Genome architecture enables local adaptation of Atlantic cod despite high connectivity

Julia M.I. Barth[1], Paul R. Berg[1,2], Per R. Jonsson[3], Sara Bonanomi[4,5], Hanna Corell[3], Jakob Hemmer-Hansen[4], Kjetill S. Jakobsen[1], Kerstin Johannesson[3], Per Erik Jorde[1], Halvor Knutsen[1,6,7], Per-Olav Moksnes[8], Bastiaan Star[1], Nils Chr. Stenseth[1,7], Henrik Svedäng[9], Sissel Jentoft[1,7] and Carl André[3]

[1] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, 0316 Oslo, Norway

[2] Centre for Molecular Medicine Norway (NCMM), Faculty of Medicine, University of Oslo, 0318 Oslo, Norway

[3] Department of Marine Sciences - Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden

[4] Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, 8600 Silkeborg, Denmark

[5] National Research Council (CNR), Institute of Marine Sciences (ISMAR), Fisheries Section, Largo Fiera della Pesca, 60125 Ancona, Italy

[6] Institute of Marine Research, Flødevigen, 4817 His, Norway

[7] Centre for Coastal Research, Department of Natural Sciences, University of Agder, 4604 Kristiansand, Norway

[8] Department of Marine Sciences, University of Gothenburg, 40530 Gothenburg, Sweden

[9] Swedish Institute for the Marine Environment (SIME), 40530, Gothenburg, Sweden

Keywords: population divergence, gene flow, ecological adaptation, *Gadus morhua*, chromosomal inversion.

Correspondence: Julia M.I. Barth, CEES, Department of Biosciences, University of Oslo, PB 1066 Blindern, 0316 Oslo, Norway, jmi.barth@gmail.com; and Sissel Jentoft, CEES, Department of Biosciences, University of Oslo, PB 1066 Blindern, 0316 Oslo, Norway, sissel.jentoft@ibv.uio.no

Running Title: Adaptation and connectivity of Atlantic cod

Abstract

Adaptation to local conditions is a fundamental process in evolution; however, mechanisms maintaining local adaptation despite high gene flow are still poorly understood. Marine ecosystems provide a wide array of diverse habitats that frequently promote ecological adaptation even in species characterized by strong levels of gene flow. As one example, populations of the marine fish Atlantic cod (*Gadus morhua*) are highly connected due to immense dispersal capabilities but nevertheless show local adaptation in several key traits. By combining population genomic analyses based on 12K single-nucleotide polymorphisms with larval dispersal patterns inferred using a biophysical ocean model, we show that Atlantic cod individuals residing in sheltered estuarine habitats of Scandinavian fjords mainly belong to offshore oceanic populations with considerable connectivity between these diverse ecosystems. Nevertheless, we also find evidence for discrete fjord populations that are genetically differentiated from offshore populations, indicative of local adaptation, the degree of which appears to be influenced by connectivity. Analyses of the genomic architecture reveal a significant overrepresentation of a large ~5 Mb chromosomal rearrangement in fjord cod, previously proposed to comprise genes critical for the survival at low salinities. This suggests that despite considerable connectivity with offshore populations, local adaptation to fjord environments may be enabled by suppression of recombination in the rearranged region. Our study provides new insights into the potential of local adaptation in high gene flow species within fine geographical scales and highlights the importance of genome architecture in analyses of ecological adaptation.

Introduction

Local adaptation characterizes populations that experience higher inherited fitness in their native habitat compared to members of other populations transferred to the same environment (Kawecki & Ebert 2004). The degree of such ecological adaptation depends on the directional selection of advantageous traits and is counteracted by high connectivity and resulting homogenizing gene flow, implicating a limited potential for local adaptation in populations experiencing high gene flow (Wright 1931; Dobzhansky 1937; Mayr 1942). Although environmental adaptation can also involve gene expression-

induced plastic responses such as morphological, physiological or behavioral changes, these occur without genotypic changes (Via *et al*. 1995; Reusch 2014).

Most marine fish populations have traditionally been regarded as large panmictic entities with high connectivity due to the apparent lack of geographical barriers, high dispersal capabilities, and slow genetic drift as a result of large effective population sizes (DeWoody & Avise 2000; Waples & Gaggiotti 2006; Allendorf *et al.* 2010). However, this assumption is challenged by an increasing number of genetic studies reporting high levels of local adaptation in marine fish populations despite substantial gene flow (Nielsen *et al.* 2009; Clarke *et al.* 2010; Limborg *et al.* 2012; Therkildsen *et al*. 2013; Milano *et al.* 2014). Simulation studies have demonstrated that local adaptation can arise in these situations through selection on tightly linked divergent alleles rather than on many single loci (Yeaman & Whitlock 2011). In line with these expectations, the occurrence of linked alleles (e.g., in the form of chromosomal rearrangements) in locally adapted populations has been reported in studies addressing the genome architecture of fish species such as stickleback (Jones *et al.* 2012; Roesti *et al*. 2015), Atlantic herring (Martinez-Barrio *et al.* 2016, Lamichhaney *et al.* 2017), and Atlantic cod (Bradbury *et al.* 2013; Hemmer-Hansen *et al.* 2013; Bradbury *et al.* 2014; Berg *et al.* 2015; 2016; Sodeland *et al.* 2016; Kirubakaran *et al.* 2016; Barney *et al.* 2017). Chromosomal rearrangements that physically combine genes residing within 'supergene clusters' and promote adaptation in connected populations are well known in plants (Lowry and Willis 2010), and insects (Joron *et al.* 2011; Cheng *et al.* 2012) and are widely discussed to play a role in speciation and evolution (Hoffmann and Rieseberg 2008; Schwander et al. 2014). However, the relative importance of this mechanism in highly connected marine populations on small geographical scales remains poorly understood.

Atlantic cod (*Gadus morhua* Linnaeus, 1758) is a benthopelagic, high-fecundity, predatory fish of great commercial and ecological value occurring in a variety of habitats in the North Atlantic and hence constitutes an ideal model for the investigation of local adaptation. Molecular studies inferring the potential for local adaptation in Atlantic cod have a long history, which began with the discovery of adaptive allelic variation in the oxygen-binding protein hemoglobin (Sick 1961) and the observation of a latitudinal gradient in the distribution of its isoforms (Sick 1965; for recent reviews see Andersen (2012) and Ross *et al.* (2013)). Since then, extensive research has contributed to the description of several genetically, phenotypically, and behaviorally distinct populations

occurring in a wide range of different ecosystems (Lilly *et al.* 2008). One of the best-investigated examples for apparent local adaptation despite high connectivity is the co-occurrence of two ecotypes of Atlantic cod, the migratory Northeast Arctic cod (NEAC) and the stationary Norwegian coastal cod (NCC), at the same spawning areas along the northern Norwegian coast (Neuenfeldt *et al.* 2013). While genetic differences between NEAC and NCC were already described in the 1960's (Moller 1966), the mechanism maintaining differentiation despite ongoing gene flow is still a controversial subject (Hemmer-Hansen *et al.* 2013; Karlsen *et al.* 2013). The releases of two successive Atlantic cod genome assemblies (Star *et al.* 2011; Tørresen *et al.* 2017) facilitated the investigation of such mechanisms, revealing the presence of large chromosomal rearrangements likely permitting differentiation of these ecotypes despite ongoing gene flow (Berg *et al.* 2016; Kirubakaran *et al.* 2016).

On a much smaller spatial scale within the Skagerrak and Kattegat, two confined seas connecting the brackish Baltic Sea with the saline North Sea (Fig. 1), evidence has recently accumulated for the presence of yet another pair of coexisting Atlantic cod ecotypes (Rogers *et al.* 2014; Sodeland *et al.* 2016; André *et al.* 2016). These coexisting fish are characterized by distinct lifestyles, with mobile oceanic (offshore) individuals foraging along the coast but possibly returning to North Sea or offshore Skagerrak spawning sites, and sedentary coastal individuals that remain close to the coast and local spawning sites at all times (Knutsen *et al.* 2007; Espeland *et al.* 2008; Neuenfeldt *et al.* 2013; Rogers *et al.* 2014). In line with this observation, differentiated Atlantic cod has been described between estuarine western Skagerrak fjords and offshore areas, as well as between individual fjords (Knutsen *et al.* 2003; Olsen *et al.* 2004; Jorde *et al.* 2007; Knutsen *et al.* 2011). In these cases, the maintenance of differentiation has been associated with seascapes, coastal topography, and hydrographic features such as salinity gradients (Howe *et al.* 2010; Ciannelli *et al.* 2010; Knutsen *et al.* 2011; Rogers *et al.* 2014). Limited migration of coastal cod (Espeland *et al.* 2007; 2008), spawning site fidelity (Espeland *et al.* 2007; Skjæraasen *et al.* 2011), and pronounced natal homing behavior (Svedäng *et al.* 2007; André *et al.* 2016; Bonanomi *et al.* 2016) could further aid differentiation of coastal and oceanic ecotypes by reducing the potential for gene flow. Interestingly, allelic frequency shifts of large chromosomal rearrangements have recently been described between western Skagerrak cod residing in coastal versus oceanic environments (Sodeland *et al.* 2016). In contrast, studies have so far failed to delineate

124 genetic structuring of coastal and locally adapted populations within the fine
125 geographical scale along the eastern Skagerrak-Kattegat coast and fjords (Svedäng *et*
126 *al.* 2010; André *et al.* 2016), although spawning site fidelity was supported by otolith
127 chemistry (Svedäng *et al.* 2010).
128 Whether the hitherto observed sedentary coastal Atlantic cod correspond to locally
129 adapted fjord populations and whether similarly differentiated ecotypes are also present
130 at the eastern Skagerrak coast remain to be investigated. It is also unclear whether the
131 oceanic genotype constitutes of North Sea cod, and whether connectivity and gene flow
132 between these groups exist - and if, whether the exceptional genomic architecture of
133 Atlantic cod contributes to the potential of local adaptation on such fine geographical
134 scales. Answering these questions to improve our knowledge about the mechanism by
135 which local adaptation can be maintained despite high connectivity and gene flow is
136 becoming increasingly relevant in a globally changing world (Pinsky & Palumbi 2014;
137 Savolainen *et al.* 2013; Bernatchez 2016).
138 By using a genome-wide 12K single-nucleotide polymorphism (SNP) array in
139 combination with a comprehensive sampling scheme including several fjords as well as
140 adjacent populations, complemented with biophysical modeling to predict the potential
141 for gene flow among areas, we here address the following research questions: 1.) Can
142 we detect the presence of differentiated cod ecotypes on small spatial scales using
143 genome wide data, and 2.) does the genomic architecture of Atlantic cod contribute to
144 the potential for local adaptation?

145 ## Materials and Methods

146 ### Sample collection

147 Samples of 350 Atlantic cod were obtained from 10 different locations in the Skagerrak-
148 Kattegat area. For comparison, 177 specimens were further sampled from adjacent, but
149 well-differentiated reference locations: English Channel, North Sea, and Danish straits
150 (western Baltic) (Fig. 1, for details see Table S1, Supporting information). Adult fish were
151 all collected during the spawning period from January to April (except ~60% of Grenland
152 fjord individuals collected in November) by trawling or gill net, and care was taken to
153 choose mature fish that were at or close to spawning. Juvenile 0-group cod were
154 collected either in June or September by beach seine. Muscle tissue or fin clips were

stored in ethanol. All cod samples used were collected in compliance with EU Directive 2010/63/EU and the national legislations in Sweden, Denmark, and Norway.

Genotyping and filtering

DNA was extracted from muscle tissue using standard DNA extraction kits and normalized to 100 ng/μl as described elsewhere (Berg *et al.* 2015; 2016). All samples were individually genotyped for 10,913 SNPs using a custom Illumina Infinium II 12K SNP array following the manufacturer's instructions (Illumina, San Diego, USA). The custom chip was designed based on eight individuals representing the global variety of the species, and the Atlantic cod genome (Star *et al.* 2011). Quality control was performed using the genotyping module in GENOMESTUDIO v2011.1 (Illumina Inc.) and the software PLINK v1.07 (Purcell *et al.* 2007) leading to a high-quality SNP set of 7,783 SNPs (for details see Note S1 and Table S2, Supporting information). Variants were further filtered based on linkage to conform with the expectations of models employed in our genetic analyses: the correlation of allele frequencies (r^2) was calculated based on genotypic allele counts and 1,125 SNPs with an $r^2 > 0.1$ were excluded, resulting in a final dataset of 6,658 unlinked SNPs.

A second dataset including SNPs with detected linkage was generated to investigate the importance of large chromosomal rearrangements containing tightly linked SNPs that may play important roles in the divergence and adaptation of Atlantic cod (Bradbury *et al.* 2013; Hemmer-Hansen *et al.* 2013; Bradbury *et al.* 2014; Berg *et al.* 2015; 2016; Sodeland *et al.* 2016; Kirubakaran *et al.* 2016; Barney *et al.* 2017; see section 'Chromosomal rearrangements' below). All format conversions were either accomplished with in-house scripts, or by using the software PGDSPIDER v2.0.8.0 (Lischer & Excoffier 2012).

Genetic differentiation

The population structure was investigated to delineate genetic differentiation and admixture of fjord samples and diverged populations, as well as to test for an isolation by distance (IBD) pattern as described earlier in the western North Atlantic cod (Pogson *et al.* 2001; Beacham *et al.* 2002). Individual ancestry and the number of genetic clusters (*K*) was assessed using a hierarchical framework in STRUCTURE v2.3.2 (Pritchard *et al.* 2000) under the admixture model with correlated allele frequencies for closely related populations or highly migratory species (Falush *et al.* 2003). Five replicates of 100,000

(Monte Carlo Markov chain (MCMC) iterations (discarding the first 10,000 iterations as burn-in) were performed per model, each testing for K=1 to K=5. Convergence was confirmed by consistent results in all five replicates (see Table S3, Supporting information). In addition, principal component analyses were performed to display the largest variances in the genotype data (PCA, Note S2, Table S4, Supporting information).

In an assignment approach to distinguish between mechanical mixture and admixture of individuals (Porras-Hurtado *et al.* 2013), STRUCTURE analyses were conducted with the USEPOPINFO model, using the North Sea and Kattegat samples as representatives of two potential source populations. Enabling of PFROMPOPFLAGONLY ensured that allele frequency estimates depend only on the reference samples, while MIGRPRIOR was set to 0.05 to allow some misclassification of individuals. Per location *q*-values (estimated ancestry) were log normalized (log(data/(1-data)) and analyzed for modality using Hartigans' dip statistic (Hartigan & Hartigan 1985) implemented in the package diptest v0.75-6 (Mächler 2014) for R v3.1.0 (R Core Team, R Foundation for Statistical Computing 2016). Test results were corrected for multiple testing by applying a false discovery rate (FDR) of < 0.05 using the R package qvalue v1.43.0 (Storey, 2004). The ancestry of fjord samples was quantified by their hybrid indices (*H*) employing Bayesian genomic cline analysis as implemented in BGC v1.03 (Gompert & Buerkle 2012). Based on the probability that an individual has inherited a genetic marker from one of the two source populations North Sea and Kattegat, *H* was estimated using two cline parameters that describe the bias (α) and rate (β) of locus-specific introgression into an admixed genomic background (Gompert & Buerkle 2012). As the full set of 6,658 SNPs was too large to allow convergence, the 50 SNPs with the highest fixation indices (F_{ST}) values between the source populations were selected as ancestry informative markers using BAYESCAN v2.1 (Foll & Gaggiotti 2008) (Note S2**,** Table S5, Supporting information). Ten replicates, each using 100,000 MCMC iterations (discarding the first 20,000 iterations as burn-in) were performed. Convergence of the MCMC chain was assessed using TRACER v1.6 (Rambaut *et al.* 2014) and by comparison of the replicates, which produced qualitatively similar results. The replicate with the best fit (highest mean log-likelihood) was selected to present the results.

Pairwise F_{ST} values (Weir & Cockerham 1984) were calculated using ARLEQUIN v3.5 and ARLECOREMAC_64BIT v3.5.2.2 (Excoffier & Lischer 2010), and their significance was

assessed using 10,000 permutation steps. *p*-values were adjusted for multiple testing by applying the FDR approach for non-independent tests by Benjamini and Yekutieli (2001). Pairwise F_{ST} values were plotted by means of classic multidimensional scaling (MDS) using the "cmdscale" method implemented in the R package stats (R Core Team, R Foundation for Statistical Computing 2016) after negative F_{ST} values were set to 0 and a minimal constant (10^{-5}) was added to prevent negative eigenvalues. F_{ST} 95% confidence intervals (200 bootstrap replicates) as well as pairwise genetic and geographic distance matrices for tests of IBD were calculated using the R packages diveRsity v1.9.73 (Keenan *et al.* 2013) and fossil v0.3.7 (Vavrek 2011). Least-cost path distances were obtained using the R package marmap v0.9.2 (Pante & Simon-Bouhet 2013) with bathymetric data from the ETOPO1 1 Arc-Minute Global Relief Model (Amante & Eakins 2009), and Mantel tests of IBD were performed using the R package vegan v2.3.0 (Dixon 2003).

Biophysical connectivity modeling

Physical transport and connectivity of Atlantic cod eggs and larvae was quantified using a biophysical model to explore gene flow potential and connectivity by predicting the most important sources of larvae settling along the Skagerrak coast and the Kattegat. A full description of the biophysical model is given in Jonsson *et al*. (2016). Briefly, the dispersal of eggs and larvae was modeled with a Lagrangian particle-tracking routine in off-line mode driven by flow fields from an ocean circulation model (BaltiX; Hordoir *et al.* 2013). The oceanographic model covers the Baltic Sea, the Kattegat, the Skagerrak, and most of the North Sea with a horizontal resolution of 2 nautical miles (3.7 km) and 84 levels in the vertical, ranging from 3 m at the surface to 23 m in the deepest parts. The model has a free surface, and the atmospheric forcing is a dynamic downscaling of the ERA40 data set (Uppala *et al.* 2006). Freshwater runoff is forced with climatological data from a composite of databases for the Baltic Sea and the North Sea (Meier 2007; O'Dea *et al*. 2014). A previous validation of the BaltiX model showed that it is able to correctly represent the sea-surface height, both tidally induced and wind driven (Hordoir *et al.* 2013). The velocity, temperature, and salinity were updated for all grid boxes in the model domain every three hours, and the trajectory calculations were done with a 15-min time step. To simulate dispersion of cod larvae we used an individual-based drift model with a wide range of combinations of spawning time, egg and larval drift depth, as well

as pelagic larval duration time (for a detailed description see Jonsson *et al.* 2016). Briefly, eggs were simulated to drift at depths between 5 and 15 m and hatched after 20 days. Subsequently, the larvae drifted for another 40 or 70 days at depths between 5 and 30 m. Drifting eggs were started on the 15th of January, February, March, and April in a number of spawning areas in the North Sea, Skagerrak, Kattegat, and the Danish straits (Fig. S1, Supporting information). No mortality was included since little information about temporal and spatial differences in mortality rates is available. Larval drift simulations were repeated for 6 years (1995, 1996, 1998, 2000, 2001, and 2002), which represent negative, neutral, and positive periods of the North Atlantic oscillation winter index (NAO, National Center for Atmospheric Research, 2015), since winter NAO is known to correlate well with variations in the circulation pattern (Marshall *et al.* 2001). To include as much variation as possible, results were based on the average of all spawning times, drift depths, drift durations, and years with a total of ~100M individual drift trajectories. Because of model domain limitations, the North Sea spawning areas did not include the Viking Bank east of Shetland. Connectivity between the spawning areas and the larval settlement areas (western and eastern Skagerrak, and Kattegat) was calculated as the proportion of eggs spawned in area *i* and settling as larvae in area *j*. Furthermore, dispersal patterns from the spawning areas to western Skagerrak fjords were also assessed. As the spatial resolution of the biophysical model is not sufficient to represent the full geomorphology of the inner fjords, only the coastal waters close to the fjord mouths were considered (Soppekilen was not included since the connectivity model cannot resolve this site from the closely situated Hellefjord). The measure of connectivity of the biophysical model only predicts the probability per egg to be transported from *i* to *j*. To obtain a relative estimate of the abundance of eggs reaching a settlement area, we also scaled the inferred connectivity with recent estimates of the spawning stock biomass (SSB, for calculations see Jonsson *et al.* 2016).

Chromosomal rearrangements

The genomic architecture was examined to study the impact of large chromosomal rearrangements on population divergence and adaptation. The physical locations of SNPs within chromosomes (here: linkage groups; LGs) were inferred by mapping the flanking regions of all SNPs to the gadMor2 genome assembly (Tørresen *et al.* 2017) using BLAST v2.2.26+ (Camacho *et al.* 2009). Querying 10,913 flanking region pairs

resulted in 10,804 blast hits, which were subsequently filtered according to the following quality thresholds: identity between query and hit > 90%, E-value < 1.0×10^{-42}, and minimum length > 100 bp. SNPs not meeting these criteria (n=182) and SNPs on unplaced contigs (n=526) were removed. Of the remaining SNPs, the exact positions were retrieved only for high-quality SNPs included in this study (7,783, including linked SNPs, see above). Of these, 506 SNPs could not be mapped, leaving 7,277 SNPs with known position for analysis of the chromosomal rearrangements. The R package inveRsion (Cáceres *et al.* 2012) was used to approximate the start and end points of rearranged regions. A block size of 3 SNPs was used to flank each side of the breakpoint, the minimum minor allele frequency was set to 0.1, and rearrangements were scanned with fixed window sizes from 1 to 13 Mbp. All predictions with Bayesian Information Criterion (BIC) > 0 were scored (Table S6, Supporting information), and breakpoints were defined as the position of the SNP closest to the mean value between breakpoint maxima. The allelic state of each individual (homozygote collinear, heterozygote, or variant rearranged homozygote, as defined by nucleotide diversity in Berg *et al.* (2016)) was inferred using PCA as implemented in the R package adegenet v1.4-1 (Jombart 2008), similar to the approach described by Ma & Amos (2012). Bootstrapping (Efron 1979) (sample size 1,000,000) of individual genotypes was used to calculate the probability of an over- or underrepresentation of the presumably rearranged allele within sampling sites and within western (Tvedestrand, Soppekilen, Hellefjord, Grenland) and eastern (Iddefjord, Gullmarsfjord, Havstensfjord) fjords under the null hypothesis that the frequency of rearranged alleles within a population corresponds to its overall frequency across all populations. Sequential Bonferroni correction was applied to correct for multiple tests (Rice 1988).

Results

Genetic differentiation

The software STRUCTURE was used to investigate population differentiation and the most likely number of clusters (*K*) by applying the admixture model in a hierarchical framework. All samples were tested for their cluster membership in up to five clusters, based on which *K*=2 (Fig. 2a) and *K*=3 (Fig. 2b) were supported as the most likely numbers of populations present (for likelihood values see Table S3, Supporting

315 information). According to Evanno's ΔK statistic, an ad hoc quantity based on the rate of
316 change of the likelihood function (Evanno *et al.* 2005), *K*=2 received most support. In a
317 hierarchical STRUCTURE analysis, the most differentiated clusters are excluded to allow
318 for a more precise analysis of the remaining samples (Vähä *et al.* 2007). Assuming *K*=2,
319 the two most differentiated clusters were composed of the English Channel (ENG), North
320 Sea (NOR), Oslofjord (OSL), and Skagerrak (SKA) (henceforth summarized as North
321 Sea-like group), and the Kattegat (KAT), Öresund (ORE), and Belt Sea (BEL) (from now
322 on summarized as western Baltic-like group) (Fig. 2a). Accordingly, these samples were
323 analyzed in separate runs, but no hidden sub-structure was detected (Fig. S2, for
324 likelihood values see Table S3, Supporting information). Likewise, separate analyses of
325 the remaining fjord sampling sites (Tvedestrand (TVE), Soppekilen (SOP), Hellefjord
326 (HEL), Grenland (GRE), Iddefjord (IDD), Gullmarsfjord (GUL), Havstensfjord (HAV))
327 revealed no further sub-structure and resulted in very similar likelihoods for *K*=2 and *K*=3
328 (Fig. S2 and Table S3, Supporting information). In contrast to the well-differentiated
329 groups, the fjord samples (except OSL, see above) consisted of either North Sea-like, or
330 western Baltic-like individuals when *K*=2 (Fig. 2a), or a distinct third genetic cluster when
331 *K*=3, which was mainly present in western Skagerrak fjords, and of these predominately
332 found in the samples Hellefjord (HEL) and Grenland (GRE) (Fig. 2b). This pattern is
333 concordant with the results of the principal component analysis (PCA), where the largest
334 variance was found between North Sea-like and western Baltic-like groups, and the
335 second-largest variance separated these groups from western Skagerrak fjord samples
336 (Note S2 and Fig. S3, Supporting information). Differentiation between North Sea and
337 Baltic-like groups was also evident based on neutral markers; however, this was not the
338 case for the third western fjord cluster (Fig. S3, Supporting information). In contrast,
339 using only diversifying SNPs, only randomly selected SNPs on larger scaffolds, or
340 excluding the most differentiated groups had no major influence on the three-cluster
341 pattern (Note S2 and Fig. S3, Supporting information).

342 All eastern, as well as many western Skagerrak fjord individuals were found either in the
343 North Sea-like or the western Baltic-like group, indicating a mechanical mix of individuals
344 from different sources. To differentiate between mechanical mixture and admixture, we
345 therefore applied an assignment approach as a second test in STRUCTURE, using the
346 well-differentiated North Sea and Kattegat samples as source populations. Per location

kernel density estimates showed unimodality, suggesting a single source of ancestry, for the well-differentiated populations: English Channel (ENG) (North Sea-like, *dip* = 0.040, $p > 0.05$), Skagerrak (SKA) (North Sea-like, *dip* = 0.068, $p > 0.05$), Oslofjord (OSL) (North Sea-like, *dip* = 0.039, $p > 0.05$), Öresund (ORE) (western Baltic-like, *dip* = 0.044, $p > 0.05$), and Belt Sea (BEL) (western Baltic-like, *dip* = 0.031, $p > 0.05$) (Fig. 2c, d). Significant bimodality suggesting ancestry from both source populations (NOR and KAT) was found for the western fjord sampling sites Tvedestrand (TVE) (*dip* 0.096, $p = 0.001$) and Soppekilen (SOP) (*dip* 0.107, $p < 0.01$), as well as the eastern fjord Iddefjord (IDD) (*dip* 0.095, $p = 0.001$) (Fig. 2c, d). Nevertheless, these three sampling sites also include individuals with genotypes intermediate between the two clusters with $q \sim 0.5$ (Fig. 2c). The two eastern Skagerrak fjords Gullmarsfjord (GUL) and Havstensfjord (HAV) also showed bimodal distributions; however, support for bimodality was non-significant (GUL: *dip* 0.050, $p > 0.05$; HAV: *dip* 0.083, $p > 0.05$). Samples from Hellefjord (HEL) and Grenland (GRE) were characterized by rather unimodal ancestry distributions, indicating a western Baltic-like origin (HEL: *dip* 0.052, $p > 0.05$; GRE: *dip* 0.909, $p > 0.05$). Whether these individuals are truly of Kattegat/western Baltic origin, or whether they originate from another non-sampled source population cannot be distinguished with this method.

To quantify genomic admixture of the two source populations within the fjord individuals by their hybrid indices (H), we performed Bayesian genomic cline analysis. The obtained hybrid indices largely corroborate the results of the STRUCTURE assignment approach (Fig. 2e and Table S7, Supporting information). By applying thresholds of $H \leq 0.25$ and ≥ 0.75, individuals were classified as pure North Sea or Kattegat ancestry. Based on these thresholds, Hellefjord (HEL) and Grenland (GRE) are unique as they possess the lowest proportions of individuals with inferred pure North Sea ancestry compared to all other fjords (HEL 0%, GRE 10.6%), the largest percentages of admixed individuals (GRE 59.6%, HEL 52.9%), and the largest proportions of individuals with inferred pure Kattegat ancestry (HEL 47.1%, GRE 29.8 %) (Table S8, Supporting information). In general, all fjords possess admixed individuals, albeit at lower proportions in Tvedestrand (TVE 34%), Soppekilen (SOP 32.1%), Iddefjord (IDD 34.8%), Gullmarsfjord (GUL 48.9%), and Havstensfjord (HAV 41.7%). In these fjords, mechanical mixing of individuals with different ancestries seems to dominate the population structure.

Pairwise fixation indices (F_{ST}) were calculated to characterize the population structure

between the different sampling sites and to assess the connectivity through isolation by distance (IBD) estimates. F_{ST} estimates were generally low (average pairwise F_{ST} 0.0031) but significant in almost three fourths of comparisons (Fig. 3a and Table S9, Supporting information). Comparatively high differentiation was estimated between the North Sea (NOR) and the western Baltic (ORE, BEL) samples (F_{ST} 0.0080-0.0084), but genetic differentiation between the English Channel (ENG) and the North Sea was weak (F_{ST} 0.0005) and not significant. The largest differentiation was found between the western Skagerrak sampling site Hellefjord (HEL) and the North Sea (F_{ST} 0.0130), but Hellefjord was similarly strongly differentiated from the English Channel, Skagerrak (SKA), and Oslofjord (OSL), as well as significantly differentiated from the western Baltic (F_{ST} 0.0030-0.0033) and eastern Skagerrak fjords (F_{ST} 0.0042-0.0068). Applying multidimensional scaling (MDS) to pairwise F_{ST} distances, this separation is evident by Hellefjord being furthest off both axes (Fig. 3b). The visualization of F_{ST} distances by MDS also revealed genetic distinction of the western Skagerrak fjord samples Soppekilen (SOP) and Grenland (GRE) in addition to Hellefjord (Fig. 3b), whereas the eastern Skagerrak fjord samples HAV and GUL are found intermediate between North Sea and Baltic-like groups. No significant differentiation could be detected between the western Baltic and the Kattegat (KAT) samples. In the MDS plot, this high similarity is apparent by the close proximity of these three locations (Fig. 3b).

Isolation by distance was assessed using a Mantel test among fjord sampling sites only, or including the reference populations, and considering either direct geographic distances between sampling coordinates or least-cost paths restricted to marine and shelf areas. However, no significant correlation was detected for any of the comparisons (Fig. S4, Supporting information). In summary, these results describe the presence of differentiated western Skagerrak fjord cod, and a mixed occurrence of North Sea and Kattegat cod within eastern Skagerrak fjords.

Biophysical connectivity modeling

The biophysical model of egg and larval dispersal suggested substantial and intermediate larval supply from the spawning areas in the North Sea to the western and the eastern Skagerrak coast, respectively, but low dispersal to the Kattegat (Fig. 4a, for spawning areas see Fig. S1, Supporting information). In contrast, the Kattegat and the small but relatively productive spawning areas in the Danish straits (belonging to the

western Baltic, see Fig. 1) may provide a large proportion of competent larvae along the eastern Skagerrak coast, but less dispersal to the western Skagerrak coast (Fig. 4a). The Kattegat itself appeared to largely rely on local spawning areas and import from the Danish straits (Fig. 4a). Similarly, local recruitment was also predicted along the western Skagerrak coast, although these values may be underestimates since the model does not resolve the complex geomorphology with high retention within fjords. No local recruitment was assumed for the eastern Skagerrak coast where spawning stocks are negligible (see Jonsson *et al.* 2016).

The fjords along the western Skagerrak coast received competent larvae from all considered spawning areas (Fig. 4b); however, the model predicted particularly large larval supply from the North Sea to the Oslofjord (OSL). This North Sea influence varies greatly between years (indicated by the SD in Fig. 4b) and is particularly strong during years with positive NAO winter index. There may also be larger connectivity of Tvedestrand (TVE) with the North Sea as compared to the Hellefjord (HEL) and Grenland (GRE). Notably, the model also predicted a substantial supply of Kattegat/Danish straits larvae to all studied western Skagerrak fjords (Fig. 4b). These results indicate that larval connectivity considerably influences the genetic population structure and that high connectivity and resulting gene flow may be negatively correlated with the potential for local adaptation.

Chromosomal rearrangements

Large genomic regions exhibiting strong linkage disequilibrium (LD) on several Atlantic cod chromosomes (linkage groups; LG) have recently been reported (Berg *et al.* 2015; 2016; Sodeland *et al.* 2016; Kirubakaran *et al.* 2016). Likely all of these regions represent large chromosomal inversions as suggested in previous studies (Berg *et al.* 2016; Sodeland *et al.* 2016), and empirically demonstrated for the linked region on LG1 (Kirubakaran *et al.* 2016). As our dataset was filtered for LD using a strict filtering cut-off ($r^2 > 0.1$), most SNPs within the rearranged regions were removed due to strong signals of LD, with the remaining ones not influencing the genetic structure (Fig. S5, Supporting information). However, as these genomic regions have been suggested to carry genes responsible for local adaptation to low salinity, temperature, and oxygen levels (Bradbury *et al.* 2010; Berg *et al.* 2015), these linked SNPs were used in separate analyses to investigate the occurrence and segregation of the chromosomal

443 rearrangements between sampling sites. Our data revealed three of the four putative
444 inversions previously described by Berg *et al.* (2015): LG2 (position 18,609,260 –
445 23,660,985; ~5.05 Mbp), LG7 (position 13,622,710 – 23,181,520; ~9.56 Mbp), and LG12
446 (position 426,531 – 13,445,150; ~13.02 Mbp). The inversion on LG1 has so far
447 exclusively been found in comparisons with the Northeast Arctic cod (Berg *et al.* 2016;
448 Kirubakaran *et al.* 2016), and was not detected in our data using the R package
449 inveRsion. However, a comparison of SNPs within the linked region on LG1 in our data
450 with the previously published data from Berg *et al.* (2016) revealed four heterozygous
451 individuals (0.76%) carrying both the inverted and the collinear allele (two from OSL, one
452 each from GRE and NOR).
453 Based on a bootstrap analysis, a significant overrepresentation of the rearranged allele
454 on LG2 was detected for the western Skagerrak fjords Hellefjord (HEL, $p < 0.001$) and
455 Grenland (GRE, $p < 0.001$), as well as for the Öresund (ORE, $p < 0.001$) (Fig. 5a). The
456 rearranged allele on LG7 was not found to be significantly overrepresented in any of the
457 sampling sites (Fig. 5b). However, the rearranged allele on LG12 was significantly
458 overrepresented in the North Sea (NOR, $p < 0.001$), the Oslofjord (OSL, $p < 0.001$), and
459 also the Skagerrak (SKA, $p < 0.05$; not significant after correction for multiple
460 comparisons) (Fig. 5c). In addition, the geographically most distant English Channel
461 (ENG) exhibited a significant underrepresentation of the rearranged alleles for all three
462 LGs ($p < 0.001$). Comparisons of the occurrence of the rearranged alleles in all western
463 fjords (TVE, SOP, HEL, GRE) and all eastern fjords (IDD, GUL, HAV) revealed a
464 significant overrepresentation of the rearranged allele on LG2 within western fjords ($p <$
465 0.001), but not within eastern fjords. Since the Oslofjord clustered with the North Sea
466 group it was excluded from this comparison; however, the rearranged allele on LG2 was
467 also significantly overrepresented ($p < 0.01$) when the Oslofjord was included within
468 the western fjords. In summary, these findings suggest that the particular genomic
469 architecture of Atlantic cod contributes to the potential for local adaptation to a low
470 salinity environment.

471 **Discussion**

472 How local adaptation can be maintained in the face of gene flow is a long-standing
473 question in evolutionary biology, which we are now beginning to understand owing to the

474 profound advances in sequencing technology and genomic analysis tools (Tigano &
475 Friesen 2016). While it is well recognized that chromosomal inversions can play an
476 important role as drivers of evolution (reviewed in Hoffmann and Rieseberg 2008), there
477 are still few studies investigating the role of chromosomal rearrangements in high gene
478 flow species. Marine organisms provide ideal models to study this question, owing to
479 their varied habitats and the lack of physical barriers. By combining genomic analyses of
480 ecologically distinct Atlantic cod populations with biophysical modeling of dispersal, we
481 were not only able to unravel cryptic population structure and detect ecologically
482 differentiated populations, but also identified chromosomal rearrangements as a
483 potential mechanism enabling local adaptation despite high connectivity.

484 **Western Skagerrak fjords possess locally differentiated Atlantic cod despite high**
485 **connectivity and a mix of North Sea and Kattegat cod**.
486 The ecological peculiarity of the low saline Baltic Sea and the transition zone connecting
487 it with the saline North Sea have led to the evolution of unique linages (Johannesson &
488 André 2006). Nevertheless, based on unlinked SNPs, the overall population
489 differentiation of Atlantic cod within this area was weak, as also shown in earlier studies
490 and explained by large effective population sizes and high gene flow (Nielsen *et al.*
491 2005; Knutsen *et al.* 2011). Comparatively strong differentiation was detected between
492 North Sea/English Channel/Skagerrak and Kattegat/western Baltic samples, reflecting
493 the geographical separation (Fig. 1) as well as a separation resulting from adaptation to
494 low-salinity as shown previously for Atlantic cod, but also many other species of the
495 eastern Baltic Sea (Johannesson & André 2006; Lamichhaney *et al*. 2012; Berg *et al.*
496 2015; Sjöqvist *et al*. 2015). However, no genetic differentiation was detected within these
497 strongly separated North Sea-like and western Baltic-like groups (Note S3, Supporting
498 information).
499 Contrary to these well defined populations, the eastern Skagerrak fjords appeared to be
500 composed of a mix between North Sea-like and western Baltic-like individuals, indicating
501 that these fjords are part of the distributional area of the two major evolutionary units
502 detected in this study. These fjords may experience larval recruitment through a strong
503 influx of central North Sea water into the Skagerrak, as well as less-saline Kattegat water
504 entering along the coast (Danielssen *et al.* 1997; Knutsen *et al.* 2004; Stenseth *et al.*
505 2006; André *et al.* 2016; Jonsson *et al.* 2016). In agreement with these predominant

ocean currents, a large fraction of individuals from the eastern Skagerrak fjords appeared to be of North Sea origin (Fig. 2), while our biophysical model suggested greater larval connectivity with the Kattegat and western Baltic. However, the model did not include the North Sea Viking bank spawning ground which has significantly increased its contribution during the last decades (Jonsson *et al.* 2016), suggesting that the influence of the North Sea spawning areas to the eastern Skagerrak is larger than shown in our modeling. We did not detect genetically differentiated individuals that would be indicative for a distinct fjord population in eastern Skagerrak fjords, although differentiation between Atlantic cod larvae inside and outside Gullmarsfjord was previously found (Øresland & André 2008). It is unknown if recent reductions in abundance along the eastern Skagerrak coast (Svedäng & Bardon 2003; Svedäng & Svenson 2006) indicate the loss or severe decimation of a genetically differentiated population in this region.

In contrast, the western Skagerrak fjord samples included varying levels of genetically differentiated individuals that clustered neither with the North Sea-like nor the western Baltic-like group (Fig. 2b), indicative of the existence of a local western Skagerrak coastal or fjord cod population(s). The existence of such local populations is also supported by the biophysical model results, which explained a large fraction of larval supply by local recruitment (Fig. 4). Local fjord cod has previously also been assumed to exist at the northern Norwegian coast (Jørstad & Naevdal 1989; Myksvoll *et al.* 2014), and differentiation between fjord, coastal, or oceanic cod has been shown in two closely related gadiids, the Pacific cod (*Gadus macrocephalus*) and the polar cod (*Boreogadus saida*) (Cunningham *et al.* 2009; Madsen *et al.* 2015).

Fjord systems represent semi-enclosed ecosystems where water exchange is restricted by a narrow connection with the outer sea, often further reduced by a tall entrance sill, thus creating an inner estuarine circulation (Howe *et al.* 2010). Such conditions have been shown to hamper gene flow as a result of stationary behavior with reduced adult migration and restricted egg and larval dispersal (Knutsen *et al.* 2007; Bergstad *et al.* 2008; Espeland *et al.* 2007; Espeland *et al.* 2008; Ciannelli *et al.* 2010; Jung *et al.* 2012; Rogers *et al.* 2014). Consequently, the strongest genetic differentiation and the largest fraction of local western Skagerrak fjord individuals was found in the particularly isolated Hellefjord (Molvær *et al.* 1978) and Grenland fjord (Danielssen & Føyn 1973) (Fig. 2b). Although the differentiation of the Hellefjord sample might be overestimated due to the

small sample size and collection of juveniles, these results were strongly supported by the Grenland fjord sample, consisting of a large sample of adults collected during both spawning and non-spawning periods. However, weaker genetic differentiation was estimated for the Tvedestrand and Soppekilen samples, which may be attributed to bathymetric and temporal differences (Note S4, Supporting information).

Interestingly, the majority of the Oslofjord individuals were assigned a North Sea origin in the ancestry analysis (Fig. 2e), a pattern largely supported by the biophysical model (Fig. 4b). However, strong contribution from the Kattegat/western Baltic was also predicted by the model but was not as evident in the genetic results, possibly indicating the lack of the North Sea Viking bank spawning ground in the model. In contrast to the Oslofjord, all western Skagerrak fjords showed a lower percentage of individuals with North Sea origin and about one quarter were assigned Kattegat/western Baltic origin. This result supports the suggestion that spawning areas in the Danish straits and especially in the Öresund may constitute an important source of Atlantic cod larvae for both the eastern and the western Skagerrak (Jonsson *et al.* 2016).

Do chromosomal rearrangements facilitate ecological adaptation of Atlantic cod?

Atlantic cod can be found in a variety of different habitats, ranging from the relatively warm waters in the Bay of Biscay, from small sheltered coastal and fjord ecosystems, to low-saline seas like the Baltic Sea, and to open oceanic environments and very cold waters in the Barents Sea (Lilly *et al.* 2008), an environmental flexibility that likely required the acquisition of locally adaptive traits. It has recently been described that such adaptations, especially in highly connected organisms like oceanic fish, can arise through the segregation of chromosomal rearrangements, where recombination is suppressed and important functional genes are inherited together (Feder *et al.* 2012; Thompson & Jiggins 2014, Tigano & Friesen 2016). While empirical evidence for this theory is still scarce, it is well supported by studies on stickleback (Jones *et al.* 2012, Roesti *et al.* 2015). Recently, haplotype blocks associated with ecological adaptation were also detected in the Atlantic herring, but it is unclear if inversions are the causative mechanism (Martinez-Barrio *et al.* 2016; Lamichhaney *et al.* 2017). In contrast, a series of recent studies employing genome-wide data to dissect Atlantic cod population differentiation, discovered exceptionally large chromosomal rearrangements that are likely to be inversions on several linkage groups (LGs), which were suggested to play a

major role for the adaptive abilities of Atlantic cod (Bradbury *et al.* 2013; Hemmer-Hansen *et al.* 2013; Bradbury *et al.* 2014; Berg *et al.* 2015; 2016; Sodeland *et al.* 2016; Kirubakaran *et al.* 2016; Barney *et al.* 2017). These recent studies, including this study, therefore contribute remarkable examples in the marine environment to a growing body of literature identifying chromosomal rearrangements and inversions as an important mechanism to maintain contrasting ecotypes in intermixing populations (Hoffmann and Rieseberg 2008; Lowry and Willis 2010; Joron *et al.* 2011; Cheng *et al.* 2012).

For example, adaptation to low saline and hypoxic environments as occurring in the Baltic Sea strongly depends on the ability for osmoregulation and effective oxygen management (Andersen *et al.* 2009; Berg *et al.* 2015). Berg and coauthors (2015) compared North and Baltic Sea cod and found several SNPs within genes important for salinity and oxygen regulation, of which the majority reside within a rearranged region on LG2, implicating an essential role of this rearranged region for the Atlantic cod's ability to adapt to the environmental conditions in the Baltic Sea. Such genetic-environment correlations may also be due to intrinsic genetic incompatibilities that merely coincide with ecological barriers (Bierne *et al.* 2011). However, similar patterns of genes involved in oxygen- or osmoregulation were also associated with salinity clines in studies of Atlantic herring (Limborg *et al.* 2012; Martinez-Barrio *et al.* 2016), indicating the presence of true local adaptation.

Remarkably, fjord ecosystems have notable similarities with the Baltic Sea: both originated by glacial retreat, represent enclosed estuaries with high freshwater input and restricted exchange with saline oceanic water leading to estuarine circulations, and both feature deep basins with mostly hypoxic conditions (Howe *et al*. 2010; Harff *et al*. 2011). Thus, similar adaptations may be required for successful colonization of the Baltic Sea and fjord ecosystems. Indeed, our ancestry analyses showed that local western Skagerrak fjord individuals are genetically more similar to the Kattegat/western Baltic population (an area discussed as a transition zone between the North Sea and the eastern Baltic Sea (Nielsen *et al.* 2003)) than to the North Sea population. In addition, we found a significant overrepresentation of the rearranged LG2 allele in the Hellefjord and Grenland fjord samples (Fig. 5a), an allelic shift that has recently also been described between oceanic and coastal cod groups (Sodeland *et al.* 2016). Both fjords have high freshwater influx, causing a low-saline surface layer above oceanic water with 25-30‰ salinity (Danielssen & Føyn 1973; Molvær *et al.* 1978), comparable to salinity

gradients in the Kattegat/western Baltic (Madsen & Højerslev 2009). As an adaptation to low-saline conditions, Atlantic cod inhabiting the Baltic Sea produce highly hydrated eggs that are neutrally buoyant between ~14‰ (eastern Baltic Sea) and ~21‰ (Danish straits) (Nissling & Westin 1997, for a recent review see Hüssy *et al.* 2011), a mechanism that for example prevents lethal sinking of the eggs to the hypoxic deeper layers in the Baltic Sea. In contrast, the eggs of marine Atlantic cod populations are neutrally buoyant at salinities of ~33‰ (Thorsen *et al.* 1996). Similar to Baltic cod, eggs of fjord cod are neutrally buoyant in the low-saline water layers of fjords, which not only prevents sinking of the eggs to hypoxic layers, but also retains the eggs inside the sheltered fjord area (Espeland *et al.* 2007; Knutsen *et al.* 2007; Ciannelli *et al.* 2010; Jung *et al.* 2012). Egg buoyancy can be regulated by the in- and efflux of solutes (Reading *et al.* 2012), and many SNPs in or close to genes coding for membrane trafficking proteins have been identified within the rearranged region on LG2 (Berg *et al.* 2015). This accumulation of adaptive variation could be explained by diversifying selection shaping the rearranged region in the likely absence of recombination between the alleles. In ecosystems where regulation of egg buoyancy provides an evolutionary advantage, an increase in the frequency of the rearrangement might be expected.

In addition to our samples from Hellefjord and Grenland fjord, our Öresund sample from the western Baltic also shared a significant overrepresentation of the rearranged allele on LG2, which occurs at very high frequency in eastern Baltic cod (Berg *et al.* 2015). However, our Belt Sea and Kattegat samples did not show an increased occurrence of the rearranged LG2 allele although the genetic structure analyses suggested genetic similarity between the Kattegat and western Baltic samples, indicative for additional adaptive variation outside the large rearrangements. Interestingly, the rearranged LG12 allele was found to be significantly overrepresented in our North Sea and Oslofjord samples, with high occurrences also in the eastern Skagerrak sample (Fig. 5c). Concordantly, this allele was recently found to occur at higher frequency in oceanic compared to coastal Atlantic cod populations and was suggested to play a role in ecological adaptation (Sodeland *et al.* 2016). It has previously also been associated with an adaptation to temperature (Bradbury *et al.* 2010; Berg *et al.* 2015), which could thus be relevant with regard to survival and abundance of Atlantic cod in the face of global warming (Drinkwater 2005). However, similar to the Kattegat/western Baltic samples, which shared most genetic variation but showed a distinct pattern in the occurrence of

637 the rearranged LG2 allele, the North Sea, Oslofjord, Skagerrak, and English Channel
638 samples were not distinguishable based on SNPs outside the rearranged regions, but
639 showed a distinct distribution of the rearranged LG12 allele. This contrast between the
640 genome-wide profile that rather reflects connectivity and geography, and the
641 chromosomal rearrangements that seem to cluster according to environment, indicates
642 that despite the high gene flow between Atlantic cod populations important genes under
643 adaptive divergent selection likely reside within rearranged regions.

644 **Significance and summary of the study**

645 Because of their relatively higher fitness in their native habitat compared to introduced
646 populations, locally adapted populations are often irreplaceable once vanished (Kawecki
647 & Ebert 2004; Reiss *et al.* 2009). Human activity has led to the collapse of several fish
648 stocks (Myers *et al.* 1996; Pinsky *et al.* 2011) and populations of Atlantic cod regionally
649 suffer from overexploitation and population decline (Svedäng & Bardon 2003; Svedäng &
650 Svenson 2006; Bartolino *et al.* 2012; Bonanomi *et al.* 2015), causing predator-prey shifts
651 and imbalance of sensible ecosystems (Baden *et al.* 2012; Östman *et al.* 2016). Thus,
652 priorities are high to clarify the potential and occurrence of local adaptation in such high
653 gene flow species, as well as to improve our understanding of the genetic mechanism for
654 adaptation to conserve genetic resources in a globally changing world.
655 Our study showed that: 1.) The here described North Sea, Kattegat/western Baltic, and
656 western Skagerrak fjord cod genotypes most likely correspond to the previously
657 identified oceanic and coastal ecotypes, respectively, thus shedding light on the long-
658 standing question whether local fjord ecotypes exist, and 2.) western Skagerrak fjord
659 cod, despite high connectivity with the North Sea, may possess adaptations facilitating a
660 life in a low-salinity environment similar to Atlantic cod from the Baltic Sea. The genes
661 encoding these adaptations are suggested to partially reside in large chromosomal
662 rearrangements, regions that due to their reduced recombination are known to promote
663 adaptive population divergence (Kirkpatrick & Barton 2006; Feder & Nosil 2009;
664 Thompson & Jiggins 2014).
665 In contrast, no locally differentiated fjord cod was detected in the eastern Skagerrak
666 fjords, supporting the absence or suspected loss of local populations along the Swedish
667 coast (Svedäng & Bardon 2003). We thus emphasize the importance of taking genome

architecture into account when characterizing ecological adaptation, particularly for species characterized by high gene flow.

Acknowledgements

We thank Mariann Arnyasi, Matthew P. Kent, and Sigbjørn Lien (Norwegian University of Life Sciences, CIGENE) for SNP genotyping, and Michael M. Hansen, Daniel Ruzzante, and two anonymous reviewers for valuable input that greatly helped to improve the manuscript. Initial sequencing for SNP identification was provided by the Norwegian Sequencing Centre. JMIB thanks Michael Matschiner for valuable input to the manuscript and careful proofreading, Trond Reitan for support with statistical analyses, and the Centre for Ecological and Evolutionary Synthesis (CEES) at the University of Oslo (UiO) for funding. This work was performed on the Abel Cluster owned by the UiO and the Norwegian metacenter for High Performance Computing (NOTUR) and operated by the UiO Department for Research Computing, and was supported by funds of the Interreg projects “CodS - restoration and management of cod populations in Skagerrak-Kattegat” (#168975) and “MarGen” (#175806), and by the Centre for Marine Evolutionary Biology at the University of Gothenburg (CeMEB).

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.

Amante C, Eakins BW (2009) ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. *NOAA Technical Memorandum*, NESDIS NGDC-24.

Andersen Ø (2012) Hemoglobin polymorphisms in Atlantic cod - A review of 50 years of study. *Marine Genomics*, **8**, 59–65.

Andersen Ø, Wetten OF, De Rosa MC *et al.* (2009) Haemoglobin polymorphisms affect the oxygen-binding properties in Atlantic cod populations. *Proceedings of the Royal Society B*, **276**, 833–841.

André C, Svedäng H, Knutsen H *et al.* (2016) Population structure in Atlantic cod in the eastern North Sea-Skagerrak-Kattegat: early life stage dispersal and adult migration. *BMC Research Notes*, **9**, 63.

Baden S, Emanuelsson A, Pihl L, Svensson CJ, Åberg P (2012) Shift in seagrass food web structure over decades is linked to overfishing. *Marine Ecology Progress Series*, **451**, 61–73.

Barney BT, Munkholm C, Walt DR, Palumbi SR (2017) Highly localized divergence within supergenes in Atlantic cod (*Gadus morhua*) within the Gulf of Maine. *BMC Genomics*, **18**, 271.

703 Bartolino V, Cardinale M, Svedäng H *et al.* (2012) Historical spatiotemporal dynamics of
704 eastern North Sea cod. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**,
705 833–841.
706 Beacham TD, Brattey J, Miller KM, Le KD, Withler RE (2002) Multiple stock structure of
707 Atlantic cod (*Gadus morhua*) off Newfoundland and Labrador determined from
708 genetic variation. *ICES Journal of Marine Science*, **59**, 650–665.
709 Benjamini Y and Yekutieli D (2001) The control of the false discovery rate in multiple
710 testing under dependency. *The Annals of Statistics*, **29**, 1165–1188.
711 Berg PR, Jentoft S, Star B *et al.* (2015) Adaptation to low salinity promotes genomic
712 divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biology and Evolution*, **7**,
713 1644–1663.
714 Berg PR, Star B, Pampoulie C *et al.* (2016) Three chromosomal rearrangements
715 promote genomic divergence between migratory and stationary ecotypes of Atlantic
716 cod. *Scientific reports*, **6**, 23246.
717 Bergstad OA, Jørgensen T, Knutsen JA, Berge JA (2008) Site fidelity of Atlantic cod
718 *Gadus morhua* L. as deduced from telemetry and stable isotope studies. *Journal of*
719 *Fish Biology*, **72**, 131–142.
720 Bernatchez L (2016) On the maintenance of genetic variation and adaptation to
721 environmental change: considerations from population genomics in fishes. *Journal of*
722 *Fish Biology*, **89**, 2519–2556.
723 Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why
724 genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**,
725 2044–2072.
726 Bonanomi S, Overgaard Therkildsen N, Retzel A *et al.* (2016) Historical DNA documents
727 long-distance natal homing in marine fish. *Molecular Ecology*, **25**, 2727–2734.
728 Bonanomi S, Pellissier L, Therkildsen NO *et al.* (2015) Archived DNA reveals fisheries
729 and climate induced collapse of a major fishery. *Scientific Reports*, **5**, 15395.
730 Bradbury IR, Bowman S, Borza T *et al.* (2014) Long distance linkage disequilibrium and
731 limited hybridization suggest cryptic speciation in Atlantic cod. *PLoS ONE*, **9**,
732 e106380.
733 Bradbury IR, Hubert S, Higgins B *et al.* (2010) Parallel adaptive evolution of Atlantic cod
734 on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the*
735 *Royal Society B*, **277**, 3725–3734.
736 Bradbury IR, Hubert S, Higgins B *et al.* (2013) Genomic islands of divergence and their
737 consequences for the resolution of spatial structure in an exploited marine fish.
738 *Evolutionary Applications*, **6**, 450–461.
739 Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and
740 applications. *BMC Bioinformatics*, **10**, 421.
741 Cáceres A, Sindi SS, Raphael BJ, Cáceres M, González JR (2012) Identification of
742 polymorphic inversions from genotypes. *BMC Bioinformatics*, **13**, 28.
743 Cheng C, White BJ, Kamdem C et al. (2012) Ecological genomics of *Anopheles gambiae*
744 along a latitudinal cline: a population-resequencing approach. *Genetics*, **190**, 1417–
745 1432.
746 Ciannelli L, Knutsen H, Olsen EM *et al.* (2010) Small-scale genetic structure in a marine
747 population in relation to water circulation and egg characteristics. *Ecology*, **91**,
748 2918–2930.
749 Clarke LM, Munch SB, Thorrold SR, Conover DO (2010) High connectivity among locally
750 adapted populations of a marine fish (*Menidia menidia*). *Ecology*, **91**, 3526–3537.
751 Cunningham KM, Canino MF, Spies IB, Hauser L (2009) Genetic isolation by distance

752 and localized fjord population structure in Pacific cod (*Gadus macrocephalus*):
753 limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of*
754 *Fisheries and Aquatic Sciences*, **66**, 153–166.
755 Danielssen DS, Føyn L (1973) Frierfjorden - en vurdering av fjordsystemets
756 vannutskiftning. *Fisken og Havet*, **6**, 1–19.
757 Danielssen DS, Edler L, Fonselius S *et al.* (1997) Oceanographic variability in the
758 Skagerrak and Northern Kattegat, May-June, 1990. *ICES Journal of Marine Science*,
759 **54**, 753–773.
760 DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and
761 anadromous fishes compared with other animals. *Journal of Fish Biology*, **56**, 461–
762 473.
763 Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of*
764 *Vegetation Science*, **14**, 927–930.
765 Drinkwater KF (2005) The response of Atlantic cod (*Gadus morhua*) to future climate
766 change. *ICES Journal of Marine Science*, **62**, 1327–1337.
767 Dobzhansky T (1937) Genetics and the origin of species. *Columbia University Press,*
768 *New York, USA*.
769 Efron B (1979) Bootstrap methods: another look at the jackknife. *The Annals of*
770 *Statistics*, **7**, 1-26.
771 Espeland SH, Gundersen AF, Olsen EM *et al.* (2007) Home range and elevated egg
772 densities within an inshore spawning ground of coastal cod. *ICES Journal of Marine*
773 *Science*, **64**, 920–928.
774 Espeland SH, Olsen EM, Knutsen H *et al.* (2008) New perspectives on fish movement:
775 kernel and GAM smoothers applied to a century of tagging data on coastal Atlantic
776 cod. *Marine Ecology Progress Series*, **372**, 231–241.
777 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals
778 using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–
779 2620.
780 Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to
781 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
782 *Resources*, **10**, 564–567.
783 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using
784 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*,
785 **164**, 1567–1587.
786 Feder JL, Nosil P (2009) Chromosomal inversions and species differences: when are
787 genes affecting adaptive divergence and reproductive isolation expected to reside
788 within inversions? *Evolution*, **63**, 3061–3075.
789 Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in*
790 *Genetics*, **28**, 342–350.
791 Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate
792 for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**,
793 977–993.
794 Gompert Z, Buerkle CA (2012) bgc: Software for Bayesian estimation of genomic clines.
795 *Molecular Ecology Resources*, **12**, 1168–1176.
796 Harff J, Björck, S, Hoth, P (2011) The Baltic Sea Basin. In: Central and Eastern
797 European Development Studies. Eds Harff J, Björck, S, Hoth, P. *Springer-Verlag*
798 *Berlin, Heidelberg, Germany*, 0-449.
799 Hartigan JA, Hartigan PM (1985) The dip test of unimodality. *The Annals of Statistics*,
800 **13**, 70–84.

Hemmer-Hansen J, Nielsen EE, Therkildsen NO *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, **22**, 2653–2667.
Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, **39**, 21–42.
Hordoir R, Dieterich C, Basu C, Dietze H, Meier HEM (2013) Freshwater outflow of the Baltic Sea and transport in the Norwegian current: A statistical correlation analysis based on a numerical experiment. *Continental Shelf Research*, **64**, 1–9.
Howe JA, Austin WEN, Forwick M *et al.* (2010) Fjord systems and archives: a review. In: Fjord systems and archives. Eds Howe JA, Austin WEN, Forwick M, Paetzel M. *Geological Society, London, Special Publications*, **344**, 5–15.
Hubert S, Higgins B, Borza T, Bowman S (2010) Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genomics*, **11**, 191.
Hüssy K (2011) Review of western Baltic cod (*Gadus morhua*) recruitment dynamics. *ICES Journal of Marine Science*, **68**, 1459–1471.
Johannesson K, André C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, **15**, 2013–2029.
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.
Jonsson PR, Corell H, André C, Svedäng H, Moksnes P (2016) Recent decline in cod stocks in the North Sea–Skagerrak–Kattegat shifts the sources of larval supply. *Fisheries Oceanography*, **25**, 210–228.
Jorde PE, Knutsen H, Espeland SH, Stenseth NC (2007) Spatial scale of genetic structuring in coastal cod *Gadus morhua* and geographic extent of local populations. *Marine Ecology Progress Series*, **343**, 229–237.
Joron M, Frezal L, Jones RT et al. (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–206.
Jung K-M, Folkvord A, Kjesbu OS *et al.* (2012) Egg buoyancy variability in local populations of Atlantic cod (*Gadus morhua*). *Marine Biology*, **159**, 1969–1980.
Jørstad KE, Naevdal G (1989) Genetic variation and population structure of cod, *Gadus morhua* L, in some fjords in northern Norway. *Journal of Fish Biology*, **35**, 245–252.
Karlsen BO, Klingan K, Emblem Å *et al.* (2013) Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, **22**, 5098-5111.
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.
Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, **4**, 782–788.
Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, **173**, 419–434.
Kirubakaran TG, Grove H, Kent MP *et al.* (2016) Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, **25**, 2130–2143.
Knutsen H, André C, Jorde PE *et al.* (2004) Transport of North Sea cod larvae into the Skagerrak coastal populations. *Proceedings of the Royal Society B*, **271**, 1337–

1344.
Knutsen H, Jorde PE, André C, Stenseth NC (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology*, **12**, 385–394.
Knutsen H, Olsen EM, Ciannelli L *et al.* (2007) Egg distribution, bottom topography and small-scale cod population structure in a coastal marine system. *Marine Ecology Progress Series*, **333**, 249–255.
Knutsen H, Olsen EM, Jorde PE *et al.* (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.
Lamichhaney S, Fuentes-Pardo AP, Rafati N *et al.* (2017) Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proceedings of the National Academy of Sciences*, **114**, E3452–E3461.
Lamichhaney S, Martinez-Barrio A, Rafati N *et al.* (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, **109**, 19345–19350.
Lilly GR, Wieland K, Rothschild BJ *et al.* (2008) Decline and recovery of Atlantic cod (*Gadus morhua*) stocks throughout the North Atlantic. In: Resiliency of gadid stocks to fishing and climate change. Eds Kruse GH, Drinkwater K, Ianelli JN, Link JS, Stram DL, Wespestad V and Woodby D. *University of Alaska Fairbanks, Alaska Sea Grant*, 39–66.
Limborg MT, Helyar SJ, De Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, **21**, 3686–3703.
Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.
Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology*, **8**, e1000500.
Ma J, Amos CI (2012) Investigation of inversion polymorphisms in the human genome using principal components analysis. *PLoS ONE*, **7**, e40224.
Mächler M (2014) diptest v0.75-6, available from https://cran.r-project.org/web/packages/diptest.
Madsen KS, Højerslev NK (2009) Long-term temperature and salinity records from the Baltic Sea transition zone. *Boreal Environment Research*, **14**, 125–131.
Madsen ML, Nelson RJ, Fevolden SE, Christiansen JS, Præbel K (2015) Population genetic analysis of Euro-Arctic polar cod *Boreogadus saida* suggests fjord and oceanic structuring. *Polar Biology*, **39**, 969-980.
Mayr E (1942) Systematics and the origin of species from the viewpoint of a zoologist. *Columbia University Press, New York, USA*.
Meier HEM (2007) Modeling the pathways and ages of inflowing salt- and freshwater in the Baltic Sea. *Estuarine, Coastal and Shelf Science*, **74**, 610–627.
Marshall J, Kushnir Y, Battisti D *et al.* (2001) North Atlantic climate variability: phenomena, impacts and mechanisms. *International Journal of Climatology*, **21**, 1863–1898.
Martinez-Barrio A, Lamichhaney S, Fan G *et al.* (2016) The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *eLife*, **5,** e12081.
Milano I, Babbucci M, Cariani A *et al.* (2014) Outlier SNP markers reveal fine-scale

genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, **23**, 118–135.
Moller D (1966) Genetic differences between cod groups in the Lofoten area. *Nature*, **212**, 824.
Molvær J, Green N, Baalsrud K (1978) Orienterende hydrokjenisk og biologisk undersøkelse av Hellefjorden, Kragerø. *Norsk Institutt for vannforsning (NIVA)*, 1–49.
Myers RA, Hutchings JA, Barrowman NJ (1996) Hypotheses for the decline of cod in the North Atlantic. *Marine Ecology Progress Series*, **138**, 293–308.
Myksvoll MS, Jung K-M, Albretsen J, Sundby S (2014) Modelling dispersal of eggs and quantifying connectivity among Norwegian coastal cod subpopulations. *ICES Journal of Marine Science*, **71**, 957–969.
Neuenfeldt S, Righton D, Neat F *et al.* (2013) Analysing migrations of Atlantic cod *Gadus morhua* in the north-east Atlantic Ocean: then, now and the future. *Journal of Fish Biology*, **82**, 741–763.
Nielsen EE, Grønkjaer P, Meldrup D, Paulsen H (2005) Retention of juveniles within a hybrid zone between North Sea and Baltic Sea Atlantic cod (*Gadus morhua*). *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 2219–2225.
Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Grønkjaer P (2003) Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Molecular Ecology*, **12**, 1497–1508.
Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.
Nissling A, Westin L (1997) Salinity requirements for successful spawning of Baltic and Belt Sea cod and the potential for cod stock interactions in the Baltic Sea. *Marine Ecology Progress Series*, **152**, 261–271.
O'Dea EJ, Arnold AK, Edwards KP *et al.* (2014) An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf. *Journal of Operational Oceanography*, **5**, 3-17.
Olsen EM, Knutsen H, Gjøsæter J *et al.* (2004) Life-history variation among local populations of Atlantic cod from the Norwegian Skagerrak coast. *Journal of Fish Biology*, **64**, 1725–1730.
Östman Ö, Eklöf J, Eriksson BK *et al.* (2016) Top-down control as important as nutrient enrichment for eutrophication effects in North Atlantic coastal ecosystems. *Journal of Applied Ecology*, **53**, 1138–1147.
Pante E, Simon-Bouhet B (2013) marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS ONE*, **8**, e73051.
Pinsky ML, Jensen OP, Ricard D, Palumbi SR (2011) Unexpected patterns of fisheries collapse in the world's oceans. *Proceedings of the National Academy of Sciences*, **108**, 8317–8322.
Pinsky ML, Palumbi SR (2014) Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, **23**, 29–39.
Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution*, **55**, 131–146.
Porras-Hurtado L, Ruiz Y, Santos C *et al.* (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, **4**, 98.

948 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
949 multilocus genotype data. *Genetics*, **155**, 945–959.
950 Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome
951 association and population-based linkage analyses. *American Journal of Human*
952 *Genetics*, **81**, 559–575.
953 R Core Team, R Foundation for Statistical Computing (2016) R: A Language and
954 Environment for Statistical Computing. Available from https://www.r-project.org/.
955 Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6, available from
956 http://beast.bio.ed.ac.uk/Tracer
957 Reading BJ, Chapman RW, Schaff JE *et al.* (2012) An ovary transcriptome for all
958 maturational stages of the striped bass (*Morone saxatilis*), a highly advanced
959 perciform fish. *BMC Research Notes*, **5**, 111.
960 Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ (2009) Genetic population structure of
961 marine fish: mismatch between biological and fisheries management units. *Fish and*
962 *Fisheries*, **10**, 361–395.
963 Reusch TBH (2014) Climate change in the oceans: evolutionary versus phenotypically
964 plastic responses of marine animals and plants. *Evolutionary Applications*, **7**, 104–
965 122.
966 Rice WR (1988) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.
967 Roesti M, Kueng B, Moser D, Berner D (2015) The genomics of ecological vicariance in
968 threespine stickleback fish. *Nature Communications*, **6**, 8767.
969 Rogers LA, Olsen EM, Knutsen H, Stenseth NC (2014) Habitat effects on population
970 connectivity in a coastal seascape. *Marine Ecology Progress Series*, **511**, 153–163.
971 Ross SD, Behrens JW, Brander K, Methling C, Mork J (2013) Haemoglobin genotypes in
972 cod (*Gadus morhua* L): their geographic distribution and physiological significance.
973 *Comparative Biochemistry and Physiology, Part A*, **166**, 158–168.
974 Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation.
975 *Nature Reviews Genetics*, **14**, 807–820.
976 Schwander T, Libbrecht R, Keller L (2014) Supergenes and complex phenotypes.
977 *Current Biology*, **24**, R288–R294.
978 Sick K (1961) Haemoglobin polymorphism in fishes. *Nature*, **192**, 894–696.
979 Sick K (1965) Haemoglobin polymorphism of cod in the North Sea and the North Atlantic
980 Ocean. *Hereditas*, **54**, 49–69.
981 Sjöqvist C, Godhe A, Jonsson PR, Sundqvist L, Kremp A (2015) Local adaptation and
982 oceanographic connectivity patterns explain genetic differentiation of a marine
983 diatom across the North Sea-Baltic Sea salinity gradient. *Molecular Ecology*, **24**,
984 2871–2885.
985 Skjæraasen JE, Meager JJ, Karlsen Ø, Hutchings JA, Fernö A (2011) Extreme
986 spawning-site fidelity in Atlantic cod. *ICES Journal of Marine Science*, **68**, 1427–
987 1477.
988 Sodeland M, Jorde PE, Lien S *et al.* (2016) "Islands of Divergence" in the Atlantic cod
989 genome represent polymorphic chromosomal rearrangements. *Genome Biology and*
990 *Evolution*, **8**, 1012–1022.
991 Star B, Nederbragt AJ, Jentoft S *et al.* (2011) The genome sequence of Atlantic cod
992 reveals a unique immune system. *Nature*, **477**, 207–210.
993 Stenseth NC, Jorde PE, Chan K-S *et al.* (2006) Ecological and genetic impact of Atlantic
994 cod larval drift in the Skagerrak. *Proceedings of the Royal Society B*, **273**, 1085–
995 1092.
996 Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation

and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B*, **66**, 187–205.
Svedäng H, Bardon G (2003) Spatial and temporal aspects of the decline in cod (*Gadus morhua* L.) abundance in the Kattegat and eastern Skagerrak. *ICES Journal of Marine Science*, **60**, 32–37.
Svedäng H, Svenson A (2006) Cod *Gadus morhua* L. populations as behavioural units: inference from time series on juvenile abundance in the eastern Skagerrak. *Journal of Fish Biology*, **69**, 151–164.
Svedäng H, André C, Jonsson P, Elfman M, Limburg KE (2010) Migratory behaviour and otolith chemistry suggest fine-scale sub-population structure within a genetically homogenous Atlantic cod population. *Environmental Biology of Fishes*, **89**, 383–397.
Svedäng H, Righton D, Jonsson P (2007) Migratory behaviour of Atlantic cod *Gadus morhua*: natal homing is the prime stock-separating mechanism. *Marine Ecology Progress Series*, **345,** 1-12.
Therkildsen NO, Hemmer-Hansen J, Als TD *et al*. (2013) Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. Molecular Ecology, **22**, 2424–2440.
Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. *Heredity*, **113**, 1–8.
Thorsen A, Kjesbu OS, Fyhndr HJ, Solemdal P (1996) Physiological mechanisms of buoyancy in eggs from brackish water cod. *Journal of Fish Biology*, **48**, 457–477.
Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular Ecology*, **25**, 2144–2164.
Tørresen OK, Star B, Jentoft S *et al.* (2017) An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics*, **18**, 95.
Uppala SM, Kållberg PW, Simmons AJ *et al.* (2006) The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961–3012.
Vavrek MJ (2011) fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, **14,** 1T:16p.
Vähä J-P, Erkinaro J, Niemelä E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*, **16**, 2638–2654.
Via S, Gomulkiewicz R, De Jong G et al. (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, **10**, 212–217.
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution*, **65**, 1897–1911.
Øresland V, André C (2008) Larval group differentiation in Atlantic cod (*Gadus morhua*) inside and outside the Gullmar Fjord. *Fisheries Research*, **90**, 9–16.

Data accessibility

All SNPs are referred to by their database of Single Nucleotide Polymorphisms (dbSNP) accession numbers, available from: http://www.ncbi.nlm.nih.gov/SNP. Individual genotype data is available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.3f1c8. The nomenclature of linkage groups in this paper follows Hubert *et al.* (2010).

Author contribution

The study was conceived and designed by CA, JMIB, PRB, JHH, KSJ, SJ, KJ, PEJ, HK, PM, BS, NCS, HS. Assessment of genotypes and data quality was done by JMIB, PRB, SB, JHH. Genomic analyses were performed by JMIB. Oceanographic modeling was carried out by HC, PRJ, PM. Samples were provided by CA, JHH, SJ, HK, HS. The manuscript was written by JMIB with contributions from PRJ. All authors read and revised the manuscript.

Figure legends

Fig. 1 Sampling sites of Atlantic cod (colored points). Dotted lines indicate boundaries between seas (North Sea, Skagerrak, Kattegat, and western Baltic Sea) and arrows delineate main water currents. ENG, English Channel; NOR, North Sea; TVE, Tvedestrand; SOP, Soppekilen; HEL, Hellefjord; GRE, Grenland; OSL, Oslofjord; IDD, Iddefjord; SKA, Skagerrak; GUL, Gullmarsfjord; HAV, Havstensfjord; KAT, Kattegat; ORE, Öresund; BEL, Belt Sea.

Fig. 2 Population differentiation, admixture and ancestry analyses. (a,b) Hierarchical STRUCTURE analysis. Individual population assignment is shown by colored and gray horizontal bars (*q*-values), black bars separate sampling locations (for sampling site abbreviation see legend Fig.1). Individuals are ordered within sampling sites according to their assignment proportions. (a) K=2, (b) K=3, see Supporting information for analyses in which the most differentiated groups were hierarchically excluded. (c, d) STRUCTURE ancestry analysis. (c) Inference of mechanical mixture versus genetic admixture using the source populations North Sea (0.0) and Kattegat (1.0). (d) Kernel density estimates of (c) displaying multimodal (mechanical mixture) versus unimodal (admixed) patterns. Hartigans' dip-test indicated three significantly multimodal sampling sites (marked with asterisks): TVE, SOP, and IDD. (e) Admixture quantified as hybrid index (H) for each individual using BGC cline analysis with the North Sea (0.0) and the Kattegat (1.0) samples as source populations. Points represent the means of posterior distributions, indicating North Sea (red, $H \leq 0.25$), Kattegat (blue, $H \geq 0.75$), and admixed individuals (black). Gray bars indicate 95% credibility intervals.

Fig. 3 F_{ST} estimates of genetic differentiation. (a) Heat map of pairwise F_{ST} comparisons. NS, non-significant. (b) Classic multidimensional scaling (MDS) plot of pairwise F_{ST} comparisons. For sampling site abbreviations see legend Fig. 1.

Fig. 4 Biophysical model of larval connectivity. (a) Modeled connectivity from four spawning areas to the eastern and western Skagerrak coasts, and to the Kattegat, expressed as the proportional larval supply. Larval supply was calculated as the probability of larval dispersal from the spawning areas scaled with the respective

spawning stock biomass (SSB). (b) Modeled connectivity from the same four spawning areas to western Skagerrak fjords expressed as larval supply by scaling the dispersal probability with the respective SSB, and normalized to the target area. For TVE, HEL and GRE, only the fjord mouths were included in the model. Error bars show the standard deviation of simulations for six years (1995-2002). For abbreviations of fjords see legend Fig. 1.

Fig. 5 Distribution of chromosomal rearrangements. Per sampling site, individuals were scored for three chromosomal rearrangements on linkage groups (LG) 2, 7, and 12. The proportion of individuals being homozygous for the presumed collinear allele is shown in white, and proportions of individuals heterozygous or homozygous for the rearranged allele are shown in light and dark gray, respectively. Sampling sites representing a significant overrepresentation of the rearranged allele are marked with an asterisk.

Supporting information

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

Supporting Notes

Note S1 Genotype quality control and filtering.

Note S2 Principal component analysis.

Note S3 Genetic differentiation within North Sea-like and western Baltic-like groups.

Note S4 Genetic differentiation of the western Skagerrak samples Tvedestrand and Soppekilen.

Supporting Figures

Fig. S1 Spawning areas used as sources of cod eggs in the biophysical connectivity model.

Fig. S2 Hierarchical structure analysis

Fig. S3 Principal component analysis (PCA).

Fig. S4 Isolation by distance (IBD).

Fig. S5 Principal component analysis (PCA) excluding SNPs in rearranged regions.

Supporting Tables

Table S1 Detailed information about Atlantic cod samples included in this study.

Table S2 Mean observed (H_o) and expected (H_E) heterozygosity.Table S3 Mean ln STRUCTURE likelihood values.

Table S4 Characteristics of neutral and selective loci.

Table S5 Diversifying SNPs between NOR and KAT.

Table S6 Chromosomal rearrangement breakpoints.Table S7 Admixture quantified as hybrid index (*H*).

Table S8 Individuals assigned to NOR, KAT, or being admixed, based on the hybrid index (*H*).

Table S9 Pairwise F_{ST} estimates of genetic differentiation.

Supporting References

Figures

Genome architecture enables local adaptation of Atlantic cod despite high connectivity

Barth *et al.* (j.m.i.barth@ibv.uio.no)

Figure 1 Sampling sites of Atlantic cod.

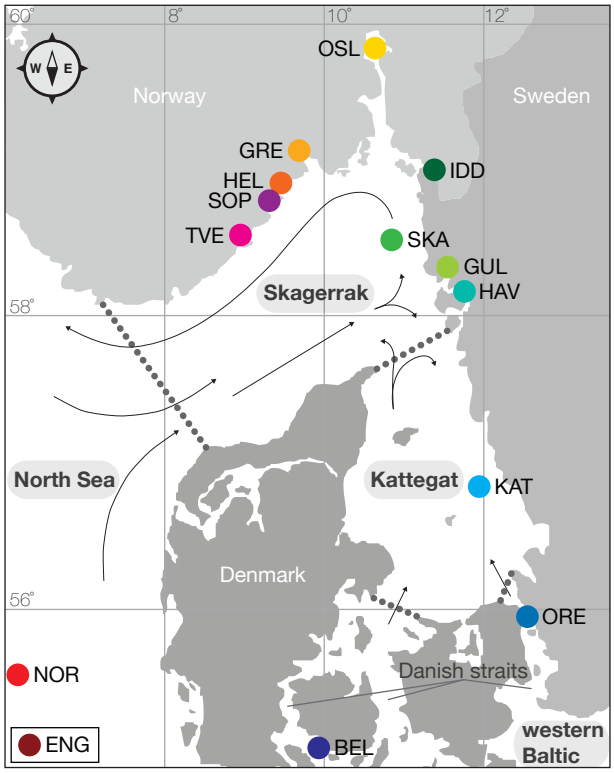

Figure 2 Population differentiation, admixture and ancestry analyses.

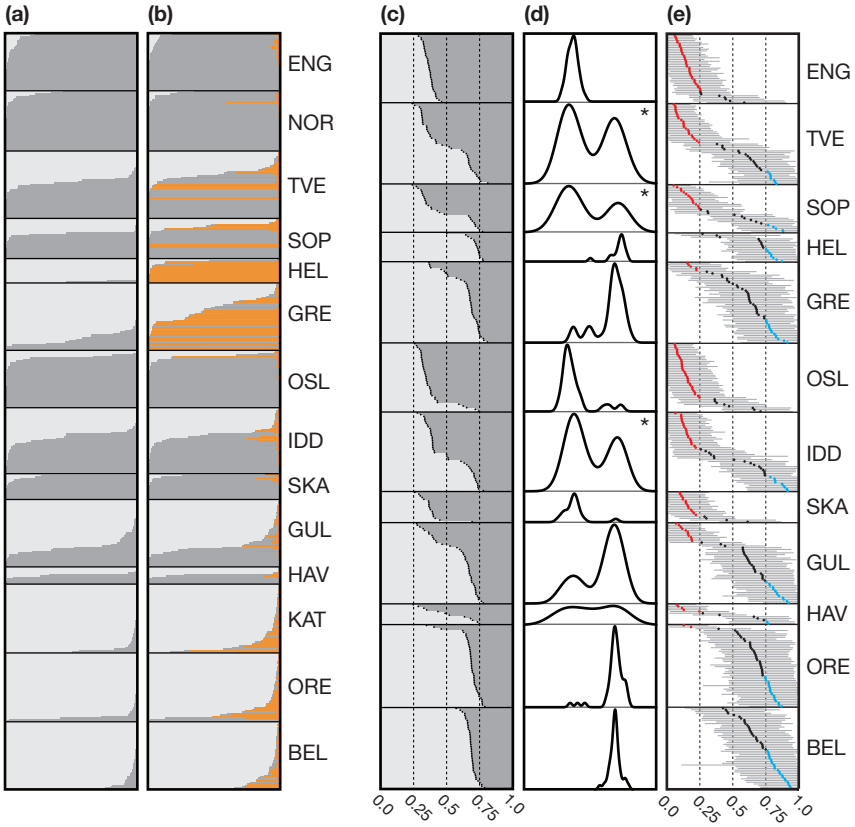

Figure 3 F_{ST} estimates of genetic differentiation.

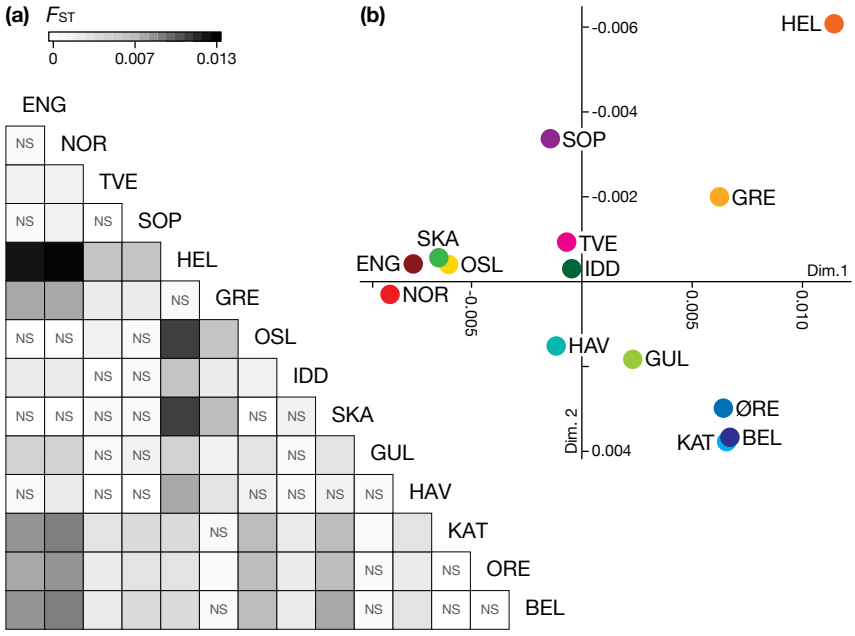

Figure 4 Biophysical model of larval connectivity.

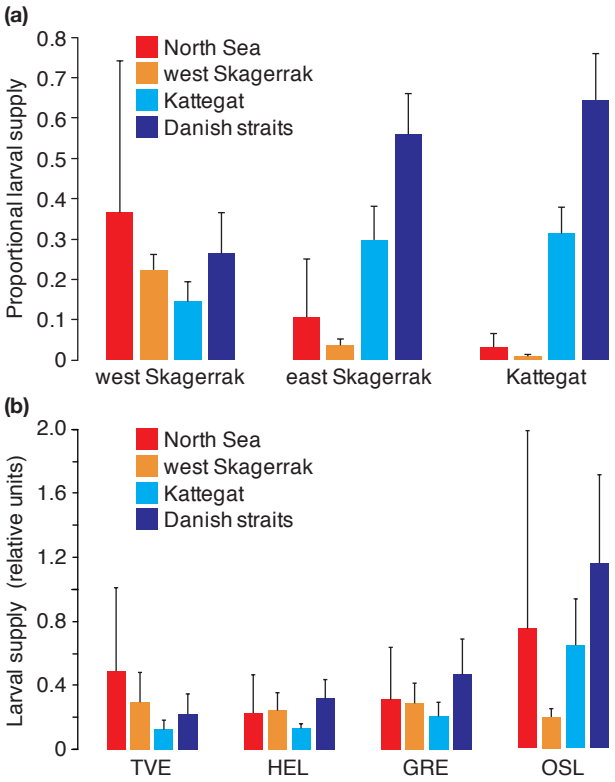

Figure 5 Distribution of chromosomal rearrangements.