Adaptation to seasonal reproduction and temperature-associated factors drive temporal and spatial differentiation in northwest Atlantic herring despite gene flow

Running title: Local adaptation in northwest Atlantic herring

Angela P. Fuentes-Pardo[1,8], Ryan Stanley[2], Christina Bourne[3], Rabindra Singh[4], Kim Emond[5], Lisa Pinkham[6], Jenni L. McDermid[7], Leif Andersson[8,9], Daniel E. Ruzzante[1]

Affiliations

[1] Department of Biology, Dalhousie University, Halifax, NS, Canada.
[2] Fisheries and Oceans Canada, Maritimes Region, Dartmouth, NS, Canada.
[3] Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St John's, NL, Canada.
[4] Fisheries and Oceans Canada, St. Andrews Biological Station, St. Andrews, NB, Canada.
[5] Fisheries and Oceans Canada, Maurice Lamontagne Institute, Mont-Joli, QC, Canada.
[6] Department of Marine Resources, West Boothbay Harbor, Maine, USA.
[7] Fisheries and Oceans Canada, Gulf Fisheries Centre, Moncton, NB, Canada.
[8] Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden.
[9] Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, USA.

Corresponding authors: Angela P. Fuentes-Pardo (apfpgen@gmail.com) and Daniel E. Ruzzante (daniel.ruzzante@dal.ca).

Abstract

Natural selection can maintain local adaptation despite the presence of gene flow. However, the genomic basis and environmental factors involved in adaptation at fine-spatial scales are not well understood. Here, we address these questions using Atlantic herring (*Clupea harengus*), an abundant, migratory, and widely distributed marine species with substantial genomic resources including a chromosome-level genome assembly and genomic data from the eastern Atlantic and Baltic populations. We analysed whole-genome sequence and oceanographic data to examine the

genetic variation of 15 spawning aggregations across the northwest Atlantic Ocean (~1,600 km of coastline) and the association of this variation with environmental variables. We found that population structure lies in a small fraction of the genome involving adaptive genetic variants of functional importance. We discovered 10 highly differentiated genomic regions distributed across four chromosomes. Two of these loci appear to be private to the northwest, four loci share a large number of adaptive variants between northwest and northeast Atlantic, and four shared loci exhibit an outstanding diversity in haplotype composition, including a novel putative inversion on chromosome 8. Another inversion on chromosome 12 underlies a latitudinal genetic pattern discriminating populations north and south of a biogeographic transition zone on the Scotian Shelf. Our genome-environment association analysis indicates that sea water temperature during winter is the environmental factor that best correlates with the latitudinal pattern of this inversion. We conclude that the timing and geographic location of spawning and early development are under diverse selective pressures related to environmental gradients. Natural selection appears to act on early-life performance traits with differential fitness across environments. Our study highlights the role of genomic architecture, ancestral haplotypes, and selection in maintaining adaptive divergence in species with large population sizes and presumably high gene flow.

1. INTRODUCTION

Understanding how organisms adapt to their habitat and identifying which genes and environmental factors underpin this process are key questions in evolutionary biology and conservation. Such knowledge can help illustrate how biological diversification occurs (Nosil & Feder, 2012) and provides a genetic framework to direct management actions towards the conservation of intraspecific diversity (Hohenlohe et al., 2021). In marine species with high gene flow and large population sizes, it has been difficult to establish the extent of population structure and local adaptation by only analyzing neutral genetic markers (Hauser & Carvalho, 2008) or a small fraction of the genome. However, with the increased ability to examine both neutral and adaptive genetic variation enabled by high-throughput sequencing, recent genomic studies have revealed unprecedented levels of genetic differentiation at loci putatively under selection in marine species with seemingly undifferentiated populations at neutral loci [e.g., Atlantic herring (Lamichhaney et al., 2012), cod (Johansen et al., 2020), American lobster (Dorant et al., 2022)].

Theory proposes that gene flow can counteract adaptation by introducing maladaptive genetic variants into locally adapted populations but, at the same time, can bring in alleles that may prove beneficial in the local environment (Lenormand, 2002). This begs the question as to how local adaptation can arise and be maintained in marine species with presumably high gene flow. Empirical evidence indicates that gene flow in marine species can be restricted by a variety of ecological factors (Palumbi, 1994). These factors include differences in life history and physiological tolerance (e.g., timing of reproduction, or thermal preference/avoidance) (Lowerre-Barbieri et al., 2011), adult reproduction and behavioural phenotype favoring natal homing (Thorrold, 2001) or local recruitment (Levin, 2006), oceanography (e.g., gradients in salinity or temperature or divergent circulation patterns), and natural selection against migrants (Hendry et al., 2004; Nosil et al., 2005). When gene flow is restricted, genetic divergence can arise due to genetic drift, mutation, and selection. In marine species with large population sizes, the role of genetic drift is assumed to be minor, facilitating the detection of signatures of selection. Therefore, abundant migratory marine species inhabiting diverse environments offer ideal systems for the study of ecological adaptation with gene flow. Such is the case of Atlantic herring.

Atlantic herring is a pelagic fish that inhabits temperate waters of the north Atlantic, including fully marine environments, fjords, and the brackish waters of the Baltic Sea. It forms large schools, often comprising millions of individuals. Juveniles and adults undertake extensive annual migrations between feeding, overwintering, and spawning areas. In the northwest Atlantic (NW), spawning occurs from southern Labrador to the Gulf of Maine at relatively predictable times and locations near shore, peaking in spring (April-May) and fall (September-October) (McQuinn, 1997; Stephenson et al., 2009; Wheeler & Winters, 1984). Such diversity in reproductive phenology and spawning behaviour can promote temporal and spatial reproductive isolation between populations that are likely exposed to contrasting environments and selective pressures. These characteristics make this species an ideal model for investigating the genetic basis and mechanisms involved in ecological adaptation.

Herring plays a critical role in the marine food chains as a forage species that feeds on plankton and is an important food source for other species of fish, mammals, and sea birds. In addition, it sustains large fisheries throughout the north Atlantic (FAO, 2019), some of which

91 have experienced periods of severe decline in the last century (Engelhard & Heino, 2004;
92 Overholtz, 2002; Simmonds, 2007). In the NW Atlantic, declining abundance has impacts on
93 both commercial landings and the various fisheries that use forage fish such as Atlantic herring
94 and Atlantic mackerel (*Scomber scombrus*) as bait. This is the case of Atlantic lobster (*Homarus*
95 *americanus*), which has become the most valuable fishery resource in North America. In 2022,
96 following a lack of recovery, Fisheries and Oceans Canada (DFO) announced a moratorium on
97 directed commercial and bait fisheries for the spring stock of Atlantic herring in the southern
98 Gulf of St. Lawrence (DFO, 2022). Though the status of individual stocks is variable across the
99 Atlantic Canadian region (e.g., from cautious to critical in the Gulf of St Lawrence, or stable in
100 Newfoundland), herring abundance overall has been decreasing or sustained at low levels, which
101 is largely attributed to high adult mortality and low recruitment (Turcotte et al., 2021). Therefore,
102 in recent years there has been increased interest in understanding the population dynamics of
103 herring stocks using diverse approaches, including genetics.
104 Numerous research efforts have aimed to estimate the level of population structure of
105 herring using diverse genetic markers and at various spatial scales, primarily in the northeast
106 (NE) Atlantic. Initial genetic studies analyzed a few dozens of neutral markers and reported
107 panmixia or low population differentiation (Andersson et al., 1981; André et al., 2011;
108 Bekkevold et al., 2005; Jørgensen et al., 2005; Ruzzante et al., 2006). Subsequent studies based
109 on genetic variants derived from the transcriptome (Limborg et al., 2012) or reduced-
110 representation sequencing (Guo et al., 2016) or that used genome sequencing and an exome
111 assembly (Lamichhaney et al., 2012) showed that genetic differentiation occurs at outlier loci
112 likely associated with environmental gradients in the Baltic. Whole-genome studies using a
113 scaffolded reference genome revealed a number of genes associated with ecological adaptation
114 (Martinez Barrio et al., 2016), some of which are shared between European and Canadian
115 populations (Lamichhaney et al., 2017). The recent assembly of a chromosome-level reference
116 genome for the species enabled the possibility to examine structural rearrangements and more
117 complete gene models (Pettersson et al., 2019), a breakthrough in the ability to study the genetic
118 basis of adaptation. The most recent whole-genome study analysed 53 locations, 48 from the NE
119 and 5 from the NW Atlantic, and disclosed hundreds of loci underlying ecological adaptation
120 (Han et al., 2020). About 30 loci showed consistent association with adaptation to salinity, seven

121 to spawning time, and four putative chromosomal inversions are presumably related to adaptation
122 to water temperature during spawning, which is higher in waters around southern UK and
123 Ireland. Candidate genes associated with spawning time include the thyroid-stimulating hormone
124 receptor (*tshr*), *sox11b* transcription factor (*sox11b*), calmodulin (*calm1b*), and estrogen receptor
125 2 (*esr2a*), all located on chr 15 and with a known or presumed role in reproductive biology. The
126 gene *myhc* (myosin heavy chain) on chr 12 is putatively involved in myogenesis and plasticity of
127 seasonal development in herring (Johnston et al., 2001).

128 A few initiatives have focused on the population structure of NW herring Atlantic. Some
129 studies analysed a dozen of microsatellite markers (McPherson, Stephenson, O'Reilly, & Jones,
130 2001; McPherson, Stephenson, & Taggart, 2003; McPherson, O'Reilly, & Taggart, 2004). One
131 compared samples collected at the same location 10 years apart and found temporal stability of
132 allele frequencies at a subset of SNPs related with spawning time (Kerr et al., 2019). Others have
133 used whole genome approaches primarily focused on NE Atlantic populations in addition to the
134 same five Canadian populations (Han et al., 2020; Lamichhaney et al., 2017). Low genome-wide
135 differentiation at presumed selectively neutral loci was found across studies. However, one of the
136 whole-genome studies showed that about 25% of SNPs associated with seasonal reproduction are
137 shared between NW and NE Atlantic populations (Lamichhaney et al., 2017). The fact that a
138 large proportion of outlier loci are likely not shared between populations across the ocean, raises
139 the question of whether herring display local adaptation to NW Atlantic environments.

140 Here, we test this hypothesis through an exhaustive examination of genome-wide
141 variation of 15 spawning aggregations of herring in the NW Atlantic. In particular, we address
142 three focal questions: 1) what is the temporal and spatial scale of population structuring in the
143 focal area? 2) what is the genomic basis of population divergence and local adaptation? and 3)
144 which evolutionary mechanisms and environmental factors are likely involved in population
145 divergence? We integrate pooled DNA whole-genome sequencing (pool-seq) and oceanographic
146 data to address these questions. Given the role of temperature and photoperiod in regulating the
147 onset of seasonal reproduction in numerous temperate fish species (Migaud et al., 2010) and the
148 relatively high predictability of spawning season and location in Atlantic herring (Sinclair &
149 Tremblay, 1984), we expect that genetic differences among aggregations are largely due to
150 natural selection.

151

152

153 **2. MATERIALS AND METHODS**

154 **2.1 Study area and sample collection**

155 The study area comprises the reproductive range of *C. harengus* in the NW Atlantic, from
156 southern Labrador in Canada to the Gulf of Maine in the U.S. (Figure 1, Table 1). We collected
157 individuals from 10 locations across this area. Sample collection took place between 2012 and
158 2016 during or near the peak of spawning at each site, mainly in spring (April to May), summer
159 (June) and fall (August to October) seasons. We targeted individuals actively spawning or ready
160 to spawn (gonadal maturity stages 5 and 6, respectively) when possible, to have a representative
161 sample of distinct reproductive units (or populations) and to minimize the chance of sampling
162 non-spawning migrants. Most of the fish were actively spawning or ready to spawn, except in the
163 Bras D'Or lake (BDO-M), in which only 15% of fish were mature and the rest were in resting
164 condition (Figure S1). Therefore, this sample was considered mixed. Gonadal maturity was
165 visually assessed following classifications of Bucholtz, Tomkiewicz, & Dalskov (2008). Muscle
166 or fin tissues were collected from each individual and preserved in 95% ethanol at -20 ºC until
167 processing.

168

169 **2.2 DNA extraction and pool-sequencing**

170 Total genomic DNA was isolated from individual tissues using a standard phenol chloroform
171 protocol (Sambrook & Russell, 2006). DNA concentration (in nanograms per microliter, ng/µl)
172 was measured in triplicates using the Quant-iT PicoGreen dsDNA assay (Thermo Fisher
173 Scientific, U.S.) and the Roche LightCycler 480 Instrument (Roche Molecular Systems, Inc.,
174 Germany). DNA integrity was inspected using 0.8% agarose gel electrophoresis with 0.5x TBE
175 buffer and a 1 kilo base pairs (Kbp) molecular weight ladder.

176 Sequence data of each spawning aggregation was obtained using the pool DNA whole
177 genome sequencing (pool-seq) approach. This technique involves sequencing the combined DNA
178 of several individuals from a population using a single-barcoded library to generate population-
179 level whole genome data for a fraction of the cost of sequencing individuals to high depth, at the
180 expense of missing individual genotypes (Schlötterer et al., 2014). We aimed to mix in a pool

181 equal amount of DNA of 50 individuals collected at a given spawning site and season. Individual
182 DNA were normalized to a common concentration of 30 ng/µl and combined to a single tube
183 using the liquid handling robot epmotion 5407 (Eppendorf, Germany). Sequencing library
184 preparation and high-throughput sequencing were performed by The Centre for Applied
185 Genomics of the Hospital for Sick Children, Canada. A TruSeq Nano Illumina DNA library with
186 an insert size of 550 bp was built for each DNA pool. Illumina paired-end reads (2 x 126 bp)
187 were generated for each library with a HiSeq-2500 sequencer, targeting a 50x average depth of
188 coverage per pool. We combined this dataset with pool-seq data from five populations that were
189 studied before (Lamichhaney et al., 2017) (Table 1). Therefore, the northwest Atlantic was
190 represented by a total of 15 pool samples.
191
192 **2.3 Sequence filtering, alignment, and variant calling**
193 Sequence quality of raw reads per pool was examined with *FastQC* v0.11.5 (Andrews, 2010),
194 and a single report for all pools was generated with *MultiQC* v.1.3 (Ewels et al., 2016). Illumina
195 adapters, low quality bases (Phred score < 20), reads shorter than 40 bp, and single-end reads
196 were removed from our dataset using *Trimmomatic* v.0.36 (Bolger et al., 2014) [parameters:
197 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDOW:5:20 MINLEN:40].
198 The filtered paired-end sequences were mapped against the chromosome-level genome
199 assembly of Atlantic herring (Pettersson et al., 2019) using the Burrows-Wheeler Aligner (*BWA*)
200 v0.7.12-r1039 [default parameters, MEM algorithm] (Li, 2013). Read mapping quality was
201 assessed with *Qualimap* v.2.2.1 (Okonechnikov et al., 2015). Mapped reads in the form of BAM
202 files were sorted, PCR duplicates marked, and read groups added using *Picard tools* v2.10.2
203 (Broad Institute, 2018). An index file for each BAM file was generated with *SAMtools* v1.5 (Li &
204 Durbin, 2009).
205 We called variants using the *UnifiedGenotyper* algorithm implemented in the program
206 *GATK* v3.8 (McKenna et al., 2010). Indels were removed and biallelic SNPs were kept for further
207 filtering. We required SNPs passed *GATK* variant quality filters, which assess sequence and
208 mapping quality, strand bias, and variant position on reads to identify high quality
209 polymorphisms and discard spurious ones. We established cut-offs values for the *GATK* quality
210 filters based on density plots (Figure S2) [filters applied: FS > 60.0, SOR > 3.0, MQ < 40.0,

211 MQRankSum < -12.5, ReadPosRankSum < -8.0] (for a description of each annotation see (Broad
212 Institute, 2016)). Using *BCFtools* (Danecek et al., 2021), we retained SNPs that had a missing
213 rate < 20%, a minimum minor allele count of 2, and were polymorphic. We retained the SNPs
214 within the 5-99 percentile of the coverage distribution per pool (Figure S3), thus excluding
215 spurious SNPs in repetitive regions and copy number variants that commonly have excessively
216 high coverage.

217 Allele frequencies in pool-seq are derived from the read counts of each allele per variant
218 site. Read counts can vary among pools and along the genome due to technical biases during
219 sequencing (Dohm et al., 2008; Kolaczkowski et al., 2011). To account for sampling error of the
220 pool during sequencing before calculating allele frequencies, we rescaled the raw read counts to
221 the effective sample size n_{eff} using a python script that implemented the formula:

222 $$n_{\text{eff}} = \frac{(n * RD - 1)}{n + RD}$$

223 where *RD* is the raw read depth and *n* is the number of chromosomes in a pool (e.g., *n* =
224 2**number of individuals in a pool*, for diploid species). n_{eff} represents an estimate of the effective
225 number of chromosomes in a pool adjusted by the read depth (Bergland et al., 2014; Feder et al.,
226 2012; Kolaczkowski et al., 2011). Based on the rescaled read counts, we calculated allele
227 frequencies per pool.

228

229 **2.4 Population structure and genetic diversity**

230 We estimated the genomic differentiation between pool samples using pairwise-F_{ST} and a
231 Principal Components Analysis (PCA). We computed the unbiased F_{ST} for pools ($\hat{F}_{\text{ST}}^{\text{pool}}$) between
232 all pairs of spawning aggregations using the *R* package *poolfstat* (Hivert, 2018). This algorithm
233 calculates *F*-statistics equivalent to Weir & Cockerham (1984) and accounts for sampling error in
234 pool-seq.

235 To compare population structure patterns of undifferentiated (presumed neutral) and
236 highly differentiated loci (presumed under selection), we separately performed PCA on two SNP
237 sets defined from the empirical distribution and standard deviation (SD) of allele frequencies of
238 all markers, as applied in Han et al. (2020). The set of undifferentiated markers corresponded to
239 SNPs with allele frequency below and close to the mean value (0.03 < allele frequency SD ≤

240 0.09), while the highly differentiated markers had an allele frequency SD ≥ 0.2 from the mean
241 (Figure S4). To reduce the effect of physical linkage among loci, we retained one SNP every 1
242 Kbp for the undifferentiated markers, and one SNP every 10 Kbp for the differentiated markers
243 assuming higher linkage in regions under selection.
244 To characterize the genetic diversity within populations, we estimated the unbiased
245 nucleotide diversity (π) and Tajima's D for each pool using the program PoPoolation 1.2.2
246 (Kofler et al., 2011). In brief, a pileup file was generated from the BAM file of each pool using
247 samtools v.1.10 (Li, 2011). Indels and SNPs 5 bp around indels were removed. To account for
248 coverage variation due to sequencing, we subsampled without replacement the depth of coverage
249 of each pileup file to a uniform value, corresponding to the 5% quantile of the coverage
250 distribution (the minimum coverage of a SNP to be retained). SNPs with read depth between 5-
251 99% of the empirical coverage distribution per pool, minimum base and mapping quality of 20,
252 and a minor allele count of 2 were retained for further analyses. The statistics π and Tajima's D
253 were calculated in sliding windows of 10 Kbp with a step size of 2 Kbp. Windows with a
254 coverage fraction ≥ 0.5 were used for further analyses. All statistical tests and graphics were
255 performed in *R* (R Core Development Team, 2023).
256
257 **2.5 Detection of genomic regions under selection**
258 To identify outlier genomic regions that are likely targets of selection, we calculated the absolute
259 difference in allele frequencies per SNP (delta allele frequency, dAF) between paired contrasts of
260 pools, using the formula: *dAF = absolute[meanAF(group1) – meanAF(group2)]*. Based on PCA
261 clustering patterns and prior ecological knowledge of spawning stocks, the two most informative
262 contrasts were established to compare: (i) spring and fall spawners, and (ii) northern and southern
263 fall spawners. This analysis was complemented with the calculation of the moving average (or
264 rolling mean) of dAF values over 100 consecutive makers, which smooths out the signal of
265 divergence while ruling out the effect of single outlier SNPs that might result from sequencing
266 bias.
267 To identify novel loci in the NW Atlantic that are likely related with local adaptation, we
268 compared the outlier SNPs found in this study with those reported in Han et al. (2020), showing
269 strong and consistent association with adaptation to spawning time and salinity among northeast

Atlantic and Baltic populations, and within four inversions (on chr 6, 12, 17, and 23). We visually inspected the allele frequencies of outlier SNPs in our 15 NW Atlantic populations and in previously published data of one Pacific herring and 47 Atlantic herring pools (Han et al., 2020) (Table S3). In addition, we generated a neighbor-joining tree based on the allele frequencies of outlier SNPs for all 63 populations.

2.6 Genome-environment association and isolation-by-distance

To identify which environmental variables are strongly correlated with putatively adaptative SNPs underlying a latitudinal genetic pattern, we performed a redundancy analysis (RDA). The environmental dataset consisted of hours of day light (dayLightHours), sea surface temperature (SST) and salinity (SSS) during the months of spawning at each location, and for the winter and summer seasons, representing the coldest and warmest periods of the year. These variables were chosen because they are relevant in fish physiology and have been linked to population structuring of numerous marine species in the NW Atlantic (Stanley et al., 2018). Temperature and salinity values for each sample location were derived from monthly environmental data layers of SST and SSS developed for the North Atlantic between 2008-2017 based on the Bedford Institute of Oceanography North Atlantic Model (BNAM), a high-resolution numerical ocean model. A detailed description of oceanic (Madec et al., 1998) and sea ice (Fichefet & Maqueda, 1997) components of the model can be found in Wang, Brickman, Greenan, & Yashayaev (2016) and Brickman, Hebert, & Wang (2018). Data layers were converted to an ASCII grid with a NAD83 projection (ellipse GRS80), and had a nominal resolution of 1/12° (~5 km^2). Binned data for the months of spawning at each location (spring: April-May, fall: September-October), and for the winter (January-February-March) and summer (July-August-September) seasons, were averaged across 9 years in order to capture long-term trends of oceanographic variation. Extractions (mean, min, and max values) were taken from these averaged layers for each sample location. Prior to performing RDA, environmental data were standardized to zero mean and unit variance using R. Collinearity between environmental variables was assessed using two methods, with pairwise correlation coefficients computed with the function *pairs.panels* of the R package *psych* (Revelle, 2018) (Figure S5), and with variance inflation factors (VIF) obtained from RDA models built with the R package *vegan* (Dixon, 2003).

300 The most collinear variables were removed, and remaining collinear variables were identified and
301 excluded one-by-one in consecutive RDA runs based on their VIF value. The variable with the
302 highest VIF was discarded in each run until all variables had a VIF < 5, following
303 recommendations of Zuur, Ieno, & Elphick (2010). The uncorrelated set of environmental
304 variables consisted of summer sea surface temperature (SST_{Summer}), winter sea surface
305 temperature (SST_{Winter}), sea surface temperature during spawning (SST_{Spawn}), and day light hours
306 (Figure S6).

307 For RDA, we used the uncorrelated set of environmental data as constraining factors for
308 the population allele frequencies of the SNPs within the chromosomal inversion on chr 12
309 (chr12:17823410-25605433). RDA runs were performed with the R package *vegan*.
310 Environmental variables that best explained the genetic variance were confirmed with a bi-
311 directional stepwise permutational ordination method (1000 iterations) using the function
312 *ordistep*. Significance of the overall RDA model and the most explanatory environmental
313 variables was assessed with an analysis of variance (ANOVA) using 9999 permutations.

314 To evaluate whether the latitudinal genetic pattern follows an isolation-by-distance (IBD)
315 pattern, we calculated the significance of the linear relationship between geographic and genetic
316 distances between pairs of locations for all SNPs and for only the ones associated with the
317 latitudinal genetic pattern (within the chr 12 inversion). For this, we performed separate Mantel
318 tests (Mantel 1967) with 9999 permutations for each set of loci using the R package *ade4* (Dray
319 & Dufour, 2007). Genetic distances were linearized ($\hat{F}_{\mathrm{ST}} = \frac{\hat{F}_{\mathrm{ST}}}{1-\hat{F}_{\mathrm{ST}}}$) (Rousset, 1997) and $\hat{F}_{ST}$
320 estimates were computed either using all SNPs identified across the genome, or solely the outlier
321 SNPs strongly associated with latitudinal divergence. Geographic distances were estimated with
322 the R package *CartDist* (Stanley & Jeffery, 2017) as the least-coast oceanic distance in
323 kilometers (km) where land is treated as a barrier to movement.
324
325
326 **3. RESULTS**
327 **3.1 Population-scale whole-genome sequencing**

328 We collected 697 individuals from 10 spawning aggregations throughout the NW Atlantic
329 (Figure 1A). We pooled DNA of 41-50 individuals per location and generated paired-end short
330 reads for the whole genome using a HiSeq Illumina sequencer. Our sequencing effort yielded a
331 total of ~200 Giga base pairs (Gb) of data. The median depth of coverage per pool ranged
332 between 57X and 77X (Table S1). We combined the pool-seq data of these 10 populations with
333 data of five Canadian populations previously published (Han et al., 2020), for a total of 15 pools
334 constituting the dataset of northwest Atlantic herring. After mapping reads against the genome
335 assembly of the Atlantic herring (Pettersson et al., 2019), calling variants, and applying quality
336 filters to the raw variants, we obtained a total of 5,264,683 high quality SNPs.
337
338 **3.2 Population structure**
339 We investigated the population structure among samples using pairwise-F_{ST} and PCA. The F_{ST}
340 values of pools ($\hat{F}_{\mathrm{ST}}^{\mathrm{pool}}$) ranged between 0.001 and 0.018 (Figure 1B, Table S2), indicating low
341 genomic differentiation among spawning aggregations across ~1,600 km of coastline. This result
342 is consistent with previous studies of Atlantic herring (Han et al., 2020). Despite low
343 differentiation, we detected two subtle patterns of genetic structure between: (i) spring and fall
344 spawners (SPH-S, IBC-S, NTS-S vs. others, mean $\hat{F}_{\mathrm{ST}}^{\mathrm{pool}}$ 0.010 ± 0.001); (ii) the southernmost
345 sample in the Gulf of Maine (ME4-F) and the other fall spawners (mean $\hat{F}_{\mathrm{ST}}^{\mathrm{pool}}$ 0.007 ± 0.001).
346 To evaluate whether population structure patterns were mainly explained by
347 undifferentiated (putatively neutral) or highly differentiated loci (putatively selective), we
348 separately performed PCA based on the allele frequencies of a subset of neutral (n = 135,950,
349 Figure 1C) or outlier markers (n = 545, Figure 1D) (see Methods for details). In the PCA with
350 neutral markers, most populations formed a single tight cluster with the exception of two samples
351 from the Gulf of St Lawrence (IBC-S and NTS-F), which appeared as outliers. The first two
352 principal components (PCs) explained 23.7% of the genetic variance, which are largely driven by
353 these two samples.
354 In stark contrast, in the PCA based on presumably adaptive markers, spring and fall
355 spawning populations separated on PC1, and fall spawners separated following a latitudinal
356 pattern along PC2. These two PCs explained a large proportion of the genetic variance (82.0%).

357 On PC2, the "northern" samples included populations from Labrador (LAB-F), Newfoundland
358 (BLS-F, TRB-F, NDB-F), the Gulf of St Lawrence (NTS-F, PEI-F), and the Bay of Fundy (SCB-
359 F), while the "southern" samples comprised populations from the Scotian Shelf (GEB-F, MUS-F)
360 and the Gulf of Maine (ME4-F). On PC1, populations from Bras D'Or lake (BDO-M) and Sept-
361 Îles (SIL-U), the first being a mixed sample and the latter was collected in the summer, appear to
362 be closer to fall spawning samples. On PC2, spring samples were separated from each other and
363 from the summer sample from Sept-Îles (SIL-U), in concordance with the pairwise $\hat{F}_{\mathrm{ST}}^{\mathrm{pool}}$
364 estimates.
365
366 **3.3 Genomic regions under selection**
367 Despite low genome-wide population structure, we uncovered genomic regions displaying
368 significant differentiation between populations. These regions were detected through genome
369 scans based on the absolute difference in allele frequencies (dAF) of 5,264,683 SNPs between
370 pool samples representing distinctive ecotypes or clusters in the PCA. The informative
371 comparisons were: (i) spring vs. fall spawners, and (ii) northern vs. southern fall spawners.
372 We discovered that the genomic regions showing the strongest association with spawning
373 time were confined to nine major loci distributed on four chromosomes (chr): chr 8 (one locus),
374 chr 12 (one locus), chr 15 (five loci), and chr 19 (two loci) (n = 2137 SNPs, Figure 2A). Four of
375 these loci are considered "novel", as they are not among the loci reported in Han et al. (2020) as
376 significantly associated with ecological adaptation (chr8:23040136-30729461, chr15:6750000-
377 7000000, chr15:9200000-9350000, chr19:23155000-23300000). While some SNPs in these loci
378 have been associated with spawning time before (Han et al., 2020) (n = 387, blue triangles in
379 Figure 2A), we discovered a large number of outlier SNPs (n = 1748, 81.8%) (Figure 3, 4, red
380 dots) that showed a stronger genetic differentiation in the west than in the east Atlantic. Most of
381 these SNPs were located in a putative structural variant (SV) on chr 8 (n = 1349, 77% of novel
382 SNPs) (Figure 3) that has not been reported previously. Other novel outliers SNPs were found in
383 loci with known association with spawning time on chr 12, 15, and 19 (Figure 4).
384 Our data also revealed that the latitudinal genetic pattern discriminating populations that
385 spawn north or south of a biogeographic transition zone on the eastern Scotian Shelf (Figure 1A)

386 is underlined by a large 8-Mbp long SV. This locus corresponds to a known chromosomal
387 inversion on chr 12 (n = 1106 SNPs with dAF ≥ 0.42, Figure 2B) (Pettersson et al., 2019).
388

389 **3.4 A candidate inversion on chr 8**

390 The new putative SV on chr 8 is located towards one end of the chromosome, in a region with
391 heterogeneous rate of recombination (Figure 3). The genetic differentiation at this locus shows
392 the typical profile of a chromosomal inversion, "block" shaped with sharp "edges" defined by a
393 drastic drop in dAF (Figure 3A) implying reduced recombination. The majority of outlier SNPs
394 at this locus are novel, except for six known spawning time- and two salinity-associated markers
395 (Han et al., 2020) (Figure 3B). This putative inversion was not highlighted in Han et al. (2020),
396 but an examination of the patterns of segregation shows that its variant haplotype is present in
397 some populations in the northeast Atlantic as well (Figure S7).

398 Alternative haplotypes of the inversion differentiate most spring and fall spawning
399 populations, and intermediate allele frequencies were observed in the summer spawning
400 population from Sept Îsles (SIL-U), the mixed sample from Bras D'Or lake (BDO-M), and the
401 fall spawning population from Northumberland Strait (NTS-F). A comparison of allele
402 frequencies between west and east Atlantic herring revealed that while opposite alleles
403 distinguish spring and fall spawners among Canadian samples, this was not always the case for
404 other Atlantic and Baltic populations (Figure S7A). For instance, the haplotype predominant
405 among Canadian fall spawners is also in high frequency in the summer-spawning sample from
406 Greenland and in two fall-spawning populations from the northwest of England (Orkney, Cape
407 Wrath, Hebrides, Figure S7B). This haplotype is in intermediate frequency in Iceland and in both,
408 spring and fall spawners in the North Sea. The opposite haplotype is prevalent among most
409 spring and winter spawners across the Atlantic, but in the Baltic Sea there is a recombinant
410 haplotype that is shared between autumn spawners and the Pacific herring.

411 The most differentiated SNPs were located in the interval chr8:27,500,000-27,900,000
412 (Figure 3C), which harbors some novel and previously reported SNPs. Genes at this locus include
413 a lncRNA (long-non-coding RNA) and *fgfr4* (fibroblast growth factor receptor 4). Another
414 interesting region is located in the interval chr8:30,150,000-30,750,000 (Figure 3D), which
415 shows a significant drop in nucleotide diversity and negative Tajima's D among fall spawners.

416 This locus harbors several genes, two of which include missense mutations, *ubr1* (ubiquitin
417 protein ligase E3 component n-recognin 1) and *entpd5a* (ectonucleoside
418 triphosphate/diphosphohydrolase 5). A description of outlier SNPs and the closest genes at
419 divergent genomic regions can be found in Table S4.
420

421 **3.5 Novel and previously reported signals of selection**

422 The selection signal on chr 12 is located upstream of a previously described inversion (Pettersson
423 et al., 2019), and extends over 1 Mbp towards the middle of the chromosome in a region of low
424 recombination (Figure 4A, 4D, S8A). Notably, most outlier SNPs in this locus are novel, except
425 for two spawning time-associated markers (Figure 4D, first track, S8B), but they occur in a
426 region showing strong signal of differentiation in previous studies (Han et al., 2020). Alternative
427 alleles were close to fixation between spring and fall spawners, except in the spring-spawning
428 sample from Stephenville (SPH-S, southwest of Newfoundland) that shows intermediate allele
429 frequencies (Figure 4B, second track, S8B). The profile of genetic diversity metrics for the most
430 divergent SNPs revealed that this region likely experienced a selective sweep in the fall
431 spawners, supported by the lower nucleotide diversity and more negative Tajima's D than in the
432 spring spawners (Figure 4D, S8B). Genes in this region include: five genes of the myosin heavy
433 chain family (i.e., *myhc*-like, *myhz1.1* and three *myh2*), *scd5* (stearoyl-CoA desaturase 5),
434 *tmem150c* (transmembrane protein 150C), *sbno1* (strawberry Notch Homolog 1), and *denr*
435 (density-regulated protein). Opposite alleles at this locus distinguish Canadian and Baltic spring
436 and fall spawners, but this is not applicable to all north Atlantic populations (Figure S8). The
437 "fall" Canadian allele is prevalent in all other oceanic populations (Greenland, Ireland-UK, North
438 Sea, and Norwegian fjords) regardless of their spawning season, and in the Baltic fall spawners.
439 There is a gradient of allele frequencies in the transition zone between the Baltic Sea and the NE
440 Atlantic Ocean.

441 The signal of selection on chr 15 is 4 Mbp long (Figure 4B, 4E, S9) and it is located near
442 the middle of the chromosome in a region of low recombination. It consists of five distinct loci,
443 two more than detected in the east Atlantic (Han et al., 2020). In genomic order, the first locus is
444 novel, it is approximately 250 Kbp long (chr15:6,750,000-7,000,000), and it harbours the genes
445 *rab15* (member RAS Oncogene Family 15) and *sipa1l1* (signal induced proliferation associated 1

like 1) (Figure S9C). The second locus extends over 200 Kbp (chr15:7,650,000-7,850,000) and contains the gene *sox11b* (Figure S9D), which has been associated with spawning time before (Han et al., 2020) but the two most differentiated SNPs are novel (dAF > 0.7), as they did not show consistent association between northeast Atlantic and Baltic populations. The third locus is about 530 Kbp long (chr15:8,540,000-9,070,000) and harbours a large number of SNPs and genes with a known association with spawning time such as *tshr* and *calm1b* (Figure S9E). The nucleotide diversity and Tajima's D profiles support a selective sweep in the *tshr* region, in concordance with previous studies (Chen et al., 2021) (Figure S9E). The fourth locus is novel, extends over 150 Kbp (chr15:9,200,000-9,350,000), and contains two genes, *flrt2* (fibronectin leucine-rich transmembrane protein 2) and a *lncRNA* (Figure S9F). The fifth locus extends 280 Kbp (chr15:10,820,000-11,100,000) and consists of two peaks, one spans two well-characterized genes, *esr2a* and *syne2* (Figure S9G), and the other one encompasses several *lncRNA* genes and *l3hypdh* (trans-L-3-hydroxyproline dehydratase). Interestingly, most SNPs in the *esr2a-syne2* region are novel, while the SNPs in the *lncRNAs-l3hypdh* region have been previously associated with spawning time (Han et al., 2020). Across all five loci in this region, the spring and fall spawners from Canada have opposite alleles in high frequency (Figure S9B, second track). When comparing allele frequencies between west and east Atlantic herring, a similar pattern distinguishing spring and fall spawners is observed in three of the five loci, harbouring *tshr*, *sox11b*, and *flrt2-lncRNA* genes, and in the second peak of the *esr2a-syne2-lncRNAs-l3hypdh* locus (*lncRNAs-l3hypdh*) (Figure S10C, S10E). In the first peak of the latter locus, corresponding to *esr2a-syne2*, the "fall" Canadian allele is predominant among populations from the transition zone and the Baltic Sea (Figure S10C, S10D). In the *rab15-sipa1l1* locus, the alternative alleles appear to be unique of Canadian spring spawners (Figure S10A, S10B).

The selection signal on chromosome 19 is located towards one end of the chromosome, in a region of low recombination, and it contains two loci (Figure 4C, 4F, S11). The first locus is 410 Kbp long (19:20,290,000-20,700,000) and harbors the genes *herpud2*, *adgrb1b* and a *lncRNA* (Figure S11C). Most SNPs and genes at this locus have been previously associated with spawning time (Han et al., 2020). The second locus is novel, extends over 145 Kbp (19:23,155,000-23,300,000), and includes a *lncRNA* and *sgk-like* genes. All the SNPs in this locus are novel (Figure S11B). The low nucleotide diversity and negative Tajima's D profiles at

476 this region suggest that it constitutes a selective sweep (Figure S11D). Indeed, the variant alleles
477 at this locus are unique for the Canadian fall spawners and the summer-spawning sample from
478 Greenland (Figure 11E, S11F).
479 Overall, this study revealed 10 genomic regions showing highly significant genetic
480 differentiation among the population samples from the northwest Atlantic. Only two of these loci
481 are unique to the northwest Atlantic, and do not appear to be segregating in northeast populations
482 of Atlantic herring (Figure 5, Table S6). These are the loci harboring *rab15-sipa1l1* on
483 chromosome 15 (Figure S9C) and the most distal region on chromosome 19 harboring *lncRNA-*
484 *sgk-like* (Figure S11D). At some of the shared loci between northwest and northeast populations,
485 like *tshr* and *herpud2* (Figure S9E and S11C, respectively), the alternate haplotypes segregating
486 at these loci are nearly identical and the ranking of top SNPs are almost identical in northwest
487 and northeast herring (Figure 5B, Table S6). In contrast, at some other shared loci there is a clear
488 difference in haplotype composition (Figure 5C). The best example is the region containing a
489 cluster of myosin heavy chain genes on chromosome 12 where numerous SNPs show strong
490 genetic differentiation in northwest Atlantic but low differentiation in northeast populations,
491 whereas a smaller set of other SNPs show the opposite trend suggesting that different haplotypes
492 must be favored by selection in the two sides of the Atlantic (Figure 5C, S8).
493
494 **3.6 The inversion on chromosome 12 likely associated with minimal temperature survival**
495 The putative inversion on chr 12 that underlies the north-south genetic pattern contains a large
496 number of outlier SNPs as previously reported (Han et al., 2020), but some of the most
497 differentiated ones are novel for the west Atlantic populations (Figure 6A, 6B). The "north"
498 haplotype of the inversion, as defined in Han et al. (2020), is prevalent among northern samples
499 from Labrador (LAB-F), Newfoundland (NDB-F, TRB-F, SPH-S), Gulf of St. Lawrence (BLS-F,
500 SIL-U, PEI-F, NTS-S), Bras D'Or lake (BDO-M), and Scots Bay in the Bay of Fundy (SCB-F)
501 (Figure S12). The "south" haplotype is in high frequency in the sample from the Gulf of Maine
502 (ME4-F), the southernmost location included in this study. Intermediate haplotype frequencies
503 were common in two samples from the Scotian Shelf (MUS-F, GEB-F) (Figure 6B, S12A).
504 The current hypothesis is that this inversion is likely associated with adaptation to
505 temperature during spawning (Pettersson et al., 2019). To test this hypothesis, we performed a

506 Genome-Environment Association (GEA) analysis based on redundancy analysis (RDA). We
507 included as environmental variables the number of hours of day light, average sea surface
508 temperature and salinity during the spawning months (SST_{Spawn}) and for the winter (SST_{Winter})
509 and summer (SST_{Summer}) seasons, as a proxy for the most extreme annual climatic conditions.
510 This analysis indicated that sea surface winter temperature (SST_{Winter}) is the environmental factor
511 that best explains the genetic variation of outlier loci in the chr 12 inversion (*P*-value = 0.005,
512 alpha = 0.01) (Figure 6C, S13). Indeed, this was the only variable that was statistically significant
513 for an alpha level of 0.01 (Table S5), and salinity measures were not informative (Figure S14). In
514 the RDA plot, spawning aggregations were separated according to SST_{Winter} along RDA 1, axis
515 that explained 58.1% of the total genetic variance. Notably, the "north" haplotype of the
516 inversion is prevalent among populations that seem to be exposed to much colder temperatures
517 during the winter than the southern ones in the Scotian Shelf and the Gulf of Maine (Figure 6D).
518 Latitudinal genetic clines can result from natural selection or historical demographic
519 processes, such secondary contact of allopatric populations that diverged during a geologic event
520 and aligned with a contemporary gradient. When natural selection is acting on populations, it
521 would be expected that the fraction of the genome under selection shows the clinal pattern. In
522 contrast, in a scenario where historical divergence leads to limited gene flow, the entire genome
523 should show the cline and it would resemble an isolation-by-distance (IBD) pattern. To examine
524 whether the genetic pattern of the chr 12 inversion followed IBD, we performed a Mantel test in
525 two separate datasets, one consisting of all SNPs and another one of only outlier SNPs in the chr
526 12 inversion. The test based on all SNPs disclosed that there is no linear relationship between
527 geographic and genetic distances genome-wide (R^2 = 0.04, P = 0.11, 9999 replicates). The test
528 performed solely on the SNPs showing the latitudinal genetic cline suggested weak IBD (R^2 =
529 0.30, P = 6e-04, 9999 replicates) and sub-structuring (Figure S14). Taken together, these results
530 suggest that demographic processes do not seem to explain the pattern of latitudinal divergence,
531 although it could coincide with IBD to some extent.
532
533
534 **4. DISCUSSION**
535 ***Population structure only at loci under selection***

536 We found that population structure of NW Atlantic herring mostly lies in a number of genomic
537 regions of adaptive importance, in otherwise weakly differentiated genomes (mean pairwise-F_{ST}
538 between 0.001 and 0.018) (Figures 1 and 2). This observation is in agreement with previous
539 studies on Atlantic herring (Bekkevold et al., 2015; Gaggiotti et al., 2009; Guo et al., 2015; Han
540 et al., 2020; Lamichhaney et al., 2012; Limborg et al., 2012; Martinez Barrio et al., 2016;
541 Teacher et al., 2013). The population structure observed is mainly driven by natural selection
542 acting in contrasting environments varying between seasons (spring and fall) and geographic
543 regions (northern and southern) where reproduction and early life stages occur. Our study
544 indicates that the strength of selection must be high, such that adaptive genetic variants display
545 striking allele frequency differences between large populations (in the order of hundreds of
546 thousands to millions of individuals). While herring population structure associated with seasonal
547 reproduction has been reported previously (Han et al., 2020; Lamichhaney et al., 2012;
548 Lamichhaney et al., 2017; Martinez Barrio et al., 2016), our study reveals spatial substructure in
549 the northwest Atlantic that had not yet been identified. Taken together, this work contributes to
550 the growing body of evidence supporting the hypothesis that population structure in marine
551 species is often adaptive and underlies ecological diversity (Benestan et al., 2015; Bradbury et al.,
552 2013; Han et al., 2020; Kess et al., 2021; Lehnert et al., 2018).
553 This study demonstrates that the majority of loci (8 out of 10) under strong selection in
554 northwest Atlantic herring are also major loci underlying genetic differentiation among
555 populations of herring in the northeast Atlantic. However, there is no perfect overlap regarding
556 which SNPs at these loci show the strongest genetic differentiation between populations. This
557 could be explained by (i) genetic drift due to isolation by distance or by (ii) local genetic
558 adaptation. We conclude that the local adaptation is the most likely explanation because drift
559 should affect all loci in a similar way but this is not the case. We find that the pattern of genetic
560 differentiation is remarkably similar at some loci (in particular *tshr* and *herpud2* both strongly
561 associated with spawning time) whereas other loci, in particular the cluster of myosin heavy
562 chain genes on chromosome 12, show considerable heterogeneity between the west and the east
563 (Figure 5). This result is consistent with our previous conclusion that genetic adaptation in
564 Atlantic herring is to a large extent based on ancestral haplotype blocks, some associated with
565 inversions, that have persisted over hundreds of thousand years (Han et al., 2020; Lamichhaney et

566 al., 2017; Pettersson et al., 2019). However, these haplotype blocks are not static over time, they
567 continue to accumulate causal changes contributing to genetic adaptation, as observed in this
568 study. This is a mode of evolution also noted in the Darwin's finches and in other adaptive
569 radiations (Rubin et al., 2022).
570
571 ***Genomic regions associated with adaptation to temporal environmental variation***
572 Populations that predominantly spawn in spring or fall seasons differ at nine major loci
573 distributed in four chromosomes: 8 (one locus), 12 (one locus), 15 (five loci), and 19 (two loci)
574 (Figure 2A, 3, 4). These loci vary in size, ranging from a few Kbp up to several Mbp, as it is the
575 case of the 8-Mbp long structural variant (SV) on chr 8, first described here for the species
576 (Figure 3). This SV is also in high frequency in the sample from Greenland and among fall-
577 spawning oceanic populations. A recombinant variant of the putative inversion is present in
578 Baltic fall-spawning samples (Figure S7). This SV contains numerous genes, for which it is hard
579 to infer its function without further experiments. A region within the SV (chr8:30,150,000-
580 30,750,000) showing low nucleotide diversity and negative Tajima's D (Figure 3B, 3D), harbors
581 genes that contain missense mutations, *ubr1* (ubiquitin protein ligase E3 component n-recognin
582 1) and *entpd5a* (ectonucleoside triphosphate/diphosphohydrolase 5). *ubr1* modulates digestive
583 organ development (He et al., 2017), and *entpd5a* plays a critical role in bone formation as well
584 as in the regulation of phosphate homeostasis (Huitema et al., 2012), both in zebrafish.
585 Signatures of selection on chr 12 (one locus), chr 15 (five loci), and chr 19 (two loci)
586 encompass both known and novel genetic variants and candidate genes (Figure 4, S8, S9, S11)
587 with respect to those previously associated with ecological adaptation (Han et al., 2020). In
588 agreement with previous research (Han et al., 2020; Lamichhaney et al., 2017; Martinez Barrio et
589 al., 2016), we found that several genetic variants in genes such as *tshr, herpud2, sox11b, esr2a*
590 and *syne2* are shared between NW and NE Atlantic populations. This result confirms that these
591 genes are strong candidates to be involved in seasonal reproduction in Atlantic herring, consistent
592 with the fact that several of them have a known function in reproduction in birds and mammals
593 (Bondesson et al., 2015; Melamed et al., 2012; Ono et al., 2008). Moreover, we discovered novel
594 genetic variants, in particular, in *myhc* in-tandem copies (i.e., *myhc*-like, *myhz1.1* and *myh2*),
595 which are involved in myogenesis (Figure 4D), and in *esr2a* (estrogen receptor beta), which is

596 essential for female reproduction (Figure 4E) (Lu et al., 2017). Interestingly, muscle development
597 and estrogen action are strongly affected by temperature and/or photoperiod (Jin et al., 2010;
598 Johnston et al., 2001), two of the most contrasting environmental conditions between spring and
599 fall seasons. Overall, the spring season is characterized by colder temperatures and increasing
600 daylength in contrast to the fall season, which is characterized by relatively warmer temperatures
601 and decreasing daylength. Considering the potential role of these genes and the presence of
602 genetic variants with particular allele frequency patterns and some seemingly private (Figure
603 S11), we infer that herring populations are locally adapted to the environmental conditions in the
604 NW Atlantic.

605 Local adaptation to the NW Atlantic is further supported by the discovery of new loci and
606 candidate genes associated with ecological adaptation, with potential important roles in
607 embryological development and gene expression regulation (Figure S8, S9, S11). For example,
608 on chr 12, *sbno1* (strawberry notch homolog 1) is a gene required for normal development of the
609 zebrafish central nervous system (Takano et al., 2010), and *denr* (density-regulated protein) is a
610 highly conserved gene involved in translation initiation (Skabkin et al., 2010). In chr15, *rab15*
611 (member RAS oncogene family 15) presumably participates in cellular response to insulin
612 stimulus and protein metabolism (Bradford et al., 2022), *sipa1l1* (signal induced proliferation
613 associated 1 like 1) is likely involved in vertebrate embryogenesis (Tsai et al., 2007), and *flrt2*
614 (fibronectin leucine-rich transmembrane protein 2) regulates embryonic heart morphogenesis in
615 mice (Müller et al., 2011). In chr19, a long-non coding RNA (lncRNA) gene likely participates in
616 the regulation of gene expression and is characteristic of the NW Atlantic. Indeed, lncRNA genes
617 were often present in several of the genomic regions under selection (Figure 3C, 3D, S8G, S8F,
618 S11C, S11D). lncRNA genes encode RNA molecules longer than 200 bp that are not translated
619 into protein. They are involved in gene expression regulation by directly intervening in chromatin
620 remodeling, transcriptional activation, transcriptional interference, RNA processing, and mRNA
621 translation (Statello et al., 2021; Zhang et al., 2019). Thus, an interesting venue for future
622 research could focus on understanding the role of lncRNAs in local adaptation of Atlantic
623 herring.
624
625 ***Genomic regions associated with adaptation to spatial environmental variation***

Fall spawning populations breeding north or south of the biogeographic transition zone on the Scotian shelf (~44.61°N) (Stanley et al., 2018), differ in genetic variants spanning a large single locus on chr 12 that corresponds to a known 7.8-Mbp long chromosomal inversion (Figure 6). Spring spawning populations were restricted to the Gulf of St Lawrence, an area north of the transition zone. All spring and fall spawning populations north of the transition zone had the "north" haplotype in high frequency, the "south" haplotype was prevalent in the southernmost location in Maine (ME4-F), and populations south of the transition zone in Musquodoboit harbor (MUS-F) and German Banks (GEB-F) had intermediate frequencies (Figure 6A, 6B).

While the role of this inversion is unknown, it has been proposed that it is associated with adaptation to temperature at spawning (Pettersson et al., 2019). Our Genome-Environment Association analysis indicated that minimal temperature during winter is the most concordant environmental factor with the spatial pattern of genomic differentiation driven by the putative inversion on chromosome 12 (Figure 6C, S10, S11) ($P = 0.005$, alpha = 0.01). The strong thermal gradient in this area results in those populations north of the biogeographic break experiencing significantly colder temperatures on average than those south of the break (Figure 6D). This pattern agrees with other studies comparing genetic and environmental gradients in this area (Lehnert et al., 2018; Stanley et al., 2018). Temperature is a critical factor influencing fish physiology (Reynolds & Casterlin, 1980) and distribution (García Molinos et al., 2016). Temperature has a particularly strong influence on the physiology of early life stages, which, due physiological constrains and high mortality during this phase, are particularly sensitive to thermal conditions (Dahlke et al., 2020; Marr, 1956). Correspondingly, our results suggest that selection associated with early life survival during the coldest months of the year (i.e., post-settlement mortality) may underlie patterns of genetic structure. Connectivity during the early life history of herring in the NW Atlantic has been generally assumed to be limited, corresponding to fine spatial scales of population structure (Stephenson et al., 2009). Our results correspond to and further illustrate this hypothesis, whereby emergent genetic patterns associated with adult fish correspond to processes likely imparting selection on earlier life history phases.

The prevalence of intermediate frequencies near the transition zone suggests that climatic conditions may vary between years in relation to oceanographic regional trends (Townsend et al., 2004). In a study of genetic structure of the invasive European green crab (*Carcinus meanus*),

656 Jeffery et al. (2018) similarly identified intermediate populations proximate to the biogeographic
657 break within our focal area. It is possible that populations at these locations experience significant
658 inter-annual environmental fluctuations during winter, depending on the strength of the warm
659 Gulf Stream flowing north or of the cold Labrador Current flowing south. Thus, we infer that
660 balancing selection may be favoring both inversion haplotypes at locations near the transition
661 zone.
662
663 ***Implications***
664 The results of this work have numerous implications in evolutionary biology and fisheries
665 management of Atlantic herring, and for other marine species with comparable life histories. This
666 study expands our understanding of the genomic basis of adaptation with gene flow by: (i)
667 contributing a list of candidate genetic variants and genes underlying intraspecific diversity; (ii)
668 providing genetic evidence supporting the hypothesis that strong selection can maintain
669 molecular divergence at loci underlying variation of phenotypic traits and fitness in large
670 populations despite the presence of gene flow; (iii) illustrating how such selection can be driven
671 by ecological and environmental factors varying at different temporal and spatial scales; and (iv)
672 confirming the central role of structural variants in adaptation with gene flow. This knowledge is
673 important to improve our ability to predict adaptive responses of wild populations in a scenario of
674 global warming. This study also illustrates the complexity predicting how climate change will
675 influence marine populations that exhibit fine-scale, environmentally mediated population
676 structure, whereby changes in temperature could elicit asymmetric responses among population
677 subgroups and potentially disrupting synchrony between reproductive timing, larval emergence
678 and seasonal prey availability (sensu (Cushing, 1990) – 'Match-Mismatch hypothesis').
679 Our results highlight the genetic underpinnings of the fine-scale population structure that
680 has long-been associated with herring in the NW Atlantic (Stephenson et al., 2009). In a period of
681 declining abundance, a robust assessment of genetic differentiation provides critical information
682 for management to conserve intraspecific variation. As we now have the molecular basis to
683 distinguish various herring ecotypes, a subset of outlier SNP markers described here can be used
684 for assessing stock composition. Such approach could help to monitor mixed stock fishing

amongst management regions helping to minimize the risk of overexploitation of vulnerable components (Bekkevold et al., 2015; Hemmer-Hansen et al., 2019; Larson et al., 2014).

Acknowledgements

We thank staff at Fisheries and Oceans Canada, the Maine Department of Resources, Comeau's Seafoods Ltd, Cape Breeze Seafoods Ltd, and fishers Crystal and Donald Kent, Gordon McKay, and many others for their valuable contribution in sample and tissue collection. Thanks to Gregory McCracken for assistance in sampling and DNA extraction, to Gavin Douglas, Emma Sylvester, Sarah Lehnert, Simone Fior, Alan Bergland, Miguel Carneiro, Michael Blum, and Mathieu Gautier for data analysis assistance. We thank Zeliang Wang and David Brickman from the Bedford Institute of Oceanography for access to BNAM model outputs. Thanks to Sean Rogers and Rob Beiko for constructive comments that helped improve this manuscript. Thanks to Luis Fuentes, Edith Pardo, Gloria Fuentes, Scott Campbell, Cathy Campbell, Meagan Campbell, John Sherlock, and the Ruzzante Lab for enriching discussions and support. Computations were mainly carried out on the supercomputer Mp2 of the University of Sherbrooke, managed by Calcul Québec and Compute Canada and funded by the Canada Foundation for Innovation (CFI), the ministère de l'Économie, de la science et de l'innovation du Québec (MESI) and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT). The analyses comparing northeast and northwest Atlantic populations were performed using the computational infrastructure provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973. A.P.F.P. and D.E.R. thank the Killam Trust. A.P.F.P. thanks the Vanier Canada Graduate Scholarship, the President's Award of Dalhousie University, the Nova Scotia Graduate Scholarship, the Lett Fund and a Strategic grant to D.E.R for graduate studies funding. This study was funded by NSERC Discovery and Strategic grants to D.E.R.

Data Archiving Statement

Sequence data generated in this study is available in the BioProject PRJNA930418 of the NCBI. Oceanographic data per location used in GEA and population-allele frequencies per SNP and pool are available in Dryad *(code will be provided upon manuscript approval)*. Custom scripts

are available in the Github repository:

https://github.com/apfuentes/2023_NWAtlanticHerring_PopGen.

Competing interest statement. The authors declare no competing interest.

REFERENCES

Andersson, L., Ryman, N., Rosenberg, R., & Ståhl, G. (1981). Genetic variability in Atlantic herring (Clupea harengus harengus): description of protein loci and population data. *Hereditas*, *95*(1), 69–78. https://doi.org/10.1111/j.1601-5223.1981.tb01330.x

André, C., Larsson, L. C., Laikre, L., Bekkevold, D., Brigham, J., Carvalho, G. R., Dahlgren, T. G., Hutchinson, W. F., Mariani, S., Mudde, K., Ruzzante, D. E., & Ryman, N. (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*(2), 270–280. https://doi.org/10.1038/hdy.2010.71

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bekkevold, D., André, C., Dahlgren, T. G., Clausen, L. A. W., Torstensen, E., Mosegaard, H., Carvalho, G. R., Christensen, T. B., Norlinder, E., & Ruzzante, D. E. (2005). Environmental correlates of population differentiation in Atlantic herring. *Evolution*, *59*(12), 2656–2668. https://doi.org/10.1111/j.0014-3820.2005.tb00977.x

Bekkevold, D., Helyar, S. J., Limborg, M. T., Nielsen, E. E., Hemmer-Hansen, J., Clausen, L. A. W., & Carvalho, G. R. (2015). Gene-associated markers can assign origin in a weakly structured fish, Atlantic herring. *ICES Journal of Marine Science: Journal Du Conseil*, *72*(6), 1790–1801. https://doi.org/10.1093/icesjms/fsu247

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). *Molecular Ecology*, *24*(13), 3299–3315. https://doi.org/10.1111/mec.13245

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS Genetics*, *10*(11), e1004775. https://doi.org/10.1371/journal.pgen.1004775

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bondesson, M., Hao, R., Lin, C.-Y., Williams, C., & Gustafsson, J.-Å. (2015). Estrogen receptor signaling during vertebrate development. *Biochim Biophys Acta*, *1849*(2), 142–151. https://doi.org/http://dx.doi.org/10.1016/j.bbagrm.2014.06.005

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., & Bentzen, P. (2013). Genomic islands of divergence and their consequences for the

resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, *6*(3), 450–461. https://doi.org/10.1111/eva.12026

Bradford, Y. M., van Slyke, C. E., Ruzicka, L., Singer, A., Eagle, A., Fashena, D., Howe, D. G., Frazer, K., Martin, R., Paddock, H., Pich, C., Ramachandran, S., & Westerfield, M. (2022). Zebrafish information network, the knowledgebase for *Danio rerio* research. *Genetics*, *220*(4). https://doi.org/10.1093/genetics/iyac016

Brickman, D., Hebert, D., & Wang, Z. (2018). Mechanism for the recent ocean warming events on the Scotian Shelf of eastern Canada. *Continental Shelf Research*, *156*, 11–22. https://doi.org/10.1016/j.csr.2018.01.001

Britten, G. L., Dowd, M., & Worm, B. (2016). Changing recruitment capacity in global fish stocks. *Proceedings of the National Academy of Sciences*, *113*(1), 134–139. https://doi.org/10.1073/pnas.1504709112

Broad Institute. (2016). *Understanding and adapting the generic hard-filtering recommendations*. https://gatkforums.broadinstitute.org/gatk/discussion/6925/understanding-and-adapting-the-generic-hard-filtering-recommendations

Broad Institute. (2018). *Picard tools*. http://broadinstitute.github.io/picard/

Bucholtz, R. H., Tomkiewicz, J., & Dalskov, J. (2008). Manual to determine gonadal maturity of herring (Clupea harengus L.). In *DTU Aqua-report*. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Manual+to+determine+gonadal+maturity+of+herring+(Clupea+harengus+L.)#0

Chen, J., Bi, H., Pettersson, M. E., Sato, D. X., Fuentes-Pardo, A. P., Mo, C., Younis, S., Wallerman, O., Jern, P., Molés, G., Gómez, A., Kleinau, G., Scheerer, P., & Andersson, L. (2021). Functional differences between TSHR alleles associate with variation in spawning season in Atlantic herring. *Communications Biology*, *4*(1). https://doi.org/10.1038/s42003-021-02307-7

Cushing, D. H. (1990). Plankton Production and Year-class Strength in Fish Populations: an Update of the Match/Mismatch Hypothesis. *Advances in Marine Biology*, *26*, 249–293. https://doi.org/10.1016/S0065-2881(08)60202-3

Dahlke, F. T., Wohlrab, S., Butzin, M., & Pörtner, H.-O. (2020). Thermal bottlenecks in the life cycle define climate vulnerability of fish. *Science*, *369*(6499), 65–70. https://doi.org/10.1126/science.aaz3658

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

DFO. (2022, March 30). *Rebuilding key forage fish stocks for healthier East Coast fisheries*. Https://Www.Canada.ca/En/Fisheries-Oceans/News/2022/03/Rebuilding-Key-Forage-Fish-Stocks-for-Healthier-East-Coast-Fisheries.Html.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*(6), 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, *36*(16), e105–e105. https://doi.org/10.1093/nar/gkn425

Dorant, Y., Laporte, M., Rougemont, Q., Cayuela, H., Rochette, R., & Bernatchez, L. (2022). Landscape genomics of the American lobster (*Homarus americanus*). *Molecular Ecology*. https://doi.org/10.1111/mec.16653

Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(4). https://doi.org/10.18637/jss.v022.i04

Engelhard, G. H., & Heino, M. (2004). Maturity changes in Norwegian spring-spawning herring before, during, and after a major population collapse. *Fisheries Research*, *66*(2–3), 299–310. https://doi.org/10.1016/S0165-7836(03)00195-4

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

FAO. (2019). *Species Fact Sheets: Clupea harengus (Linnaeus, 1758)*. http://www.fao.org/fishery/species/2886/en

Feder, A. F., Petrov, D. A., & Bergland, A. O. (2012). LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data. *PLoS ONE*, *7*(11), e48588. https://doi.org/10.1371/journal.pone.0048588

Fichefet, T., & Maqueda, M. A. M. (1997). Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *Journal of Geophysical Research: Oceans*, *102*(C6), 12609–12646. https://doi.org/10.1029/97JC00480

Gaggiotti, O. E., Bekkevold, D., Jørgensen, H. B. H., Foll, M., Carvalho, G. R., Andre, C., & Ruzzante, D. E. (2009). Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, *63*(11), 2939–2951. https://doi.org/10.1111/j.1558-5646.2009.00779.x

García Molinos, J., Halpern, B. S., Schoeman, D. S., Brown, C. J., Kiessling, W., Moore, P. J., Pandolfi, J. M., Poloczanska, E. S., Richardson, A. J., & Burrows, M. T. (2016). Climate velocity and the future global redistribution of marine biodiversity. *Nature Climate Change*, *6*(1), 83–88. https://doi.org/10.1038/nclimate2769

Guo, B., DeFaveri, J., Sotelo, G., Nair, A., & Merilä, J. (2015). Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biology*, *13*, 19. https://doi.org/10.1186/s12915-015-0130-8

Guo, B., Li, Z., & Merilä, J. (2016). Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular Ecology*, *25*(12), 2833–2852. https://doi.org/10.1111/mec.13657

Han, F., Jamsandekar, M., Pettersson, M. E., Su, L., Fuentes-Pardo, A. P., Davis, B. W., Bekkevold, D., Berg, F., Casini, M., Dahle, G., Farrell, E. D., Folkvord, A., & Andersson, L. (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *ELife*, *9*, 1–20. https://doi.org/10.7554/eLife.61076

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

He, J., Yang, Y., Zhang, J., Chen, J., Wei, X., He, J., & Luo, L. (2017). Ribosome biogenesis protein Urb1 acts downstream of mTOR complex 1 to modulate digestive organ development in zebrafish. *Journal of Genetics and Genomics*, *44*(12), 567–576. https://doi.org/10.1016/j.jgg.2017.09.013

Hemmer-Hansen, J., Hüssy, K., Baktoft, H., Huwer, B., Bekkevold, D., Haslob, H., Herrmann, J., Hinrichsen, H., Krumme, U., Mosegaard, H., Nielsen, E. E., Reusch, T. B. H., Storr-Paulsen, M., Velasco, A., von Dewitz, B., Dierking, J., & Eero, M. (2019). Genetic analyses reveal complex dynamics within a marine fish management area. *Evolutionary Applications*, *November 2018*, eva.12760. https://doi.org/10.1111/eva.12760

Hendry, A. P., Morbey, Y. E., Berg, O. K., & Wenburg, J. K. (2004). Adaptive variation in senescence: reproductive lifespan in a wild salmon population. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1536), 259–266. https://doi.org/10.1098/rspb.2003.2600

Hivert, V. (2018). *Measuring genetic differentiation from Pool-seq data*.

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62–82. https://doi.org/10.1111/mec.15720

Huitema, L. F. A., Apschner, A., Logister, I., Spoorendonk, K. M., Bussmann, J., Hammond, C. L., & Schulte-Merker, S. (2012). *Entpd5* is essential for skeletal mineralization and regulates phosphate homeostasis in zebrafish. *Proceedings of the National Academy of Sciences*, *109*(52), 21372–21377. https://doi.org/10.1073/pnas.1214231110

Jeffery, N. W., Bradbury, I. R., Stanley, R. R. E., Wringe, B. F., van Wyngaarden, M., Lowen, J. ben, McKenzie, C. H., Matheson, K., Sargent, P. S., & DiBacco, C. (2018). Genomewide evidence of environmentally mediated secondary contact of European green crab (Carcinus maenas) lineages in eastern North America. *Evolutionary Applications*, *11*(6), 869–882. https://doi.org/10.1111/eva.12601

Jin, Y., Shu, L., Sun, L., Liu, W., & Fu, Z. (2010). Temperature and photoperiod affect the endocrine disruption effects of ethinylestradiol, nonylphenol and their binary mixture in zebrafish (Danio rerio). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *151*(2), 258–263. https://doi.org/10.1016/j.cbpc.2009.11.004

Johansen, T., Besnier, F., Quintela, M., Jorde, P. E., Glover, K. A., Westgaard, J., Dahle, G., Lien, S., & Kent, M. P. (2020). Genomic analysis reveals neutral and adaptive patterns that challenge the current management regime for East Atlantic cod *Gadus morhua L*. *Evolutionary Applications*, *13*(10), 2673–2688. https://doi.org/10.1111/eva.13070

Johnston, I., Vieira, V., & Temple, G. (2001). Functional consequences and population differences in the developmental plasticity of muscle to temperature in Atlantic herring Clupea harengus. *Marine Ecology Progress Series*, *213*, 285–300. https://doi.org/10.3354/meps213285

Jørgensen, H. B. H., Hansen, M. M., Bekkevold, D., Ruzzante, D. E., & Loeschcke, V. (2005). Marine landscapes and population genetic structure of herring (Clupea harengus L.) in the Baltic Sea. *Molecular Ecology*, *14*(10), 3219–3234. https://doi.org/10.1111/j.1365-294X.2005.02658.x

Kerr, Q., Fuentes-Pardo, A. P., Kho, J., McDermid, J. L., & Ruzzante, D. E. (2019). Temporal stability and assignment power of adaptively divergent genomic regions between herring (Clupea harengus) seasonal spawning aggregations. *Ecology and Evolution*, *9*, 500–510. https://doi.org/10.1002/ece3.4768

Kess, T., Einfeldt, A. L., Wringe, B., Lehnert, S. J., Layton, K. K. S., McBride, M. C., Robert, D., Fisher, J., le Bris, A., den Heyer, C., Shackell, N., Ruzzante, D. E., Bentzen, P., & Bradbury, I. R. (2021). A putative structural variant and environmental variation associated

with genomic divergence across the Northwest Atlantic in Atlantic Halibut. *ICES Journal of Marine Science*, *78*(7), 2371–2384. https://doi.org/10.1093/icesjms/fsab061

Kofler, R., Orozco-terWengel, P., de Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE*, *6*(1), e15925. https://doi.org/10.1371/journal.pone.0015925

Kolaczkowski, B., Kern, A. D., Holloway, A. K., & Begun, D. J. (2011). Genomic Differentiation Between Temperate and Tropical Australian Populations of Drosophila melanogaster. *Genetics*, *187*(1), 245–260. https://doi.org/10.1534/genetics.110.123059

Lamichhaney, S., Barrio, A. M., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E. R., Berglund, J., Wetterbom, A., Laikre, L., Webster, M. T., Grabherr, M., Ryman, N., & Andersson, L. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, *109*(47), 19345–19350. https://doi.org/10.1073/pnas.1216128109

Lamichhaney, S., Fuentes-Pardo, A. P., Rafati, N., Ryman, N., McCracken, G. R., Bourne, C., Singh, R., Ruzzante, D. E., & Andersson, L. (2017). Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proceedings of the National Academy of Sciences*, *114*(17), E3452–E3461. https://doi.org/10.1073/pnas.1617728114

Larson, W. A., Seeb, J. E., Pascal, C. E., Templin, W. D., & Seeb, L. W. (2014). Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (Oncorhynchus tshawytscha) from western Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, *71*(5), 698–708. https://doi.org/10.1139/cjfas-2013-0502

Lehnert, S. J., DiBacco, C., van Wyngaarden, M., Jeffery, N. W., ben Lowen, J., Sylvester, E. V. A., Wringe, B. F., Stanley, R. R. E., Hamilton, L. C., & Bradbury, I. R. (2018). Fine-scale temperature-associated genetic structure between inshore and offshore populations of sea scallop (Placopecten magellanicus). *Heredity*, 1–12. https://doi.org/10.1038/s41437-018-0087-9

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Levin, L. A. (2006). Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, *46*(3), 282–297. https://doi.org/10.1093/icb/icj024

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv*, *00*(00), 3. https://doi.org/arXiv:1303.3997 [q-bio.GN]

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Limborg, M. T., Helyar, SarahJ., de Bruyn, M., Taylor, M. I., Nielsen, E. E., Ogden, R., Carvalho, G. R., & Bekkevold, D. (2012). Environmental selection on transcriptome-derived

SNPs in a high gene flow marine fish, the Atlantic herring (Clupea harengus). *Molecular Ecology*, *21*(15), 3686–3703. https://doi.org/10.1111/j.1365-294X.2012.05639.x

Lowerre-Barbieri, S. K., Ganias, K., Saborido-Rey, F., Murua, H., & Hunter, J. R. (2011). Reproductive Timing in Marine Fishes: Variability, Temporal Scales, and Methods. *Marine and Coastal Fisheries*, *3*(1), 71–91. https://doi.org/10.1080/19425120.2011.556932

Lu, H., Cui, Y., Jiang, L., & Ge, W. (2017). Functional Analysis of Nuclear Estrogen Receptors in Zebrafish Reproduction by Genome Editing Approach. *Endocrinology*, *158*(7), 2292–2308. https://doi.org/10.1210/en.2017-00215

Madec, G., Delecluse, P., Imbard, M., & Levy, C. (1998). *OPA8.1 Ocean general Circulation Model reference manual*.

Marr, J. C. (1956). The "Critical Period" in the Early Life History of Marine Fishes. *ICES Journal of Marine Science*, *21*(2), 160–170. https://doi.org/10.1093/icesjms/21.2.160

Martinez Barrio, A., Lamichhaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H., Dainat, J., Ekman, D., Höppner, M., Jern, P., Martin, M., Nystedt, B., Liu, X., Chen, W., Liang, X., Shi, C., Fu, Y., Ma, K., Zhan, X., … Andersson, L. (2016). The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *ELife*, *5*(MAY2016), 1–32. https://doi.org/10.7554/eLife.12081

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McPherson, A. A., O'Reilly, P. T., & Taggart, C. T. (2004). Genetic Differentiation, Temporal Stability, and the Absence of Isolation by Distance among Atlantic Herring Populations. *Transactions of the American Fisheries Society*, *133*(2), 434–446. https://doi.org/10.1577/02-106

McPherson, A. A., Stephenson, R. L., & Taggart, C. T. (2003). Genetically different Atlantic herring Clupea harengus spawning waves. *Marine Ecology Progress Series*, *247*, 303–309. https://doi.org/10.3354/meps247303

McPherson, A., Stephenson, R. L., O'Reilly, P. T., & Jones, M. W. T. C. T. (2001). Genetic diversity of coastal Northwest Atlantic herring populations: implications for management. *Journal of Fish Biology*, *59*(SUPPL. A), 356–370. https://doi.org/10.1006/jfbi.2001.1769

McQuinn, I. H. (1997). Metapopulations and the Atlantic herring. *Reviews in Fish Biology and Fisheries*, *7*(3), 297–329. https://doi.org/10.1023/A:1018491828875

Melamed, P., Savulescu, D., Lim, S., Wijeweera, A., Luo, Z., Luo, M., & Pnueli, L. (2012). Gonadotrophin-Releasing Hormone Signalling Downstream of Calmodulin. *Journal of Neuroendocrinology*, *24*(12), 1463–1475. https://doi.org/10.1111/j.1365-2826.2012.02359.x

Migaud, H., Davie, A., & Taylor, J. F. (2010). Current knowledge on the photoneuroendocrine regulation of reproduction in temperate fish species. *Journal of Fish Biology*, *76*(1), 27–68. https://doi.org/10.1111/j.1095-8649.2009.02500.x

Müller, P.-S., Schulz, R., Maretto, S., Costello, I., Srinivas, S., Bikoff, E., & Robertson, E. (2011). The fibronectin leucine-rich repeat transmembrane protein Flrt2 is required in the epicardium to promote heart morphogenesis. *Development*, *138*(7), 1297–1308. https://doi.org/10.1242/dev.059386

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342. https://doi.org/10.1098/rstb.2011.0263

Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, *59*(4), 705–719. https://doi.org/10.1111/j.0014-3820.2005.tb01747.x

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2015). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), btv566. https://doi.org/10.1093/bioinformatics/btv566

Ono, H., Hoshino, Y., Yasuo, S., Watanabe, M., Nakane, Y., Murai, A., Ebihara, S., Korf, H.-W., & Yoshimura, T. (2008). Involvement of thyrotropin in photoperiodic signal transduction in mice. *Proceedings of the National Academy of Sciences, USA*, *105*(47), 18238–18242. https://doi.org/10.1073/pnas.0808952105

Overholtz, W. J. (2002). The Gulf of Maine–Georges Bank Atlantic herring (Clupea harengus): spatial pattern analysis of the collapse and recovery of a large marine fish complex. *Fisheries Research*, *57*(3), 237–254. https://doi.org/10.1016/S0165-7836(01)00359-9

Palumbi, S. R. (1994). Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics*, *25*(1), 547–572. https://doi.org/10.1146/annurev.es.25.110194.002555

Pettersson, M. E., Rochus, C. M., Han, F., Chen, J., Hill, J., Wallerman, O., Fan, G., Hong, X., Xu, Q., Zhang, H., Liu, S., Liu, X., Haggerty, L., Hunt, T., Martin, F. J., Flicek, P., Bunikis, I., Folkvord, A., & Andersson, L. (2019). A chromosome-level assembly of the Atlantic herring genome—detection of a supergene and other signals of selection. *Genome Research*, *29*(11), 1919–1928. https://doi.org/10.1101/gr.253435.119

R Core Development Team. (2023). *R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing*.

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, Illinois, USA. https://cran.r-project.org/package=psych

Reynolds, W. W., & Casterlin, M. E. (1980). The Role of Temperature in the Environmental Physiology of Fishes. In *Environmental Physiology of Fishes* (pp. 497–518). Springer US. https://doi.org/10.1007/978-1-4899-3659-2_19

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

Rubin, C.-J., Enbody, E. D., Dobreva, M. P., Abzhanov, A., Davis, B. W., Lamichhaney, S., Pettersson, M., Sendell-Price, A. T., Sprehn, C. G., Valle, C. A., Vasco, K., Wallerman, O., Grant, B. R., Grant, P. R., & Andersson, L. (2022). Rapid adaptive radiation of Darwin's finches depends on ancestral genetic modules. *Science Advances*, *8*(27). https://doi.org/10.1126/sciadv.abm5982

Ruzzante, D. E., Mariani, S., Bekkevold, D., André, C., Mosegaard, H., Clausen, L. A. W., Dahlgren, T. G., Hutchinson, W. F., Hatfield, E. M. C., Torstensen, E., Brigham, J., Simmonds, E. J., Laikre, L., Larsson, L. C., Stet, R. J. M., Ryman, N., & Carvalho, G. R. (2006). Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1593), 1459–1464. https://doi.org/10.1098/rspb.2005.3463

Sambrook, J., & Russell, D. W. (2006). Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harbor Protocols*, *2006*(1). https://doi.org/10.1101/pdb.prot4455

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763. https://doi.org/10.1038/nrg3803

Simmonds, E. J. (2007). Comparison of two periods of North Sea herring stock management: success, failure, and monetary value. *ICES Journal of Marine Science*, *64*(4), 686–692. https://doi.org/10.1093/icesjms/fsm045

Sinclair, M., & Tremblay, M. J. (1984). Timing of Spawning of Atlantic Herring (Clupea harengus harengus) Populations and the Match–Mismatch Theory. *Canadian Journal of Fisheries and Aquatic Sciences*, *41*(7), 1055–1065. https://doi.org/10.1139/f84-123

Skabkin, M. A., Skabkina, O. v., Dhote, V., Komar, A. A., Hellen, C. U. T., & Pestova, T. v. (2010). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes & Development*, *24*(16), 1787–1801. https://doi.org/10.1101/gad.1957510

Stanley, R. R. E., DiBacco, C., Lowen, B., Beiko, R. G., Jeffery, N. W., van Wyngaarden, M., Bentzen, P., Brickman, D., Benestan, L., Bernatchez, L., Johnson, C., Snelgrove, P. V. R., Wang, Z., Wringe, B. F., & Bradbury, I. R. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, *4*(3), eaaq0929. https://doi.org/10.1126/sciadv.aaq0929

Stanley, R. R. E., & Jeffery, N. W. (2017). *CartDist: Re-projection tool for complex marine systems*. https://doi.org/10.5281/zenodo.802875

Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, *22*(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Stephenson, R. L., Melvin, G. D., & Power, M. J. (2009). Population integrity and connectivity in Northwest Atlantic herring: a review of assumptions and evidence. *ICES Journal of Marine Science*, *66*(8), 1733–1739. https://doi.org/10.1093/icesjms/fsp189

Takano, A., Zochi, R., Hibi, M., Terashima, T., & Katsuyama, Y. (2010). Expression of strawberry notch family genes during zebrafish embryogenesis. *Developmental Dynamics*, *239*(6), 1789–1796. https://doi.org/10.1002/dvdy.22287

Teacher, A. G., André, C., Jonsson, P. R., & Merilä, J. (2013). Oceanographic connectivity and environmental correlates of genetic structuring in Atlantic herring in the Baltic Sea. *Evolutionary Applications*, *6*(3), 549–567. https://doi.org/10.1111/eva.12042

Thorrold, S. R. (2001). Natal Homing in a Marine Fish Metapopulation. *Science*, *291*(5502), 297–299. https://doi.org/10.1126/science.291.5502.297

Townsend, D. W., Thomas, A. C., Mayer, L. M., Thomas, M. A., & Quinlan, J. A. (2004). Oceanography of the Northwest Atlantic Shelf (1, W). In A. R. Robinson & K. H. Brink (Eds.), *The Sea: The Global Coastal Ocean: Interdisciplinary Regional Studies and Syntheses* (pp. 1–57). Harvard University Press.

Tsai, I.-C., Amack, J. D., Gao, Z.-H., Band, V., Yost, H. J., & Virshup, D. M. (2007). A Wnt-CKIε-Rap1 Pathway Regulates Gastrulation by Modulating SIPA1L1, a Rap GTPase Activating Protein. *Developmental Cell*, *12*(3), 335–347. https://doi.org/10.1016/j.devcel.2007.02.009

Turcotte, F., Swain, D. P., McDermid, J. L., & DeJong, R. A. (2021). *Assessment of the NAFO Division 4TVn southern Gulf of St. Lawrence Atlantic Herring (Clupea harengus) in 2018-2019. DFO Can. Sci. Advis. Sec. Res. Doc. 2021/030. xiv + 158 p.*

Wang, Z., Brickman, D., Greenan, B. J. W., & Yashayaev, I. (2016). An abrupt shift in the Labrador Current System in relation to winter NAO events. *Journal of Geophysical Research: Oceans*, *121*(7), 5338–5349. https://doi.org/10.1002/2016JC011721

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358. https://doi.org/10.2307/2408641

Wheeler, J. P., & Winters, G. H. (1984). Homing of Atlantic herring (Clupea harengus harengus) in Newfoundland waters as indicated by tagging data. *Can. J. Fish. Aquat. Sci.*, *41*, 108–117.

Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., & Shen, F. (2019). Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. *International Journal of Molecular Sciences*, *20*(22), 5573. https://doi.org/10.3390/ijms20225573

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

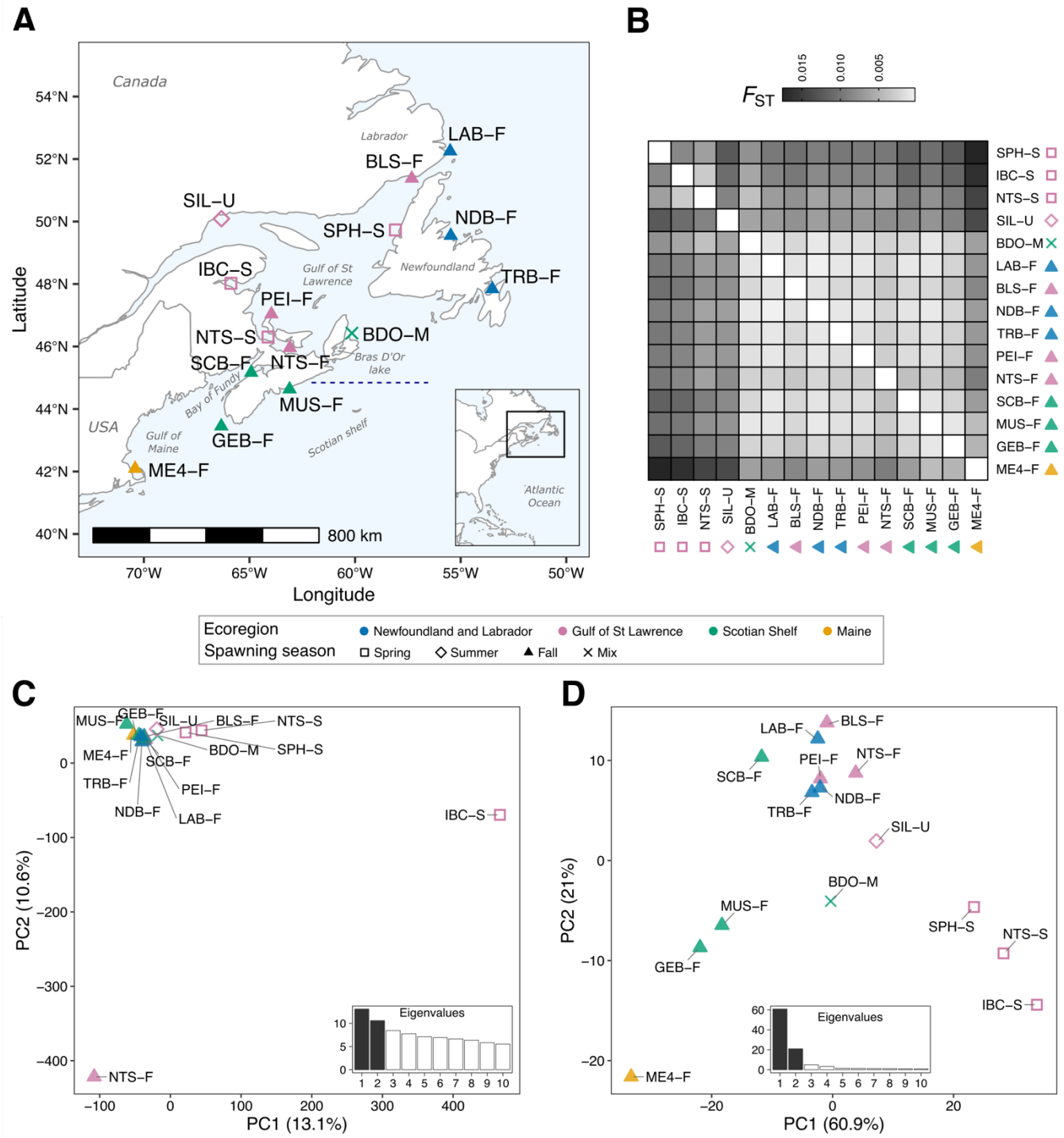

Figure 1. Sampling locations and population structure of northwest Atlantic herring. (**A**) Map depicting collection sites. Location name abbreviations as described in Table 1, in which the spawning season is indicated with the suffix "-S" for spring, "-F" for fall, "-U" for summer, and "-M" for mixed. Each point corresponds to a pool sample. Point colors indicate the designation of the sample to one of the major biogeographic units described for the Canadian Atlantic Ocean (DFO, 2009). Symbol shapes represent the predominant spawning season based on individual gonadal maturation status at the time of collection (Figure S1). The horizontal blue dashed line on the map indicates the approximate location of

a biogeographic transition zone in the eastern Scotian Shelf, 44.61°N ± 0.25 (Stanley et al., 2018). (**B**) Heatmap plot representing pairwise F_{ST} estimates based on pool allele frequencies of 5,073,572 SNPs (Table S2). Samples are ordered by season, and within season, by latitude. Cell shading represents the degree of genomic divergence between a pair of pool samples, which goes from lack (white) to high (black) differentiation. Principal component analysis plot based on undifferentiated (**C**) (n = 135 950) and highly differentiated (**D**) markers (n = 545). In both plots the first two axes or principal components (PCs) are shown, and the inset bar plots indicate the percentage of the variance explained by each of the first 10 PCs. Same as before, each point corresponds to a pool sample, its color indicates the designated biogeographic unit and its shape, the corresponding spawning season.

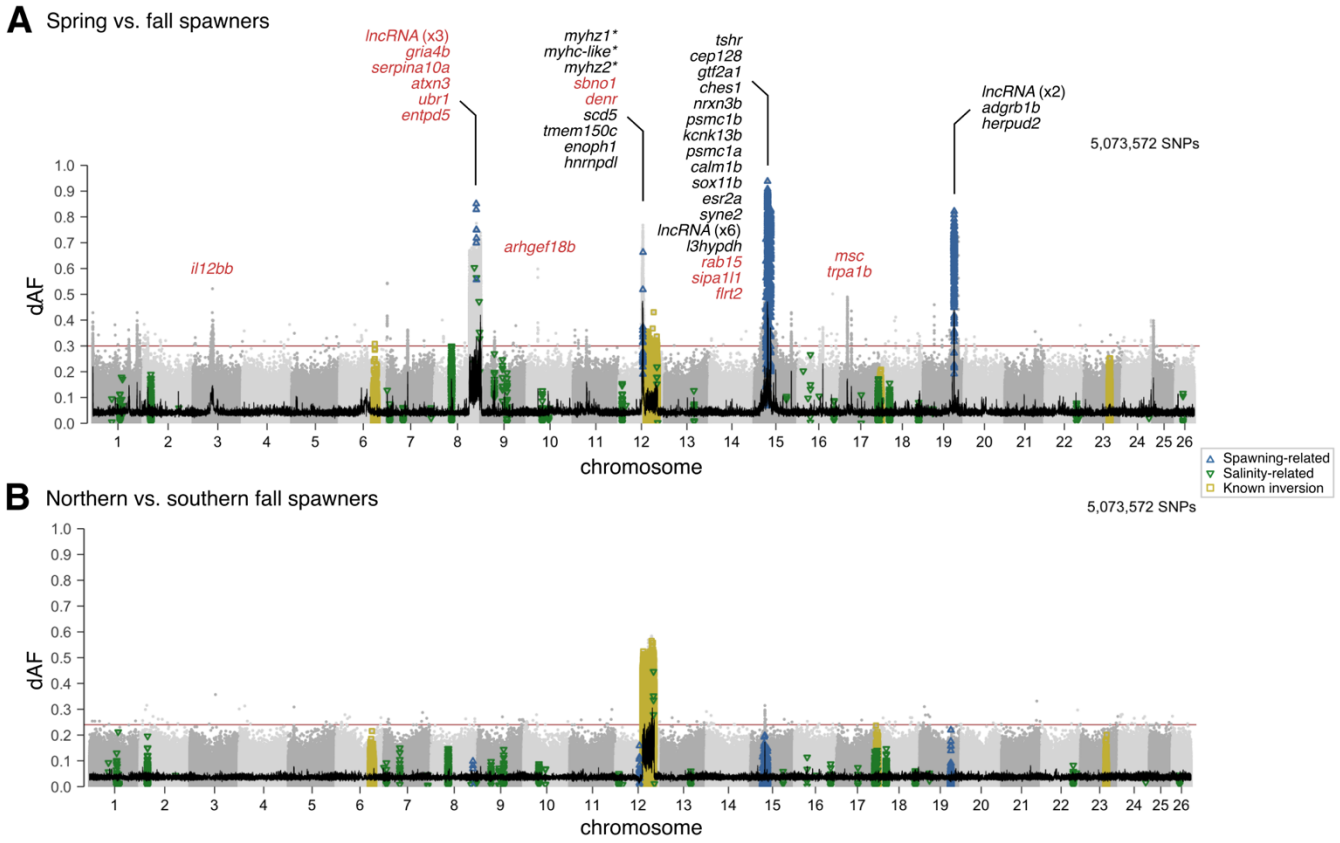

Figure 2. Genomic regions associated with adaptation to seasonal reproduction and a latitudinal environmental cline. Genetic differentiation (dAF) across the genome (**A**) between spring and fall spawners, and (**B**) between fall spawners in the north vs. south of the transition zone in eastern Scotian Shelf (Stanley et al., 2018). Each dot represents a single SNP. SNPs previously reported in (Han et al., 2020) as strongly associated with spawning season, salinity, and four known inversions (on chr 6, 12, 17, and 23) are denoted as empty blue upward triangles, green downward triangles, and yellow squares, respectively. Consecutive chromosomes are colored in intercalating shades of gray. The 100 SNP-rolling average of dAF is shown as a black line and the Bonferroni significance value across the genome is shown as a horizontal red line. Genes within ± 40 Kbp of the most divergent SNPs are shown on top of each informative locus. Gene names with an were inferred from homology with orthologous genes. Gene names colored in red correspond to genes within the associated genomic region.

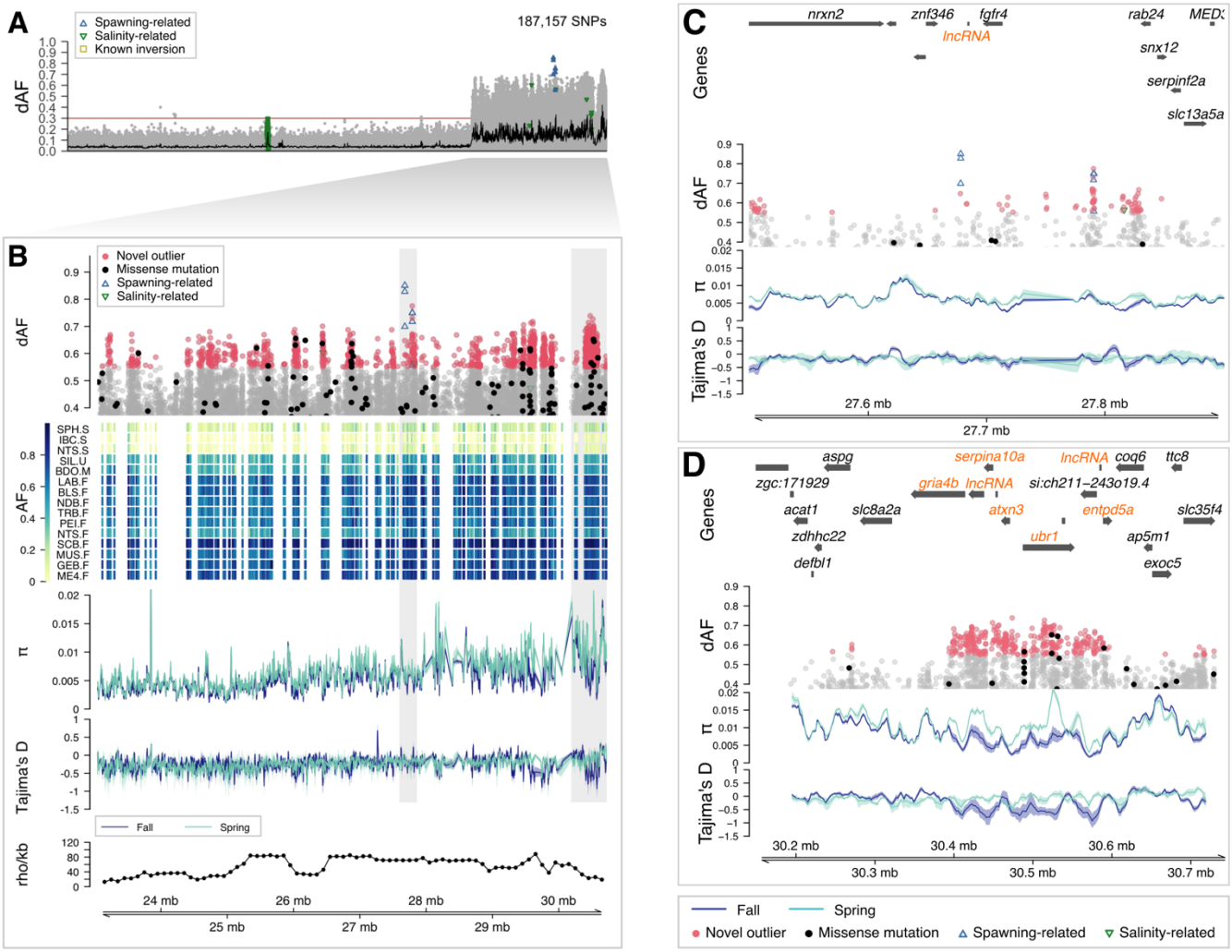

Figure 3. Novel putative inversion on chromosome 8 associated with seasonal reproduction. (**A**) Genetic differentiation (dAF) between spring and fall spawners across chromosome 8. Each dot is a single SNP. The black line is the rolling average of dAF over 100 SNPs, and the horizontal red line is the Bonferroni significance value. SNPs with known association with spawning, salinity, or four known inversions (in chr 6, 12, 17, and 23) (Han et al., 2020) are denoted with empty blue upward triangles, green downward triangles, and yellow squares, respectively. (**B**) Close-up plot to the putative inversion. This plot has five tracks, which show from top to bottom: genetic differentiation between spring and fall spawners for SNPs with dAF ≥ 0.4; heatmap plot depicting the minor allele frequency per population (rows) for the novel outlier SNPs (columns); average nucleotide diversity (π) and Tajima's D (window size 10 Kbp, step size 2 Kbp) for spring and fall spawners, in light and dark blue lines, respectively; and estimate of recombination rate (rho/Kbp) every 100 Kbp (Pettersson et al., 2019). Novel outlier SNPs (dAF ≥ 0.55) are denoted as red filled circles, missense mutations as filled black circles, and other SNPs are gray circles. Zoom-in into two informative intervals, (**C**) chr8:27,500,000-27,900,000, which contains the most differentiated SNPs, and (**D**) chr8:30,150,000-30,750,000. Genes in the associated region are highlighted in orange.

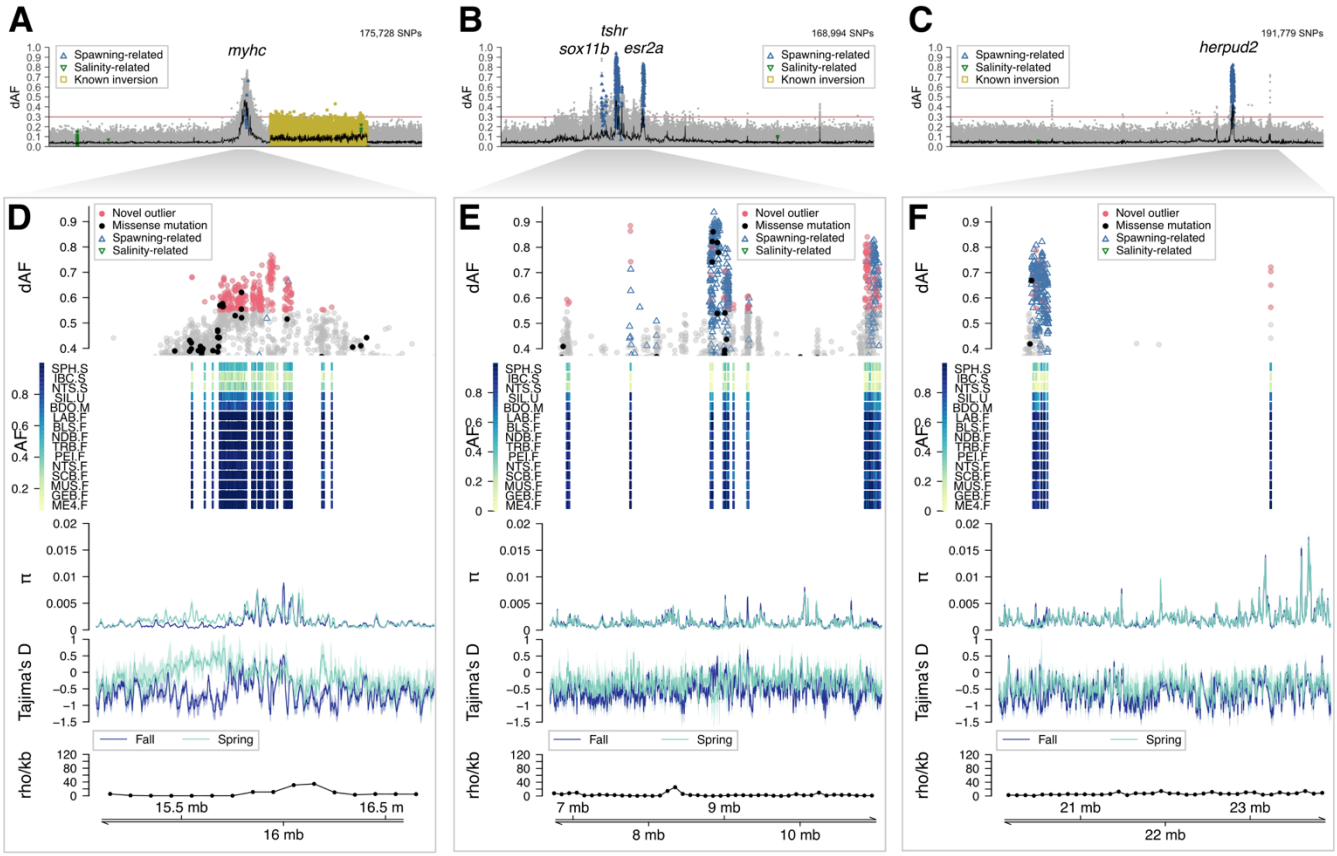

Figure 4. Signatures of selection associated with seasonal reproduction. Genetic differentiation (dAF) between spring and fall spawners across chromosomes (**A**) 12, (**B**) 15, and (**C**) 19. Each dot represents a single SNP. The rolling average of dAF over 100 SNPs is shown as a horizontal black line, and the Bonferroni significance value is shown as a horizontal red line. SNPs with described association with spawning, salinity, or four inversions (in chr 6, 12, 17, and 23) in Han et al. (2020) are indicated with empty blue upward triangles, green downward triangles, and yellow squares, respectively. Close-up to the signatures of selection in chromosomes (**D**) 12, (**E**) 15, (**F**) 19. Each plot has five tracks that show from top to bottom: genetic differentiation between spring and fall spawners for SNPs with dAF ≥ 0.4; heatmap plot depicting the minor allele frequency per population (rows) for the novel outlier SNPs (columns); average nucleotide diversity (π) and Tajima's D (window size 10 Kbp, step size 2 Kbp) for spring and fall spawners, in light and dark blue lines, respectively; and estimate of recombination rate (rho/Kbp) every 100 Kbp (Pettersson et al., 2019). Novel outlier SNPs (dAF ≥ 0.55) are shown as red filled circles, and missense mutations, as filled black circles. Other SNPs are shown in gray. For zoom-in plots to each peak of divergence including gene names see Supplementary Figures S8, S9, S11.

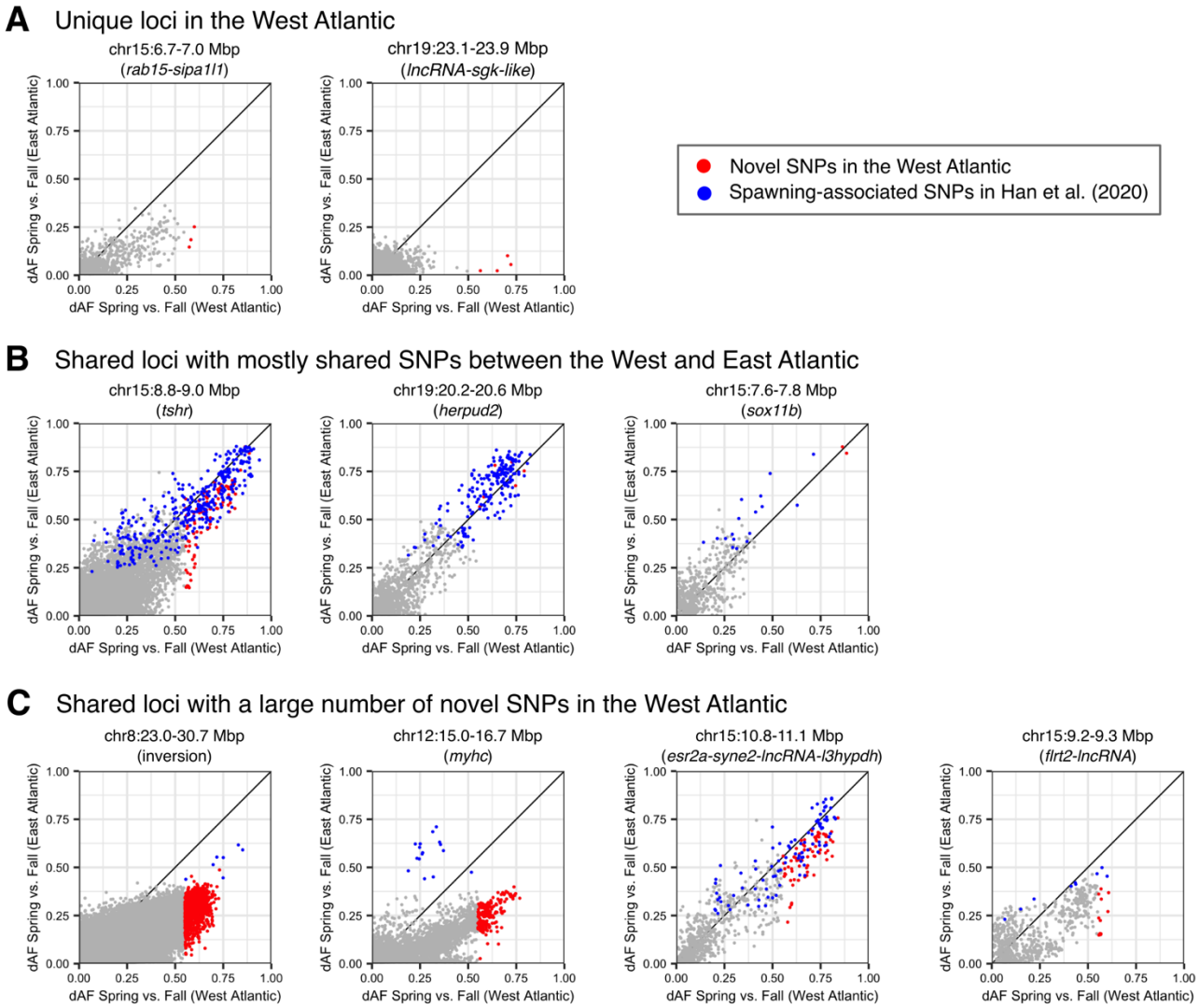

Figure 5. Comparison of the absolute difference in allele frequencies (dAF) between Spring and Fall spawners in the West and East Atlantic for nine genomic regions strongly associated with spawning time. (**A**) Unique loci in the West Atlantic. (**B**) Shared loci with mostly shared SNPs between the West and East Atlantic. (**C**) Shared loci with a large number of novel SNPs in the West Atlantic. Each dot represents a single SNP. Red dots indicate novel SNPs in the West Atlantic and blue dots indicate SNPs with known association with spawning time as reported in Han et al. (2020). The rest of SNPs in the region (not outliers, dAF in the West Atlantic <= 0.55) are shown in gray. A summary of the percentage of unique and shared loci and SNPs is shown in Table S6.

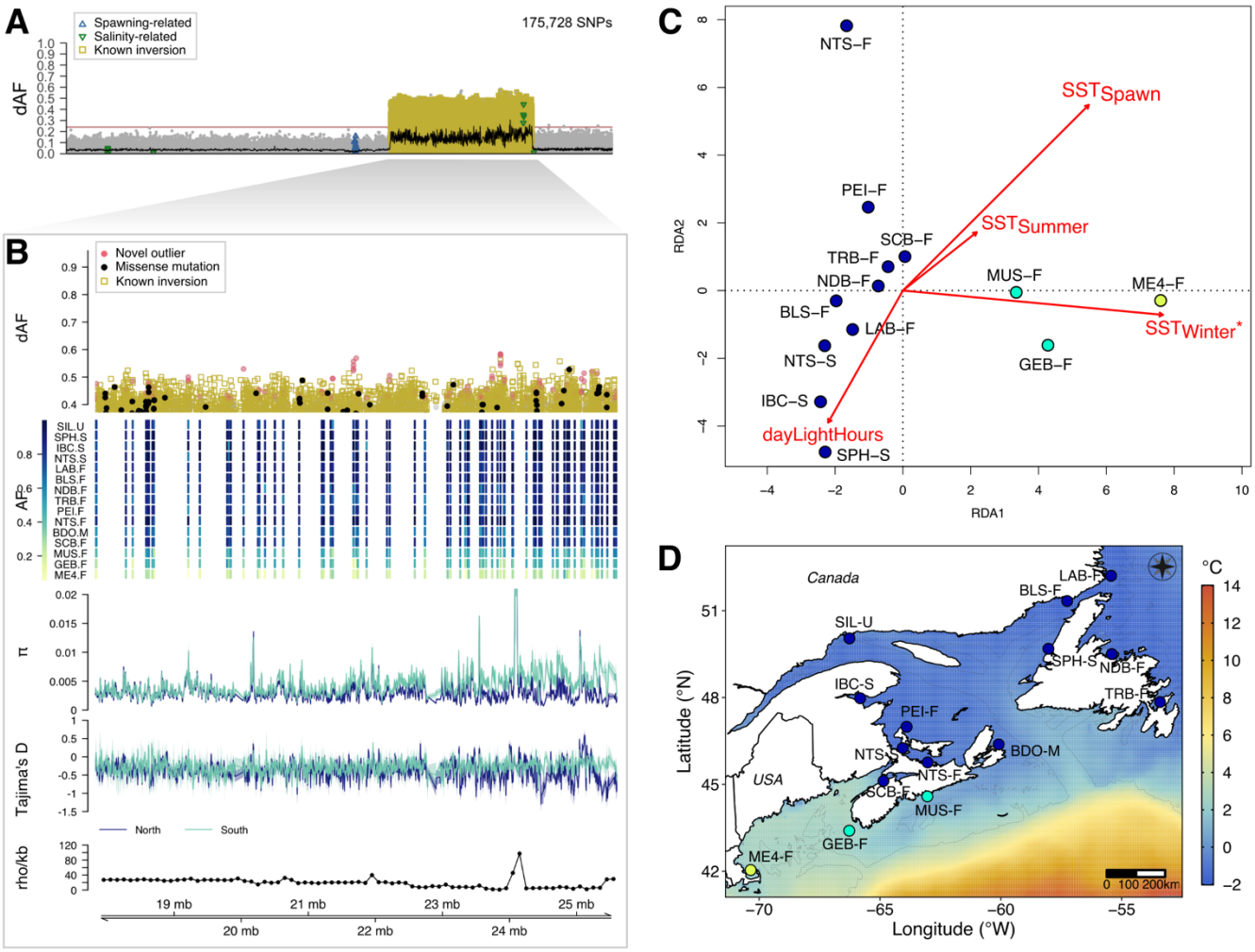

Figure 6. Chromosomal inversion on chromosome 12 associated with spatial genetic divergence along a latitudinal environmental cline. (**A**) Genetic differentiation (dAF) along the chromosome. (**B**) Close-up plot to the putative inversion (chr12:17826318-25603093) consisting of five tracks. The first track is the genetic differentiation for SNPs with dAF ≥ 0.4. Each dot is a single SNP, its shape indicates whether it is a novel outlier (red filled circle), a missense mutation (black circle), or is in one of four inversions reported in Han et al. (2020). The second track depicts the pool-minor allele frequency of the novel outlier SNPs, where each row is a single pool sample, and each column is a SNP. The third and fourth tracks show the profile of the average nucleotide diversity (π) and Tajima's D of spring and fall spawners, calculated in 10 Kbp sliding windows with a 2 Kbp step. The fifth track shows the recombination rate (rho/Kbp) every 100 Kbp (Pettersson et al., 2019). (**C**) Redundancy analysis plot representing the association between four uncorrelated environmental variables and population allele frequencies of the top outlier SNPs within the inversion on chr 12 (dAF > 0.55). Each circle corresponds to a spawning aggregation and their color indicates their predominant population allele frequency. Sample abbreviations and names as in Table 1. The red arrows (and their length) indicate the level of correlation of each environmental variable with genetic variation in the first two axes. Results of ANOVA to test statistical significance are shown in Table S5. (**D**) Map depicting average winter sea surface temperature and the predominant population allele frequencies at diagnostic SNPs within the inversion on chr 12 for the 15 spawning aggregations included in this study. Each circle corresponds to a

spawning aggregation and their color indicates their predominant population allele frequency as per in B) second track.

Tables

Table 1. Characteristics of the 15 herring spawning aggregations included in this study.

Locality	ID code	Sample size (N)	Longitude	Latitude	Sampling date	Spawning season	Source	BioSample accession
Seven Islands (Sept-Îles)	SIL-U	50	-66.33	50.09	06 June 2012	Summer	This study	SAMN33007718
Inner Bay Chaleur	IBC-S	41	-65.85	48.00	08 May 2014	Spring	(Lamichhaney et al., 2017), WInB	SAMN05589078
Stephenville	SPH-S	48	-57.94	49.73	30 May 2012	Spring	This study	SAMN33007720
Northumberland Strait	NTS-S	49	-64.12	46.30	06 May 2014	Spring	(Lamichhaney et al., 2017), WNsS	SAMN05589080
Northumberland Strait	NTS-F	50	-62.60	45.73	16 Sep 2014	Fall	(Lamichhaney et al., 2017), WNsF	SAMN05589079
Labrador	LAB-F	50	-55.50	52.25	24 Aug 2014, 22 Aug 2015	Fall	This study	SAMN33007723
Blanc Sablon	BLS-F	49	-57.31	51.38	13 Aug 2014	Fall	This study	SAMN33007724
Notre Dame Bay	NDB-F	50	-55.47	49.55	26 Oct 2015	Fall	This study	SAMN33007725
Trinity Bay	TRB-F	49	-53.47	47.84	28 Sep 2014	Fall	(Lamichhaney et al., 2017), WBob	SAMN05589075
Prince Edward Island	PEI-F	50	-63.96	47.04	25 Aug 2014	Fall	This study	SAMN33007727
Bras D'Or lake	BDO-M	50	-60.85	45.93	20 April 2016	Spring	This study	SAMN33007728
Scots Bays	SCB-F	50	-64.92	45.17	24 Aug 2015	Fall	This study	SAMN33007729
Musquodoboit	MUS-F	50	-63.10	44.63	28 Oct 2015	Fall	This study	SAMN33007730
German Banks	GEB-F	48	-66.18	43.16	28 Aug 2014	Fall	(Lamichhaney et al., 2017), WGeB	SAMN05589077
Maine fishing area 514	ME4-F	50	-70.41	42.09	19 Oct 2015	Fall	This study	SAMN33007732