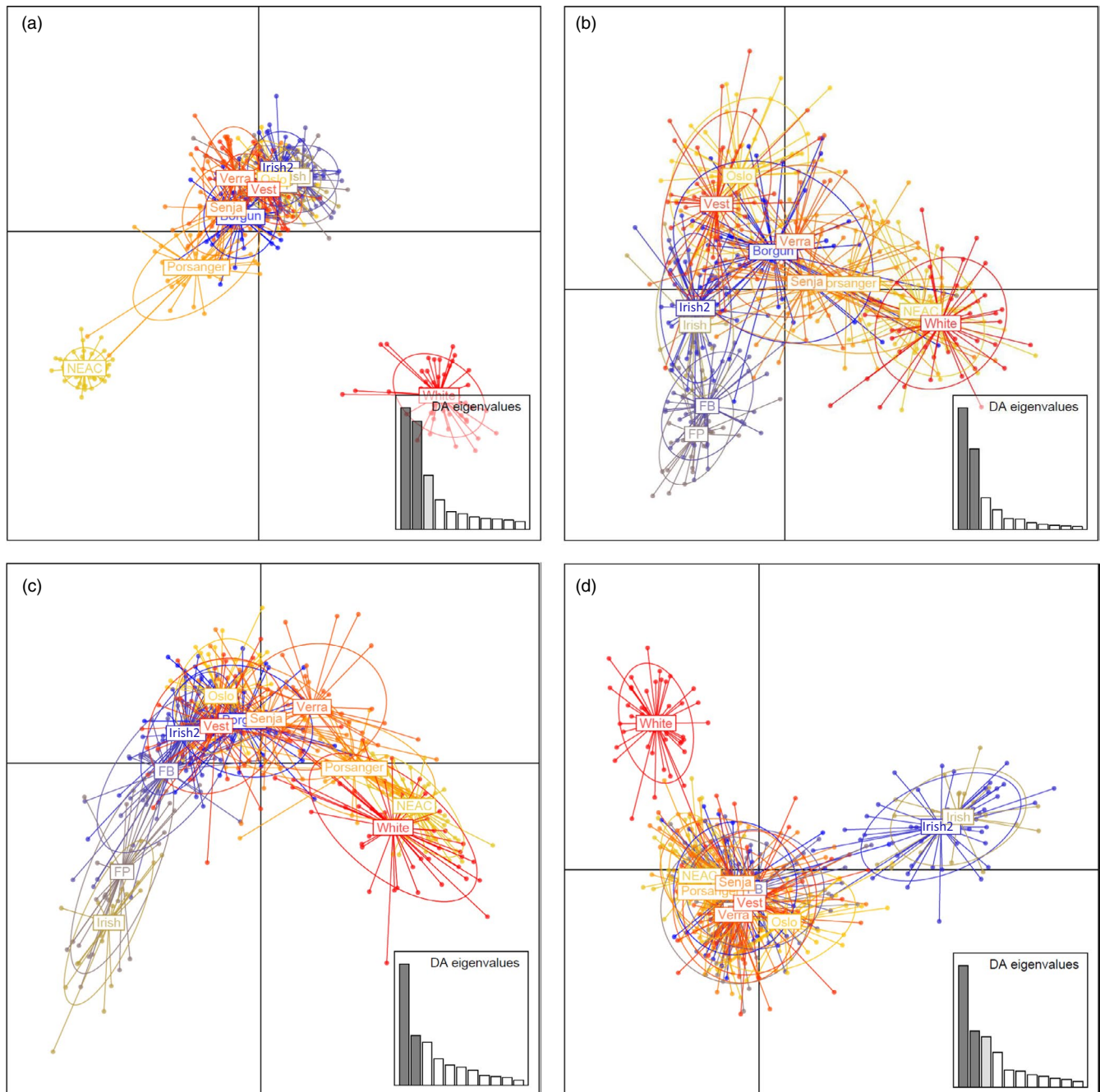


**FIGURE 5** Discriminant analyses of principal components (DAPC) of the samples from the twelve sites based on SNPs under positive selection on: (a) LG1, (b) LG2, (c) LG7 and (d) LG12

(Figure 5c) show latitudinal gradients in the Norwegian samples, the NEAC and White Sea samples are overlapping.

## 3.3 | Geographic patterns in reconstructed haplotypes

Haplotypes were constructed for each sample along 15-SNP sliding windows for each of the four LGs displaying markers under selection (Table 2, Figure S6). For each LG, one or several haplotypes were observed in moderate to high frequencies in all the samples. For example, in LG1 (Figure S6), the 15-SNP haplotype starting at SNP position 223 was found in moderate to high frequencies ranging from 24% in Oslo up to 96% in NEAC (Table 2). Interestingly, this region of highly shared haplotypes was very narrow around SNP no. 223 in all samples, apart from NEAC, where the highly shared haplotypes seemed to cover a large portion of the chromosome (between SNP nos. 120 and 400, Figure S6). Similarly, in LG2, LG7 and LG12, at least one locus presented a haplotype with a high frequency in all or most of the regions (Table 2).

For each genomic position where a haplotype was reported (see Table 2), haplotypes present in ≥ 10% of the individuals in at least one sampling site were visualized in Figures 6-7. In LG1, the frequency of the one haplotype was almost fixed in the NEAC sample (Pink: Figure 6a), but gradually decreased in frequency southwards in the NCC to finally disappear in the two NCS sites. A second haplotype (white) was present in all samples but NEAC, whereas the blue haplotype was shared by the coastal cod from Verra (NCC) to Oslofjord (NCS). The White Sea sample shared haplotypes (black/white) with the NCC and NCS except for the yellow haplotype also found in the one NCC from Senja.

In LG2, haplotypes were less conserved than in LG1 (Figure 6b). In some regions (Faroe Bank and Irish samples), all haplotypes were present at frequencies less than 10% (shown in black). One haplotype (in red) was only present in the White Sea and NEAC samples. Haplotype in green was mostly represented in the NCC, NEAC and White Sea samples, and the frequency gradually decreased southwards, whereas the white haplotype was present in >28% of the Oslo (NCS) sample and gradually decreased in frequency towards the north.

In LG12, haplotype 1 (white) was the most common in all northern samples between White Sea and Vest, but completely absent from the Irish samples. In the southern samples, haplotypes in green and red were present in high frequencies in the NCS sample from Oslofjord, but completely absent in NCC, NEAC and White Sea samples (Figure 6c).

LG7 displayed several genomic regions where haplotypes were found in high frequencies (Table 2). The geographic distribution of haplotypes varied greatly from one locus to another on LG7, which is illustrated in four different haplotype maps (Figure 7a-d). The first 15-SNP haplotype that spanned from SNP 205 to SNP 219 on LG7 was almost fixed in NEAC, that is the same sequence of 15 SNPs was found in all individuals from the NEAC sample (Figure 7a), whereas the NCC samples displayed a higher haplotype diversity. The same pattern was found in the haplotype that spanned from SNP 287 to 301 (Figure 7c). The two other loci (SNP252-266 and 330–344) displayed a higher haplotype diversity with 3–6 distinct SNP sequences that were found both in the NEAC and NCC samples (Figure 7b, d). NCC and NCS are more variable in Figure 7a, c, while unique regional haplotypes were found in NCC (Figure 7d).

## 3.4 | Genetic differentiation and Isolation by distance

Global $F_{ST}$ for neutral loci was low, albeit statistically significant: 0.007 ($p < .0001$) for Neutrals-A and 0.010 ($p < .0001$) for Neutrals-B. Following expectations, global $F_{ST}$ for loci under positive selection within the different LGs ranged from 0.122 to 0.253 ($p < .0001$). Pairwise $F_{ST}$ matrices for both sets of neutral loci can be found in Table 3 and in Table S2a–d for each LG. At the neutral loci,

**TABLE 2** Frequencies of the most common haplotype (%) within each geographic sample

| | | Frequency of the most common haplotype (%) | | | | | | | | | | | |
| | | | | NCC | | | | NCS | | Faroe Islands | | Ireland | |
| LG | SNP frame | White Sea | NEAC | Porsanger | Senja | Verrabotn | Borgundfjord | Vest | Oslo | Bank | Plateau | Irish 1 | Irish 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LG1 | 223–237 | 31 | 96 | 35 | 30 | 35 | 27 | 28 | 24 | 39 | 33 | 30 | 36 |
| LG2 | 337–351 | 40 | 52 | 34 | 33 | 22 | 25 | 20 | 24 | 12 | 11 | 17 | 16 |
| LG7 | 205–219 | 61 | 91 | 78 | 41 | 59 | 25 | 20 | 24 | 18 | 18 | 42 | 22 |
| LG7 | 252–266 | 25 | 20 | 18 | 51 | 42 | 71 | 78 | 65 | 82 | 85 | 83 | 95 |
| LG7 | 287–301 | 70 | 85 | 78 | 54 | 55 | 65 | 77 | 68 | 85 | 96 | 97 | 94 |
| LG7 | 330–344 | 45 | 62 | 59 | 56 | 42 | 72 | 78 | 70 | 87 | 97 | 99 | 96 |
| LG12 | 29–43 | 88 | 83 | 79 | 70 | 66 | 79 | 50 | 32 | 59 | 52 | 45 | 38 |

*Note:* One row represents one genomic region or "locus," defined by linkage groups (LG) and SNP frame (or numbers) within LG, and may contain different haplotypes in different samples. See also Figures 6 and 7 and text for details.
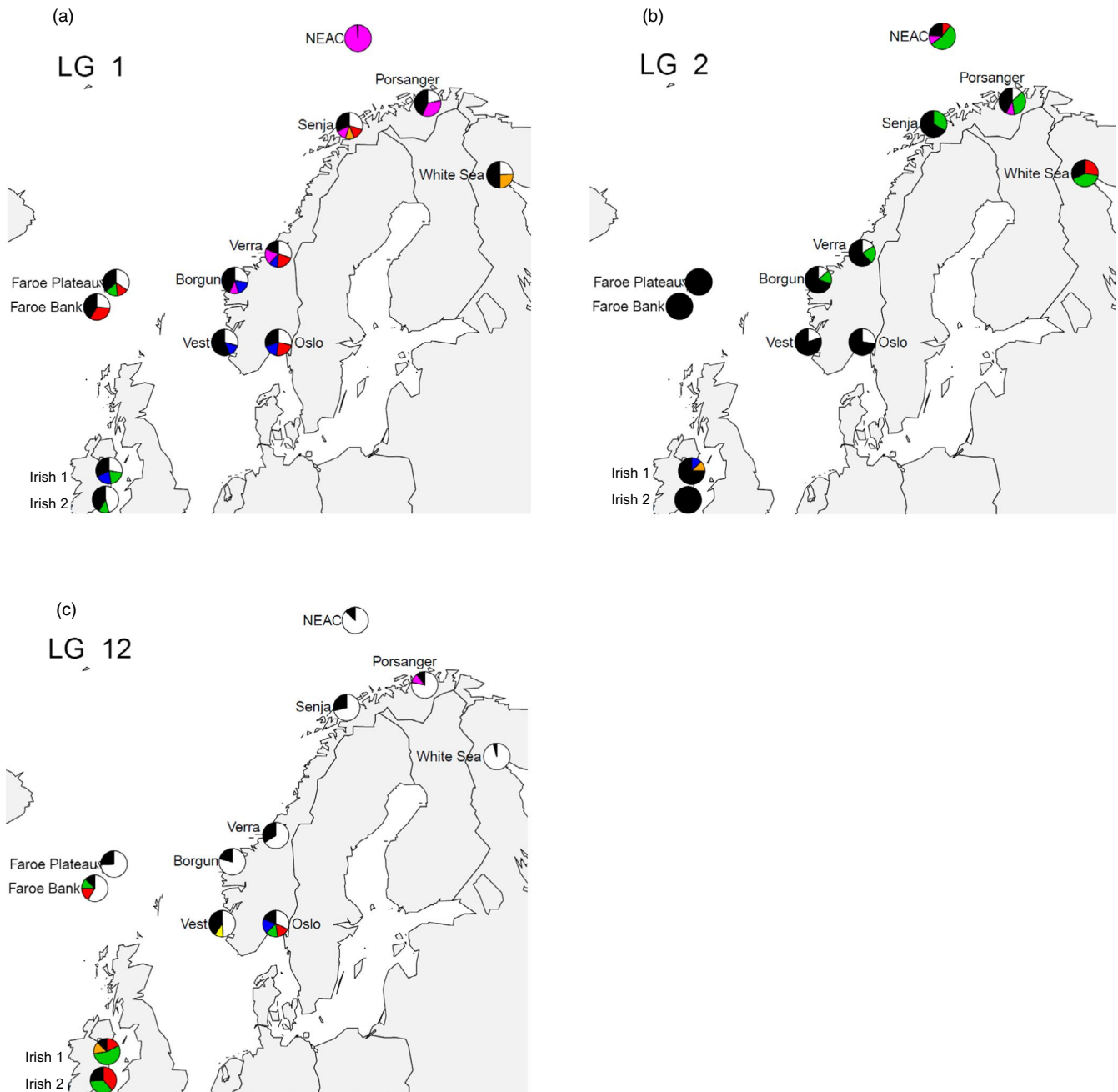
**FIGURE 6** Geographic distribution of haplotypes: (a) LG1: (b) LG2: (c) LG12: geographic repartition of the haplotypes with frequency ≥ 10% in at least one region. Each haplotype is identified by the same given colour across sampling sites. Black represents the sum of all other haplotypes displaying < 10% frequencies. (i.e. all frequencies always sum to 100%)

the White Sea deviated highly from the remaining samples including NCC and NCS. The samples within NCC were significantly different from the NCS for both sets of neutral markers, but no differentiation was found within NCC by neutral SNPs. The average degree of differentiation between NEAC and any of the other samples was 0.62, whereas the average $F_{ST}$ of the remaining comparisons was 0.02 (28-fold difference).

A series of two-tailed Mantel tests were conducted for each set of SNPs in different subsets of sampling sites to assess the correlation between geographic and genetic distance. The total suite of twelve sites significantly followed IBD expectations for all the

set of loci (range from $M = 0.439–0.662$, $p = .033–.000$) except the markers under selection within LG1 ($m = 0.136$, $p = .160$; but see Table S3). When excluding the White Sea from the analyses, the IBD pattern was lost from the neutral markers but present in the four sets of outliers. The third series of tests, conducted only with Norwegian coastal cod from the management units NCC and NCS, showed a relatively strong IBD in all data sets (Neutral-A: $m = 0.657$, $p = .024$; Neutral-B: $m = 0.550$, $p = .039$; LG1: $m = 0.732$, $p = .002$; LG12; $m = 0.572$, $p = .043$) except for LG2 ($0.369$, $p = .120$) and LG7 ($m = 0.502$, $p = .064$). The dispersal distance estimated for neutral markers, when considering the full set of samples was 320 km,
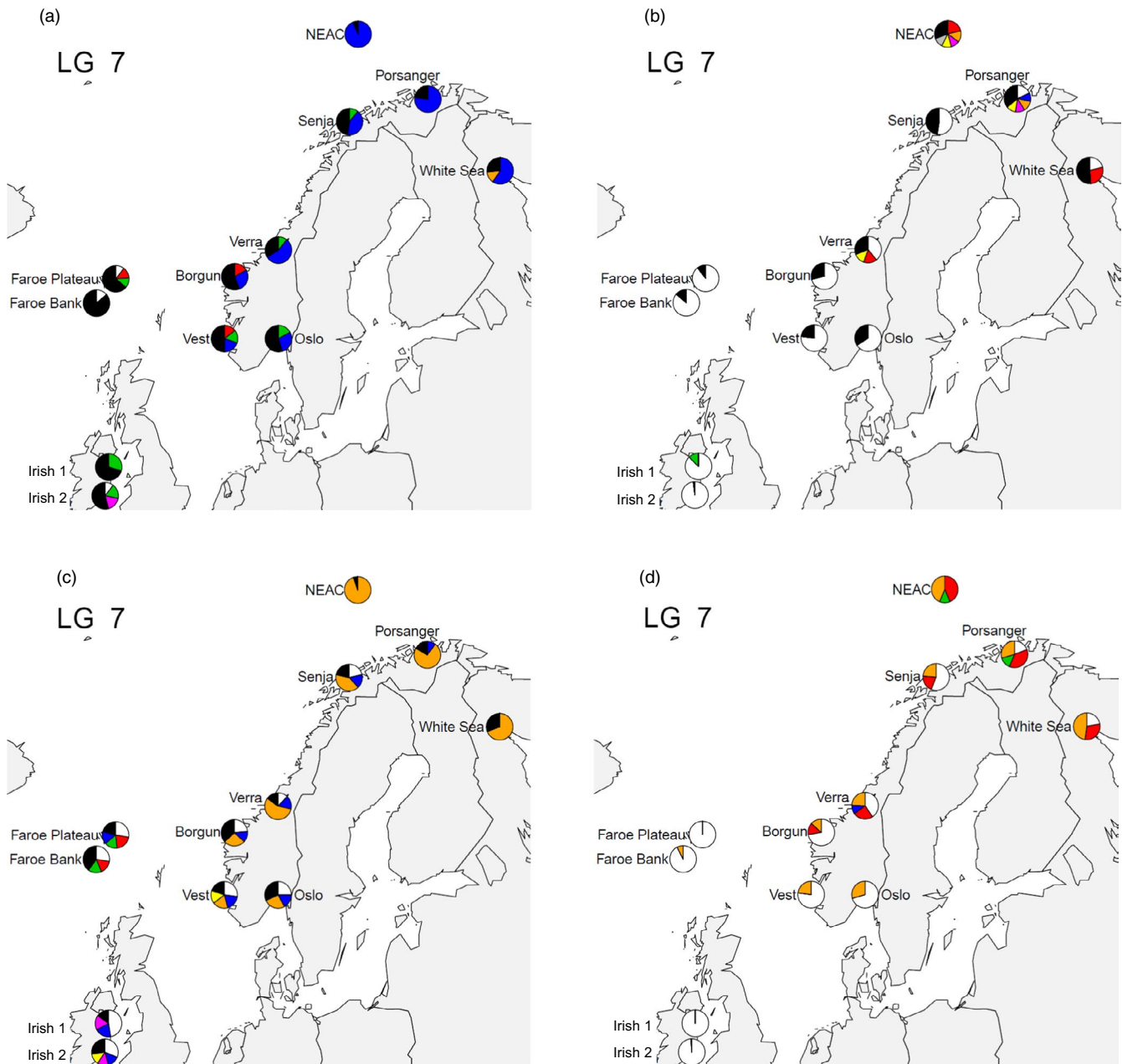
**FIGURE 7** Geographic distribution haplotypes. The four plots represent different subregions of LG7 (cf. Table 2). (a) LG7 SNP frame 205–219, (b) LG7 SNP frame 252–266, (c) LG7 SNP frame 287–301 and (d) LG7 SNP frame 330–344. Each haplotype is identified by the same given colour across sampling sites. Black represents the sum of all other haplotypes displaying < 10% frequencies

whereas the dispersal for loci under selection within LG was significantly smaller (<50 km).

## 3.5 | Detection of loci associated with environmental variables

Loci showing significant association with sea temperature in March were mostly found within LG7, followed by LG12 (91% and 80%, respectively) conversely to the 34% of the loci in LG1 and 55% in LG2. In all cases, percentages slightly increased when testing for sea temperature in July. However, in terms of the strength of the

association, LG7 singled out in both comparisons. By setting the threshold of $-\log_{10}(PO)=15$, 69% of the loci in LG7 showed association with temperature in March and 84% in July. Among the rest of the linkage groups, the highest value was 6% of loci overcoming this threshold in July at LG12 (see Figure S7a,b).

## 4 | DISCUSSION

This is the first study to investigate genetic structure of Norwegian coastal cod across the two management units using a population genomics SNP approach. We observed pairwise genetic differences

**TABLE 3** $F_{ST}$ values between pairs of samples calculated with ARLEQUIN based on Neutrals-A data set (below diagonal) and Neutrals-B (above diagonal)

| | White Sea | NEAC | NCC | | | | NCS | | Faroe Islands | | Ireland | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | Porsanger | Senja | Verrabotn | Borgundfjord | Vest | Oslo | Bank | Plateau | Irish 1 | Irish 2 |
| White Sea | | 0.022 | 0.024 | 0.028 | 0.038 | 0.034 | 0.034 | 0.037 | 0.021 | 0.030 | 0.026 | 0.028 |
| NEAC | 0.019 | | 0.002 | 0.002 | 0.007 | 0.004 | 0.008 | 0.011 | 0.004 | 0.004 | 0.000 | 0.005 |
| Porsanger | 0.021 | 0.001 | | 0.001 | 0.004 | 0.004 | 0.005 | 0.007 | 0.009 | 0.009 | 0.000 | 0.008 |
| Senja | 0.023 | 0.002 | 0.000 | | 0.003 | 0.001 | 0.004 | 0.007 | 0.006 | 0.006 | 0.000 | 0.006 |
| Verrabotn | 0.026 | 0.003 | 0.003 | 0.000 | | 0.001 | 0.006 | 0.007 | 0.014 | 0.015 | 0.001 | 0.008 |
| Borgundfjord | 0.021 | 0.002 | 0.002 | 0.000 | 0.000 | | 0.003 | 0.008 | 0.012 | 0.013 | 0.000 | 0.004 |
| Vest | 0.028 | 0.008 | 0.008 | 0.003 | 0.003 | 0.002 | | 0.003 | 0.016 | 0.016 | 0.000 | 0.008 |
| Oslo | 0.027 | 0.009 | 0.006 | 0.003 | 0.001 | 0.003 | 0.000 | | 0.014 | 0.016 | 0.001 | 0.012 |
| Faroe Bank | 0.022 | 0.003 | 0.003 | 0.006 | 0.006 | 0.003 | 0.011 | 0.010 | | 0.004 | 0.000 | 0.009 |
| Faroe Plateau | 0.026 | 0.003 | 0.005 | 0.005 | 0.006 | 0.008 | 0.010 | 0.010 | 0.000 | | 0.005 | 0.011 |
| Irish 1 | 0.025 | 0.002 | 0.004 | 0.004 | 0.004 | 0.000 | 0.008 | 0.005 | 0.002 | 0.005 | | 0.000 |
| Irish 2 | 0.021 | 0.003 | 0.003 | 0.002 | 0.002 | 0.002 | 0.006 | 0.007 | 0.004 | 0.003 | 0.001 | |

*Note:* Boldface type depicts values significantly different from zero at $\alpha = 0.05$ (after 10,000 permutations).

between all six coastal cod populations, from Porsanger fjord in the far north to Oslo fjord in the south-east. Although the degree of differentiation among samples depended on the set of markers being considered, population genetic differences were observed in all four genomic regions under selection (i.e. LG1, LG2, LG7 and LG12), as well as in both sets of neutral markers (i.e. Neutrals-A and Neutrals-B). The present ICES management regime for coastal cod in Norway is formulated around the 62°N latitude divide, that is Norwegian coastal cod is at present managed as two stocks. Our data demonstrate that while the present management division at 62°N does capture break in gene flow to the south of this border, it does not sufficiently represent genetic structuring in the north and is in need of revision. In order to aid in defining further management boundaries for coastal cod, finer geographic sampling is needed.

Norwegian coastal cod spawn in sheltered fjords and more open coastal areas (Jakobsen, 1987). Although spawning varies greatly in time and space (Johansen et al., 2017; Otterå, Agnalt, & Jørstad, 2006; Otterå et al., 2018), NEAC and NCC overlap on the spawning grounds in some areas. Depending on the markers considered, genetic differentiation observed herein largely followed a north-to-south gradient of similarity with the NEAC sample, whereby the sample from Oslo fjord was the most differentiated to NEAC, and the sample from Porsanger fjord was the least differentiated to NEAC. This gradient is apparent from the population structuring analysis (STRUCTURE, DAPC and IBD: Figures 3-5, Figures S3–S4, respectively). A similar trend of genetic isolation by distance merging towards NEAC in the north has previously been observed (Dahle et al., 2018). Collectively, these studies demonstrate a combination of mixing and gene flow between NEAC and coastal cod that follows a north-to-south gradient, thus driving at least partly the observed population genetic structure. Previous studies have demonstrated

gene flow between NEAC and NCC (Berg et al., 2016; Rodríguez-Ramilo et al., 2019), but not its spatial pattern.

The observed population genetic structure of Norwegian coastal cod could have arisen by different mechanisms but ultimately implies spatially restricted dispersal and gene flow. Taken at face value, the results of the Mantel tests indicate a dispersal distance of Norwegian coastal cod of some 500–800 km (Table S3). Although these estimates should be treated with caution, they align somewhat to the observed migratory distance of NCC that has been estimated to be up to ~300 km on the basis of results from tagging experiments (Michalsen et al., 2014). NCC also show spawning site fidelity (Jakobsen, 1987; Michalsen et al., 2014). The observed gradient of relatedness to NEAC may then have arisen in situ since the postglacial colonization of the coastline by a combination of genetic drift and limited dispersal distance. Another, more likely mechanism is that coastal and NEAC cod would have already diverged prior to colonization and the gradient developed through secondary contact. Such a hypothesis has been suggested when studying population divergence in flatfish from the Baltic and the North Sea as demographic history revealed that the age of the Baltic lineage was actually older than the Baltic Sea itself (Le Moan et al., 2019). Within the latter interpretation, NCC seems more introgressed with NEAC than NCS, a pattern that has some tentative support from the Structure analyses of neutral variation (Figure 3).

## 4.1 | Patterns in non-neutral SNPs and reconstructed haplotypes

Regions with putative selected genes coincided with the four previously known chromosome inversions on LG1, LG2, LG7 and LG12. Spatial patterns of genetic variation at these regions revealed

potentially useful information regarding population structuring in Norwegian coastal cod and adjacent waters. It appears that each genomic region reflects a different pattern. Briefly, LG1 largely distinguishes NEAC and White Sea from the rest (cf. Figure 5a): one single haplotype dominates the whole NEAC sample (96%), while the same haplotype is present in other samples along the Norwegian coast in a decreasing gradient of frequency towards the south (Figure 6a). This pattern appears to reflect a gradient of introgression of NEAC into NCC from north to south. In the southern populations (the Irish Sea and Faroe), other haplotypes are present in high frequency, indicating additional selective forces operating at the same "locus" at LG1 independently from the influence of NEAC.

Selected SNP at LG2 and LG7 regions largely coincided with the neutral ones in describing a pattern of genetic differentiation that followed geographic positions (Figure 5b-c), while also give indication of positive selection between most pairs of samples. The most novel pattern was observed in LG7, where at least four regions or "loci" (starting at SNP positions 205, 252, 287 and 330) appeared to be carrying highly frequent haplotypes across several regions. Representing the frequency of the common haplotypes on each sample (Figure 7a-d) also revealed that each locus reflects a different pattern. Haplotypes starting on SNP position 205 (Figure 7a) and 287 (Figure 7c) are characterized by one unique haplotype present in >80% of the NEAC samples and an increasing gradient of haplotype variety towards the south. On SNP positions 252 (Figure 7b) and 330 (Figure 7d), we observe the opposite phenomenon, with a unique haplotype dominating most of the southern sample with frequency >75% and an increasing gradient of haplotype diversification towards the north. This could reflect different adaptation to temperature in the different areas (Clucas et al., 2019). Of all the LG we tested for local adaptation, only LG7 was linked to temperature in March and July reflecting spawning and larval growth (Figure S7a,b). More detailed investigation of the specific SNP is warranted future research.

Finally, LG12 separated from all others the White Sea and southern samples (cf Figure 5d) while also showed positive selection in the latter. LG12 has previously been found to differentiate between North Sea and coastal cod in the southern Norway (Sodeland et al., 2016).

## 4.2 | White Sea cod

This is the first study to compare the genetics of White Sea cod with neighbouring Norwegian coastal cod. Cod from the White Sea are unique as they are the only coastal cod population inhabiting the Arctic waters of the North Atlantic. Although its life history is poorly understood, recent studies have suggested it to be resident in the White Sea and spawn under the ice at water temperatures as cold as −1.8°C (Makhotin, 2016; Yershov, Marschal, Ereskovsky, & Vishnyakov, 2016). Earlier, White Sea cod has been ascribed to a subspecies due to its divergent biology (*Gadus morhua marisalbi*: Derjugin 1920), and genetic analyses using mtDNA polymorphisms

have confirmed its recent divergence from *Gadus morhua* (Zelenina, Makeenko, Volkov, & Mugue, 2016). Here, we detected highly significant genetic differences between the White Sea cod and Norwegian coastal cod, and between the White Sea cod and all other samples. These differences were highly distinct. For example, pairwise $F_{ST}$ values between the White Sea sample and all others using the Neutral-A and Neutral-B sets of SNPs gave values ranging from 0.019 to 0.028 and 0.021 to 0.038, respectively (Table 3). In contrast, pairwise $F_{ST}$ values between all other population pairs excluding the White Sea sample ranged from 0 to 0.011 and 0 to 0.016, respectively. The White Sea sample also displayed differentiation in the markers under selection on LG1, LG2, LG7 and LG12. Collectively, these analyses demonstrate that White Sea cod are highly divergent to other cod analysed and thus that this population have been isolated longer if not a subspecies.

## 4.3 | The outgroup samples: Ireland and Faroe Islands

Our analysis demonstrated that samples from Ireland and Faroe Islands were distinct from each other and from all the Norwegian samples including NEAC. The Faroe Island cod stocks divided into the Faroe Island Plateau cod and the Faroe Bank cod showed weak divergence between the two, supporting a hypothesis of panmixia in this region. The cod from Faroe bank together with the cod from the Celtic Sea are among the fastest growing cod (Magnussen, 2007). Microsatellites have found low but significant genetic difference between the Bank and the Faroe Plateau cod (Nielsen, Wright, et al., 2009). The Faroe Bank cod have shown resemblance to the NEAC cod in haemoglobin polymorphism (Fyhn, Brix, Nævdal, & Johansen, 1994) but did not show any such association in the present study. The Bank cod experience higher temperature than cod on the Faroe Plateau (Magnússon, Bergstad, Hareide, Magnússon, & Reinert, 1997), which is higher than that experienced by the Norwegian coastal cod and NEAC. The Faroe Bank cod mature later and at larger size than cod on the Faroe Plateau (Brander, 1995). In spite of these differences in life history, only subtle differences were found between them by our SNP panels. The SNP markers included on the cod SNP array (Berg et al., 2016) were identified in samples collected from a broad geographic distribution of Norwegian waters. Nevertheless, the inversions located in LG7 and LG12 were present for these outgroup samples and do indicate differentiation in these stocks that needs to be characterized.

## 5 | CONCLUSIONS

We have revealed a complex genetic variability within coastal cod. This variability is not currently been reflected in the management plan, with its potential consequences for depletion of genetic diversity. The current dividing of Norwegian coastal cod into only two management units, north (NCC) and south (NCS) of 62°N, is not

sufficient to reflect the true biological units and needs to be revised. Although 62°N can be the natural border between north and southern populations, most likely there is an even finer structure both north and south of this border. It will be challenging to devise an optimal management strategy that adequately reflects the patterns of genetic variability within coastal cod. The apparent north–south cline in large parts of the genome does not easily lend itself to interpretation in distinct units, and much more geographic fine-scaled approach seems necessary to resolve genetic units. There are still issues to look further into, as the potential areas under selection, that could maybe help in changes in management plan in the future.

This study has shown that coastal cod in Norway, while highly heterogenous, is also genetically distinct from neighbouring stocks in the north (NEAC, White Sea), west (Faroe Island) and the south (Irish samples). We further found that the White Sea cod are highly divergent from other cod, possibly yielding support the earlier notion of subspecies rank. The other two outlier samples were clearly different from the Norwegian coastal cod and NEAC but showed only subtle differences within, which could be influenced by ascertainment bias.

## DATA AVAILABILITY STATEMENT

Data for this study are available at: available on request from lead author and in supplement.

## ORCID

*Torild Johansen* https://orcid.org/0000-0002-9858-052X
*François Besnier* https://orcid.org/0000-0002-7186-2121
*María Quintela* https://orcid.org/0000-0003-4762-2192
*Per Erik Jorde* https://orcid.org/0000-0001-5515-7257
*Kevin A. Glover* https://orcid.org/0000-0002-7541-1299
*Jon-Ivar Westgaard* https://orcid.org/0000-0002-9627-8124
*Geir Dahle* https://orcid.org/0000-0002-6070-8073
*Sigbjørn Lien* https://orcid.org/0000-0002-5159-1950
*Matthew P. Kent* https://orcid.org/0000-0001-8822-8533

## REFERENCES

Allendorf, F. W., England, P. R., Luikart, G., Ritchie, P. A., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in Ecology and Evolution*, 23, 327–337. https://doi.org/10.1016/j.tree.2008.02.008

Antao, T., Lopes, A., Lopes, R., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a $F_{ST}$-outlier method. *BMC Bioinformatics, 9,* 323.

Barney, B. T., Munkholm, C., Walt, D. R., & Palumbi, S. R. (2017). Highly localized divergence within supergenes in Atlantic cod (*Gadus morhua*) within the Gulf of Maine. *BMC Genomics*, 18, 271. https://doi.org/10.1186/s12864-017-3660-3

Barth, J. M. I., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., ... André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*, 26, 4452–4466. https://doi.org/10.1111/mec.14207

Becker, R. A., & Wilks, R. A. (2018). R-package maps: Draw Geographical Maps. Retrieved from https://CRAN.R-project.org/package=maps

Berg, E., & Albert, O. T. (2003). Cod in fjords and coastal waters of North Norway: Distribution and variation in length and maturity at age. *ICES Journal of Marine Science*, 60, 787–797. https://doi.org/10.1016/S1054-3139(03)00037-7

Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., ... Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity*, 119, (6), 418–428. http://dx.doi.org/10.1038/hdy.2017.54

Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., ... Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, 6, 23246. https://doi.org/10.1038/srep23246

Besnier, F., & Glover, K. A. (2013). ParallelStructure: A R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PLoS One*, 8, e70651.

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., ... Bentzen, P. (2013). Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, 6, 450–461. https://doi.org/10.1111/eva.12026

Brander, K. M. (1995). The effect of temperature on growth of Atlantic cod (*Gadus morhua* L.). *ICES Journal of Marine Science*, 52, 1–10. https://doi.org/10.1016/1054-3139(95)80010-7

Clucas, G. V., Kerr, L. A., Cadrin, S. X., Zemeckis, D. R., Sherwood, G. D., Goethel, D., ... Kovach, A. I. (2019). Adaptive genetic variation underlies biocomplexity of Atlantic Cod in the Gulf of Maine and on Georges Bank. *PLoS One*, 14, e0216992. https://doi.org/10.1371/journal.pone.0216992

Dahle, G., Quintela, M., Johansen, T., Westgaard, J.-I., Besnier, F., Aglen, A., ... Glover, K. A. (2018). Analysis of coastal cod (*Gadus morhua* L.) sampled on spawning sites reveals a genetic gradient throughout Norway's coastline. *BMC Genetics*, 19, 42. https://doi.org/10.1186/s12863-018-0625-8

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online*, 1, 47–50.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993. https://doi.org/10.1534/genetics.108.092221

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30, 1687–1699. https://doi.org/10.1093/molbev/mst063

Fyhn, U. E. H., Brix, O., Nævdal, G., & Johansen, T. (1994). New variants of the haemoglobins of Atlantic cod: A tool for discriminating between coastal and Arctic cod populations. *ICES Marine Science Symposia*, 198, 666–670.

Garrod, D. J., & Schumacher, A. (1994). North Atlantic cod: The broad canvas. *ICES Marine Science Symposia*, 198, 59–76.

Gerritsen, H. (2018). R-package map plots: Data Visualization on Maps. Retrieved from https://CRAN.R-project.org/package=mapplots

Hauser, L., Adcock, G. J., Smith, P. J., Ramírez, J. H. B., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences*, 99(18), 11742–11747. https://doi.org/10.1073/pnas.172242899

Hauser, L., Waples, R. S., & Carvalho, G. R. (2008). Editorial. *Fish and Fisheries*, 9, 331–332. https://doi.org/10.1111/j.1467-2979.2008.00307.x

ICES (2012). *Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK), 4–10 May 2011* (p. 1197). ICES: Advisory comitty. http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2012/WGNSSK

ICES (2015). *Cod (Gadus morhua) in Subareas I and II (Northeast Arctic)*, ICES: Advice. http://www.ices.dk/sites/pub/Publication%20Reports/Advice/2015/2015/cod-arct.pdf

ICES (2018). *Blue Ling (Molva dyptergia) in the Northeast Atlantic. Report on the Working Group on the Biology and Assessment of Deep-sea Fisheries Resources.* ICES WGDEEP Report 2018. ICES, Denmark, p. 773). WGDEEP advice: ICES. http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2018/WGDEEP/01%20WGDEEP%20Report.pdf

ICES (2019). *Arctic Fisheries Working Group (AFWG)*, ICES: Scientific Reports, 930 pp. https://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/Fisheries%20Resources%20Steering%20Group/2019/AFWG/AFWG2019.pdf

Jakobsen, T. (1987). Coastal cod in Northern Norway. *Fisheries Research*, 5, 223–234. https://doi.org/10.1016/0165-7836(87)90042-7

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.

Johansen, T., Westgaard, J.-I., Seliussen, B. B., Nedreaas, K., Dahle, G., Glover, K. A., ... Aglen, A. (2018). "Real-time" genetic monitoring of a commercial fishery on the doorstep of an MPA reveals unique insights into the interaction between coastal and migratory forms of the Atlantic cod. *ICES Journal of Marine Science*, 75(3), 1093–1104.

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94. https://doi.org/10.1186/1471-2156-11-94

Jorde, P. E., Knutsen, H., Espeland, S. H., & Stenseth, N. C. (2007). Spatial scale of genetic structuring in coastal cod Gadus morhua and geographic extent of local populations. *Marine Ecology Progress Series*, 343, 229–237.

Kerr, L. A., Cadrin, S. X., & Kovach, A. I. (2014). Consequences of a mismatch between biological and management units on our perception of Atlantic cod off New England. *ICES Journal of Marine Science*, 71, 1366–1381. https://doi.org/10.1093/icesjms/fsu113

Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L. W., Dickey-Collas, M., Goethel, D. R., ... Nash, R. D. M. (2016). Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. *ICES Journal of Marine Science*, 74, 1708–1722. https://doi.org/10.1093/icesjms/fsw188

Kinlan, B. P., & Gaines, S. D. (2003). Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology*, 84, 2007–2020.

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., ... Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, 25, 2130–2143. https://doi.org/10.1111/mec.13592

Knutsen, H., Jorde, P. E., André, C., & Stenseth, N. C. (2003). Fine-scaled geographic population structuring in a highly mobile marine species: The Atlantic cod. *Molecular Ecology*, 12, 385–394.

Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., AndrÉ, C., & Stenseth, N. C. (2011). Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular Ecology*, 20, 768–783. https://doi.org/10.1111/j.1365-294X.2010.04979.x

Kovach, A. I., Breton, T. S., Berlinsky, D. L., Maceda, L., & Wirgin, I. (2010). Fine-scale spatial and temporal genetic structure of Atlantic cod off the Atlantic coast of the USA. *Marine Ecology Progress Series*, 410, 177–195. https://doi.org/10.3354/meps08612

Le Moan, A., Gaggiotti, O., Henriques, R., Martinez, P., Bekkevold, D., & Hemmer-Hansen, J. (2019). Beyond parallel evolution: when several species colonize the same environmental gradient. *bioRxiv*, 662569. https://doi.org/10.1101/662569

Li, Y.-L., & Liu, J.-X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18, 176–177.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, 23, 2178–2192.

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17, 67–77.

Magnussen, E. (2007). Interpopulation comparison of growth patterns of 14 fish species on Faroe Bank: Are all fishes on the bank fast-growing? *Journal of Fish Biology*, 71, 453–475. https://doi.org/10.1111/j.1095-8649.2007.01502.x

Magnússon, J. V., Bergstad, O. A., Hareide, N. R., Magnússon, J., & Reinert, J. (1997). *Ling, Blue Ling and Tusk of the Northeast Atlantic.* Nordic Project Report, 535, p. 64. Nordic Council of Ministers, TemaNord.

Makhotin, V. V. (2016). Embryonic and early larval development of White Sea cod *Gadus morhua marisalbi* (Gadidae). *Journal of Ichthyology*, 56, 259–280. https://doi.org/10.1134/S0032945216020119

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.

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. https://doi.org/10.1093/icesjms/fsu083

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11(Suppl. 1), 184–194.

Nielsen, E. E., Hemmer-Hansen, J., Poulsen, N. A., Loeschcke, V., Moen, T., Johansen, T., ... Carvalho, G. R. (2009). Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, 9, 276. https://doi.org/10.1186/1471-2148-9-276

Nielsen, E., Wright, P., Hemmer-Hansen, J., Poulsen, N., Gibb, I., & Meldrup, D. (2009). Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: The role of sampling loci and individuals. *Marine Ecology Progress Series*, 376, 213–225. https://doi.org/10.3354/meps07798

Nordeide, J. T., Johansen, S. D., Jørgensen, T. E., Karlsen, B. O., & Moum, 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. https://doi.org/10.3354/meps09232

Olsen, E., Aanes, S., Mehl, S., Holst, J. C., Aglen, A., & Gjøsæter, H. (2010). Cod, haddock, saithe, herring, and capelin in the Barents Sea and adjacent waters: A review of the biological value of the area. *ICES Journal of Marine Science*, 67, 87–101. https://doi.org/10.1093/icesjms/fsp229

Olsen, E. M., Knutsen, H., Gjøsæter, J., Jorde, P. E., Knutsen, J. A., & Stenseth, N. C. (2004). Life-history variation among local

populations of Atlantic cod from the Norwegian Skagerrak coast. *Journal of Fish Biology*, *64*, 1725–1730. https://doi.org/10.1111/j.0022-1112.2004.00402.x

Otterå, H., Agnalt, A.-L., & Jørstad, K. E. (2006). Differences in spawning time of captive Atlantic cod from four regions of Norway, kept under identical conditions. *ICES Journal of Marine Science*, *63*, 216–223. https://doi.org/10.1016/j.icesjms.2005.11.004

Otterå, H., Heino, M., Sørvik, A. G. E., Svåsand, T., Karlsen, Ø., Thorsen, A., & Glover, K. A. (2018). Growth of wild and domesticated Atlantic cod *Gadus morhua* reared under semi-commercial conditions. *Aquaculture Environment Interactions*, *10*, 187–200. https://doi.org/10.3354/aei00262

Palumbi, S. R. (2003). Population genetics, demographic connectivity and the design of marine reserves. *Ecological Applications*, *13*, S146–S158.

Pampoulie, C., Daníelsdóttir, A. K., Storr-Paulsen, M., Hovgård, H., Hjörleifsson, E., & Steinarsson, B. Æ. (2011). Neutral and nonneutral genetic markers revealed the presence of inshore and offshore stock components of Atlantic cod in Greenland waters. *Transactions of the American Fisheries Society*, *140*, 307–319. https://doi.org/10.1080/00028487.2011.567850

Pampoulie, C., Daníelsdóttir, A. K., Thorsteinsson, V., Hjörleifsson, E., Marteinsdóttir, G., & Ruzzante, D. E. (2012). The composition of adult overwintering and juvenile aggregations of Atlantic cod (*Gadus morhua*) around Iceland using neutral and functional markers: A statistical challenge. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*, 307–320.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–995.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: Mismatch between biological and fisheries management units. *Fish and Fisheries*, *10*, 361–395. https://doi.org/10.1111/j.1467-2979.2008.00324.x

Rodríguez-Ramilo, S. T., Baranski, M., Moghadam, H., Grove, H., Lien, S., Goddard, M. E., ... Sonesson, A. K. (2019). Strong selection pressures maintain divergence on genomic islands in Atlantic cod (*Gadus morhua* L.) populations. *Genetics Selection Evolution*, *51*, 61. https://doi.org/10.1186/s12711-019-0503-5

Rollefsen, G. (1933). The otoliths of the cod. *Fiskeridir. Skr. Ser. Havunders.*, *4*(3), 1–18.

Rosenberg, M. S., & Anderson, C. D. (2011). PASSaGE: Pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution*, *2*, 229–232.

Russello, M. A., Kirk, S. L., Frazer, K. K., & Askey, P. J. (2012). Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, *5*, 39–52. https://doi.org/10.1111/j.1752-4571.2011.00206.x

Ruzzante, D. E., Taggart, C. T., & Cook, D. (1999). A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St Lawrence. *Fisheries Research*, *43*, 79–97. https://doi.org/10.1016/S0165-7836(99)00067-3

Saha, A., Hauser, L., Kent, M., Planque, B., Neat, F., Kirubakaran, T. G., ... Johansen, T. (2015). Seascape genetics of saithe (Pollachius virens) across the North Atlantic using single nucleotide polymorphisms. *ICES Journal of Marine Science*, *72*(9), 2732–2741. http://dx.doi.org/10.1093/icesjms/fsv139

Skarstein, T. H., Westgaard, J. I., & Fevolden, S. E. (2007). Comparing microsatellite variation in north-east Atlantic cod (*Gadus morhua* L.) to genetic structuring as revealed by the pantophysin (*Pan* I) locus. *Journal of Fish Biology*, *70*, 271–290. https://doi.org/10.1111/j.1095-8649.2007.01456.x

Sodeland, M., Jorde, P. E., Lien, S., Jentoft, S., Berg, P. R., Grove, H., ... Knutsen, H. (2016). "Islands of Divergence" in the Atlantic cod

genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution*, *8*, 1012–1022.

Spies, I., Hauser, L., Jorde, P. E., Knutsen, H., Punt, A. E., Rogers, L. A., & Stenseth, N. C. (2018). Inferring genetic connectivity in real populations, exemplified by coastal and oceanic Atlantic cod. *Proceedings of the National Academy of Sciences*, *115*, 4945. https://doi.org/10.1073/pnas.1800096115

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., ... Jakobsen, K. S. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, *477*, 207–210. https://doi.org/10.1038/nature10342

Stephens, M., Smith, N. J., & Donnelly, P. (2004). Documentation for PHASE, version 2.1. Retrieved from http://stephenslab.uchicago.edu/phase/download.html

Tørresen, O. K., Star, B., Jentoft, S., Reinar, W. B., Grove, H., Miller, J. R., ... Nederbragt, A. J. (2017). An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics*, *18*, 95. https://doi.org/10.1186/s12864-016-3448-x

de Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*, 1248–1258.

Vitalis, R., Dawson, K., & Boursot, P. (2001). Interpretation of variation across marker loci as evidence of selection. *Genetics*, *158*, 1811–1823.

Ward, R. D., Woodwark, M., & Skibinski, D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology*, *44*, 213–232. https://doi.org/10.1111/j.1095-8649.1994.tb01200.x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Wennevik, V., Jørstad, K. E., Dahle, G., & Fevolden, S.-E. (2008). Mixed stock analysis and the power of different classes of molecular markers in discriminating coastal and oceanic Atlantic cod (*Gadus morhua* L.) on the Lofoten spawning grounds, Northern Norway. In J. Davenport, G. M. Burnell, T. Cross, M. Emmerson, R. McAllen, R. Ramsay, & E. Rogan (Eds.), *Challenges to Marine Ecosystems: Proceedings of the 41st European Marine Biology Symposium* (pp. 7–25). Dordrecht: Springer.

Yaragina, N. A., Aglen, A., & Sokolov, K. M. (2011). Cod. In T. Jakobsen et al. (Eds.), *The Barents Sea; Ecosystem, Resources, Management. Half a century of Russian-Norwegian cooperation* (pp. 223–270). Tapir Academic Press.

Yershov, P. N., Marschal, C., Ereskovsky, A. V., & Vishnyakov, A. E. (2016). New data on the longevity of coastal cod *Gadus morhua* Linnaeus, 1758 in the White Sea. *Journal of Applied Ichthyology*, *32*, 350–352.

Zelenina, D. A., Makeenko, G. A., Volkov, A. A., & Mugue, N. S. (2016). Mitochondrial DNA polymorphism of Atlantic cod of the Barents and White seas. *Biology Bulletin*, *43*, 235–243. https://doi.org/10.1134/S1062359016030134

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.