Finding coarse and fine scale population structure in a coastal species: population demographics meets genomics

Diana Catarino[1,2], Per Erik Jorde[3], Lauren Rogers[4], Jon Albretsen[3], Marlene Jahnke[5], Ida Mellerud[3], Marte Sodeland[1], Carl Andre[5], Halvor Knutsen[1, 3]

[1]Centre for Coastal Research, Department of Natural Sciences, University of Agder, 4630 Kristiansand, Norway

[2] Okeanos – Institute of Marine Sciences, University of the Azores, Rua Professor Doutor Frederico Machado 4, 9901-862 Horta, Portugal

[3]Institute of Marine Research, Nye Flødevigveien 20, 4817 His, Norway

[4]Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA 98115, USA

[5]Department of Marine Sciences – Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden

***Correspondence:**

e-mail: dscatarino@gmail.com; diana.catarino@uia.no

Abstract (about 250 words)

Often population genetic studies focus on patterns at a regional scale and use spatially aggregated samples to draw inferences about population structure and drivers, potentially masking ecologically relevant population sub-structure and dynamics. In this study we use a multidisciplinary approach combining genomic, demographic, and habitat data with an oceanographic particles drift model, to unravel the patterns of genetic structure on different scales in the black goby (*Gobius niger*) along the Norwegian coast. Using a high-density sampling protocol, we observed restricted gene flow both at a surprisingly fine (kilometres) and large (100km) scale. Our results showed a pattern of isolation by distance related to the level of ocean exposure along the Skagerrak coast, where sheltered sampling stations had an overall level of genetic divergence about three times higher (F_{ST} =0.0046) than that among exposed samples (F_{ST} =0.0015). These results were also supported by demographic analyses which showed that synchrony decreases with increasing distance at much smaller scales for sheltered samples (20 km) that for exposed ones (80 km), suggesting higher levels of gene flow among the latter. We also found a pronounced genetic discontinuity between populations along the Norwegian west and east coasts with sharp increasing in genetic divergence around the southern tip of Norway, likely driven by both lack of habitat and oceanography.

Keywords: connectivity; Isolation by distance; habitat discontinuity; oceanography; marine fish

Introduction

Many marine species are broadcast spawners and disperse by early life stages drifting with ocean currents. Traditionally, these species were expected to show little or no genetic population structure across small (kms) or even large geographic scales (ocean regions; i.e. Longmore et al., 2014). Although for some species this holds true, in the past decades this paradigm has considerably shifted (Palumbi, 2004; Hauser & Carvalho 2008), since for many species it has been shown that phenomena such as ocean currents, bathymetry, local adaptation, habitat fragmentation or reduced migration, limit genetic connectivity across different spatial and temporal levels (i.e. Catarino et al., 2015; Kelley et al., 2016; Barth et al., 2019).

Additionally, many studies may have missed existing population divergence, because population genetic studies usually focus on patterns at a regional scale and use spatially aggregated samples to draw inferences about population structure and drivers, potentially masking ecologically relevant population sub-structure and dynamics. Given that most marine species have wide ranges, spatial sampling for genetic studies are often geographically coarse, since financial and technical restrains limit the number of sampled specimens and their geographical coverage. Apparent homogeneity over large areas in an initial screening may therefore yield the impression of panmixia and discourage further studies at finer geographic scales. While a coarse-scale sampling strategy may be adequate for uncovering the major population genetic patterns in a species, it risks overlooking finer scale patterns that may hint at population processes of biological relevance for local populations.

Many species, marine or terrestrial, live in habitats that in the context of individual movement and dispersal may be considered two-dimensional, meaning that dispersing individuals may reach neighbouring population(s) along several routes. Theoretical analyses have found that in two dimensional habitats, gene flow is particularly effective in counteracting genetic

divergence (Kimura & Weiss 1964; Rousset, 1997; Slatkin, 1994), implying that population genetic structure may be weak and difficult to detect in the face of sampling noise (Waples, 1998; Palumbi, 2004). A more powerful approach is to consider species that occupy a linear, one-dimensional habitat where theory predicts that gene flow is a less potent homogenizing force and where genetic structure should be more pronounced than in a two-dimensional habitat, for a given level of gene flow (Rousset, 1997). In the marine environment species living along the coastline approximate such a linear habitat, in contrast to the open ocean or in more complex seascapes which are more two-dimensional. In linear habitats, limited gene flow caused by habitat discontinuities, environmental drivers, or by geographic distance, is expected to result in more pronounced and thereby more easily detectable genetic structures. Understanding and identify the geographic scales of connectivity is a crucial step for basic and applied ecology of marine species (Cowen et al., 2006; Selkoe et al., 2016), since this information is essential to improving conservation efforts such as the design of marine reserve networks at suitable scales (Almany et al., 2009; Mertens et al., 2018) and to define fish stocks for fisheries management (Bernatchez et al., 2017).

In the present study we adopted a multidisciplinary approach to unravel gene flow and population connectivity at large and fine geographic scales of a coastal fish species. Using a high-density sampling protocol, we combine genomic data (ddRAD Seq) of the black goby with spatial population dynamics from a time-series dataset and oceanographic modelling of pelagic larval drift to identify features that influence local- and regional-scale population connectivity and genetic structure.

Material and Methods

Study Species

The black goby, *Gobius niger* Linnaeus, 1758 is widely distributed in Eastern Atlantic and Mediterranean Sea, from Mauritania (North Africa) to the eastern Atlantic coast northwards to Norway and Baltic Sea (Froese & Pauly, 2021). Black goby is a small coastal fish growing up to 17 cm total length, and they can be found mostly on sandy and muddy bottoms but also hard bottom among algae (Moen & Svensen, 2000). The species can be found from 2 to 70 m depth, but is most common in depths less than 30m, and it can also inhabit lagoons with brackish water down to 10‰ salinity (Penthon, 2005). Sexual maturity is reached at around 2 years old, followed by a sedentary benthic adult life (Miller, 1986). Black goby adult males build and protect a nest where the females deposit their eggs. Spawning begins when the water temperature reaches 12°C, and the incubation of the eggs at this temperature will last for about 20 days (Vaas et al., 1975). The eggs will be guard by the male until they hatch. After hatching, the duration of the planktonic larval phase is approximately 28 days (Planes, 1998). The species has no commercial value, however it is ecologically important in the Norwegian food web complex as prey for commercially important species such as cod, haddock, or trout (e.g. Hop et al., 1992; Knutsen et al., 2001).

Study area and sampling

In this study we focused on the Norwegian coastal system from the eastern border with Sweden, along the southern coast of Norway, and up to Ålesund (62° N). Most of the sampling sites are on the Skagerrak coast, which is a stretch of sea between North Sea and Kattegat, bordered by Denmark, Sweden, and southern Norway (Figure 1). The Norwegian coast is characterized by an indented landscape with numerous islands, islets and enclosed bays, and cut by several fjords that can extent several kilometres “inland”. The rough coastline and the scattered landmasses create a dynamic and complex oceanography along the coast, but with

111 the general pattern of the main flow westwards along the Norwegian Skagerrak coast
112 (Norwegian Coastal Current (NCC); Albretsen et al., 2012).

113 Since the 1919, the Institute of Marine Research has conducted an autumn beach seine survey
114 at a set of fixed stations along the Norwegian Skagerrak coast (Stenseth et al., 1999; Table S1).
115 The beach seine is 40 m long, with a stretched mesh size of 1.5 cm, and samples an area up to
116 700 m2 of nearshore (< 15 m depth) habitat (see Barcelo et al. (2016) and Rogers et al. (2014)).
117 Black goby abundance has been enumerated as number of individuals per haul yearly since
118 1989 at over 130 stations distributed from Torvefjorden (TOR) in southeast Norway, to the
119 Swedish border (Figure 1) and aggregated in 13 main locations (Table 1). In addition, biogenic
120 habitat type (e.g., eelgrass, green or brown macroalgae) and coverage (scaled from 1-5
121 subjectively: 1 = bare substrate to 5 = full coverage) have been characterized at each haul
122 location (Table S1). Within Skagerrak, stations range from sheltered locations inside fjords to
123 exposed areas on the open Skagerrak coast, and therefore were classified as "sheltered" or
124 "exposed" according to their degree of exposure to the open sea (see Rogers et al. 2014).

125 Genetic samples of *G. niger* were obtained from a total of 17 different locations along the
126 Norwegian coast (Figure 1, Table 1 and S1) between June and October of 2018. Fin clips were
127 taken from each fish and stored in 96% ethanol until extraction. Fine-scale sampling was
128 obtained from the Skagerrak coast during the annual beach seine survey. During this survey,
129 black gobies were sampled in 37 stations at 13 main locations (Table 1 and S1; Figure 1), and
130 these samples are hereafter referred as Skagerrak samples (SK). Further sampling, in addition
131 to the 13 main locations in the Skagerrak, was conducted for comparison purposes in Egersund
132 (EGE), and along the Norwegian West coast (WC), including in Stavanger (STA), Manger (MAN)
133 and Ålesund (ALE).

134 The collection of fish samples was conducted in accordance with Norwegian regulations and
135 had all permits required by the Norwegian Government. The studied species (*G. niger*) is not

136 listed in CITES and it is listed as “Least Concern” by the International Union for Conservation of
137 Nature (IUCN; Carpenter et al., 2015).

138 **Laboratory procedures**

139 Genomic DNA was extracted using the E-Z® 96 Tissue DNA kit (Omega, USA) following
140 manufacturer’s procedures. DNA was quantified on a Fluoroskan FL (Thermo Scientific, USA)
141 using the Quant-iT Broad Range dsDNA assay kit (Thermo Scientific, USA), and integrity was
142 checked by gel electrophoresis on a Multiskan SkyHigh spectrophotometric microplate reader
143 (Thermo Scientific, USA). Double-digest RADseq libraries (ddRAD, Peterson et al., 2012) were
144 prepared using restriction enzymes Sbf1 HF/Sph1 HF. The 95 samples (with 3 replicates) and
145 one blank per plate were digested at 37°C for 45 minutes, and then combined in a ligation
146 mixture with T4 ligase, a general P2 adapter and P1 barcoded adapter and incubated at 16°C
147 for 1 hour 30 minutes, 22°C for 1 hour and 30 minutes and heat inactivated at 65°C for 10
148 minutes. Successfully ligated samples were size selected (250-600bp) on a Blue Pippin (Sage
149 Science), followed by the AMPure magnetic bead clean-up (Beckman Coulter, USA) and then
150 amplified by PCR: 30s at 98°C, 30s at 98°C followed by 14 cycles of 10 seconds at 98°C, 30
151 seconds at 65°C, and 30 seconds at 72°C, followed by a final extension of 2 minutes at 72°C.
152 The resulting PCR was cleaned with the AMPure magnetic bead clean-up (Beckman Coulter,
153 USA). Library preparation success was then evaluated by running the libraries on a Bioanalyzer
154 (Agilent, USA), and quantified on Qubit (Thermo Scientific, USA). Libraries were diluted to
155 14pM, spiked with 5% PhiX and sequenced on an in-house Illumina MiSeq platform (160bp PE),
156 by using one lane per plate and the MiSeq v3 reagents kit.

157

158 **Bioinformatics and statistical analyses**

159 Sequencing reads for individual specimens were retrieved using the process_radtags module in
160 the Stacks software package (Catchen et al., 2013). De-novo assembly of reads from all
161 specimens, as well as variant detection and initial SNP filtering was conducted using the

162 ustacks, cstacks, sstacks and populations modules in the same software package. An average
163 of 549'980 PE 150bp reads were utilized per specimen in stacks formation. Minimum stack
164 depth (m) for SNP detection was set to 2.

165 Using VCFtools (Danecek et al., 2011), a total of five different SNPs datasets were generated,
166 each allowing different levels of missing data: 10, 30, 50, 70, 80%. Comparing the analyses of
167 these datasets showed that the main genetic pattern found did not change with different
168 levels of missing data. Therefore, the final dataset was filtered to retain SNPs with a minimum
169 read depth of 8, a minimum allele count of 2 and a maximum of 50% missing data for locus
170 and individuals. Loci showing deviations of HWE at <0.001 were also removed and the dataset
171 was further filtered to remove loci exhibiting linkage disequilibrium (LD at 0.5 level) with PLINK
172 1.9 (Purcell et al., 2007). In the end we retained a total of 611 SNPs and 292 individuals for
173 downstream analyses.

174 Summary statistics, sample sizes, observed and expected heterozygosity, and F_{IS} were
175 calculated using the "diveRsity" (Keenan et al., 2013) R package (R Core Team, 2021). Variation
176 in heterozygosity across sites was investigated using a linear model. BayeScan (v 2.1., Foll &
177 Gaggiotti 2008) and PCAdapt (Luu, Bazin, & Blum, 2017; Privé et al., 2020) were used to test if
178 loci deviated from neutrality across localities, and between west coast and Skagerrak samples.
179 BayeScan was run by setting sample size to 10,000 and the thinning interval to 50 as suggested
180 by Foll & Gaggiotti (2008). In PCAdapt population structure was first assessed via PCA, and
181 outliers were thereafter detected with respect to how they relate to population structure. The
182 number of principal components (K=2) to identify SNPs deviating from neutrality was then
183 selected by Cattell's rule (Cattell, 1966). The resulting *P*-values were corrected for false
184 positives using false-discovery rate (FDR) approach at a 0.05 threshold in both cases (Benjamini
185 & Hochberg, 1995). Following a consensus approach, only those loci flagged as outliers by both
186 methods were considered as being under selection (Narum & Hess, 2011).

187 Pairwise F_{ST} (Weir & Cockerham 1984) was calculated using Genepop R package (v1.1.7,
188 Rousset, 2008), and statistical significance of pairwise F_{ST} tests was assessed by G-tests for
189 allele frequency differences, using 10 000 dememorizations and batches, and 1000 iterations
190 per batch. *P*-values in multiple tests were evaluated for significance using the FDR approach.
191 Genetic clustering was investigated by discriminant analyses of principal components (DAPC,
192 Jombart et al., 2010) using the adegenet R package (Jombart & Ahmed, 2011), to provide a
193 graphic description of the genetic divergence among populations in multivariate space. The
194 analyses were performed for the whole dataset and for Skagerrak+Egersund separately. The
195 number of principal components kept was determined from the number of PCAs needed to
196 explain approximately 80% of variation. Optimal clustering was retrieved with the function
197 find.clusters where the best number of clusters is indicated by the lowest Bayesian information
198 criterion (BIC; Schwarz, 1978), inferred by the *k*-means algorithm with 10^6 simulations.

199 Isolation by distance (IBD) was analysed for the whole dataset and for Skagerrak only, by
200 testing for a correlation between genetic ($F_{ST}/(1 - F_{ST})$) and geographic distances (overwater
201 distances) using Mantel tests in Genepop R package, with 10 000 permutations. Overwater
202 distance was calculated using the marmap (Pante & Simon-Bouhet, 2013) and fossil (Vavrek,
203 2011) R packages, and using bathymetric data (Amante & Eakins, 2009) hosted by NOAA
204 (https://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/docs/ETOPO1.pdf). To better
205 visualize the sharp increase in F_{ST} around the southern tip of Norway, we fit a sigmoid curve to
206 the F_{ST} pairwise comparisons with the easternmost location (Hvaler, HVA), by using the *nls*
207 function in R to fit parameters to data.

208 The geographical fine-scale genetic structure of the black goby was characterized using
209 pairwise genetic differences between individuals (Rousset, 2000), as calculated by Genepop R
210 package (estimator ê). This analysis was restricted to the more densely sampled Skagerrak
211 coast and performed at station level. Possible effects of ocean currents on genetic structure
212 were tested by classifying beach seine stations as "exposed" and "sheltered" with respect to
213 the coastal current, following Rogers et al. (2014), and carrying out the analysis separately for

sheltered and exposed pairs of individuals (Table S1, Figure S1). All individuals sampled from the same station were assigned the same classification. The geographic distance between individuals were calculated as distance overwater between beach seine stations. Genetic differences (ê-estimates) for pairs of individuals from sheltered stations were then regressed (linear regression, lm function in R) against geographic distance, and similar for exposed pairs. The slopes of the regressions were tested for significance with permutations tests. The tests randomly permuted the genetic distance (ê) matrices repeatedly (10000 times) and regression slopes were calculated for the permuted data. *P*-values were calculated as the fraction of permutations that resulted in a slope that was as large or larger than the observed one.

Spatial population synchrony and decorrelation scale

Spatial population synchrony, defined as the spatial covariation in population density fluctuations (Bjørnstad et al., 1999), was calculated by taking the pairwise correlation in natural log-transformed catches among each pair of sampling stations, and relating pairwise correlations to the distance between stations. A constant of 1 was added to each count to avoid taking the logarithm of zero. Comparisons were limited to those stations sampled a minimum of 15 years and having at least 10 years with positive (non-zero) catches (n = 132 stations).

An exponential decay model (Bjørnstad et al., 1999) was fitted to pairwise correlations by distance, measured as distance over water between stations:

$$\rho(d) = \rho_r + \rho_0 e^{-\left(\frac{d}{v}\right)},$$

where ρ(d) is the pairwise correlation at distance d, ρ_r is the asymptotic correlation, or background regional correlation in recruitment, and $\rho_r + \rho_0$ is the estimated correlation at zero distance, following Rogers et al. (2014). The parameter v ("decorrelation scale") describes distance at which the pairwise correlation between time series is reduced to ~37% (e^{-1}) of that

at zero distance, relative to the background regional level of correlation. To test the hypothesis that abundance at exposed stations would covary over longer distances than sheltered stations due to differences in connectivity, models were also fit separately to pairs of exposed stations and pairs of sheltered stations. Confidence intervals (90%) were estimated based on 10,000 bootstrap replicates, following Bjørnstad et al. (1999).

Habitat associations

Habitat associations for black goby were investigated using the Skagerrak beach seine time-series (Table S1). Specifically, we tested for effects of habitat type and percent cover on abundance of black goby in beach seine hauls using a negative binomial generalized mixed effects model which included a random intercept by station. Data from all stations surveyed for at least 15 years (n = 137) were used in this analysis, for a total of 3930 individual observations of black goby abundance. Although recorded as a categorical variable with five categories (“bare substrate” to “full cover”), coverage was treated in the model as a continuous variable (ranging from 1 to 5) after preliminary model runs indicated an approximately linear relationship between abundance and coverage. Models were fit using the lme4 package (Bates et al., 2015) in R.

Oceanographic modelling

A combination of high-resolution hydrodynamical modelling and particle-tracking was applied to investigate the potential spread of larvae from all beach seine locations, and additional sites around the southern tip of Norway, and in the west encompassing the genetic sampling stations between Egersund and Stavanger (see Table S1). Our circulation model system consisted of three high-resolution fjord models run in parallel with input along the open boundaries from the Norwegian coastal model NorKyst800 (Asplin et al., 2020). The highest

resolution models all applied 160m resolution in the horizontal, and similar experiments and validation analysis are presented in e.g. Dalsøren et al., 2020. We applied the open-source Regional Ocean Modeling System (ROMS) (Haidvogel et al., 2008; Shchepetkin & McWilliams, 2005) (see also http://myroms.org), which is a state-of-the-art, three-dimensional, free-surface, hydrostatic, primitive equation ocean model that uses generalized terrain-following s-coordinates in the vertical. The results from the three model grids using 160m resolution covering the inner Skagerrak, the southernmost part of the Norwegian coast and the Ryfylke/Rogaland fjords, respectively, were merged into one large grid before used as input in the particle-tracking model.

Our simulations of drift of larvae-particles were conducted by using the open access Lagrangian Advection and Diffusion Model (LADIM; https://github.com/bjornaa/ladim). The model solves the Lagrangian equation of motion using a 2nd-order Runge-Kutta scheme. Due to limited knowledge on larval swimming behaviour, we initialized the particles in 1, 2, 3 and 4m depth and left them without any vertical movement. 250 particles were released in each depth at all 204 locations which totally gives 1000 particles per station. To include slightly different physical environments, we performed three identical simulations initiating June 2nd, July 1st and August 1st, all in 2018 as a representative year, and followed every particle 25 days. A small horizontal random walk was also implemented to ensure variability between the particles and to represent small-scale turbulence.

284 **Results**

285 **Population genetic patters**

286 The final dataset consisted in 292 individuals from 17 main localities genotyped for a total of
287 611 SNPs, with no deviations from HWE or LD. Genetic variability (He) ranged from 0.139 for
288 SAN locality to 0.156 at MAN (Table 1) and displayed a significant increasing trend from east to
289 west (linear regression of He on distance over water from the easternmost locality, Hvaler:
290 slope = $1.285*10^{-5}$-; R^2 = 0.3731; *P* = 0.0055).

291 BayeScan identified three possible outlier loci out of the 611 SNPs analysed for the whole
292 dataset, while no outlier loci were identified when comparing West Coast vs. Skagerrak.
293 PCAdapt identified a total of 15 outlier loci. Following a consensus approach, two loci (SNP
294 9075_98 and 13326_52) were identified by both methods as outlier candidates under
295 selection. Pairwise F_{ST}, DAPC and IBD analysis were performed both with and without the two
296 candidate outlier loci and since no significant differences were observed in the results, all loci
297 were retained for all analyses (Figure S2).

298 Cluster analyses (DAPC) comprising the whole dataset clearly showed two main clusters
299 separated along DF1 (Figure 2a), one grouping together the West Coast samples and
300 Stavanger, and a second grouping all Skagerrak samples and Egersund. This result is supported
301 by the lowest BIC, which indicated that *K*=2 was the most likely number of clusters (Figure S3).
302 Using only Skagerrak and Egersund samples in a separated DAPC analyses (Figure 2b), no
303 further pattern stood out with samples showing no clear separation, and many locations
304 overlapping. This genetic pattern is also supported by F_{ST} analyses, where significant genetic
305 differentiation was found for overall sampling localities (F_{ST} overall =0.0181, *P*<0.001). Pairwise
306 F_{ST} values ranged from -0.0065 to 0.0657 for the 611 SNP considered, with the highest values
307 observed for the pairwise comparisons between West Coast localities and Skagerrak/Egersund,
308 which were highly significant (Table 2). On the contrary, pairwise comparisons showed no
309 significant differences among West Coast localities and Stavanger neither among Skagerrak

310 and Egersund (Table 2). Analysis of IBD showed that for the entire Norwegian coast, genetic
311 divergence ($F_{ST}/1-F_{ST}$) increased with increasing distance (Figure 3a; Mantel, P <0.001), while
312 no such relationship was observed within Skagerrak samples only (Figure 3a, Mantel, P>0.05).
313 This suggests that the divergence pattern found is mostly driven by the West Coast samples
314 and is illustrated with the sigmoid curve in Figure 3b, evidencing a sharp increase in F_{ST} around
315 the southern tip of Norway.

316 **Fine-scale genetic structure patterns**

317 Although pairwise comparisons among Skagerrak sampling localities were non-significant
318 (Table 2), and low levels of genetic divergence were found among the Skagerrak sampling
319 localities across all loci, the overall divergence was however highly significant (overall F_{ST}
320 =0.0040, P<0.001), suggesting the possibility of some cryptic genetic structure within
321 Skagerrak. In order to look for fine-scale patterns of genetic divergence within Skagerrak, this
322 dataset was analysed separately for IBD among individuals. Although the Skagerrak samples
323 did not show any patterns of IBD when using the whole dataset with the individuals grouped in
324 main sampled localities (Figure 3a, P> 0.05; Table 1), when separating the individuals according
325 to the level of exposure for each station, two different patterns of genetic differentiation
326 emerged. First, the overall level of genetic divergence among sheltered samples were about
327 three times higher (overall sheltered F_{ST} =0.0046) than those among exposed samples (overall
328 exposed F_{ST} =0.0015). Second, analysing IBD separately for sheltered and exposed stations, we
329 found that genetic divergence increases significantly with distance for shelter stations (Figure
330 4; P = 0.0018; slope = $3.446*10^{-5}$) but not for exposed stations (Figure 4; P = 0.3555; slope =
331 $2.298*10^{-6}$).

332 **Spatial decorrelation scale**

333 Station-specific fluctuations of black goby abundance in the Skagerrak were positively
334 correlated (i.e., synchronous) at nearby stations, with synchrony decreasing with increasing
335 distance between stations (Figure 5). The spatial decorrelation scale (v) considering both

exposed and sheltered stations was estimated as 27.9 km (90% CI = 13.8 - 94.8 km). When only sheltered stations were considered, the spatial decorrelation scale was 20 km (90% CI = 2.9 – 266 km), which was substantially smaller than the decorrelation scale among exposed stations of 81 km (90% CI = 15.1 – 254 km), although note that confidence intervals were overlapping.

Habitat associations

The best model for black goby abundance in the Skagerrak beach seine survey included effects of biogenic habitat type and cover. Black goby catches significantly varied by habitat type (Chisq = 44.2, df = 5, $P < 0.001$), with highest catches in eelgrass and mixed eelgrass/brown macroalgae habitats, and lowest catches in habitats lacking eelgrass, but containing brown macroalgae or green algae (Figure 6). Black goby catches were also significantly associated with habitat cover, with higher catches at sites with higher coverage percentage ($z = 8.42$, $P < 0.001$).

Oceanographic modelling

The oceanographic modelling shows that the larvae-particles released in the surface layer inshore and along the Skagerrak coast, from the Swedish border in the East to Egersund in the West, are subject to three different destinies (Figure 7). A substantial portion is mixed and spread widely by the NCC, which has a persistent westward flow out of the Skagerrak along the Norwegian coast. The NCC in the Skagerrak is relatively strong and often has a distinct core close to the coast. A large portion of the larvae-particles entering the NCC have a stronger offshore drift and were not able to settle along the coast at all. Additionally, a significant and important portion of the larvae hatched inside the fjords likely end up in a local settlement. The probability for onshore settlement West of Egersund from larvae-particles released East of Egersund is, however, very small. The localities around Stavanger provide the area with some larvae retention even though most of the larvae-drift from the localities West of Egersund is offshore.

361 **Discussion**

362 Using a multidisciplinary approach, we combined genetic, demographic and habitat data from
363 a long-term data series and an oceanographic particles drift model, to unravel the patterns of
364 genetic structure on different scales in the black goby along the Norwegian coast. While
365 geneflow and weak population structure is expected for marine species with high dispersal
366 ability, we observed restricted gene flow both at a surprisingly fine (kilometres) and large
367 (100km) scale. Our results show a pattern of isolation by distance along the Skagerrak coast
368 and a pronounced genetic discontinuity between populations along the Norwegian south and
369 west coasts, likely driven by both lack of habitat and oceanography.

370 Our sampling design focused on the Skagerrak coast, where time series for annual abundance
371 are available, giving the opportunity to investigate fine-scale spatial population patterns both
372 at the genetic and demographic level. At a first glimpse, the general overall pattern showed
373 high levels of connectivity across Skagerrak for both genetic and demographic data. In the
374 Skagerrak, we found no genetic differentiation for pairwise comparisons among locations
375 (Table 2), no IBD, and no clusters were perceived in the DAPC analyses. This is in accordance
376 with the predictions of the oceanographic model where the Norwegian coastal current boosts
377 passive dispersion of larvae promoting connectivity. However, although the overall level of
378 genetic divergence among Skagerrak sample localities was low, it was statistically significant
379 ($SK_{overall}$ F_{ST} =0.0040, P<0.001), suggesting possible hidden genetic structure. Previous studies
380 also found low levels of genetic structuring along the Skagerrak coast in other species (Knutsen
381 et al. 2013; 2014; Blanco Gonzalez et al., 2016). The dense, fine-scale sampling design allowed
382 us enough spatial resolution to examine samples at the individual level and compare stations
383 that are largely sheltered from the coastal current (i.e., located inside small fjords or bays, or
384 behind skerries) and more exposed areas. On the supposition that ocean currents transport
385 pelagic larvae along the coast, we expect that exchange of larvae will be higher among the
386 more exposed localities and, further, if this current-driven larval dispersal is the dominant
387 mechanisms for gene flow in the species, that gene flow should also be higher among exposed

388 localities than among more sheltered ones. These expectations were indeed met, as the
389 genetic IBD analyses yielded a steeper slope for sheltered than for exposed localities (cf. Figure
390 4), indicating less gene flow among sheltered localities.

391 The observed spatial correlation in abundance, or population synchrony, corroborates the
392 genetic findings. First, nearby stations only a few km apart displayed higher correlations in
393 abundance than stations situated farther apart on the coast (cf. Figure 5). This finding gives
394 evidence that nearby populations are indeed connected demographically, as expected if they
395 also exchange genes. Second, the sheltered stations displayed a more rapid decline in
396 correlation with distance, i.e. a shorter "decorrelation" scale, than did the exposed areas (20
397 and 81 km, respectively), indicating that dispersal is more restricted among sheltered areas.
398 This finding is also in agreement with genetic data, suggesting a higher level of connectivity
399 among exposed stations than sheltered ones. Unfortunately, our particle drift model did not
400 have enough resolution at such small spatial scales to test the finding at the demographic and
401 genetic level. However, in the light of the results obtained it is feasible that exposed stations
402 are more affected by the Norwegian coastal current promoting higher rates of dispersal
403 translating in a higher level of population homogenization, while sheltered areas likely are less
404 influenced by the NCC and experience localized retention processes (i.e. Ciannelli et al., 2010).
405 This correspondence between genetic and demographic patterns is relevant for the long-
406 standing discussion about the use of genetic data for inferring connectivity among natural
407 populations (Lowe & Allendorf 2010; Knutsen et al., 2011), and yields support to the notion
408 that genetic markers provide information of ecological as well as evolutionary relevance.

409 At a larger scale, when considering the entire study area, we found a strong genetic
410 discontinuity around the southern tip of Norway. We identified two distinct genetic clusters,
411 one comprising the west coast samples and the other comprising all samples from Egersund
412 and the Skagerrak coast, with increased levels of pairwise divergence (F_{ST} =0.0400) between
413 West coast and Skagerrak samples. Lower levels of genetic diversity were found within the
414 Skagerrak samples, suggesting restricted geneflow between these two populations. Genetic

discontinuities, or “genetic breaks,” are characterized by rapid increase in genetic divergence over shorth geographic distances. This particular “genetic break”, around the southern tip of Norway, has been previously identified for other fish species such as corkwing wrasse (*Symphodus melops*; Knutsen et al., 2013; Blanco Gonzalez et al., 2016; Mattingsdal et al., 2020) and ballan wrasse (*Labrus bergylta*; Seljestad et al., 2020). It has been shown that this area is characterized by a vast and bare stretch of sandy beaches (Jæren, about 26 km; Blanco Gonzalez et al., 2016), between Egersund and Stavanger, resulting in a habitat discontinuity for some species. For wrasse species, hard bottom substrate is an essential habitat, required to build nests where females lay their eggs (Darwall et al., 1992; Villegas-Ríos et al., 2013). Furthermore, telemetry studies showed that wrasses have small home ranges and individual movements are likely to be restricted to less that 300m (Halvorsen et al., 2021), suggesting that individual migration across Jæren is unlikely. For the black goby, not much is known about individual movements, although they share similar life history traits with wrasses: adults black gobies are mostly sedentary, and males are also territorial, building and guarding nests where females lay their sticky eggs. Therefore, for both, black goby and wrasses, dispersal is assumed to be mostly promoted by the pelagic larvae (Knutsen et al., 2013; Kara & Quignard, 2019). Regarding habitat, the black goby has been described in the literature as preferring soft bottoms and unvegetated areas (Perry et al., 2018), and with strong habitat selection shifts based on the presence of predators (Kruschel & Schultz, 2011). However, although black gobies are occasionally captured in bare substrates, our habitat analysis clearly showed that the black goby is more abundant in areas with higher vegetation coverage and has a preference for eelgrass and mixed eelgrass/brown macroalgae habitats (see Figure 6). This suggests that habitat discontinuity may also be a good explanation for the lack of gene flow across the bare stretches of sandy beaches at Jæren for the black goby population, similarly to what has been found for wrasse species (Blanco Gonzalez et al., 2016; Seljestad et al., 2020). However, habitat discontinuity alone does not explain how a species with a long pelagic larval stage (about 28 days) is unable to drift over a few dozens of kilometres of the stretch of sandy

beaches. Our oceanographic model suggests that particles have a dispersal potential of several kilometres, and therefore larval transport should be high enough to overcome the physical barrier of lack of habitat across Jæren. The results obtained with the oceanographic drift model offers a clear picture of the drivers promoting the genetic break and restricted geneflow, by showing that particles that are released nearby the southern tip of Norway drift offshore towards deep water, where recruitment of juveniles is likely not possible. This patter is driven by the Norwegian coastal current, which flow westwards along the Skagerrak coast (Albretsen et al., 2012), but after the southern tip of Norway the current takes its course drifting away from shore. The oceanographic particle drift model corroborates this. The model also shows that particles released close to Stavanger and Jæren will continue westwards and will not mix with Skagerrak, prevented genetic connectivity and likely strengthening the barrier effect. The model shows that few released from western end of Skagerrak will reach Egersund area and this is likely sufficient to homogenising (probably together with some individual migration) the gene pool and making Egersund more genetically similar to Skagerrak samples.

Our multidisciplinary approach and fine-scale sampling design allows for detecting the importance of retention of early life history forms in shaping the genetic structure at finer geographic scale in coastal waters. Further, oceanographic circulation also seems to play an important role in the creation of the genetic divergence observed between West coast and Skagerrak samples. Therefore, our results using these combined approaches, indicated a more complex scenario of gene-flow restriction suggesting a combination of both habitat discontinuity but also an important oceanographic component.

References

Albretsen J, Aure J, Sætre R, Danielssen DS (2012) Climatic variability in the Skagerrak and coastal waters of Norway, ICES Journal of Marine Science, 69, 758-763.

Almany GR, Connolly SR, Heath DD, Hogan JD, Jones GP, McCook LJ, Mills M, Pressey RL, Williamson DH (2009) Connectivity, biodiversity conservation, and the design of marine reserve networks for coral reefs. Coral Reefs, 28, pages339–351.

Amante C, Eakins BW (2009) ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24, 19 pp.

Asplin L, Albretsen J, Johnsen IA, Sandvik AD (2020) The hydrodynamic foundation for salmon lice dispersion modeling along the Norwegian coast, Ocean Dynamics, 70, 1151–1167.

Barceló C, Ciannelli L, Olsen EM, Johannessen T, Knutsen H (2016) Eight decades of sampling reveal a contemporary novel fish assemblage in coastal nursery habitats. Global Change Biology, 22, 1155– 1167.

Barth JMI, Villegas-Ríos D, Freitas C, Moland E, Bastiaan S., André C, Knutsen H, Bradbury I, Dierking JP, Christoph R, Metcalfe D, Jakobsen J, Olsen KS, Olsen EM, Jentoft S (2019) Disentangling structural genomic and behavioral barriers in a sea of connectivity. Molecular Ecology, 28, 1394-1411.

Bates D, Maechler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67, 1-48.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B (Methodological), 57, 289–300.

Bernatchez L, Wellenreuther M, Araneda C, Ashton DT, Barth JMI, Beacham TD, et al (2017) Harnessing the power of genomics to secure the future of seafood. Trends in Ecology & Evolution, 32(9), 665–680.

Bjørnstad ON, Ims RA, Lambin X (1999) Spatial population dynamics: Analyzing patterns and processes of population synchrony. Trends in Ecology & Evolution, 14, 427–432.

Blanco Gonzalez E, Knutsen H, Jorde PE (2016) Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. Plos One 11, e0163052.

Carpenter KE, Smith-Vaniz WF, de Bruyne G, de Morais L (2015) Gobius niger. The IUCN Red List of Threatened Species 2015: e.T198570A21913403. https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T198570A21913403.en. Downloaded on 16 August 2021.

Cattell R B (1966) The Scree test for the number of factors. Multivariate Behavioral Research, 1(2), 245–276.

Catarino D, Knutsen H, Verissimo A, Olsen EM, Jorde PE, Menezes G, et al (2015) The Pillars of Hercules as a bathymetric barrier to gene flow promoting isolation in a global deep-sea shark (*Centroscymnus coelolepis*). Molecular Ecology, 24, 6061 - 6079.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology, 22,3124–3140.

Ciannelli L, Knutsen H, Moland E, Espeland SH, Asplin L,Jelmert A, Knutsen JA, 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, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. Science, 311, 522– 527.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST (2011) The Variant Call Format and VCFtools. Bioinformatics, 2715, 2156–2158.

Dalsøren S, Albretsen J, Asplin L (2020) New validation method for hydrodynamic fjord models applied in the Hardangerfjord, Norway. Estuarine Coastal and Shelf Science, 246, 107028.

Darwall WRT, Costello MJ, Donnelly R, Lysaght S (1992) Implication of life-history strategies for a new wrasse fishery. Journal of Fish Biology, 41: 111–123.

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.

Froese and Pauly, (2021) Editors. 2021. FishBase. World Wide Web electronic publication. www.fishbase.org, version (06/2021).

Haidvogel DB, Arango H, Budgell WP, Cornuelle BD, Curchitser E, Di Lorenzo E, et al. (2008) Ocean forecasting in terrain-following coordinates: formulation and skill assessment of the Regional Ocean modelling system. Journal of Computational Physics, 227, 3595–3624.

Halvorsen KT, Larsen T, Browman HI, Durif C, Aasen N, Vøllestad LA, et al (2021) Movement patterns of temperate wrasses (Labridae) within a small Marine Protected area. Journal of Fish Biology, 99, 1513-1518.

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

Hop H, Gjøsæter J, Danielssen DS (1992) Seasonal feeding ecology of cod (*Gadus morhua* L.) on the Norwegian Skagerrak coast. ICES Journal of Marine Sciences, 49, 453–461.

Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics, 27(21), 3070–3071.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics, 11(1), 94.

Kara MH, Quignard JP (2019) Fishes in lagoons and estuaries in the Mediterranean 2. Sedentary fish. ISTE-Wiley Ed. 418 pp.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) DiveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution, 4, 782–788.

Kelley JL, Brown AP, Therkildsen NO, Foote AD (2016) The life aquatic: advances in marine vertebrate genomics. Nature Reviews Genetics, 17, 523–534.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics, 49, 561-576.

Knutsen JA, Knutsen H, Gjøsæter J, Jonsson B (2001) Food of anadromous brown trout at sea. Journal of Fish Biology, 59, 533– 543.

Knutsen H, Olsen EM, Jorde PE, Espeland SH, André C, Stenseth NC (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.

Knutsen H, Jorde PE, Blanco Gonzalez E, Robalo J, Albretsen J, Almada V (2013) Climate Change and Genetic Structure of Leading Edge and Rear End Populations in a Northwards Shifting Marine Fish Species, the Corkwing Wrasse (*Symphodus melops*). Plos One, 8, e67492.

Knutsen H, Jorde PE, Blanco Gonzalez E, Eigaard OR, Pereyra RT, Sannæs H, Dahl M, André C, Søvik G (2014) Does population genetic structure support present management regulations of the northern shrimp (*Pandalus borealis*) in Skagerrak and the North Sea? ICES Journal of Marine Sciences, 72, 863–871.

Kruschel C, Schultz ST (2011) Juvenile *Gobius niger* avoids seagrass in the presence and uncertain absence of seagrass-inhabiting predators. Journal of Experimental Marine Biology and Ecology, 409, 240–246.

Longmore C, Trueman CN, Neat F et al. (2014) Oceanic scale connectivity and life cycle reconstruction in a deep-sea fish. Canadian Journal of Fisheries and Aquatic Sciences, 71, 1312–1323.

Lowe WH, Allendorf F (2010) What can genetics tell us about population connectivity? Molecular Ecology, 19, 3038-3051.

Luu K, Bazin E, Blum MGB (2017) pcadapt: An R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources, 17(1), 67–77.

Mattingsdal M, Jorde PE, Knutsen H, Jentoft S, Stenseth NC, Sodeland M, Robalo JI, Hansen MM, André C, Blanco Gonzalez E (2020) Demographic history has shaped the strongly differentiated corkwing wrasse populations in Northern Europe. Molecular Ecology, 29, 160-171.

Mertens LE, Treml EA, Von der Heyden S (2018) Genetic and biophysical models help define marine conservation focus areas. Frontiers in Marine Science, 6(5), 268.

Miller PJ (1986) Gobiidae, Fishes of the North-eastern Atlantic and Mediterranean, vol. 3, Paris, UNESCO.

Moen FE, Svensen E (2000) Dyreliv i havet – Håndbok i norsk marin fauna. 2nd edition. Kom Forlag. Kristiansund, Norway. 576pp.

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. Molecular Ecology Resources, 11, 184–194.

579 Palumbi SR (2004) Marine reserves and ocean neighborhoods: the spatial scale of marine
580 populations and their management. Annual Review of Environment and Resources, 29, 31-68.

581 Pante E, Simon-Bouhet B (2013) marmap: a package for importing, plotting and analyzing
582 bathymetric and topographic data in R. PLoS ONE, 8, e73051.

583 Penthon P (2005) Aschehougs store Fiskebok. 2nd edition. H. Aschehoug & Co, Kristiansand,
584 Norway, 447pp.

585 Perry D, Staveley TAB, Gullström M (2018) Habitat Connectivity of Fish in Temperate Shallow-
586 Water Seascapes. Frontiers in Marine Science, 4, 440.

587 Peterson BK, Weber JN, Kay EH, Fisher HS Hoekstra HE (2012) Double Digest RADseq: An
588 Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non Model
589 Species. Plos One, 7, doi: 10.1371/journal.pone.0037135

590 Planes S (1998) Genetic diversity and dispersal capabilities in marine fish, in M.K. HECHT et al.
591 (eds), Evolutionary Biology, Plenum Press, New York.

592 Privé F, Luu K, Vilhjálmsson BJ, Blum MG (2020) Performing highly efficient genome scans for
593 local adaptation with R package pcadapt version 4. Molecular Biology and Evolution, 37, 2153–
594 2154.

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

597 R Core Team (2021) R: A language and environment for statistical computing. R Foundation for
598 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

601 Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under
602 isolation by distance. Genetics, 145, 1219-1228.

603 Rousset F (2000) Genetic differentiation between individuals. Journal of Evolutionary Biology,
604 13, 58-62.

605 Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for
606 Windows and Linux. Molecular Ecology Resources, 8, 103–106.

607 Schwarz G (1978) Estimating the dimension of a model. Annals of Statistics, 6, 461–464.

608 Seljestad GW, Quintela M, Faust E, Halvorsen KT, Besnier F, Jansson E, Dahle G, Knutsen H,
609 André C, Folkvord A, Glover KA (2020) "A cleaner break": Genetic divergence between
610 geographic groups and sympatric phenotypes revealed in ballan wrasse (*Labrus bergylta*).
611 Ecology and Evolution, 10, 6120-6135.

612 Selkoe KA, D'Aloia CC, Crandall ED, Iacchei M, Liggins L, Puritz JB, et al (2016) A decade of
613 seascape genetics: Contributions to basic and applied marine connectivity. Marine Ecology
614 Progress Series, 554, 1– 19.

Shchepetkin AF, McWilliams JC (2005) The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Modelling, 9, 347–404.

Slatkin M (1994) Gene Flow and Population Structure. In: Real, L., Ed., Ecological Genetics, Princeton University Press, Princeton, 3-17.

Stenseth NC, Bjørnstad ON, Falck W, Fromentin JM, Gjøsæter J, Gray J (1999) Dynamics of coastal cod populations: intra- and intercohort density dependence and stochastic processes. Proceedings of the Biological Society, 266, 1645–1654.

Vaas KF, Vlasblom AG, De Koeijer P (1975) Studies on the Black Goby (*Gobius niger*, Gobiidae, Pisces) in the Veerse Meer, S.W. Netherlands. Nertherlands Journal of Sea Research, 9 (1), 56–68.

Vavrek MJ (2011) Fossil: palaeoecological and palaeogeographical analysis tools. Palaeontologia Electronica, 14, 1T

Villegas-Ríos D, Alós J, March D, Palmer M, Mucientes G, Saborido-Rey F (2013) Home range and diel behavior of the ballan wrasse, *Labrus bergylta*, determined by acoustic telemetry. Journal of Sea Research, 80, 61–71.

Waples R (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species, Journal of Heredity, 89, 438-450.

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

Acknowledgments

This study was funded by the ECOGENOME project (grant nr. grant number: 280453), through the Norwegian Research Council. Further funding was provided by MarGen II, an Intereg project under the Øresund-Kattegat-Skagerrak program, by the Ministry of Trade, Industry and Fisheries, and from University of Agder, CCR. We thank to the scientific team participating in the Beach seine survey. We also thank to Ricardo Medeiros (IMAR, University of the Azores) for making the map and to Karen Martinez-Swatson (IMR) for helping with the laboratory methodology description. DC was funded by the University of Agder and at later stage by FCT – Foundation for Science and Technology, I.P., under the project UIDP/05634/2020 granted to Okeanos. Okeanos is funded by through the FCT, I.P., under the project UIDB/05634/2020 and UIDP/05634/2020 and through the Regional Government of the Azores through the initiative to support the Research Centers of the University of the Azores and through the project M1.1.A/REEQ.CIENTÍFICO UI&D/2021/010.

Table 1. Collection sites and summary statistics for black goby samples. Geographic area in the Norwegian coast (WC, West Coast; SK, Skagerrak), sampling site name and code, coordinates in decimal degrees (the coordinated showed is an average value when more than one station was sampled). N1 refers to the total number of individuals initially genotyped, combined over sampling stations; N2 is the number of black gobies remaining after filtering and used for downstream analyses (292 in total); Stations referrer to the total number of stations used within a sampling site; expected (He) and observed (Ho) heterozygosity, and inbreeding coefficient (F_{is}) averaged over 611 SNPs.

Area	Sampling site	Site code	Latitude (N)	Longitude (E)	N1	N2	Stations	H_E	H_O	F_{IS}
WC	**Ålesund**	**ALE**	62,338	6,581	23	19	1	0.155	0.157	0.017
	Manger	**MAN**	60,664	5,095	21	18	2	0.156	0.156	0.020
	Stavanger	**STA**	59,114	5,727	24	20	1	0.152	0.149	0.047
	Egersund	**EGE**	58,421	6,014	23	18	1	0.148	0.144	0.052
SK	**Torvefjorden**	**TOR**	58,064	7,764	16	14	4	0.152	0.152	0.009
	Topdalsfjor-den	**TOP**	58,144	8,050	24	21	2	0.155	0.155	0.024
	Lillesand	**LIL**	58,206	8,308	24	21	5	0.146	0.148	0.004
	Grimstad	**GRM**	58,283	8,509	17	9	2	0.145	0.150	-0.036
	Tvedestrand	**TVE**	58,623	9,058	24	17	3	0.145	0.142	0.042
	Risør	**RIS**	58,737	9,314	24	11	4	0.141	0.150	-0.038
	Kragerø	**KRA**	58,835	9,319	24	18	4	0.143	0.139	0.046
	Langesund	**LAN**	59,030	9,808	24	17	3	0.147	0.145	0.038
	Sandefjord	**SAN**	59,088	10,227	13	12	2	0.139	0.137	0.023
	Vrengen	**VRE**	59,165	10,377	25	20	3	0.153	0.146	0.056
	Håholmen	**HAH**	59,139	10,434	24	20	1	0.144	0.138	0.054
	Oslofjord	**OSF**	59,766	10,712	24	18	1	0.147	0.143	0.036
	Hvaler	**HVA**	59,264	10,731	25	19	3	0.147	0.148	-0.010

Table 2. Pairwise F_{ST} with significant pairwise comparisons in bold and level of significance after correction denoted by asterisk: * p<0.05; ** p<0.01; *** p<0.001. See Table 1 for sample abbreviations.

	ALE	MAN	STA	EGE	TOR	TOP	LIL	GRM	TVE	RIS	KRA	LAN	SAN	VRE	HAH	OSF
MAN	0.0041															
STA	0.0131	0.0116														
EGE	**0.04229*** **	**0.04596*** **	0.0234													
TOR	**0.04447*** **	**0.04364*** **	0.0267	-0.0019												
TOP	**0.03818*** **	**0.04046*** **	**0.03305*** **	0.0036	-0.0003											
LIL	**0.0430*** **	**0.05144*** **	**0.04781*** **	0.0085	0.0122	0.0051										
GRM	**0.06464*** **	**0.06473*** **	0.0361	-0.0014	0.0000	0.0115	0.0235									
TVE	**0.04023*** **	**0.04606*** **	**0.03695*** **	0.0011	0.0082	0.0102	0.0041	0.0079								
RIS	**0.03769** **	**0.04602*** **	0.0377	0.0035	0.0045	0.0010	-0.0008	0.0107	-0.0040							
KRA	**0.04836*** **	**0.05177*** **	**0.03289*** **	0.0045	0.0049	0.0072	0.0018	0.0058	-0.0017	-0.0064						
LAN	**0.04338*** **	**0.04557*** **	0.0260	0.0032	0.0025	-0.0010	0.0091	0.0009	0.0045	0.0025	-0.0013					
SAN	**0.05817*** **	**0.0657*** **	**0.04529*** **	0.0027	0.0007	-0.0036	0.0113	0.0065	0.0137	0.0122	0.0080	0.0028				
VRE	**0.04393*** **	**0.0502*** **	**0.03388*** **	-0.0031	0.0021	-0.0025	0.0110	-0.0029	0.0084	0.0028	0.0021	0.0006	-0.0006			
HAH	**0.05965*** **	**0.06464*** **	**0.03829*** **	0.0018	0.0000	0.0085	0.0193	-0.0065	0.0101	0.0051	0.0069	0.0000	0.0031	0.0002		
OSF	**0.04801*** **	**0.05363*** **	**0.04173*** **	-0.0006	0.0035	-0.0034	0.0041	0.0081	0.0054	0.0011	0.0051	0.0005	-0.0025	-0.0015	0.0060	
HVA	**0.04554*** **	**0.05165*** **	**0.03563*** **	0.0038	0.0071	-0.0012	0.0089	0.0099	0.0041	0.0044	0.0031	0.0041	-0.0012	-0.0021	0.0063	0.0006

Figure 1. Map showing the main sampling sites where black goby samples were collected for the genetic analyses. Dashed lines delimitate the Skagerrak region between Norway, Denmark and Sweden. Sampled sites abbreviations are given in Table 1.

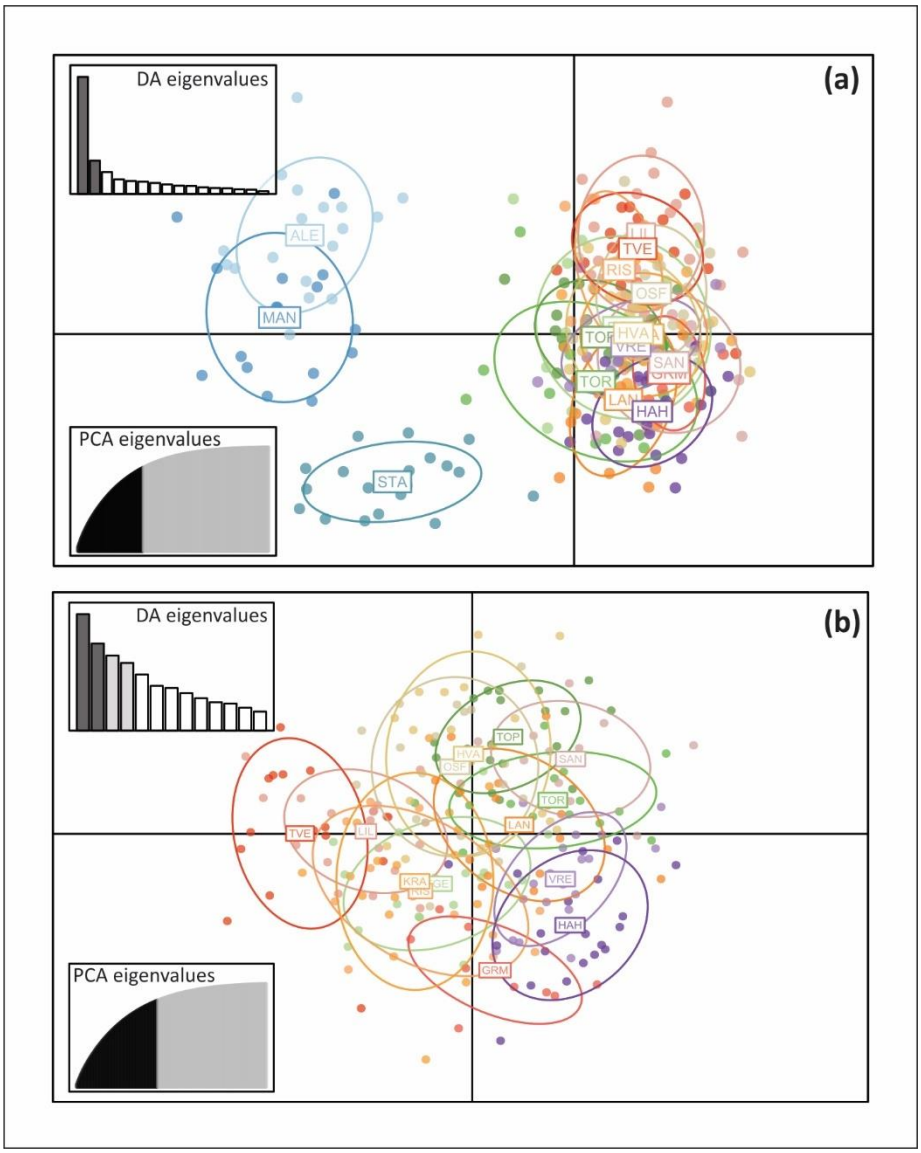
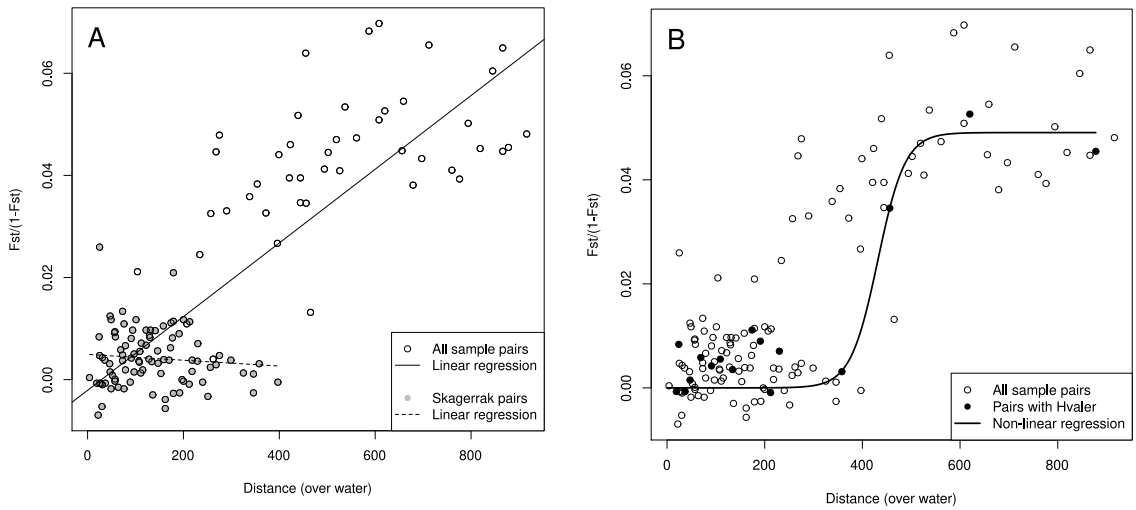


Figure 2. Results from the discriminant analysis of principal components (DAPC) showing (a) two clusters, one corresponding to West coast localities and other to Skagerrak+ Egersund samples; and showing (b) no distinct cluster pattern within Skagerrak +Egersund only. The DAPC was conducted based on 100 PCs that conserved more than 80% variation.



674

675 **Figure 3.** Isolation by distance (IBD) for the black goby. Effect of geographic distance (km over
676 water) on linearized observed genetic divergence, $F_{ST}/(1-F_{ST})$, (a) among all black goby sampling
677 localities (slope = $7.2*10^{-5}$, $P<0.001$) and for Skagerrak localities only (slope= $3.8*10^{-6}$, $P>0.05$);
678 (b) among all pairwise comparisons with Hvaler in bold (eastern most sample) and by fitting a
679 sigmoid curve to the data points including Hvaler highlighting the sharp increase in F_{ST} from the
680 east to the West coast of Norway.

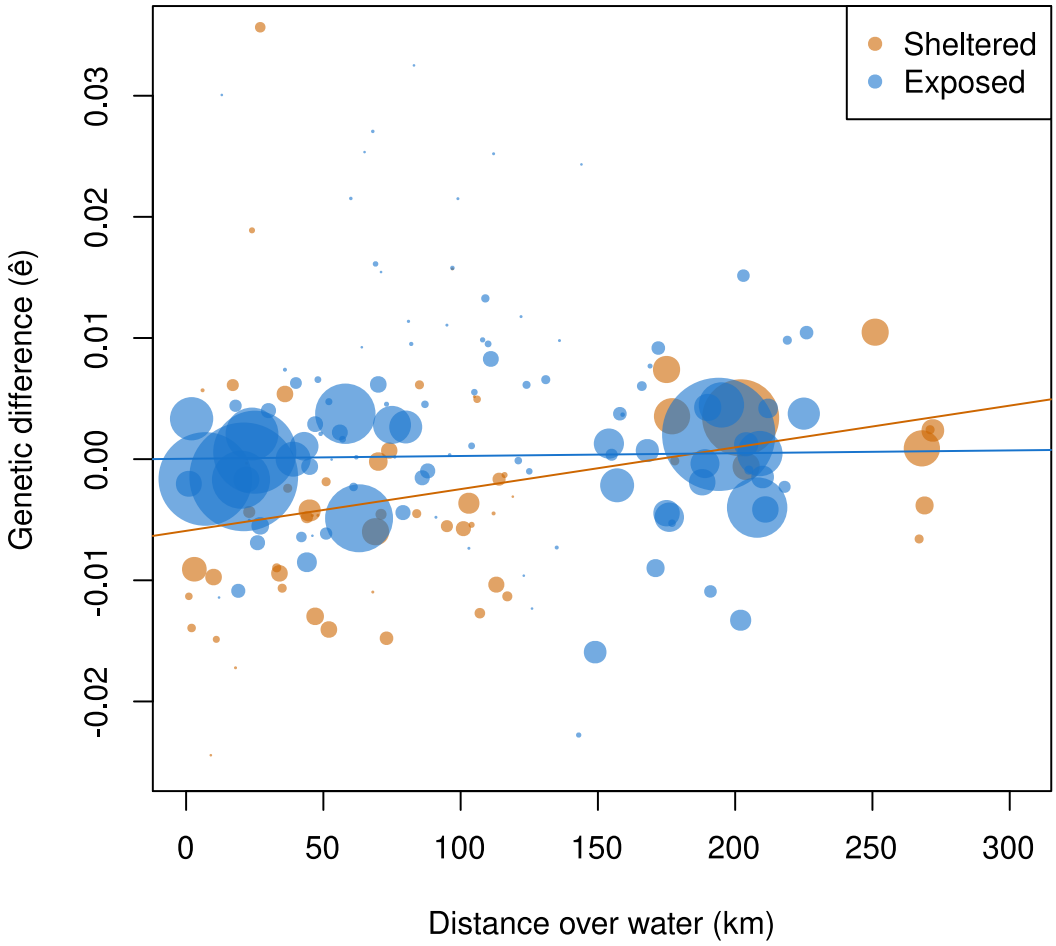


Figure 4. Isolation by distance (IBD) for the black goby sheltered (orange) and exposed (blue) sampling stations in Skagerrak. Sampling stations were binned into distance classes of 1 km and size of the dots were scaled to number of individual pairs within each distance class. Effect of linear distance on the genetic difference (ê) calculated at individuals' level, showing a significant increase with distance for sheltered stations (P = 0.0018, slope = $3.446*10^{-5}$) and non-significate trend for exposed stations (P = 0.3555, slope = $2.298*10^{-6}$).

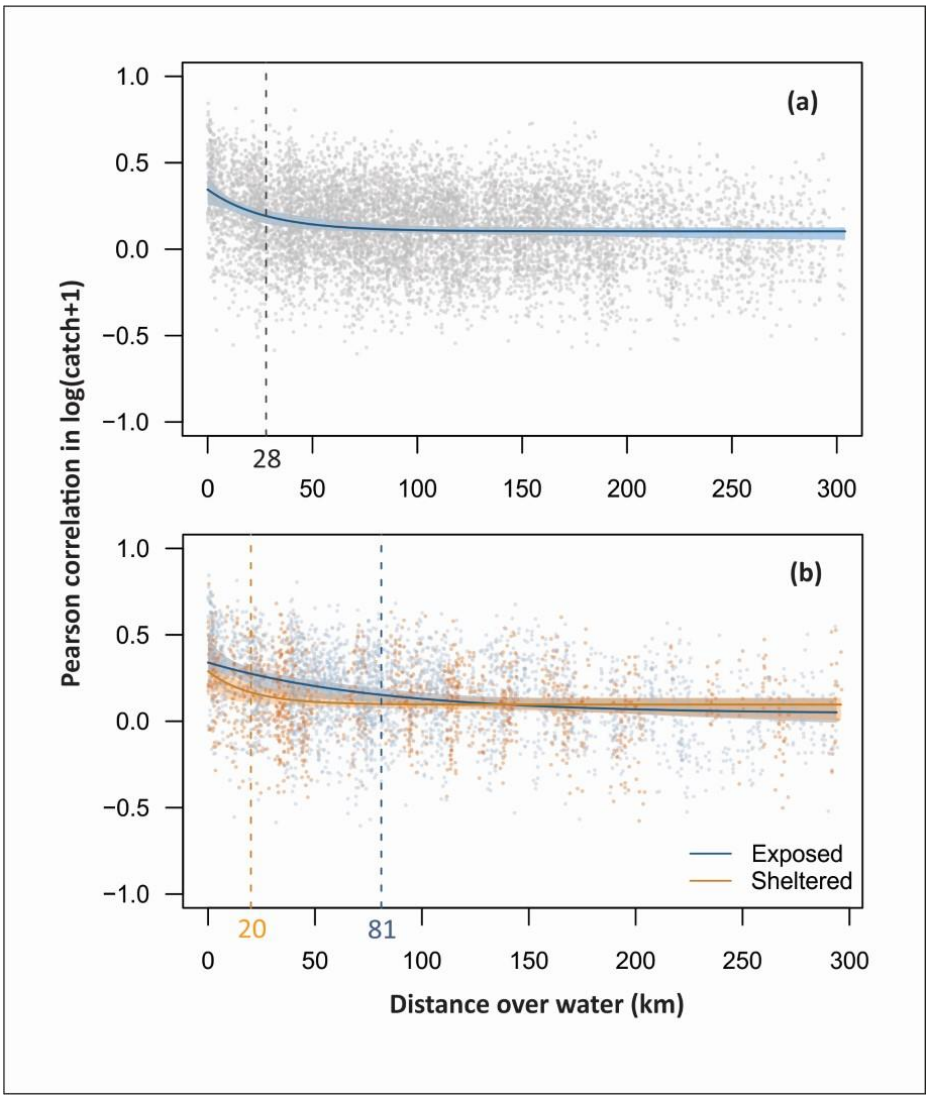


Figure 5. Spatial decorrelation in abundance fluctuations of black goby at beach seine stations along the Skagerrak coast. Results are shown for (a) all stations and (b) models fit to sheltered (orange) and exposed (blue) stations separately. Points show pairwise correlations among stations, and solid lines show a fitted exponential decay model along with bootstrapped 90% confidence intervals (shaded areas). The estimated decorrelation scale (v) is shown as a vertical dashed line.

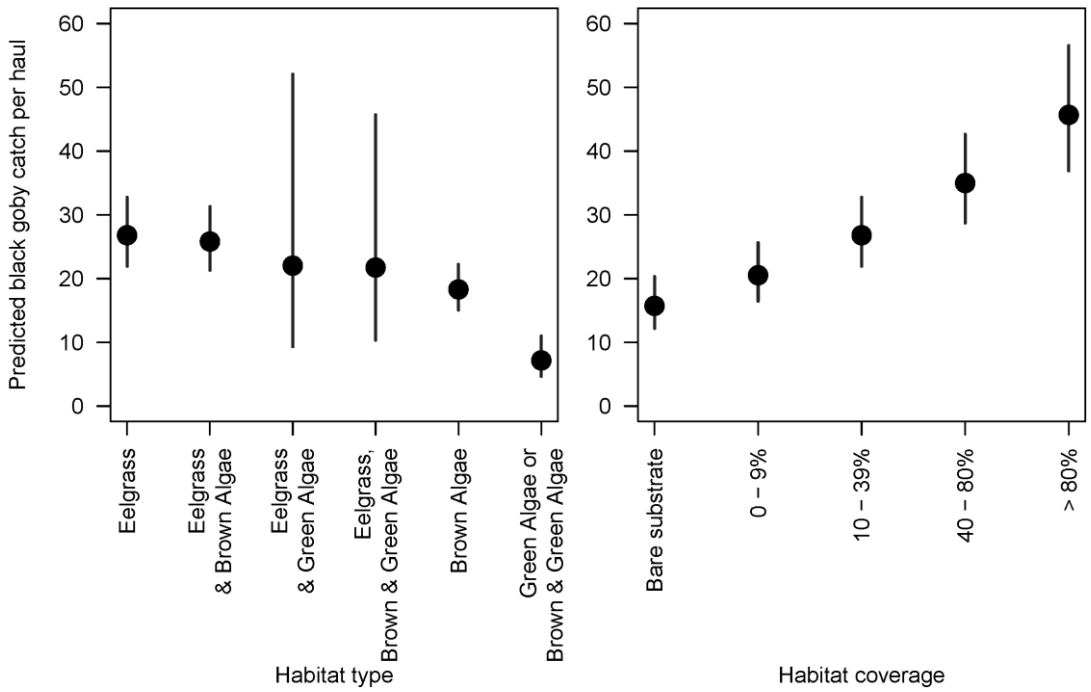


Figure 6. Predicted effects of biogenic habitat type and percent coverage on catches of black goby in the Skagerrak beach seine survey (with 95% prediction intervals). Predictions are shown for intermediate coverage (10-39%) in the first panel, and for eelgrass habitat in the second panel.

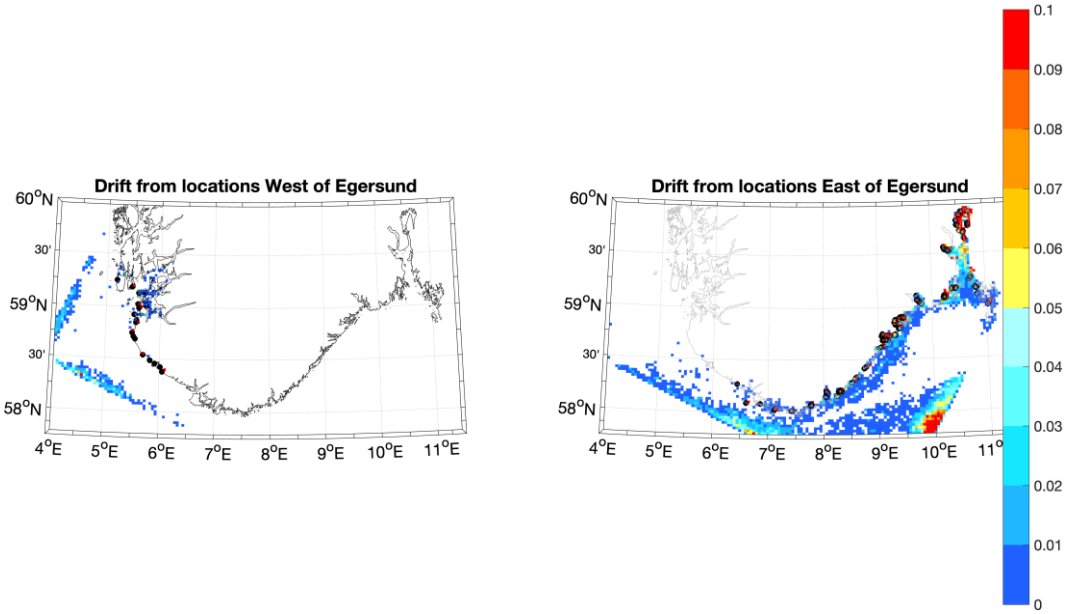


Figure 7. Maps from the oceanographic drift model showing the density of particles (colours) after 25 days drift from the locations West (left panel) and East of Egersund (right panel). The locations where particles are released are denoted with black circles. The colour scale denotes probability in % relative to the total number of particles released from all locations East and West of Egersund. The broad, zones of particle densities offshore are linked to the open boundary of the high-resolution fjord models that inhibits further drift offshore.