



OPEN

Genomic architecture and population structure of *Boreogadus saida* from Canadian waters

Trevor T. Bringloe¹✉, Audrey Bourret¹, David Cote², Roux Marie-Julie¹, Jennifer Herbig³, Dominique Robert⁴, Maxime Geoffroy^{3,5} & Geneviève J. Parent¹✉

The polar cod, *Boreogadus saida*, is an abundant and ubiquitous forage fish and a crucial link in Arctic marine trophic dynamics. Our objective was to unravel layers of genomic structure in *B. saida* from Canadian waters, specifically screening for potential hybridization with the Arctic cod, *Arctogadus glacialis*, large chromosomal inversions, and sex-linked regions, prior to interpreting population structure. Our analysis of 53,384 SNPs in 522 individuals revealed hybridization and introgression between *A. glacialis* and *B. saida*. Subsequent population level analyses of *B. saida* using 12,305 SNPs in 511 individuals revealed three large (ca. 7.4–16.1 Mbp) chromosomal inversions, and a 2 Mbp region featuring sex-linked loci. We showcase population structuring across the Western and Eastern North American Arctic, and subarctic regions ranging from the Hudson Bay to the Canadian Atlantic maritime provinces. Genomic signal for the inferred population structure was highly aggregated into a handful of SNPs (13.8%), pointing to potentially important adaptive evolution across the Canadian range. Our study provides a high-resolution perspective on the genomic structure of *B. saida*, providing a foundation for work that could be expanded to the entire circumpolar range for the species.

Keywords Arctic cod, Polar cod, Inversion, Hybridization, Introgression

Polar cod (*Boreogadus saida*) is an abundant and ubiquitously distributed, circumpolar gadid. It is an important consumer of a variety of zooplanktonic species^{1–4}, and is itself an important food source for Arctic predators, contributing disproportionately to energy transfer to higher trophic levels in some areas^{5–7}. Aggregations of *B. saida* are an important food source for narwhal (*Monodon monoceros*⁸), beluga (*Delphinapterus leucas*⁹), ringed seal (*Pusa hispida*¹⁰), and various marine birds such as thick-billed murres (*Uria lomvia*) and northern fulmars (*Fulmarus glacialis*⁵). *B. saida* is found throughout Arctic ice shelves (particularly earlier life stages^{11,12}), where it can flourish due to the synthesis of antifreeze glycoproteins¹³. Given the central role played by *B. saida* in Arctic food webs, its polar distribution, and potential adaptation for Arctic conditions, climate change is anticipated to impact this species, precipitating profound changes to ecological services in Arctic waters¹². Potential non-mutually exclusive threats include the introduction of new predators and competitors and shifting prey distributions through the borealization of Arctic waters, diminishing sea ice extent and thickness that serves as habitat for early life stages, warming surface waters impacting egg development and hatching success, and increased human activity in a more seasonally open Arctic ocean^{12,14,15}. Robust baseline understanding of the diversity and connectivity of *B. saida* in Arctic and subarctic regions is needed to better anticipate the impacts of a changing climate and anthropogenic stressors on the species and marine communities.

Early genetic studies in *B. saida* showed weak population structure and high connectivity across global populations. Among the earliest molecular perspectives, Fevolden et al.¹⁶ found no evidence for differentiation between populations of Greenland (Disko Bay and Denmark Strait), Svalbard, and Northern Norway using the

¹Fisheries and Oceans Canada, Maurice Lamontagne Institute, Mont-Joli, QC G5H 3Z4, Canada. ²Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John's, NL A0G 2M0, Canada. ³Centre for Fisheries Ecosystems Research, Fisheries and Marine Institute of Memorial, University of Newfoundland, St. John's A1C 5R3, Canada. ⁴Institut Des Sciences de La Mer, Université du Québec à Rimouski, Rimouski, QC G5L 3A1, Canada. ⁵Department of Arctic and Marine Biology, UiT The Arctic University of Norway, 9036 Tromsø, Norway. ✉email: tbringloe@gmail.com; genevieve.parent@dfo-mpo.gc.ca

nuclear locus *Syp1* and 38 randomly amplified polymorphic nuclear loci, a result further confirmed by Pálsson et al.¹⁷ using mitochondrial markers. In more recent work, Maes et al.¹⁸ also confirmed no population structure between Svalbard and the Eurasian basin using nine microsatellites, which the authors attribute to high gene flow between study regions. Quintela et al.¹⁹ found similar results using 116 single nucleotide polymorphic (SNP) loci, but did detect isolation by distance between the East Siberian and Laptev Seas, and the Eurasian Basin locations. Low levels to complete absence of population structure were further reported in the Chukchi Sea and adjacent Beaufort Sea populations using mitochondrial *cytb* and microsatellites^{20,21}, and in Russian Arctic seas using microsatellites²². In contrast, differentiation between fjord-dwelling and shelf-dwelling *B. saida* populations have been reported from the East Atlantic using microsatellites²³, suggesting local divergence could play a primary role in driving genetic structure in *B. saida*. Lending credence to this hypothesis, Hill et al.²⁴ demonstrated that fjord and offshore dwelling specimens differentially expressed approximately 2000 genes. In contrast to these works, a global-scale study of *B. saida* populations using nine microsatellites was published by Nelson et al.²⁵, who revealed four populations corresponding to (1) the East Canadian Arctic, (2) the West Canadian Arctic (i.e. East Beaufort Sea), (3) the US Arctic Seas (Bering, Chukchi and West Beaufort Seas), and (4) the European Arctic Seas (Greenland, Iceland, and Laptev Sea). Besides demonstrating that population structure did exist at global scales, the authors also noted our understanding of *B. saida* population structure would benefit from higher density marker datasets, which can be provided by genomics.

Interpreting genomic structure in *B. saida* is likely to come with specific challenges. Among the complications with resolving population structure in *B. saida* are potential hybridizations. These events can have profound evolutionary consequences, ranging from adding new opportunities for adaptive evolution to eroding the integrity of species barriers, resulting in loss of biodiversity^{26,27}. In gadids, extensive introgression potentially facilitated early evolution in some species, as evidenced by non-random sharing of alleles between walleye pollock (*Gadus chalcogrammus*), Pacific cod (*Gadus macrocephalus*), and Greenland cod (*Gadus ogac*)²⁸ (as noted by the authors Pacific cod and Greenland cod may represent the same species²⁹). More relevant for population level analyses are known contemporary hybridization between *B. saida* and the Arctic cod *Arctogadus glacialis*, which overlaps with *B. saida* in the Canadian Arctic³⁰ and has similar ice-associated habitat³¹. At least one hybrid individual has been confirmed with molecular data by Wilson et al.³². Identifying *B. saida* and removing individuals with introgressed *A. glacialis* information are important steps to ensure species level information is not conflated with population level dynamics, and for better understanding the evolutionary trajectory of both species. In addition to hybridizations, the genomic architecture of Canadian *B. saida* is likely to feature large inverted chromosomal segments as reported for gadid species in general³³, which results in reduced recombination between inverted genotypes that creates highly differentiated regions between genomes. This differentiation can subsequently offer important genomic variability on which adaptive evolution can act. These inversions are hypothesized to provide a template for adaptive divergence in cod populations, a mechanism explored in the Atlantic cod, *Gadus morhua*^{34,35}. By mapping genomic data for *B. saida* to the reference genome of *G. morhua*, Einarsson et al.³⁶ suggested the presence of inversions on chromosomes 1 and 7.

In this study, we describe the genomic structure of *B. saida* over wide latitudinal and longitudinal gradients covering Canadian Arctic and subarctic regions (from 48.25° to 76.39° N, and 53.71° to 138.65° W). Given the complexities detailed above, we proceeded in a stepwise fashion, starting with the objective to determine whether *B. saida*/*A. glacialis* hybrids were present in our dataset, and if so, the extent and distribution of introgressed information. Second, we screened our data for highly differentiated regions of the genome, in particular sex-linked regions and inverted genomic blocks acting as potential supergenes. Third, and once the aforementioned factors were removed from our dataset, we investigated patterns of population structuring for *B. saida* in Canadian waters. Our expectation, based on available genomic evidence for Gadidae^{32–36}, was that the genomic structure of *B. saida* would be layered and complex. Our analysis of 1000 s of SNPs provided further insights related to introgression events between *B. saida* and *A. glacialis*, large inverted genomic blocks similar to *G. morhua*, sex-linked loci of the genome, and new details about population structure across Canadian Arctic and subarctic waters.

Results

To infer the genomic structure of *B. saida*, we analysed 522 specimens from 92 sampling sites from the Western Arctic to the Gulf of St. Lawrence marine ecoregions (Fig. 1A). The initial analysis of 53,384 SNPs indicated the presence of nine highly differentiated individuals that were confirmed as *A. glacialis* based on mitochondrial DNA SNPs (Fig. 1B). Eight of these individuals were sampled in the Western Arctic while the remaining one originated from the Eastern Arctic (Table 1). Another 19 individuals belonging to a larger cluster identified as *B. saida* also had *A. glacialis* mitochondrial DNA (Fig. 1B). Individuals with *B. saida* nuclear DNA and *A. glacialis* mitochondrial DNA were sampled from all marine ecoregions investigated except for the Hudson Bay Complex, with the highest proportions in the Western Arctic (Table 1). Three individuals exhibited admixed proportions between *B. saida* and the *A. glacialis* bearing an ~ 11% membership probability in favour of the other species cluster (Fig. 1C). Such admixed individuals were collected in the Western Arctic ecoregion (Table 1).

A subsequent set of analyses to detect chromosomal inversions and including SNPs from only non-hybrid *B. saida* individuals (n = 511, SNPs = 12,305) revealed clustering patterns lacking any obvious geographical structure among the ecoregions considered. These patterns included two widespread clusters on PCA axis 3, and a further two clusters on axis 6 (Fig. 2A,B). Axis loadings also revealed highly aggregated signals on chromosomes 9 and 5, respectively (Fig. 2C,D). The loadings on chromosome 9 featured two notable peaks between positions 7,350,621–11,315,580 and 17,597,431–23,417,670 (Fig. 2C). A third genomic region was also identified as biased towards sex, though no geographic associations were found. A significant departure from random sex association was detected on a segment of chromosome 13 at positions 16,060,190–17,944,656 Mbp, which was biased toward

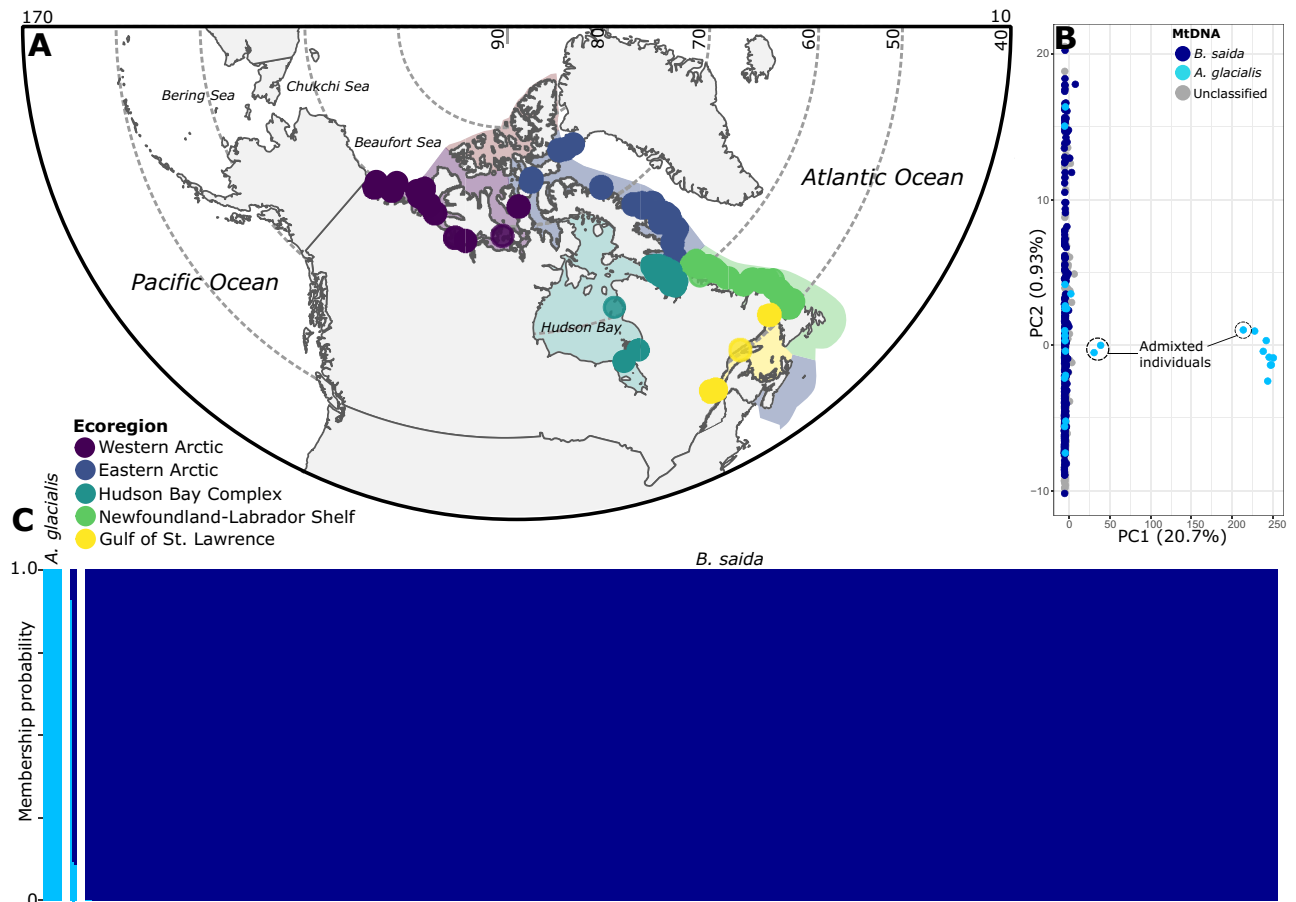


Figure 1. (A) Sampling locations for *Boreogadus saida* and *Arctogadus glacialis* specimens considered in this study. Specimens were sampled across five marine ecoregions in Canadian Arctic and subarctic waters⁷³. (B) Principal Component Analysis of 53,384 SNPs in 522 *B. saida* and *A. glacialis* individuals, with some *B. saida* individuals (left) harboring *A. glacialis* (right) MtDNA (admixed individuals noted); note, dots overlap, and some individuals were unable to be classified with mitochondrial SNPs. (C) ADMIXTURE analysis of the same 522 individuals. Map polygons were sourced from *rnaturalearth* R package⁹⁸ and set to polar projection.

Ecoregion	Sample size	Percentage of <i>A. glacialis</i>	Percentage of introgressed (<i>A. glacialis</i> x <i>B. saida</i>)	Percentage of <i>B. saida</i> with <i>A. glacialis</i> mtDNA
Western Arctic	117	6 (7)	2.5 (3)	7.7 (9)
Eastern Arctic	168	0.6 (1)	0	1.8 (3)
Hudson Bay Complex	58	0	0	0
Newfoundland-Labrador Shelf	118	0	0	2.5 (3)
Gulf of St. Lawrence	61	0	0	3.3 (2)

Table 1. Percentage of specimens from Canadian ecoregions that were *Arctogadus glacialis*, putative introgressed between *A. glacialis* and *Boreogadus saida*, and *B. saida* with *A. glacialis* mitochondrial DNA. Sample sizes are provided in brackets.

males (heterogametic sex, Fig. 2E). The aggregated signals on the three chromosomes were further corroborated with linkage disequilibrium measurements (Fig. 2F). Other notable peaks of linkage disequilibrium were also present on other chromosomes (Fig. 2F). A PCA of SNPs located within the putative inverted segments of chromosome 5 and 9 both featured three clusters (Fig. S2): two clusters with low heterozygosity, i.e., homozygous genotypes for the ancestral or the inverted segment, on each extremity of the first axis, and a third cluster with high heterozygosity, i.e., genotype heterozygous with an ancestral and inverted segment, close to the middle of the first axis (Fig. S2). All inversion genotypes lacked geographic specificity. Despite investigating 428, 634, and 680 Ensembl gene codes for the biased regions on chromosomes 5, 9, and 13, respectively, g:Profiler only revealed a few instances of potential gene enrichment for ribosomal functions and RNA processing (Fig. S3).

The analysis of population structure in *B. saida* without SNPs from the biased regions of chromosomes 5, 9, and 13 (N = 11,233 SNPs) revealed three main clusters in the study area. The PCA revealed that the majority of

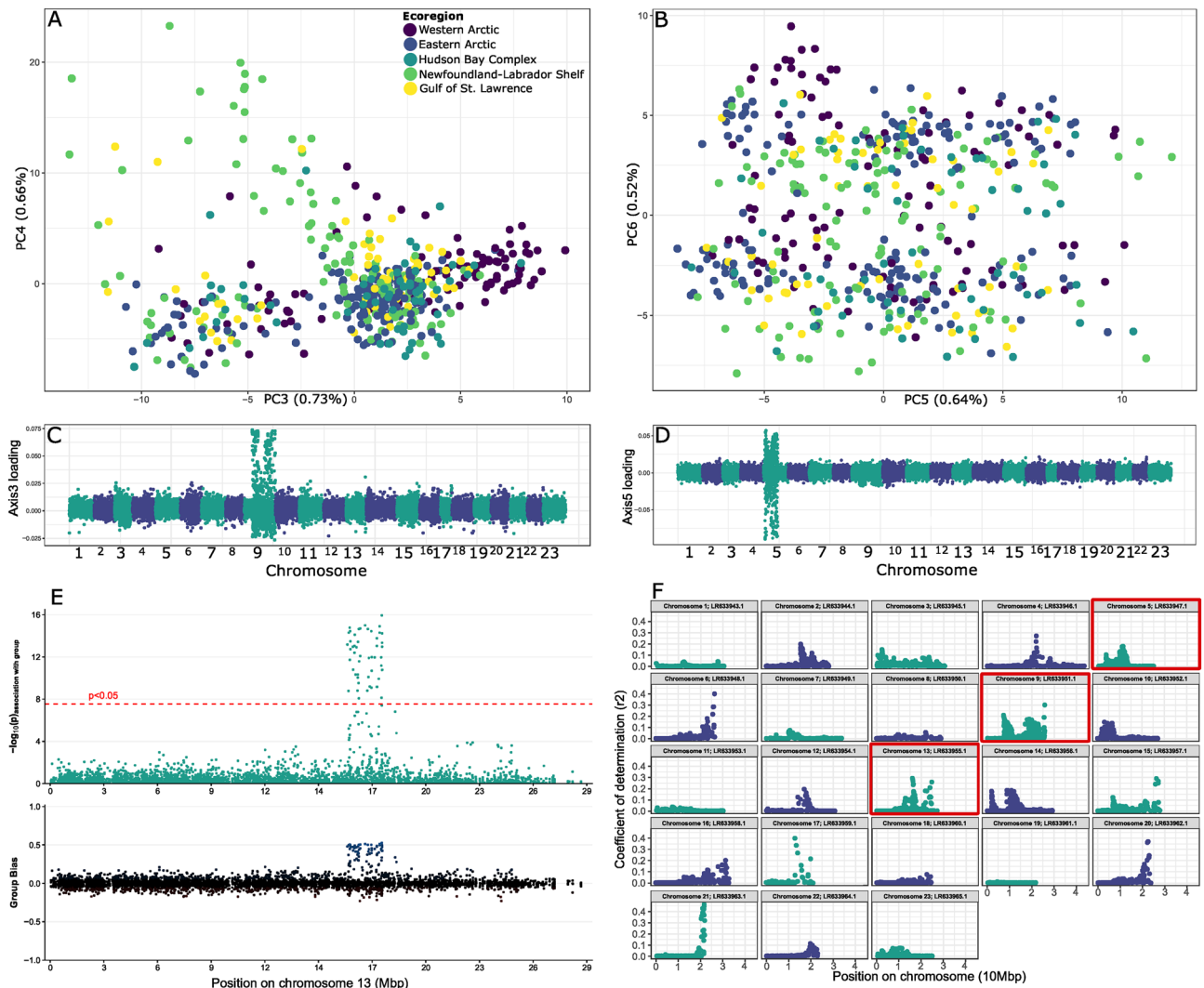


Figure 2. Initial population results of Canadian *Boreogadus saida* based on genome-wide SNPs ($n = 12,305$ SNPs). (A) Principal Component Analysis (PCA) axes 3 and 4 demonstrating lack of geographic specificity to clustering (PCs 1 and 2 depict structure and allele loadings in Fig. 3). (B) PCA axes 5 and 6 demonstrating similar results. Allele loading profiles (C and D) linked to chromosomes 9 (LR633951.1) and 5 (LR633955.1), respectively. (E) RADSex results showing sex bias on chromosome 13 (note x-axis indicates beginning position of mapped reads within the chromosome). Upper panel presents the probability of association with sex, with the dashed line indicating a p -value of 0.05 after Bonferroni correction, while the lower panel indicates the direction of bias, where positive values indicate the bias toward males and negative, toward females. (F) Linkage disequilibrium (LD) averaged for all loci pairs separated by 100 kbp–5 Mbp. Spikes in LD indicate sites are correlated with adjacent sites to within 5 Mbp, suggesting these regions are inherited as a unit. Chromosomes 5, 9, and 13 are highlighted by red boxes, but note LD spikes are present on other chromosomes.

individuals from the Hudson Bay Complex, Newfoundland-Labrador Shelves, and some individuals from the Gulf of St. Lawrence, differed from individuals of the Western and Eastern Arctic on the first axis. Specimens from the Western and Eastern Arctic ecoregions were differentiated on PCA axis 2 (Fig. 3A), though there was considerable overlap. This clustering pattern was replicated with Discriminant Analysis of Principal Components (DAPC), and allele loadings were aggregated to some extent across most chromosomes (Fig. 3A). In line with these results, while all pairwise F_{ST} comparisons were significantly different from 0 (all $p < 0.001$), the pairwise comparison of Western Arctic and Newfoundland and Labrador Shelves ecoregions was the most differentiated ($F_{ST} = 0.0112$), followed by comparably high differentiation between the Western and Eastern Arctic ($F_{ST} = 0.0096$), between the Western Arctic and Hudson Bay complex ($F_{ST} = 0.0086$), and between the Eastern Arctic and the Newfoundland and Labrador Shelves ($F_{ST} = 0.0095$; Fig. 3B). The Gulf of St. Lawrence was the least differentiated ecoregion across all comparisons, while the Hudson Bay Complex and Newfoundland and Labrador Shelves similarly showed low differentiation (Fig. 3B). Inbreeding coefficients indicated heterozygosity was significantly different across ecoregions ($p = 0.002$). Specifically, heterozygosity appeared to be highest in one of the subarctic ecoregions, Newfoundland and Labrador Shelves. Dunn's post hoc test confirmed F was significantly less in Newfoundland and Labrador Shelves compared to Hudson Bay Complex ($p < 0.001$; Fig. 3C),

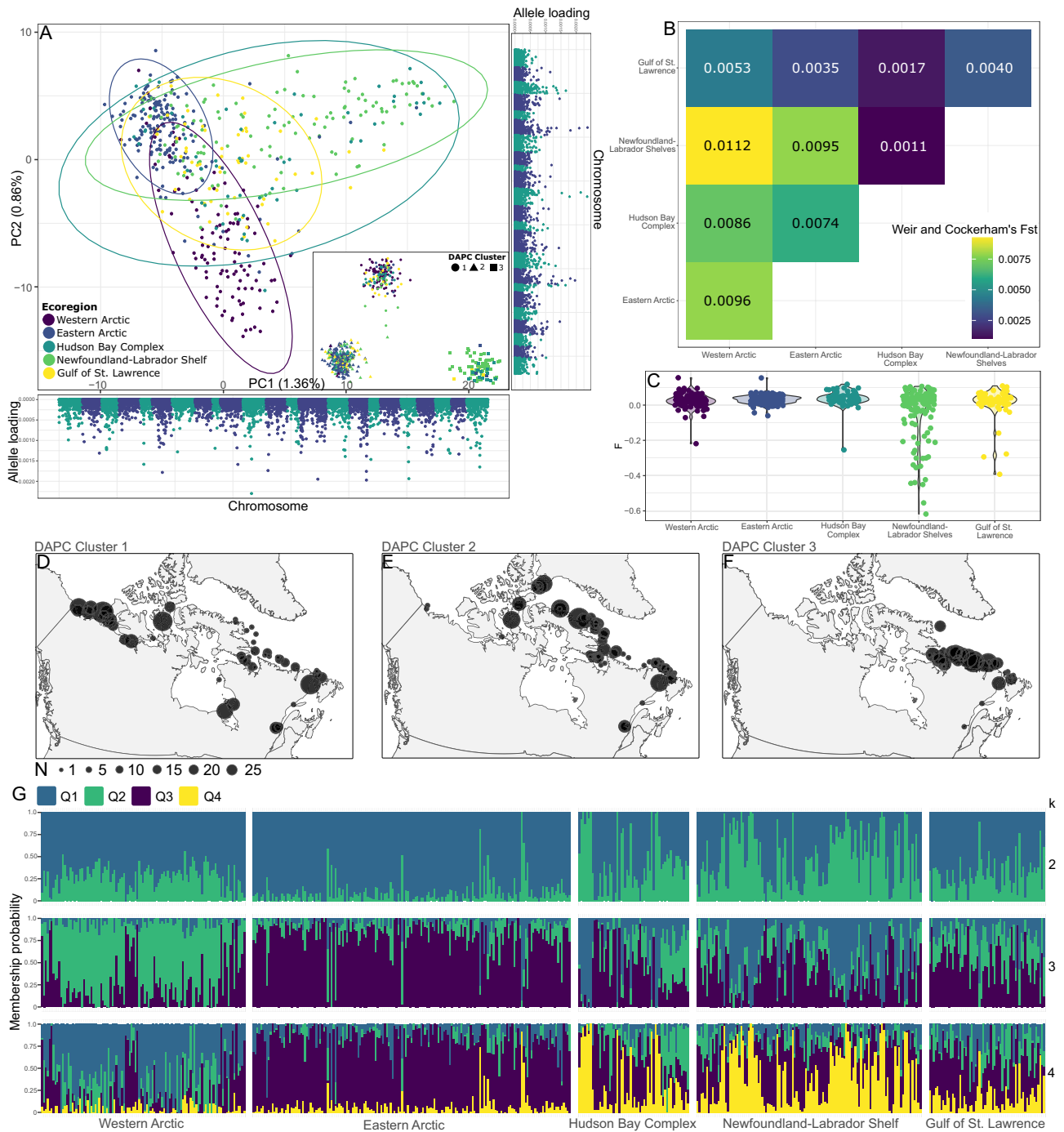


Figure 3. Population structure of Canadian *Boreogadus saida* ($n = 11,233$ SNPs). Results include: **(A)** Principal Component Analysis with Discriminant Analysis of Principal Components (DAPC) inset and discriminant analysis allele loading profiles depicted for each axis (assuming 3 clusters and ca. 80% of variation retained during analysis); **(B)** Weir and Cockerham's F_{ST} pairwise measurements for each ecoregion (all pairwise estimations were significantly different from 0; all $p < 0.001$); **(C)** Inbreeding coefficients; geographic distribution of DAPC clusters 1–3 (**D–F**, respectively), where N refers to number of individual specimens genotyped; and **(G)** ADMIXTURE results at $k = 2–4$ ($k = 3$ was best supported based on cross-validation error, Fig. S1).

but did not confirm significant differences in other pairwise comparisons. Though all DAPC clusters were distributed throughout the study region, geographic patterns were still notable. Most Western Arctic specimens were associated with cluster 1 (Fig. 3D), while most specimens from the Eastern Arctic were associated with cluster 2 (Fig. 3E). The third cluster was only present in the eastern Canada distribution with a majority of samples collected along the Labrador coast (Fig. 3F). These three clusters were likely not completely reproductively isolated since higher levels of heterozygosity appeared to aggregate at the facing margins of each cluster (Fig. S4). ADMIXTURE corroborated the above structure inferred from PCA and F_{ST} , with distinct structure

corresponding to the Western Arctic, the Eastern Arctic, and the continuum from Hudson Bay Complex through to the Gulf of St. Lawrence (Fig. 3G).

Allele loadings were notably aggregated in the above analyses (Fig. 3A), so outlier analyses were used to corroborate these results. Outlier loci contributing excessively to observed patterns in population structure were identified using both PCAdapt (Fig. 4) and BayeScan. Assuming a 5% false discovery rate, 13.8% of loci were identified as outliers with PCAdapt ($n = 1553$; Fig. 4A) and 4.1% with BayeScan ($n = 465$), with 4.0% ($n = 443$) identified with both approaches. Using a more conservative Bonferroni corrected threshold, 4.3% ($n = 487$) of loci were identified as outliers, with 71.3% ($n = 347$) in common with BayeScan. Loci were aggregated within chromosomes (Fig. 4A) and mirrored peaks in linkage disequilibrium (Fig. 2F). The outlier loci identified with PCAdapt replicated the population structure of the full dataset using both conservative and normal thresholds (Fig. S5). The PCA of the putative neutral loci (i.e., not identified as outlier by PCAdapt, $n = 9680$) identified clusters typical of inversion genotypes in PC axes 1 and 2 (Fig. 4B) and supported the inferred population structure depicted in Fig. 3 for PC axes 3 and 4 (Fig. 4C). The allele loadings indicated the signal remained aggregated in a few genomic positions, except for chromosome 14 between positions 8,874,942–16,337,747 (Fig. 4D). A PCA of the chromosome 14 biased region revealed a clustering pattern related to levels of heterozygosity or inversion genotypes (Fig. S2). The final dataset with the biased region of chromosome 14 removed, and retaining 1 SNP/inter- and intra-chromosomal linkage group ($n = 7133$) again reflected to a weaker extent the previously inferred population structure, with Newfoundland-Labrador Shelves and Hudson Bay Complex individuals differentiated on PCA axes 1 and 2 (Fig. 4E), and Western Arctic individuals differentiated on axis 3 (Fig. 4F). Allele loadings confirmed the aggregated signal was largely removed (Fig. 4G). Results for population analyses were nearly identical for the unlinked loci (Fig. S6), except clusters were less well defined, and only a single cluster was supported by DAPC and ADMIXTURE (Fig. S1).

Discussion

Previous genetic studies relying on mitochondrial DNA, microsatellites, and small-scale SNP datasets (100+) have suggested population structure in polar cod, *Boreogadus saida*, is generally absent or weak owing to high connectivity across its range, though global scale structure has been inferred in some cases^{19,25}. Higher resolution datasets promise to reveal genomic structure potentially missed in previous analyses, but many complications owing to the biology of gadids, such as putative chromosomal inversions, must be carefully ruled out before interpreting particular factors affecting structure such as gene flow. Using 1000s of SNPs for higher resolution, our objective was to peel back layers of genomic structure in *B. saida* from five Canadian Arctic and subarctic ecoregions, specifically screening for potential hybridizations with *Arctogadus glacialis*, and putative large genomic inversions and sex-linked regions, prior to interpreting population structure. Our analysis of 53,384 SNPs confirmed introgression of genomic information between *B. saida* and *A. glacialis*, suggesting hybrids backcross with both parental species. We also unravelled important genomic architecture suggesting inverted genomic segments on chromosomes 5, 9, and 14, with another segment linked to sex differences on chromosome 13. We found that population structure in Canadian waters is driven by a small subset of outlier loci aggregated in numerous regions throughout the genome, which are potentially the result of adaptive divergence across the North American Arctic. Once all confounding factors were removed from the SNP dataset, population structure was weak but reaffirmed patterns observed prior to removing putative inversions and adaptive loci. Our study provides new insights into the biology of *B. saida*, and relevant genomic context for protecting this species and its associated ecological functions under a changing climate.

Interbreeding of *Boreogadus saida* and *Arctogadus glacialis*

B. saida co-occurs with another gadid *A. glacialis* in the Canadian Arctic. *A. glacialis* is less widely distributed and abundant, but does overlap with *B. saida* on Arctic continental shelves, mainly in the Northern Chukchi Sea, Northwest reaches of the Canadian Arctic Archipelago, Northern Greenland, and Arctic Russian waters^{30,37}. The two species can be distinguished morphologically based on total gill rakers on the anterior arch, and the number of rays of the second dorsal fin and first anal fin³⁰. Mitochondrial DNA³⁸ and a single microsatellite marker (Gmo8) have been previously reported to distinguish the two species³⁹, though as explored below, this mitochondrial DNA signal may not be an entirely reliable diagnostic criterion.

Despite differences in ecology, morphology, and genetics^{7,30,39}, we found evidence of introgression between *B. saida* and *A. glacialis* based on principal component and ADMIXTURE analyses. It was possible to discriminate these admixed individuals since both parental species were genotyped in our dataset. The initial results identified nine highly differentiated individuals caught in the Western and Eastern Arctic ecoregions that were identified as *A. glacialis* on the basis of nuclear loci and mitochondrial haplotypes (Fig. 1A,B). A further 19 individuals clustering with *B. saida* for nuclear loci were genotyped for *A. glacialis* for mitochondrial haplotypes, suggesting directional introgression of *A. glacialis* mitochondrial DNA into *B. saida*. While this observation could stem from shared ancestral polymorphisms, the presence of three admixed individuals strongly suggests these patterns are due to introgression of mitochondrial DNA. Introgression of cytoplasmic DNA is not uncommon in lineages with the potential to interbreed, and has been widely reported in animals and plants undergoing range expansions⁴⁰. If the patterns reported here are due to introgression, this was potentially initiated during recolonization of the Arctic following the last glacial maximum ca. 12 ka⁴¹ once overlapping species ranges were established, though time calibrated analyses are needed before this can be confirmed. Given both species spawn under sea ice, with sympatric development of eggs and larvae³¹, the potential for interbreeding appears biologically highly plausible. Because organellar genomes are uniparentally inherited, the *A. glacialis* haplotype has potentially become a divergent sequence in the landscape of mitochondrial genetics of *B. saida*. As such, it is imperative that the identification of specimens does not rely on mitochondrial data alone (e.g. common application of cytochrome

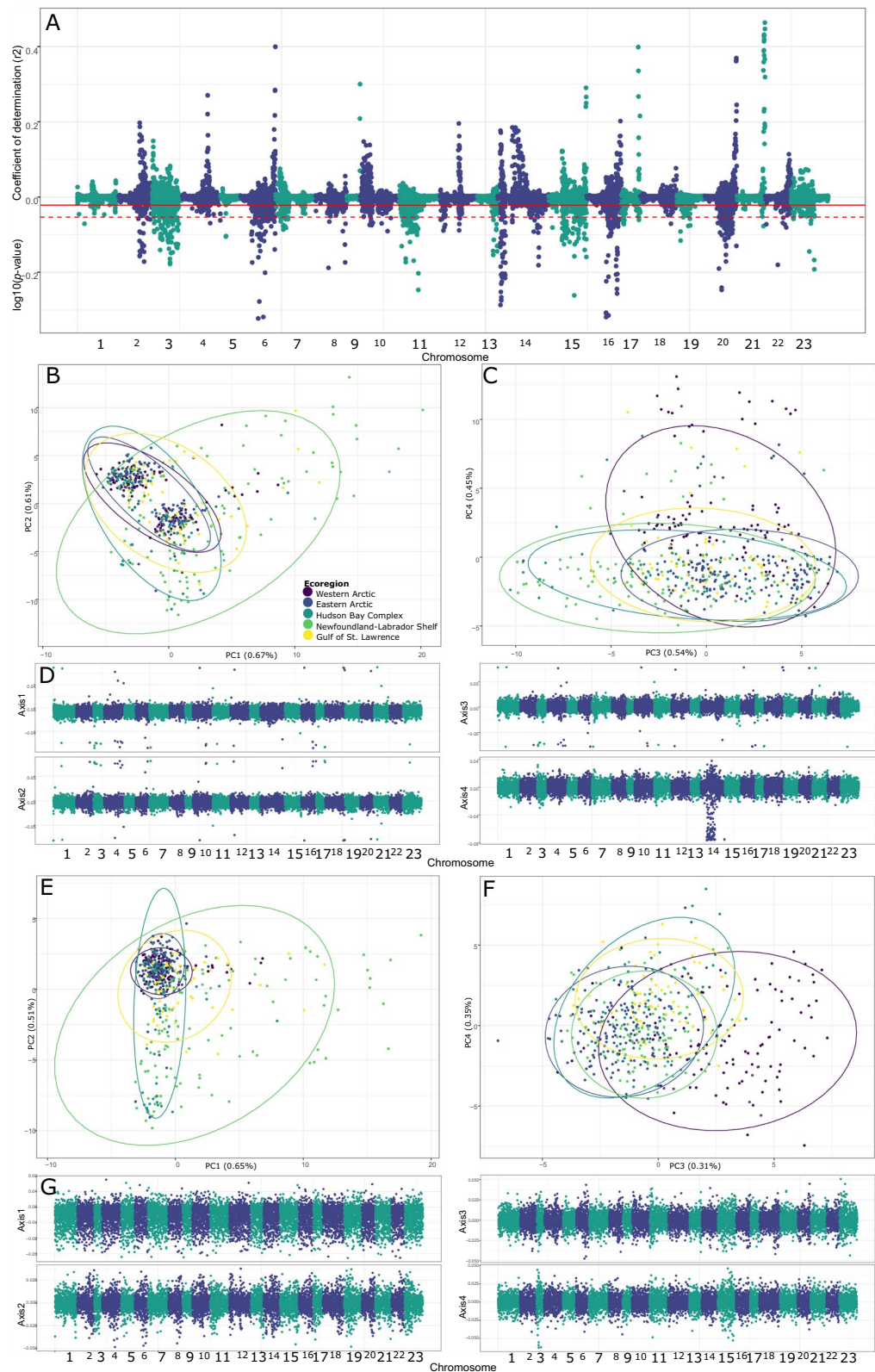


Figure 4. Population structure of outlier and neutral loci identified by PCAdapt in Canadian *Boreogadus saida*. (A) Manhattan plot of linkage disequilibrium averaged for all loci pairs separated by 100 kbp-5 Mbp (positive values) and PCAdapt $\log_{10}(p)$ -values, negative values) for all loci. The solid red line depicts a threshold for candidate loci under selection based on q-values assuming a false discovery rate of 5%, whereas the dashed red line is a more conservative Bonferroni corrected threshold. (B) Principal Component Analysis (PCA) axes 1 and 2 based on neutral loci identified using the less conservative q-value threshold, along with (C) axes 3 and 4 ($N=9680$ SNPs). (D) Allele loadings of neutral loci after filtering for linkage disequilibrium. PCA of unlinked loci ($n=7133$ SNPs), depicting axes 1-2 (E) and 3-4 (F). (G) Allele loadings of the PCAs based on unlinked loci.

c oxidase I in environmental DNA surveys) given the potential to conflate the two species, a limitation that can be overcome with nuclear markers³⁹.

We also found evidence for hybridization and introgression of nuclear alleles in at least three *B. saida* individuals, two sampled in the Coronation Gulf of the Kitikmeot Region of Nunavut (Western Arctic ecoregion), and the third sampled between Prince of Wales Island and the southern tip of Somerset Island within the Eastern Arctic ecoregion. Previous work by Wilson et al.³² similarly identified a potential late generation hybrid with *A. glacialis* mtDNA and *B. saida* nuclear genotype sampled close to the Mackenzie River Delta, Beaufort Sea (Western Arctic ecoregion). The authors also simulated assignment probabilities following hybridization, and showed admixture proportions are practically negligible after more than 5 generations of backcrossing. Our proportions detected using maximum-likelihood ancestry models (i.e. ADMIXTURE; ca. 11/89%) were larger than those in the analysis of Wilson et al.³², suggesting hybridization between the two species is recent in Canada's Western and Eastern Arctic ecoregions, despite millions of years of evolution between the two species⁴². Moreover, introgression of nuclear genomic information appears to flow in both directions across species, as suggested by the smaller *B. saida* proportion in one hybrid individual, and the smaller *A. glacialis* proportions in the remaining two hybrid individuals (Fig. 1C). We also noted around 3% (17 out of 511) of specimens collected as *B. saida* featured the mitochondrial DNA of *A. glacialis* (Table 1). In contrast, Nelson et al.²⁵ removed a greater proportion of specimens with *A. glacialis* mitochondrial DNA (12%; 300 out of 2587), while Wilson et al.³² sampled 115 *B. saida* and 12 *A. glacialis* from Northern Alaska and the Western Canadian Arctic, with no mixing of mitochondrial and nuclear signals. Multiple reasons could explain these differences in reported introgression of *A. glacialis* mitochondrial DNA into *B. saida*, such as taxonomic expertise, sample sizes, and the extent/study areas, all of which varied considerably among studies. We hypothesize that differences in introgression rates may vary across the entire distribution range of *B. saida* (Table 1); confirming this would necessitate a broader study using a similar genome-wide genotyping approach.

Inverted and sex-linked genomic regions in *Boreogadus saida*

We detected several genomic regions that contributed to important clustering patterns in population analyses of *B. saida*. The first dubious clustering pattern we detected on PC axes 3 and 4 of our initial population level PCA (Fig. 2A) was partially driven by two ca. 4–6 Mbp regions on chromosome 9 (Fig. 2C). Allele loadings (i.e. contributions to genomic signal) on axis 5 also pointed to a differentiated region on chromosome 5, and our downstream analyses pointed to a differentiated region on chromosome 14 (Fig. 4B,D). Given these regions were differentiated, with sites in high linkage disequilibrium (i.e. inherited as a unit; Fig. 2F), without any obvious spatial structure, and featured obvious clustering corresponding to high and low heterozygosity genotypes (Fig. S2), we concluded these regions represent large (i.e. Mbp) inverted regions of the genome. Large chromosomal inversions reduce recombination due to disruption of crossing over events in heterozygotes⁴³, leading to regions of the genome with elevated linkage disequilibrium and *F_{ST}* (as observed here; Fig. 2F), and notable departures in heterozygosity between individuals lacking an inverted region (putatively ancestral AA and recently derived BB genotypes) and hosting one (AB; Fig. S2). We hypothesize the BB genotype represents two copies of the inverted region, given higher levels of homozygosity are expected when the large accumulation of alleles is brought back together. On chromosome 9, given two notable peaks, we speculate the ancestral state was a single inverted segment that broke into smaller inverted regions with varying levels of recombination (Fig. 2C,F, S2). Alternatively, these inverted regions were potentially established together at the same time. Chromosomal inversions have been reported in Atlantic cod (*Gadus morhua*), with similar PCA patterns in Northwest Atlantic populations reported by Puncher et al.⁴⁴, and adaptive evolution inferred from these regions first reported by Bradbury et al.⁴⁵. Sodeland et al.³⁴ also link these “islands of divergence” to adaptive evolution in Atlantic cod, and hypothesize linkage groups are tailored to coastal and oceanic environments. Berg et al.³⁵ also link these inversions to adaptive divergence in Northeast Arctic, Norwegian Coastal, and North Sea cod ecotypes. In Atlantic herring, a large inversion on chromosome 12 is associated with ecological adaptation to water temperatures and gonadal development, spawning, or early larval development⁴⁶, while in capelin large chromosomal inversion(s) are also reported but not yet linked to ecological traits⁴⁷. For *B. saida* in Canadian Arctic and subarctic waters, we found no obvious geographic specificity or ecological adaptations associated with the occurrence of the inverted repeat regions (Fig. S2). Nonetheless, chromosomal inversions suppress recombination in particular regions of the genome, creating the conditions for adaptive divergence in species with high gene flow, as appears to be the case in *B. saida*.

Inversions were reported for *B. saida* on chromosomes 1 and 7 by Einarrsson et al.³⁶, which are possibly derived from a common ancestor with Atlantic cod, or an introgression event. Curiously, however, we did not observe any evidence for divergent chromosomal segments corresponding to the inverted regions reported by Einarrsson et al.³⁶, including genomic regions in linkage disequilibrium on chromosomes 1 and 7 (Fig. 2F). Our results therefore indicate the inversions on chromosomes 1 and 7 are absent from Canadian populations, either because the inversion was lost or fixed, or because the inversions are potentially exclusive to Greenlandic populations (the study area of Einarrsson et al.³⁶). A larger scale analysis incorporating specimens from the entire circumpolar distribution range of *B. saida* is needed to determine the pervasiveness of these inversions. Einarrsson et al.³⁶ did not report on the putative inverted regions reported here from chromosomes 5, 9, and 14, and these regions were not previously reported to house chromosomal inversions in *G. morhua*. Our results therefore suggest *B. saida* chromosomal inversions have independently evolved, and are spatially segregated across its circumpolar distribution.

We also confirmed a region on chromosome 13 was linked to sex biased divergence (Fig. 2E). Determining factors of sexual differentiation in fish are highly variable, ranging from genetic sex determination with distinct chromosomes to diffuse loci, to determination through environmental conditions⁴⁸. Here, we show that sex

determination in *B. saida* is at least partially genetically driven through an XY mating system, though other factors are likely playing an important role given males and females did not cluster in our PCAs. In Atlantic cod, sex determining loci are found across five linkage groups, one of which closely approximated an XY mating system⁴⁹. None of the linkage groups reported by Star et al.⁴⁹, however, correspond to the region identified here, suggesting a rapid turnover of sexual determination in *B. saida* or *G. morhua*, a phenomenon frequently reported in fish^{50–52}.

Aside from nucleotide differences, we did not note any noteworthy functional differences specific to the inversion and sex-linked genomic clusters (e.g. Fig. S3). Future work would need to document or include information on habitat use/association, the occurrence of ontogenetic migrations, and prevailing environmental conditions (e.g. near-shore vs open ocean, brackish vs marine), as well as biological information (e.g. specimen size and timing of reproduction), to potentially unravel environmental effects and/or behavioral, morphological, or life history trait differences indicative of adaptation. The inverted regions (chromosomes 5 and 9) were associated with a wide variety of Gene Ontology terms, ranging from heme binding to cytoskeletal motor activity, while the sex-linked cluster contained genes associated with ion channel activity and metalloendopeptidase inhibition, though none of these regions were significantly enriched with particular functions (Fig. S3). Whether these genomic regions are of any adaptive significance requires further analysis, ideally leveraging a reference genome for *B. saida* rather than *G. morhua*. Note also, given the linkage disequilibrium results (Fig. 3F), more inversions could be confirmed in future analyses.

Population structure of Canadian *B. saida* reflects regional scale structuring and potential adaptive evolution

Our analysis of population structure in *B. saida* revealed clear spatial patterns across the species range within Canada (Fig. 3A), with notable structure broadly corresponding to three ecoregions/areas: (1) the Western Arctic, (2) the Eastern Arctic, and (3) the area comprising the Hudson Bay Complex, the Newfoundland-Labrador Shelves, and the Gulf of St. Lawrence ecoregions (Fig. 3D–F). This structure supports the findings of Nelson et al.²⁵, wherein microsatellites reflected similar population differentiation in the Eastern versus Western North American Arctic. Our study moreover confirmed that some specimens from the Hudson Bay Complex through to subarctic waters of the Newfoundland and Labrador shelves and Gulf of St. Lawrence ecoregions are also differentiated from higher Arctic locations (Fig. 3). Some population structure (i.e. Q4, Fig. 3G) was especially associated with the Newfoundland-Labrador shelves, where heterozygosity was also notably high (Fig. 3C). Several hypotheses might explain these patterns. The Arctic environment was recently covered with thick-ice sheets during the last glacial maximum⁴¹, which presumably extirpated the highly productive under ice algae and zooplankton that sustain oceanic Arctic ecosystems⁶. As such, *B. saida* populations were likely also extirpated from the Arctic during glaciations, surviving in glacial refugia isolated in the North Pacific and North Atlantic oceans, as inferred in widespread marine lineages including various invertebrates⁵³, sea stars⁵⁴, polychaetes⁵⁵, urchins⁵⁶, and seaweeds^{57–59}. Isolation into separate Atlantic and Pacific refugia could explain the divergence between Western Arctic and Newfoundland and Labrador Shelves ecoregions (Fig. 3B). The Eastern Arctic, meanwhile, could be the result of recent range expansions, either from Atlantic or Pacific refugia. The overall low FST values (Fig. 3B), weak structure globally^{16–23,25}, and high migration potential, does not exclude the possibility of circumpolar expansion from a single basin following glaciation, followed by differentiation of globally established populations, or some combination of the above scenarios. Increased geographic sampling of populations and demographic modelling are needed to unravel the series of events and timeline for evolution in *B. saida* populations.

The high heterozygosity of the subarctic Canadian population (Hudson Bay Complex, Newfoundland and Labrador Shelves and Gulf of St. Lawrence), however, does not support the hypothesis of recent Arctic recolonization from a single source population, as these results suggest the subarctic population is outbred. Glaciation is also known to have pushed marine populations into East vs West Atlantic refugia^{60–62}, and the work of Nelson et al.²⁵ indicated European Arctic populations are differentiated from Canadian populations. Bigg et al.⁶³ also support the view of refugia scattered across the Atlantic during the last glacial maximum for *G. morhua*, with solid agreement between paleoecological models and a variety of genetic data. The signal unique to *B. saida* inhabiting subarctic waters of Canada's East coast, combined with the observation of high heterozygosity in some individuals, points to potential interbreeding with another differentiated population. One possibility is interbreeding between DAPC clusters 2 and 3 (Fig. 3A), which is somewhat supported by high heterozygosity at the facing margins of clusters (Fig. S4). Refugial Greenlandic populations are also worth considering, a hypothesis supported in kelp populations both through ecological niche modelling⁶⁴, and genomic datasets⁶⁵. A global scale analysis of genome-wide SNPs is needed to determine the origin of the unique population signal and high heterozygosity encountered in specimens from the three subarctic ecoregions.

We also observed that the population structure reported here was excessively driven by a small number of loci, largely aggregated in 13.8% of the SNPs and into particular genomic regions scattered across chromosomes, a finding also reflected in the linkage disequilibrium results (Fig. 4A). Though the reported population structure remains in the neutral loci (Fig. 4E,F, S6), the observation that population structure is disproportionately driven by a relatively small number of loci suggests adaptive evolution is playing an important role in globally distributed *B. saida* populations⁶⁶. One possibility is that these adaptive regions may also represent smaller chromosomal inversions, meaning many, potentially more recent, inversions are playing a role in the evolution of *B. saida* populations. Examples of adaptive evolution located in inversions are suggested in Atlantic cod^{45,67}, and islands of adaptive divergence facilitated by inversions are reported in freshwater and marine threespine sticklebacks (*Gasterosteus aculeatus*)⁶⁸. In the case of *B. saida*, divergence has been reported between fjord-dwelling and shelf-dwelling populations in the East Atlantic²³, but this mechanism for adaptive evolution, among others explored here, is supported with little direct evidence (though see Hill²⁴). A hypothesis worth exploring is

whether specimens from the subarctic ecoregions (Hudson Bay Complex, Newfoundland and Labrador Shelves and Gulf of St. Lawrence) are adapted to a shorter ice season and thus a more pelagic habitat. A more conservative, less adaptation oriented conclusion of our results, is that genetic drift plays a minor role in structuring populations given there is little random structure present in Canadian populations. Explanations for this include consistently large effective population sizes possibly mediated by complex source-sink dynamics or recently expanded populations (see above discussion on Arctic recolonization; see also positive impact of expansion on false discovery rate of outlier SNPs in Luu et al.⁶⁹). Of course, a dominant role of adaptive evolution^{45,66–68}, or a combination of neutral and non-neutral factors are viable possibilities. An analysis of functional differentiation of populations would benefit immensely from a reference genome for *B. saida*, along with whole genome sequence information to pinpoint peaks in adaptive signal^{70–72}, and environmental and biometric/life history data that could be coupled with an analysis of functional variation among specimens.

Conclusions

Boreogadus saida is a key, mid-trophic level species modulating energy flows between lower and higher levels in Arctic ecosystems. Our genome wide analysis of population SNPs provides important context for understanding the potential vulnerability of this species to ongoing climate change and direct anthropogenic stressors. In particular, if West vs East Arctic and subarctic regions are functionally differentiated, we could expect different responses to climate change across the Canadian range. The introgression between *B. saida* and *A. glacialis* could also have functional implications in the future, particularly in the Western Arctic where hybridization occurs. If populations are not yet functionally differentiated, ample substrate exists for selection given three putative chromosomal inversions (chromosomes 5, 9, and 14) and the linked outlier loci across chromosomes in Canadian *B. saida* populations. A reference genome for *B. saida*, along with sampling from the entire circum-polar species distribution range, corroborated with environmental and biometric/life history data, are needed to confirm the positions of SNPs and inversions and to delve further into the functional relevance of the patterns observed here. Doing so will provide further genomic context for the protection of the species and ecosystems that depend on *B. saida*.

Methods

Data collection, DNA extraction, and sequencing

Individuals were collected with bottom or pelagic trawls during scientific surveys (CCGS *Amundsen*, CCGS *Teleost*, CCGS *Leim*, the Aqviq and the Katsshshuk II of the Northern Shrimp Research Foundation—Fisheries and Oceans Canada) in five marine ecoregions (i.e., Canadian Marine Areas⁷³): the Western Arctic (2021, n = 123) and the Eastern Arctic (2021, n = 175), the Hudson Bay Complex (2005, 2021, n = 63), the Newfoundland and Labrador Shelf (2021, n = 149) and the Gulf of St. Lawrence (2020, 2022, n = 66; Fig. 1A; Table S1). A few individuals were also sampled from ice fishing in the Saguenay Fjord (2019, n = 26), considered here as the Gulf of St. Lawrence ecoregion. The ecoregions used for analyses are high-level spatial units primarily based on oceanographic and bathymetric similarities⁷³. Individuals were measured and sex was determined visually. Muscle tissue or fin clips were preserved in EtOH 95% prior to the DNA extraction.

DNA was extracted with DNeasy Blood and Tissue Kits (Qiagen). Libraries of ddRAD were prepared by the Plateforme d'analyse génomique (IBIS, Université Laval) using *Pst*I and *Mst*I restriction enzymes and 20 ng of DNA per sample [see detailed protocol in⁷⁴]. Six libraries of 96 samples (570 individuals and 6 duplicates) were sequenced on one lane of NovaSeq 6000 S4 PE150 (NovaSeq 6000 S4 Reagent Kit v1.5, 300 cycles) by Génome Québec (Montréal, Canada) with 10% PhiX. The targeted sequencing depth was > 10x.

Read processing, SNP calling and filtering

Illumina adaptors were first removed using Trimmomatic v0.39⁷⁵. Then, demultiplexing and quality filtering were performed with the *process_radtag* module of Stacks v2.64^{76,77} with a truncation at 90 bp and check for restriction cut sites. Given the lack of reference genome in *B. saida* and previous work in *B. saida* mapping to the reference genome of Atlantic cod³⁶, reads were aligned to the chromosomal level reference assembly GadMor3 (GCF_902167405) with *mem* from BWA-MEM⁷⁸ and default parameters. Ten samples with low numbers of mapped reads (< 10 millions) were discarded.

The *gstacks* module was used to create a SNP catalog, then the *population* module was run over high quality samples (mean coverage > 10x) to compile into a Variant Call Format (VCF) file. SNPs with a minor allele frequency greater than 1%, and present in over 75% of samples were kept in the dataset, while one sample with > 30% missing values was removed from the dataset. Next, SNPs with > 10% missing values, with expected heterozygosity > 0.6, and with low (< 10x) and high (> 95x) median coverage were removed with VCFtools⁷⁹. Relatedness coefficients were checked using the *relatedness2* function of VCFtools and one individual each from two pairs of related individuals ($\Phi < 0.3$) were removed. Finally, to avoid SNPs in linkage disequilibrium (i.e. in close proximity on the genome), only the first SNP of each ddRAD loci was kept (Table 2).

Genomic analyses

SNPs located within the mitochondrial genome (i.e. haplotypes) were explored to assess mitochondrial identity, particularly for potential misidentification with *A. glacialis*. Raw reads were first aligned to the *B. saida* reference mitogenome (NC_010121.1). Then, haplotypes starting in three regions with high density observations (position 4894, 5138, 9134) were blasted to a local version of NCBI-nt (downloaded 2022-01-04) to assign them a taxonomic identity (Top hit approach, 95% identity threshold). Five individuals were targeted for the sequencing of the Folmer region of cytochrome C oxidase using the LCO1490 and HCO2198 primers⁸⁰ to confirm which haplotypes were discriminants. We determined that one haplotype was discriminant for *B. saida* (position 5138)

Filtration step	SNPs	loci	N
Initial samples			576
Mapped reads > 10 millions			566
SNPs catalog creation (gstacks)		1,046,093	566
Minimum 10 × coverage		1,046,093	532
Minor Allele Frequency (MAF) > 1% and SNP detected in at least 75% of individuals	227,266	84,220	532
Individuals with < 30% missing SNPs, and SNPs with < 10% missing values	191,001	56,312	526
Observed heterozygosity lower than 60%	190,049	56,049	526
SNPs with coverage between 10 and 105x (approx. 1–99% percentile)	184,620	53,550	526
Relatedness (Φ) < 0.3	184,620	53,550	522
Species level analyses			
1 SNP per locus, MAF > 1%	53,384	53,384	522
Analyses of inversions and sex-biased loci			
Minor allele frequency > 5% and <i>Arctogadus glacialis</i> and hybrid individuals removed	12,305	12,305	511
Sites with elevated axis/allele loadings on chromosome 5 removed (i.e. putative inversion)	11,991	11,991	511
Sites with elevated axis/allele loadings on chromosome 9 removed (i.e. putative inversion)	11,428	11,428	511
Sex linked loci on chromosome 13 removed	11,233	11,233	511
Analyses of outlier SNPs			
Outlier loci removed based on PCAdapt q-values	9,680	9,680	511
Sites with elevated axis/allele loadings on chromosome 14 removed (i.e. putative inversion)	9,458	9,458	511
Analyses of neutral population structure			
One SNP per linkage group ($r^2 > 0.05$ within 10 Mbp; $r^2 > 0.10$ inter-chromosome) retained	7,133	7,133	511

Table 2. Counts of ddRADseq single nucleotide polymorphisms (SNPs), loci, and samples after filtering steps. SNPs/loci are carried forward at each step to progressively filter data, thus, each row includes filtering steps above. Filtering steps indicate criteria to retain SNPs, loci or samples.

while another was discriminant for *A. glacialis* (position 4894). These two haplotypes were used to genotype mitogenomes of all specimens (Table S2).

Nuclear SNP information was also used to confirm species identity and remove *A. glacialis* and potential hybrids prior to population analyses of *B. saida*. Principal component analyses (PCAs) were conducted on the full dataset (522 individuals, 53,384 SNPs) using the *glPca* function of *adegenet* v.2.1.10^{81,82} in R⁸³, with alleles centered and missing values converted to the genotype mean across all individuals. Ancestral proportions were estimated with fivefold cross-validation using ADMIXTURE v.1.3.0⁸⁴. Individuals with nuclear DNA corresponding to *A. glacialis* or potential hybrids were removed using VCFtools v.0.1.13, leaving 12,305 SNPs for analysis (also removing SNPs with a minor allele frequency < 0.05; Table 2).

In order to detect biased regions of the genome, PCA was performed on the data subset consisting of only *B. saida* individuals (n = 511 individuals and 12,305 loci). Several clusters lacking geographic specific associations were noted, and allele loadings were inspected for patterns indicative of inversions (i.e. loadings aggregated to a specific region). Demultiplexed reads were also used to identify genomic regions associated with the sex of individuals. RADSex⁸⁵ was used to sort unaligned raw reads and determine their presence/absence in each individual, allowing the identification of sex-biased reads. We used forward (R1) reads as input for RADSex along with sex identity of 255 individuals (144 females, 111 males) to identify reads significantly biased toward one sex (min depth = 5x, with Bonferroni correction). The sorted reads were also mapped to the Atlantic cod reference genome (GadMor3; GCF_902167405) to identify the genomic region(s) of sex-biased reads. These patterns were further confirmed by screening genomic regions for linkage disequilibrium, calculated using plink v1.9⁸⁶. For linkage disequilibrium, r^2 values were averaged at the site level, pairing sites at a distance of 100 kbp–5 Mbp.

Inversions and sex-linked regions were removed from the dataset (chromosome 5/LR633947.1, positions 3,622,906–13,963,306; chromosome 9/LR633951.1, positions 7,350,621–11,315,580 and 17,597,431–23,417,670; chromosome 13/LR633955.1, positions 11,053,431–27,975,038), leaving 11,233 SNPs. Functions of the removed regions were investigated using Ensembl gene codes and Gene Ontology (GO) terms^{87,88}. First, the above biased gene regions and corresponding annotations were extracted from the GadMor3 reference genome (GCF_902167405.1) using NCBI's genome data viewer and gene search engine. Enrichment analyses were performed using Ensembl gene identifiers and the *G. morhua* organism option and g:SCS threshold correction for multiple testing in g:Profiler⁸⁹. Molecular function GO terms from the g:Profiler output were then reduced using Multidimensional Scaling using the default SimRel similarity measure in REVIGO v.1.8.1 (REduce + VISualize Gene Ontology⁹⁰). The scatter plot output was further edited in R.

In order to interpret population structure in *B. saida*, PCA, DAPC, and ADMIXTURE were performed on the subsequent dataset with inversions and sex-linked regions removed (n = 11,233 loci). PCA was performed as above using the *glPca* function of *adegenet*, while ADMIXTURE was performed from $k = 1-10$. For DAPC, the *find.clusters* function of *adegenet* was used to define groups, retaining 400 PCs (80% of variation) and three clusters based on Bayesian Information Criterion (Fig. S1A,B). Allele loadings for the discriminant functions were then inspected on axes 1 and 2.

Further suspect clustering patterns were noted, so SNPs were screened for potential outliers, i.e., SNPs explaining larger than expected proportions of population structure. We first used PCAdapt v4.3.5⁹¹ and an individual based-method to identify outlier SNPs in the two first PC axes. We also used Bayescan v2.1⁹² to identify SNPs with higher than expected F_{ST} values among the five marine ecoregions. For both approaches, SNPs with a q -value < 0.05 were identified as outliers, and outliers based on Bonferroni correction were also explored in the PCAdapt results. Plink was also used to remove linked sites with an r^2 value > 0.05 and within 10 Mbp, and with an r^2 value > 0.1 across chromosomes. A final PCA was performed as above on the resulting no-hybrids, no-inversions, no-sex-linked segment, and no-outlier dataset ($n = 7133$ loci), which was used to interpret population structure in *B. saida*.

Heterozygosity was calculated for each individual using the `-het` function in VCFtools. Pairwise ecoregion F_{ST} was estimated using the `glfst.pop` function of `dartR`^{93,94}, using 1000 bootstrap replicates to test for significant departures from 0. Kruskal–Wallis tests were used to test for significant differences in heterozygosity among ecoregions using the `stats` R package, followed by Dunn's post hoc tests to determine what ecoregions, if any, were significantly different from each other using the `dunn.test` R package⁹⁵. Plots were constructed using `ggplot2`⁹⁶. Length and weight measurements were tested for significant differences across regions using the `agricolae` package in R⁹⁷. Departures across regions were tested using Tukey's Post Hoc test.

Data availability

All sequence data can be accessed via NCBI's BioProject PRJNA1062734. Workflow commands are available via Github: https://github.com/GenomicsMLI-DFO/Boreogadus_ddRADseq_2022.

Received: 30 April 2024; Accepted: 8 August 2024

Published online: 20 August 2024

References

- Lønne, O. J. & Gulliksen, B. Size, age and diet of polar cod, *Boreogadus saida* (Lepechin 1773), in ice covered waters. *Pol. Biol.* **9**, 187–191 (1989).
- Majewski, A. R. *et al.* Distribution and diet of demersal Arctic Cod, *Boreogadus saida*, in relation to habitat characteristics in the Canadian Beaufort Sea. *Pol. Biol.* **39**, 1087–1098 (2016).
- Bouchard, C. & Fortier, L. The importance of *Calanus glacialis* for the feeding success of young polar cod: A circumpolar synthesis. *Pol. Biol.* **43**, 1095–1107 (2020).
- Maes, S. M. *et al.* Comparative visual and DNA-based diet assessment extends the prey spectrum of polar cod *Boreogadus saida*. *Mar. Ecol. Prog. Ser.* **698**, 139–154 (2022).
- Welch, H. E. *et al.* Energy flow through the marine ecosystem of the Lancaster Sound Region, Arctic Canada. *Arctic* **45**, 343–357 (1992).
- Darnis, G. *et al.* Current state and trends in Canadian Arctic marine ecosystems: II: Heterotrophic food web, pelagic-benthic coupling, and biodiversity. *Clim. Change* **115**, 179–205 (2012).
- Mueter, F. J., Nahrgang, J., Nelson, R. J. & Berge, J. The ecology of gadid fishes in the circumpolar Arctic with a special emphasis on the polar cod (*Boreogadus saida*). *Pol. Biol.* **39**, 961–967 (2016).
- Finley, K. J. & Gibb, E. J. Summer diet of the narwhal (*Monodon monoceros*) in Pond Inlet, northern Baffin Island. *Can. J. Zool.* **60**, 3353–3363 (1982).
- Loseto, L. L. *et al.* Summer diet of beluga whales inferred by fatty acid analysis of the eastern Beaufort food web. *J. Exp. Mar. Biol. Ecol.* **374**, 12–18 (2009).
- Bengtsson, O., Lydersen, C., Kovacs, K. M. & Lindstöm, U. Ringed seal (*Pusa hispida*) diet on the west coast of Spitsbergen, Svalbard, Norway: during a time of ecosystem change. *Pol. Biol.* **43**, 773–788 (2022).
- Melnikov, I. A. & Chernova, N. V. Characteristics of under-ice swarming of polar cod *Boreogadus saida* (Galidae) in the Central Arctic Ocean. *J. Ichthyol.* **53**, 7–15 (2013).
- Geoffroy, M. *et al.* The circumpolar impacts of climate change and anthropogenic stressors on Arctic cod (*Boreogadus saida*) and its ecosystem. *Elementa Sci. Anthropocene* **11**, 1–44 (2023).
- Chen, L., DeVries, A. L. & Cheng, C.-H.C. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc. Nat. Acad. Sci.* **94**, 3817–3822 (1997).
- Meredith, M. *et al.* I: Polar regions, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate (2019).
- Cote, D. *et al.* Forecasted shifts in thermal habitat for cod species in the Northwest Atlantic and Eastern Canadian Arctic. *Front. Mar. Sci.* **8**, 764072 (2021).
- Fevolden, S.-E., Martinez, I. & Christiansen, J. S. RAPD and scnDNA analyses of polar cod, *Boreogadus saida* (Pisces, Gadidae), in the North Atlantic. *Sarsia* **84**, 99–103 (1999).
- Pálsson, S., Kállman, T., Paulsen, J. & Árnason, E. An assessment of mitochondrial variation in Arctic gadoids. *Polar Biol.* **32**, 471–479 (2008).
- Maes, S. M. *et al.* High gene flow in polar cod (*Boreogadus saida*) from West-Svalbard and the Eurasian Basin. *J. Fish Biol.* **99**, 49–60 (2021).
- Quintela, M. *et al.* Distinct genetic clustering in the weakly differentiated polar cod, *Boreogadus saida*, 1774 from East Siberian Sea to Svalbard. *Pol. Biol.* **44**, 1711–1724 (2021).
- Wilson, R. E. *et al.* Micro-geographic population genetic structure within Arctic cod (*Boreogadus saida*) in Beaufort Sea Alaska. *ICES J. Mar. Sci.* **76**, 1713–1721 (2019).
- Emelianova, O. R., Grigorov, I. V., Orlov, A. M. & Orlova, S. Y. Polymorphism of mtDNA gene *cyt b* of the Chukchi Sea polar cod, *Boreogadus saida* (Galidae, Gadiformes). *Deep-Sea Res. Pt. II* **206**, 105212 (2022).
- Gordeeva, N. V. & Mishin, A. V. Population genetic diversity of Arctic cod (*Boreogadus saida*) of Russian Arctic seas. *J. Ichthyol.* **59**, 246–254 (2019).
- Madsen, M. L., Nelson, R. J., Fevolden, S.-E., Christiansen, J. S. & Præbel, K. Population genetic analysis of Euro-Arctic polar cod *Boreogadus saida* suggest fjord and oceanic structuring. *Pol. Biol.* **39**, 969–980 (2016).
- Hill, G. Transcriptomic basis for differentiation of fjord and offshore *Boreogadus saida* (Polar cod) populations. UiT The Arctic University of Norway (2021). PhD Thesis.
- Nelson, R. J. *et al.* Circumpolar genetic population structure of polar cod. *Boreogadus saida*. *Pol. Biol.* **43**, 951–961 (2020).
- Seehausen, O., Takimoto, G., Roy, D. & Jokela, J. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* **17**, 30–44 (2008).
- Abbott, R. *et al.* Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246 (2013).

28. Árnason, E. & Halldórsdóttir, K. Codweb: whole-genome sequencing uncovers extensive reticulations fueling adaptation among Atlantic, Arctic, and Pacific gadids. *Sci. Adv.* **5**, eaat8788 (2019).
29. Carr, S. M., Kivlichan, D. S., Pepin, P. & Crutcher, D. C. Molecular systematics of gadid fishes: Implications for the biogeographic origins of Pacific species. *Can. J. Zool.* **77**, 19–26 (1999).
30. Boulva, J. Comparison of the Arctic cod (*Boreogadus saida*), the polar cod (*Arctogadus glacialis*), and the toothed cod (*A. borisovi*). *Canadian Atlantic Fisheries Scientific Advisory Committee Research Document*. **79/50** (1979).
31. Bouchard, C., Mollard, S., Suzuki, K., Robert, D. & Fortier, L. Contrasting the early life histories of sympatric Arctic gadids *Boreogadus saida* and *Arctogadus glacialis* in the Canadian Beaufort Sea. *Pol. Biol.* **39**, 1005–1022 (2016).
32. Wilson, R. E. *et al.* Low levels of hybridization between sympatric cold-water-adapted Arctic cod and Polar cod in the Beaufort Sea confirm genetic distinctiveness. *Arctic Sci.* **8**, 1082–1093 (2022).
33. Matschiner, M. *et al.* Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* **6**, 469–481 (2022).
34. Sodeland, M. *et al.* “Islands of divergence” in the Atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biol. Evol.* **8**, 1012–1022 (2016).
35. Berg, P. R. *et al.* Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci. Rep.* **6**, 23246 (2016).
36. Einarsson, A. W., Árnason, E., & Halldórsdóttir, K. Genomic evidence for large inversions in the polar cod *Boreogadus saida* at similar chromosomal locations as found in Atlantic cod *Gadus morhua*. *Preprint* (2022).
37. Møller, P. R., Jordan, A. D., Gravlund, P. & Steffensen, J. F. Phylogenetic position of the cryopelagic codfish genus *Arctogadus* Drjagin, 1932 based on partial mitochondrial cytochrome b sequences. *Pol. Biol.* **25**, 342–349 (2002).
38. Coulson, M. W., Marshall, H. D. & Carr, S. M. Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome* **49**, 1115–1130 (2006).
39. Nelson, R. J. *et al.* Microsatellite loci for genetic analysis of the arctic gadids *Boreogadus saida* and *Arctogadus glacialis*. *Conserv. Genet. Resour.* **5**, 445–448 (2013).
40. Currat, M., Ruedi, M., Petit, R. J. & Excoffier, L. The hidden side of invasions: massive introgression by local genes. *Evolution* **62**, 1908–1920 (2008).
41. Miller, G. H. *et al.* Temperature and precipitation history of the Arctic. *Quat. Sci. Rev.* **29**, 1679–1715 (2010).
42. Baalsrud, H. T. *et al.* Evolution of hemoglobin genes in codfishes influenced by ocean depth. *Sci. Rep.* **7**, 7956 (2017).
43. Stivison, L. S., Hoehn, K. B. & Noor, M. A. F. Effects of inversions on within- and between-species recombination and divergence. *Genome Biol. Evol.* **3**, 830–841 (2011).
44. Puncher, G. N. *et al.* Chromosomal inversions in the Atlantic cod genome: implications for management of Canada’s Northern cod stock. *Fish. Res.* **216**, 29–40 (2019).
45. Bradbury, I. R. *et al.* Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proc. R. Soc. B* **277**, 3725–3734 (2010).
46. Pettersson, M. E. *et al.* A chromosome-level assembly of the Atlantic herring genome: Detection of a supergene and other signals of selection. *Genome Res.* **29**, 1919–1928 (2024).
47. Cayuela, H. *et al.* Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Mol. Ecol.* **29**, 2379–2398 (2020).
48. Devlin, R. H. & Nagahama, Y. Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture* **208**, 191–364 (2002).
49. Star, B. *et al.* Genomic characterization of the Atlantic cod sex-locus. *Sci. Rep.* **6**, 31235 (2016).
50. Kitano, J. & Peichel, C. L. Turnover of sex chromosomes and speciation in fishes. *Environ. Biol. Fish.* **94**, 549–558 (2012).
51. Heule, C., Salzburger, W. & Böhne, A. Genetics of sexual development: An evolutionary playground for fish. *Genetics* **196**, 579–591 (2014).
52. Taher, A. E., Ronco, F., Matschiner, M., Salzburger, W. & Böhne, A. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. *Sci. Adv.* **7**, eabe8215 (2021).
53. Hardy, S. M. *et al.* Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools. *Mar. Biodivers.* **41**, 195–210 (2011).
54. Foltz, D. W. *et al.* Global population divergence of the sea star *Hippasteria phrygiana* corresponds to the onset of the last glacial period of the Pleistocene. *Mar. Biol.* **160**, 1285–1296 (2013).
55. Carr, C. M., Hardy, S. M., Brown, T. M., Macdonald, T. A. & Hebert, P. D. N. A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. *PLoS ONE* **6**, e22232 (2011).
56. Addison, J. A. & Hart, M. W. Colonization, dispersal, and hybridization influence phylogeography of North Atlantic sea urchins (*Strongylocentrotus droebachiensis*). *Evolution* **59**, 532–543 (2005).
57. Laughinghouse, H. D. IV. *et al.* Evolution of the Northern Rockweed, *Fucus distichus*, in a regime of glacial cycling: Implications for Benthic Algal Phylogenetics. *PLoS ONE* **10**, e0143795 (2015).
58. Neiva, J. *et al.* Glacial vicariance drives phylogeographic diversification in the amphiboreal kelp *Saccharina latissima*. *Sci. Rep.* **8**, 1112 (2018).
59. Bringloe, T. T., Saunders, G. W. & Verbruggen, H. Unique biodiversity in Arctic marine forests is shaped by diverse recolonization pathways and far northern glacial refugia. *Proc. Natl. Acad. Sci. USA* **117**, 22590–22596 (2020).
60. Wares, J. P. & Cunningham, C. W. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55**, 2455–2469 (2001).
61. Maggs, C. A. *et al.* Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89**, S108–S122 (2008).
62. Bringloe, T. T. & Saunders, G. W. Mitochondrial DNA sequence data reveal the origins of postglacial marine macroalgae in the Northwest Atlantic. *Mar. Ecol. Prog. Ser.* **589**, 45–58 (2018).
63. Bigg, G. R. *et al.* Ice-age survival of Atlantic cod: Agreement between paleoecology models and genetics. *Proc. Biol. Sci.* **275**, 163–172 (2008).
64. Assis, J., Araújo, M. B. & Serrão, E. A. Projected climate changes threaten ancient refugia of kelp forests in the North Atlantic. *Glob. Change Biol.* **24**, e55–e66 (2018).
65. Bringloe, T. T. *et al.* Whole genome population structure of North Atlantic kelp confirms high-latitude glacial refugia. *Mol. Ecol.* **31**, 6473–6488 (2022).
66. Hauser, L. & Carvalho, G. R. Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. *Fish. Fish.* **9**, 333–362 (2008).
67. Clucas, G. V., Lou, R. N., Therkildsen, N. O. & Kovach, A. I. Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evol. Appl.* **12**, 1971–1987 (2019).
68. Jones, F. C. *et al.* The genomic basis for adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
69. Luu, K., Bazin, E. & Blum, M. G. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* **17**, 67–77 (2017).
70. Fuentes-Pardo, A. P. & Ruzzante, D. E. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol. Ecol.* **26**, 5369–5406 (2017).
71. Lowry, D. B. *et al.* Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Mol. Ecol. Resour.* **17**, 366–369 (2017).

72. McKinney, G. J., Larson, W. A., Seeb, L. W. & Seeb, J. E. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: Comment on Breaking RAD by Lowry et al. (2016). *Mol. Ecol. Resour.* **17**, 356–361 (2017).
73. DFO. Development of a framework and principles for the biogeographic classification of Canadian marine areas. DFO Canadian Science Advisory Secretariat Science Advisory Report (2009/056).
74. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of high-density genetic maps for Barley and Wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7**, e32253 (2012).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
76. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
77. Rochette, N. C., Rivera-Colón, A. G. & Catchen, J. M. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* **28**, 4737–4754 (2019).
78. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
79. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
80. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
81. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
82. Jombart, T. & Ahmed, I. adegenet 1.3–1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
83. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2022).
84. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
85. Feron, R. et al. RADSex: A computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol. Ecol. Resour.* **21**, 1715–1731 (2021).
86. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
87. Ashburner, M. et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
88. The Gene Ontology Consortium. The gene ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
89. Kolberg, L. et al. g:Profiler - interoperable web service for function enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–212 (2023).
90. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
91. Privé, F., Luu, K., Vilhjálmsson, B. J. & Blum, M. G. B. Performing highly efficient genome scans for local adaptation with R package PCADAPT version 4. *Mol. Biol. Evol.* **37**, 2153–2154 (2020).
92. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993 (2008).
93. Gruber, B., Unmack, P. J., Berry, O. F. & Georges, A. dartR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* **18**, 691–699 (2018).
94. Mijangos, J. L., Berry, O. F., Pacioni, C. & Georges, A. dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. *Methods Ecol. Evol.* **13**, 2150–2158 (2022).
95. Dinno, A. Dunn's test of multiple comparisons using rank sums. R package version 1.3.5.
96. Wickham, H. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.
97. de Mendiburu, F., & Yaseen, M. Agricolae: Statistical procedures for agricultural research. R package version 1.4.0.
98. Massicotte, P., & South, P. rnaturlaearth: World map data from natural earth. R package version 1.0.1.9.

Acknowledgements

Field work for this research was conducted aboard the CCGS Amundsen, the CCGS Teleost, the CCGS Leim, the Aqviq, the Katsheshuk II vessels and was supported by their crews as well as technical staff from the Marine Institute of Memorial University, Fisheries and Oceans Canada (DFO), Amundsen Science, the Nunatsiavut Government, the Université du Québec à Rimouski, and the Saguenay St-Lawrence Marine Park, Northern Shrimp Research Foundation. We thank Wojciech Walkusz (DFO) for organizing and managing the sample collection in the Hudson Strait region. The technical expertise for the laboratory work of Grégoire Cortial, Gabriel Bardaxoglou, Cloé Lepage, and Sara Khan (DFO) was also crucial to the success of this project.

Author contributions

D.C., M.G., M.J.R., and G.P. conceived the study and provided samples. J.H. and D.R. provided samples. G.P. generated the data. T.T.B. and A.B. analysed the data. T.T.B. wrote the manuscript. T.T.B., G.P., A.B., D.C., R.M.J., M.G., J.H., and D.R. revised the manuscript.

Funding

Funding was provided by Fisheries and Ocean Canada's ISECOLD, ISICLE, and Competitive Research Science Funds projects, and grants from Amundsen Science, the Canada Foundation for Innovation (CFI), ArcticNet a Network of Centres of Excellence Canada. The Natural Sciences and Engineering Research Council (NSERC) and Crown-Indigenous Relations and Northern Affairs Canada (CIRNAC). Some of the data presented herein were collected by the Canadian research icebreaker CCGS Amundsen and made available by the Amundsen Science program, which was supported by the Canada Foundation for Innovation and Natural Sciences and Engineering Research Council of Canada. Research permits for work aboard the Amundsen were obtained through ArcticNet (<https://amundsen-science.com/expeditions/past-expeditions/>).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-69782-w>.

Correspondence and requests for materials should be addressed to T.T.B. or G.J.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2024