ORIGINAL RESEARCH

Range-wide genomic data synthesis reveals transatlantic vicariance and secondary contact in Atlantic cod

Robert Fairweather[1,2]* [iD] | Ian R. Bradbury[3]* [iD] | Sarah J. Helyar[4] [iD] | Mark de Bruyn[2,5] [iD] |
Nina O. Therkildsen[6] [iD] | Paul Bentzen[1] [iD] | Jakob Hemmer-Hansen[7]† [iD] |
Gary R. Carvalho[2]† [iD]

[1]Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada

[2]School of Biological Sciences, Bangor University, Bangor, UK

[3]Science Branch, Department of Fisheries, St John's, Newfoundland and Labrador, Canada

[4]Institute of Global Food Security, Queen's University Belfast, Belfast, UK

[5]School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia

[6]Department of Natural Resources, Cornell University, Ithaca, New York

[7]Section for Marine Living Resources, National Institute for Aquatic Resources, Technical University of Denmark, Silkeborg, Denmark

Correspondence
Robert Fairweather, Department of Biology, Dalhousie University, Halifax, NS, Canada.
Email: robert.fairweather@dal.ca

Abstract

Recent advances in genetic and genomic analysis have greatly improved our understanding of spatial population structure in marine species. However, studies addressing phylogeographic patterns at oceanic spatial scales remain rare. In Atlantic cod (*Gadus morhua*), existing range-wide examinations suggest significant transatlantic divergence, although the fine-scale contemporary distribution of populations and potential for secondary contact are largely unresolved. Here, we explore transatlantic phylogeography in Atlantic cod using a data-synthesis approach, integrating multiple genome-wide single-nucleotide polymorphism (SNP) datasets representative of different regions to create a single range-wide dataset containing 1,494 individuals from 54 locations and genotyped at 796 common loci. Our analysis highlights significant transatlantic divergence and supports the hypothesis of westward post-glacial colonization of Greenland from the East Atlantic. Accordingly, our analysis suggests the presence of transatlantic secondary contact off eastern North America and supports existing perspectives on the phylogeographic history of Atlantic cod with an unprecedented combination of genetic and geographic resolution. Moreover, we demonstrate the utility of integrating distinct SNP databases of high comparability.

KEYWORDS
Atlantic cod, fish, *Gadus morhua*, marine, phylogeography, synthesis

1 | INTRODUCTION

The contemporary distribution of marine species has been shaped by both current and historical patterns in ocean climate and habitat availability (Kettle, Morales-Muñiz, Roselló-Izquierdo, Heinrich, & Vøllestad, 2011; Perry, Low, Ellis, & Reynolds, 2005). Genetic divergence across ocean basins has been reported in many marine species

consistent with the isolation of populations during the Pleistocene (Hewitt, 2003; Maggs et al., 2008), when periods of climatic cooling and glacial advance acted to segregate marine populations into glacial refugia (Provan & Bennett, 2008). For example, the last glacial maximum (LGM,~21 kya) saw the loss of most subarctic shelf sea habitat in the North Atlantic due to advancing glacier and lower sea levels (Lambeck, Esat, & Potter, 2002; Mitrovica, 2003; Peltier, 1994).

Many shelf sea species persisted in glacial refugia mostly at more southerly latitudes in the East and West Atlantic, before expanding north during glacial retreat (Hewitt, 2003; Maggs et al., 2008; Provan & Bennett, 2008). Signatures of glacial isolation, subsequent range expansion, and secondary contact have been observed in a variety of Atlantic taxa including fish (Chevolot, Hoarau, Rijnsdorp, Stam, & Olsen, 2006; Souche et al., 2015), benthic invertebrates (Jolly, Viard, Gentil, Thiébaut, & Jollivet, 2006; Young, Torres, Mack, & Cunningham, 2002), and macroalgae (Hoarau, Coyer, Veldsink, Stam, & Olsen, 2007; Provan, Wattier, & Maggs, 2005). As a result of human-induced climate change, contemporary distributions of marine species are anticipated to shift dramatically once again, with profound ecological and economic consequences (Stanley et al., 2018; Wisz et al., 2015). The analysis of climatically induced changes in past distributions can provide valuable insights into predicted future responses (Beaugrand, Edwards, Raybaud, Goberville, & Kirby, 2015).

Phylogenetic analysis of contemporary populations has become a valuable tool in assessing changes in distributions in marine species (Maggs et al., 2008). However, for marine fishes, large-scale range-wide assessments of population structure and post-glacial colonization are often challenging due to the logistical difficulties of sampling marine species at ocean basin scales and uncertain comparability of length-based genetic markers (Lahood, Moran, Olsen, Grant, & Park, 2002). The last decade has seen a shift in the approaches employed from microsatellites and mitochondrial DNA to panels of single-nucleotide polymorphisms (SNPs). In addition to their increasingly low cost and genome-wide information content (Fuentes-Pardo & Ruzzante, 2017; Helyar et al., 2011), the simple format of essentially binary SNP data also facilitates the posthoc integration of independently collected genotype datasets (Helyar et al., 2011; Morin, Luikart, & Wayne, 2004). Genetic data synthesis has already been used in human phylogeography to supplement existing datasets with genotypes from previously unrepresented regions (Jinam et al., 2012). Similarly, in marine fish conservation, the ease of SNP synthesis has been used to encourage the use of universal genetic databases for taxa with transboundary ranges, such as tunas (Albaina et al., 2013) and Atlantic salmon (Jeffery et al., 2018). The increasing number of population genetic studies for many species and the prospective ease with which combined datasets may be generated presents an opportunity for collaborative approaches to address range-wide phylogeographic questions.

Atlantic cod (*Gadus morhua*) is a benthopelagic gadid found throughout the shelf seas of the North Atlantic, ranging from the Norwegian Arctic to as far south as the Bay of Biscay in the East Atlantic and North Carolina in the west (Robichaud & Rose, 2004; Rose, 2007). Ancestry of contemporary populations can be traced to the pre-Wisconsin interglacial period 200–250 kya (Lait, Marshall, & Carr, 2018). During the LGM, cod were separated into East and West Atlantic glacial refugia before advancing northward during glacial retreat, with recolonization of Greenland hypothesized to have resulted from westward dispersal from the East Atlantic (Bigg et al., 2008). Previous analyses of range-wide population structure in cod

undertaken using mtDNA, allozymes, microsatellites, and SNPs suggested pronounced genetic differentiation between contemporary East and West Atlantic populations (Berg et al., 2017; Bradbury et al, 2010; Bradbury et al., 2013; Carr & Marshall, 2008; Lait et al., 2018; Mork, Ryman, Ståhl, Utter, & Sundnes, 1985;). However, earlier studies have been limited by genomic and geographic coverage (Mork et al., 1985; O'Leary, Coughlan, Dillane, McCarthy, & Cross, 2007; Pogson, Mesa, & Boutilier, 1995), while others tend to lean heavily on either one or the other side of the Atlantic (Berg et al., 2017; Carr & Marshall, 2008; Lait et al., 2018).

Here, we address these limitations by using a genetic data-synthesis approach to investigate range-wide contemporary population structure in Atlantic cod. In our study, we address the difficulties of gathering representative samples from the entire species distribution through a synthesis of data from several independent regional studies. Using these data, we examine phylogeographic history in cod throughout the Atlantic at unprecedented geographic resolution. In particular, we aim to (a) re-examine the evidence of a genetic break between the East and West Atlantic, (b) verify the evidence for recolonization of Greenland from the East Atlantic, and (c) investigate secondary contact between North American and European cod. Additionally, our analysis provides a case study of the utility of a genomic data-synthesis approach for future ecological studies.

2 | METHODS

2.1 | Data synthesis

A dataset of 1,494 individuals from 54 sampling replicates was assembled from datasets provided by the FishPopTrace (FPT) Consortium (Nielsen et al., 2012), the Cod Genomics Broodstock Project (CGP) (Bradbury et al., 2013), and the authors of Therkildsen et al. (2013), and covering the East Atlantic, West Atlantic, and Greenland, respectively (Figure 1a). Individuals from all datasets had been genotyped for 972 shared SNPs originally identified in a Canadian broodstock (Hubert, Bussey, Higgins, Curtis, & Bowman, 2009). Sampling took place between 1996 and 2010 as part of scientific surveys and commercial harvests, and fish were typically in spawning condition, except for Arctic samples where fishing was restricted to summer months. All samples comprised blood or tissue from between 9 and 40 individuals per location. Further details of sampling and genotyping protocols can be found in Nielsen et al. (2012), Bradbury et al. (2013) and Therkildsen et al. (2013), and references therein.

Sequence strand designations between CGP and FPT components of the data were found to be inconsistent (a known issue with the Illumina Golden Gate platform used for Genotyping the FPT data) and, in the absence of genotyping records, ambiguous loci (A/T and C/G polymorphisms) were corrected manually by comparison of minor allele frequencies (MAF) between CGP and FPT samples from geographically proximate locations (IRE/IRS, GST/GSL, and NOR/BAS). We ground truthed this approach by comparing allele frequencies of unambiguous loci between
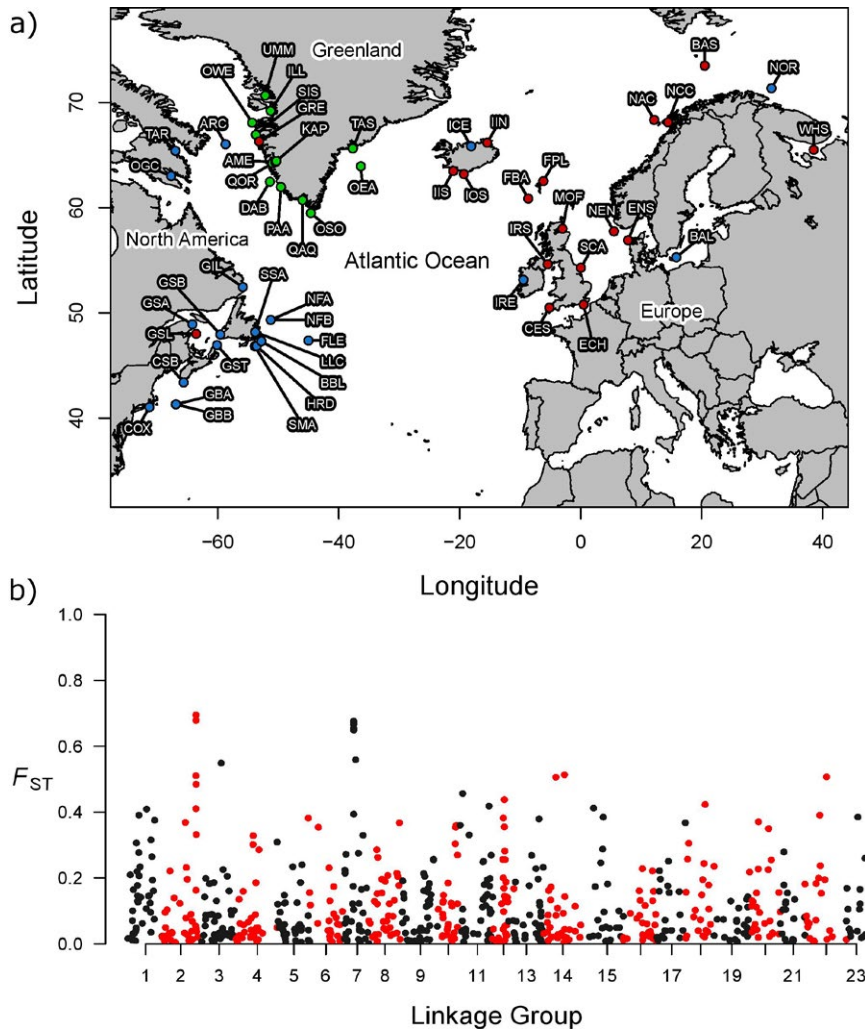
FIGURE 1 An overview of the data presented in this study. (a) Geographic distribution of samples color coded by dataset of origin, RED = FishPopTrace (FPT), BLUE = Cod Genomics Broodstock Project (CGP), GREEN = Therkildsen et al. (2013). Sample attributes are in the Supporting information Appendix S1. (b) Genomic distribution of 796 SNP panel against per-locus F_{ST}. Alternating colors indicate transitions between linkage groups. Note that not all SNPs are mapped to a linkage group (further details in the Supporting information Appendix S1)

datasets to determine the rate of matched minor alleles at cumulative 0.05 MAF intervals. The probability of incorrect assignment at each interval was calculated as the cumulative count of loci occurring between 0 and interval x with a CGP frequency >0.5, divided by the cumulative total up to that interval. Error rates were subtracted from 1 and averaged across all three locales to estimate confidence, and a traditional confidence interval of 0.95 was used to define the MAF threshold. Thus, where the frequency of a given allele was greater or less than the threshold at both locations, matched strands were assumed, while where allele frequency was less than the threshold in one dataset and greater in the other, mismatched strands were corrected. Final inclusion depended on a consistent consensus and an average confidence interval >0.95 at all three locations. Missing data were calculated per locus using GenAlEx (Peakall & Smouse, 2012) and loci exhibiting high concentrations of missing data (≤60% of loci successfully genotyped in any sample) or missing significant amounts of data (≥15%) in more than five samples were discarded. Departure from Hardy–Weinberg

equilibrium (HWE) was tested in all samples using the Monte Carlo Markov Chain procedure implemented in Pegas (Paradis, 2010) with 105 permutations. Multiple testing was corrected for using the R-package QVALUE (Storey & Tibshirani, 2003). SNPs that remained significantly divergent in more than three samples following FDR correction and these were discarded. To reduce bias attributable to inconsistencies between datasets, three pairs of samples representing geographic replicates between the FPT and CGP datasets were scanned for high per-locus F_{ST} and the degree of common occurrence in the 5% upper F_{ST} quantiles was computed using VENNDIAGRAM (Chen & Boutros, 2011). Loci occurring in the upper 5% in more than one replicate, with high occurrence of missing data, or deviating significantly from Hardy–Weinberg expectations were discarded. Due to more restricted spatial overlap between the Greenland dataset and other samples, this analysis could not be repeated with the Greenland data. However, as Greenland samples were genotyped in the same lab and according to the same protocol as FPT data and comprise a direct subset of

the FPT markers as well as sharing some individuals, strong agreement was anticipated even in the absence of this step.

2.2 | Population genetics

Expected heterozygosity and observed heterozygosity were calculated in all samples using GenAlEx. Pairwise F_{ST} was calculated between all samples via the Weir and Cockerham (1984) method implemented by hierfstat (Goudet, 2005). Isolation by distance (IBD) was evaluated using a Mantel test to test the strength of relationship between $F_{ST}/(1-F_{ST})$ and shelf sea distance. To control for the possibility that IBD detected in the full dataset is a product of geographic distance between divergent regions, IBD was evaluated individually within the Europe, Greenland/Iceland, and North America. Mean F_{ST} was calculated within each region and compared between regions using a Welch t test. Population structure was further investigated by running STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) with the admixture model. A burn-in period of 100,000 followed by an MCMC run of 500,000 iterations was used. Values of K ranged from 1 to 6 with three runs per K-value. Optimal K was identified by the change in likelihoods (ΔK, Evanno, Regnaut, & Goudet, 2005), and results were analyzed using CLUMPACK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). STRUCTURE was run in a hierarchal fashion, evaluating first all populations, and then East and West Atlantic separately. For East and West runs, we experimented with the inclusion and exclusion of samples admixed in the range-wide run, including the Davis Strait, Flemish Cap, and Arctic Lakes. Discriminate Analysis of Principle Components (DAPC) (Jombart, Devillard, & Balloux, 2010) was implemented using the R-package ADEGENET (Jombart, 2008). Clusters were identified by running successive K-means clustering for K = 1:60, and the best supported number of clusters was identified through comparison of Bayesian Information Criteria (BIC). DAPC was then used to describe the relationship between the inferred clusters. To examine the effects of ambiguous loci retention, population structure results were compared between the full panel and unambiguous loci only.

To filter the effects of selection and focus on neutral evolution, F_{ST} outlier loci were filtered using two approaches. First, a Bayesian approach was implemented using the software BAYESCAN v.2.1 (Foll & Gaggiotti, 2008) with default settings, 10,000 iterations, and a burn-in period of 500,000. Initially, the program was run on the entire dataset, resulting in a very large number of outliers (nearly half the loci). In a subsequent run, we accounted for transatlantic population structure by analyzing European and North American samples separately. Individuals from the Flemish Cap were included with Europe as population structure analysis indicated European affiliation. Samples from Lake Tariujarusiq and Lake Ogac were excluded from BAYESCAN analysis to avoid confounding our result with genetic drift. Second, a hierarchal island model was implemented using AR-LEQUIN v.3.5.2.2 (Excoffier & Lischer, 2010). The default numbers of simulations and demes per group were used, and the number of groups was set to the maximum permitted (50) to approximate the 54 samples in the data. As ARLEQUIN permits a priori input regarding

expected neutral structuring, eight groups were defined using the eight clusters approximated in the DAPC analysis.

2.3 | Migration

To examine migration across the Atlantic, samples were grouped according to geographic proximity and their connectivity estimated through population structure analysis, although care was taken not to overaggregate samples. Direction and magnitude of migration based on G_{ST} and number of migrants (Alcala, Goudet, & Vuilleumier, 2014) were estimated using the divMigrate method (Sundqvist, Keenan, Zackrisson, Prodöhl, & Kleinhans, 2016) implemented in DIVERSITY (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013) with 1,000 bootstrap replicates. Additionally, gene flow was estimated for the same groups using TREEMIX (Pickrell & Pritchard, 2012) to model evolutionary history with 1,000 bootstrap replicates and up to four prospective migration events compared based on residuals.

3 | RESULTS

3.1 | Data synthesis

When combined, 2,392 unique loci were found across the three datasets, of which 972 were common to all. For these, 106 loci exhibiting high concentrations of missing data (≤60% of loci successfully genotyped in any sample) or missing significant amounts of data (≥15%) in more than five samples were discarded. Out of 42,984 tests for HWE, 1,259 proved significant ($p \leq 0.05$), with the highest concentrations in samples from GSL, OWE, and QAQ (55, 46, and 39 loci, respectively). However, only 81 observations remained significant after correction for FDR ($q \leq 0.05$) and these were mostly distributed evenly among samples. Only three SNPs remained significantly divergent in more than three samples following FDR correction, and these were discarded. Average F_{ST} between the same geographic replicates was expectedly low (0.012, 0.012, and 0.009 for pairs representing the Gulf of Saint Lawrence, Irish Sea, and Barents Sea, respectively), indicating an overall homogeneity between datasets. Five loci that appeared in the upper 5% F_{ST} quantile in more than one of these locations were discarded. Comparison of allele frequencies in unambiguous loci at the same locations revealed that an MAF threshold of <0.45 retained an average confidence of 97.09% and >95% confidence at all three locations individually. Strand correction of ambiguous (A/T and C/G markers) based on this threshold allowed the retention of 163 loci.

3.2 | Population genetics

Population genetic analysis commenced with 796 genome-wide markers (Figure 1b). Heterozygosity within samples ranged from 0.355 on Georges Bank to 0.207 in the semi-isolated Baltic Sea and was also especially low in Lake Ogac and Lake Tariujarusiq (0.253 and 0.232, respectively). A decline in heterozygosity was observed from North America to Europe, with a particularly marked reduction
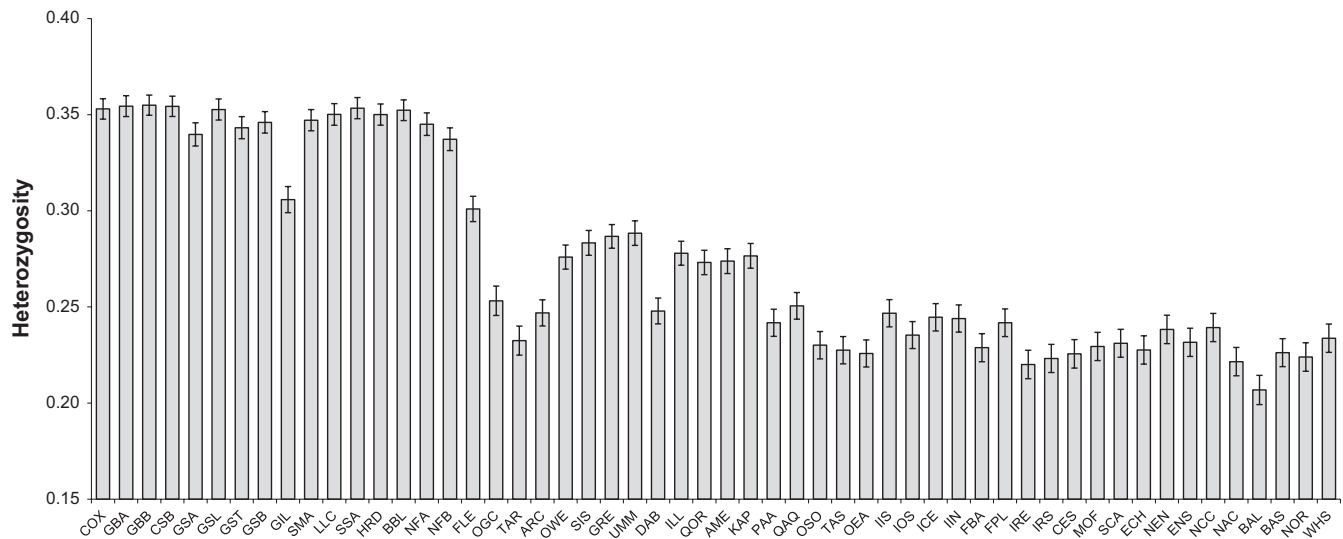
FIGURE 2 Observed heterozygosity per sample ±1 _SD_. Samples arranged approximately west to east

in the transition between Atlantic Canada and Europe (Figure 2). Heterozygosity for unambiguous loci only was tightly correlated with observations for the full panel (r^2 = 0.9986, p < 0.001) and was consistently higher for each sample, although this difference was not significant (Welch t test; t = −1.094, df = 105.78, p = 0.277). Pairwise comparisons of F_{ST} with and without ambiguous loci were also tightly correlated (r^2 = 0.9982, p < 0.001). Differences were always smaller than 0.006 and were less than 0.001 in 66% of cases. Pairwise F_{ST} ranged from 0.006 to 0.148 and was generally higher (>0.06) between Europe and North America. High F_{ST} was observed between isolated populations such as Gilbert Bay, Lake Ogac, and Lake Tariujarusiq and all other samples. IBD was detected across the range (r = 0.682, p < 0.001) and remained robust to regional evaluation (western Atlantic r = 0.527, p < 0.001; Greenland/Iceland with the exclusion of the Arctic lakes r = 0.494, p < 0.001; East Atlantic r = 0.693, p < 0.001) (Figure 3). Observations between the West Atlantic and other regions, however, exhibited notably higher F_{ST} for any given distance than those within the West Atlantic or Greenlandic/Icelandic/East Atlantic systems. Welch t test comparison of these two groups revealed significantly divergent F_{ST} means (t = −62.996, df = 953.2, p < 0.001).

For DAPC, the first 62 PCs were retained, indicated to be the optimal number by the `optim.a.score` function and representing 49.4% of total variation in the dataset. Eight clusters displayed the most favorable BIC score. The first and most informative discriminate function, which describes 65% of the variance (Figure 4a), distinctly separates clusters in the North America from those in the Europe. The North American clusters approximate the Gulf of Maine and Atlantic Canada, respectively, while three European clusters approximately represent West Greenland, northern Europe (including Iceland), and southern Europe, while a fourth cluster, termed Offshore Europe, is found most commonly among samples among offshore European and Greenlandic samples (Figure 5a). Lake Ogac and Lake Tariujarusiq form their own clusters located intermediate along the first discriminate function to the North American and
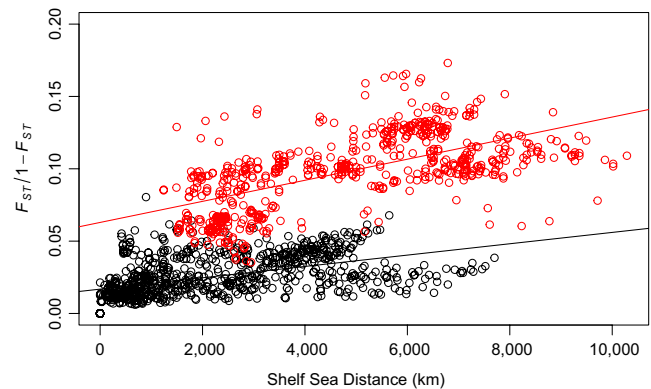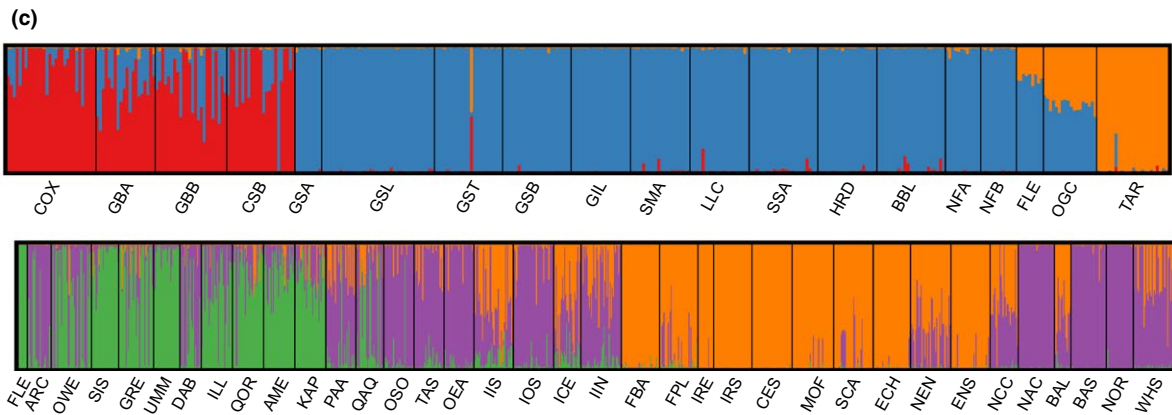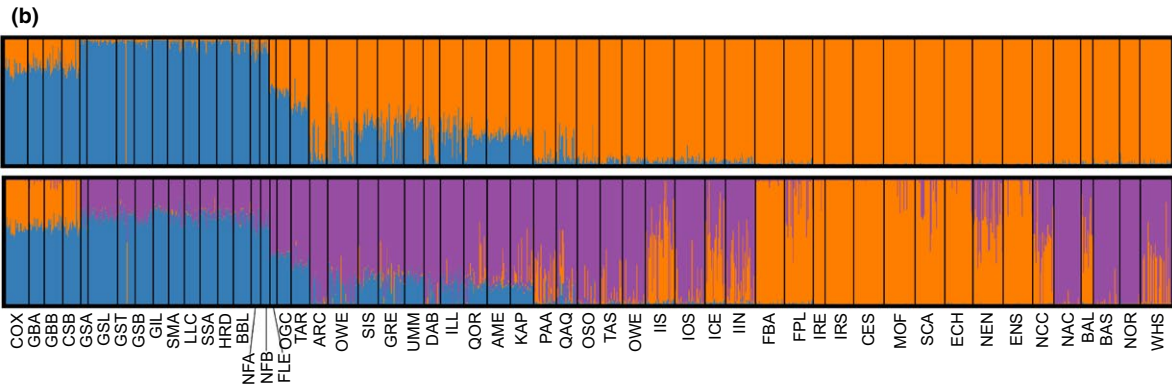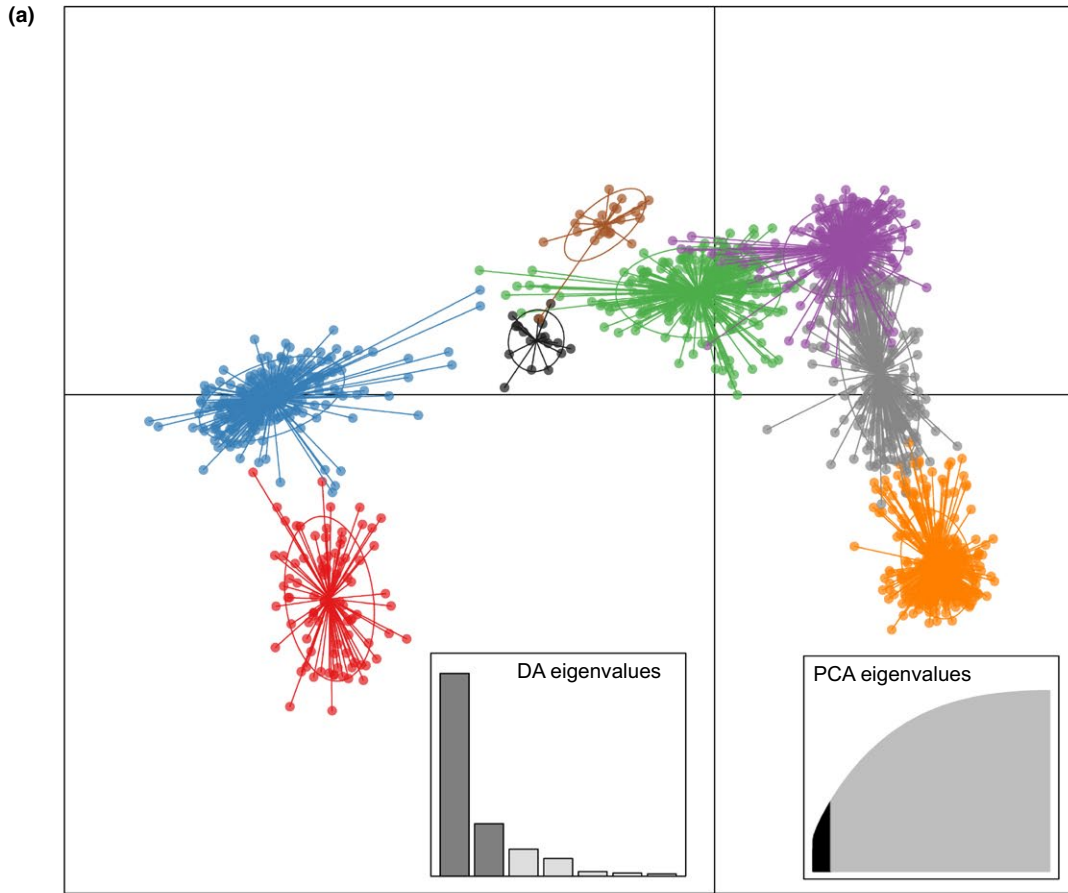

FIGURE 3 Pairwise FST/(1-FST) plotted against shelf sea distance. Comparisons between the West and East Atlantic are in red. Comparisons within the West or East Atlantic are in black

European clusters, while individuals from the Flemish Cap cluster almost exclusively with West Greenland.

In our structure analyses, the best supported K was K = 2, which indicated division between the North America and Europe and a clinal introgression of North American genotypes in Greenland (Figure 4b). For K = 3, southern Europe is additionally separated from northern Europe and Greenland in a pattern qualitatively similar to the DAPC results, although support for K = 3 is significantly lower. Hierarchical structure analysis in the West Atlantic best supported two clusters that were reflective of a North–South divide between the Gulf of Maine and Atlantic Canada (Figure 4c and 5b). A more complex northwest–southeast structure is evident in the East Atlantic, where West Greenland samples separate from Offshore Europe to form a second and third cluster in addition to southern Europe (Figure 4c). There is strong admixture within coastal samples in Norway and Iceland but interestingly offshore samples from these regions share ancestry with West Greenland, as do the Flemish Cap and the Davis Strait.

(a)

DA eigenvalues

PCA eigenvalues

(b)

COX GBA GBB CSB GSA GSL GST GSB GIL SMA LLC SSA HRD BBL NFA NFB FLE OGC TAR ARC OWE SIS GRE UMM DAB ILL QOR AME KAP PAA QAQ OSO TAS OWE IIS IOS ICE IIN FBA FPL IRE IRS CES MOF SCA ECH NEN ENS NCC NAC BAL BAS NOR WHS

(c)

COX GBA GBB CSB GSA GSL GST GSB GIL SMA LLC SSA HRD BBL NFA NFB FLE OGC TAR

FLE ARC OWE SIS GRE UMM DAB ILL QOR AME KAP PAA QAQ OSO TAS OEA IIS IOS ICE IIN FBA FPL IRE IRS CES MOF SCA ECH NEN ENS NCC NAC BAL BAS NOR WHS

FIGURE 4 Analyses of population structure. a) Scatter plot of the first and second discriminant functions of DAPC analysis of 8 clusters delineated by the find.clusters function in ADEGENT. Clusters are approximately geographically representative and are color coded as RED = Gulf of Maine, BLUE = Atlantic Canada, GREEN = West Greenland, PURPLE = Offshore Europe, GRAY = northern Europe, ORANGE = southern Europe, BLACK = Ogac, and BROWN = Tariujarusiq. (b) Admixture plots from STRUCTURE analysis showing all samples at K = 2 (top) and K = 3 (bottom). (c) Hierarchical STRUCTURE analysis in North America (top) and Europe (bottom). Note that the Flemish Cap (FLE) is included in both hierarchal runs for comparison

F_{ST} for 96 loci was above the 95% confidence interval, and 44 above the 99% confidence interval following analysis in ARLEQUIN. Initially, BAYESCAN detected 424 outlier loci, declining to 181 after controlling for transatlantic structure. Outliers detected by BAYESCAN included the majority of ARLEQUIN outliers (75% at a = 0.05, 93% at a = 0.01). Outliers chosen for further analyses comprised a set of 68 loci common to either of the BAYESCAN runs and ARLEQUIN analyses at a = 0.05. Population structure with putatively neutral

loci displayed a corresponding east–west structuring pattern as with the full panel. The main effect of outlier filtering was a breakdown of North–South population structure on both sides of the Atlantic, which resulted in a reduction from six DAPC clusters to four (excluding the Arctic lakes). These four clusters included a single West Atlantic cluster, a European cluster, and equally strong separation between inshore Greenland and offshore Arctic populations.
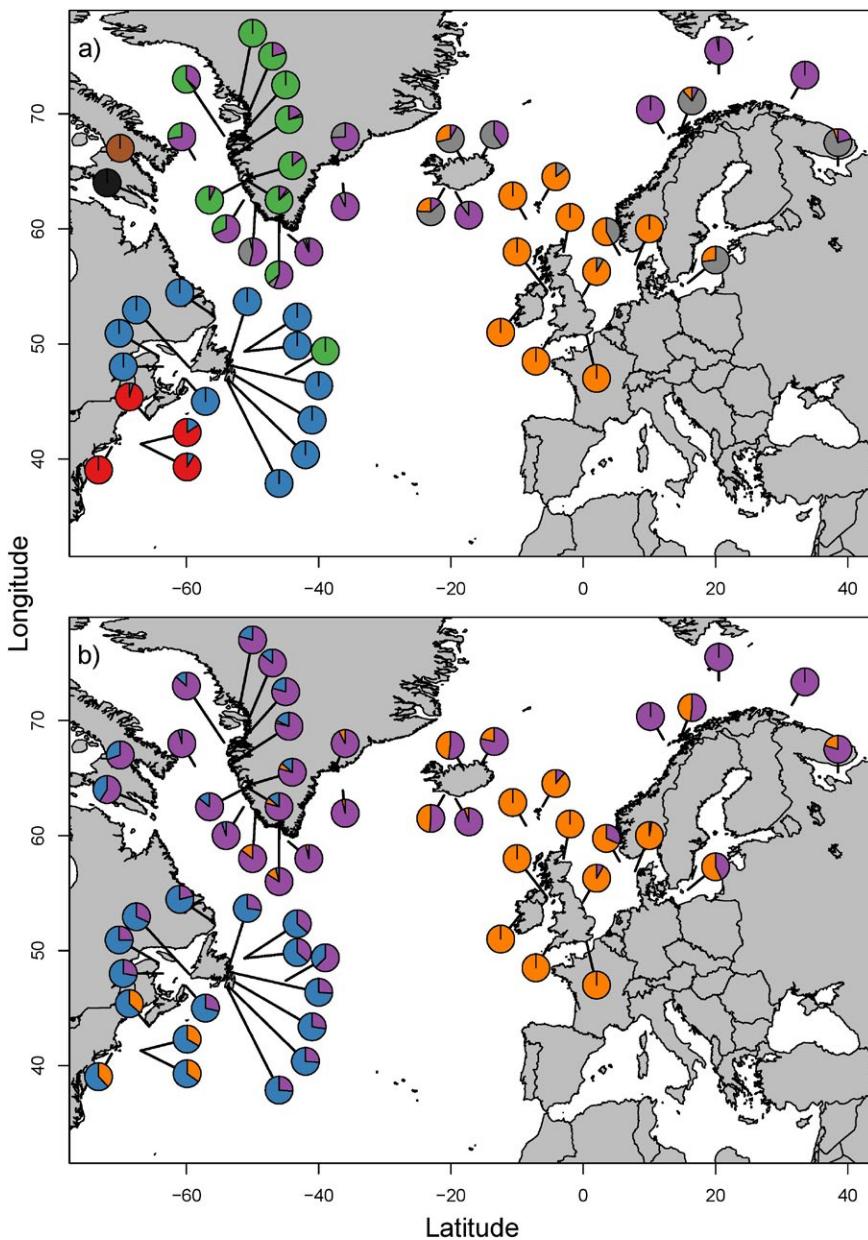
FIGURE 5 Geographic distribution of clusters from a) DAPC and b) STRUCTURE (K = 3) analyses. DAPC colors are RED = Gulf of Maine, BLUE = Atlantic Canada, GREEN = West Greenland, PURPLE = Offshore Europe, GRAY = northern Europe, ORANGE = southern Europe, BLACK = Ogac, and BROWN = Tariujarusiq
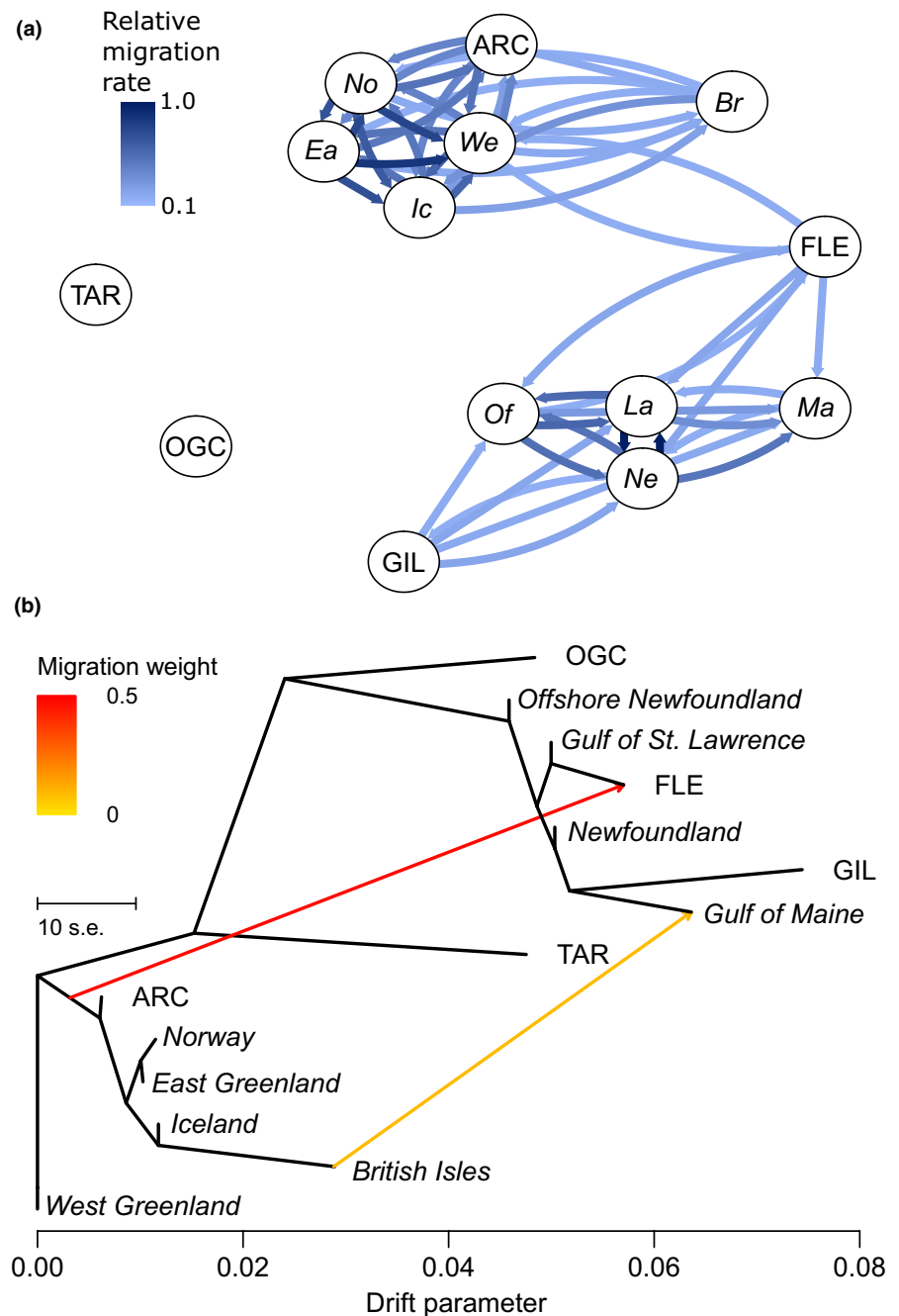
FIGURE 6 Migration estimated by (a) divMigrate based on GST and (b) TᴙᴇᴇMıx based on two migrations. Samples are grouped geographically as follows: Gulf of Maine (Ma) = COX, GBA, GBB, CSB; Gulf of St. Lawrence (La) = GSA, GSL, GST, GSB; Newfoundland (Ne) = SMA, LLC, SSA, HRD, BBL; Offshore Newfoundland (Of) = NFA, NFB; West Greenland (We) = OWE, SIS, GRE, UMM, DAB, ILL, QOR, AME, KAP, PAA, QAQ, OSO; East Greenland (Ea) = TAS, OEA; Iceland (Ic) = IIS, IOS, ICE, IIN; British Isles (Br) = FBA, FPL, IRE, IRS, CES, MOF, SCA, ECH, NEN, ENS, NCC, BAL; Norway (No) = NAC, BAS, NOR, WHS

3.3 | Migration

Analysis of asymmetric migration further supported the presence of North American and European clusters; divMigrate showed high levels of migration between all groups within North America and Europe while TᴙᴇᴇMıx splits Europe from North America at the terminal node of the dendrogram (Figure 6). TᴙᴇᴇMıx results suggest that the Flemish Cap comprises North American ancestry but are subject to a strong migration event from the Europe (Figure 6b). Further, a relatively strong (>0.1) migration is detected directly between West Greenland and the Flemish Cap by divMigrate, along with continued migration from the Flemish Cap into various West Atlantic populations (Figure 6a). In the best supported TᴙᴇᴇMıx model, a second,

weaker migration is also detected from southern Europe to the Gulf of Maine (Figure 6b).

4 | DISCUSSION

The Pleistocene spanned a period of repeated glacial cycles that resulted in the retreat of many North Atlantic marine species into glacial refugia, leaving distinct phylogenetic signatures (Maggs et al., 2008; Provan & Bennett, 2008). During the LGM, one of the Pleistocene's most severe climatic events, Atlantic cod populations are thought to have retreated into refugia in the East and West Atlantic, and a subsequent westward recolonization of Greenland

occurred from the East (Bigg et al., 2008). Previous studies using allozyme, microsatellites, and mitochondrial markers have detected pronounced genetic differentiation between East and West Atlantic cod (Carr & Marshall, 2008; Mork et al., 1985; O'Leary et al., 2007; Pogson et al., 1995) that is consistent with the separation of cod populations into glacial refugia during the LGM. In our study, glacial vicariance remains strongly supported by the differentiation between North America and Europe that dominates all population structure analyses. This separation is robust to the removal of outliers, suggesting that while latitudinal structure may be influenced by parallel adaptation, transatlantic structure is largely governed by neutral processes (Bradbury et al, 2010). Where previous studies supporting transatlantic vicariance may have been confounded by the possibility of IBD due to large geographic distances between samples (Pampoulie, Stefánsson, Jörundsdóttir, Danilowicz, & Danielsdottir, 2008; Pogson et al., 1995), our higher resolution of sampling sites suggests greater differentiation between North American and European populations than within those regions, even when shelf sea distance is controlled for. Similar patterns are observed by Lait et al. (2018), who, using analysis of the mitogenome, observe significant differentiation between many European and North American samples and an absence of population structure or IBD within regions. The role of glacial refugia in contemporary population structure is supported by East–West structure in other North Atlantic shelf species. For example, capelin (*Mallotus villosus*), a pelagic species with high larval dispersal, exhibit deep divergence between the West Atlantic and the East and Central Atlantic (Dodson, Tremblay, Colombani, Carscadden, & Lecomte, 2007). Similarly, East and West Atlantic population structure is observable in Atlantic wolffish (*Anarhichas lupus*), although phylogeographic history may be only one of numerous influences (McCusker & Bentzen, 2010).

At the end of the LGM, increased availability of cod habitat is thought to have opened the Greenlandic coast to expansion from Europe (Bigg et al., 2008). Correspondingly, in all our analyses Greenlandic samples show a high affinity with Europe while exhibiting a much lower affinity to North America (Figure 5). In DAPC, in particular, there is a longitudinal gradient between Greenland and Europe, while separation from North America is essentially complete (Figures 4a and 5a). The two DAPC clusters identified in Greenland were previously identified by Therkildsen et al. (2013) using the same data, but our inclusion of a broader dataset places them in context with the rest of the North Atlantic. For example, the East Greenland cluster identified by Therkildsen et al. (2013) (and termed the Offshore cluster herein) extends almost across the breadth of the European range (Figure 5a) and appears to link offshore populations of Northeast Arctic cod, Icelandic Frontal cod, and Davis Strait cod. This homogeneity may imply shared ancestry or ongoing connectivity among offshore populations (see also Bonanomi et al., 2016), which may have significantly influenced westward expansion following the LGM. Colonization of the subarctic may have extended beyond Greenland, leading to secondary contact between European and North American populations.

One of the most remarkable examples of putative secondary contact is the Flemish Cap, a shallow (125–700 m) remote bank off the easternmost tip of the Newfoundland Shelf (Stein, 2007). The link between the Flemish Cap and European populations is evident from Bentzen, Taggart, Ruzzante, and Cook (1996) and Bradbury et al. (2013), who find affiliations with several East Atlantic populations, although purely North American haplotypes are found using the mitochondrial genome (Lait et al., 2018). Using DAPC and STRUCTURE on a wider range of European data, we find that the Flemish Cap population is dominated by a Greenlandic cluster found at no other North American locations, implying that the Flemish Cap population represents an offshoot of Greenlandic ancestry in North America. Such observations may result from cod eggs and larvae from Greenland that occur in the Davis Strait (Wieland & Hovgard, 2002) being transported south by the Labrador Current and becoming trapped by an oceanic gyre surrounding the Cap (Gil, Sánchez, Cerviño, & Garabana, 2004; Stein, 2007). Subsequently, adults may then be prevented from leaving by the gyre and the 1,000 m deep Flemish Pass that separates the Flemish Cap from the Newfoundland Shelf (Stein, 2007). TREEMIX results find strong signal of migration between the Davis Strait and the Flemish Cap, but also indicates that the Flemish Cap population is largely of North American ancestry (Figure 6b). This is supported by the presence of North American genotypes at the Flemish Cap detected by STRUCTURE. Other results suggest that migration between Greenland and the Flemish Cap is via some route other than the Davis Strait. For example, the DAPC Offshore cluster is dominant at the Davis Strait but absent from the Flemish Cap (Figure 5a), while divMigrate suggests migration is restricted to that occurring directly between West Greenland and the Flemish Cap. Thus, while our results indicate strong support for Greenlandic origin of the Flemish Cap population, they remain inconclusive as to the exact route or mechanism of migration.

Strong European influence has also previously been found among cod in meromictic lakes on Baffin Island, including Lake Ogac and Lake Tariujarusiq (Bradbury et al., 2013; Carr & Marshall, 2008; Lait et al., 2018). The resident populations occurring in Ogac and Tariujarusiq were established 5 to 8 kya following a period of comparatively warm climate, rising sea levels, and isostatic rebound that enabled cod to inhabit the then fully marine sites (Hardie, Gillett, & Hutchings, 2006). Westward colonization following glacial retreat supports the hypothesis of at least some colonization from Europe (Lait et al., 2018; Therkildsen et al., 2013). However, our results concerning these populations were inconclusive. In DAPC analysis, each lake affiliated to its own cluster, indicating high genetic distinctiveness that is likely the result of small population size and strong genetic drift (Hardie et al., 2006, 2008). Although slightly closer to European clusters, these clusters were ultimately intermediate to Europe and North America (Figure 4a). Conversely, at *K* = 2 both lakes exhibit North American affinity in STRUCTURE analysis, although *K* = 3 introduces some admixture from Greenland. Lait et al. (2018) find evidence of paraphyly in the geographically and ecologically similar Lake Qasigialiminiq, and hypothesize multiple colonization events from both the East and West Atlantic, but note that

further study is required for a better understanding of the history of meromictic lake populations.

Our study focuses largely on genome-wide patterns; however, more recent research emphasizes the roles of genomic architecture, in particular chromosomal inversions on linkage groups (LGs) 01, 02, 07, and 12, in shaping both historical and contemporary population genomic processes. For example, ecotype associated inversions likely play a significant role in different migratory behaviors between Northeast Arctic cod and Norwegian coastal cod (Berg et al., 2016), and are associated with migration behavior in numerous populations (Barney, Munkholm, Walt, & Palumbi, 2017; Berg et al., 2017; Sinclair-Waters et al., 2017). Although the inversions on both sides of the Atlantic are thought to have common ancestral origin (Berg et al., 2017), the shared ancestry between offshore (migratory) ecotypes implied by our results is contradicted by the findings of Kirubakaran et al. (2016), who find the nonmigratory Norwegian coastal cod to represent the ancestral population (in Norway and for the LG01 inversion at least). This implies that our finding of a common Offshore cluster may be the product of parallel adaptation driven by one or more inversions. Interestingly, despite using a North American SNP panel, we are less able to reproduce migration associated divergence in North America than in Europe (e.g., Barney et al., 2017; Sinclair-Waters et al., 2017). In some cases, such as the northern cod, this may reflect the finding that population structure at inversion sites seems to operate at the individual level (i.e., populations homozygous for one of the other form of an inversion appear to exist in sympatry (Kess et al., 2018). Other patterns observed in our data, for example, in Baltic Sea and north Sea divergence, may reflect contemporary local adaptation to a salinity and temperature gradient (Berg et al., 2015), and we found that removing putatively adaptive (outlier) loci from our analysis resulted in the North Sea and Baltic Sea clustering together. Adaptation to temperature and salinity has again been linked to chromosomal inversions on LGs 02, 07, and 12 and can lead to adaptive divergence at both large and small geographic scales (Barth et al., 2017; Bradbury et al, 2010; Sinclair-Waters et al., 2017). Of the outliers we encounter, there is an even distribution among LGs, while in our full dataset each of LGs 01, 02, 07, and 12 is represented by only a small number of loci (11–33). The resultant stochasticity makes it difficult to draw meaningful conclusions regarding the phylogeography of the inversions from our dataset alone. However, chromosomal inversions clearly play a significant role in the phylogeographic history of Atlantic cod and future research will surely continue to explore this role.

The study of phylogeographic histories is relevant to a diversity of conservation and management applications in marine species (e.g., Sinclair-Waters et al., 2017; von der Heyden, 2017), but studying phylogeographic patterns in species with oceanic ranges is a significant challenge. Here, we address this challenge in Atlantic cod by synthesizing three pre-existing SNP genotype datasets to improve representation across the species' range. Our dataset includes 1,494 individuals taken from 54 Atlantic-wide locations, representing one of the most geographically and genetically extensive datasets for a marine fish to date. (Note, however, that Arnason (2004) achieves greater geography coverage than us, albeit using a much-reduced genome representation, and that despite lower geographic coverage, modern population genetics routinely employs a larger genomic coverage.) Our approach circumvents the expense of novel data collection and illustrates the potential for collaborative synthesis to address phylogeographic questions (Helyar et al., 2011; Morin et al., 2004). The benefits of a synthesis approach are especially apparent in marine fishes, in which the acquisition of data is both challenging and expensive (Albaina et al., 2013; Bernatchez et al., 2017).

Our final dataset of 796 SNP markers is moderate by contemporary standards and, aside from the limits of the source data, was primarily limited by the degree of marker overlap. The use of the CGP dataset resulted in all common markers in our analysis originating from a North American broodstock, increasing the risk of ascertainment bias affecting our results. We observe a marked decline in heterozygosity from North America to Europe that we attribute to ascertainment bias. In general, however, the impacts of ascertainment bias in high gene flow marine species are thought to be modest (Bradbury et al., 2011). Further, we would therefore expect a bias toward greater diversity represented in North America to favor detection of North American variants in Europe, but instead, we primarily see the reverse (European genotypes in North America), giving us confidence that the bias has a negligible impact on our results. Another potential source of bias was filtering for strand mismatch. In the absence of records that would have allowed us to track the sequenced strand, we mitigated data loss by correcting ambiguous mismatched strands (i.e., A/T and C/G flips, for which alternate strands were confounded with alternate alleles) by comparing minor allele frequencies between samples of the same putative populations between datasets. We found that only variants with a MAF of ≤0.45 could be reliably identified this way, resulting in a slight systematic bias toward reduced heterozygosity when ambiguous markers were included. However, all duplicated analyses yielded nearly identical between the inclusion and exclusion of MAF resolved genotypes, again suggesting that the impact of this bias was negligible.

5 | CONCLUSIONS

The LGM was a crucial event in the phylogeographic history of North Atlantic marine species, as indicated by the observed strong phylogenetic signals. We investigated these signals in Atlantic cod using a data-synthesis approach to generate a single range-wide dataset. In doing so, we reveal new details of the genetic impacts of glacial vicariance, such as secondary contact via the Davis Strait and the Flemish Cap. Further, we demonstrate the utility of a data-synthesis approach to genetics and ecology. Current "whole genome" studies now often employ millions of markers, while the deployment of hundreds to thousands of SNPs has become more routine in ecology and conservation research. Synthesis of common datasets, as

demonstrated here, may be increasingly applicable because of increased accessibility of genomic approaches and archived datasets across target species.

AUTHOR CONTRIBUTIONS

The study was conceived by GRC in discussion with SJH, JHH, and IRB, with contributions to the design by all authors. GRC and MdB supervised RF's research program. Data were provided by IRB, JHH, NOT, SJH, and PB. Data synthesis and analysis were carried out by RF with assistance from IRB, SJH, and JHH and suggestions from all other authors. The manuscript was drafted by RF and IRB and was read and revised by all authors.

DATA ACCESSIBILITY

The SNPs analyzed in this study are available in GenBank. Accession numbers are listed in the Supporting information Appendix S2. Genotypes are available on Dryad (https://doi.org/10.5061/dryad.66jp1m9).

ORCID

Robert Fairweather ⬥ https://orcid.org/0000-0001-7726-773X

Ian R. Bradbury ⬥ https://orcid.org/0000-0002-8152-4943

Sarah J. Helyar ⬥ https://orcid.org/0000-0002-1600-997X

Mark de Bruyn ⬥ https://orcid.org/0000-0003-1528-9604

Nina O. Therkildsen ⬥ https://orcid.org/0000-0002-6591-591X

Paul Bentzen ⬥ https://orcid.org/0000-0003-4288-7730

Jakob Hemmer-Hansen ⬥ https://orcid.org/0000-0002-1033-2565

Gary R. Carvalho ⬥ https://orcid.org/0000-0002-9509-7284

REFERENCES

Albaina, A., Iriondo, M., Velado, I., Laconcha, U., Zarraonaindia, I., Arrizabalaga, H., ... Estonba, A. (2013). Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure. _Animal Genetics_, _44_(6), 678–692. https://doi.org/10.1111/age.12051

Alcala, N., Goudet, J., & Vuilleumier, S. (2014). On the transition of genetic differentiation from isolation to panmixia: What we can learn from GST and D. _Theoretical Population Biology_, _93_, 75–84.

Arnason, E. (2004). Mitochondrial cytochrome b DNA variation in the high-fecundity Atlantic cod: Trans-Atlantic clines and shallow gene genealogy. _Genetics_, _166_(4), 1871–1885. https://doi.org/10.1534/genetics.166.4.1871

Barney, B. T., Munkholm, C., Walt, D. R., & Palumbi, S. R. (2017). Highly localized divergence within supergenes in Atlantic cod (_Gadus morhua_) within the Gulf of Maine. _BMC Genomics_, _18_(1), 271–785. https://doi.org/10.1186/s12864-017-3660-3

Barth, J. M., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., ... Moksnes, P. O. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. _Molecular Ecology_, _26_(17), 4452–4466. https://doi.org/10.1111/mec.14207

Beaugrand, G., Edwards, M., Raybaud, V., Goberville, E., & Kirby, R. R. (2015). Future vulnerability of marine biodiversity compared with contemporary and past changes. _Nature Climate Change_, _5_(7), 695–701. https://doi.org/10.1038/nclimate2650

Bentzen, P., Taggart, C. T., Ruzzante, D. E., & Cook, D. (1996). Microsatellite polymorphism and the population structure of Atlantic cod (_Gadus morhua_) in the northwest Atlantic. _Canadian Journal of Fisheries and Aquatic Sciences_, _53_(12), 2706–2721.

Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., ... Andre, C. (2015). Adaptation to low salinity promotes genomic divergence in Atlantic cod (_Gadus morhua_ L.). _GenomeBiology and Evolution_, _7_(6), 1644–1663. https://doi.org/10.1093/gbe/evv093

Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., ... Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. _Heredity_, _119_(6), 418–428. https://doi.org/10.1038/hdy.2017.54

Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M., Knutsen, H., ... Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. _Scientific Reports_, _6_, 232–246. https://doi.org/10.1038/srep23246

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., ... Ovenden, J. R. (2017). Harnessing the power of genomics to secure the future of seafood. _Trends in Ecology and Evolution_, _32_(9), 665–680. https://doi.org/10.1016/j.tree.2017.06.010

Bigg, G. R., Cunningham, C. W., Ottersen, G., Pogson, G. H., Wadley, M. R., & Williamson, P. (2008). Ice-age survival of Atlantic cod: Agreement between palaeoecology models and genetics. _Proceedings of the Royal Society of London B: Biological Sciences_, _275_, 163–173. https://doi.org/10.1098/rspb.2007.1153

Bonanomi, S., Overgaard Therkildsen, N., Retzel, A., Berg Hedeholm, R., Pedersen, M. W., Meldrup, D., ... Nielsen, E. E. (2016). Historical DNA documents long-distance natal homing in marine fish. _Molecular Ecology_, _25_(12), 2727–2734.

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., ... Hutchings, J. A. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. _Proceedings of the Royal Society of London B: Biological Sciences_, _277_, 3725–3734.

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., ... Hutchings, J. A. (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

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Paterson, I. G., Snelgrove, P. V., ... Bentzen, P. (2011). Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, _Gadus morhua_. _Molecular Ecology Resources_, _11_, 218–225.

Carr, S. M., & Marshall, H. D. (2008). Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod

(*Gadus morhua*): origins of the "codmother", transatlantic vicariance and midglacial population expansion. *Genetics*, *180*(1), 381–389.

Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, *12*, 35.

Chevolot, M., Hoarau, G., Rijnsdorp, A. D., Stam, W. T., & Olsen, J. L. (2006). Phylogeography and population structure of thornback rays (*Raja clavata* L., Rajidae). *Molecular Ecology*, *15*(12), 3693–3705. https://doi.org/10.1111/j.1365-294X.2006.03043.x

Dodson, J. J., Tremblay, S., Colombani, F., Carscadden, J. E., & Lecomte, F. (2007). Trans-Arctic dispersals and the evolution of a circumpolar marine fish species complex, the capelin (*Mallotus villosus*). *Molecular Ecology*, *16*(23), 5030–5043. https://doi.org/10.1111/j.1365-294X.2007.03559.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. (2010). ARLEQUIN suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Foll, M., & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, *26*(20), 5369–5406. https://doi.org/10.1111/mec.14264

Gil, J., Sánchez, R., Cerviño, S., & Garabana, D. (2004). Geostrophic circulation and heat flux across the Flemish Cap, 1988–2000. *Journal of Northwest Atlantic Fishery Science*, *34*, 63–83. https://doi.org/10.2960/J.v34.m510

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Hardie, D. C., Gillett, R. M., & Hutchings, J. A. (2006). The effects of isolation and colonization history on the genetic structure of marine-relict populations of Atlantic cod (*Gadus morhua*) in the Canadian Arctic. *Canadian Journal of Fisheries and Aquatic Sciences*, *63*(8), 1830–1839.

Hardie, D. C., Renaud, C. B., Ponomarenko, V. P., Mukhina, N. V., Yaragina, N. A., Skjæraasen, J. E., & Hutchings, J. A. (2008). The isolation of Atlantic cod, *Gadus morhua* (Gadiformes), populations in northern meromictic lakes—a recurrent Arctic phenomenon. *Journal of Ichthyology*, *48*(3), 230–240. https://doi.org/10.1134/S0032945208030053

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., … Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, *11*, 123–136. https://doi.org/10.1111/j.1755-0998.2010.02943.x

Hewitt, G. (2003). Ice ages: Species distributions, and evolution. In L. J. Rothschild, & A. M. Lister (Eds.), *Evolution on Planet Earth: Impact of the physical environment* (pp. 339–361). London, UK: Academic Press.

Hoarau, G., Coyer, J. A., Veldsink, J. H., Stam, W. T., & Olsen, J. L. (2007). Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. *Molecular Ecology*, *16*(17), 3606–3616.

Hubert, S., Bussey, J. T., Higgins, B., Curtis, B. A., & Bowman, S. (2009). Development of single nucleotide polymorphism markers for Atlantic cod (Gadus morhua) using expressed sequences. *Aquaculture*, *296*(1–2), 7–14. https://doi.org/10.1016/j.aquaculture.2009.07.027

Jeffery, N. W., Wringe, B. F., McBride, M. C., Hamilton, L. C., Stanley, R. R., Bernatchez, L., … Bentzen, P., (2018). Range-wide regional assignment of Atlantic salmon (*Salmo salar*) using genome wide single-nucleotide polymorphisms. *Fisheries Research*, *206*, 163.

Jinam, T., Nishida, N., Hirai, M., Kawamura, S., Oota, H., Umetsu, K., … Tanabe, H. (2012). The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *Journal of Human Genetics*, *57*(12), 787–796.

Jolly, M. T., Viard, F., Gentil, F., Thiébaut, É., & Jollivet, D. (2006). Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology*, *15*(7), 1841–1855. https://doi.org/10.1111/j.1365-294X.2006.02910.x

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94–109. https://doi.org/10.1186/1471-2156-11-94

Kess, T., Bentzen, P., Lehnert, S. J., Sylvester, E. V. A., Lien, S., Kent, M. P., … Bradbury, I. R. (2018). A migration-associated supergene reveals loss of biocomplexity in Atlantic Cod *bioRxiv*, 361691. https://doi.org/10.1101/361691

Kettle, A. J., Morales-Muñiz, A., Roselló-Izquierdo, E., Heinrich, D., & Vøllestad, L. A. (2011). Refugia of marine fish in the northeast Atlantic during the last glacial maximum: Concordant assessment from archaeozoology and palaeotemperature reconstructions. *Climate of the past*, *7*, 181–201.

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., … Sonesson, A. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, *25*(10), 2130–2143.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191.

LaHood, E. S., Moran, P., Olsen, J., Grant, W. S., & Park, L. K. (2002). Microsatellite allele ladders in two species of Pacific salmon: Preparation and field-test results. *Molecular Ecology Notes*, *2*(2), 187–190. https://doi.org/10.1046/j.1471-8286.2002.00174.x

Lait, L. A., Marshall, H. D., & Carr, S. M. (2018). Phylogeographic mitogenomics of Atlantic cod Gadus morhua: Variation in and among trans-Atlantic, trans-Laurentian, Northern cod, and landlocked fjord populations. *Ecology and Evolution*, *8*, 6420–6437.

Lambeck, K., Esat, T. M., & Potter, E. K. (2002). Links between climate and sea levels for the past three million years. *Nature*, *419*(6903), 199–206.

Maggs, C. A., Castilho, R., Foltz, D., Henzler, C., Jolly, M. T., Kelly, J., … Viard, F. (2008). Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, *89*(sp11), S108–S122. https://doi.org/10.1890/08-0257.1

McCusker, M. R., & Bentzen, P. (2010). Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolffish (*Anarhichas lupus*), across the North Atlantic Ocean. *Molecular Ecology*, *19*(19), 4228–4241. https://doi.org/10.1111/j.1365-294X.2010.04806.x

Mitrovica, J. X. (2003). Recent controversies in predicting post-glacial sea-level change. *Quaternary Science Reviews*, *22*(2–4), 127–133. https://doi.org/10.1016/S0277-3791(02)00225-1

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, *19*(4), 208–216. https://doi.org/10.1016/j.tree.2004.01.009

Mork, J., Ryman, N., Ståhl, G., Utter, F., & Sundnes, G. (1985). Genetic variation in Atlantic cod (*Gadus morhua*) throughout its range. *Canadian Journal of Fisheries and Aquatic Sciences*, *42*(10), 1580–1587.

Nielsen, E. E., Cariani, A., Mac Aoidh, E., Maes, G. E., Milano, I., Ogden, R., … Bekkevold, D. (2012). Gene-associated markers

provide tools for tackling illegal fishing and false eco-certification. *NatureCommunications*, 3, 851–858.

O'Leary, D. B., Coughlan, J., Dillane, E., McCarthy, T. V., & Cross, T. F. (2007). Microsatellite variation in cod *Gadus morhua* throughout its geographic range. *Journal of Fish Biology*, 70, 310–335. https://doi.org/10.1111/j.1095-8649.2007.01451.x

Pampoulie, C., Stefánsson, M. Ö., Jörundsdóttir, T. D., Danilowicz, B. S., & Danielsdottir, A. K. (2008). Recolonization history and large-scale dispersal in the open sea: The case study of the North Atlantic cod, *Gadus morhua* L. *Biological Journal of the Linnean Society*, 94(2), 315–329. https://doi.org/10.1111/j.1095-8312.2008.00995.x

Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28, 2537–2539.

Peltier, W. R. (1994). Ice age paleotopography. *Science*, 265(5169), 195–201.

Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *Science*, 308(5730), 1912–1915.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pogson, G. H., Mesa, K. A., & Boutilier, R. G. (1995). Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: A comparison of allozyme and nuclear RFLP loci. *Genetics*, 139(1), 375–385.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.

Provan, J., & Bennett, K. D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, 23(10), 564–571. https://doi.org/10.1016/j.tree.2008.06.010

Provan, J., Wattier, R. A., & Maggs, C. A. (2005). Phylogeographic analysis of the red seaweed Palmaria palmata reveals a Pleistocene marine glacial refugium in the English Channel. *Molecular Ecology*, 14(3), 793–803. https://doi.org/10.1111/j.1365-294X.2005.02447.x

Robichaud, D., & Rose, G. A. (2004). Migratory behaviour and range in Atlantic cod: Inference from a century of tagging. *Fish and Fisheries*, 5(3), 185–214. https://doi.org/10.1111/j.1467-2679.2004.00141.x

Rose, G. A. (2007). *Cod: An ecological history of the North Atlantic fisheries*. St John's: Breakwater books.

Sinclair-Waters, M., Bradbury, I. R., Morris, C. J., Lien, S., Kent, M. P., & Bentzen, P. (2017). Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic Cod in the northwest Atlantic. *Molecular Ecology*, 27(2), 339–351. https://doi.org/10.1111/mec.14442

Souche, E. L., Hellemans, B., Babbucci, M., MacAoidh, E., Guinand, B., Bargelloni, L., ... Volckaert, F. A. (2015). Range-wide population structure of European sea bass *Dicentrarchus labrax*. *Biological Journal of the Linnaean Society*, 116(1), 86–105.

Stanley, R. R., DiBacco, C., Lowen, B., Beiko, R. G., Jeffery, N. W., Van Wyngaarden, M., ... Johnson, C. (2018). A climate-associated multispecies cryptic cline in the northwest. *Atlantic. Science Advances*, 4(3), eaaq0929. https://doi.org/10.1126/sciadv.aaq0929

Stein, M. (2007). Oceanography of the Flemish Cap and adjacent waters. *Journal of Northwest Atlantic Fishery Science*, 37(2), 135–146. https://doi.org/10.2960/J.v37.m652

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16), 9440–9445. https://doi.org/10.1073/pnas.1530509100

Sundqvist, L., Keenan, K., Zackrisson, M., Prodöhl, P., & Kleinhans, D. (2016). Directional genetic differentiation and relative migration. *Ecology and Evolution*, 6(11), 3461–3475. https://doi.org/10.1002/ece3.2096

Therkildsen, N. O., Hemmer-Hansen, J., Hedeholm, R. B., Wisz, M. S., Pampoulie, C., Meldrup, D., ... Nielsen, E. E. (2013). Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. *Evolutionary Applications*, 6(4), 690–705.

von der Heyden, S. (2017). Making evolutionary history count: Biodiversity planning for coral reef fishes and the conservation of evolutionary processes. *Coral Reefs*, 36(1), 183–194. https://doi.org/10.1007/s00338-016-1512-2

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. https://doi.org/10.2307/2408641

Wieland, K., & Hovgard, H. (2002). Distribution and drift of Atlantic cod (*Gadus morhua*) eggs and larvae in Greenland offshore waters. *Journal of Northwest Atlantic Fishery Science*, 30, 61–76.

Wisz, M. S., Broennimann, O., Grønkjær, P., Møller, P. R., Olsen, S. M., Swingedouw, D., ... Pellissier, L. (2015). Arctic warming will promote Atlantic-Pacific fish interchange. *Nature Climate Change*, 5(3), 261.

Young, A., Torres, C., Mack, J., & Cunningham, C. (2002). Morphological and genetic evidence for vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit crab *Pagurus longicarpus*. *Marine Biology*, 140(5), 1059–1066. https://doi.org/10.1007/s00227-002-0780-2

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.