DR. JULIA M I BARTH (Orcid ID : 0000-0001-5125-819X)

DR. IAN BRADBURY (Orcid ID : 0000-0002-8152-4943)

Article type : Special Issue

Disentangling structural genomic and behavioral barriers in a sea of connectivity

Running title: Isolating barriers in Atlantic cod

Julia M.I. Barth[1,2], David Villegas-Ríos[3,4], Carla Freitas[5,6,7], Even Moland[5.6], Bastiaan Star[1], Carl André[8], Halvor Knutsen[1,5,6], Ian Bradbury[9], Jan Dierking[10], Christoph Petereit[10], David Righton[11], Julian Metcalfe[11], Kjetill S. Jakobsen[1] & Esben M. Olsen[5,6], Sissel Jentoft[1]

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

2 Zoological Institute, University of Basel, Switzerland

3 Mediterranean Institute for Advanced Studies (IMEDEA), Department of Ecology and Marine Resources, Esporles, Spain

4 Marine Research Institute (IIM CSIC), Department of Ecology and Marine Resources, Vigo, Spain

5 Institute for Marine Research, Flødevigen, Norway

6 Centre for Coastal Research, University of Agder, Norway

7 Oceanic Observatory of Madeira, Funchal, Portugal

8 Department of Marine Sciences - Tjärnö, University of Gothenburg, Sweden

9 Science Branch, Fisheries and Oceans Canada, St. John's, Canada

10 GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

11 Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Lowestoft, UK

Corresponding author: Julia M.I. Barth (jmi.barth@gmail.com)

ABSTRACT

Genetic divergence among populations arises through natural selection or drift and is counteracted by connectivity and gene flow. In sympatric populations, isolating mechanisms are thus needed to limit the homogenizing effects of gene flow to allow for adaptation and speciation. Chromosomal inversions act as an important mechanism maintaining isolating barriers, yet their role in sympatric populations and divergence with gene flow is not entirely understood. Here, we revisit the question whether inversions play a role in the divergence of connected populations of the marine fish Atlantic cod, by exploring a unique dataset combining whole-genome sequencing data and behavioral data obtained with acoustic telemetry. Within a confined fjord environment, we find three genetically differentiated Atlantic cod types belonging to the oceanic North Sea population, the western Baltic population, and a local fjord-type cod. Continuous behavioral tracking over four years revealed temporally stable sympatry of these types within the fjord. Despite overall weak genetic differentiation consistent with high levels of gene flow, we detected significant frequency shifts of three previously identified inversions, indicative for an adaptive barrier to gene flow. In addition, behavioral data indicated that North Sea cod and individuals homozygous for the LG12 inversion had lower fitness in the fjord environment. However, North Sea and fjord-type cod also occupy different depths, possibly contributing to prezygotic reproductive isolation and representing a behavioral barrier to gene flow. Our results provide the first insights into a complex interplay of genomic and behavioral isolating barriers in Atlantic cod and establish a new model system towards an understanding of the role of genomic structural variants in adaptation and diversification.

Keywords: chromosomal rearrangements, gene flow, sympatric divergence, adaptation, behavioral traits, Atlantic cod

1 INTRODUCTION

How new species arise and adapt to their environments is a fundamental question in the field of evolutionary biology. Yet, our understanding of the genetic mechanisms behind speciation with gene flow is far from complete (Ravinet et al. 2017, Jorde et al. 2018a). Within the last decade, it has become accepted that population

divergence, adaptation, and speciation in the face of gene flow is no rare exception (Hey 2006; Nosil 2008), and advances in sequencing technology have begun to reveal the underlying genomic architecture of this complex process (Seehausen et al. 2014; Tigano & Friesen 2016; Wolf & Ellegren 2017; Wellenreuther & Bernatchez 2018). Several studies recently highlighted the role of genomic structural variants such as inversions in adaptation and diversification (Kirkpatrick & Barton 2006; Lowry & Willis 2010; Jones et al. 2012; Fishman et al. 2013; Lohse et al. 2015, among others). Such rearranged regions may constitute intrinsic postzygotic barriers to gene flow (through genetic incompatibilities), extrinsic postzygotic barriers (where hybrids suffer reduced fitness in the parental environment), or act through ecological adaptive barriers where sets of locally adaptive alleles are captured and protected against recombination, giving a selective advantage to the individuals carrying the rearranged regions and leading to their spread in the respective environment (Rieseberg 2001; Kirkpatrick & Barton 2006; Feder et al. 2012b). The presence of structural rearrangements can therefore promote early stages of ecological divergence, which may eventually lead to speciation (Feder et al. 2014a). Although theory predicts that in the face of gene flow, few large-effect alleles may similarly spread and contribute to divergence if selection is strong enough, adaptive alleles rather persist if architecturally linked (Yeaman &Otto 2011; Yeaman & Whitlock 2011). Indeed, comparisons of sister species of rodents and birds revealed that sympatric sister species are more likely to differ by chromosomal rearrangements than allopatric ones (Castiglia 2013; Hooper & Price, 2017).

In the marine fish Atlantic cod (*Gadus morhua*), four large (5 – 17 Mb) chromosomal rearrangements, each possessing high internal linkage disequilibrium (LD) have been detected (Berg et al. 2016; Sodeland et al. 2016; Kirubakaran et al. 2016), which show clinal distributions amongst the majority of populations throughout the entire species range in the North Atlantic Ocean, indicating a role in adaptation to local environments (Sodeland et al. 2016; Kirubakaran et al. 2016; Berg et al. 2016; Barth et al. 2017a; Star et al. 2017; Berg et al. 2017). It is likely that all of these rearrangements represent old chromosomal inversions (~2 million years based on divergence time estimates of closely related species (Matschiner, personal communication 2015; Kirubakaran et al. 2016), hereafter for simplicity termed inversions), which may have evolved in allopatric refugia during Pleistocene glaciation cycles (Kirubakaran et al. 2016). As a marine batch spawner without any brood care, releasing thousands to millions of pelagic eggs to be distributed by ocean currents, Atlantic cod reside in environments that a priori seem to provide little boundaries to restrict connectivity amongst populations (Hutchings et al. 1999; Munk et al. 1999; Cowen & Sponaugle 2009; but see also Espeland et al. 2015). The inverted regions have therefore been associated with ecological adaptation of Atlantic cod to temperature (Bradbury et al. 2010; Therkildsen et al. 2013), oxygen and salinity (Berg et al. 2015), coastal environments (Sodeland et al. 2016; Barth et al. 2017a), and with migration behavior (Hemmer-Hansen et al. 2013; Karlsen et al. 2013; Berg et al. 2016; Kirubakaran et al. 2016; Sinclair-Waters et al. 2018). While in some populations one or several of the inverted arrangements have reached near fixation, other represent strong frequency shifts relative to the ancestral arrangements, yet

other frequently possess both arrangements in similar numbers (Berg et al. 2016; Sodeland et al. 2016; Kirubakaran et al. 2016; Barth et al. 2017a; Star et al. 2017; Berg et al. 2017).

In addition to structural rearrangements of the genome acting as postzygotic barriers to restrict gene flow between sympatric populations, prezygotic mechanisms including fitness advantages in the local environment, temporal, spatial or ecological shifts during the breeding season (causing allopatric reproduction), as well as behavioral differences (e.g., assortative mating), may also play a role (Coyne & Orr 2004, Jones 2006). For Atlantic cod, behavioral features that may contribute to prezygotic isolation such as spawning site fidelity, natal homing (Espeland et al. 2007; Skjæraasen et al. 2011; Bonanomi et al. 2016; Villegas-Ríos et al. 2017), and mate choice, as well as territorial behavior (Hutchings et al. 1999) have been described, questioning the hitherto outlined prominent role of ecological adaptation through the chromosomal inversions in the ecotype divergence of this species.

To investigate the relative roles of different barrier mechanisms for the potential of adaptation and diversification in sympatric populations characterized by homogenizing gene flow, we here focus on a topographically restricted coastal fjord ecosystem in which the occurrence of three genetically differentiated types of Atlantic cod, a resident fjord-type, an oceanic North Sea-type, and a western Baltic-type have been reported (Barth et al. 2017a; Knutsen et al. 2018). Sympatric populations that are connected and exposed to gene flow, but exhibit distinct phenotypic or behavioral differences, represent well-suited systems to investigate genomic signatures of divergence since in these systems differentiated sites are expected to be rare and restricted to regions involved in adaptation (Coyne & Orr 2004). High connectivity through geographic overlap between the different cod types has previously been demonstrated (André et al. 2016; Barth et al. 2017a), and -despite low overall genetic differentiation- significant shifts in frequency of the inverted arrangements have been observed (Sodeland et al. 2016; Barth et al. 2017a). However, whether the adaptive properties of the inversions constitute the only barrier to gene flow, or whether other barriers may also play a role, is not known. For example, yet undetected additional genomic structural variation in the form of large-effect alleles (Yeaman &Otto 2011), or physical barriers such as seascapes and salinity, oxygen, or temperature gradients (Howe et al. 2010; Ciannelli et al. 2010; Rogers et al. 2014), as well as behavioral differences could act as additional mechanisms restricting connectivity and gene flow.

Along the convoluted Norwegian coast, Atlantic cod show high site fidelity and restricted movement in studies using acoustic telemetry and longer-term mark-recapture data (Olsen and Moland 2011; Rogers et al. 2014; Aalvik et al. 2015; Freitas et al. 2016; Villegas-Ríos et al. 2017), limiting the potential for adult dispersal. However, long-distance spawning migrations have also been recorded (Skjæraasen et al. 2011; Neat et al. 2014). Resident local behavioral units with spawning aggregations along the coast and inside the fjords, as well as physical retention of pelagic eggs through a fjord-inward flow (Ciannelli et al. 2010) may thus explain the

occurrence of different cod types and the genetic structuring documented in previous studies (Jorde et al. 2007; Knutsen et al. 2011). Yet, the degree to which settled juveniles stay and recruit to the local adult population is not known. Nevertheless, the relatively stable coexistence of early-stage eggs, larvae, and adults in spawning condition of at least two distinct Atlantic cod types indicate that several populations use the convoluted coastline for spawning, as nursery, and for long-term residence (Rogers et al. 2014; Barth et al. 2017a; Knutsen et al. 2018; Jorde et al. 2018b).

By using whole-genome sequencing data of more than 200 Atlantic cod specimens from one fjord and six adjacent locations, we here gain comprehensive insights into the genomic variation and population relationships of the fjord cod community. In order to clarify how these highly connected cod types maintain differentiation, we characterize potential barriers to gene flow in the genomic landscape. Additionally, we test whether behavioral differences, potentially acting as prezygotic barriers to gene flow, are present among the cod types and whether these are correlated with the inversion states.

2 MATERIALS AND METHODS

2.1 Sampling and bioinformatics

A total of 204 individuals of Atlantic cod (*Gadus morhua*) from seven sites were sampled (Figure 1a, Table S1): Averøya, North Atlantic (AVE, N = 20), North Sea (LOW, N = 24; NOR, N = 24), Skagerrak Tvedestrand fjord (TVE, N = 70), and the western (ORE, N = 21; KIE, N = 22) and eastern Baltic Sea (BOR, N = 23). For all TVE individuals, both genomic and behavioral data were collected. DNA extraction, library preparation (Illumina Truseq DNA PCR-free kit, combinatorial dual index adapters), and sequencing (Illumina HiSeq 2500, V4 chemistry, 2 x 125 bp paired-end) was performed at the Norwegian Sequencing Centre according to the Centre's protocols (for a brief description see Barth et al. 2017b). Index-hopping (Kircher et al. 2012) should not constitute a problem given the sequencing strategy used. Mapping and genotype calling were performed following the method described by Star et al. (2017), employing the PALEOMIX v1.5 (Schubert et al. 2014) pipeline for read processing and mapping (BWA MEM v0.7, Li & Durbin 2009) against the GadMor2 genome (Star et al. 2011; Tørresen et al. 2017), and the GATK HAPLOTYPECALLER v3.4.46 (McKenna et al. 2010) for variant calling. Average read depth per sample was 9.84 +/-1.16 (Table S1). Filtering was performed using BCFTOOLS v1.3 (Li 2011), VCFTOOLS v0.1.14 (Danecek et al. 2011), and PLINK v1.9 (Purcell et al. 2007) according to the following thresholds: FS < 60, MQRankSum > -12.5, ReadPosRankSum > -8, QD > 2, MQ > 40, SnpGap = 10, number of indels = 0, number of alleles = 2, meanDP < 30, GQ > 20, DP > 3, missing data < 20%, minor allele count > 2, minor allele frequency (MAF) > 0.03, heterozygous excess (p < 0.001), and intrachromosomal LD (window size

10 kb; pairwise r^2 threshold 0.8). Additionally, repetitive regions were excluded (Tørresen et al. 2017), leading to a dataset of 1,258,658 SNPs. For population genomic and phylogenetic analyses, linkage groups (LG) 01, 02, 07, and 12 were excluded due to the presence of large inversions (Berg et al. 2016; Sodeland et al. 2016; Kirubakaran et al. 2016). Phylogenetic analyses were performed without MAF filtering. Because of biased call probability of heterozygotes at low sequencing depth (Nielsen et al. 2011), which does not affect individually based analyses (Figure S1) but can influence the estimates of population genetic parameters due to an overall higher or lower coverage per population, analyses of the genomic landscape were performed using a dataset where genotypes with less than seven reads (DP < 7) for a sample were set missing.

2.2 Population differentiation and phylogenetic analyses

Genome-wide population structure was inferred by performing hierarchical Principal Component Analyses (PCA) using smartPCA in EIGENSOFT v.6.0.1 (Patterson et al. 2006) including the function "lsqproject", and through model-based clustering using ADMIXTURE v1.3 (Alexander et al. 2009). Mean standard errors for ADMIXTURE point estimates were calculated using 100 block-bootstrap resamplings. Three replicates, each testing for 1 to 6 clusters (*k*) and 10-fold cross-validation was performed. Maximum likelihood phylogenetic inference was performed using RAxML v8.2.4 (Stamatakis 2014) under the GTRCAT model with ascertainment bias correction and 100 bootstrap replicates on a dataset including five randomly selected individuals per population. Only variable SNPs (781,038) were included and heterozygous SNPs were translated to ambiguity codes. The sister species of the Atlantic cod, the Alaska pollock (*Gadus chalcogrammus*; Malmstrøm et al. 2016) was used to root the tree. Homologous sites of *G. chalcogrammus* were identified through mapping against GadMor2 as described above and retrieving the consensus sequence using "mpileup" in SAMTOOLS v1.3 (Li 2011) setting MQ > 40. Five Canadian Atlantic cod individuals were included for subsequent rooting of the SNAPP phylogeny (see below). The species tree and population sizes were estimated using the multispecies coalescence approach in the add-on SNAPP v1.3.0 (Bryant et al. 2012) for BEAST2 v2.4.7 (Bouckaert et al. 2014), using the same individuals as for the maximum likelihood analysis (except *G. chalcogrammus*), but allowing no missing data. As SNAPP runtimes are very long on larger SNP sets, the dataset was reduced by applying a minimum distance of 70,000 bp between SNPs, resulting in 3,307 SNPs. The script "snapp_prep.rb" (Stange et al. 2018) was used to prepare the input XML retaining the original settings except that theta values were not linked among branches. Three replicate analyses with random starting trees and a length of 1 million Markov-chain Monte Carlo (MCMC) steps were carried out. Convergence was assessed by effective sample size (ESS) values > 200 (Rambaut et al. 2018) after discarding the first 5% of each chain as burnin and merging the posterior distributions. The summary

tree cloudogram (see Figure 1c) represents the entire tree set (28,503 trees), while for the representation of topology uncertainties only every 1,000th tree was sampled.

2.3 Genomic landscape

To identify structural genomic barriers to gene flow, we used an equal number of 20 individuals (randomly downsampled) from each of the identified co-occurring North Sea (TVEn), western Baltic (KIE), and fjord-type (TVEf) populations (where TVEn and TVEf are the TVE specimens showing the North Sea and the fjord genotype, respectively, see results section 3.1; KIE was chosen as representative population for the western Baltic since only three western Baltic-type cod were collected in the fjord, see Figure 1b). Non-overlapping 50 and 100 kb windowed chromosome scans to calculate the following measurements were performed using custom scripts: the pairwise fixation index (F_{ST}; as in Weir & Cockerham 1984), the pairwise between population sequence divergence (d_{xy}; as in Ruegg et al. 2014), the proportion of fixed differences (d_f; see Ruegg et al. 2014), and the nucleotide diversity (π). To detect divergent regions < 50 kb, we additionally ran BAYESCAN v2.1 (Foll and Gaggiotti 2008) for each LG, evaluating the population pairs TVEf-TVEn, TVEf-KIE, and TVEn-KIE using default settings, but adjusting the prior odds for the neutral model (PO) to 1,000 due to the large number of SNPs. Convergence was assessed by ESS > 200 using the CODA package v0.18 (Plummer 2006) in R v3.3.3 (R core team 2018). Genes and their associated Gene Ontology (GO) terms in diverged regions (for the inverted regions on LG02, 07, and 12 see Barney et al. 2017) were identified based on the GadMor2 genome annotation (Tørresen et al. 2017), considering all SNPs detected as outliers in the BAYESCAN analysis ($\log_{10}$(PO) > 1, False Discovery Rate (FDR) 0.05) as well as genes within 5 kb of such outlier SNPs. Enrichment tests for GO terms were performed for genes detected in comparisons TVEf-TVEn and TVEf-KIE, but not TVEn-KIE using Fisher's exact test with the algorithm "weight01" in the TOPGO package v2.26 (Alexa & Rahnenfuhrer 2016) for R. All genes located in the highly differentiated region on LG16 (see results) were included in the GO-term test. The squared correlation between SNPs as a measure of LD (r^2; using PLINK) was calculated using all 204 individuals. SNPs in regions of putative inversions (LG01: positions 9,114,741-26,192,386; LG02: positions 18,609,260-23,660,985; LG07: positions 13,622,710-23,019,113; LG12 positions: 426,531-13,445,150; see Barth et al. 2017a; Berg et al. 2016) were extracted, and the inversion state of all specimens was analyzed for each of these regions using smartPCA. Bootstrapping (Efron 1979; sample size 1,000,000) of individual genotypes was used to assess a possible overrepresentation of the inverted arrangement (or the ancestral arrangement, referring to the overall frequency among all samples) of all inverted regions, under the null hypothesis that the frequency of the tested arrangement within a population corresponds to its overall frequency across all populations. Bonferroni correction was applied in R to correct for multiple comparisons (Rice 1988). Maximum

likelihood phylogenetic inferences were performed using RAxML as described above, including only sequences of individuals homozygous for either the respective ancestral or inverted arrangement.

2.4 Behavioral analyses

A total of 70 Atlantic cod (mean body length 46 cm, range 30-75 cm) were captured and tagged in the Tvedestrand fjord in May 2011-2013 as described elsewhere (Olsen et al. 2012). Briefly, fish were captured using fyke-net fishing for 1-3 days. Fish selected for tagging were anesthetized, equipped with acoustic transmitter (Vemco V9P-2L, 508-660 days battery life), programmed to send the current depth and a fish identification code (fish ID) every 110-250 s. Transmitters recorded on average 16,402 (+5,489) locations per fish, in 2011 and 2012 a maximal depth of 100 m (0.44 m resolution and 5 m accuracy), and in 2013 a maximum depth of 50 m (0.22 m resolution and 2.5 m accuracy). A fin-clip was taken and stored in ethanol for DNA analysis. Fish were released at their initial capture location. Previous experiments showed no mortality in the tagging process (Olsen & Moland 2011; Olsen et al. 2012). An array of 33 underwater ultrasonic receivers (Vemco VR2W, 69kHz) were used to record and log transmitter signals (Villegas-Ríos et al. 2017).

For each tagged fish, centers of activity (COA) were calculated in 30 min time bins following Simpfendorfer et al. (2002). Depth data time series and COA latitude/longitude plots were used to identify and remove all detections recorded after cessation of movement, i.e., death of the fish (Harrison et al. 2015). Code collisions and false detections were eliminated using a minimum of two detections per 24 h period. Diel vertical migration was estimated as the difference between average depth during the day and average depth at night, averaged over months (Freitas et al. 2015). Monthly home range was estimated as the kernel utilization distribution with a probability level of 95% using all COA from that month (fish were required to be present in the array during at least 20 days in any particular month). Monthly mean depth during daytime was calculated by averaging daily daytime mean depths.

Individual fate was assigned based on detection patterns. Fish were classified as either: 1) alive within the study area (i.e., multiple detections indicated horizontal and vertical movements), 2) dispersed from the study area (i.e., directional movement towards the outermost receivers followed by an absence of detections for the rest of the study), or 3) dead within the study area (i.e., when the fish either stopped transmitting while inside the study area or started transmitting continuously at the same depth).

Linear mixed effects models were used to analyze variation in cod behavioral traits (*BT*). The study area (a no-take marine reserve) holds a key spawning locality for local cod (Ciannelli et al. 2010) and we specifically explored whether Atlantic cod displayed contrasting behavior during the spawning season (January-April) compared to the feeding season (Espeland et al. 2007; Roney et al. 2018). The full linear mixed effects model, prior to model selection, included fixed effects of season (*S*), inversion state for the LG02, LG07, and LG12 inverted regions (*LG02*, *LG07*, *LG12*), body size (*L*), and genetically determined sex (*GS*, Star et al. 2016). To explore our working hypothesis, we also included an interaction effect between season and inversion state:

$$BT = c_0 + c_1,S + c_2,LG02 + c_3,LG07 + c_4,LG12 + c_5L + c_6,GS + c_7,S\times LG02 + c_8,S\times LG07 + c_9,S\times LG12 + c_{10},S\times L + c_{11},S\times GS$$

where c_0 is the intercept. Season was modeled as a factor with two levels (spawning and feeding), cod inversion states as factors with three levels (ancestral, heterozygous and inverted), body size as a continuous variable, and genetic sex as a factor with two levels (female and male).

We included the following interaction effects: (i) between season and each of the inversions to explore whether any behavioral changes between feeding and spawning season depended on genotype, (ii) between season and body size because smaller fish might not be sexually mature and therefore not participate in spawning, and (iii) between season and sex since spawning behavior can differ between females and males, where males are known to defend territories associated with seafloor features during the spawning season (Meager et al. 2009; 2012). Three behavioral traits were analyzed in separate models: (i) monthly home range size, (ii) monthly mean depth use, and (iii) monthly mean diel vertical migration. Home range size was log-transformed to stabilize the variance. Fish ID was included as a random effect to account for repeated (monthly) observations on fish behavior (among-individual variance).

Next, we used the Lande-Arnold linear regression approach (Lande & Arnold 1983) to model fitness as relative longevity (days survived / mean days survived, *S*). Fish that dispersed permanently from the study area were excluded from this analysis since their fate cannot be determined. The full model included effects of inversion states (*LG02, LG07, and LG12*), body size (*L*), and genetic sex (*GS*):

$$S = c_0 + c_1,LG02 + c_2,LG07 + c_3,LG12 + c_4L + c_5,GS$$

In a final set of analyses, we replaced the inversion factor with a genotype factor where fish were defined as either North Sea-type (TVEn) or fjord-type (TVEf) based on genetic results (see Results section 3.1). A total of 22 individuals could not be reliably defined as either of the two types (i.e., were classified as “intermediates”) and were excluded from these analyses.

We used the AIC criterion for model selection (Burnham and Anderson 1998). In a two-step process, we first decided on the most parsimonious model structure for effects of body size and sex on behavior, while maintaining all inversion state effects of interest. Next, we selected the most parsimonious model structure for inversion state effects on behavior.

3 RESULTS

3.1 Genomic variation and population relationships

Genomic variation and relationships among all sampled specimens (Figure 1a) were explored using multivariate and clustering analyses, as well as phylogenetic methods. Applying PCA revealed the largest differentiation to be found between the eastern Baltic Sea sample (BOR) and all other samples, while the second largest variation was found between a combined North Sea cluster (NOR, LOW) and some of the fjord specimens (TVE) (Figure 1b). The western Baltic samples (ORE, KIE) formed a separate, intermediately placed cluster, while the Norwegian coastal individuals (AVE) were recovered between the North Sea, the western Baltic, and the fjord (TVE) cluster. Individuals collected in the fjord (TVE) showed a split distribution, with about half of the specimens (N = 21) grouping with the North Sea cluster, while the remaining ones (N = 27) formed a well-defined private cluster (Figure 1b, fjord). Three TVE specimens were also placed within the western Baltic cluster, one within the Norwegian coastal cluster, and some individuals (N = 18) were distributed between the main clusters on PC2 and could not be clearly assigned to a particular cluster (referred to as 'intermediate' in 3.3, see also Table S2). Hierarchical exclusion of the most differentiated clusters (BOR, NOR, LOW), as well as separate analyses of the TVE specimens and the North Sea cluster, respectively, confirmed this overall pattern of population differentiation (Figure S1b-f). Using the software ADMIXTURE for maximum likelihood model-based inference of genomic variation, the best-supported models (lowest cross-validation error) were obtained assuming one to three clusters. As in the PCA results, the ADMIXTURE ancestry proportions yielded three main clusters represented by the eastern Baltic Sea sample, the North Sea samples, and half of the TVE sample, while the ancestry proportions of the western Baltic and Norwegian coastal samples appeared admixed (Figure S2a-c).

To obtain estimates of between-population relationships, we constructed a phylogeny using the multispecies coalescence approach. Since the population genomic analyses above suggested the existence of distinct genotype units within the TVE sample, we subdivided the TVE specimens into a "fjord-type" termed TVEf, and a North Sea-type termed TVEn according to their position in the PCA (Figure 1c inset, Table S2). As a closely related outgroup, we chose five Atlantic cod specimens from the western North Atlantic (Twillingate, Canada; CAN), which were identified as sister population to our eastern North Atlantic samples based on a maximum

likelihood phylogeny rooted using the sister species of Atlantic cod, the Alaska pollock (Figure S3). The summary tree (Figure 1c) shows two well-supported monophyletic clades: one for the combined group of samples from the eastern Baltic Sea (BOR) and the western Baltic Sea (ORE, KIE), which also included the fjord-type TVE specimens (TVEf), and another for the North Sea populations (NOR, LOW), which included the North Sea-type TVE individuals (TVEn). Relative divergence times indicate that the eastern Baltic Sea population is evolutionary older than the TVEf and the western Baltic populations. Relative effective population sizes are larger in the North Sea (theta = 1.14) and western Baltic (theta = 1.43, KIE; and 1.34, ORE) than in the eastern Baltic Sea (theta = 0.04) and within the TVEf cod (theta = 0.52; Figure 1c). Alternative topologies indicate topological uncertainty due to incomplete lineage sorting or existing gene flow between the North Sea populations and the Baltic populations (Figure 1c).

3.2 Barriers to gene flow in the genomic landscape

Our population genomic analyses described the spatial and temporal co-occurrence of three genetically differentiated types of Atlantic cod within the Tvedestrand fjord ecosystem (TVE North-Sea type (TVEn), TVE fjord-type (TVEf), and TVE western Baltic-type cod). To identify genomic regions that may act as structural barriers to restrict gene flow between these groups, we performed genome scans testing for LD and differentiation. We identified three previously described large regions of strong LD and high inter-population divergence on LG 02, 07, and 12 (Figure 2a, see Figure S4 for all chromosomes and measurements). Besides these three regions, no other large regions of elevated LD accompanied by genetic divergence indicative for chromosomal rearrangements, were detected. However, several genome-wide distributed differentiation peaks, including a 117 kb region on LG16 (15,467,682-15,584,985) showing no signs of strong LD, but consistently high F_{ST} (> 0.6, Figure S5), indicate the existence of further smaller regions of fjord-type divergence. Fixed SNPs, commonly observed between diverging species, were not detected among the TVEn, TVEf, and KIE individuals. As regions of tightly linked loci have been proposed as promoters of divergence in sympatric populations (Feder et al. 2012a), SNPs within the three high LD regions were extracted and the inversion state for all regions, specimens, and populations, as well as their evolutionary relationships were analyzed. Applying PCA segregated specimens into three clusters on the first principal component axis (Figure 2b), visualizing the bi-allelic nature of the inversions in which homozygous individuals occurred in two groups, with heterozygous individuals located as intermediate, irrespective of sampling site. On the second axis, the ancestral arrangement showed relatively little divergence between populations (except for the eastern Baltic specimens on LG07), whereas higher divergence was observed between individuals carrying the inverted arrangement. Similar to the whole-genome analysis, the arrangements of TVE individuals clustered with either North Sea or western Baltic individuals, as well as within a private cluster for the inversion on LG02. A phylogenetic approach revealed that these TVE individuals form a monophyletic group, which was well separated from the eastern Baltic individuals (Figure 2c). All of these TVE

individuals belong to the fjord-type (TVEf). Similarly, TVEf individuals were mostly recovered within shared or neighboring clades for the ancestral and inverted arrangements of the inverted regions on LG07 and LG12. In contrast, TVEf individuals were recovered well dispersed among all other individuals sharing the ancestral LG02 arrangement (Figure 2c).

Frequency comparisons of the ancestral and inverted arrangements among populations revealed a significant overrepresentation of the ancestral LG02 arrangement in LOW ($p < 0.001$) as well as in TVEn ($p < 0.01$), and an overrepresentation of the inverted LG02 arrangement in BOR ($p < 0.001$) as well as in TVEf ($p < 0.01$) (Figure 2d, Table S3). On LG07, the ancestral arrangement was also found to be overrepresented in LOW ($p < 0.01$), whereas the inverted arrangement on LG12 was overrepresented in NOR ($p < 0.01$) and AVE ($p < 0.001$). For TVEf, we also observed an overrepresentation of the inverted LG07 ($p = 0.0043$) and the ancestral LG12 arrangement ($p = 0.0041$); however, both comparisons were not significant after correction for multiple comparisons. No significant yearly difference was detected within the TVE sample (Figure S6), and none of the inverted regions deviated from Hardy-Weinberg equilibrium (HWE) ($p > 0.05$; in LOW the inverted LG02 arrangement is not present). All of the inverted regions segregated independently, and all but three of the 27 possible combinations of homozygous ancestral, heterozygous, and homozygous inverted arrangements on the three LGs were present. The non-sampled combinations are: (i) homozygous for all inverted arrangements, (ii) homozygous for the inverted arrangement on LG02 and LG07 but heterozygous on LG12, and (iii) heterozygous on LG02 but homozygous for the inverted arrangement on LG07 and 12. However, due to few inverted arrangements, the chance of sampling all combinations within 204 individuals is small (< 0.00001, based on the observed frequencies at each inverted region).

Several genome-wide distributed differentiation peaks not showing high LD were identified (Figure S4), which may indicate genes under positive selection among the cod types. Bayesian analyses to detect candidate loci under positive selection identified significant SNPs ($q < 0.05$) between TVEf-TVEn and TVEf-KIE, excluding SNPs also detected in TVEn-KIE, in 20 genes on four LGs (Table S4). The GadMor2 genome assembly includes 14,060 predicted genes associated with GO terms (Tørresen 2017), of which 13,977 had called SNPs within a region of 5,000 bp up- or downstream from the genes' coding sequence. GO enrichment analyses of the detected genes identified 12 significantly enriched GO terms ($p < 0.05$); however, none of these remained significant after FDR correction (Table S5).

3.3 Behavioral barriers to gene flow

To investigate whether possessing the inverted arrangements, or genetically belonging to the fjord-type (TVEf) was correlated with behavioral differences constituting an additional prezygotic barrier to gene flow, we employed data from acoustically tagged specimens to calculate home range, mean depth, diel vertical migration, and survival.

3.3.1 Home range

A mixed effects model supported two-way interaction effects between the LG07 inversion state and season, on cod home range size (Table 1, Figure 3a, see Figure S7 for individual raw values). A simplified model excluding the LG07 inversion factor, as well as a more complex model including an effect of the LG02 inversion state had similar support (Table 1). Specifically, all inversion states and body sizes tended to increase their home range during the spawning season compared to the feeding season, while this effect was stronger for fish having the inverted LG07 arrangement and for larger fish (Table 2, Figure 3a). A total of 25% of the variation in home range was associated with fish ID (among-individual variance), while the fixed effects explained 18% of variance. Alternative models, where the inversion state was replaced with a factor describing genotype (TVEf vs. TVEn), revealed no significant effects of genotype, either as part of a two-way interaction with season or as a simple additive effect on home range ($p > 0.50$).

3.3.2) Mean depth during daytime

A mixed effects model supported two-way interaction effects of between body size and season, as well as sex and season on cod mean depth use during daytime (Table 1, Figure 3b). Alternative models including inversion state had little support (Table 1). Smaller cod tended to occupy deeper waters during the spawning season compared to larger cod (Table 2, Figure 3b). On average, males occupied deeper waters compared to females. During the spawning season, females shifted to somewhat shallower depths while this effect was the opposite for males (Table 2, Figure 3b). A total of 23% of the variation in mean depth use was associated with fish ID (among-individual variance), while the fixed effects explained 9% of variance. Alternative models, where the inversion state was replaced with a factor describing genotype (TVEf vs. TVEn) revealed no significant interaction between genotype and season ($p = 0.85$), while there was statistical support for an additive effect of genotype where TVEf individuals tended to occupy deeper waters compared to the TVEn individuals ($\beta_{TVEf} = 3.67$, standard error = 1.40, $p = 0.012$, fixed effects explained 12% of the variance), accounting for effects of body size and sex (Figure 3b).

3.3.3.) Diel vertical migration

As for mean depth, a mixed effects model supported two-way interaction effects between body size and season, as well as sex and season, on cod diel vertical migrations (Table 1, Figure 3c). Alternative models including inversion states had little support (Table 1). On average, smaller fish had more extensive diel vertical migrations compared to larger fish and maintained this movement pattern during both feeding and spawning seasons (Table 2, Figure 3c). In contrast, larger fish only performed noticeable diel vertical migrations during the feeding season. On average, females displayed more extensive diel vertical migrations compared to males, especially during the spawning season where males showed signs of a reversed migration pattern (Table 2, Figure 3c). A total of 22% of the variation in diel vertical migration was associated with fish ID (among-individual variance), while 18% was explained by fixed effects. Alternative models where the inversion state was replaced with a factor describing genotype (TVEf vs. TVEn) revealed no significant effects of genotype, either as part of a two-way interaction with season or as a simple additive effect on diel vertical migration ($p > 0.48$).

3.3.4) Survival

A total of six fish permanently left the fjord for the outer coast during the four-year study period ($N_{North\ Sea} = 3$, $N_{fjord} = 1$, $N_{indermediate} = 2$). Of those fish that did not leave ($N_{dead} = 32$, $N_{alive} = 32$), a generalized linear model supported effects of the LG12 inversion state and body size on cod survival, although the support for an effect of body size was marginal (Table 3, Figure 3d). A total of 17% of the variation in survival was explained by the selected model. Alternative models including an effect of the LG02 and LG07 inversion states as well as sex received little support (Table 3). Predicted cod survival was lower for cod having the inverted LG12 arrangement and was also marginally lower for smaller fish compared to larger fish (Table 4). Alternative models, where the inversion state was replaced with a factor describing genotype (TVEf vs. TVEn) predicted survival to be lower for the TVEn compared to the TVEf individuals ($\beta_{TVEn} = -0.27$, standard error = 0.11, $p = 0.013$, Figure 3d). A total of 15% of the variation in survival was explained by the selected genotype model.

4 DISCUSSION

What are the processes and mechanisms behind hampered gene flow in sympatric populations? Do inversions play a role in sympatric adaptation and diversification, and which other barriers may be important to reduce gene flow? These questions are central to the understanding of the mechanisms leading to speciation with gene flow.

Using a unique dataset combining genomic with behavioral data of highly connected wild Atlantic cod populations within a local fjord system, we here show that significant frequency shifts of chromosomal inversions among cod types exist, one of which was also correlated with survival within the fjord environment, suggesting the existence of adaptive barriers to gene flow (i.e., a fitness effect associated with the environment). However, many weakly differentiated loci across the genome possibly under positive selection, as well as behavioral differences of depth usage between North Sea and fjord types suggest that further prezygotic barriers to gene flow may also exist.

Temporally stable sympatric occurrence of three genetically differentiated Atlantic cod types

Based on whole-genome sequencing data (excluding the inversion-carrying LGs), we demonstrated that the Atlantic cod community in the southern Norwegian fjord of Tvedestrand consists of three co-occurring genetically differentiated cod types, of which two types are associated with the adjacent Atlantic cod populations in the North Sea and the western Baltic Sea, while the third forms a distinct unit that was only detected within the fjord (Figure 1b). Several observations suggest that this community structure is temporally stable. First, all three types were sampled from two consecutive years. Moreover, our behavioral data showed that only six out of the 70 tagged cod permanently left the fjord during their tracking period (~18 months) within the four-year study period, suggesting that the different cod types do not just visit the fjord occasionally during the mobile feeding season -as shown for Atlantic cod populations in the North Sea area (Neat et al. 2014)- but they are instead present in the fjord for extended periods of time. Furthermore, stable coexistence of juvenile North Sea and fjord-type cod has been documented over a timescale of 14 years at several locations along the southern Norwegian coast (Knutsen et al. 2018). In addition, sympatric spawning is supported through the shared spawning period for North Sea and western Baltic cod from January to April (Hüssy 2011; Neat et al. 2014), the occurrence of North Sea, western Baltic, and fjord-type adult cod in spawning condition in other southern Norwegian fjords (Barth et al. 2017a), and the simultaneous presence of young-stage cod eggs from at least North Sea and fjord-type cod in the fjords (Knutsen et al. 2007; Ciannelli et al. 2010, Jorde et al. 2018b). Finally, male cod will typically defend territories at specific sites connected to the fjord bottom during the spawning period and thus reduced diel vertical migration can be expected for individuals that take part in spawning (Meager et al. 2009; 2012). We detected reduced diel vertical migration during the spawning season for the fjord-type cod (TVEf) as well as the North Sea-type cod (TVEn), indicating that both types are active in spawning. All these observations of previous and the current study indicate a stable sympatric cod community within the Tvedestrand fjord. Although we found a very low abundance of the western Baltic-type in the fjords, an earlier study where several southern Norwegian fjords and a larger number of individuals were sampled detected almost equal numbers of North Sea, western Baltic, and fjord-type cod (Barth et al. 2017a).

Two questions then arise: Which barriers limit gene flow between sympatric cod types and maintain genetic differentiation? How does genetic exchange between North Sea and western Baltic-type cod in the fjords with the respective offshore populations occur? Our phylogenetic relationships showed that the fjord (TVEf) and North Sea-type (TVEn) appear in two different well-supported clades, with TVEn nested among the North Sea populations with large topological uncertainties, underlining that these cod belong to the same genotypic unit (Figure 1c). The genetic similarity of TVEf and the western Baltic cod may further indicate an origin from the western Baltic population, which could have evolved as a response to random or non-random dispersal, driven, for example, by habitat matching due to similar habitats of reduced salinity (Edelaar & Bolnick 2012). These data support a scenario of either recent (sympatric) divergence, or of allopatric divergence and secondary contact with (i) insufficient time for homogenization or with (ii) reduced gene flow between fjord-type (TVEf) with North Sea (TVEn) and western Baltic-type cod. The latter possibly due to the presence of reproductive barriers, while genetic exchange among oceanic populations is maintained through, for example, the regular supply of fjords with larvae of offshore origin, which has already been documented to occur through oceanic drift of pelagic eggs and larvae (Stenseth et al. 2006; Jonsson et al. 2016; André et al. 2016; Barth et al. 2017a). However, in this last scenario, it is still uncertain whether the chromosomal inversions or other barriers maintain genetic separation.

Frequency shifts of inversions do not fully explain differentiation of Atlantic cod types

Chromosomal inversions can promote divergence through the suppression of recombination (Kirkpatrick & Barton 2006), and ample investigations in several organisms such as insects (Ayala et al. 2014; Lohse et al. 2015), and plants (Lowry & Willis 2010; Twyford & Friedman 2015), but also fish (Jones et al. 2012; Fan & Meyer 2014; Kirubakaran et al. 2016) have described a central role of inversions in divergence with gene flow. This view is also supported by recent simulation studies (Yeaman 2013; Feder et al. 2014b). Genome scans of North Sea (TVEn), western Baltic (KIE), and fjord-type (TVEf) cod revealed low genomic differentiation across most of the genome with the exception of three previously identified inverted regions on LG02, 07, and 12 (Figure 2a, Berg et al. 2016; Sodeland et al. 2016; Kirubakaran et al. 2016). Thus, these three inversions are prime candidates for barrier mechanism that might act through either intrinsic (genome incompatibilities) or extrinsic (fitness effect associated with the environment) isolation.

Genetic characterization of the inversions revealed frequency shifts of the ancestral and inverted arrangements between TVEf and TVEn and bi-allelic segregation, in which the ancestral arrangement shows less divergence among populations than the inverted arrangement (Figure 2b). Such bi-allelic segregation can be attributed to diversifying selection acting on the inverted arrangement, or it could indicate that the ancestral, more common arrangement is subject to ongoing homogenization by gene flow, while recombination in the less frequent

inverted arrangement is reduced. Interestingly, phylogenetic analyses revealed population-specific clustering of the ancestral and inverted arrangements of fjord-type cod (TVEf) on LG07 and 12, but not on LG02 (Figure 2c), suggesting that the inversions have independent properties, where the ancestral arrangement on LG02 experiences more exchange among populations than the ancestral arrangement of the other two LGs. Thus, sets of co-adaptive alleles may be captured in one or several of the inverted regions and protected from recombination, creating an adaptive barrier to gene flow through the fitness advantage of individuals carrying the inversion in the local environment (Kirkpatrick & Barton 2006). Indeed, for eastern Baltic cod living in low salinity conditions, key genes important for osmoregulation have been described in and around the inversion on LG02 (Berg et al. 2015), which could also provide adaptive value in fjord environments (Barth et al. 2017a). Furthermore, temperature adaptations through physiological adjustments and oxygen consumption (Grenchik et al. 2013) may be important properties in the fjord environment, which is characterized by stable and stratified temperatures and decreased oxygen concentrations in deeper layers (Saetre 2007; Halvorsen 2013). Notably, the inversions on LG02 and LG12 have been shown to contain genes associated with temperature and oxygen regulation (Bradbury et al. 2010; Therkildsen et al. 2013; Berg et al. 2015). However, GO enrichment analyses of genes within the inverted region on LG12 showed no significant enrichment, while in the inversion on LG02 genes involved in DNA/chromatin structuring were found to be significantly enriched, and the inversion on LG07 showed a significant enrichment of genes in signaling and metabolic processes (Barney et al. 2017). Adaptive genes residing in inversions have also been described in other species, for example the threespine stickleback (Jones et al. 2012) or the willow warbler (Lundberg et al. 2017). Yet, such correlations between inversions and the environment may also be caused by intrinsic genetic incompatibilities that merely coincide with ecological barriers (Bierne et al. 2011). In conflict with a hypothesized intrinsic postzygotic barrier is the fact that all possible 27 combinations of the three inverted regions have been observed in wild Atlantic cod (data not shown). In our dataset, all but three of these 27 combinations occured. However, due to the low frequency of inverted arrangements, the chance of sampling all combinations within ~200 individuals is small. On the other hand, an intrinsic postzygotic barrier where problems during meiotic chromosome pairing in heterozygotes may lead to sterility or underdominance cannot be ruled out, but seems unlikely given that none of the inversions was found to be fixed within a population, all rearrangement haplotypes conform to HWE expectations, and heterozygotes are abundant. In addition, extrinsic postzygotic barriers where hybrids are unfit in the local environment seem similarly unlikely, since we did not observe decreased fitness (survival) of heterozygotes for any of the inversions. However, significantly reduced fitness in the fjord environment of individuals homozygous for the LG12 inversion (Fig. 3, Table 4) may indicate immigrant inviability (Nosil et al. 2005) and selection against North Sea migrants, which were also shown to have an overrepresentation of the LG12 inverted arrangement (Fig. 2d).

Although frequency shifts of the inverted arrangement indicates adaptive properties, and individuals homozygous for the LG12 inversion showed lower fitness in the fjord environment, the lack of fixed inverted arrangements implies that the inversions are not purely diagnostic for genotype fate. This is in contrast to the high degree of fixation of the LG01 inversion in Northeast Arctic cod, which has been linked to migratory behavior (Berg et al. 2016; Kirubakaran et al. 2016). Behavioral traits, especially traits related to reproductive behavior, have previously also been associated with inversions in the white-throated sparrow (Tuttle et al. 2016) and the ruff (Küpper et al. 2016). However, in line with the lack of fixed inversions, support for a correlation between the inversions and the tested behavioral traits was weak and restricted to a tendency for larger home ranges during the spawning season in individuals having the LG07 inverted arrangement.

The lack of such direct relation between fjord-type cod (TVEf) and the inverted arrangements suggests that adaptive alleles may perhaps have manifested elsewhere in the genome. In our genome scans, we did not detect other large regions (more than few kb) with tightly linked loci (indicating reduced recombination) that are also differentiated between TVEf and TVEn/KIE and thus may be protected from gene flow and have a barrier effect. However, it has been shown that chromosomal rearrangements are not always required to maintain a barrier to gene flow (Davey et al. 2017), and large-effect alleles that are persisting gene flow could as well be located within other genomic regions. We indeed found several smaller differentiation peaks on different LGs, suggesting that genomic differentiation between TVEn and TVEf is more widespread across the whole genome than what was previously expected due to the sympatric occurrence and connectivity (Barth 2017a). On the other hand, fixed alleles were not detected. The processes leading to differentiation peaks are complicated and difficult to interpret, and further research including cline analysis, the identification of introgression and the direction and strength of gene flow, identification of selection axes, and analysis of the demographic history (Noor & Bennett 2009; Cruickshank & Hahn 2014; Ravinet et al. 2017) will be required to fully determine their importance.

Localization of many weakly differentiated loci across the genome could also be an indication for polygenic adaptation (Pritchard & Di Rienzo 2010), which would, however, be expected to be quickly broken up through gene flow and recombination if selection is not strong enough, leading to maladapted intermediate genotypes (Lenormand 2002; Yeaman &Otto 2011; Yeaman & Whitlock 2011). Nevertheless, our outlier analyses detected several candidate loci possibly under positive selection, mostly located within regions that have also been identified as differentiation peaks in the genome scans (Figure S4). No significant GO term enrichment was found, but predicted gene models underlying the candidate loci included various genes associated with functions in salinity, temperature, and oxygen adaptation: A solute carrier gene (*SLCO1C1*) on LG16 controlling in- and efflux of solutes, which may be important for egg buoyancy regulation under different salinities (Berg et al. 2015), an ion channel (*KCNA10*) on LG01, found to be up-regulated under salinity stress in blue mussels (Lockwood & Somero, 2011), as well as the phosphatidylethanolamine *N*-methyl transferase (*PEMT*) gene located on LG18,

which is involved in synthesizing phosphatidylcholine, a major component of membranes shown to occur in different compositions related to salinity and temperature (Farkas et al. 2001; Athamena et al. 2011). Two additional genes (*Abhd15, PDE3A*, both on LG16) that have previously been associated with temperature adaptation (Scott et al. 2012; Dikmen et al. 2013), were also detected. Lastly, a nitric oxide (NO) binding gene (*THAP4*) located on LG08, possibly acting as NO-dependent sensor and transcriptional regulator (Bianchetti et al. 2011), was identified. The conversion of nitrite to NO has been shown to occur in fish under hypoxic conditions (Jensen 2009), thus suggesting a function of *THAP4* in adaptation to low oxygen levels. Interestingly, adaptation to salinity in the Atlantic herring, a marine fish with similarly high levels of connectivity as Atlantic cod, seems to be immensely polygenic (Lamichhaney et al. 2012) and associated with 10-200 kb haplotype blocks that are unlikely to be inversions, but instead predicted to have evolved through the accumulation of multiple causal variants maintained by selection (Martinez-Barrio et al. 2016). Similarly, in a recently diverged ecotype pair of resident stream and migratory lake sticklebacks that reproduce in sympatry, 19 differentiated regions averaging 267 kb have been suggested to facilitate adaptation (Marques et al. 2016).

Differentiated haplotype blocks may, along with environmentally adaptive alleles, also contain alleles that are associated with prezygotic isolating mechanisms, such as behavioral differences, which are important for the reduction of gene flow when postzygotic barriers are incomplete (Kopp et al. 2018). Since different barriers can accumulate during the isolation process and/or act sequentially over the entire life cycle of an organism (Coyne & Orr 2004; Kulmuni & Westram 2017), multiple sites may manifest as polygenic traits. In addition, behavioral barriers do not necessarily need to be established as genomic differences, but can as well be caused by epigenetic modifications and the subsequent differential expression of genes (i.e., plastic responses), making them more difficult to detect in genome analyses (Ledon-Rettig et al. 2013; Reusch 2014).

Behavioral differences between Atlantic cod types suggest the existence of additional barriers

Behavioral differences that may act as barriers to gene flow include spawning habitat preferences (ecological or spatial isolation, i.e., allopatric reproduction), temporal isolation (e.g., differences in spawning time), as well as assortative mating. In Atlantic cod, spawning site fidelity (Skjæraasen et al. 2011), natal homing (Bonanomi et al. 2016), and mate choice (Hutchings et al. 1999; Nordeide & Folstad 2000; Rudolfsen et al. 2005) have been described and discussed to be involved in fitness, local adaptation and population divergence.

Our main behavioral findings showed that fjord-type (TVEf) cod utilize generally deeper habitats within the fjord as compared to North Sea-type (TVEn) cod (Figure 3). Differences in habitat use may lead to a minimization of encounters among the cod types, which could create a prezygotic barrier to gene flow during the spawning season through allopatric reproduction. Contrasting use of different habitat depths is also known from the recently diverged sympatric species pair of *Pundamilia* cichlids (Meier et al. 2017; 2018), while in Atlantic cod so far only

temporal differences in spawning ground usage among populations have been described (Hüssy 2011; Hüssy et al. 2016). Interestingly, tagging experiments of two Icelandic Atlantic cod ecotypes showed shared depth usage during the spawning season, but different depth ranges during the feeding period (Pampoulie et al. 2007). Furthermore, for juvenile Northeast Arctic cod and coastal cod, which are co-occurring at northern Norwegian spawning grounds, different settlement depths have been described (Fevolden et al. 2012). However, the degree to which the here detected differences contribute to reproductive isolation and represent a barrier to gene flow cannot be unequivocally determined due to overlap in space and habitat use. Moreover, limited spatial resolution of acoustic telemetry makes it difficult to locate the individuals at the scale at which reproductive events likely occur. Nevertheless, since cod behavior is highly variable, both within a single individual (over time) and among individuals (Villegas-Ríos et al. 2017), the observed differences between cod-types are likely grounded on deep ecological differences. Behavioral monitoring at a finer scale (e.g., using a VPS system, Freitas et al. 2016) during the spawning season will be needed to conclude if the behavioral differences reported in this study translate into actual allopatric reproduction.

Behavioral differences between cod types during the spawning period might also be an indication for assortative mating where alike individuals mate with their kind, leading to reduced gene flow between different ecotypes (Kopp et al. 2018). Significant differences in size between North Sea-type and fjord-type cod have been found (Knutsen et al. 2018), generating an opportunity for size-selective assortative mating (Taggart et al. 2001; Rueger et al. 2016). Consistent with this, we found behavioral differences between larger and smaller individuals (Figure 3); however, since individuals were not aged, age-specific behavior cannot be excluded. Further investigation through, for example, network statistics (Jacoby & Freeman 2016) is needed to identify stable interactions between individuals to shed more light on the role of assortative mating in maintaining ecotypes.

Nevertheless, assortative mating is also hypothesized to increase the likelihood of producing fit offspring and has been observed in many animals (Jiang et al. 2013). Under captive conditions, mate choice and a resulting increase in fitness have also been described for Atlantic cod (Hutchings et al. 1999; Rudolfsen et al. 2005), and male reproductive success was shown to be dependent on the magnitude of the size difference between the female and the male (Bekkevold et al. 2002). In our study, fjord-type (TVEf) cod showed higher survival, indicative for a higher fitness within the fjord environment as compared to North Sea-type (TVEn) cod. Such fitness advantage could arise through natural selection as described above, or through fisheries-induced selection where Atlantic cod residing at lower depths are more likely to be harvested (Olsen & Moland 2011; Olsen et al. 2012). However, since the central part of the study area is a no-take marine reserve where no fishing is allowed, harvesting mortality is unlikely. In a recent paper by Jorde et al (2018b), newly spawned eggs of North Sea-type cod were found inside two fjords, however, with a variable pattern between the fjords. One of the fjords remained structured throughout the season with fjord-type cod dominating inside the fjord, while the other fjord

showed fluctuations in cod-type frequency, eventually leading to a dominance of fjord-type cod inside, and North Sea-type cod outside. This latter observation suggests that fjord-type cod possibly has a fitness advantage in sheltered fjord habitats, which could for example arise through adaptive advantages for oocyte growth and survival in warmer water layers (Bradbury et al. 2001; Kjesbu et al. 2010). The Tvedestrand fjord contains several deep basins divided by shallow sills and the water column consists of a surface freshwater layer and saline water underneath where salinity increases with depth, but oxygen decreases (Ciannelli et al. 2010; Halvorsen 2013). The water temperature is generally more stable and stratified in the fjord as compared to outer coastal or oceanic areas, with deeper layers inside the fjord being comparatively warm in winter and relatively cold in summer (Saetre 2007). In contrast, exposed outer-fjord areas may experience extensive mixing that causes the temperature in deeper water to vary more with the seasons (Saetre 2007). Cod individuals would thus experience less temperature variation inside the fjords, than outside. The reduced mean depth of fjord-type (TVEf) cod indicates residence in water layers with generally lower temperature and oxygen, which could require special adaptations of adults and/or eggs and larvae. In line with this assumption, our outlier analyses detected genes associated with adaptation to temperature and hypoxic conditions (see above). Furthermore, many of the detected genes were also associated with a function in osmoregulation, possibly regulating drift depth of the eggs, where eggs neutrally buoyant in less saline upper layers are retained within the fjord (Ciannelli et al. 2010; Jung et al. 2012). Alternatively, North Sea-type cod may be better adapted to a life in warmer water layers, while the fjord-type cod have to seek more stable layers (Freitas et al. 2016). However, adaptation to such habitat has also been discussed in relation to the inversions (see above), indicating that multiple barriers to gene flow may exist.

Concluding remarks

Isolating barriers to gene flow are best studied in differentiated populations that occur in sympatry, where such barriers actively prevent admixture. Here, we demonstrate extensive evidence for the temporally stable sympatric occurrence of genetically and behaviorally differentiated Atlantic cod types within a confined fjord environment. We show that these differences are likely to be maintained through a combination of structural and behavioral barriers to gene flow, both of which may reflect a fitness advantage in local environments. Our study thus emphasizes the high value of genomic analyses for conservation and fisheries management (Bernatchez et al. 2017), while simultaneously highlighting the role of prezygotic, behavioral mechanisms in shaping community structures in the sea of connectivity.

5 ACKNOWLEDGEMENTS

We thank H. T. Baalsrud and M. Matschiner for assistance in sampling, M. Skage, S. Kollias, A. Tooming-Klunderud, M. H. Selander Hansen, and H. Rydbeck from the Norwegian Sequencing Centre for sequencing and processing of samples, and four anonymous reviewers for their constructive comments that greatly improved this manuscript. J.M.I.B. thanks W. Salzburger for providing a stimulating workspace. The computational genomic work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at the University of Oslo IT-department (USIT). This work was funded by the Research Council of Norway (RCN) through "The Aqua Genome Project" (221734/O30) to K.S.J. Additional support was also obtained by the RCN "Ecogenome project" (280453/E40) to H.K., by the Centre for Marine Evolutionary Biology at Göteborg University to C.A., and a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme, project #625852 (BE-FISH) granted to D.V.R.

6 REFERENCES

Aalvik IM, Moland E, Olsen EM, Stenseth NC (2015) Spatial ecology of coastal Atlantic cod *Gadus morhua* associated with parasite load. *Journal of Fish Biology, 87, 449–464*.

Alexa A, Rahnenfuhrer J (2016) topGO: Enrichment Analysis for Gene Ontology. *R package v2.26.0.*

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19, 1655–1664*.

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*.

Athamena A, Brichon G, Trajkovic-Bodennec S *et al.* (2011) Salinity regulates *N*-methylation of phosphatidylethanolamine in euryhaline crustaceans hepatopancreas and exchange of newly-formed phosphatidylcholine with hemolymph. *Journal of comparative physiology. B, 181, 731–740*.

Ayala D, Guerrero RF, Kirkpatrick M (2013) Reproductive isolation and local adaptation quantified for a chromosome inversion in a malaria mosquito. *Evolution, 67, 946–958.*

Ayala D, Ullastres A, González J (2014) Adaptation through chromosomal inversions in Anopheles. *Frontiers in genetics, 5:129*.

Bianchetti CM, Bingman CA, Phillips GN Jr. (2011) Structure of the C-terminal heme-binding domain of THAP domain containing protein 4 from Homo sapiens. *Proteins, 79, 1337–1341*.

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*

Barth JMI, Berg PR, Jonsson PR et al. (2017a) Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology, 26, 4452-4466*

Barth JMI, Damerau M, Matschiner M, Jentoft S, Hanel R (2017b) Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna (*Thunnus albacares*) populations. *Genome Biology and Evolution, 9, 1084–1098*.

Bekkevold D, Hansen MM, Loeschcke V (2002) Male reproductive competition in spawning aggregations of cod (*Gadus morhua*, L.). *Molecular Ecology, 11, 91–102*.

Berg PR, Jentoft S, Star B et al. (2015) Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biology and Evolution, 7, 1644–1663*.

Berg PR, Star B, Pampoulie C et al. (2016) Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific reports, 6:23246*.

Berg PR, Star B, Pampoulie C et al. (2017) Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity, 119, 418-428*.

Bernatchez L, Wellenreuther M, Araneda C et al. (2017) Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution, 32, 665–680*.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology, 20, 2044–2072*.

Bonanomi S, Overgaard Therkildsen N, Retzel A et al. (2016) Historical DNA documents long-distance natal homing in marine fish. *Molecular Ecology, 25, 2727–2734*.

Bouckaert R, Heled J, Kühnert D et al. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology, 10:e1003537*.

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

Bradbury IR, Snelgrove PVR, Fraser S (2001) The influence of temperature on advective loss of Atlantic cod (*Gadus morhua*) eggs from the inshore environment. *Fisheries Oceanography, 10, 342–352*.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution, 29, 1917–1932*.

Burnham KP and Anderson DR. 1998. Model selection and inference. Springer, New York.

Castiglia R (2013) Sympatric sister species in rodents are more chromosomally differentiated than allopatric ones: implications for the role of chromosomal rearrangements in speciation. *Mammal Review, 44, 1–4.*

Ciannelli L, Knutsen H, Olsen EM et al. (2010) Small-scale genetic structure in a marine population in relation to water circulation and egg characteristics. *Ecology, 91, 2918–2930*.

Cowen RK, Sponaugle S (2009) Larval dispersal and marine population connectivity. *Annual Review of Marine Science, 1, 443–466*.

Coyne JA, Orr HA (2004) Speciation. *Sinauer Associates Incorporated, Massachusetts*.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology, 23, 3133-3157*.

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. *Bioinformatics, 27, 2156–2158*.

Davey JW, Barker SL, Rastas PM et al. (2017) No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions*. Evolution Letters, 1, 138–154*.

Dikmen S, Cole JB, Null DJ, Hansen PJ (2013) Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. *PLoS ONE, 8:e69202*.

Edelaar P, Bolnick DI (2012) Non-random gene flow: anunderappreciated force in evolutionand ecology. *Trends in Ecology & Evolution, 27, 659–665.*

Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics, 7, 1–26*.

Espeland SH, Albretsen J, Olsen EM, Bodvin T (2015) Modelling drift of pelagic offspring: the importance of egg surveys in providing a realistic model initialization. *ICES Journal of Marine Science, 72, 2578–2589*.

Espeland SH, Gundersen AF, Olsen EM et al. (2007) Home range and elevated egg densities within an inshore spawning ground of coastal cod. *ICES Journal of Marine Science, 64, 920–928*.

Fan S, Meyer A (2014) Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Frontiers in genetics, 5:163*.

Farkas T, Fodor E, Kitajka K, Halver JE (2001) Response of fish membranes to environmental temperature. *Aquaculture Research, 32, 645–655.*

Feder JL, Egan SP, Nosil P (2012a) The genomics of speciation-with-gene-flow. *Trends in Genetics, 28, 342–350*.

Feder JL, Gejji R, Yeaman S, Nosil P (2012b) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B, 367, 461–474*.

Feder JL, Nosil P, Wacholder AC *et al.* (2014a) Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *Journal of Heredity, 105, 810–820*.

Feder JL, Nosil P, Flaxman SM (2014b) Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. *Frontiers in genetics, 5:295*.

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

Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP (2013) Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). *Evolution, 67, 2547–2560*.

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*.

Freitas C, Olsen EM, Knutsen H, Albretsen J, Moland E (2016) Temperature-associated habitat selection in a cold-water marine fish. *Journal of Animal Ecology, 85, 628–637*.

Freitas C, Olsen EM, Moland E, Ciannelli L, Knutsen H (2015) Behavioral responses of Atlantic cod to sea temperature changes. *Ecology and Evolution, 5, 2070–2083*.

Grenchik MK, Donelson JM, Munday PL (2012) Evidence for developmental thermal acclimation in the damselfish, *Pomacentrus moluccensis*. *Coral Reefs, 32, 85–90.*

Halvorsen MD (2013) The distribution of Skagerrak coastal cod (*Gadus morhua*) in relation to oxygen depletion, temperature and salinity, studied by acoustic telemetry in the Tvedestrand fjord in south-eastern Norway. *Master Thesis, 1–58*.

Harrison PM, Gutowsky LFG, Martins EG et al. (2015) Personality-dependent spatial ecology occurs independently from dispersal in wild burbot (*Lota lota*). *Behavioral Ecology, 26, 483–492*.

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*.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development, 16, 592–596*.

Hooper DM, Price TD (2017) Chromosomal inversion differences correlate with range overlap in passerine birds. *Nature Ecology & Evolution, 1, 1526–1534*.

Howe JA, Austin WEN, Forwick M et al. (2010) Fjord systems and archives: a review. *Geological Society, London, Special Publications, 344, 5–15*.

Hutchings JA, Bishop TD, McGregor-Shaw CR (1999) Spawning behaviour of Atlantic cod, *Gadus morhua*: evidence of mate competition and mate choice in a broadcast spawner. *Canadian Journal of Fisheries and Aquatic Sciences, 56, 97–104*.

Hüssy K (2011) Review of western Baltic cod (*Gadus morhua*) recruitment dynamics. *ICES Journal of Marine Science, 68, 1459–1471.*

Hüssy K, Hinrichsen HH, Eero M et al. (2016) Spatio-temporal trends in stock mixing of eastern and western Baltic cod in the Arkona Basin and the implications for recruitment. *ICES Journal of Marine Science, 73, 293–303.*

Jacoby DMP, Freeman R (2016) Emerging network-based tools in movement ecology. *Trends in Ecology & Evolution, 31, 301–314*.

Jensen FB (2009) The role of nitrite in nitric oxide homeostasis: a comparative perspective. *Biochimica et Biophysica Acta, 1787, 841–848*.

Jiang Y, Bolnick DI, Kirkpatrick M (2013) Assortative mating in animals. *The American Naturalist, 181, E125–38*.

Jones FC, Brown C, Pemberton TJ, Braithwaite VA (2006) Reproductive isolation in a threespine stickleback hybrid zone. *Journal of Evolutionary Biology, 19, 1531–1544*.

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*.

Jorde PE, Andersson A, Ryman N, Laikre L (2018a) Are we underestimating the occurrence of sympatric populations? *Molecular Ecology*, *27, 4011–4025*.

Jorde PE, Synnes A-E, Espeland SH, Sodeland M, Knutsen H (2018b) Can we rely on selected genetic markers for population identification? Evidence from coastal Atlantic cod. *Ecology and Evolution, in press.*

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*.

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

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research, 40:e3*.

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*.

Kjesbu OS, Righton D, Krüger-Johnsen M *et al.* (2010) Thermal dynamics of ovarian maturation in Atlantic cod (*Gadus morhua*). *Canadian Journal of Fisheries and Aquatic Sciences, 67, 605–625*.

Knutsen H, Jorde PE, Hutchings JA et al. (2018) Stable coexistence of genetically divergent Atlantic cod ecotypes at multiple spatial scales. *Evolutionary Applications, early view*.

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*.

Kopp M, Servedio MR, Mendelson TC et al. (2018) Mechanisms of assortative mating in speciation with gene flow: Connecting theory and empirical research. *The American Naturalist, 191, 1–20*.

Kulmuni J, Westram AM (2017) Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Molecular Ecology, 26, 3093–3103*.

Küpper C, Stocks M, Risse JE *et al.* (2016) A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics, 48, 79–83*.

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*.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution, 37, 1210–1226*.

Ledon-Rettig CC, Richards CL, Martin LB (2013) Epigenetics for behavioral ecologists. *Behavioral Ecology, 24, 311–324*.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution, 17, 183–189*.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27, 2987–2993*.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25, 1754–1760*.

Lockwood BL, Somero GN (2011) Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Molecular Ecology, 20, 517–529*.

Lohse K, Clarke M, Ritchie MG, Etges WJ (2015) Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution, 69, 1178–1190*.

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*.

Lundberg M, Liedvogel M, Larson K et al. (2017) Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters, 1, 155–168*.

Malmstrøm M, Matschiner M, Tørresen OK et al. (2016) Evolution of the immune system influences speciation rates in teleost fishes. *Nature Genetics, 48, 1204–1210*.

Marques DA, Lucek K, Meier JI et al. (2016) Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics, 12:e1005887*.

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*.

Matschiner M, Star B, Berg PR et al. (2015) Inversions or introgression? *Poster, Congress of the European Society for Evolutionary Biology (ESEB), http://evoinformatics.eu/presentations.htm.*

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

Meager JJ, Skjæraasen JE, Fernö A *et al.* (2009) Vertical dynamics and reproductive behaviour of farmed and wild Atlantic cod *Gadus morhua*. *Marine Ecology Progress Series, 389, 233–243*.

Meager JJ, Skjæraasen JE, Karlsen Ø et al. (2012) Environmental regulation of individual depth on a cod spawning ground. *Aquatic Biology, 17, 211–221*.

Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O (2018) Genomics of parallel ecological speciation in Lake Victoria cichlids. *Molecular Biology and Evolution, 35, 1489–1506*.

Meier JI, Sousa VC, Marques DA et al. (2017) Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. *Molecular Ecology, 26, 123–141*.

Munk P, Larsson PO, Danielssen DS, Moksness E (1999) Variability in frontal zone formation and distribution of gadoid fish larvae at the shelf break in the northeastern North Sea. *Marine Ecology Progress Series, 177, 221–233*.

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics, 12, 443–451*.

Neat FC, Bendall V, Berx B et al. (2014) Movement of Atlantic cod around the British Isles: implications for finer scale stock management. *Journal of Applied Ecology, 51, 1564–1574*.

Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity, 103, 439–444*.

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

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology, 17, 2103–2106*.

Olsen EM, Moland E (2011) Fitness landscape of Atlantic cod shaped by harvest selection and natural selection. *Evolutionary Ecology, 25, 695–710*.

Nosil P, Vines TH, Funk DJ (2005) Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution, 59, 705–719*.

Olsen EM, Heupel MR, Simpfendorfer CA, Moland E (2012) Harvest selection on Atlantic cod behavioral traits: implications for spatial management. *Ecology and Evolution, 2, 1549–1562*.

Pampoulie C, Jakobsdóttir KB, Marteinsdóttir G, Thorsteinsson V (2007) Are Vertical Behaviour Patterns Related to the Pantophysin Locus in the Atlantic Cod (*Gadus morhua* L.)? *Behavior Genetics, 38, 76–81*.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics, 2:e190*.

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News, 6, 7-11.*

Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics, 11, 665–667*.

Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81, 559–575*.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology,67, 901-904*.

Ravinet M, Faria R, Butlin RK et al. (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology, 30, 1450–1477*.

R core team (2018) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, http://R-project.org*

Reusch TBH (2014) Climate change in the oceans: evolutionary versus phenotypically plastic responses of marine animals and plants. *Evolutionary Applications, 7, 104–122*.

Rice WR (1988) Analyzing tables of statistical tests. *Evolution, 43, 223–225*.

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

Rogers LA, Olsen EM, Knutsen H, Stenseth NC (2014) Habitat effects on population connectivity in a coastal seascape. *Marine Ecology Progress Series, 511, 153–163*.

Roney NE, Oomen RA, Knutsen H, Olsen EM, Hutchings JA (2018) Temporal variability in offspring quality and individual reproductive output in a broadcast-spawning marine fish. *ICES Journal of Marine Science, 75, 1353-1361*.

Rudolfsen G, Figenschou L, Folstad I, Nordeide JT, Soreng E (2005) Potential fitness benefits from mate selection in the Atlantic cod (*Gadus morhua*). *Journal of Evolutionary Biology, 18, 172–179*.

Rueger T, Gardiner NM, Jones GP (2016) Size matters: male and female mate choice leads to size-assortative pairing in a coral reef cardinalfish. *Behavioral Ecology, 27, 1585–1591*.

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology, 23, 4757–4769*.

Saetre R (2007) The Norwegian coastal current: Oceanography and climate. *Tapir Academic Press, Trondheim.*

Schubert M, Ermini L, Sarkissian Der C et al. (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols, 9, 1056–1082*.

Scott GR, Johnston IA (2012) Temperature during embryonic development has persistent effects on thermal acclimation capacity in zebrafish. *Proceedings of the National Academy of Sciences, 109, 14247–14252*.

Seehausen O, Butlin RK, Keller I et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics, 15, 176–192*.

Simpfendorfer CA, Heupel MR, Hueter RE (2002) Estimation of short-term centers of activity from an array of omnidirectional hydrophones and its use in studying animal movements. *Canadian Journal of Fisheries and Aquatic Sciences, 59, 23–32*.

Sinclair-Waters M, Bradbury IR, Morris CJ et al. (2018) Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic cod in the northwest Atlantic. *Molecular Ecology, 27, 339–351*.

Skjæraasen JE, Meager JJ, Karlsen Ø, Hutchings JA, Fernö A (2011) Extreme spawning-site fidelity in Atlantic cod. *ICES Journal of Marine Science, 68, 1427–1477*.

Sodeland M, Jorde PE, Lien S et al. (2016) "Islands of Divergence" in the Atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution, 8, 1012–1022*.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30, 1312–1313*.

Stange M, Sánchez-Villagra MR, Salzburger W, Matschiner M (2018) Bayesian divergence-time estimation with genome-wide SNP data of sea catfishes (*Ariidae*) supports miocene closure of the Panamanian isthmus. *Systematic Biology, 67, 681-699*.

Star B, Boessenkool S, Gondek AT et al. (2017) Ancient DNA reveals the Arctic origin of Viking Age cod from Haithabu, Germany. *Proceedings of the National Academy of Sciences, 114, 9152–9157*.

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

Star B, Tørresen OK, Nederbragt AJ *et al.* (2016) Genomic characterization of the Atlantic cod sex-locus. *Scientific reports, 6, 31235*.

Stenseth NC, Jorde PE, Chan K-S et al. (2006) Ecological and genetic impact of Atlantic cod larval drift in the Skagerrak. *Proceedings of the Royal Society B, 273, 1085–1092*.

Taggart JB, McLaren IS, Hay DW, Webb JH, Youngson AF (2001) Spawning success in Atlantic salmon (*Salmo salar* L.): a long-term DNA profiling-based study conducted in a natural stream. *Molecular Ecology, 10, 1047–1060*.

Therkildsen NO, Hemmer-Hansen J, Hedeholm RB et al. (2013) Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. *Evolutionary Applications, 6, 690–705*.

Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular Ecology, 25, 2144–2164*.

Tuttle EM, Bergland AO, Korody ML et al. (2016) Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology, 26, 344–350*.

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

Tørresen OK, Star B, Jentoft S et al. (2017) An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics, 18:95*.

Villegas-Ríos D, Moland E, Olsen EM (2016) Potential of contemporary evolution to erode fishery benefits from marine reserves. *Fish and Fisheries, 18, 571–577*.

Villegas-Ríos D, Réale D, Freitas C, Moland E, Olsen EM (2017) Individual level consistency and correlations of fish spatial behaviour assessed from aquatic animal telemetry. *Animal Behaviour, 124, 83–94*.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution, 38, 1358–1370*.

Wellenreuther M, Bernatchez L (2018) Eco-Evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution, 33, 427-440*.

Wolf JBW, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics, 18, 87–100*.

Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences, 110, 1743–1751*.

Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution, 65, 2123–2129*.

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution, 65, 1897–1911*.

7 DATA ACCESSIBILITY

The filtered SNP dataset, individual behavioral data, and custom scripts for sliding window analyses are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.d9c48b6. Whole sequencing read data have been deposited on the European Nucleotide Archive (ENA) under accession number PRJEB29231.

8 AUTHOR CONTRIBUTIONS

The study was conceived and designed by J.M.I.B., S.J., E.M.O, and facilitated by S.J. and K.S.J. Genomic data were compiled by B.S. and J.M.I.B. Genomic analyses were performed by J.M.I.B. Behavioral analyses were carried out by C.F., E.M., D.V.-R. and E.M.O. Samples were provided by C.A., H.K., I.B., J.D., C.P., D.R., J.M., and E.M.O. The manuscript was written by J.M.I.B. with contributions from C.F., D.V.-R., E.M., E.M.O., and H.K. All authors read and critically revised the manuscript.

9 FIGURES AND TABLES

Figure 1. Genomic variation and population relationships. (a) Map of the study area with sampling sites depicted as colored points: AVE Averøya, LOW Lowestoft, NOR North Sea, TVE Tvedestrand, ORE Öresund, KIE Kiel Bight, BOR Bornholm. Colors match sampling sites in all figures. (b) PCA of genotypes excluding linkage groups LG01, 02, 07, and 12. Shown are the two main principal component axes (PC1, PC2); main clusters are identified as belonging to the North Sea, eastern Baltic Sea, western Baltic Sea, and fjord-type cod. (c) Phylogenetic relationships inferred using the multispecies coalescence model. Canadian individuals (CAN) were used to root the tree; the Tvedestrand sample was divided into North Sea-type individuals (TVEn), and fjord-type individuals (TVEf, see inset). Node support is shown as Bayesian posterior probabilities (PP), estimated population sizes are shown by point size (theta), the white line follows the maximum clade credibility summary tree, thin colored lines show a subsample of all inferred trees.

Figure 2. Barriers to gene flow in the genomic landscape. (a) Windowed measurements of linkage disequilibrium (LD, genotype correlation, r^2), pairwise population divergence (fixation index, F_{ST}), pairwise between population sequence divergence (d_{xy}), and nucleotide diversity (π) for linkage groups (LG) 02, 07, and 12. Gray boxes outline the regions of large inversions. (b) PCA of variants within the inversions. Shown are the two main axes (PC1, PC2); main clusters are outlined by gray circles depicting individuals homozygous for the ancestral arrangement (anc), heterozygous (het), or homozygous for the inverted arrangement (inv). (c) Maximum likelihood trees of homozygous sequences within the inversions. The gray outline indicates placement of fjord-type (TVEf) specimens. The gray double bar signifies discontinuation of the branch connecting ancestral (left) and inverted (right) arrangements. (d) Frequency of homozygous ancestral (white), homozygous inverted (black), and heterozygous (gray) individuals per sampling site, as well as within fjord-type (TVEf) and North Sea-type (TVEn) fish. Colors match sampling sites in Figure 1.

Figure 3. Survival and behavioral barriers to gene flow. Mean predictions (black lines) from models describing cod behavior (a-c) and survival (d), also showing the 95% confidence bands (gray fields) and partial residuals (black dots). For each model, the predictions are scaled against a reference level (zero). Survival is estimated as relative longevity (days survived/mean days survived). TVEf = fjord-type, TVEn = North Sea-type, F = female, M = male, anc = ancestral, het = heterozygous, inv = inverted.

Table 1. Comparison of linear mixed models for predicting Atlantic cod behavior (home range, mean depth, and diel vertical migration), showing the structure and Akaike Information Criterion (AIC) of each model. Inversion state (*LG02*, *LG07*, and *LG12*), season (*S*), and genetically determined sex (*GS*) were included as factors, while body length (*L*) was included as a linear covariate. Fish ID was included as a random effect (not shown). The most parsimonious model selected for inference is shown in bold.

Home range (*HR*) model	AIC
HR = S´LG02 + S´LG07 + S´LG12 + S´L + S´GS	1689.3
HR = S´LG02 + S´LG07 + S´LG12 + S´L + GS	1687.3
HR = S´LG02 + S´LG07 + S´LG12 + S´L	1685.7
HR = S´LG02 + S´LG07 + S´LG12 + L	1695.0
HR = S´LG02 + S´LG07 + LG12 + S´L	1683.4
HR = S´LG02 + S´LG07 + S´L	1679.9
HR = S´LG02 + LG07 + S´L	1683.3
HR = LG02 + S´LG07 + S´L	1677.8
HR = S´LG07 + S´L	**1677.0**
HR = S´L	1677.3

Mean depth (*D*) model	AIC
D = S´LG02 + S´LG07 + S´LG12 + S´L + S´GS	5539.4
D = S´LG02 + S´LG07 + S´LG12 + S´L + GS	5580.1
D = S´LG02 + S´LG07 + S´LG12 + L + S´GS	5543.3
D = S´LG02 + S´LG07 + LG12 + S´L + S´GS	5538.1
D = S´LG02 + S´LG07 + S´L + S´GS	5537.3
D = S´LG02 + LG07 + S´L + S´GS	5533.6
D = S´LG02 + S´L + S´GS	5530.3
D = LG02 + S´L + S´GS	5528.3
D = S´L + S´GS	**5525.5**

Diel vertical migration (*DVM*) model	AIC
DVM = S´LG02 + S´LG07 + S´LG12 + S´L + S´GS	4507.6
DVM = S´LG02 + S´LG07 + S´LG12 + S´L + GS	4555.0
DVM = S´LG02 + S´LG07 + S´LG12 + L + S´GS	4507.7
DVM = S´LG02 + S´LG07 + LG12 + S´L + S´GS	4507.0
DVM = S´LG02 + S´LG07 + S´L + S´GS	4505.9
DVM = S´LG02 + LG07 + S´L + S´GS	4506.8
DVM = LG02 + S´LG07 + S´L + S´GS	4506.4
DVM = S´LG07 + S´L + S´GS	4503.3
DVM = S´L + S´GS	**4499.1**

Table 2. Parameter estimates (Par) with standard errors (SE) for the fixed effects included in the model selected for inference about variation in Atlantic cod behavior. Cod having the ancestral chromosomal arrangement, the female sex and the feeding season were coded as zero in the model (reference levels). Het = heterozygous, inv = inverted.

Home range model	**Par**	**SE**	***p*-value**
Intercept	-3.018	0.221	< 0.001
$LG07_{het}$	0.054	0.104	0.606
$LG07_{inv}$	-0.368	0.238	0.125
Body length	0.014	0.005	0.006
$Season_{spawning}$	-0.308	0.246	0.212
$LG07_{het}$ ´ $Season_{spawning}$	-0.098	0.112	0.383
$LG07_{inv}$ ´ $Season_{spawning}$	0.557	0.24	0.02
Body length ´ $Season_{spawning}$	0.019	0.005	< 0.001

Mean depth model	**Par**	**SE**	***p*-value**
Intercept	14.759	2.073	< 0.001
Body length	0.015	0.043	0.735
$Season_{spawning}$	3.694	2.343	0.115
Sex_{male}	1.264	0.893	0.161
Body length ´ $Season_{spawning}$	-0.135	0.047	0.004
Sex_{male} ´ $Season_{spawning}$	6.541	0.969	< 0.001

Diel vertical migration model	**Par**	**SE**	***p*-value**
Intercept	7.18	1.125	< 0.001
Body length	-0.082	0.023	0.001
$Season_{spawning}$	2.55	1.287	0.048
Sex_{male}	-0.404	0.485	0.408
Body length ´ $Season_{spawning}$	-0.063	0.026	0.016
Sex_{male} ´ $Season_{spawning}$	-4.071	0.532	< 0.001

Table 3. Comparison of linear models for predicting Atlantic cod survival (*S*), showing the structure and Akaike Information Criterion (AIC) of each model. Inversion state (*LG02*, *LG07* and *LG12*) and genetically determined sex (*GS*) were included as factors, while body length (*L*) was included as a linear covariate. The most parsimonious model selected for inference is shown in bold.

Cod survival model	**AIC**
S = LG02 + LG07 + LG12 + L + GS	51.4
HR = LG02 + LG07 + LG12 + L	51.9
HR = LG02 + LG07 + LG12	52.5
HR = LG02 + LG12 + L	47.9
HR = LG12 + L	**44.5**
HR = LG12	44.8
HR = L	47.7

Table 4. Parameter estimates (Par) with standard errors (SE) from the model selected for inference about variation in Atlantic cod survival. Individuals having the ancestral chromosomal arrangement and the female sex were coded as zero in the model (reference levels). Het = heterozygous, inv = inverted.

Cod survival	**Par**	**SE**	***P*-value**
Intercept	0.702	0.211	0.002
$LG12_{het}$	0.070	0.087	0.421
$LG12_{inv}$	-0.372	0.160	0.024
Body length	0.007	0.004	0.140

10 SUPPORTING INFORMATION

Supporting Figures

Figure S1 Hierarchical principal component analysis

Figure S2 Maximum likelihood model-based ancestry clustering

Figure S3 Maximum likelihood based phylogenetic inference

Figure S4 Per chromosome genome scans

Figure S5 Genome scans of LG16

Figure S6 Yearly differences in inversion frequencies

Figure S7 Behavioral traits and fate of individual Atlantic cod

Supporting Tables

Table S1 Detailed information about Atlantic cod (*Gadus morhua*) specimens

Table S2 Classification of specimens

Table S3 Absolut allele counts

Table S4 Genes underlying SNPs detected to be under divergent selection

Table S5 GO term enrichment

9 FIGURES AND TABLES

Disentangling structural genomic and behavioral barriers in a sea of connectivity

Julia M.I. Barth, David Villegas-Ríos, Carla Freitas, Even Moland, Bastiaan Star, Carl André, Halvor Knutsen, Ian Bradbury, Jan Dierking, Christoph Petereit, David Righton, Julian Metcalfe, Kjetill S. Jakobsen & Esben M. Olsen, Sissel Jentoft

Figure 1. Genomic variation and population relationships.

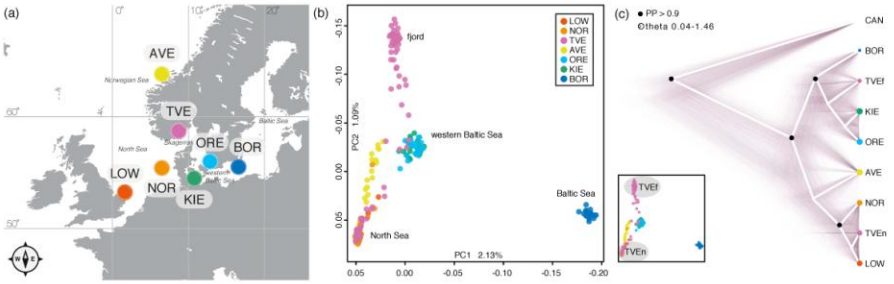

Figure 2. Barriers to gene flow in the genomic landscape.

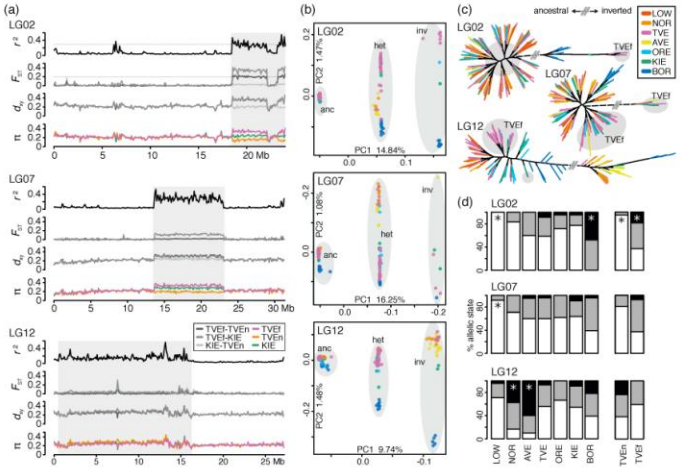

Figure 3. Survival and behavioral barriers to gene flow