GBE

Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (*Gadus morhua* L.)

Paul R. Berg[1], Sissel Jentoft[1], Bastiaan Star[1], Kristoffer H. Ring[2], Halvor Knutsen[1,3,4], Sigbjørn Lien[5], Kjetill S. Jakobsen[1,*] and Carl André[6]

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

[2]Centre for Development and the Environment (SUM), University of Oslo, Norway

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

[4]University of Agder, Kristiansand, Norway

[5]Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Aas, Norway

[6]Department of Biology and Environmental Sciences-Tjärnö, University of Gothenburg, Strömstad, Sweden

*Corresponding author: E-mail: k.s.jakobsen@ibv.uio.no.

Accepted: May 17, 2015

Data deposition: All new SNPs in this article have been deposited at dbSNP (ss1712295757-ss1712305448 and ss1713750232-ss1713751191) and are described by their respective ss# throughout the article.

Abstract

How genomic selection enables species to adapt to divergent environments is a fundamental question in ecology and evolution. We investigated the genomic signatures of local adaptation in Atlantic cod (*Gadus morhua* L.) along a natural salinity gradient, ranging from 35‰ in the North Sea to 7‰ within the Baltic Sea. By utilizing a 12 K SNPchip, we simultaneously assessed neutral and adaptive genetic divergence across the Atlantic cod genome. Combining outlier analyses with a landscape genomic approach, we identified a set of directionally selected loci that are strongly correlated with habitat differences in salinity, oxygen, and temperature. Our results show that discrete regions within the Atlantic cod genome are subject to directional selection and associated with adaptation to the local environmental conditions in the Baltic- and the North Sea, indicating divergence hitchhiking and the presence of genomic islands of divergence. We report a suite of outlier single nucleotide polymorphisms within or closely located to genes associated with osmoregulation, as well as genes known to play important roles in the hydration and development of oocytes. These genes are likely to have key functions within a general osmoregulatory framework and are important for the survival of eggs and larvae, contributing to the buildup of reproductive isolation between the low-salinity adapted Baltic cod and the adjacent cod populations. Hence, our data suggest that adaptive responses to the environmental conditions in the Baltic Sea may contribute to a strong and effective reproductive barrier, and that Baltic cod can be viewed as an example of ongoing speciation.

Key words: Atlantic cod, Baltic Sea, ecological divergence, genomic adaptation, population genomics, SNPs, speciation.

Introduction

Teleost fishes occupy a range of aquatic habitats from freshwater environments to extreme marine environments with salinities well above 40‰. In order to spawn in marine habitats, teleosts have eggs adapted to a hyper-osmotic environment, resulting in less permeable and highly hydrated eggs that compensate for water outflow until osmoregulatory organs develop (Fyhn et al. 1999; Finn and Kristoffersen 2007). After the egg stage, osmoregulation primarily takes place in the intestines (Whittamore 2011; Laverty and Skadhauge 2012) and through salt excretion from specialized ionocyte cells in the gill epithelium (Evans 2008; Hiroi and McCormick 2012). Adaptation to specific osmotic conditions thus requires a wide range of molecular and physiological modifications in order to maintain water homeostasis both at the egg stage and in adult fish. Several studies have demonstrated population-specific acclimatization to different salinity environments through differential gene expression (see, e.g., Boutet et al. 2006; Larsen et al. 2011; Papakostas et al. 2014). Although many genes involved in the physiological response to variation in salinity have been identified, the genetic basis of broad salinity tolerance remains unclear (Tine et al. 2014). In addition to

the characteristic physical environment (salinity, temperature, and oxygen), Baltic cod also experiences a different parasite load compared with Kattegat and North Sea cod populations (Mellergaard and Lang 1999; Perdiguero-Alonso et al. 2008). However, the role of parasites in promoting genetic divergence has mixed support (Karvonen and Seehausen 2012).

The Baltic Sea is one of the world's largest semienclosed brackish seas, originating approximately 8,000 years ago (Zillén et al. 2008). Since then, the Baltic Sea has been colonized by both freshwater and marine teleosts (Ojaveer and Kalejs 2005). Limited water exchange and almost no tidal flow allow for stable salinity conditions, ranging from nearly fresh water in the northern Baltic Sea to around 30‰ at the border to the North Sea (fig. 1). Fossil records suggest that all contemporary marine teleost fish species have descended from a freshwater ancestry (Romer and Grove 1935; Griffith 1994; Long 1995; Fyhn et al. 1999), reflected by their hypo-osmotic state at the adult stage and their eggs and sperm in seawater (Griffith 1987; Fyhn et al. 1999; Finn and Kristoffersen 2007). The adaptation of neutral egg buoyancy toward those salinity levels found in the marine environment could arguably impede the successful colonization of less saline environments, as hyper-osmotic eggs would sink to the bottom, requiring an even higher degree of oocyte hydration, which is the case for Atlantic cod adapted to the Baltic environment (Nissling et al. 1994). Despite such considerations, some marine fishes with pelagic eggs, such as sprat, plaice and Atlantic cod, have colonized and successfully reproduce in low saline waters such as the Baltic Sea. Indeed, Atlantic cod colonized the central Baltic Sea some 8–6,000 YBP (Schmölcke 2006) when the surface salinity was 12–14‰. Since then, surface salinity has gradually decreased to around 7‰ today and around 14‰ at the spawning depth for Atlantic cod (50–100 m; Ignatius et al. 1981). Hence, the selection pressure to adapt to the low saline waters is a major force influencing the spawning success of cod in the Baltic Sea (Westin and Nissling 1991; Nissling et al. 1994), limited by factors such as egg buoyancy, sperm motility, and general osmoregulation. Importantly, Baltic cod spawns in a different season, compared with the North Sea and Western Baltic cod stocks, likely promoting reproductive isolation (Brander 1993; Wieland 2000; Brander 2005). The adaptation of Atlantic cod to low saline conditions is thus likely of relatively recent evolutionary origin, even though we cannot exclude the possibility that the founders of the Baltic population were already adapted to low salinity. This provides an excellent opportunity to study the genomic architecture behind salinity adaptation in a natural environment. As the ecological adaptation of Atlantic cod to a low-saline environment may contribute to reduced gene flow and thereby promote population divergence (cf. Nosil 2012), investigation of the genomic architecture of Baltic cod may give insights into ecological speciation in nature, and especially the genetic link between adaptation and reproductive isolation (cf. Seehausen et al. 2014).

Atlantic cod has high fecundity, pelagic eggs and larvae and large effective population sizes (Ward et al. 1994; Poulsen et al. 2006; Therkildsen et al. 2010), suggesting that selection, rather than genetic drift, is the main driver for genetic differentiation (Allendorf et al. 2010). Dispersal and gene flow could however prevent genetic structuring caused by local adaptation. Nevertheless, low levels of genetic structuring may not necessarily extend to genomic regions of functional importance (Weir et al. 2005; Nosil et al. 2009; Orsini et al. 2013) and highly differentiated genomic regions have been found in marine fish populations (Cano et al. 2008; Nielsen, Hemmer-Hansen, Larsen, et al. 2009; Nielsen, Hemmer-Hansen, Poulsen, et al. 2009; Hohenlohe et al. 2010; André et al. 2011; Shimada et al. 2011; Jones et al. 2012; Hemmer-Hansen et al. 2013). It has been shown in several cases that adaptive genetic variation may be selectively maintained, despite homogenizing effects of gene flow (see, e.g., Cano et al. 2008; Hauser and Carvalho 2008; Nielsen, Hemmer-Hansen, Larsen, et al. 2009; Bradbury et al. 2013; Defaveri, Jonsson, et al. 2013; Teacher et al. 2013; reviewed in Orsini et al. 2013). To identify genomic regions under selection and to get a better understanding of the evolutionary processes that generate such regions are important as the sheer number of genes and the number of potential traits evolving together are related to the size of the "islands of divergence" (see, e.g., Via 2009; Feder and Nosil 2010; Feder, Gejji, et al. 2012; Nosil and Feder 2012; Via et al. 2012; Flaxman et al. 2013).

So far, the genomic studies targeting the Baltic cod have used a relatively restricted number of genetic markers (e.g., Nielsen et al. 2003, 2012; Poulsen et al. 2006; Nielsen, Hemmer-Hansen, Poulsen, et al. 2009; Bradbury et al. 2010, 2013; Larsen et al. 2011; Hemmer-Hansen et al. 2013), limiting inference of the genomic architecture underlying salinity adaptation. Here, we investigate divergence patterns using 8,809 polymorphic single nucleotide polymorphisms (SNPs) distributed throughout the entire genome of Atlantic cod (based on an Illumina 12 K SNPchip described in The Cod SNP Consortium, in preparation), providing the first individually genotyped whole genome-wide approach to date on this species. We used different statistical outlier approaches on four different populations across the steep salinity gradient from the fully marine North Sea to the low saline Baltic Sea, identifying SNPs and genomic regions under selection. These genomic regions were also identified by a landscape genomic analysis, corroborating that certain genomic regions with underlying SNPs are influenced by defined environmental variables.

Materials and Methods

Sample Collection

Fin clips or spleen from adult Atlantic cod specimens ($n = 194$) were collected from seven localities and stored on ethanol or RNAlater (Qiagen). Sample sizes per location ranged from 8 to

Fig. 1.—Sampling locations of Atlantic cod specimens. Samples were obtained between 2002 and 2008 using trawl. Only mature specimens were selected for genetic analysis. See table 1 for a detailed description of the samples. Average surface salinity is denoted in ‰.

Table 1

Location and Details of the Atlantic Cod Samples Included in This Study, Combined with Basic Population Genetic Parameters

Population ID	Location	Sampling Time	Lat.	Long.	Spawning	Sample Size	Avg. Call Rate	# Polymorphic Loci	H_o (SD)	H_e (SD)
North Sea	North Sea	Mar 2002	N55.60	E05.85	Yes	42	0.983	8,659	0.362 (0.149)	0.361 (0.138)
Kattegat	Kattegat	Feb 2004	N56.90	E12.15	Yes	48	0.995	8,690	0.360 (0.145)	0.360 (0.135)
Öresund	Öresund	Feb 2003	N55.95	E12.70	Yes	48	0.995	8,715	0.363 (0.146)	0.363 (0.134)
Baltic	Baltic Sea					56	0.993	8,517	0.344 (0.155)	0.345 (0.147)
Baltic	Bornholm 05	June 2005	N55.50	E16.00	Yes	8	0.989	7,605	0.381 (0.188)	0.382 (0.135)
Baltic	Bornholm 04	June 2004	N55.59	E16.30	Yes	8	0.993	7,389	0.379 (0.192)	0.378 (0.137)
Baltic	Gotland	Nov 2008	N57.15	E18.78	No	20	0.994	8,267	0.359 (0.168)	0.356 (0.144)
Baltic	Öland	Nov 2008	N57.40	E17.00	No	20	0.994	8,301	0.353 (0.163)	0.357 (0.143)

Note.—All sampled individuals were adults. Estimates of observed (H_o) and expected heterozygosity (H_e) were calculated using ARLEQUIN 3.5.1.3. SD, standard deviation. Latitude and longitude values are given in degrees and minutes. The two samples from Bornholm are denoted as Bornholm 04 and 05 based on the sampling year.

48 individuals (fig. 1 and table 1). All individuals were collected during spawning, except for the Öland and the Gotland samples that presumably belong to the Bornholm basin spawning population. The samples collected within the Baltic Sea were treated as a one single population (see below).

DNA Extraction and Genotyping

DNA was extracted using the E.Z.N.A Tissue DNA kit (Omega Bio-Tek) and normalized to 100 ng/μl using a NanoDrop DN1000 instrument (Thermo Scientific Inc.) prior to genotyping, using a 12 K Illumina SNP-chip (The Cod SNP Consortium, in preparation). Out of the 10,913 SNPs on the SNPchip, 8,809 SNPs were polymorphic, showed good clustering/separation, had a call rate of greater than 95% and also showed Mendelian segregation among family individuals (data not shown). Each individual SNP locus was manually inspected and clusters were adjusted if necessary, using the Genome Studio 2011.1 software (Illumina). Of these, 262 SNPs are previously published (Moen et al. 2008; Hubert et al. 2010), 648 SNPs target 313 selected candidate genes (initially selected for a family-based study on growth and maturity, hence these markers could be treated as random genes in this study), 1,554 SNPs are nonsynonymous whereas the remaining 6,345 SNPs are randomly distributed throughout the 23 different LGs in the Atlantic cod genome (the source of

each SNP is listed in supplementary table S4, Supplementary Material online). All SNPs are referred to by their rs# or ss#, available in dbSNP (www.ncbi.nlm.nih.gov/SNP, last accessed May 1, 2015). The LGs are numbered as described by Hubert et al. (2010) and the order of SNPs is based on preliminary linkage data (Lien S, unpublished data).

Population Genetics

Within each population, estimates of observed (H_o) and expected heterozygosity (H_e) were calculated in ARLEQUIN 3.5.1.3 (Excoffier and Lischer 2010), whereas allele frequencies were calculated for all SNPs in all populations using Convert 1.31 (Glaubitz 2004). Departure from HWE was tested locus by locus in each population using the exact test by Guo and Thompson (1992) implemented in ARLEQUIN with 100,000 iterations and a Markov Chain of 1 million. We corrected for multiple testing by computing the q value for each locus, using the QVALUE package (Storey 2002) in R (R Core Team 2012), using a q value of 0.05 as a threshold for significance.

Based on results from the outlier analyses (see below), each SNP was categorized either as an outlier or as a neutral marker. Consequently, three data sets were constructed and used in the F_{ST} and STRUCTURE analyses. The first data set contains all 8,809 SNP loci and is referred to as the "full data set." The second data set contains 6,913 loci that were not under selection in any of the outlier tests (prior to corrections for multiple testing), and is referred to as the "neutral data set." The third data set, the "outlier data set," contains 233 physically unlinked outlier SNPs from all 23 LGs. Hence, outlier SNPs in high LD ($r^2 > 0.8$) with each other were represented by a single tag SNP in this data set, selected using PLINK v1.07 (Purcell et al. 2007). To minimize the use of false positives in this data set, only loci detected to be under directional selection by two or more outlier tests, after corrections for multiple testing, were included.

Locus specific F_{ST} values for all pairwise population comparisons and the weighted average F_{ST} values (Weir and Cockerham 1984) between all population pairs for all data sets were calculated in ARLEQUIN, using 10,000 permutations for significant testing. We used the Bayesian clustering model implemented in the program STRUCTURE v2.3.4 (Pritchard et al. 2000) to identify major genetic clusters (see supplementary text S3, Supplementary Material online, for details).

Linkage Disequilibrium

We evaluated the presence of LD among all 8,809 SNPs, using a data set consisting of all populations combined, calculating both inter- and intrachromosomal LD between all SNPs. LD was quantified in PLINK with the r^2 estimate, reporting all r^2 values for the intrachromosomal calculations and $r^2 > 0.2$ for the interchromosomal calculations.

Outlier Detection

Roesti *et al.* (2012) suggested that uninformative markers at low frequency be excluded from outlier analyses to increase the power of genome scans. The initial selection of SNPs on the Illumina SNP chip, based on a sequencing of eight individuals, resulted in a relatively high minor allele frequency of most of the SNPs, hence no further SNPs were excluded prior to the outlier analyses. There are several challenges in outlier studies, including detection of false positives, false negatives, and complications due to underlying population structuring. For this reason, three independent methods were used to identify candidate loci under selection.

First, we used a Bayesian regression approach implemented in BAYESCAN v2.1 (Foll and Gaggiotti 2008) which, based on F_{ST} coefficients, measures the discordance between global and population-specific allele frequencies. The degree of differentiation, based on F_{ST}, is decomposed into a locus-specific component (α), shared by all populations, and a population-specific component (β), shared by all loci. Selection is assumed when alpha is necessary to explain the observed pattern of diversity. To control for variation in the BF distribution caused by randomness and imperfection in each independent run of BAYESCAN, the median value of ten independent runs was calculated for each SNP. We carried out the simulations on the global data set, including all four populations, using stringent criteria, assuming selection to be 10%. The FDR was set to 0.01. We also calculated empirical P values, as described in Lotterhos and Whitlock (2014), based on 3,137 neutral SNPs that are more than 5 kb away from any annotated gene. We report both the median $\log_{10}$ values of the posterior odds (PO) as well as the q values and the empirical P values, only accepting $\log_{10}$(PO) above 2.0 and 1.0 as "decisive" and "strong" evidence for selection (according to Jeffreys 1961) and empirical P values < 0.01 as significant.

Second, we used the FDIST2 approach by Beaumont and Nichols (1996) implemented in the software LOSITAN (Antao et al. 2008). When using this approach, comparisons are made of F_{ST} values in relation to heterozygosity of individual loci, based on a neutral distribution, generated by means of coalescence simulations in a symmetric island migration model at mutation–drift equilibrium. We carried out the simulations on all four populations together under the Infinite Allele Method with 1 million simulations, a confidence interval of 0.99 with an FDR of 0.01, using the neutral mean F_{ST} option and forcing mean F_{ST} option. We report that the median value of ten independent runs of LOSITAN. FDIST2 is known to be robust to a wide range of nonequilibrium conditions, but it can be sensitive to demographic variations among populations as well as hierarchical genetic structure that may result in the detection of false outliers (Storz 2005). We corrected for multiple testing by computing the q value for each locus, using the QVALUE package (Storey 2002) in R (R Core Team 2012) and calculated empirical P values, as described above. We report

both the q values and the empirical P values, using 0.01 as a threshold for significance.

Third, outlier tests were performed pairwise between all four population pairs, as overall divergence based on global F_{ST} values may not detect candidates that are under selection in only some of the populations (Vitalis et al. 2001) and also to identify which populations that are under directional selection for each locus. By performing pairwise comparisons of populations, we omit some of the methodological weakness associated with population structure/demographic processes in the data sets. BAYESCAN was used as this is considered to be the most conservative test (Narum and Hess 2011). We used the same stringent criteria as in the global analyses, reporting both the median $\log_{10}$(PO) of ten independent BAYESCAN runs as well as the q values and the empirical P values, only accepting $\log_{10}$(PO) above 2.0 and 1.0 as decisive and strong evidence for selection (according to Jeffreys 1961) and empirical P values < 0.01 as significant.

Environmental Association

Correlations between allele frequencies and the environmental variables, salinity (‰), temperature (°C) and oxygen concentration, all at surface and at spawning depth, were tested using the Bayesian method by Coop et al. (2010) implemented in the software BAYENV 2.0 (Günther and Coop 2013). The environmental variables were retrieved from ICES and Helcom, as detailed in table 3. For the combined Baltic population, data from the proposed spawning area of the Bornholm Basin were used in the analyses. The rationale behind the method by Coop *et al.* (2010) is to estimate a neutral covariance matrix based on a large set of control loci. Second, a test for covariance between the environmental variables and the population specific allele frequencies at each SNP is performed, using the neutral covariance matrix as a reference (null model) to control for shared population history and gene flow. For each test, a BF is calculated based on the ratio of the posterior probabilities between the two models. A high BF indicates a support for the alternative model where the environmental variable has a linear effect on the locus of interest. As a basis for the null model, we used 3,137 neutral SNP loci that are more than 5 kb away from any annotated gene as control loci. BAYENV 2.0 was run for 500,000 Markov chain Monte Carlo (MCMC) iterations and the covariance matrix was created using an average of all matrices estimated by the program (output every 500 iteration). Next, all SNPs were tested for correlation with the environmental variables using 500,000 MCMC iterations each. Following Blair *et al.* (2014), to control for variation in the BF distribution caused by randomness in the MCMC algorithm and imperfection in the null model, 32 independent runs were carried out. The starting point for the algorithm (random seed) was drawn randomly for each run. Based on the BAYENV 2.0 results, q values were calculated according to de Villemereuil et al. (2014). Empirical p values were calculated as described above. SNPs showing median $\log_{10}(BF) > 2$ indicating "decisive evidence" and median $\log_{10}(BF) > 1$ indicating "strong evidence" according to Jeffreys (1961) were regarded as significant results when also significant at $q < 0.01$.

SNP Annotation

The original SNP chip design was based on the ATLCOD1A genome assembly (www.codgenome.no, last accessed May 1, 2015) that has a different coordinate system than the genome assembly for which Ensembl annotation is available (ATLCOD1C). Therefore, the location of each SNP was determined by aligning 200 bp of flanking sequence to the ATLCOD1C assembly (Star et al. 2011) using BWA aln -o 1 -n 0.04, seeding disabled (Li and Durbin 2009). Only unique hits were selected. The locations of annotated genes were downloaded from the Ensembl database Gadus_morhua.gadMor1.73 using Biomart (Kasprzyk 2011). The distance of the flanking sequence of the SNPs from genes (exons) was determined using BEDclosest including the options -t "first" and -d (BEDtools v2.16.2; Quinlan and Hall 2010). The protein transcripts of Ensembl genes that were associated with the SNPs through this approach were annotated with BLAST2GO (Conesa et al. 2005) using public database b2g_sep13. Protein transcripts were aligned to the refseq_protein data using the BLASTP algorithm allowing a maximum of 20 hits with a minimum e value of 1 E-3, as implemented in BLAST2GO. Apart from setting the evidence code weight of IEA (electronic annotation evidence) to 1, default weights were used. Annotation was augmented using the Annex function in BLAST2GO.

Results

SNP-Set and Samples

Based on a 12 K SNPchip, we analyzed a set of 8,809 polymorphic SNPs (see Materials and Methods for details), distributed over all 23 linkage groups (LGs) with an average distance of 94,000 bp between SNPs, based on a genome size of 830 Mb (Star et al. 2011). A total of 5,565 SNPs were located within 5,000 bp of 4,522 Ensembl annotated genes (supplementary table S4, Supplementary Material online). A final number of 194 individuals of Atlantic cod (individual call rate $> 95\%$) from seven different localities were investigated (fig. 1, table 1). As the four Baltic localities likely comprised individuals from the same breeding population, these data were pooled into a single sample, which was justified by assignment testing (see Materials and Methods and supplementary text S1, Supplementary Material online) and further supported by tests for deviation from Hardy–Weinberg equilibrium (HWE) (supplementary text S2, Supplementary Material online). Only one SNP locus was significantly out of HWE after false discovery rate (FDR) correction ($q < 0.05$) in

the pooled Baltic sample, indicating no sign of a Wahlund effect. The number of polymorphic loci, observed- and expected heterozygosity (H_o and H_e) were similar in all populations (table 1).

Population Genetic Structuring

In addition to the full data set, SNPs were categorized either as outliers or as neutral, based on outlier analyses (see Outlier Detection and Environmental Association section). In the full- and neutral data set, all average pairwise F_{ST} comparisons were statistically significant except between Kattegat and Öresund, indicating little genetic differentiation between these two populations (table 2). In comparison, based on 233 unlinked SNPs from the outlier data set, all pairwise F_{ST} comparisons were significant, indicating some genetic differentiation also between the Kattegat and Öresund populations.

Locus-specific F_{ST} values for all pairwise population comparisons indicate a genome-wide pattern of high F_{ST} when comparisons are made to the Baltic Sea, whereas few discrete regions of the genome show elevated F_{ST} values in the remaining comparisons (fig. 2). A total of 40 private alleles (supplementary table S2, Supplementary Material online) were detected among the populations, 38 of which were rare in frequency (<0.04). The highest number of private alleles (23) was observed in the North Sea population and the lowest number (3) was observed in the Baltic population. The two private alleles of higher frequency (0.179 and 0.232) were found in the North Sea population, in complete linkage disequilibrium (LD) with each other (ss1712302407 and ss1712304757, located 241 bp apart, in LG3). In total, 334 fixed allele frequencies were detected in any population; however, only 84 of these showed frequency differences of more than 0.1 in any pairwise comparison. Out of these, 59 were unique to the Baltic population (19 potentially under selection) whereas 11, 2 and 1 were unique to the North Sea, Kattegat and Öresund populations, respectively. In addition, 59 outlier SNPs were close to fixation (frequency > 0.95; supplementary table S2, Supplementary Material online).

Bayesian cluster analyses (STRUCTURE analysis; supplementary text S3, Supplementary Material online) strongly support a separation between the Baltic population and the other populations for all three data sets (supplementary fig. S1, Supplementary Material online). At higher K values, the Kattegat and the Öresund populations, representing the transition zone between the North Sea and the Baltic Sea, form an intermediate group with more similarity to the North Sea than to the Baltic population (supplementary fig. S2, Supplementary Material online). The highest delta K value was detected for $K = 2$ for all three data sets, indicating a deep-rooted structuring in the data and that the Baltic cod is separated by a larger genomic distance to all the other populations. The identified clusters correspond well with the

Table 2

Pairwise F_{ST} Values among Atlantic Cod Populations, Using a Full-, Neutral-, and Outlier Data Set (See Text for Details)

	North Sea	Kattegat	Öresund	Baltic
Full data set (8,809 SNPs)				
North Sea	—	**0.01595**	**0.01467**	**0.07095**
Kattegat		—	0.00026	**0.05140**
Öresund			—	**0.04824**
Baltic				—
Neutral data set (6,913 SNPs)				
North Sea	—	**0.00962**	**0.00996**	**0.04840**
Kattegat		—	0.00007	**0.03138**
Öresund			—	**0.03061**
Baltic				—
Outlier data set (233 SNPs)				
North Sea	—	**0.12065**	**0.10278**	**0.40130**
Kattegat		—	**0.00825**	**0.32685**
Öresund			—	**0.29778**
Baltic				—

Note.—Values in bold are significant values (P values < 0.05), calculated using ARLEQUIN 3.5.1.3.

Table 3

Environmental Conditions at the Atlantic Cod Sampling Locations

Population ID	Location	Spawning Month	Spawning Depth (m)	Salinity Surface	Salinity Spawning Depth	Temp. Surface	Temp. Spawning Depth	Oxygen Surface	Oxygen Spawning Depth	Source
North Sea	North Sea	Feb.–April	40–60	34.7	34.8	6.0	6.1	6.8	6.7	ICES
Kattegat	Kattegat	Feb.–April	60–100	24.2	34.1	3.5	6.3	8.0	6.5	Helcom
Öresund	Öresund	Feb.–April	40–53	16.4	30.9	3.4	6.4	8.5	6.2	Helcom
Baltic	Bornholm 05	June–Aug.	50–82	7.4	13.6	13.6	6.2	7.2	4.2	Helcom
Baltic	Bornholm 04	June–Aug.	50–90	7.4	13.6	13.6	5.7	7.4	4.5	Helcom
Baltic	Gotland	June–Aug.	50–90	6.9	8.6	9.3	4.4	7.3	4.1	Helcom
Baltic	Öland	June–Aug.	50–82	6.7	8.4	8.4	4.8	7.6	3.4	Helcom

Note.—Salinity, temperature, and oxygen values were obtained online from ICES or Helcom. Salinity is denoted in ‰, temperature in °C and oxygen in concentration. The two samples from Bornholm are denoted as Bornholm 04 and 05 based on the sampling year.

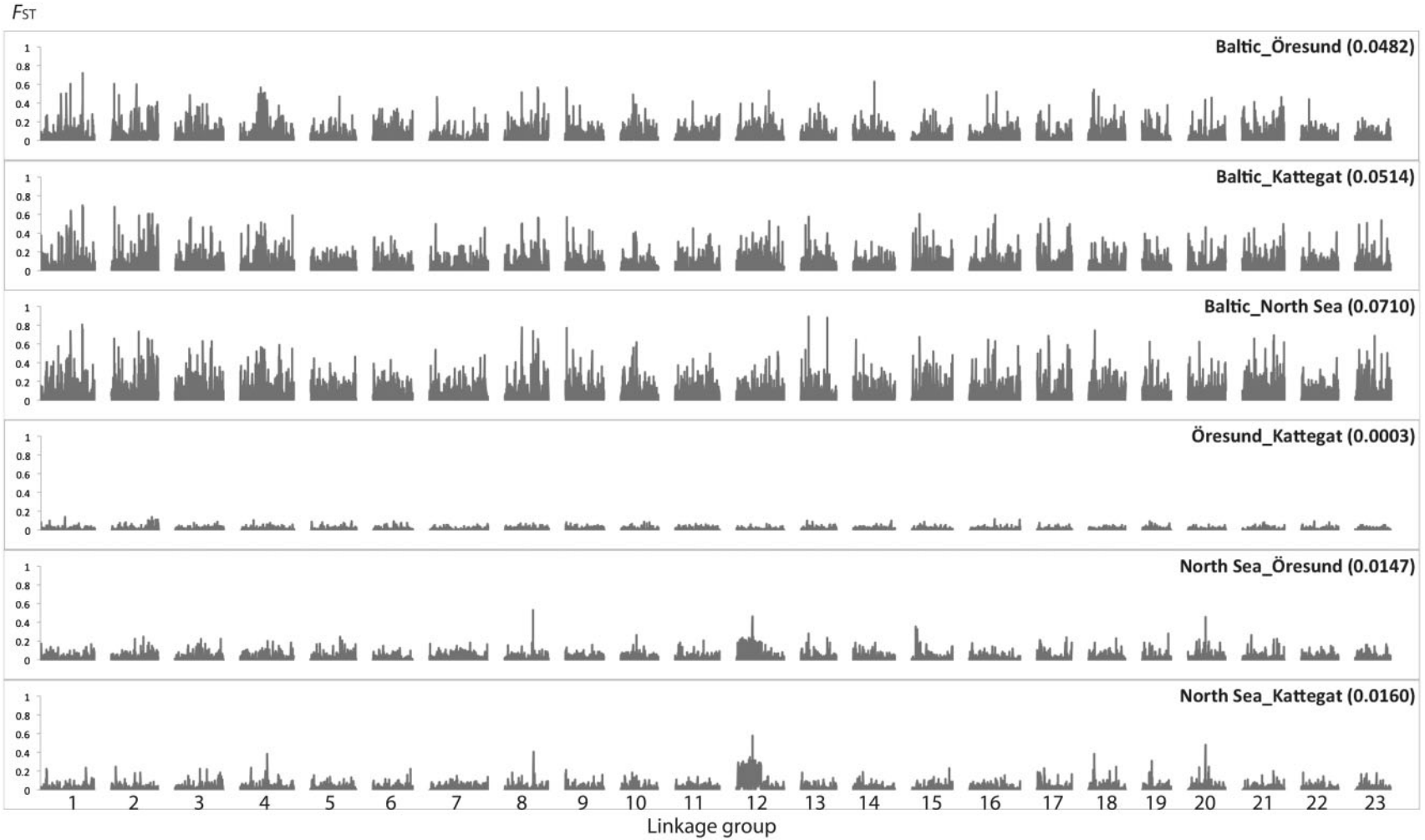

FIG. 2.—Pairwise F_{st} values between Atlantic cod populations using 8,809 SNPs in 23 LGs. SNPs are ordered according to LG and position within LGs based on the results of a preliminary SNP linkage map (Lien S, unpublished data). LG nomenclature follows that of Hubert et al. (2010). Median F_{st} estimates between the population pairs are denoted in parentheses.

overall population structure estimated by pairwise F_{ST} comparisons (table 2).

Linkage Disequilibrium

LD was evaluated among all 8,809 SNPs, independent of LG localization to detect both inter- and intrachromosomal LD. As expected, we find SNPs with high r^2 values (>0.75) within all LGs (supplementary table S3, Supplementary Material online). The distributions of the intrachromosomal r^2 values among the different LGs show that SNPs with high r^2 values often reside within the same scaffolds and hence are physically close. However, for LG2, 7 and 12, an extensive number of SNPs with high LD (fig. 3) covers multiple scaffolds including 15 and 5 scaffolds in two close but separated blocks in LG2, 14 scaffolds in LG7, and 31 scaffolds in LG12 (supplementary table S3, Supplementary Material online). The LD pattern (fig. 3) also shows less distinct blocks of elevated r^2 values in LG4, 10, and 17. Interestingly, relatively high levels of r^2 values (>0.3) also occur between SNPs on different LGs, particularly between LG1 and 2 and between LG2 and 4 (supplementary table S3, Supplementary Material online).

Outlier Detection and Environmental Association

We used two outlier detection approaches, BAYESCAN and LOSITAN, combined with a method that explores the correlation between environmental variables and allele frequencies (BAYENV 2.0) to uncover various aspects of the genomic regions under selection (see Materials and Methods for details).

In the global outlier analyses comprising all four populations, we identified 123 (1.4%) and 114 (1.3%) SNPs as candidates for divergent selection ($\log_{10}$(Bayes factor [BF]) > 1, empirical $P < 0.01$), using BAYESCAN and LOSITAN, respectively. Pairwise outlier analyses using BAYESCAN, revealed 228 (2.6%) SNPs as candidates for divergent selection (supplementary fig. S3 and table S4, Supplementary Material online). However, the pattern and the number of outliers in each pairwise comparison varied (fig. 4*b*–*g*). In total, 266 SNPs (3.0%) were detected as candidates for divergent selection using global or pairwise outlier test. LG2 and 12 had the highest proportion of SNPs potentially under selection (36 and 98, respectively; supplementary table S4, Supplementary Material online). These two LGs contain large LD blocks, where most, but not all, of these outliers reside (fig. 3).

Correlations between allele frequencies and salinity (‰), temperature (°C) and oxygen concentration at surface and at spawning depth (table 3) were investigated using a landscape genomic approach (BAYENV 2.0). Significant association (median $\log_{10}(BF) > 1$ and $q < 0.01$) with at least one environmental variable was detected in 234 (2.6%) of the 8,809 examined SNP loci, distributed across all LGs except LG6 (supplementary fig. S3 and table S4, Supplementary Material online). The pattern and the number of SNPs with correlation to each environmental variable varied (fig. 5a–f). An association with salinity and oxygen level at spawning depth is clearly observed at a genome-wide scale, and in particular toward the end of LG2, whereas temperature shows a weak correlation in LG12 (fig. 5). There is a strong correlation between SNPs associated with salinity and oxygen, whereas SNPs associated with temperature are not strongly correlated with salinity or oxygen SNPs (fig. 5).

In total, 326 SNPs (3.7%) were candidates for divergent selection or were associated with an environmental variable (supplementary fig. S3, Supplementary Material online). The overlap between the outlier and the landscape genomic data sets reveals 174 SNPs with significant evidence of selection in both the outlier approach (global or pairwise comparisons) and the landscape genomic approach, distributed on 102 different scaffolds on most LGs, except for LG6, and 23 (supplementary table S4, Supplementary Material online). Of these 174 SNPs, 129 loci were located in or within 5 kb of an annotated gene, of which 61 were located in exons and 48 were nonsynonymous substitutions causing amino acid changes (supplementary table S4, Supplementary Material online). Further, 123 of the 174 outlier SNPs reside in five LGs (LG1, 2, 4, 8, and 12; supplementary fig. S3, Supplementary Material online). In LGs where outliers are detected, they do not seem to be clustered within the LGs, except for the outliers in LG2, 12 and to some extent in LG4, 8, 10, 17, and 18 (fig. 3).

Identifying Genes and Genomic Regions under Selection

Three LGs (LG2, 7, and 12) have substantial long-distance LD (fig. 3) and the LD pattern is similar within the different populations (data not shown). In LG2, 51 SNPs were candidates for selection, out of which 32 were located in two tightly linked regions (consisting of 22 and 10 SNPs, respectively) toward the end of the LG (fig. 3). All SNPs that were candidates for selection on this LG were associated with oxygen and salinity (primarily at spawning depth). Further, all of these outliers were detected as pairwise outliers relative to the Baltic Sea, but not in any of the other pairwise comparisons, except for an SNP (ss1712297946) close to a C-type lectin gene (*CLEC16A*) toward the beginning of the LG. Combined, these results indicate that the outliers on LG2 are of particular relevance for the adaptation to the low saline conditions in the Baltic Sea. In LG7, one single outlier SNP (ss1712305126) with decisive evidence of selection was located within an exon in an Na+/K+ ATPase gene (*ATP1A1*) and was not associated with the large LD region present on LG7 (fig. 3). In LG12 the long-distance LD area contains 37 decisive outliers, but notably only detected in the pairwise comparison between North Sea and Kattegat/North Sea and Öresund. No decisive pairwise outliers were detected in any of the comparisons with the Baltic population (supplementary table S4, Supplementary Material online). However, the majority of outliers detected on LG12 appears to be under a less stringent selection regime as the

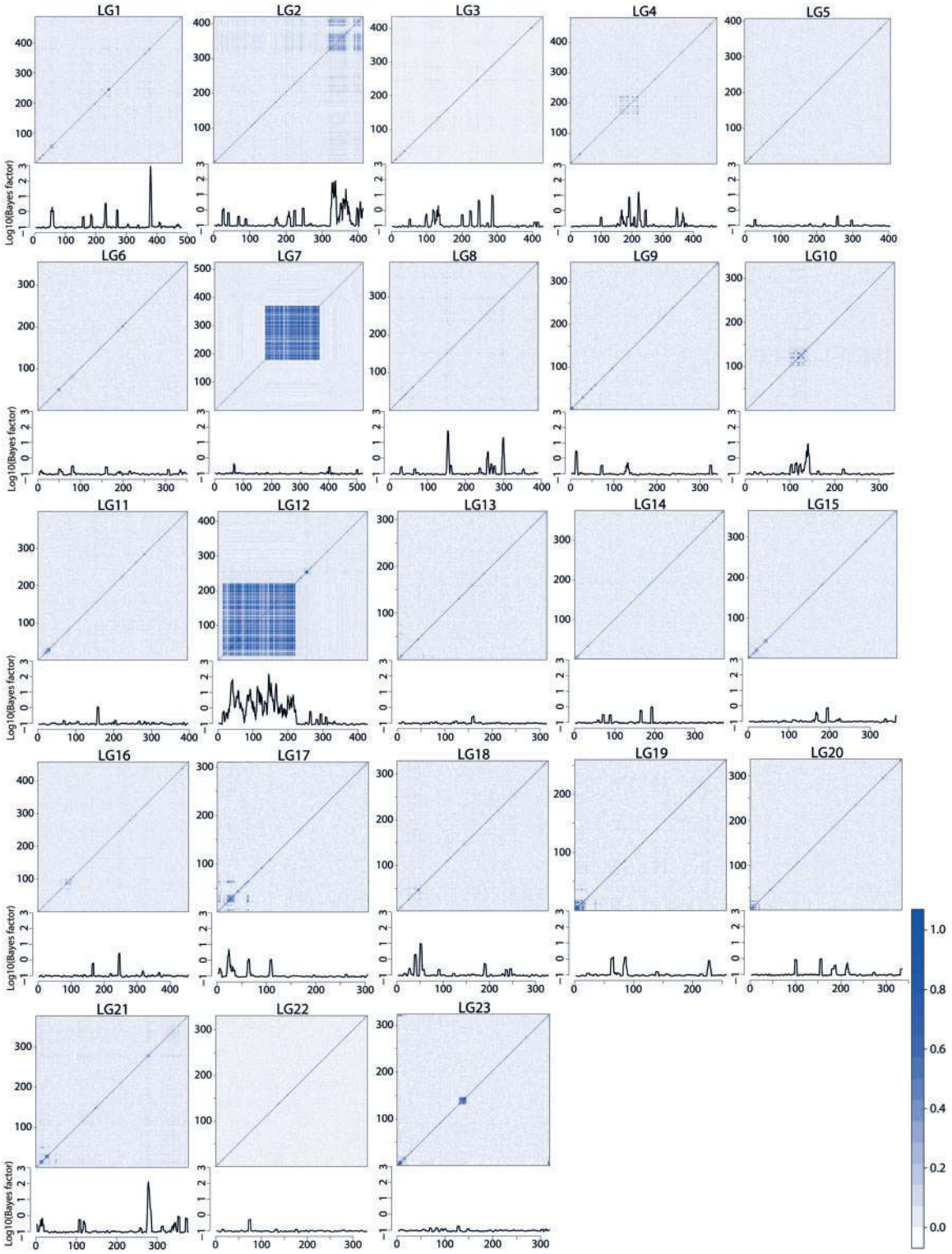

Fig. 3.—Global outlier pattern and LD in Atlantic cod. The extent of pairwise LD among loci within 23 LGs, measured as *r2*, is estimated using all individuals and populations. A color bar on the right hand side denotes the strength of the *r2* values. A global outlier pattern (underneath each LD plot) based on the same SNP data is visualized using a moving average of the median log10(BF), calculated using the global BAYESCAN outlier analyses (see text for details). The SNPs are plotted according to LG and their respective position within the LGs, based on the results of a preliminary SNP linkage map (Lien S, unpublished data). LG nomenclature follows Hubert et al. (2010).

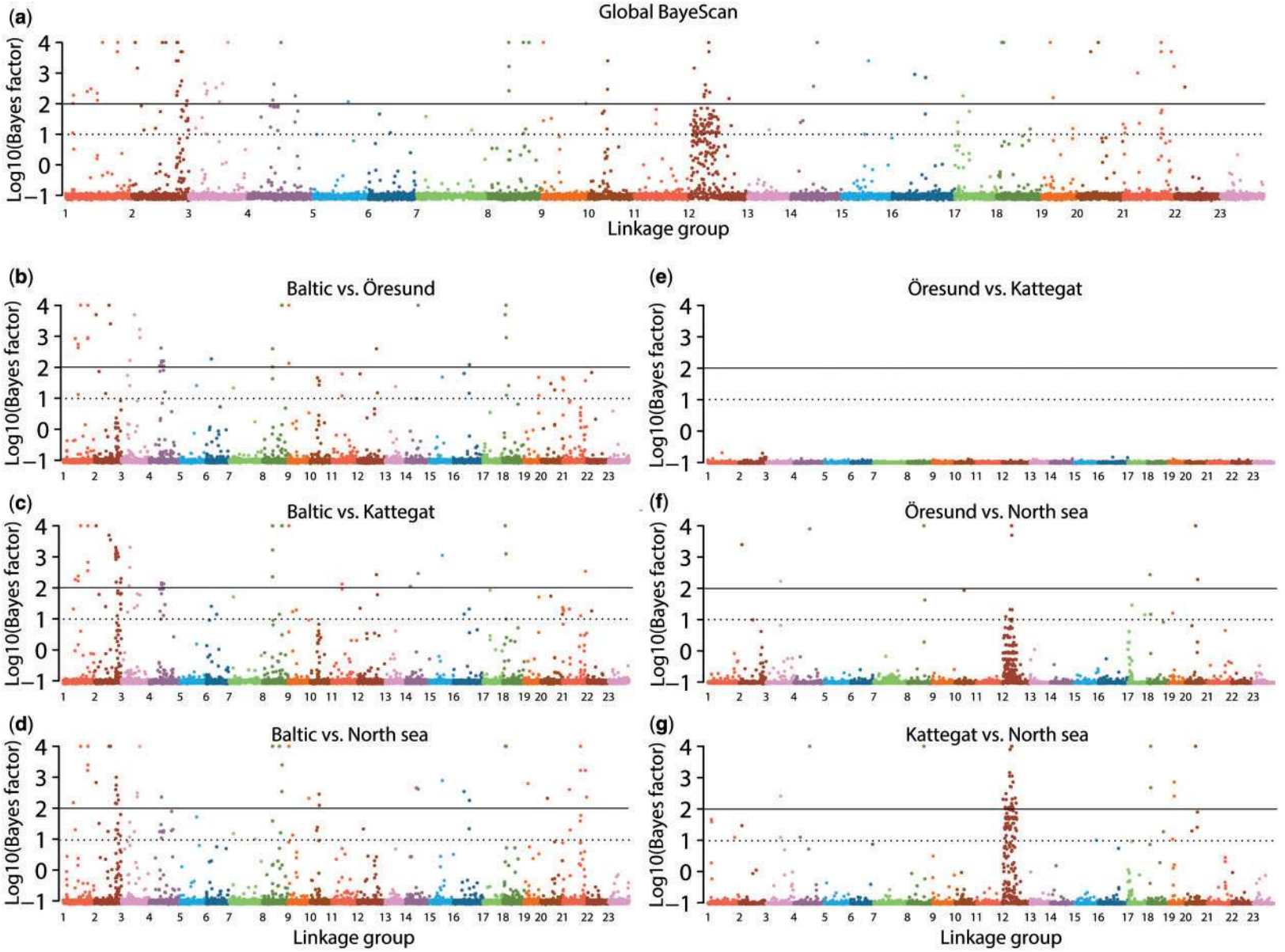

FIG. 4.—Global and pairwise outlier patterns in Atlantic cod. Manhattan plots of global and pairwise outlier analyses based on median log10(BF) from ten replicate runs of BAYESCAN. The SNPs are plotted according to LG and their respective position within the LGs along the *X* axis based on the results of a preliminary SNP linkage map (Lien S, unpublished data). LG nomenclature follows Hubert et al. (2010). The dotted line at log10(BF) = 1 indicates "strong association" and the solid line at 2 indicates "decisive association" according to Jeffreys (1961).

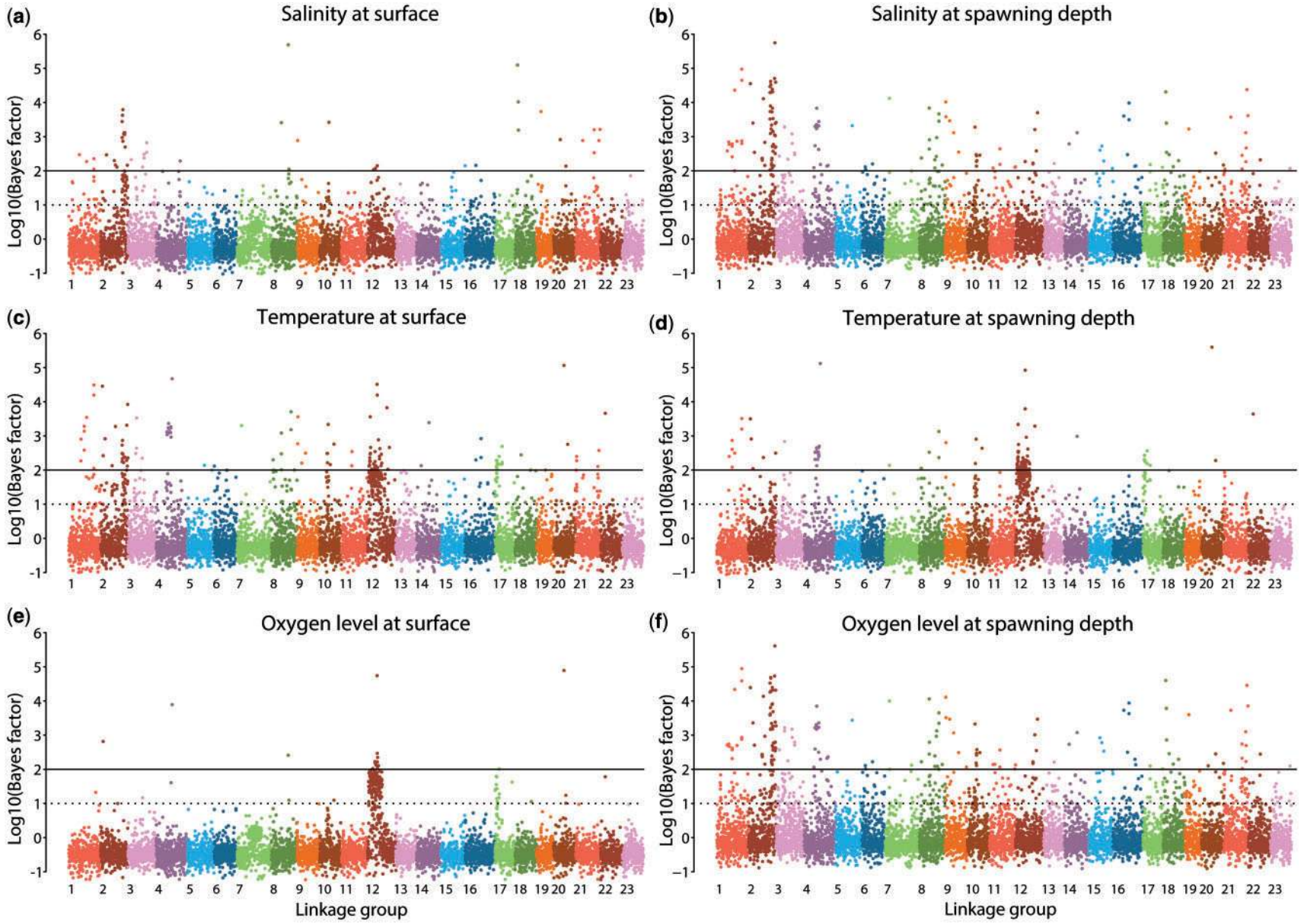

FIG. 5.—SNPs associated with environmental variables in Atlantic cod. Manhattan plots of SNP association with salinity, temperature, and oxygen level (all at surface and at spawning depth, see table 1) based on median log10(BF) from 32 independent runs of BAYENV. The SNPs are plotted according to LG and their respective position within the LGs along the *X* axis, based on the results of a preliminary SNP linkage map (Lien S, unpublished data). LG nomenclature follows Hubert et al. (2010). The dotted line at log10(BF) = 1 indicates strong association and the solid line at 2 indicates decisive association according to Jeffreys (1961).

number of outliers rise to 103, using $\log_{10}(BF) > 1$ and empirical $P < 0.01$ as a threshold for significance. Similarly, there is a clear selection pattern within the linked region in LG12, associated with temperature—but not with salinity or oxygen (fig. 5 and supplementary table S4, Supplementary Material online).

In addition, we find three less distinct LD regions (LG4, 10, and 17) that also harbor several outlier candidates. In LG4, the LD region contains 11 of the 15 decisive outlier SNPs. All of these SNPs showed signs of selection in the pairwise comparisons with the Baltic population, but not with any of the other populations and were all decisively associated with salinity and oxygen, primarily at spawning depth (fig. 5 and supplementary table S4, Supplementary Material online). Hence, there are strong indications that also the genomic region on LG4 has relevance for adaptation to the low saline conditions in the Baltic Sea. Notably, SNPs within the LD block on LG4 are linked to SNPs on the LD blocks on LG2 ($0.3 < r^2 < 0.4$; supplementary table S3, Supplementary Material online), possibly indicating a common selection pressure on the two genomic regions. In LG10, the main outlier peak lies just outside the LD area, with some less significant outlier SNPs residing within the LD area (fig. 3). All SNPs under selection on LG10 were associated with salinity and oxygen at spawning depth (supplementary table S4, Supplementary Material online). In LG17, seven SNPs displayed decisive evidence of selection and four of these SNPs, associated with oxygen at surface, were within the LD region (fig. 3).

A total of 51 SNPs, within 5 kb of 43 annotated genes (see Discussion) were decisively associated with salinity and oxygen and were specific to the Baltic Sea, relative to the other three sampling locations (table 4). These pairwise outliers are likely to identify the strongest genomic differentiation, shaped by the environment in the Baltic Sea. Surrounding these strong outliers, other less pronounced outlier SNPs were also detected (figs. 4 and 5 and supplementary table S4, Supplementary Material online).

Discussion

We identified a set of directionally selected loci with allele frequencies strongly correlated with habitat differences in salinity, oxygen, and temperature. These loci reside within several discrete regions within the Atlantic cod genome, suggesting that these regions are influenced by divergence hitchhiking and indicate presence of genomic islands of divergence.

Patterns of Genomic Diversification Are Caused by Adaptation

Neutral and selective evolutionary forces shape the genetic makeup among populations and it is important to disentangle these effects. Using the set of presumed neutral SNPs we show that the Baltic population is divergent relative to the other three populations (table 2 and supplementary fig. S1, Supplementary Material online). Although effects of selection on some of these SNPs cannot be totally excluded, our results are in line with microsatellite data, showing low levels of genetic differentiation among North Sea, Kattegat and Öresund populations compared with the Baltic population (Nielsen et al. 2003) which could possibly be a secondary effect of local adaptation to the Baltic conditions. The Baltic cod population is on the margin of its ecological and geographical range (Johannesson and André 2006), suggesting that the observed divergence between the Baltic population and the other populations could be partly caused by decreased population density at its distribution margin (Bridle and Vines 2007; Orsini et al. 2013). This is less likely however, as the effective population size estimates of Baltic cod still remain relatively high (Poulsen et al. 2006). Poulsen et al. (2006) estimated N_e for Baltic and North Sea cod using temporal analyses of allele frequencies in microsatellites, and found N_e in the Baltic population to be only slightly smaller than in the North Sea population (both in the thousands), but still large enough to not loose its evolutionary potential (Franklin 1980). Poulsen et al. (2006) concluded by urging more attention to selection than to bottleneck effects on genetic composition of Baltic cod. Hence, there are strong indications that N_e in the Baltic population is sufficiently large and that bottleneck effects should not be a major issue, supporting the view that demographic processes alone are not the ultimate driver for the outlier patterns that we observe. Rather, a scenario where adaptation drives neutral structure and "isolation by adaptation" or "isolation by environment" is more likely (Orsini et al. 2013; Wang and Bradburd 2014).

Outlier Detection and Methodological Considerations

The consistency of the different approaches for outlier detection (supplementary fig. S3, Supplementary Material online) and the functional relevance of the underlying genes strongly suggest that the majority of the identified loci and their associated genomic regions is subject to divergent selection (see however Bierne et al. 2013). Nevertheless, it is inherently difficult to define a biologically meaningful cutoff value in the various outlier tests, and identified outliers remain candidates. Here, we attempted to achieve robust conclusions by combining cutoff values based on $\log_{10}(BF/PO)$ values (Jeffreys 1961) with q values and empirical P values (for a detailed discussion, see supplementary text S4, Supplementary Material online).

It has been suggested that outlier tests may have high false positive rates due to the effects of population structure and bottlenecks (see, e.g., Narum and Hess 2011; de Villemereuil et al. 2014; Lotterhos and Whitlock 2014). One way to reduce the effect of population structure is to perform outlier analyses between pairs of populations (cf. Vitalis et al. 2001). Another way to compensate for population structuring is by calculating empirical P values based on a set of putatively neutral SNP

markers (determined a priori) to create a null distribution to test all SNP markers against (cf. Lotterhos and Whitlock 2014). We employed both of these methods to minimize the effects of population structuring in the data (supplementary text S4, Supplementary Material online).

Recent simulation studies have suggested that correlation-based approaches in many instances outperform more traditional population genomic approaches in accurately identifying loci under divergent selection (de Mita et al. 2013; de Villemereuil et al. 2014; Lotterhos and Whitlock 2014). By also analyzing our data using BAYENV 2.0, which accounts for demographic signals, we were able to disentangle the genetic signals caused by selection rather than demography with more confidence (details in supplementary text S4, Supplementary Material online).

Genomic Divergence Implies Ongoing Speciation

In the process of population divergence and speciation, heterogeneous genomic divergence can be formed (Nosil et al. 2009; Smadja and Butlin 2011), observed as elevated levels of divergence in selected regions, surrounded by physically linked loci through divergence hitchhiking (cf. Charlesworth et al. 1997; Via 2009, 2012). Divergent selection may further promote extrinsic reproductive isolation where migrants between different environments have reduced fitness (Feder, Egan, et al. 2012). For Atlantic cod, ecological adaptation seems to restrict gene flow across the salinity barrier, and in the Baltic Sea population we observe a genome-wide pattern of divergence, with SNPs that are candidates for selection in all LGs except LG23 (fig. 4*b*–*d*). This is consistent with the theory of isolation by adaptation and the later stages of ecological speciation where gene flow is small or nonexistent (Feder, Egan, et al. 2012). Further, most of these outlier SNPs cluster in genomic regions in high LD (fig. 3 and supplementary table S3, Supplementary Material online) suggesting that mutations in close proximity (high LD) hitchhike with the selected genomic region, increasing the size of the local genomic islands (Smith and Haigh 1974; Feder, Egan, et al. 2012; Feder, Gejji, et al. 2012). However, recombination will reduce LD across the genome over time, resulting in larger patterns of genome-wide divergence, although heterogeneity among regions may still be present due to varying degree of selection and recombination. Eventually, ecologically favored alleles will predominate in one habitat and neutral- and universally favored alleles will potentially be present in all habitats (see, e.g., Savolainen et al. 2013). At this point, diagnostically fixed differences can be observed between populations (Feder, Egan, et al. 2012; Seehausen et al. 2014). In our data set, fixed allele frequencies were detected in 19 SNPs potentially under selection in the Baltic population and 59 SNPs were nearly fixed (frequency > 0.95), suggesting strong ongoing diversification. These results are comparable to those in other fish species such as three-spine sticklebacks (*Gasterosteus aculeatus*) (Roesti et al. 2012, 2014) and lake whitefish (*Coregonus clupeaformis*) (Gagnaire et al. 2013) for which similar patterns of divergence have been described.

Among the North Sea, Kattegat, and Öresund populations, where obvious environmental differences are small, we observe a pattern consistent with the early stages of genetic divergence, where a substantial amount of gene flow is still prevalent, and one would expect the presence of relatively few but potentially large genomic islands of divergence (Feder, Egan, et al. 2012). In these populations, we detected regions that are subject to divergent selection in few LGs (fig. 4*e*–*g*) compared with the Baltic population and most of these SNPs reside in LG12, in high LD with each other (fig. 3 and supplementary table S3, Supplementary Material online). The finding of selection in relatively few LGs agrees with previous findings in Atlantic cod (Bradbury et al. 2010, 2013; Hemmer-Hansen et al. 2013; Karlsen et al. 2013). This pattern of early divergence has also been shown in the studies of pea aphids (*Acyrthosiphon pisum pisum*), African malaria mosquitoes (*Anopheles gambiae*), and *Heliconius* butterflies (Turner et al. 2005; Via and West 2008; Via 2009; Nadeau et al. 2012; Via et al. 2012). In some instances, studies of the same species under different environmental conditions have arrived at different results regarding the distribution of the divergence pattern (see, e.g., Hohenlohe et al. 2011; Roesti et al. 2012; Gagnaire et al. 2013). This seems to be the case also in our investigation, where the observed divergence patterns for the Baltic population comparisons (fig. 4*b*–*d*) differ markedly to that observed in the North Sea, Kattegat, and Öresund populations (fig. 4*e*–*g*). In some species strong signatures of selection have been shown for chromosome inversions, although few studies have identified the actual target genes for selection (Kirkpatrick and Kern 2012). Nevertheless, our current data cannot accurately resolve if the clustering of outliers in LD regions is due to inversions.

Genomic Regions under Divergent Selection in the Baltic Sea

Studying the genetic changes that contribute to reproductive isolation in partly reproductively isolated populations (before they become confounded by additional genetic differences after speciation is complete) may reveal important aspects of the speciation process (Via 2009). Our sampling design enabled us to do so and our results indicate that several discrete regions of the Atlantic cod genome are candidates for directional selection and most are associated with adaptation to the local conditions in the Baltic Sea. It is likely that salinity and oxygen levels are strong evolutionary forces for Baltic cod (figs. 4 and 5) and our results agree with findings that natural selection can shape population structure on short spatial scales, despite the high dispersal capacity of marine organisms (see, e.g., Gaggiotti et al. 2009; Bradbury et al. 2010, 2013; André et al. 2011; Lamichhaney et al. 2012; Nielsen et al.

2012; Corander et al. 2013; Defaveri, Shikano, et al. 2013; Hemmer-Hansen et al. 2013; Vandamme et al. 2014).

During speciation in the presence of gene flow, the establishment and maintenance of genomic regions that sufficiently can resist gene flow is only likely if divergent selection (or sexual selection) is strong and hence the initial barriers to gene flow are likely to evolve quickly (Via 2001; Hendry et al. 2007). As a result of such strong selection, the genomic regions causing the reproductive isolation become particularly distinctive relative to the remaining genome, facilitating its discovery in empirical analyses (Via 2009). We detect several distinct genomic regions in which a number of SNPs in high LD appear as candidates for selection. Several of these regions have been associated with environment variables in other studies. For instance, we identify 32 SNPs in linked regions toward the end of LG2 as candidates for selection. The same regions have previously been associated with variation in temperature (Bradbury *et al.* 2010). Nevertheless, in our study these SNPs appear more associated with salinity and oxygen (at spawning depth) than with temperature (fig. 5 and supplementary table S4, Supplementary Material online). Similarly, one SNP located 337 bp from the prolactin (*PRL*, in LG2) gene, lying outside the region defined by Bradbury *et al.* (2010), was also associated with temperature but more strongly so with salinity and oxygen (supplementary table S4, Supplementary Material online). The gene for hemoglobin beta-1 (H*b*-β*1*) is also located in LG2 (in contig_03031) and identified to be 17.8 cM away from the outlier region by Bradbury *et al.* (2010). It has been proposed by Andersen *et al.* (2009), but see also Sick (1961), that the Baltic cod is adapted to the environmental hypoxia and temperature in the Baltic Sea by possessing the high-affinity Val-Ala form of H*b*-β1. In the same scaffold as H*b*-β1 an outlier in a nonsynonymous SNP (ss1712304723) was identified, which is associated with salinity (at surface and spawning depth) and oxygen level (at spawning depth). If this defined outlier SNP is linked to H*b*-β1, it may reflect the hemoglobin polymorphism in the Baltic Sea.

Moreover, we detected 17 SNPs as candidates for selection in LG1. The same region has been associated with a migratory ecotype by Hemmer-Hansen *et al.* (2013) and by Karlsen *et al.* (2013) between migratory Northeast arctic cod and stationary Norwegian coastal cod. This region contains the well-known pantophysin gene polymorphism (*Pan*I, in scaffold_09065) that has been used to determine individuals as either stationary or migratory (Fevolden and Pogson 1997; Pogson and Fevolden 2003). This marker is known to be nearly monomorphic for the coastal type (*Pan*I^{A}) in North Sea and Baltic populations (Case et al. 2005) which was also the case for the closest genotyped SNP in this study (ss1712302787, 5,148 bp away; supplementary table S2, Supplementary Material online). Notably, within LG1, all pairwise outlier SNPs relative to the Baltic Sea, and loci environmentally associated with salinity and oxygen (table 4), are located within the same scaffolds (supplementary table S4, Supplementary Material online) as Hemmer-Hansen *et al.* (2013) describes as associated with a migratory ecotype. The fact that we detect outliers in the same region that has previously been identified as associated with a migratory ecotype, but also between the Baltic- and the North Sea (Hemmer-Hansen et al. 2013), in the Baltic population, might indicate that this region is not selected by migratory behavior per se.

Furthermore, in LG12, we observe an outlier pattern driven by the North Sea/Kattegat comparison and to some extent also the North Sea/Öresund comparison (fig. 4). The landscape genomic analysis suggests a temperature association within the long-range LD region on LG12 (figs. 3 and 5) that corresponds to the temperature-associated region identified by Bradbury *et al.* (2010). This correlation is not as robust as the correlations found for salinity and oxygen level at the other LGs, as decisive association was only detected in eight SNPs (using $q < 0.01$ as a cutoff). These results suggest that LG12 is not central in the adaptation to the Baltic Sea environment, but rather the selection in this genomic region is driven by adaptation to temperature variation, not encountered by the Baltic Sea cod. In the tightly linked outlier block on LG12, the allele frequencies within the Baltic Sea and within the North Sea are similar (supplementary table S2, Supplementary Material online) possibly reflecting that the detected outlier pattern either predates the Baltic split or is subject to similar selection pressure in both areas. In Hemmer-Hansen *et al.* (2013) it is evident that SNPs in this region shows elevated F_{ST} values in a Norwegian coastal cod/North Sea comparison, which may indicate that it is the North Sea that is driving this outlier pattern. However, it is also possible that the Öresund and Kattegat populations are experiencing specific selection pressure in this genomic region or share a common ancestry with the Norwegian coastal cod populations.

Apart from outliers in long LD regions, we also identify single SNPs as significant candidates for selection, associated with at least one environmental variable (figs. 4 and 5 and supplementary table S4, Supplementary Material online) indicating that several smaller genomic regions also play a role in the adaptation to the different environments under study. As expected, we did not identify all previously known outliers regions in Atlantic cod; however, we did identify some novel regions under selection. For instance, in LG7 we detected a single decisive outlier SNP, located within an exon in the Na^+/K^+-ATPase gene (*ATP1A1*); see discussion on physiological adaptation to low salinity. Nevertheless, no SNPs under selection were detected in the large LD region on LG7 (fig. 3) which was previously described as associated with temperature and behavior (Bradbury et al. 2010; Hemmer-Hansen et al. 2013).

Bradbury *et al.* (2014) identified strong interchromosomal LD (ILD) among the large LD regions in LG2, 7 and 12, supporting a hypothesis of divergence hitchhiking, transitioning to genome hitchhiking with reproductive isolation. Even though we find some evidence of high ILD between SNPs among the same regions, the most notable findings of ILD

Table 4

SNPs Closer than 5 kb to an Annotated Gene with Decisive Association (i.e., median $\log_{10}(BF) > 2$ and Empirical $P < 0.01$, see text for details) to Salinity and Oxygen at Spawning Depth that Are Also Strong Pairwise Outliers across the Salinity Barrier

					BAYENV 2.0 $\log_{10}$(BF)		BAYESCAN $\log_{10}$(PO)					
Locus	LG	Distance to Exon	HGNC	Gene ID	Salinity Depth	Oxygen Depth	Baltic Öresund	Baltic Kattegat	Baltic North Sea	North Sea Kattegat	North Sea Öresund	Öresund Kattegat
ss1712302658	1	0	*TSHB*	ENSGMOG00000011010	**2.78**	**2.62**	**2.75**	**2.38**	0.37	−1.02	−1.03	−1.02
ss1712302659	1	244	*TSHB*	ENSGMOG00000011010	**2.72**	**2.57**	**2.64**	**2.22**	0.38	−1.02	−1.02	−1.00
ss1712299135	1	21 (I)	*GRIP2A*	ENSGMOG00000017983	**2.58**	**2.71**	−0.37	**1.32**	**2.18**	−1.02	−0.73	−0.99
ss1712303317	1	0 (NS)	*FBLN2*	ENSGMOG00000017143	**2.77**	**2.68**	**2.92**	**2.27**	**1.30**	−1.01	−1.03	−1.00
ss1712303323	1	0 (NS)	*COL2A1*	ENSGMOG00000009692	**4.36**	**4.34**	>4	>4	>4	−1.02	−1.02	−1.00
ss1712300221	1	0 (NS)	—	ENSGMOG00000013560	**4.98**	**4.94**	>4	>4	>4	−1.04	−1.03	−1.00
ss1712304198	1	0 (NS)	—	ENSGMOG00000013560	**2.80**	**2.88**	**2.92**	**2.55**	**3.22**	−1.01	−1.01	−1.00
ss1712304199	1	0 (NS)	—	ENSGMOG00000013560	**2.88**	**2.94**	**2.96**	**2.82**	**3.40**	−1.02	−1.02	−1.03
ss1712304200	1	0 (NS)	—	ENSGMOG00000013560	**4.64**	**4.59**	>4	>4	>4	−1.03	−1.03	−1.02
ss1712304720	2	0 (NS)	*ZFAND2A*	ENSGMOG00000004883	**4.11**	**4.14**	>4	**3.70**	>4	−1.02	−1.02	−1.02
ss1712296776	2	832 (I)	*TMEM235*	ENSGMOG00000013183	**2.71**	**2.62**	−0.88	**1.82**	−0.31	−1.02	−1.00	−0.84
ss1712298606	2	120 (I)	*TMEM104*	ENSGMOG00000018126	**3.40**	**3.36**	**1.86**	**1.78**	**1.51**	−1.02	−1.02	−1.00
ss1712302705	2	621	*SSTR3*	ENSGMOG00000013330	**2.68**	**2.63**	−0.86	**1.48**	−0.18	−1.01	−1.00	−0.91
ss1712302707	2	1,058	*SSTR3*	ENSGMOG00000013330	**3.38**	**3.38**	−0.91	**1.31**	0.27	−1.02	−0.86	−0.88
rs119055764	2	0	*SLC25A39*	ENSGMOG00000009149	**4.31**	**4.31**	0.05	**3.00**	**2.27**	−1.02	−0.67	−0.89
ss1712302819	2	337	*PRL*	ENSGMOG00000017998	**4.55**	**4.39**	**3.70**	>4	**2.82**	−1.03	−1.01	−1.02
ss1712299603	2	202	*PDGFAB*	ENSGMOG00000013650	**3.89**	**3.80**	0.21	**2.92**	**1.30**	−1.01	−0.98	−0.92
ss1712298457	2	4,924	*PACSIN1A*	ENSGMOG00000000364	**4.39**	**4.39**	0.20	**3.11**	**2.42**	−1.02	−0.67	−0.88
ss1712304723	2	0 (NS)	*MPG*	ENSGMOG00000004225	**2.83**	**2.97**	**3.40**	**3.55**	>4	−0.98	−0.97	−1.00
ss1712300659	2	0 (NS)	*KCNH6*	ENSGMOG00000004695	**3.73**	**3.90**	−0.89	0.63	**1.29**	−1.02	0.11	−0.96
ss1712296497	2	1,603 (I)	*IQCK*	ENSGMOG00000018312	**4.62**	**4.69**	0.36	**3.22**	**3.00**	−1.02	−0.12	−0.93
ss1712296037	2	82	*HEATR2*	ENSGMOG00000004347	**3.73**	**3.62**	0.06	**3.00**	**1.02**	−1.02	−1.01	−0.90
ss1712304682	2	0 (NS)	*GRN*	ENSGMOG00000000761	**3.16**	**3.18**	−0.36	**1.84**	**1.17**	−1.02	−0.95	−0.96
ss1712298365	2	109 (I)	*COL1A1A*	ENSGMOG00000008472	**5.75**	**5.61**	−0.05	**1.80**	**1.11**	−0.94	−0.94	−0.96
ss1712296495	2	7	*CDK6*	ENSGMOG00000017917	**4.30**	**4.28**	0.20	**3.16**	**2.15**	−1.02	−0.81	−0.89
ss1712298100	2	0	*CDC27*	ENSGMOG00000013717	**4.60**	**4.32**	0.62	**2.30**	0.57	−1.01	−1.03	−0.97
ss1712301880	2	1,421	*APOL4*	ENSGMOG00000009453	**2.51**	**2.55**	0.92	**2.17**	**1.80**	−1.02	−1.02	−0.99
ss1712298561	2	43 (I)	*ADCY9*	ENSGMOG00000014968	**4.47**	**4.53**	0.20	**3.31**	**2.60**	−1.02	−0.40	−0.88
ss1712295765	3	11 (I)	*SERPING1*	ENSGMOG00000013214	**3.28**	**3.21**	**1.41**	**2.66**	**1.55**	−1.02	−1.02	−1.00
ss1712299811	3	0	*LITAF*	ENSGMOG00000014597	**2.52**	**2.79**	**3.22**	**1.78**	>4	0.70	−0.65	−1.01
ss1712304088	3	0 (NS)	*LITAF*	ENSGMOG00000014597	**2.48**	**2.71**	**2.96**	**1.77**	>4	−0.11	−0.81	−1.01
ss1712304766	3	0 (NS)	*DCHS2*	ENSGMOG00000000621	**3.08**	**3.17**	0.91	**1.81**	**2.49**	−1.02	−0.93	−1.00
ss1712304767	3	0 (NS)	*DCHS2*	ENSGMOG00000000621	**2.79**	**2.93**	0.11	**1.43**	**2.37**	−1.00	−0.81	−1.00
ss1712300044	4	3,371	*XKR6*	ENSGMOG00000011568	**3.30**	**3.22**	**2.03**	**1.96**	**1.23**	−1.01	−1.01	−1.00
ss1712297689	4	387 (I)	*PSD4*	ENSGMOG00000014165	**3.83**	**3.84**	**2.18**	**2.16**	**2.36**	−1.03	−1.02	−1.01
ss1712301665	4	4,026 (I)	*MGAT5*	ENSGMOG00000018588	**2.62**	**2.50**	**1.90**	**1.11**	−0.16	−1.00	−1.02	−1.01
ss1712297584	4	56	*CCNT2B*	ENSGMOG00000006401	**3.31**	**3.23**	**2.20**	**1.97**	**1.22**	−1.02	−1.01	−1.02
ss1712305126	7	0	*ATP1A1*	ENSGMOG00000005261	**4.12**	**4.00**	**1.34**	**1.71**	**1.18**	−1.01	−1.02	−0.99
ss1712303106	8	4,358	*ZP2L1*	ENSGMOG00000010293	**3.66**	**3.66**	>4	>4	**3.40**	−1.04	−1.02	−1.02
ss1712303105	8	2,840	*PER2*	ENSGMOG00000010256	**3.45**	**3.35**	>4	>4	**2.54**	−1.01	−1.00	−1.00
ss1712299931	8	219	*MCOLN3*	ENSGMOG00000011255	**2.91**	**2.80**	**2.01**	>4	**1.58**	−1.02	−1.01	−0.99
ss1712304112	8	0 (NS)	*MCOLN3*	ENSGMOG00000011255	**2.64**	**2.44**	**1.63**	**3.22**	−0.44	−1.02	−1.03	−0.99
ss1712302864	8	1,986	*KISS1R*	ENSGMOG00000011215	**3.84**	**4.06**	**2.60**	**2.35**	>4	−0.46	−0.58	−1.01
ss1712299109	9	0	*PSMA1*	ENSGMOG00000013869	**4.02**	**4.11**	>4	>4	>4	−0.98	−0.94	−1.01
ss1712303592	9	0 (NS)	*ITPR2*	ENSGMOG00000003496	**3.58**	**3.51**	**2.13**	0.58	0.91	−1.02	−1.02	−0.99
ss1712302672	11	0 (NS)	*TLR18*	ENSGMOG00000003793	**2.64**	**2.56**	**1.78**	**2.12**	0.69	−1.02	−1.01	−1.01
ss1712303453	14	0 (NS)	*RNF128*	ENSGMOG00000008481	**3.12**	**3.07**	>4	**2.47**	**2.62**	−1.03	−1.01	−0.98
ss1712301982	15	0	*FBXW4*	ENSGMOG00000015043	**2.73**	**2.78**	**1.68**	**3.05**	**2.89**	−1.02	−0.99	−1.00
ss1712299551	16	4,571	*KCNJ5*	ENSGMOG00000013278	**3.50**	**3.63**	**2.08**	0.55	**2.25**	−0.86	−1.01	−0.99
ss1712304231	21	0 (NS)	*SLC22A16*	ENSGMOG00000002642	**3.57**	**3.73**	0.87	**1.32**	**2.60**	−0.93	−0.81	−0.99
ss1712302811	21	188	*ESR1*	ENSGMOG00000014898	**2.45**	**2.74**	0.44	0.47	>4	0.45	0.65	−1.01

Note.—I, intron; NS, nonsynonymous. Distance to gene is in base pairs. Values in the BAYENV2 and BAYESCAN columns have negative values written in red color and $\log_{10}(BF/PO) > 1$ values are in bold text. HGNC, gene names according to the HUGO gene nomenclature committee. Gene ID refers to the annotated Atlantic cod genome, available at www.codgenome.no and Ensembl.

are between outlier SNPs on LG1 and 2 and between outlier SNPs on LG2 and 4 (supplementary table S3, Supplementary Material online). Altogether, our results correspond well with earlier work in Atlantic cod: Bradbury *et al.* (2010) and Hemmer-Hansen *et al.* (2013) separately identified outlier SNPs in LG1, 2, 7, and 12 whereas a later study by Bradbury *et al.* (2013) identified outliers in four additional LGs (LG6, 8, 13, and 22).

Outlier SNPs Indicate Genes Associated with Adaptation to Low Salinity at the Egg Stage

In the low saline environment in the Baltic Sea, cod eggs need to be highly hydrated to maintain buoyancy and avoid sinking into the deeper anoxic water layers. As a consequence, Baltic cod eggs have a neutral egg buoyancy at 14.5‰ (Nissling et al. 1994; Nissling and Westin 1997) compared with 33.0‰ for fully marine cod (Thorsen et al. 1996) and Belt Sea and Skagerrak cod (19.6‰ and 26.6‰; Nissling and Westin 1997).

The molecular and cellular mechanisms behind the extreme hydration of the oocytes are poorly understood, but recent studies have revealed an important role of aquaporins in oocyte hydration (Fabra et al. 2005; Raldúa et al. 2008; Cerdà 2009; Cerdà et al. 2013). The discovery of a teleost-specific aquaporin, *AQP1ab*, which is specifically expressed in the oocyte (Fabra et al. 2005, 2006), and the fact that hydration is inhibited when *AQP1ab* is blocked (Fabra et al. 2005, 2006) suggests a role for this AQP channel in hydration of oocytes. Further details are given in supplementary text S5, Supplementary Material online. Two outlier SNPs close to a tandem duplicate of *AQP1a* (on LG8) were associated with salinity and oxygen and were also pairwise outliers between the Baltic and the Kattegat populations, indicating that this gene could be a key player in the adaptation to low saline conditions in Atlantic cod.

We also detected nine SNPs as candidates for selection in solute carrier protein (SLC) genes (supplementary table S4, Supplementary Material online), which is significantly more than expected by chance alone (Fisher's exact test; $P < 0.0001$), belonging to seven SLC families (SLC 1, 4, 9, 22, 25, 29, and 35) where five families (SLC 4, 9, 25, 29, and 35) have previously been detected in the ovary transcripts of striped bass (Reading et al. 2012). SLCs consist of a wide range of membrane proteins that control cellular influx and efflux of solutes such as inorganic cations and anions, salts, metals, amino acids, fatty acids, and lipids (Hediger et al. 2004; He et al. 2009; Schlessinger et al. 2010). The characterization of SLC gene expression in growing oocytes and during ovarian maturation has direct importance for understanding oogenesis in teleosts (Reading et al. 2012), especially in the context of egg buoyancy in pelagophilic marine species. It is likely that a repertoire of the SLC genes in Atlantic cod acts in concert with aquaporin genes during oogenesis, supported by Bobe *et al.* (2006) showing an upregulation of *SLC26* and *AQP4* in rainbow trout (*Oncorhynchus mykiss*) during ovarian maturation. For a detailed discussion of the SLC genes, see supplementary text S5, Supplementary Material online.

During oocyte maturation, the hydrolysis of yolk proteins generates a pool of free amino acids that enables water influx into the oocyte during oocyte hydration (Finn and Fyhn 2010). Central to this process are the vitellogenins, and indeed, we do find a SNP in vitellogenin (*VTG6*), a lipid transporter, as a candidate for selection (supplementary table S4, Supplementary Material online) associated with both salinity at spawning depth and temperature. One of the outlier SNPs with strongest association to salinity and oxygen level at spawning depth and a decisive pairwise outlier in all Baltic comparisons was localized close to the zona pellucida glycoprotein-2 gene (*ZP2L1*) on LG8 (table 4). *ZP2L1* shows ovary-specific expression in zebrafish (*Danio rerio*) (Mold et al. 2001) and is liable for species-restricted binding of sperm to unfertilized eggs (Wassarman 2008). Further details are given in supplementary text S5, Supplementary Material online.

Outlier SNPs Indicate Physiological Adaptation to Low Salinity

After the egg stage, osmoregulation primarily takes place in gills, intestines and kidneys, in which the enzyme Na^+/K^+-ATPase, an enzyme that is fundamental to osmoregulation and ion exchange, is abundantly expressed. The role of this enzyme within osmoregulatory organs in teleosts is to actively pump K^+ in to the cell and Na^+ out of the cell across a concentration gradient (Geering 1990) and aiding in other transport processes associated with osmoregulatory function (McCormick 1995). We detected two different Na^+/K^+-ATPase genes (*ATP1A1* and *ATP1B1*) as candidates for selection (see supplementary text S6, Supplementary Material online). The *ATP1A1* gene (in LG7) is decisively associated with salinity and oxygen at spawning depth and surface temperature and is also a pairwise outlier only across the salinity barrier (table 4), highlighting its presumptive osmoregulatory role. This gene has been shown to have significant different gene expression in gill tissue between North Sea- and Baltic cod, suggesting adaption to local salinity conditions (Larsen et al. 2011).

A SNP (ss1712302819) located 337 bp away from the *PRL* gene, localized on LG2, is one of the SNPs with strongest environmental correlation and is a pairwise outlier in all Baltic comparisons (table 4). Prolactin is central in ion uptake and in regulating water and ion permeability in osmoregulatory surfaces in freshwater and euryhaline fish species (Manzon 2002; Sakamoto and McCormick 2006) and is known to contribute to the Na^+/K^+-ATPase regulation in for example gilthead seabream (*Sparus aurata*) (Mancera et al. 2002). It is widely accepted that *PRL* is involved in freshwater adaptation (Hirano 1986; Manzon 2002), antagonized by

growth hormone (*GH*) and insulin-like growth factor-1 (*IGF1*)—known to promote acclimatization to seawater in many teleosts (see, e.g., Sakamoto and McCormick 2006). Here, we find significant association between salinity and/or oxygen level within or close to the genes *GHRHR*, *GFBP1*, and *IGFBP3* (see supplementary text S6, Supplementary Material online), indicating central roles in salinity adaptation.

Conclusions

By combining outlier and landscape genomic analyses, we identified a set of directionally selected loci that are strongly correlated with habitat differences in salinity and oxygen in the Baltic Sea. Relative to the Baltic Sea, we detect an outlier pattern consistent with later stages of ecological divergence (Feder, Egan, et al. 2012) where most of the outlier SNPs cluster in extended genomic regions in high LD, suggesting divergence hitchhiking and the presence of genomic islands of divergence. Fixed and nearly fixed allele frequencies were detected in the Baltic population at a low frequency, suggesting that the process of diversification is ongoing. Candidate SNPs for selection was detected in all LGs, but LG2 clearly stands out with the highest number of outliers with respect to the Baltic Sea and salinity and oxygen association. A large portion of the outlier SNPs detected reside within ecologically important genes affecting egg buoyancy and general osmoregulation, and are thus likely to constitute an evolutionary response to the ecological conditions in the Baltic Sea. Such an adaptive response may contribute to a strong and effective reproductive barrier, leading to ecological speciation in the Baltic cod.

Data Access

Previously published SNPs are referred to by their rs-numbers whereas novel SNPs are referred by their ss-numbers, available in dbSNP (www.ncbi.nlm.nih.gov/SNP, last accessed May 1, 2015).

Supplementary Material

Supplementary texts S1–S6, tables S1–S4, and figures S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors are grateful to Ann-Christin Rudolphi, Yvonne Walther, and Henrik Svedäng at Swedish University of Agricultural Sciences for providing the Kattegat, Öresund, and Baltic cod samples as well as Kate Enersen and Hanne Sannæs at Institute of Marine Research (IMR, Flødevigen) for providing the North Sea cod sample. They also thank Matthew P. Kent and Mariann Arnyasi at Norwegian University of Life Sciences, CIGENE for SNP genotyping. Sequencing was provided by the Norwegian Sequencing Centre, University of Oslo. This work is part of the Cod SNP Consortium (CSC) activities—a collaboration between CIGENE, CEES, IMR, and Nofima. Funding was provided by the Research Council of Norway to K.S.J. (grant numbers 199806 and 187940) and the Centre for Marine Evolutionary Biology to C.A.

Literature Cited

Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. Nat Rev Genet. 11:697–709.

Andersen O, et al. 2009. Haemoglobin polymorphisms affect the oxygen-binding properties in Atlantic cod populations. Proc R Soc Lond B Biol Sci. 276:833–841.

André C, et al. 2011. Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity 106:270–280.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323.

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc B Biol Sci. 263:1619–1626.

Bierne N, Gagnaire P-A, David P. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. Curr Zool. 59:72–86.

Blair LM, Granka JM, Feldman MW. 2014. On the stability of the Bayenv method in assessing human SNP-environment associations. Hum Genomics. 8:1.

Bobe J, Montfort J, Nguyen T, Fostier A. 2006. Identification of new participants in the rainbow trout (*Oncorhynchus mykiss*) oocyte maturation and ovulation processes using cDNA microarrays. Reprod Biol Endocrinol. 4:39.

Boutet I, Ky CLL, Bonhomme F. 2006. A transcriptomic approach of salinity response in the euryhaline teleost, *Dicentrarchus labrax*. Gene 379:40–50.

Bradbury IR, et al. 2010. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. Proc R Soc B Biol Sci. 277:3725–3734.

Bradbury IR, et al. 2013. Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. Evol Appl. 6:450–461.

Bradbury IR, et al. 2014. Long distance linkage disequilibrium and limited hybridization suggest cryptic speciation in Atlantic cod. PLoS One 9:e106380.

Brander K. 1993. Comparison of spawning characteristics of cod (*Gadus morhua*) stocks in the North Atlantic. NAFO Sci Coun Studies. 18:13–20.

Brander K. 2005. Spawning and life history information for North Atlantic cod stocks. ICES Coop Res Rep. 274:1–152.

Bridle JR, Vines TH. 2007. Limits to evolution at range margins: when and why does adaptation fail? Trends Ecol Evol. 22:140–147.

Cano JM, Shikano T, Kuparinen A, Merila J. 2008. Genetic differentiation, effective population size and gene flow in marine fishes: implications for stock management. JIFS 5:1–10.

Case R, Hutchinson W, Hauser L, Van Oosterhout C, Carvalho G. 2005. Macro- and micro-geographic variation in pantophysin (*Pan*I) allele frequencies in NE Atlantic cod *Gadus morhua*. Mar Ecol Prog Ser. 301:267–278.

Cerdà J. 2009. Molecular pathways during marine fish egg hydration: the role of aquaporins. J Fish Biol. 75:2175–2196.

Cerdà J, Zapater C, Chauvigné F, Finn RN. 2013. Water homeostasis in the fish oocyte: new insights into the role and molecular regulation of a teleost-specific aquaporin. Fish Physiol Biochem. 39:19–27.

Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on

equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 70:155–174.
Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.
Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411–1423.
Corander J, Majander KK, Cheng L, Merilä J. 2013. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. Mol Ecol. 22:2931–2940.
de Mita S, et al. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol. 22:1383–1399.
de Villemereuil P, Frichot É, Bazin E, François O, Gaggiotti OE. 2014. Genome scan methods against more complex models: when and how much should we trust them? Mol Ecol. 23:2006–2019.
Defaveri J, Jonsson PR, Merilä J. 2013. Heterogeneous genomic differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. Evolution 67:2530–2546.
Defaveri J, Shikano T, Shimada Y, Merilä J. 2013. High degree of genetic differentiation in marine three-spined sticklebacks (*Gasterosteus aculeatus*). Mol Ecol. 22:4811–4828.
Evans DH. 2008. Teleost fish osmoregulation: what have we learned since August Krogh, Homer Smith, and Ancel Keys. Am J Physiol Regul Integr Comp Physiol. 295:R704–R713.
Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res. 10:564–567.
Fabra M, et al. 2006. Yolk proteolysis and aquaporin-1o play essential roles to regulate fish oocyte hydration during meiosis resumption. Dev Biol. 295:250–262.
Fabra M, Raldúa D, Power DM, Deen PMT, Cerdà J. 2005. Marine fish egg hydration is aquaporin-mediated. Science 307:545.
Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.
Feder JL, Gejji R, Yeaman S, Nosil P. 2012. Establishment of new mutations under divergence and genome hitchhiking. Philos Trans R Soc Lond B Biol Sci. 367:461–474.
Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64:1729–1747.
Fevolden S-E, Pogson GH. 1997. Genetic divergence at the synaptophysin (*Syp* I) locus among Norwegian coastal and north-east Arctic populations of Atlantic cod. J Fish Biol. 51:895–908.
Finn RN, Fyhn HJ. 2010. Requirement for amino acids in ontogeny of fish. Aquacult Res. 41:684–716.
Finn RN, Kristoffersen BA. 2007. Vertebrate vitellogenin gene duplication in relation to the "3R hypothesis": correlation to the pelagic egg and the oceanic radiation of teleosts. PLoS One 2:e169.
Flaxman SM, Feder JL, Nosil P. 2013. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. Evolution 67:2577–2591.
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.
Franklin IR. 1980. Evolutionary change in small populations. In: Soule ME, Wilcox BA, editors. Conservation biology—an evolutionary-ecological perspective. Sunderland (MA): Sinauer Associates. p. 135–150.
Fyhn H, Finn R, Reith M, Norberg B. 1999. Yolk protein hydrolysis and oocyte free amino acids as key features in the adaptive evolution of teleost fishes to seawater. Sarsia 84:451–456.
Gaggiotti OE, et al. 2009. Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. Evolution 63:2939–2951.
Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L. 2013. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution 67:2483–2497.
Geering K. 1990. Subunit assembly and functional maturation of Na,K-ATPase. J Membr Biol. 115:109–121.
Glaubitz JC. 2004. CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol Ecol Notes. 4:309–310.
Griffith RW. 1987. Freshwater or marine origin of the vertebrates? Comp Biochem Physiol A Mol Integr Physiol. 87:523–531.
Griffith RW. 1994. The life of the first vertebrates. Bioscience 4(6):408–417.
Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. Genetics 195:205–220.
Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372.
Hauser L, Carvalho GR. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish Fish. 9:333–362.
He L, Vasiliou K, Nebert DW. 2009. Analysis and update of the human solute carrier (SLC) gene superfamily. Hum Genomics. 3:195–206.
Hediger MA, et al. 2004. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. Pflugers Arch. 447:465–468.
Hemmer-Hansen J, et al. 2013. A genomic island linked to ecotype divergence in Atlantic cod. Mol Ecol. 22:2653–2667.
Hendry AP, Nosil P, Rieseberg LH. 2007. The speed of ecological speciation. Funct Ecol. 21:455–464.
Hirano T. 1986. The spectrum of prolactin action in teleosts. Prog Clin Biol Res. 205:53–74.
Hiroi J, McCormick SD. 2012. New insights into gill ionocyte and ion transporter function in euryhaline and diadromous fish. Respir Physiol Neurobiol. 184:257–268.
Hohenlohe PA, Bassham S, Currey M, Cresko WA. 2011. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos Trans R Soc Lond B Biol Sci. 367:395–408.
Hohenlohe PA, et al. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6:e1000862.
Hubert S, Higgins B, Borza T, Bowman S. 2010. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). BMC Genomics 11:191.
Ignatius H, Axberg S, Niemisto L, Winterhalter B. 1981. Quaternary geology of the Baltic Sea. In: Voipio A, editor. The Baltic Sea. Elsevier Oceanography Series, Elsevier Scientific, Amsterdam. p. 54–103.
Jeffreys H. 1961. Theory of probability. New-York: Oxford University Press.
Johannesson K, André C. 2006. Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. Mol Ecol. 15:2013–2029.
Jones FC, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.
Karlsen BO, et al. 2013. Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. Mol Ecol. 22:5098–5111.
Karvonen A, Seehausen O. 2012. The role of parasitism in adaptive radiations—when might parasites promote and when might they constrain ecological speciation? Int J Ecol. 2012:1–20.
Kasprzyk A. 2011. BioMart: driving a paradigm change in biological data management. Database (Oxford) 2011:bar049.
Kirkpatrick M, Kern A. 2012. Where's the money? Inversions, genes, and the hunt for genomic targets of selection. Genetics 190:1153–1155.

Lamichhaney S, et al. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. Proc Natl Acad Sci U S A. 109:19345–19350.
Larsen PF, et al. 2011. Differences in salinity tolerance and gene expression between two populations of Atlantic Cod (*Gadus morhua*) in response to salinity stress. Biochem Genet. 50:454–466.
Laverty G, Skadhauge E. 2012. Adaptation of teleosts to very high salinity. Comp Biochem Physiol A Mol Integr Physiol. 163:1–6.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
Long JA. 1995. The rise of fishes: 500 million years of evolution. Johns Hopkins University Press.
Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. Mol Ecol. 23:2178–2192.
Mancera JL, Carrión RL, del Río MDPM. 2002. Osmoregulatory action of PRL, GH, and cortisol in the gilthead seabream (*Sparus aurata* L). Gen Comp Endocrinol. 129:95–103.
Manzon LA. 2002. The role of prolactin in fish osmoregulation: a review. Gen Comp Endocrinol. 125:291–310.
McCormick SD. 1995. Hormonal control of gill Na+, K+-ATPase and chloride cell function. In: Wood CM, Shuttleworth TJ, editors. Cellular and molecular approaches to fish ionic regulation. Vol. 14. Fish physiology. Elsevier. p. 285–315.
Mellergaard S, Lang T. 1999. Diseases and parasites of Baltic cod (*Gadus morhua*) from the Mecklenburg Bight to the Estonian coast. ICES J Mar Sci. 56:164–168.
Moen T, et al. 2008. Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. BMC Genet. 9:18.
Mold DE, et al. 2001. Cluster of genes encoding the major egg envelope protein of zebrafish. Mol Reprod Dev. 58:4–14.
Nadeau NJ, et al. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. Philos Trans R Soc Lond B Biol Sci. 367:343–353.
Narum SR, Hess JE. 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. Mol Ecol Res. 11(Suppl. 1):184–194.
Nielsen EE, et al. 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat Commun. 3:851.
Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Gronkjaer P. 2003. Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. Mol Ecol. 12:1497–1508.
Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D. 2009. Population genomics of marine fishes: identifying adaptive variation in space and time. Mol Ecol. 18:3128–3150.
Nielsen EE, Hemmer-Hansen J, Poulsen NA, et al. 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evol Biol. 9:276.
Nissling A, Kryvi H, Vallin L. 1994. Variation in egg buoyancy of Baltic cod *Gadus morhua* and its implications for egg survival in prevailing conditions in the Baltic Sea. Mar Ecol Prog Ser. 110:67–74.
Nissling A, Westin L. 1997. Salinity requirements for successful spawning of Baltic and Belt Sea cod and the potential for cod stock interactions in the Baltic Sea. Mar Ecol Prog Ser. 152:261–271.
Nosil P. 2012. Ecological speciation. Oxford: Oxford University Press.
Nosil P, Feder JL. 2012. Genomic divergence during speciation: causes and consequences. Philos Trans R Soc Lond B Biol Sci. 367:332–342.
Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. Mol Ecol. 18:375–402.
Ojaveer E, Kalejs M. 2005. The impact of climate change on the adaptation of marine fish in the Baltic Sea. ICES J Mar Sci. 62:1492–1500.
Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol. 22:5983–5999.
Papakostas S, Vasemagi A, Himberg M, Primmer CR. 2014. Proteome variance differences within populations of European whitefish (*Coregonus lavaretus*) originating from contrasting salinity environments. J Proteomics. 105:144–150.
Perdiguero-Alonso D, Montero FE, Raga JA, Kostadinova A. 2008. Composition and structure of the parasite faunas of cod, *Gadus morhua* L. (Teleostei: Gadidae), in the North East Atlantic. Parasite Vector 1:23.
Pogson GH, Fevolden S-E. 2003. Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (*Pan*I) locus. Mol Ecol. 12:63–74.
Poulsen NAA, Nielsen EE, Schierup MH, Loeschcke V, Gronkjaer P. 2006. Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). Mol Ecol. 15:321–331.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81:559–575.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
R Core Team. 2012. R: A language and environment for statistical computing. 2nd ed. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://www.R-project.org/.
Raldúa D, Otero D, Fabra M, Cerdà J. 2008. Differential localization and regulation of two aquaporin-1 homologs in the intestinal epithelia of the marine teleost Sparus aurata. Am J Physiol Regul Integr Comp Physiol. 294:R993–R1003.
Reading BJ, et al. 2012. An ovary transcriptome for all maturational stages of the striped bass (*Morone saxatilis*), a highly advanced perciform fish. BMC Res Notes. 5:111.
Roesti M, Gavrilets S, Hendry AP, Salzburger W, Berner D. 2014. The genomic signature of parallel adaptation from shared genetic variation. Mol Ecol. 23:3944–3956.
Roesti M, Hendry AP, Salzburger W, Berner D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Mol Ecol. 21:2852–2862.
Romer AS, Grove BH. 1935. Environment of the early vertebrates. Am Midl Nat. 16:805–856.
Sakamoto T, McCormick SD. 2006. Prolactin and growth hormone in fish osmoregulation. Gen Comp Endocrinol. 147:24–30.
Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. Nat Rev Genet. 14:807–820.
Schlessinger A, et al. 2010. Comparison of human solute carriers. Protein Sci. 19:412–428.
Schmölcke U. 2006. Holocene vertebrate palaeontology of the Baltic Sea area. Meyniana 58:191–204.
Seehausen O, et al. 2014. Genomics and the origin of species. Nat Rev Genet. 15:176–192.
Shimada Y, Shikano T, Merilä J. 2011. A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. Mol Biol Evol. 28:181–193.
Sick K. 1961. Haemoglobin polymorphism in fishes. Nature 192:894–896.
Smadja CM, Butlin RK. 2011. A framework for comparing processes of speciation in the presence of gene flow. Mol Ecol. 20:5123–5140.
Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23:23–35.
Star B, et al. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210.

Storey JD. 2002. A direct approach to false discovery rates. J R Stat Soc B. 64:479–498.
Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol. 14:671–688.
Teacher AG, André C, Jonsson PR, Merilä J. 2013. Oceanographic connectivity and environmental correlates of genetic structuring in Atlantic herring in the Baltic Sea. Evol Appl. 6:549–567.
Therkildsen NO, Nielsen EE, Swain DP, Pedersen JS. 2010. Large effective population size and temporal genetic stability in Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. Can J Fish Aquat Sci. 67:1585–1595.
Thorsen A, Kjesbu OS, Fyhndr HJ, Solemdal P. 1996. Physiological mechanisms of buoyancy in eggs from brackish water cod. J Fish Biol. 48:457–477.
Tine M, et al. 2014. The European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun. 5:5770.
Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3:e285.
Vandamme SG, et al. 2014. Regional environmental pressure influences population differentiation in turbot (*Scophthalmus maximus*). Mol Ecol. 23:618–636.
Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol Evol. 16:381–390.
Via S. 2009. Natural selection in action during speciation. Proc Natl Acad Sci U S A. 106(Suppl. 1):9939–9946.
Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos Trans R Soc Lond B Biol Sci. 367:451–460.
Via S, Conte G, Mason-Foley C, Mills K. 2012. Localizing F_{ST} outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. Mol Ecol. 21:5546–5560.
Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol Ecol. 17:4334–4345.
Vitalis R, Dawson K, Boursot P. 2001. Interpretation of variation across marker loci as evidence of selection. Genetics 158:1811–1823.
Wang IJ, Bradburd GS. 2014. Isolation by environment. Mol Ecol. 23:5649–5662.
Ward RD, Woodwark M, Skibinski DOF. 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. J Fish Biol. 44:213–232.
Wassarman PM. 2008. Zona pellucida glycoproteins. J Biol Chem. 283:24285–24289.
Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG. 2005. Measures of human population structure show heterogeneity among genomic regions. Genome Res. 15:1468–1476.
Weir BS, Cockerham CC. 1984. Estimating *F*-Statistics for the analysis of population structure. Evolution 38:1358–1370.
Westin L, Nissling A. 1991. Effects of salinity on spermatozoa motility, percentage of fertilized eggs and egg development of Baltic cod (*Gadus morhua*), and implications for cod stock fluctuations in the Baltic. Mar Biol. 108:5–9.
Whittamore JM. 2011. Osmoregulation and epithelial water transport: lessons from the intestine of marine teleost fish. J Comp Physiol B. 182:1–39.
Wieland K. 2000. Changes in the timing of spawning of Baltic cod: possible causes and implications for recruitment. ICES J Mar Sci. 57:452–464.
Zillén L, Conley DJ, Andrén T, Andrén E, Björck S. 2008. Past occurrences of hypoxia in the Baltic Sea and the role of climate variability, environmental change and human impact. Earth-Sci Rev. 91:77–92.

Associate editor: Laurence Hurst