MOLECULAR ECOLOGY

Molecular Ecology(2012) **21**, 3686–3703

Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*)

MORTEN T. LIMBORG,*[1] SARAH J. HELYAR,†‡[1] MARK DE BRUYN,† MARTIN I. TAYLOR,† EINAR E. NIELSEN,* ROB OGDEN,§ GARY R. CARVALHO†, FPT CONSORTIUM and DORTE BEKKEVOLD*

**National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøvej 39, DK-8600 Silkeborg, Denmark, †Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Environment Centre Wales, Bangor University, Bangor LL57 2UW, UK, ‡Matís, Vínlandsleið 12, 113 Reykjavík, Iceland, §TRACE Wildlife Forensics Network, Royal Zoological Society of Scotland, Corstorphine Road, Edinburgh EH12 6TS, UK*

Abstract

High gene flow is considered the norm for most marine organisms and is expected to limit their ability to adapt to local environments. Few studies have directly compared the patterns of differentiation at neutral and selected gene loci in marine organisms. We analysed a transcriptome-derived panel of 281 SNPs in Atlantic herring (*Clupea harengus*), a highly migratory small pelagic fish, for elucidating neutral and selected genetic variation among populations and to identify candidate genes for environmental adaptation. We analysed 607 individuals from 18 spawning locations in the northeast Atlantic, including two temperature clines (5–12 °C) and two salinity clines (5–35‰). By combining genome scan and landscape genetic analyses, four genetically distinct groups of herring were identified: Baltic Sea, Baltic–North Sea transition area, North Sea/British Isles and North Atlantic; notably, samples exhibited divergent clustering patterns for neutral and selected loci. We found statistically strong evidence for divergent selection at 16 outlier loci on a global scale, and significant correlations with temperature and salinity at nine loci. On regional scales, we identified two outlier loci with parallel patterns across temperature clines and five loci associated with temperature in the North Sea/North Atlantic. Likewise, we found seven replicated outliers, of which five were significantly associated with low salinity across both salinity clines. Our results reveal a complex pattern of varying spatial genetic variation among outlier loci, likely reflecting adaptations to local environments. In addition to disclosing the fine scale of local adaptation in a highly vagile species, our data emphasize the need to preserve functionally important biodiversity.

Keywords: genome scan, haemoglobin, heat shock protein, local adaptation, salinity, single nucleotide polymorphism

Received 21 December 2011; revision received 30 March 2012; accepted 14 April 2012

Introduction

Local adaptation can evolve only if the strength of divergent selection overrides random genetic drift and the homogenizing effect of gene flow among populations (Kawecki & Ebert 2004). These premises suggest that the occurrence of local adaptation should be rare in high gene flow species such as many marine organisms (Palumbi 1994; Conover *et al.* 2006). In contrast, large effective population sizes (N_e) should enhance response to selection, and local selective pressures may be substantial considering the often immense environmental heterogeneity experienced by widely distributed marine species. A recent simulation-based study showed that

Correspondence: Morten T. Limborg, Fax: +45 35 88 31 50; E-mail: mol@aqua.dtu.dk

[1]Joint first authors.

even in the face of considerable gene flow, environmental heterogeneity may cause disruptive selection and result in local adaptation (Yeaman & Whitlock 2011). Accordingly, expectations are that genes and linked regions under the influence of divergent selection will show elevated differentiation, in comparison with selectively 'neutral' gene regions. Until now, genomic studies of high gene flow marine fish have been mostly restricted to Atlantic cod (*Gadus morhua*) (Moen *et al.* 2008; Nielsen *et al.* 2009b; Bradbury *et al.* 2010), while for other fishes, inference of genic selection has often been made from a single or few candidate genes (Hemmer-Hansen *et al.* 2007a; Gaggiotti *et al.* 2009; Larmuseau *et al.* 2009).

The task of identifying signatures of natural selection in nonmodel species has been constrained by often limited numbers of (usually) neutral genetic markers (Hauser & Seeb 2008). Next-generation sequencing (NGS) technologies have facilitated the development of large transcriptome-derived marker panels, effectively increasing the chance of detecting natural selection by studying functional genetic variation, which is expected to be more directly affected by natural selection (Allendorf *et al.* 2010). The increased genomic coverage further improves the chance of detecting loci affected by divergent selection from neutrally evolving sites by applying genome scan approaches (Beaumont 2005; Storz 2005). The adaptive significance of actual outlier loci is often elusive because they may not be the direct target of selection but rather exhibit hitchhiking with genes under selection (Maynard Smith & Haigh 1974). However, the combination of insights from known gene functions, landscape effects (Manel *et al.* 2003), replicated patterns across independent environmental clines (Schmidt *et al.* 2008) and previous findings provides stronger evidence for adaptive roles of outlier loci (Vasemägi & Primmer 2005).

Despite a predominant picture of weak population structure in most marine fishes (Ward *et al.* 1994), genomic regions under divergent selection may be more prevalent than hitherto anticipated (Nielsen *et al.* 2009a). In the present study, we use the Atlantic herring (*Clupea harengus*; hereafter 'herring') as a model to investigate spatially explicit genomic variation in a marine organism characterized by high gene flow and large effective population size (N_e). Herring is a small, highly migratory pelagic fish distributed throughout heterogeneous environments in large parts of the North Atlantic. Local populations exhibit large differences in demographic and life history parameters including growth, spawning season and migratory behaviour (Iles & Sinclair 1982; Aro 1989). Outside spawning seasons, several populations undergo long-distance migrations to communal feeding areas (e.g. Ruzzante *et al.* 2006), suggesting ample opportunities for dispersal and gene flow. In some areas, a combination of high gene flow and large N_e among herring populations presumably impedes genetic detection of local demes using neutral markers (Mariani *et al.* 2005). However, for other geographical regions, significant genetic structuring is evident, especially across the strong environmental cline separating the fully marine North Sea from the brackish Baltic Sea (Bekkevold *et al.* 2005) as well as weak, but statistically significant, patterns within the Baltic Sea (Jørgensen *et al.* 2005). More recently, signatures of selection have also been demonstrated in herring (Larsson *et al.* 2007; Gaggiotti *et al.* 2009; Andre *et al.* 2011), but these studies focus on comparisons between North Sea and Baltic Sea herring for a single microsatellite locus. Thus, despite many population genetic studies on herring, the geographical scale and pattern of adaptive divergence at genomic levels remains largely unknown.

We investigated the spatial and genomic scales at which herring populations are likely to exhibit adaptation to local environments. We conducted comprehensive sampling of herring spawning populations throughout the northeastern Atlantic, and across several environmental gradients, and applied a statistical genome scan approach to transcriptome-derived single-nucleotide polymorphism (SNP) markers. To assess the robustness of loci under selection, we use two different 'outlier tests' for identifying gene regions exhibiting statistical evidence of predominantly either neutral or divergent selection processes. Furthermore, we use a complementary 'landscape genetics' approach to identify loci under divergent selection in relation to key environmental parameters. Findings are discussed in relation to the prospects and significance of detecting functional biodiversity in high gene flow taxa through exploring genes subject to local adaptive evolution in the oceans.

Materials and methods

Samples

Twenty-one samples were collected from scientific surveys and commercial fishing vessels, representing 18 locations spanning the majority of the species' east Atlantic distribution (Fig. 1). Three samples represented temporal (range = 6–10 years) replicates within locations. Populations were targeted during the spawning season at known spawning grounds and mainly comprised spawning (ripe and running) individuals. Samples spanned latitudinal clines (reflecting temperature) both in the North Sea/North Atlantic and in the Baltic Sea (Fig. 1). Samples also covered longitudinal clines (corresponding with two low-salinity environments): one

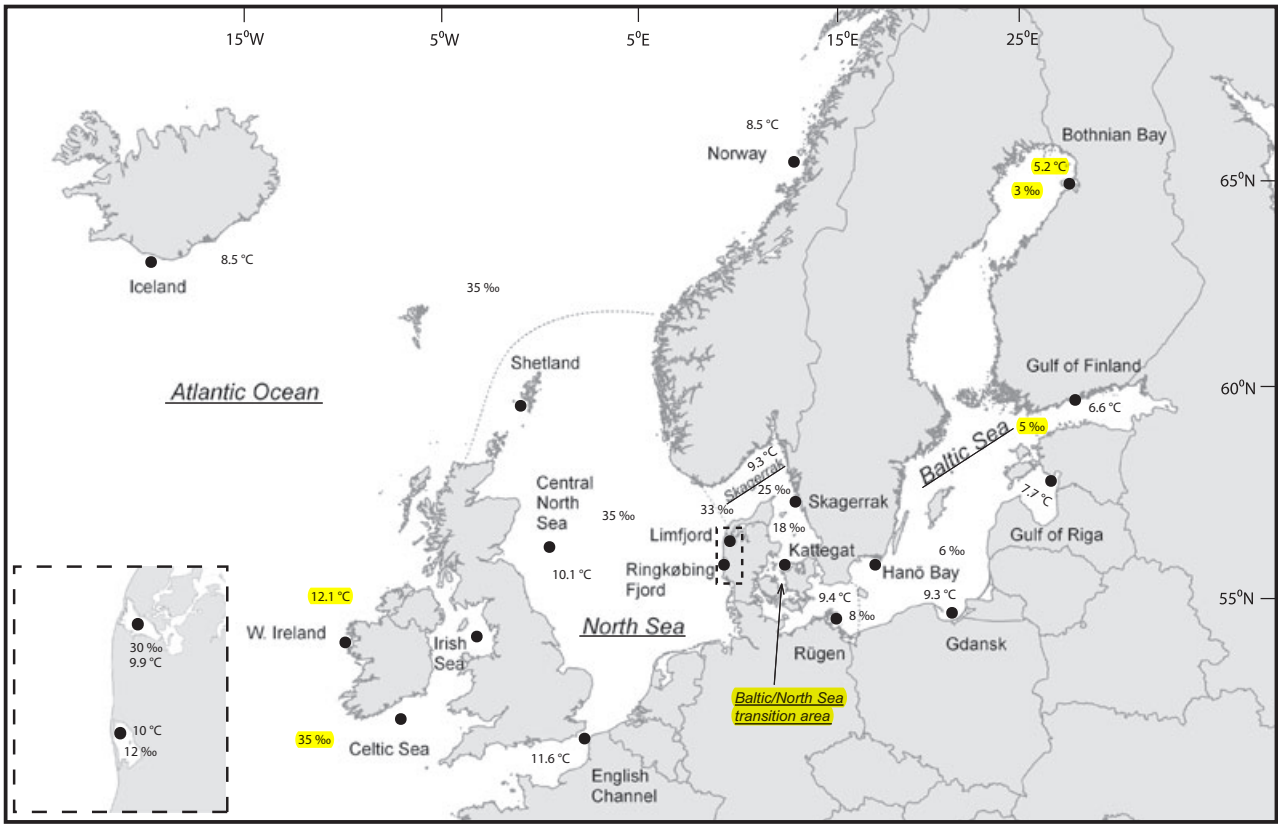

Fig. 1 Distribution of sampled locations. Average annual surface temperature (°C) and salinities (‰) are given throughout the distribution, and major regional areas are denoted in italics.

spanning the North Sea/British Isles into the Baltic Sea which is an isolated brackish sea only receiving saline waters from the North Sea through the narrow Danish Straits and one going from the North Sea/British Isles with high-salinity coastal locations, changing to more brackish spawning locations in the Ringkøbing Fjord draining into the eastern North Sea and separated from the low-saline Baltic Sea cline (Fig. 1). Spawning times of herring differ among local populations (Cushing 1967) and collections reflect this, as spring-, autumn- and winter-spawning populations are all represented (Table 1).

Molecular analyses and genotyping

DNA was extracted from gill, muscle or fin tissue stored in 96% ethanol using the E.Z.N.A. Tissue DNA kit (Omega Bio-Tek, Norcross, GA, USA) following the manufacturer's protocol. A NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc.) was used to ensure adequate quality and quantity of DNA prior to genotyping. A total of 762 individuals were screened for a panel of 310 SNPs (Table S1, Supporting information). These SNPs were selected from a larger collection of 578 SNPs (Helyar *et al.* 2012) based on robust clustering of genotypes (see below) and minimization of the number of loci affected by very strong linkage disequilibrium. Genotyping was performed using a custom Illumina Golden Gate® Assay (Fan *et al.* 2003) in Sentrix Array Matrix (SAM) format on the iScan platform. SNPs were developed from an ascertainment panel of eight individuals representing all major geographical regions studied here (i.e. Baltic Sea, English Channel, North Sea and North Atlantic) (Helyar *et al.* 2012), thus minimizing ascertainment bias (Rosenblum & Novembre 2007). Overall, 29 SNPs were discarded due to ambiguous clustering (Table S1, Supporting information), and of the remaining 281 SNPs, 70% were annotated using BLASTN (NCBI) (Helyar *et al.* 2012). Genotyping data were visualized and analysed using the GENOMESTUDIO Data Analysis Software package (1.0.2.20706; Illumina Inc.). One specimen was independently re-genotyped 12 times, which allowed the estimation of an overall genotyping error of 1.57% across all loci and samples. Specimens with low call rates (<90% loci genotyped) were discarded, leading to 607 individuals genotyped for a total of 281 SNPs (Table S1, Supporting information) in 21 sample collections (n = 17–39).

Table 1 Sample information including peak spawning times for the populations sampled

Geographical region	Sample location	Year	Month	Peak spawning	n	Latitude	Longitude	Genetic cluster*	H_e	H_o	Environmental conditions†			
											SST	SST. spawn	PSU	PSU.spawn
North Atlantic	Norway[1,2]	2009	Sep	Mar	31	65.54‡	11.26‡	North Atlantic	0.30	0.30	8.5	6.0	33.7	33.9
	Iceland[1,2]	2009	Jul	May–Sep	34	63.62	−19.62	North Atlantic	0.30	0.30	8.5	10.9	34.9	34.5
North Sea/British Isles	Shetland[1,2]	2009	Aug	Aug	34	60.35	−2.72	North Sea/British Isles	0.30	0.30	10.5	11.7	35.2	35.2
	W. Ireland[1,4,5]	2004	Nov	Nov	28	53.90	−10.36	North Sea/British Isles	0.31	0.32	12.1	13.0	35.2	35.2
	Celtic Sea[1]	2008	Oct	Oct	39	51.24	−8.26	North Sea/British Isles	0.30	0.30	12.4	12.8	34.9	35.0
	Irish Sea[1,4,5]	2009	Sep	Sep	36	54.03	−4.07	North Sea/British Isles	0.30	0.31	10.8	13.3	33.5	33.6
	English Channel[1]	1999	Nov	Nov–Jan	17	50.81	1.57	North Sea/British Isles	0.30	0.31	11.6	8.5	34.9	34.9
	English Channel[1,2]	2009	Jan	Nov–Jan	36	50.81	1.57	North Sea/British Isles	0.30	0.31	11.6	8.5	34.9	34.9
	Central North Sea[1,2,4,5]	2009	Aug	Aug	30	56.43	0.20	North Sea/British Isles	0.30	0.30	10.1	11.4	34.9	34.9
	Ringkøbing Fjord[1,5]	2009	Apr	Apr	33	55.97	8.24	B/NS transition area	0.30	0.31	10.0	12.5	9.1	9.0
	Limfjord[1]	2009	Apr	Apr	33	56.60	8.35	B/NS transition area	0.30	0.31	9.9	11.5	30.8	30.7
B/NS transition area	Skagerrak[1,4]	2009	Mar	Apr	36	57.40	11.40	North Sea/British Isles	0.30	0.31	9.3	5.6	24.9	23.8
	Kattegat[1,4]	2003	Apr	Apr	23	55.73	11.37	B/NS transition area	0.29	0.31	9.4	6.0	18.7	17.1
	Rügen[1]	2003	Apr	Mar–Apr	19	54.21	13.62	B/NS transition area	0.30	0.31	9.4	5.3	8.0	8.1
	Rügen[1,4]	2009	Mar	Mar–Apr	36	54.21	13.62	B/NS transition area	0.30	0.30	9.4	5.3	8.0	8.1
Baltic Sea	Hanö Bay[1,3]	2002	Apr	Apr	24	55.57	15.18	Baltic Sea	0.31	0.32	8.6	7.4	7.5	7.5
	Gdansk[1,3,4]	2009	Mar	Mar	17	54.37	19.67	Baltic Sea	0.27	0.30	9.3	4.5	7.3	7.2
	Gulf of Riga[1]	2002	May	May–Jun	17	57.83	22.83	Baltic Sea	0.29	0.32	7.7	6.4	5.5	5.5
	Gulf of Riga[1,3,4]	2008	Jun	May–Jun	27	57.83	22.83	Baltic Sea	0.27	0.29	7.7	6.4	5.5	5.5
	Gulf of Finland[1,3]	2009	May	May	24	60.40	26.70	Baltic Sea	0.27	0.28	6.6	4.8	5.5	5.3
	Bothnian Bay[1,3]	2009	Jun	Jun	33	65.05	24.58	Baltic Sea	0.30	0.30	5.2	8.1	2.9	2.6

Numbers after sample location names denote samples included in (1) Global and regional analyses, (2) North Sea/North Atlantic latitudinal cline, (3) Baltic Sea latitudinal cline, (4) North Sea/British Isles—Baltic Sea longitudinal cline and (5) North Sea/British Isles—Ringkøbing Fjord longitudinal cline. Also shown for each sample is the most likely of four clusters as inferred from STRUCTURE analysis (B/NS = Baltic Sea/North Sea). Expected (H_e) and observed (H_o) heterozygosities are shown for each population. Environmental conditions used in landscape genetic analyses are shown for each sample location as SST (annual mean sea surface temperature), SST.spawn (spawning period mean sea surface temperature), PSU (sea surface salinity) and PSU.spawn (spawning period mean sea surface salinity) (See text for more details).

*Based on STRUCTURE analysis including the 'full' marker set and for $K = 4$ (see text). For each sample, membership to the genetic cluster receiving highest support is shown, disregarding that some samples showed high levels of admixture (e.g. Skagerrak, Fig. 3).

†Temperature and salinity data sources for Ringkøbing Fjord and Limfjord (http://www.dmu.dk/vand/havmiljoe/mads/ctd/data) and all other samples (http://www.ices.dk/ocean/data/surface/surface.htm).

‡This population was sampled off the major spawning ground at: 70.06 N and 16.90 E; however, preliminary otolith analyses showed that this sample represents the major Norwegian spring-spawning herring population (FishPopTrace consortium), and we use coordinates for the major spawning area in the landscape genetics analyses.

Summary statistics

Within each population, loci were tested for departure from Hardy–Weinberg proportions (HWE) using ARLEQUIN 3.5 (Excoffier & Lischer 2010) with a Markov Chain (MC) of length 10^6 and 100 000 dememorizations. A false discovery rate (FDR) was calculated to correct for multiple testing using the approach by Benjamini & Yekutieli (2001). Linkage disequilibrium was tested for each marker pair in all samples with GENEPOP 4.0 (Raymond & Rousset 1995) (10 000 dememorizations, 100 batches and 5000 iterations), and the results were corrected for multiple testing as above. For each population, estimates of expected (H_e) and observed (H_o) heterozygosities were obtained using GENALEX 6.4 (Peakall & Smouse 2006).

Outlier analyses

Two independent methods were used to identify putative loci under selection. ARLEQUIN v3.5 (Excoffier & Lischer 2010) utilizes coalescent simulations to generate a null distribution of *F*-statistics, with *P*-values conditioned on observed levels of heterozygosities across loci (Excoffier *et al.* 2009). Excoffier *et al.* (2009) demonstrated that the hierarchical island model produces fewer false positives than the finite island model for species exhibiting spatial population structure. For comparison, we tested both models. The hierarchical island model was implemented by grouping population samples according to the genetic clustering analyses (Table 1), as follows: (i) the Baltic Sea, (ii) the Baltic/North Sea transition area, (iii) the North Sea/British Isles and (iv) the North Atlantic. For all analyses, the settings were 10 000 simulations, 100 demes per group, and 10 groups. Loci that fell outside the 95% quantile were regarded as candidates for selection. BAYESCAN v2.01 (Foll & Gaggiotti 2008) measures the discord between global and population-specific allele frequencies (based on F_{ST} coefficients). While this method does not take into account the population structure, simulations have shown BAYESCAN to have lower type I and II errors than ARLEQUIN (Narum & Hess 2011). Log10 values of the posterior odds (PO) >0.5 and 2.0 were taken as 'substantial' and 'decisive' evidence for selection (Jeffreys 1961). An advantage of the posterior probability approach is that it directly allows for control of the FDR; here, the FDR was set at 0.05 and 0.01, adjusting the log10(PO) significance thresholds corresponding to the 0.5 and 2.0 values considered before correction. To compare global and regionally based signatures of selection, we performed global (21 population samples) genome scans using both software packages as detailed above. Based on the combined inference from these global genome scans, each SNP was categorized as either an 'outlier' (if it came out as such with either one or both of the programs) or 'neutral' (if it showed no indication of outlier behaviour with either program). We then constructed two data sets: one including both outlier and neutrally behaving SNP loci (referred to as the 'full' marker set) and one where all loci detected as outliers in global tests were removed (referred to as 'neutral' marker set). To further increase support for potential outliers in relation to environmental adaptation, we performed local genome scans to focus on the two separate regional temperature clines and two salinity clines (see Table 1 for samples included).

Population structure

For the 'neutral' marker set, temporal stability between replicates for three locations (Table 1) was assessed through pairwise F_{ST} analyses (following Weir & Cockerham 1984) using the *Fstat* function implemented in GENELAND (Guillot *et al.* 2005) conducted in the program *R* (http://cran.r-project.org). Temporal samples not exhibiting significant differentiation ($\alpha = 0.05$) were pooled within locations for subsequent analyses. F_{ST} was computed between all pairs of samples using both 'neutral' and 'full' marker sets. For all comparisons, significance was tested by permuting individuals 10 000 times among samples followed by correction for multiple tests using the FDR ($\alpha = 0.05$) according to Benjamini & Yekutieli (2001). The statistical power of the 'neutral' marker set for detecting genetic differentiation was assessed using POWSIM (Ryman & Palm 2006). By defining a given effective population size (N_e), POWSIM simulates genetic drift within two independent populations for t generations. N_e was set to 10 000 (the maximum allowed) and t varied among simulations to obtain a range of known F_{ST} values (0.00–0.02) between two hypothetical populations. Hereafter, 40 individuals were sampled from each population, and the null hypothesis of genetic homogeneity between samples was tested using a chi-square test. Repeating this procedure 1000 times allowed the assessment of the statistical power as the proportion of significant outcomes for each level of F_{ST}.

To infer the number of major genetic clusters, we used the Bayesian MCMC clustering approach implemented in STRUCTURE 2.3.1 (Pritchard *et al.* 2000). This model clusters all individuals into a predefined number of clusters (K) by minimizing overall deviation from HW and linkage equilibrium within clusters. Considering previous findings of high levels of gene flow in herring, we used the admixture model with correlated allele frequencies to reflect the most likely pattern of population connectivity. Also, due to the sampling

design, we allowed the model to include prior information on sampling location (Hubisz *et al.* 2009). Ten independent trials were run for each predefined K value, with K = 1–10. We used a burn-in of 10 000 iterations followed by 100 000 MCMC repetitions, and consistency of the three most likely K estimates was confirmed by longer chains of 100 000 burn-in and 500 000 final repetitions. In order to identify the most likely number of genetic clusters, also considering a sound biological interpretation, we initially considered both raw probability values of $\ln P(X \mid K)$ given by the program, and the ΔK estimate (Evanno *et al.* 2005). Where two models with consecutive K values could not be statistically distinguished, we performed hierarchical AMOVA using the locus-by-locus approach and 10 000 permutations in ARLEQUIN 3.5 (Excoffier & Lischer 2010), using both the 'neutral' and 'full' marker sets.

In addition, barplots of individual admixture proportions were visually inspected to infer the biologically most meaningful value of K. For example, if increasing K by one simply added a new cluster equally represented by all individuals in the data, as opposed to the break-up of existing clusters forming a new more or less admixed cluster, the lower value of K would be considered more biologically realistic.

Environmental associations with genetic variation

To test for association between specific gene regions and environmental or landscape parameters, we applied the Bayesian approach implemented in BAYENV (Coop *et al.* 2010). This approach takes into account the effect of underlying (neutral) population structure by first estimating a covariance matrix based on neutral markers, which is subsequently used to control for demographic variation when testing landscape- and locus-specific correlations in a Bayesian framework (see Coop *et al.* 2010). For this, we estimated a neutral covariance matrix based on the 'neutral' marker set. Results are given as a Bayes factor (BF) for each landscape variable and SNP locus correlation. This BF represents a ratio of the posterior likelihoods of a model where the landscape parameter has a significant effect on the locus, over an alternative model with no effect of the tested variable. We considered log10(BF) values above 1.5 as 'very strong' evidence (Jeffreys 1961) for an effect of the tested landscape/environmental factor (or any correlated factors) on the observed SNP allele distribution. The following landscape/environmental parameters were considered: (i) latitude, (ii) longitude, (iii) mean annual surface salinity, (iv) mean annual surface temperature, (v) mean spawning period surface salinity and (vi) mean spawning period surface temperature (Table 1). The latter two parameters were tested based on the assumption that mortality selection is expected to be most important during the egg (7–14 days) and larval (*c.* 2 months) phases, when natural mortality is highest (Dahlberg 1979). Estimates for temperature and salinity were calculated as the mean value over periods ranging from 20 to 120 years (depending on data availability) for all months (annual means) or for the 3 months following the midpoint of the spawning period (data and sources are listed in Table 1). As all pairwise genetic comparisons between temporally replicated samples were nonsignificant (see Results), all within-population genotypes across years were pooled for these analyses. To test for relationships between selected genetic variation and environment across both global and local scales, we performed a global analysis including all 18 samples as well as four regionally based analyses (regions defined as per regional genome scans; Table 1). To further rule out potential false positive correlations resulting from covarying isolation-by-distance (IBD) effects, we performed partial Mantel tests (Legendre & Legendre 1998) of locus-specific pairwise F_{ST} matrices and environmental distances for all loci correlating with temperature and salinity while controlling for geographical distance (shortest waterway) using the NCF (spatial nonparametric covariance functions) package in R (http://cran.r-project.org/web/packages/ncf/index.html) and running 1000 simulations to test for significance.

Results

Summary statistics

After removing loci that were monomorphic within a population, a total of 5541 tests for deviation from HWE were performed across all samples. Prior to and following correction for multiple testing (FDR = 5%), 159 (2.6%) and 19 (0.3%) tests were significant, respectively. The latter category included 19 different markers and 12 samples with a maximum of three significant tests for the same sample. Only one significant test involved a SNP likely to be affected by selection. A total of 726 865 tests for LD within samples were performed, of which 1309 tests remained significant ($P < 0.05$) after correction for multiple testing. These ranged between 29 and 95 significant outcomes (of ~40 000 possible SNP pairs) within samples, and no SNP pair was significant in more than three of 21 samples. We thus do not expect LD or departure from HWE across the 281 loci to affect downstream analyses. Levels of H_e and H_o were similar and ranged between 0.27 and 0.32 with no clear spatial pattern of regionally differentiated levels of genetic diversity (Table 1).

Outlier tests for selected vs. neutral variation

For the global analyses, overall 16 (5.7%) and 14 (5.0%) outlier loci were suggested to be under divergent selection at the 5% and 1% thresholds, respectively, across the two genome scan approaches. All 16 loci were detected by BAYESCAN, while 15 of these were also detected by ARLEQUIN (Fig. 2a, Table 2). A comparison of genotyping error across all loci (1.57%) vs. the 16 global outlier loci (1.08%) affirmed that outliers were not expected to suffer increased genotyping error rates. No outliers for balancing selection were observed (Fig. 2a). Based on the combined inference from global outlier tests, the 'full' marker set included all loci (16 outlier loci and 265 'neutral'), whereas the 'neutral' marker set included 265 putatively neutral loci. Genome scans comprising only samples across each of the two temperature clines and each of the two salinity clines

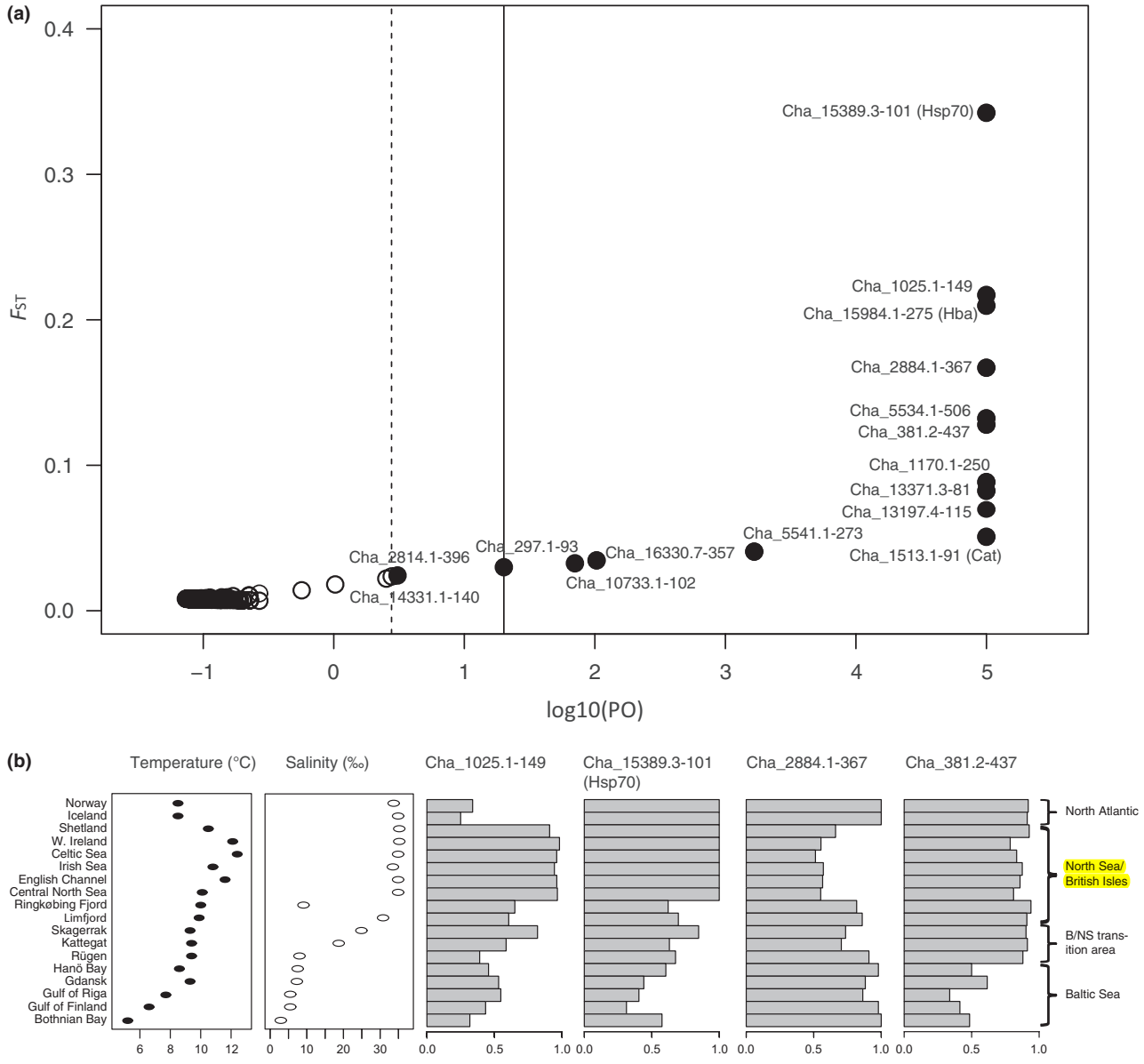

Fig. 2 Global genome scan and allele frequency plots of candidate SNPs for divergent selection. (a) Result from the global genome scan analysis using the approach by Foll & Gaggiotti (2008). Broken and solid lines represent the 5% and 1% log10(PO) thresholds for being under selection after correction with false discovery rate. Solid black circles denote SNPs that were also significant outliers using a hierarchical genome scan (Excoffier *et al.* 2009). All global outliers above the 5% level are identified by their SNP name. (b) Sample-specific values for annual average temperature, salinity and allele frequencies for a subset of four global outliers. Samples are ordered by geographical connectivity. B/NS = Baltic Sea/North Sea.

Table 2 Global results for selection including all samples

Outlier results			BAYENV results				
ARLEQUIN	BAYESCAN	SNP (Cha_)	Lat	Long	SST	SST.spawn	PSU
**	**	1025.1-149	**	**	**	(**)	(**)
*	**	10733.1-102					
**	**	1170.1-250	**		(**)		
**	**	13197.4-115		*			
**	**	13371.3-81		*	**		**
	*	14331.1-140					
**	**	1513.1-91 (Cat)		*			**
**	**	15389.3-101 (Hsp70)		**	**	(**)	**
**	**	15984.1-275 (Hba)		**	*	(**)	**
**	**	16330.7-357					
*	*	2814.1-396					
**	**	2884.1-367	**	*	**	(*)	(**)
*	**	297.1-93					
**	**	381.2-437		**	(**)		**
**	**	5534.1-506	**		**		
**	**	5541.1-273					

Overview of results from ARLEQUIN and BAYESCAN outlier tests (left of SNP names) together with all landscape association (BAYENV) results. The first two columns left of the SNP names show all detected outliers where * and ** denote outliers with $P < 0.05$ or 0.01 for the ARLEQUIN analysis. BAYESCAN outliers were detected with false discovery rates of 5% (*) and 1% (**). Statistical inference of correlations between SNPs and landscape parameters are given for relationships with log10(BF) = 1.5–2.0 (*) and log10(BF) > 2.0 (**). Corresponding correlations from partial Mantel tests that became nonsignificant when controlling for geographical distance are shown in parentheses. No relationships between neutral SNPs and tested landscape parameters had log10(BF) > 1.5 in the BAYENV tests (not shown). Lat, latitude; Long, longitude; SST, annual mean temperature, SST spawn, average spawning season temperature; PSU, annual mean salinity.

revealed between 11 and 14 outliers, with the majority only detected by ARLEQUIN (Tables 3–4). All 14 global outliers detected at the 1% thresholds were also detected in one or more regional tests. Two loci were outliers across both temperature clines (Table 3), and seven loci were outliers across both salinity clines (Table 4). In total, 39 loci were identified as outliers in one or more analyses (Tables 2–4), and of these, 28 were annotated (Helyar *et al.* 2012).

Population structure

Clustering analysis based on the 'neutral' marker set suggested a model of $K = 3$ as the statistically most likely ($\ln(K) = -145\,231 \pm 30$ SD). Most individual genotypes indicated admixture between clusters, but overall the three identified clusters corresponded with (i) the Baltic Sea, (ii) the Baltic/North Sea transition area and (iii) the North Sea/British Isles/North Atlantic (Fig. 3a). Setting $K = 4$ decreased the probability to $\ln(K) = -145\,603 \pm 43$ SD. Here, three samples from the North Sea/North Atlantic (Shetland, Norway and Iceland) showed a trend of being admixed between a North Sea/British Isles and a fourth, North Atlantic cluster (Fig. 3a). AMOVA tests for $K = 3$ and 4 revealed similar levels of variation among groups (Table 5). When using the 'full' marker set, a model with $K = 4$ was suggested as the single most likely scenario ($\ln(K) = -153\,959 \pm 43$ SD). Again, most individuals exhibited admixed genotypes, but the four clusters overall corresponded with (i) the Baltic, (ii) the Baltic/North Sea transition area, (iii) the North Sea/British Isles and (iv) the North Atlantic (Fig. 3b). The four clusters were further supported by the AMOVA revealing increased levels of variation among four groups compared to three (Table 5). For a $K = 4$ model, the two marker sets were largely similar in defining a total of four groups, and we consider this as the most likely number of groups detectable with our data. However, in comparison, the 'full' marker set was able to more clearly define the North Atlantic cluster as well as identifying admixture between the Baltic and the North Atlantic clusters in the Bothnian Bay sample (Fig. 3).

Statistical power for detecting genetic differentiation among local populations with neutral, bi-allelic markers was high (>0.89 for detecting differentiation at $F_{ST} \geq 0.005$), based on the POWSIM analysis. Genetic differentiation between three temporal within-location replicates (representing three major clusters) inferred for both the 'neutral' and 'full' marker sets was low and in all cases nonsignificant ($F_{ST} = -0.002$ to 0.005, $P > 0.05$), suggesting that the identified structure is temporally

Table 3 Regional results for selection across two latitudinal clines (reflecting temperature gradients) in the North Sea/North Atlantic (five samples) and in the Baltic Sea (five samples), respectively

North Sea/North Atlantic							
(Norway, Iceland, Shetland, Central North Sea, English Channel)							
Outlier results			BAYENV results				
ARLEQUIN	BAYESCAN	SNP (Cha_)	Lat	Long	SST	SST.spawn	PSU
**	**	1025.1-149	**		(**)		
**	**	10733.1-102					
**	**	1170.1-250	**		(*)		
**		11922.3-225					
	**	13197.4-115		*			
**	**	13371.3-81			(**)		
*		13427.1-146					
*		16060.1-279					
**	**	2884.1-367	**		(**)		
**	*	297.1-93					
**	**	462.3-102					
**	**	5534.1-506	**		**		
Baltic Sea							
(Bothnian Bay, Gulf of Finland, Gulf of Riga, Hanö Bay, Gdansk)							
Outlier results			BAYENV results				
ARLEQUIN	BAYESCAN	SNP (Cha_)	Lat	Long	SST	SST.spawn	PSU
**		10428.2-348					
**		11521.1-298					
**		12888.1-297					
*		13197.3-287					
*		13197.4-115					
*		15056.1-166					
*		15389.3-101 (Hsp70)					
**		1567.1-307					
**		15898.2-568					
**		160.1-805					
**	*	2884.1-367					
**		535.2-394					
**		5625.1-135					
**		9634.1-256					

Results are presented as in Table 2, and underlined SNPs represent replicated outliers in both transects.

stable at least within the time frame studied. Across all pairs of samples, F_{ST} estimates based on 265 neutral markers were generally low (F_{ST} = −0.002–0.012), while estimates based on all 281 SNPs were slightly higher (F_{ST} = −0.002 to 0.028; Table S2, Supporting information). Comparisons between samples from the four major population groups were generally significant for both marker sets, whereas within-group comparisons were often low (all F_{ST} < 0.008) and nonsignificant (Table S2, Supporting information). Exceptions to overall within-group homogeneity were mostly identified for the 'full' marker set and included (i) differentiation between the Gulf of Finland and Hanö Bay within the Baltic group, (ii) samples from Rügen exhibiting differentiation from the Skagerrak and Kattegat for the Baltic/North Sea transition area group, (iii) Shetland being differentiated from the Irish Sea, Celtic Sea and English Channel samples in the North Sea/British Isles group, (iv) Central North Sea and Irish Sea samples within the North Sea/British Isles group and (v) Norway being differentiated from Iceland in the North Atlantic group (Table S2, Supporting information). We cannot rule out that some loci affected by selec-

Table 4 Regional results for selection across longitudinal clines including two low-salinity environments from the North Sea/British Isles into the Baltic Sea (eight samples) and Ringkøbing Fjord (four samples), respectively.

North Sea/British Isles—Baltic Sea							
(W. Ireland, Irish Sea, Central North Sea, Skagerrak, Kattegat, Rügen, Gdansk, Gulf of Riga)							
Outlier results			BAYENV results				
ARLEQUIN	BAYESCAN	SNP (Cha_)	Lat	Long	SST	SST spawn	PSU
**	**	**1025.1-149**		**	**	**	**
*		1170.1-250					
**	**	**13371.3-81**		*	(*)		**
*		14067.1-259					
**	**	**1513.1-91 (Cat)**		*			*
**	**	**15389.3-101 (Hsp70)**		**	(**)	(**)	**
**	**	**15984.1-275 (Hba)**		**	**	**	**
**	*	16330.7-357					
**	**	2884.1-367		**		(**)	**
**	**	381.2-437					
**		3888.1-826					
*		688.1-238					
*		693.2-263					
North Sea/British Isles—Ringkøbing Fjord							
(W. Ireland, Irish Sea, Central North Sea, Ringkøbing Fjord)							
Outlier results			BAYENV results				
ARLEQUIN	BAYESCAN	SNP (Cha_)	Lat	Long	SST	SST spawn	PSU
**	**	**1025.1-149**		*			(**)
**		10733.1-102					
**		**13371.3-81**					(*)
**		**1513.1-91 (Cat)**					(*)
**	**	**15389.3-101 (Hsp70)**		**	(**)		**
**		15964.1-332					
**	**	**15984.1-275 (Hba)**		**			(**)
**		2884.1-367					
*		318.1-301					
**		5541.1-273					
*		693.2-263					
**		7456.1-168					
*		8760.1-243					

Results are presented as in Table 3, and SNPs in boldface show correlations with one or more similar landscape parameters across both clines.

tion were included in the 'neutral' marker set, because only global outliers were excluded in defining neutral markers, despite the prevalence of other loci exhibiting regional outlier status. However, here we mainly apply an intercluster comparison of neutral and selected data, and combined with the finding of mainly weak and non-significant neutral genetic differentiation within clusters (Table S2, Supporting information), we argue that the applied approach remains useful.

Environmental associations

The global test for correlations between individual loci and six landscape variables revealed significant associations with one or more landscape variables for ten of the 16 global outlier loci (Table 2). None of the 265 neutral loci were correlated with any of the tested variables. Outlier loci showed distinct allele frequency distributions potentially reflecting the effects of spatially

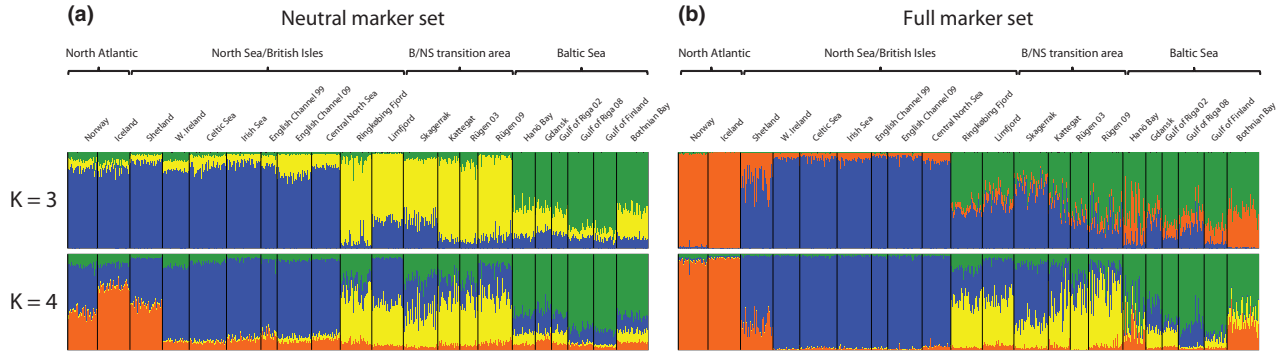

Fig. 3 Results from clustering analyses for two data sets either (a) excluding global outlier loci (neutral marker set) or (b) including all loci (full marker set) for *K* values of 3 and 4. Samples are ordered to reflect geographical connectivity illustrated by the top brackets representing the geographical origin of each sample irrespective of genetic composition. Colours designate corresponding genetic clusters between data sets: orange = North Atlantic, blue = North Sea/British Isles, yellow = Baltic/North Sea transition area and green = Baltic Sea.

Table 5 AMOVA based on neutral loci (neutral marker set) and all loci (full marker set). For each data set, an AMOVA was performed for $K = 3$ and 4 following clustering results from STRUCTURE analyses (see text for details). All variance levels are highly significant ($P < 0.001$)

		% variation	
Data set	Hierarchical level	$K = 3$	$K = 4$
Neutral marker set	Among groups (F_{CT})	0.31	0.32
	Among populations within groups (F_{SC})	0.19	0.16
	Within populations (F_{ST})	99.49	99.52
Full marker set	Among groups (F_{CT})	1.10	1.32
	Among populations within groups (F_{SC})	0.58	0.31
	Within populations (F_{ST})	98.33	98.37

varying selective forces. This is exemplified in Fig. 2b, illustrating allele frequency distributions for a subset of four global outlier loci. While allele frequencies generally vary most between the Baltic and the Atlantic, some SNPs show more similar allele frequencies between the geographically very distant North Atlantic and Baltic clusters than between geographically adjacent regions (e.g. Cha_1025.1-149 and Cha_2884.1-367). One locus (Cha_381.2-437) was mainly polymorphic in Baltic populations and was either fixed or near fixation in all other populations, whereas another locus (Cha_15389.3-101) was polymorphic throughout the Baltic and Baltic/North Sea transition area but fixed in all other populations (Fig. 2b).

All loci significantly correlated with salinity showed similar patterns with both annual and spawning period averages; thus, we only present results for annual averages below (Tables 2-4). Of seven loci globally correlated with salinity, six and five associations were also significant (log10(BF) > 1.5) across the two regional salinity clines, respectively (Fig. 4a; Table 4). Three of these loci are annotated (Helyar *et al.* 2012), including a heat-shock protein (Hsp70; Cha_15389.3-101), a haemoglobin alpha subunit gene (Hba; Cha_15984.1-275) and a gene coding for the enzyme catalase (Cat; Cha_1513.1-91).

While eight loci showed significant associations with annual mean temperature, only four of these were significantly correlated with spawning period temperature (Table 2). Likewise, none of the loci correlated with spawning period temperature across the North Sea/North Atlantic temperature cline, despite five loci showing significant associations with annual mean temperature (Fig. 4b). No landscape variables were significantly correlated with allelic variation over the Baltic temperature cline (Table 3), in spite of two outlier loci identified across both the North Sea/North Atlantic and Baltic temperature clines. Several loci that were significantly associated with temperature and/or salinity were also significantly correlated with latitude and/or longitude potentially indicating isolation-by-distance relationships covarying with environmental factors (Table 2). However, 11 of 19 global environmental correlations remained significant after controlling for geographical distance using partial Mantel tests (Table 2). Including regional tests, in total, 11 of 26 temperature correlations and 12 of 18 salinity correlations remained significant after accounting for geographical distance (Tables 2–4). Further, as the method by Coop *et al.* (2010) already controls for population demography using neutral marker information, it is not expected that isolation-by-distance effects alone explain the associations between adaptive genetic variance and environmental factors, as reported in other studies (see Vasemägi 2006).

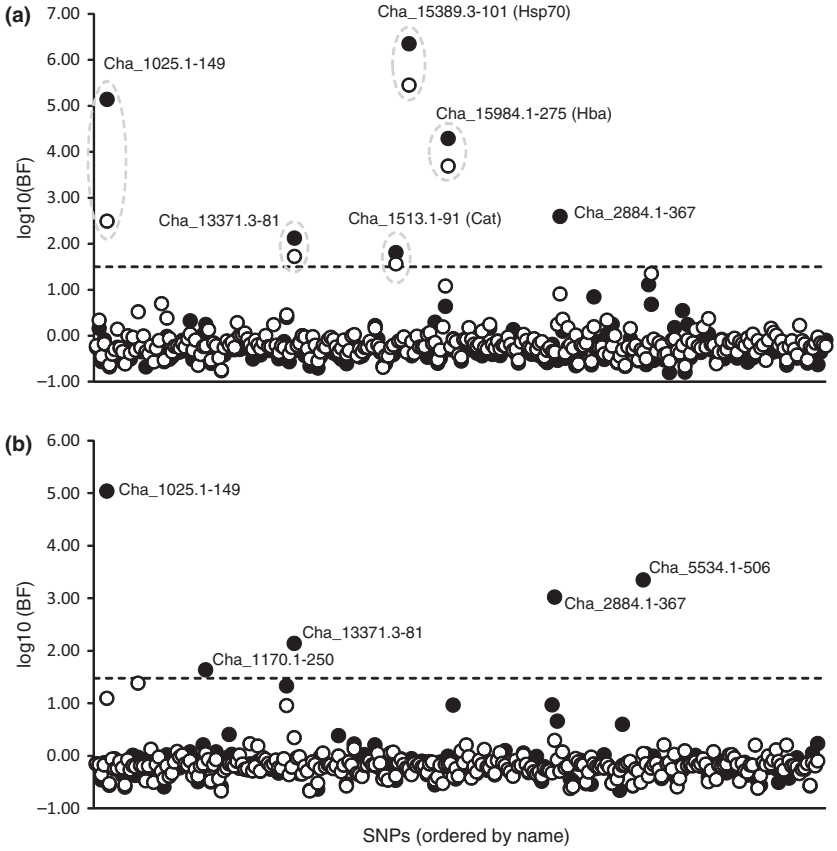

Fig. 4 Landscape association results. (a) Annual mean salinity over longitudinal clines (reflecting two low-salinity environments); 'North Sea/British Isles—Baltic Sea' (solid circles) and 'North Sea/British Isles—Ringkøbing Fjord' (open circles). See Table 4 for samples included in each analysis. Broken ellipses in grey denote five candidate loci showing significant correlations with salinity over both clines. (b) Annual (solid circles) and spawning season (open circles) mean temperatures in the North Sea/North Atlantic temperature cline. Broken horizontal lines mark lower thresholds of log10(BF) = 1.5.

Discussion

The extent and dynamics of local adaptation is key to understanding the ecological and evolutionary processes that influence biodiversity, as well as providing a spatially explicit framework for the conservation of genetic resources. While it is well recognized that opportunities for local adaptation in freshwaters due to habitat fragmentation and typically constrained dispersal (e.g. Mäkinen *et al.* 2008; Hohenlohe *et al.* 2010) are higher than in many marine fishes, recent evidence indicates remarkably small-scale adaptive variation in the latter (Moen *et al.* 2008; Poulsen *et al.* 2011). Here, by applying analyses of a large number of novel transcriptome-based SNP markers to spatio-temporal samples of herring, we identified outlier candidate genes indicating divergent selection among locally adapted populations. Moreover, candidate gene variation did not follow spatially uniform patterns across loci, suggesting that local populations have undergone multiple selective sweeps. Landscape genetic analyses suggested that environmental heterogeneity is an important driving force of divergent selection among populations, even in high gene flow organisms.

Combining neutral and selected loci to assess population structure

The highly consistent results from two statistical approaches for global outlier detection strongly suggest that the identified outlier loci or associated genomic regions are subject to divergent selection. The global outliers made up 5.7% of all analysed loci, which is within the range reported for other organisms (Nosil *et al.* 2009), including another high gene flow marine fish, Atlantic cod (Nielsen *et al.* 2009b; Bradbury *et al.* 2010). No outliers were suggested to be under balancing selection, possibly due to reduced statistical power when studying weakly structured species (Foll & Gaggiotti 2008). Substantial differences between the clustering of populations when applying the 'neutral' vs. the 'full' data set were observed. Identification of

three major clusters comprising the Baltic Sea, Baltic/North Sea transition area and the North Sea/British Isles/North Atlantic, respectively, using the 'neutral' marker set is in accordance with a previous microsatellite study (Bekkevold *et al.* 2005). An additional cluster was identified when adding outlier loci to the marker set, leading to a clear separation of North Atlantic from North Sea/British Isles populations, but also to more complex admixture patterns within populations. Such patterns of structuring clearly illustrate the increased statistical power for distinguishing weakly structured populations from the inclusion of only a few loci affected by divergent selection.

The inclusion of non-neutral markers may violate model assumptions of STRUCTURE (Pritchard *et al.* 2000) if outlier loci are under fluctuating environmental selection pressures uncoupled from the general population structuring process (migration and drift) within the species. However, differentiation at outlier loci may also elucidate evolutionary significant population units that could not be detected with neutral markers alone. Furthermore, no systematic trends of LD or deviations from HWE were observed at any of the outliers, and outliers exhibited consistent and temporally stable (over 6–10 years) patterns within and among geographical regions. Thus, we argue that careful inclusion of selected loci in a comparative approach (as presented here) remains useful for assessing spatial scales of demographically and reproductively isolated populations.

At local scales, there were clear examples of genetic separation between samples from geographically adjacent spawning locations, supported by both the clustering analysis and pairwise F_{ST} estimates for neutral markers. For example, samples from two fjords draining into the eastern North Sea (Limfjord and Ringkøbing Fjord) exhibited strong differentiation from all western North Sea locations (Fig. 3), in spite of phenotypic marker studies showing overlapping feeding habitat and large potential for mixing between populations in these areas (Rosenberg & Palmen 1982). Similarly, the western Baltic population of Rügen also exhibited clear genetic heterogeneity from the two neighbouring Baltic populations at Hanö Bay and Gdansk. In both cases, the results demonstrate that genetic variation does not follow a linear isolation-by-distance model, and corroborates natal homing as a strong driver of population structuring in herring (Gaggiotti *et al.* 2009).

Genetic divergence among samples in the North Sea/British Isles/North Atlantic was dramatically different using the 'neutral' vs. 'full' marker sets. Whereas neutral markers exhibited low, nonsignificant differentiation, agreeing with microsatellite studies (Mariani *et al.* 2005; Gaggiotti *et al.* 2009), analyses including selected loci exhibited a clear north–south separation with pairwise F_{ST} estimates of the same magnitude as between North Sea and Baltic Sea samples (Table S2, Supporting information). Such patterns could reflect selective sweeps for SNPs or associated gene regions at local scales corroborating findings in other marine fishes (Schulte 2001; Hemmer-Hansen *et al.* 2007a; Larsen *et al.* 2007; Moen *et al.* 2008; Nielsen *et al.* 2009b; Bradbury *et al.* 2010; Poulsen *et al.* 2011), and hence contribute to the notion that local selection pressures can override the homogenizing effects of high gene flow (Yeaman & Otto 2011). Indeed, results from Atlantic cod have shown similar latitudinal trends in the northeastern Atlantic Ocean to those identified for herring in this study (Nielsen *et al.* 2009b; Bradbury *et al.* 2010), and similar adaptive patterns have also been demonstrated in marine fishes inhabiting the western Atlantic (Schulte 2001; Bradbury *et al.* 2010).

The Baltic/North Sea transition area is hypothesized to constitute a hybrid zone with fish being genetically distinct from either North Sea or Baltic fish populations across several species (Nielsen *et al.* 2003; Hemmer-Hansen *et al.* 2007b; Limborg *et al.* 2009), including herring (Bekkevold *et al.* 2005). A genetically distinct cluster of herring in the Baltic/North Sea transition area was also supported here (see e.g. Fig. 3a for $K = 3$), in accordance with the findings of Gaggiotti *et al.* (2009). Further, our results suggested a relatively stronger admixture pattern for selected than for neutral loci (compare Fig. 3a and 3b). This pattern might reflect that populations in the transition area, despite exhibiting relatively closer neutral genetic relationships with Baltic than with North Sea populations for $K = 2$ (not shown), experience environmental selection pressures that are more similar to those in the North Sea. This was supported by sample-specific allele frequencies for the selected loci Cha_1025.1-149 and Cha_381.2-437, where some Baltic/North Sea transition area samples showed higher resemblance to North Sea populations (Fig. 2b).

The additional evidence provided here that adaptive divergence is marked even among potentially high gene flow species, such as herring, has wider significance. If gene flow restricts adaptive divergence, as is often assumed (Slatkin 1987), standard approaches using neutral genetic markers and landscape genetic approaches may be sufficient to get a crude estimate of adaptive variation. However, indications here, as elsewhere (Nielsen *et al.* 2009a), reinforce the notion of disparity among patterns of structuring across neutral and selected loci. Thus, if adaptive divergence does limit gene flow, genetic population structure may be poorly predicted from larval dispersal patterns, but more related to environmental heterogeneity that is sometimes obvious (Jørgensen *et al.* 2008), but sometimes not (Ha-

user & Carvalho 2008). Moreover, the implications for recruitment dynamics are considerable. Occasionally, high rates of larval influx from divergent populations may contribute little to local recruitment and may indeed be detrimental by increasing maladaptive traits (migration load). In such circumstances, selective mortality may be an important factor explaining population structure and could underlie some of the abrupt genetic discontinuities observed across hybrid zones of divergent populations, such as detected here in Baltic–North Sea herring. Documented evidence of high selective mortality in recruits to local populations (Planes & Lenfant 2002; Veliz *et al.* 2006; Vigliola *et al.* 2007) adds considerable support to this notion.

Environmental adaptation and candidate genes

Our results revealed an important role of environmental heterogeneity in shaping adaptive genetic variation at outlier genes. Specifically, the landscape genetic approach demonstrated clear associations with temperature for nine outliers and with salinity for seven outlier loci. The observation that only adaptive loci correlated with environmental factors further illustrates that divergent selection is an important force leading to locally adapted populations of herring, despite assumingly high levels of gene flow. Acknowledging the possibility that temperature or salinity is merely correlated with other environmental selection forces, our results support an evolutionary scenario with a strong environmental effect on shaping adaptive genetic variation in local herring populations.

Temperature is expected to affect a range of physiological pathways representing a multitude of underlying genes in poikilothermic organisms forced to exert innate responses to changes in ambient temperatures. Thus, it is not surprising that temperature affects a relatively large number of outlier genes including those also associated with salinity. Both temperature and salinity have also been suggested to shape adaptive genetic diversity among Atlantic cod in populations from some of the same areas as in this study (Nielsen *et al.* 2009b; Bradbury *et al.* 2010), as well as in other marine fishes (e.g. Schulte 2001; Mäkinen *et al.* 2008). However, a relatively large proportion of global outlier loci (6 of 16) did not correlate with landscape parameters, clearly suggesting an adaptive role for other (untested) selective agents such as environmental (physical, chemical, biological) factors or landscape-independent selection from intrinsic genetic incompatibilities due to, for example, epistasis (Bierne *et al.* 2011). A few outlier loci showed similar allele frequencies in the Baltic and North Atlantic samples and thus a tendency for clustering of northern Baltic and North Atlantic herring for the full marker set. These presumably adaptive signatures may reflect convergent evolution to common environmental conditions such as low temperature, which has also been shown for Atlantic cod on both sides of the Atlantic Ocean (Bradbury *et al.* 2010).

Herring are renowned for exhibiting population-specific spawning times (Cushing 1967), which might suggest local adaptation to spawning at specific seasonal temperatures. Temperature was identified as a covariant for several SNP loci in global analyses; however, these relationships were not evident at a regional scale except for the North Sea/North Atlantic latitudinal cline. This could result from reduced statistical power to detect outliers, caused by the reduced number of samples in each subanalysis, of which the method by Foll & Gaggiotti (2008) is expected to be particularly sensitive (Foll & Gaggiotti 2008). This was supported by the observation that more outlier loci were in fact detected using the approach of Excoffier *et al.* (2009). Alternatively, this observation could also be explained by increased type I and II errors and suggests that a large proportion of outliers only detected with the method of Excoffier *et al.* (2009) are in fact false positives (Narum & Hess 2011). The apparent lack of genetic covariance with temperature in the Baltic for loci exhibiting such a relationship globally contrasts with a previous microsatellite-based study (Jørgensen *et al.* 2005). This discrepancy may be attributable to the inclusion of the western Baltic Rügen population in the study of Jørgensen *et al.* (2005), in contrast to this study where it clusters with a Baltic/North Sea transition area group. Thus, weaker levels of differentiation within the Baltic proper may limit statistical power for detecting a potential relationship when excluding western Baltic samples.

Contrasting results between annual temperature and spawning temperature for the North Sea/North Atlantic temperature cline suggested that adaptation to spawning temperature was not the cause of selection at any of the loci examined here. For example, the English Channel population in the southern part of the range spawns November–January, whereas the subarctic populations spawn in April–May and August–September (Iceland) and March–May (Norway) (Table 1). As a result, Icelandic herring spawn at higher average temperatures than English Channel herring, suggesting that any temperature-related selection pressures are not specifically associated with conditions during spawning and early life stages (compared to observations in salmonids; Jensen *et al.* 2008); notably, in contrast to our findings for herring, a higher number of outlier genes correlated with spawning period (compared to annual mean) temperature in Atlantic cod (Nielsen *et al.* 2009b). These contrasting findings may reflect biological differences

between the two species. However, another important lesson learned from this study is that great caution is needed when using landscape genetic approaches, because annual estimates of environmental data may differ substantially from population-specific seasons actually affecting divergent selection. This is particularly pertinent for species exhibiting large seasonal variation in time of spawning as seen for herring. Alternatively, temperature conditions during early life stages may still impose selection at genes not associated with our marker panel. Thus, while our results support an adaptive role of temperature in general, crucial life stages and functions of outlier genes correlating with temperature remain unknown.

The clear correlation between outlier SNP variation and salinity was not driven by the Baltic populations alone, as five of six outlier loci also showed significant correlations with salinity across the Ringkøbing Fjord cline. Despite the brackish Ringkøbing Fjord population's geographical proximity to the North Sea, it has a close genetic relationship with Baltic/North Sea transition area populations likely reflecting a recent shared ancestry. Thus, we cannot rule out that the Ringkøbing Fjord population adapted to a low-saline environment as part of a larger ancestral population, with subsequent colonization of the Ringkøbing Fjord. However, whether candidate genes for salinity reflect historical adaptation in a common ancestral population, more recent parallel adaptation or a combination of the two, our results strongly indicate a general adaptive role of these genes or gene regions currently maintained in geographically isolated low-salinity environments. Three of these salinity-associated genes were annotated to known functions. A strong correlation was found between salinity and a nonsynonymous mutation (Cha_15389.3-101) in the heat-shock protein Hsp70 (Helyar *et al.* 2012), a gene family with a presumed key adaptive role in relation to environmental stress in fish (reviewed in Iwama *et al.* 1998; Basu *et al.* 2002), including European flounder (*Platichtys flesus*) (Hemmer-Hansen *et al.* 2007a; Larsen *et al.* 2008) and Atlantic cod (Nielsen *et al.* 2009b). Another outlier (Cha_15984.1-275) represented a synonymous mutation in a haemoglobin alpha subunit gene (Hba). Different variants of haemoglobin genes have been shown to be involved in local adaptation of Atlantic cod populations where different alleles are associated with divergent oxygen affinities and different temperature and hydrographical conditions (Sick 1961; Andersen *et al.* 2009). The third annotated outlier, the enzyme catalase (Cha_1513.1-91), decomposes hydrogen peroxide that is often generated at harmful levels during toxic stress responses. As such, this gene may play an important stress reaction role in marine fishes, as found for the thornfish (*Therapon jarbua*) (Nagarani *et al.* 2011). These candidate gene relationships are suggestive of environmental adaptation, albeit whether they are directly targeted by selection or exhibit hitchhiking with genomic regions of adaptive significance is not resolved.

Bierne *et al.* (2011) cautioned against interpreting significant landscape correlations as evidence for environmental adaptation at specific candidate genes. Instead, they suggested that detected outliers could represent intrinsic genetic incompatibilities uncoupled from the environment (so-called tension zones), which may become trapped in external hybrid zones driven by environmental selection. Here, candidate genes for environmental adaptation exhibited spatially distinct variation among loci, suggesting that drivers of divergence are not the same across loci and populations (Fig. 2b). Thus, we argue that it is unlikely that all outliers represent environmentally uncoupled barriers to gene flow and that a high proportion of our candidate genes indeed reflect adaptation to local environments. However, with the data at hand, we were not able to determine whether intrinsic or exogenous processes were more likely to have shaped patterns of differentiation at individual loci. To further understand the genetic architecture of fitness-related traits in these presumably locally adapted populations, studies with increasing genomic coverage (Hohenlohe *et al.* 2010; Star *et al.* 2011) and controlled rearing experiments examining genetically based fitness responses to specific environmental factors (Kawecki & Ebert 2004) are warranted.

Concluding remarks

Such local adaptation is highly relevant to fisheries management. It is not merely the conservation of genetic 'diversity' ('neutral and adaptive diversity at the DNA level') that is critical for the preservation of stocks; it is the protection of genetic 'resources' (diversity at the DNA level and its phenotypic expression at ecologically important traits). Thus, extirpation of locally adapted assemblages is of particular relevance to vulnerable species experiencing continued environmental change such as global warming or overexploitation (O'Brien *et al.* 2000). While the vulnerability of species at high trophic levels and with long generation times is widely accepted (Myers & Worm 2003), a recent study by Pinsky *et al.* (2011) showed that the majority of collapsed fisheries actually involve low-level trophic species like herring and other small pelagic fishes. Coupled with our findings, this implies that conserving the genetic 'resources' in heavily exploited species, including herring, is of paramount importance in safeguarding population resilience (Hilborn *et al.* 2003; Hauser & Carvalho 2008). Findings here constitute a basis for fur-

ther exploration of the genomic variation underlying locally adaptive traits in herring and for understanding the distribution of functionally important genetic variation in marine fishes in general.

Acknowledgements

The FishPopTrace consortium is thanked for many insightful discussions. We appreciate the assistance received from Matthieu Foll with BAYESCAN analyses and Nils Ryman for compiling a new version of POWSIM suitable for large SNP data sets. We owe many thanks to Eero Aro, Philip Coupland, Geir Dahle, Audrey Geffen, Thomas Gröhsler, Birgitta Krischansson, Ciaran O'Donnell, Henn Ojaver, Guðmundur Óskarsson, Iain Penny, Jukka Pönni, Veronique Verrez-Bagnis, Phil Watts and Miroslaw Wyszynski for providing herring samples. M.T.L. received financial support from the European Commission through the FP6 projects UNCOVER (Contract No. 022717) and RECLAIM (Contract No. 044133). The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. KBBE-212399 (FishPopTrace) and from the MariFish ERA-NET project DefineIt (ERAC-CT-2006-025989).

References

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

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

Andre C, Larsson LC, Laikre L *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.

Aro E (1989) A review of fish migration patterns in the Baltic. *Rapports et Procés-Verbaux Des Réunions Du Conseil International Pour l'Exploration de la Mer*, **190**, 72–96.

Basu N, Todgham AE, Ackerman PA *et al.* (2002) Heat shock protein genes and their functional significance in fish. *Gene*, **295**, 173–183.

Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Bekkevold D, Andre C, Dahlgren TG *et al.* (2005) Environmental correlates of population differentiation in Atlantic herring. *Evolution*, **59**, 2656–2668.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.

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

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

Conover DO, Clarke LM, Munch SB, Wagner GN (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*, **69**, 21–47.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.

Cushing DH (1967) The grouping of herring populations. *Journal of the Marine Biological Association of the United Kingdom*, **47**, 193–208.

Dahlberg MD (1979) A review of survival rates of fish eggs and larvae in relation to impact assessments. *Marine Fisheries Review*, **41**, 1–12.

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Fan JB, Oliphant A, Shen R *et al.* (2003) Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*, **68**, 69–78.

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.

Gaggiotti OE, Bekkevold D, Jørgensen HBH *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.

Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

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

Hauser L, Seeb JE (2008) Advances in molecular technology and their impact on fisheries genetics. *Fish and Fisheries*, **9**, 473–486.

Helyar S, Limborg MT, Bekkevold D *et al.* (2012) SNP discovery using next generation transcriptomic sequencing in Atlantic Herring (*Clupea harengus*). *PLoS ONE, accepted pending minor revisions*.

Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007a) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platicthys flesus L.*). *Heredity*, **99**, 592–600.

Hemmer-Hansen J, Nielsen EE, Grønkjær P, Loeschcke V (2007b) Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). *Molecular Ecology*, **16**, 3104–3118.

Hilborn R, Quinn TP, Schindler DE, Rogers DE (2003) Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 6564–6568.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD Tags. *PLoS Genetics*, **6**, 23.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Iles TD, Sinclair M (1982) Atlantic herring – stock discreteness and abundance. *Science*, **215**, 627–633.

Iwama GK, Thomas PT, Forsyth RHB, Vijayan MM (1998) Heat shock protein expression in fish. *Reviews in Fish Biology and Fisheries*, **8**, 35–56.

Jeffreys H (1961) *Theory of Probability*, 3rd edn. Oxford University Press, London, p. 432.

Jensen LF, Hansen MM, Pertoldi C *et al.* (2008) Local adaptation in brown trout early life-history traits: implications for climate change adaptability. *Proceedings of the Royal Society B-Biological Sciences*, **275**, 2859–2868.

Jørgensen HBH, Hansen MM, Bekkevold D, Ruzzante DE, Loeschcke V (2005) Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. *Molecular Ecology*, **14**, 3219–3234.

Jørgensen HBH, Pertoldi C, Hansen MM, Ruzzante DE, Loeschcke V (2008) Genetic and environmental correlates of morphological variation in a marine fish: the case of Baltic Sea herring (*Clupea harengus*). *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 389–400.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Larmuseau MHD, Raeymaekers JAM, Ruddick KG, Van Houdt JKJ, Volckaert FAM (2009) To see in different seas: spatial variation in the rhodopsin gene of the sand goby (*Pomatoschistus minutus*). *Molecular Ecology*, **18**, 4227–4239.

Larsen PF, Nielsen EE, Williams TD *et al.* (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Molecular Ecology*, **16**, 4674–4683.

Larsen PF, Nielsen EE, Williams TD, Loeschcke V (2008) Intraspecific variation in expression of candidate genes for osmoregulation, heme biosynthesis and stress resistance suggests local adaptation in European flounder (*Platichthys flesus*). *Heredity*, **101**, 247–259.

Larsson LC, Laikre L, Palm S *et al.* (2007) Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. *Molecular Ecology*, **16**, 1135–1147.

Legendre P, Legendre L (1998) *Numerical Ecology*. Elsevier, Amsterdam.

Limborg MT, Pedersen JS, Hemmer-Hansen J, Tomkiewicz J, Bekkevold D (2009) Genetic population structure of European sprat *Sprattus sprattus*: differentiation across a steep environmental gradient in a small pelagic fish. *Marine Ecology Progress Series*, **379**, 213–224.

Mäkinen HS, Cano M, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Molecular Ecology*, **17**, 3565–3582.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mariani S, Hutchinson WF, Hatfield EMC *et al.* (2005) North Sea herring population structure revealed by microsatellite analysis. *Marine Ecology Progress Series*, **303**, 245–257.

Maynard Smith J, Haigh J (1974) The hitchhiking effect of a favorable gene. *Genetical Research*, **23**, 23–35.

Moen T, Hayes B, Nilsen F *et al.* (2008) Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genetics*, **9**, 18.

Myers RA, Worm B (2003) Rapid worldwide depletion of predatory fish communities. *Nature*, **423**, 280–283.

Nagarani N, Devi VJ, Kumaraguru AK (2011) Mercuric chloride induced proteotoxicity and structural destabilization in marine fish (*Therapon jarbua*). *Toxicological and Environmental Chemistry*, **93**, 296–306.

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

Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Grønkjær P (2003) Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Molecular Ecology*, **12**, 1497–1508.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009a) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, **18**, 3128–3150.

Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009b) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 11.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

O'Brien CM, Fox CJ, Planque B, Casey J (2000) Climate variability and North Sea cod. *Nature*, **404**, 142.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547–572.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pinsky ML, Jensen OP, Ricard D, Palumbi SR (2011) Unexpected patterns of fisheries collapse in the world's oceans. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 8317–8322.

Planes S, Lenfant P (2002) Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Molecular Ecology*, **11**, 1515–1524.

Poulsen NA, Hemmer-Hansen J, Loeschcke V, Carvalho GR, Nielsen EE (2011) Microgeographical population structure and adaptation in Atlantic cod *Gadus morhua*: spatio-temporal insights from gene-associated DNA markers. *Marine Ecology Progress Series*, **436**, 231–243.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Raymond M, Rousset F (1995) Genepop (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rosenberg R, Palmen LE (1982) Composition of herring stocks in the Skagerrak–Kattegat and the relations of these stocks with those of the North Sea and adjacent waters. *Fisheries Research*, **1**, 83–104.

Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *Journal of Heredity*, **98**, 331–336.

Ruzzante DE, Mariani S, Bekkevold D *et al.* (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 1459–1464.

Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600–602.

Schmidt PS, Serrao EA, Pearson GA *et al.* (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology*, **89**, S91–S107.

Schulte PM (2001) Environmental adaptations as windows on molecular evolution. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, **128**, 597–611.

Sick K (1961) Haemoglobin polymorphism in fishes. *Nature*, **192**, 894–896.

Slatkin M (1987) Gene flow and the geographic structure of natural-populations. *Science*, **236**, 787–792.

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

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Vasemägi A (2006) The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. *Genetics*, **173**, 2411–2414.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Veliz D, Duchesne P, Bourget E, Bernatchez L (2006) Stable genetic polymorphism in heterogeneous environments: balance between asymmetrical dispersal and selection in the acorn barnacle. *Journal of Evolutionary Biology*, **19**, 589–599.

Vigliola L, Doherty PJ, Meekan MG *et al.* (2007) Genetic identity determines risk of post- settlement mortality of a marine fish. *Ecology*, **88**, 1263–1277.

Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, fresh-water, and anadromous fishes. *Journal of Fish Biology*, **44**, 213–232.

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

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

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

This study represents a part of M.T.L's PhD thesis focusing on local adaptation in marine fishes. D.B. and E.E.N. supervised M.T.L. and share a general interest in population structure and mechanisms of local adaptation in marine fishes. S.J.H. is a population geneticist who applies molecular methods to improve assessment and conservation of genetic biodiversity within exploited species. M.I.T. and G.R.C. have research interests in fisheries genetics and the evolutionary biology and conservation genetics of aquatic organisms in general. R.O. is interested in the application of genetic tools to fisheries enforcement. All authors are part of the FishPopTrace (FPT) Consortium which aims at developing gene associated SNP assays for describing and understanding spatial patterns of population structure and local adaptations in marine fish. The FPT Consortium is also interested in using loci under selection for developing cost effective traceability tools for forensic use.

Data accessibility

SNP genotypes have been deposited under the DRYAD entry doi: 10.5061/dryad.2n763.

Supporting information

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

Table S1 Names of the 310 screened SNP assays from Helyar *et al.* (2012) and information on genotyping success.

Table S2 Pairwise F_{ST} comparisons for the 'neutral' marker set (below diagonal) and for the 'full' marker set (above diagonal). Significant values ($\alpha = 0.05$) after correction for multiple tests using the FDR (Benjamini & Yekutieli 2001) are marked with an asterisk (*).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.