ORIGINAL ARTICLE

Evolutionary Applications
Open Access
WILEY

Genomic differentiation in Pacific cod using Pool-Seq

Ingrid Spies[1] | Carolyn Tarpey[2] | Trond Kristiansen[3] | Mary Fisher[2] | Sean Rohan[4] | Lorenz Hauser[1]

[1]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, Seattle, Washington, USA

[2]School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington, USA

[3]Farallon Institute, Inc., Petaluma, California, USA

[4]Resource Assessment and Conservation Engineering Division, Alaska Fisheries Science Center, Seattle, Washington, USA

Correspondence
Ingrid Spies, Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA.
Email: ingrid.spies@noaa.gov

Abstract

Patterns of genetic differentiation across the genome can provide insight into selective forces driving adaptation. We used pooled whole genome sequencing, gene annotation, and environmental covariates to evaluate patterns of genomic differentiation and to investigate mechanisms responsible for divergence among proximate Pacific cod (*Gadus macrocephalus*) populations from the Bering Sea and Aleutian Islands and more distant Washington Coast cod. Samples were taken from eight spawning locations, three of which were replicated to estimate consistency in allele frequency estimation. A kernel smoothing moving weighted average of relative divergence (F_{ST}) identified 11 genomic islands of differentiation between the Aleutian Islands and Bering Sea samples. In some islands of differentiation, there was also elevated absolute divergence (d_{XY}) and evidence for selection, despite proximity and potential for gene flow. Similar levels of absolute divergence (d_{XY}) but roughly double the relative divergence (F_{ST}) were observed between the distant Bering Sea and Washington Coast samples. Islands of differentiation were much smaller than the four large inversions among Atlantic cod ecotypes. Islands of differentiation between the Bering Sea and Aleutian Island were associated with SNPs from five vision system genes, which can be associated with feeding, predator avoidance, orientation, and socialization. We hypothesize that islands of differentiation between Pacific cod from the Bering Sea and Aleutian Islands provide evidence for adaptive differentiation despite gene flow in this commercially important marine species.

KEYWORDS
genomic differentiation, genomics, local adaptation, Pacific cod, Pool-Seq, selection, selection–migration balance

1 | INTRODUCTION

Comparison among genomes of allopatric and parapatric populations may show heterogeneous levels of differentiation, depending on the level of divergence (Nosil et al., 2009), and may provide insight into the evolution and maintenance of adaptive divergence despite gene flow (Graham et al., 2018; Monnahan et al., 2015). Studies of nonmodel species indicate that selection at specific genome regions may allow local adaption even in the presence of gene flow (Barth et al., 2017; Dennenmoser et al., 2017). Regions of significantly

Ingrid Spies and Carolyn Tarpey shared first authorship.

elevated differentiation have been referred to as 'genomic islands of differentiation' (Wolf & Ellegren, 2017) or 'genomic islands of speciation' (Turner et al., 2005). We use the phrase 'genomic islands of differentiation' because here we focus on local adaptation rather than speciation and these regions are not necessarily associated with speciation.

Natural selection in different environments may influence the emergence and persistence of islands of differentiation, resulting from a complex range of mechanisms (Yeaman, 2013). Islands of differentiation can arise from chromosomal rearrangements, selection with hitchhiking, and background selection. Low recombination in regions of genomic rearrangements may facilitate the evolution of islands of differentiation, because genetic linkage reduces recombination of adaptive combinations of alleles (Via, 2012; Yeaman, 2013). Chromosomal rearrangements, or inversions, restrict recombination in heterokaryotypes and can be adaptive or neutral (Lotterhos, 2019; Noor & Bennett, 2009). Background selection can result in increased levels of differentiation due to selection against deleterious alleles at linked loci (Charlesworth et al., 1993). Selection with hitchhiking occurs when regions of the genome subject to strong selection promote secondary changes in gene frequencies in closely linked regions (Smith & Haigh, 1974; Wolf & Ellegren, 2017).

The capacity for local adaptation rests on a balance between the strength of selection and gene flow, or migration–selection balance (Graham et al., 2018). During the emergence of local adaptation among connected populations, the majority of the genome is thought to be subject to gene flow and few divergent clusters are expected (Nosil et al., 2009; Via, 2012). Stronger selection is thought to cause islands of differentiation to increase in size (Graham et al., 2018; Nosil et al., 2009; Yeaman & Whitlock, 2011). Background selection against strongly deleterious mutations can also lead to nonneutral distortions in the allele frequency spectrum (Charlesworth et al., 1993; Cvijović et al., 2018; Good et al., 2014). In a region in which the deleterious mutation rate is elevated, there will be an excess of rare alleles because selection cannot act instantaneously (Cvijović et al., 2018; Good et al., 2014). As such, distinguishing background selection from positive selection and selection with hitchhiking might not be possible because they may both lead to an excess of rare alleles in the site frequency spectrum (Cvijović et al., 2018). Nevertheless, a comparison of relative (F_{ST}) and absolute (d_{XY}) differentiation between populations with Tajima's D and nucleotide diversity within populations may provide insights into the relative importance of selection and migration in shaping patterns of genome variability.

Selection in connected populations is particularly relevant in marine species supporting commercial fisheries to understand the biological basis behind established management units. Genetic population structure has been documented among parapatric groups of Pacific cod that spawn in ecologically and biophysically different regions across their range (Drinan et al., 2018; Spies, 2012; Spies et al., 2020). In Alaska, Pacific cod are managed in three management units corresponding to the Gulf of Alaska, Aleutian Islands, and eastern Bering Sea. A general pattern of isolation by distance among

spawning groups of Pacific cod (Cunningham et al., 2009; Drinan et al., 2018; Spies, 2012) is punctuated by locally pronounced genetic differentiation between the Aleutian Islands and eastern Bering Sea (Spies, 2012) and between the western and eastern Gulf of Alaska (Drinan et al., 2018; Spies et al., 2021). However, divergence across the genome between Aleutian Islands and eastern Bering Sea cod has not been examined. Such data may further elucidate levels of connectivity as well as mechanisms of local adaptation in these areas, which support the largest cod fishery in the United States, with an average annual gross value of $379 million between 2016 and 2020 (Fissel et al., 2020).

Examination of adaptive divergence in Pacific cod is also merited given differences observed in physical environment and prey of cod in these two regions. The Bering Sea contains a prominent and broad (500 km) continental shelf adjacent to a deep-sea basin that stretches from the Alaska Peninsula to the Bering Strait. In contrast, the Aleutian Islands consist of a narrow steep shelf formed by an 1800 km long chain of volcanic mountaintops (Logerwell et al., 2005). While Pacific cod occurs in the western and eastern Bering Sea (O'Leary et al., 2022), we focus on Pacific cod in the United States management region of eastern Bering Sea (Stevenson & Lauth, 2019). The eastern Bering Sea (EBS) continental shelf consists of three depth regions: the inner (0–50 m depth), middle (50–100 m), and outer domain (100–180 m; Stabeno et al., 2012). The extent of sea ice in the Bering Sea during winter varies interannually, but has declined in recent years (Grebmeier et al., 2006; Stevenson & Lauth, 2019). The timing of the maximum ice extent and its retreat affects the species composition across a range of taxa from the zooplankton community to higher trophic level predators (Stevenson & Lauth, 2019). The physical environment of the Aleutian Islands is dynamic; islands are separated by passes that allow the transfer of water from the North Pacific Ocean northward and the water quality and species distributions change along the chain in response to shifting currents and passes (Hunt Jr & Stabeno, 2005; Logerwell et al., 2005). For example, there is an abrupt shift in the physical environment and species composition at Samalga Pass (169° W), an area that divides warmer, nutrient-poor coastal waters of the Alaska Coastal Current to the east and the colder, nutrient-rich oceanic Alaskan Stream to the west (Hunt Jr & Stabeno, 2005).

These distinct physical environments, and the resulting shifts in ecological communities, drive diet differences between Pacific cod in the Aleutian Islands and EBS. Walleye pollock (*Gadus chalcogrammus*) are a major diet component of Pacific cod in the EBS (26%), but in the Aleutians, Atka mackerel (*Pleurogrammus monopterygius*) and sculpins are the predominant fish prey (15% each), while pollock comprise only 5% (Aydin et al., 2007). Snow and tanner crab (*Chionoecetes opilio* and *C. bairdi*) make up 9% of cod diets in the EBS, but less than 3% in the Aleutian Islands. In contrast, squids comprise over 6% of cod diets in the AI but are a negligible proportion of diets in the EBS. Myctophids (family Myctophidae) are also found only in cod diets in the Aleutian Islands, but not in the EBS even though they are present in the diets of other EBS fish species (Aydin et al., 2007; Lang & Livingston, 1996).

Lessons from the congeneric Atlantic cod (*Gadus morhua*) may help inform the genomics of Pacific cod; in Atlantic cod, several ecotypes are known, including Northeast Arctic cod (NEAC) and the Norwegian coastal cod (NCC). The genomic basis of these ecotypes appears to be related to inversions on linkage groups 1, 2, 7, and 12 (Árnason & Halldórsdóttir, 2019; Barth et al., 2017; Hemmer-Hansen et al., 2013). Genes within these inversions appear to be responsible for maintaining selective differences related to habitats and life history (Barth et al., 2017; Kirubakaran et al., 2016). It is currently unknown if such inversions also drive population structure in Pacific cod.

Some knowledge of the demographic history of Pacific cod is valuable for interpreting genomic differences. Pacific cod are believed to have originated from an Atlantic cod ancestor that moved into the Pacific Ocean 3.8 million years ago (Árnason & Halldórsdóttir, 2019). During Pleistocene glaciations, water levels fluctuated and the Bering Sea shelf was exposed during several time periods, forming the Bering land bridge as recently as 30,000–18,000 years ago (Elias et al., 1996). Glaciers extended from the Alaska Peninsula to the Aleutian chain through the southern extent of Alaska to northwestern Canada southward to Puget Sound (Batchelor et al., 2019), during which time cod likely migrated southward. Cod recolonized the North Pacific as glaciers receded, as recently as 14–15 kyr before present (Mann & Hamilton, 1995; Menounos et al., 2009). Analysis of the putative zona pellucida gene is consistent with this scenario, as the present-day southern populations are more closely related to Atlantic cod and more northerly populations are more derived at this locus (Spies et al., 2021).

The aim of this study was to compare genomic patterns of differentiation among proximate versus distant Pacific cod populations from ecologically distinct regions and evaluate regions of the genome that may be subject to local adaptation. We applied Pool-Seq to Pacific cod from the ecologically different Aleutian Islands and Bering Sea among which gene flow may occur, and to allopatric groups (EBS and Washington Coast), between which migration has not been observed despite thousands of tagging records (Bryan et al., 2021; Rand et al., 2014; Shimada & Kimura, 1994). These parapatric and allopatric groups were selected to provide examples across several levels of population divergence (Stankowski & Ravinet, 2021; Wolf & Ellegren, 2017). At the time this study was initiated, individual-based whole genome sequencing was not a cost-effective option. The power to identify departures from panmixia in marine fish with large effective population sizes increases with the number of genetic markers and individuals (Kalinowski, 2005; Vendrami et al., 2017; Waples & Gaggiotti, 2006). Therefore, we selected Pool-Seq, which offers the capacity to sequence the genome more thoroughly than sequence capture (Harvey et al., 2016) or reduced representation sequencing (e.g., RAD-seq, Andrews et al., 2016), and at a lower cost than individual-based sequencing. Pool-Seq also provides the combined contributions of individuals in the pool and has been shown to provide robust and reliable allele frequencies (Anand et al., 2016). The disadvantages to Pool-Seq stem from the need to calculate allele frequencies over all pooled samples, which masks individual genotypes and precludes its utility for tests of linkage disequilibrium or population assignment (Dorant et al., 2019). In addition, unequal contributions of individuals in the pool can increase the error of allele frequency estimates (Rode et al., 2018). Nonetheless, Pool-Seq has been used to successfully examine adaptive divergence and genetic population structure in a range of species (Dennenmoser et al., 2017; Dorant et al., 2019; Han et al., 2020; Kurland et al., 2019). This study was also well suited to Pool-Seq because previous studies have defined genetic population structure and spawning units, and all cod were presumed to be in or near spawning stage (Cunningham et al., 2009; Spies et al., 2020).

A primary hypothesis was that genomic divergence between eastern Bering Sea and Aleutian Islands cod would manifest as a limited number of highly differentiated clusters if selection were sufficiently high, because gene flow is thought to be limited (Spies, 2012; Spies & Punt, 2015; Yeaman & Whitlock, 2011). We also expected that temperature would play a role in local adaptation, as winter bottom temperatures in the Aleutian Islands are roughly 2°C warmer in the winter than the Bering Sea, and cod are adapted to a relatively narrow temperature range at all life stages (Barbeaux et al., 2020; Hurst et al., 2010; Laurel et al., 2008). For example, in the Gulf of Alaska, Pacific cod has suffered recent declines in abundance due to the impact of marine heatwaves, and increased ocean temperatures are predicted under climate change (Barbeaux et al., 2020; Capotondi et al., 2012). Specifically, we hypothesized that (1) genetic differentiation between parapatric groups which may be subject to gene flow would vary across their genomes, while differentiation would be both more pronounced and more widespread across the genomes of allopatric groups, (2) regions of elevated divergence may provide clues to the selective mechanisms responsible for divergence among the parapatric Aleutian Islands and Bering Sea, and (3) genes underlying local adaptation of parapatric groups may be correlated with the environmental factors to which adaptation occurred, with temperature most likely to play a role. To address the first two hypotheses, we examined the parapatric and allopatric sets of samples for F_{ST} outlier regions and calculated nucleotide diversity (π), d_{XY}, and Tajima's D to inform the level of gene flow and potential mechanisms shaping divergent regions. To address the final hypothesis, we identified four environmental covariates considered to impact Pacific cod: salinity, bottom temperature, surface chlorophyll, and current velocity, and quantified correlations between environmental covariates and allele frequencies. In addition, we screened for annotated genes within islands of differentiation among the Aleutian Islands and Bering Sea and measured light transmission to the seafloor in the Bering Sea and Aleutian Islands through bottom optical depth.

2 | METHODS

2.1 | Sample preparation and sequencing

We analyzed Pacific cod fin clips from eight known spawning sites from the Aleutian Islands, EBS, Gulf of Alaska, and Washington Coast collected during February–May between 2003 and 2017

(Table 1, Figure 1). Studies of genetic stock structure in Pacific cod are typically limited to spawning fish because populations may intermingle outside the spawning season (Cunningham et al., 2009; Drinan et al., 2018; Rand et al., 2014). While Pacific cod typically spawn between January and April (Neidetcher et al., 2014), the Zhemchug sample, taken May 9, 2017, was considered close enough to that time window to be considered a late spawning sample. Individual gonadal condition was not recorded, but all samples were from adult cod that were taken from known spawning areas during spawning season and likely to be reproductively mature and in spawning condition. DNA from tissues stored in 95% ethanol from 88 to 96 individuals from each unique collection location and year (Table 1) were extracted using the DNeasy 96-well Blood & Tissue Kits (Qiagen Inc.). The extracted DNA was quantified using Quant-iT PicoGreen for double-stranded DNA (Invitrogen) and visualized on a 1% agarose gel.

DNA pools were constructed of 43–48 individuals with an equimolar concentration of DNA from each individual and a total of 750 ng of DNA per pool. In three spawning areas, we included DNA replicates to examine the accuracy of SNP calling. We included a temporal replicate sampled off Kodiak Island in 2003 (pool A) and 2005 (pool B) that were not genetically different at 11 microsatellites, with a statistically nonsignificant $F_{ST} = -0.0006$ (Cunningham et al., 2009). Two pools from Pervenets Canyon consisted of 95 individuals taken at a single location with longline gear were included to determine the effect of lower versus higher DNA concentration. These samples were split by DNA quantity into pool A (48 samples with the highest concentration of DNA, 47–72 ng/μl) and pool B (47 samples with the lowest DNA concentration, 25–47 ng/μl) to determine whether DNA concentration would affect Pool-Seq results. Finally, two pools (A & B) of the Washington Coast samples contained the same 43 individuals, with five additional samples in pool B included to test sensitivity to 10% additional samples. These eight spawning locations with three replicates resulted in a total of eleven pools that were sequenced.

DNA libraries were prepared by and sequenced at the Northwest Genomics Center at the University of Washington (https://www.nwgs.gs.washington.edu/). Pooled samples were sheared using a Covaris LE220 focused ultrasonicator targeting an insert size of 350–380 base pairs (bp). The resulting sheared DNA was cleaned with Agencourt AMPure XP beads to remove sample impurities prior to library construction. A two-sided AMPure cleanup was then performed in order to further restrict the fragment sizes to the desired range. End-repair, A-tailing, and ligation were performed as directed by the KAPA Hyper Prep Kit (KR0961 v1.14) protocol without amplification. A final AMPure cleanup was performed after ligation in order to remove excess adapter dimers from the library. All library construction steps were automated on a Janus Automated workstation (Perkin Elmer, Inc.).

Final library concentration prior to sequencing was determined by triplicate qPCR using the KAPA Library Quantification Kit (KK4824), and molecular weight distributions were verified using an Agilent Bioanalyzer (Agilent Technologies). Samples were sequenced on a HiSeq X using Illumina's HiSeq X Ten Reagent Kit (v2.5). Cluster generation was performed on a cBot modified for use with the HiSeq X flow cells, and flow cells were loaded onto the HiSeq X machine for sequencing.

2.2 | Bioinformatics pipeline

The pipeline we used deviated from the standardized pipeline to account for pooled sequences, the nonmodel species, and the alignment of our sequence data to the reference genome of a related species (Atlantic cod). Within GATK (v4), we used the GenomeAnalysisToolkit v. 4.1.2.0 to identify a set of single-nucleotide polymorphisms (SNPs) and to estimate allele frequencies for each pool. We designed and built a variant calling pipeline, including filtering for minimum and maximum read depth,

TABLE 1 Collection location of samples used in each pool, month and year of collection, latitude, longitude, and the number of individuals in collection (*n*)

Location	Month/year	Latitude	Longitude	*n*	Superpool	RD1	RD2
Adak Island	Mar. 2006	51.32	−176.23	48	AI	71	63
Kiska Island	Mar. 2005	51.89	177.47	48	AI	60	50
Kodiak Island	Mar. 2003	57.92	−152.30	46	EBS	70	62
Kodiak Island	Mar. 2005	57.92	−152.30	48	EBS	79	70
Near Island	Feb. 2005	52.60	174.40	47	AI	60	51
Pervenets Canyon (A)	Mar. 2016	59.21	−177.15	48	EBS	98	87
Pervenets Canyon (B)	Mar. 2016	59.21	−177.15	47	EBS	88	79
Pribilof Island	Apr. 2017	57.78	−172.12	48	EBS	84	76
Washington Coast (A)	Feb. 2005	48.27	−125.00	43	WA	87	81
Washington Coast (B)	Feb. 2005	48.27	−125.00	48	WA	81	66
Zhemchug Canyon	May 2017	58.27	−173.80	48	EBS	71	63

Note: Duplicate pools from Pervenets Canyon and Washington State were assigned to "A" and "B" collections because they were collected during the same year. Pools were grouped into three superpools, abbreviated as follows: Aleutian Islands (AI), Bering Sea (EBS), and Washington coast (WA). Mean read depth prior to filtering (RD1) and postfiltering (RD2) are presented, rounded to the nearest whole read.

FIGURE 1 Spawning locations included in this study were named after nearby landmarks. There were three collections from the Bering Sea (Pervenets, Zhemchug, and Pribilof), three from the Aleutian Islands (Near Island, Kiska Island, and Adak Island), as well as a collection from Kodiak Island and the Washington Coast.

missing data, and variant SNPs using the GATK best practices workflow (*germline short variant discovery, SNPs + Indels*, Poplin et al., 2018; Van der Auwera et al., 2013). Base calls generated on the HiSeq X instrument (RTA 2.7.6) were demultiplexed and converted to unmapped BAM files using the Picard programs ExtractIlluminaBarcodes and IlluminaBasecallsToSam in GATK. BAM files for each pool were then aligned to the Atlantic cod reference genome assembly, GadMor2 (Tørresen et al., 2017) using BWA-MEM (Burrows-Wheeler Aligner; v0.7.10). SNPs were identified using maximum likelihood expectation (MLE) to estimate the allele frequency rather than the observed read count (allele frequency) method in GATK v4. We assumed exactly 20 alleles per pool at each locus (ploidy), so that the minimum allele frequency for a variable SNP was 0.05 and the allele counts for each SNP summed to 20. This option was selected due to computational constraints and because maximum likelihood methodology was considered a good choice for allele frequency estimation. A single set of minimum and maximum read depth filters were applied to data that aligned to the GadMor2 linkage groups and scaffolds. A second set of filters was applied to remove SNPs exceeding set ranges for read depth per pool. The minimum read depth was 20× for the linkage group and scaffold data, and the maximum read depth was 2% of the total reads of each pool. The second filtering step removed any loci with missing data in any of the pools and retained only biallelic and variant SNPs. The full pipeline is described in more detail in the Appendix S1.

2.3 | Assessing patterns of genetic differentiation

Lin's (1989, 2000) concordance correlation coefficient, ρ_c, was calculated to measure relative agreement and validate results between the estimated allele frequencies among duplicated pools, using the R package *epiR* (Stevenson, 2020). More closely related pools were expected to have higher correlation. This statistic, which ranges from 0 to 1, with larger values indicative of higher concordance, compares allele frequency calls between two pools to determine deviation from perfect concordance:

$$\rho_c = \frac{2\rho\sigma_X\sigma_Y}{\sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2},$$

where μ_X and μ_Y are the means of the allele frequencies in pools X and Y, and σ_X and σ_Y are the corresponding variances. The concordance correlation coefficient, ρ_c, was calculated for the three duplicated pools (Washington, Pervenets, and Kodiak) to determine the relative correlation among them, as well as all other pools for comparison. The allele frequency matrix with the full filtered set of 1,944,780 variant SNPs was used for this analysis.

F_{ST} and d_{XY} were used to screen genomic regions of elevated divergence. F_{ST} is a relative measure of genetic differentiation (Wright, 1931) whose expectation is inversely correlated with genetic variation in one or both populations under comparison. Elevated pairwise F_{ST} may be due to lower within-group genetic variation, rather than divergence among groups (Cruickshank & Hahn, 2014). Therefore, an absolute measure of divergence, d_{XY}, whose expectation is independent of genetic variation was also used. The statistic d_{XY} summarizes the average number of nucleotide differences between pools of samples using an equation for unphased data: $d_{XY} = (p_X{}^*q_Y) + (q_X{}^*p_Y)$, where p_X is the proportion of the reference allele in pool X and p_Y is the proportion of the reference allele in pool Y, $q_X = 1 - p_X$, $q_Y = 1 - p_Y$ (Dennenmoser et al., 2017; Schirrmann et al., 2018). Pairwise F_{ST} was calculated globally between all pools and per-SNP between the three superpools described below (EBS, AI, and WA). F_{ST} was calculated in PoPoolation2 (Kofler et al., 2011) using Nei's (1973) formula,

$$F_{ST} = \frac{\hat{H}_T - \hat{H}_S}{\hat{H}_T},$$

where $\hat{H}_T$ is the expected heterozygosity in combined pools/superpools and $\hat{H}_S$ is the expected heterozygosity in each pool/superpool (Hivert et al., 2018). The components $\hat{H}_T$ and $\hat{H}_S$ were calculated using Hivert et al. (2018), equations 17 and 18. While this measure of F_{ST} is considered biased (Hivert et al., 2018), the bias was consistent among pools because we retained similar read depth and sample size per pool (Table 1).

Pooled data were grouped into superpools according to the previously identified patterns of genetic differentiation between cod spawning in those regions (Drinan et al., 2018; Spies et al., 2020) and patterns identified in the principal components analysis (PCA) described below. Collections from Kodiak Island were included in the Bering Sea superpool based on these criteria. Three superpools

were formed: the eastern Bering Sea superpool ("EBS"), the Aleutian Islands superpool ("AI"), and the Washington Coast ("WA") superpool (Table 1). The Washington coast samples were designated as a third superpool due to their genetic distinctiveness. Allele frequencies for superpools were calculated as the arithmetic mean of the pool allele frequencies.

2.4 | Sliding window analysis and islands of differentiation

A Gaussian kernel smoothing moving weighted average (KSMWA) approach was used to assess genome-scale patterns of differentiation based on all identified loci. This analysis required pairwise comparisons between two sets of data; therefore, KSMWA was performed twice, on two pairs of superpools: (1) the Bering Sea superpool compared with the Washington coast superpool (EBS-WA), and (2) the Bering Sea superpool compared with the Aleutian Islands superpool (EBS-AI). This allowed for a relatively proximate, parapatric comparison (EBS-AI) and a spatially distant, allopatric comparison (EBS-WA) of genomic differentiation. Weighted average F_{ST} and d_{XY} was computed for all 29,825 consecutive windows along the genome for each superpool comparison.

We adapted the methodology of Hohenlohe et al. (2010) and Waters et al. (2018) for the KSMWA to accommodate pooled whole genome sequence data. The window was shifted along the genome by a step size that allowed the windows to overlap. We evaluated a range of step sizes from 7 to 50 kb and choices of window size sigma (σ) from 10 to 80 kb to balance smoothing the variation in F_{ST}, the number of SNPs per window, and reducing noise along the window. The contribution of the SNP at position p to the region average was weighted by the Gaussian function $exp(-[p-c]^2/2\sigma^2)$, where c is the window center and σ is a parameter for window size, so that SNPs in the center of the window received the highest weight according to a normal ($\mu = 0$, σ) distribution. Sliding window analysis used a step size of 20 kb and $\sigma = 30$ kb, as described in the results section. The statistic d_{XY} was reported as the weighted average over all 180 k SNPs in each window, whereas weighted average F_{ST} was only calculated for variable positions, as it is undefined for invariable positions.

The significance of each sliding window's empirical Gaussian weighted average F_{ST} was determined by comparison to a set of bootstrap distributions generated from the entire genome (EBS-AI or EBS-WA). To reduce computational load, we generated a reference set of 2000 bootstrap distributions for each of the two KSMWA we performed. Each bootstrap distribution consisted of 10^6 random draws with replacement of size n from the empirical per-SNP F_{ST} estimates across the genome for that pairwise comparison, for n in 1–2000. One million bootstrap replicates were selected because this number sufficiently sampled the range of values of F_{ST}. A Gaussian weighted average was calculated for each of the 10^6 draws, arranged in random order. p-values for each window were estimated as the proportion of bootstrap replicates higher or equal than the

empirical weighted mean F_{ST}'s. Statistics were adjusted for multiple tests using the Benjamini–Hochberg false-discovery rate method (Benjamini & Hochberg, 1995).

Windows with kernel-weighted F_{ST} averages considered significant based on the false-discovery rate are hereafter referred to as F_{ST} outlier windows. Sliding window d_{XY} values were calculated in an analogous fashion.

Outlier regions with at least four consecutive windows with weighted mean $F_{ST} > 0.03$ were considered islands of differentiation. The $F_{ST} > 0.03$ threshold was selected because it contained the upper 10% of weighted average F_{ST} values, provided an inflection point above which F_{ST} values reached an asymptote, and was generally a threshold above which outlier windows appeared prominent (Figure S1). Outlier regions of interest were defined by the center points of full windows that shifted by step sizes of 20,000 bp. Therefore, the size of outlier regions was allowed to be smaller than full window sizes (180,000 bp). The coefficient of variation of the number of significant F_{ST} outlier windows per linkage group was also quantified for the EBS-WA and the EBS-AI superpool comparisons.

Weighted mean d_{XY} and F_{ST}, and mean nucleotide diversity (π), and Tajima's D (T_D, Tajima, 1989) were summarized over all F_{ST} outlier and nonoutlier windows for each KSMWA comparison (EBS-AI and EBS-WA) and for each EBS-AI island of differentiation. Islands of differentiation for EBS-WA comparisons were not included due to the much larger scale of differentiation among these superpools. For islands of differentiation between EBS-AI superpools, we quantified the presence of windows with d_{XY} greater than the upper 5% quantile over all windows and windows with d_{XY} greater than the genome-wide average. In early stages of population differentiation, d_{XY} may not be a reliable indicator of reduced gene flow (Cheng et al., 2021; Dennenmoser et al., 2017; Feulner et al., 2015). Therefore, for the EBS-AI superpool comparison, Tajimas's D provided information on whether selection or reduced gene flow may be implicated in regions of increased divergence, and nucleotide diversity was used to assess genetic variation (Nei, 1986). Reduced nucleotide diversity (π) can result from strong directional selection, as well as background selection and selective sweeps (Booker & Keightley, 2018). Nucleotide diversity (π) was estimated in each of the 29,825 windows for each superpool as:

$$\pi = \sum \frac{2j(C - j)}{C(C - 1)} / n,$$

where j is the derived allele count (j = the reference allele frequency $\times$ ploidy), C is the ploidy, or (20) $\times$ the number of pools in each superpool, summation takes place over all sites (including monomorphic sites) in each window, and n is the size of the window (180,000 bp). Tajima's D was also estimated in each sliding window as:

$$D = \frac{\pi - \theta_W}{\sqrt{var(\pi - \theta_W)}},$$

where $\theta_W = \frac{S_n}{\sum_{i=1}^{C} \frac{1}{i}}$. Here, S_n = the number of SNPs in each sliding window. Note that each superpool had the same number of SNPs per window, but the number of SNPs per window varied.

Principal components analysis (PCA) was performed using all 1,944,780 SNPs in the full dataset using the R package *PCAdapt* (Luu et al., 2017) to visualize the genetic relationships among pools. The optimum number of principal components retained for analysis was determined based on Cattell's rule (Cattell, 1966).

We used *fst-sliding.pl* in PoPoolation2 to estimate the global pairwise comparison of F_{ST} among all pools using the following options [*pool-size 20, min-count 1, min-coverage 1, max-coverage 500, window-size 5,800,000, step-size 0*]. For this calculation, the position for each SNP was renumbered sequentially and a window size that exceeded the total number of SNPs (*5,800,000*) was specified over this large region to calculate F_{ST} over the entire genome. The pool size was 20 alleles for each pool, reflecting a ploidy of 20. The option *min count 1* represents the minimum count of the minor allele and specifies that at least one pool must have an alternate allele to identify the position as a SNP, and SNPs were identified considering all populations simultaneously. Minimum and maximum coverage criteria applied to each pool were 1 and 500 reads, to ensure that no data were excluded. Secondly, PoPoolation2 *fst-sliding.pl* was used to estimate the pairwise per-SNP F_{ST} between superpools using the following options [*--pool-size 120:60:40 –min-count 1 –min-coverage 1 –max-coverage 500 –window-size 1 --step-size 1*]. The pool size for that analysis reflected the number of pools per superpool multiplied by 20 (a ploidy of 120 for the EBS superpool, 60 for the Aleutian Islands superpool, and 40 for the Washington superpool).

2.5 | Gene annotations

The 11 EBS-AI islands of differentiation were aligned with all four GadMor2 annotation files (https://osf.io/4qsdw/; Tørresen et al., 2017) using the Bioconductor package *GenomicRanges* (Lawrence et al., 2013, Table 2). Only Bering Sea versus Aleutian Islands KSMWA comparisons were annotated, in line with the goals and hypotheses described above. Genes with known function were recorded if they contained one or more SNPs from the EBS-AI dataset. The number of SNPs and mean F_{ST} over all SNPs within these proteins were also recorded, as well as a summary of gene function (www.uniprot.org).

2.6 | Environmental correlation

We selected four environmental covariates (salinity [psu], bottom temperature [°C], chlorophyll [mg/m³], and current velocity [m/s]) that are considered to be of impact to Pacific cod. Salinity, and particularly the interaction between salinity and temperature, affects growth, egg fertilization, and regulation of fish growth hormone during exposure to stress (Bœuf & Payan, 2001; Deane & Woo, 2009).

Chlorophyll has been widely used as a proxy for marine productivity because it is indicative of phytoplankton biomass, and zooplankton, the prey of larval fish, forage on phytoplankton (Boyce et al., 2014; Hughes et al., 2018; Kristiansen et al., 2011). Recruitment in Pacific cod is more correlated with flow along and across the Bering Slope than other groundfish species, indicating that current velocity and direction are significant factors in Pacific cod early life history (Vestfals et al., 2014).

These four covariates were downloaded from the Copernicus Marine Environment Monitoring Services (CMEMS, https://resources.marine.copernicus.eu/), and values for each region were averaged over the cod spawning months January–April over the years in our study, 2003–2017. CMEMS is a global ocean eddy-resolving (1/12° horizontal resolution, 50 vertical levels) reanalysis covering the period 1993 until present day. The reanalysis assimilates existing satellite and ocean vertical observations of temperature, salinity, and sea level and represents the state-of-the-art in ocean models (Lellouche et al., 2018). Model simulations were extracted using the nearest grid point of the sample locations that was not on land with a "find nearest grid point" routine, which shifted the location by 4.5 km or less. The exception was the Adak Island sample for which the depth was outside the biological range of Pacific cod (3220 m); therefore, the sample location latitude was shifted slightly from 51.32°N to 51.62°N (33 km). Chlorophyll was extracted for surface values, but all other covariates were taken from the ocean bottom. Scalar current velocity was calculated from the east (u) and north (v) velocity components (m/s), $w = \sqrt{u^2 + v^2}$.

We used BayPass v.2.2 (Gautier, 2015) to identify SNPs that were correlated to environmental variables within a Bayesian framework. The BayPass analysis was performed separately for the EBS-WA and EBS-AI comparisons, though for each comparison all pools were considered separately, not grouped as superpools. For the BayPass analyses, the Pribilof pool was excluded because it displayed a skewed allele frequency distribution with more SNPs with low minor allele frequencies than any other pool. The Pribilof sample was retained for all other analyses because the elevated number of low-frequency SNPs had no noticeable effect on the allelic composition in superpools. The core BayPass model was run three times, with default parameters, to test for convergence. The auxiliary covariate model of BayPass associates allele frequencies with an environmental covariate while accounting for population structure. This was run with default settings, using the scaled mean values of the four covariates. To facilitate parallel computing, we used the function pooldata.subset() from the R package *poolfstat* (Hivert et al., 2018) to create 20 sets of 97,239 SNPs from our dataset. The auxiliary model was run three times for each covariate on the 20 sets of SNPs. The resulting per-SNP X^TX (a Bayesian measure of the deviation of population allele frequencies from expected values) and $\log_{10}$(Bayes factors) statistics were compared over replicated runs for each covariate by calculating Pearson's correlation coefficient among the three possible comparisons for each covariate, to ensure convergence before averaging values over runs. The average of the three scaled covariance matrices was used in the auxiliary covariate model. We considered all SNPs

TABLE 2 Number of outlier and nonoutlier windows, mean F_{ST}, d_{XY}, Tajima's D, and nucleotide diversity (π) for eastern Bering Sea (EBS) vs. Washington Coast (WA) comparisons (rows 1 and 2) and EBS vs. Aleutian Islands (AI), rows 3 and 4). Subsequent rows (3–15) provide these statistics for the 11 islands of differentiation between the EBS and AI superpools.

Row	Comparison EBS vs. WA	# win.	F_{ST}	d_{XY}	Tajima's D D_{EBS}/D_{WA}	π $\pi \times 10^{-3}$ (EBS/WA)	CHL 14,814	SAL 8848	TMP 7427	VEL 13,036
1	Nonoutlier	25,478	0.032	0.00054	−0.385/−1.179	0.547/0.530				
2	Outlier	4347	0.045	0.00059	−0.285/−1.230	0.573/0.525				
	EBS vs. AI	**# win.**	**F_{ST}**	**d_{XY} (n sig. outliers)**	**D_{EBS} (n)/D_{AI} (n)**	**$\pi \times 10^{-3}$ (EBS/AI)**	**CHL**	**SAL**	**TMP**	**VEL**
3	Nonoutlier	28,996	0.017	0.00054	−0.370/−0.909	0.552/0.529				
4	Outlier	829	0.026	0.00053	−0.409/−1.000	0.529/0.492				
5	LG02_1	8	0.031	0.00058(0)	1.446(0)/1.071(0)	0.499(0)/0.522(0)	1	12	1	8
6	LG06_1	10	0.025	0.00051(0)	−0.556(0)/−1.438(4)	0.548(0)/0.450(0)	19	10	12	27
7	LG08_1	6	0.067	0.00029(0)	−1.314(3)/−1.293(0)	0.184(5)/0.221(1)	1	29	3	38
8	LG08_2	3	0.023	0.00042(0)	0.596(0)/0.260(0)	0.368(0)/0.384(0)	0	0	2	0
9	LG08_3	19	0.026	0.00029(0)	−0.265(0)/−0.449(0)	0.252(2)/0.276(0)	1	10	3	16
10	LG12_1	47	0.049	0.00038(0)	−0.522(3)/−0.730(0)	0.313(9)/0.339(7)	35	31	1	135
11	LG14_1	27	0.026	0.00027(0)	−1.539(25)/−2.010(20)	0.272(2)/0.250(5)	8	11	5	18
12	LG16_1	12	0.030	0.00068(2)	0.008(0)/−0.650(0)	0.663(0)/0.622(0)	0	3	2	1
13	LG18_1	13	0.025	0.00055(0)	0.639(0)/−0.001(0)	0.527(0)/0.507(0)	1	1	3	10
14	LG19_1	5	0.044	0.00095(0)	−0.310(0)/−1.225(0)	0.890(0)/0.741(0)	3	0	1	7
15	LG22_1	6	0.033	0.00069(3)	0.370(0)/−0.085(0)	0.621(0)/0.626(0)	1	6	2	10

Note: Details on the 11 islands of differentiation also include the size of the island in number of windows (# win.), the number of windows in which d_{XY} was greater than the upper 5% quantile over all windows, in parentheses, mean Tajima's D (D_{EBS}/D_{AI}) and nucleotide diversity (π) from the EBS-AI superpools followed by the number of windows in which Tajima's D or π was lower than the 5% quantile over all EBS-AI windows. Gray-shaded rows indicate outlier F_{ST} regions containing windows with d_{XY} greater than the genome-wide mean d_{XY}. Finally, the last four columns provide the number of SNPs for which $\log_{10}$ Bayes factor > 30 with regard to chlorophyll (CHL), salinity (SAL), temperature (TMP), and current velocity (VEL), presented over the entire genome for EBS-WA and within islands of differentiation for EBS-AI.

TABLE 3 Pairwise F_{ST} between samples for all data combined, Wash. = Washington, Zhem. = Zhemchug, Perv. = Pervenets

	Near	Kiska	Adak	Perv. (A)	Perv. (B)	Zhem.	Pribilof	Kodiak 2003	Kodiak 2005	Wash. (A)
Near	–									
Kiska	0.0250	–								
Adak	0.0262	0.0261	–							
Perv. (A)	0.0229	0.0228	0.0238	–						
Perv. (B)	0.0236	0.0235	0.0245	0.0181	–					
Zhemchug	0.0249	0.0248	0.0257	0.0197	0.0204	–				
Pribilof	0.0250	0.0248	0.0258	0.0195	0.0202	0.0219	–			
Kodiak 2003	0.0248	0.0245	0.0255	0.0199	0.0204	0.0216	0.0219	–		
Kodiak 2005	0.0243	0.0242	0.0250	0.0193	0.0200	0.0212	0.0215	0.0205	–	
Wash. (A)	0.0377	0.0371	0.0364	0.0317	0.0322	0.0334	0.0333	0.0321	0.0314	–
Wash. (B)	0.0362	0.0358	0.0361	0.0298	0.0303	0.0315	0.0316	0.0301	0.0296	0.0179

with an averaged $\log_{10}$(Bayes factor) in decibans (dB, the corresponding weights of evidence) greater than 30 associated with the covariate to be significant, a conservative threshold under which the alternative hypothesis is 10^{30} times more likely than the null hypothesis (Baldwin-Brown & Long, 2020). We quantified the number of outlier SNPs in F_{ST} outlier windows for each covariate.

Finally, we plotted near-bottom (2–7 m above the sea floor) optical depth for the Aleutian Islands and EBS shelf during summer (June–July) to evaluate differences in water clarity between regions during 2006, 2010, 2012, 2014, 2016, and 2018. Optical depth is a natural logarithmic ratio that characterizes how much downwelling irradiance just below the sea surface reaches a given depth. Optical depth was calculated from 6657 irradiance profiles obtained during NMFS bottom trawl surveys using light-sensitive archival tags connected to the trawl net. Data were collected during years when both summer surveys (Aleutian Islands and EBS) were conducted and light data were collected. Methods for collecting irradiance data and calculating near-bottom optical depth are described in Rohan et al. (2021). To quantify differences among regions, we fit a generalized additive model (GAM) to near-bottom optical depth (response variable) using a penalized cubic regression spline of depth and region (Aleutian Islands or EBS shelf) as predictor variables. For a given depth, higher near-bottom optical depth is associated with lower water clarity (i.e., less light transmission to the seafloor). Every 2.303 units of optical depth correspond with an order of magnitude difference in light transmission.

3 | RESULTS

There were 13,993,143 SNPs in the raw dataset, which included 13,348,645 SNPs that aligned to the 23 GadMor2 linkage groups, 1740 SNPs that aligned to mitochondrial DNA, and 644,498 SNPs to scaffolds (Table A1). Initial average read depth per pool ranged from 60 to 98 (Table 1). Following filtering, the total number of SNPs that aligned to linkage groups was 1,944,780, while 161 SNPs aligned to mtDNA, and 21,546 SNPs aligned to scaffolds. Average read depth

after filtering ranged from 50 to 87 per pool (Table 1). We retained the 1,944,780 SNPs that aligned to linkage groups for downstream analysis, hereafter referred to as the "full dataset".

The concordance correlation coefficient showed that the highest similarity among all pairwise sample comparisons was between Washington Coast Pool A and Pool B ($\rho_c = 0.9912383$), followed by the Pervenets Pool A and Pool B ($\rho_c = 0.9906341$, Table A2). Samples taken in different years off Kodiak Island had the sixth highest concordance correlation coefficient (out of 46 comparisons), preceded by Pervenets Pool A compared with Kodiak Pool B, Zhemchug, and Kodiak Pool A, respectively (Table A2). Estimates of pairwise F_{ST} confirmed that Pool-Seq genotypes were sensitive to the level of similarity among duplicate pools. Pairwise F_{ST} between duplicate pools increased in the inverse order as the ρ_c, with lowest F_{ST} between Washington Coast pools A and B ($F_{ST} = 0.0179$), followed by Pervenets pools, ($F_{ST} = 0.0181$), and finally Kodiak ($F_{ST} = 0.0205$, Table 3). Estimates of F_{ST} between nearly duplicated pools are expected to be near zero, and elevated levels of pairwise F_{ST} among duplicated pools indicated some level of upward bias in F_{ST}. The highest levels of pairwise F_{ST} were observed between the most distant comparisons, Near Island versus Washington Coast replicate pools A and B ($F_{ST} = 0.0377, 0.0362$), whereas more proximate samples exhibited smaller pairwise F_{ST} (e.g., Adak vs. Kiska $F_{ST} = 0.0261$). Therefore, F_{ST} appeared to be a robust measure of genetic differentiation in a relative sense, albeit positively biased.

A PCA generated with all 1,944,780 filtered loci was optimized using only one principal component, which explained 13.3% of the variance, as indicated by Cattell's rule applied to the scree plot (Figure S2). The Aleutian Islands and Bering Sea samples clustered together, with Aleutian Islands pools to the left of Bering Sea pools (Figure 2). Aleutian Islands samples ordered by longitude in the PCA, with the furthest west, Near Islands, to the left, followed by Kiska, then Adak. The Bering Sea pools were closely grouped, and following the longitudinal pattern, Kodiak samples were furthest to the right of the cluster that included Pervenets, Pribilof, and Zhemchug. Washington pools were distant, and some separation was observed between the two Washington pools.

All 1,944,780 SNPs, all LGs
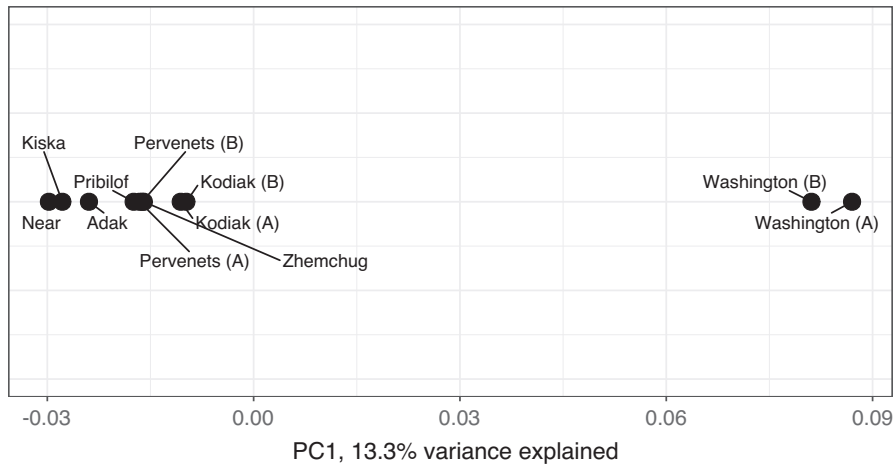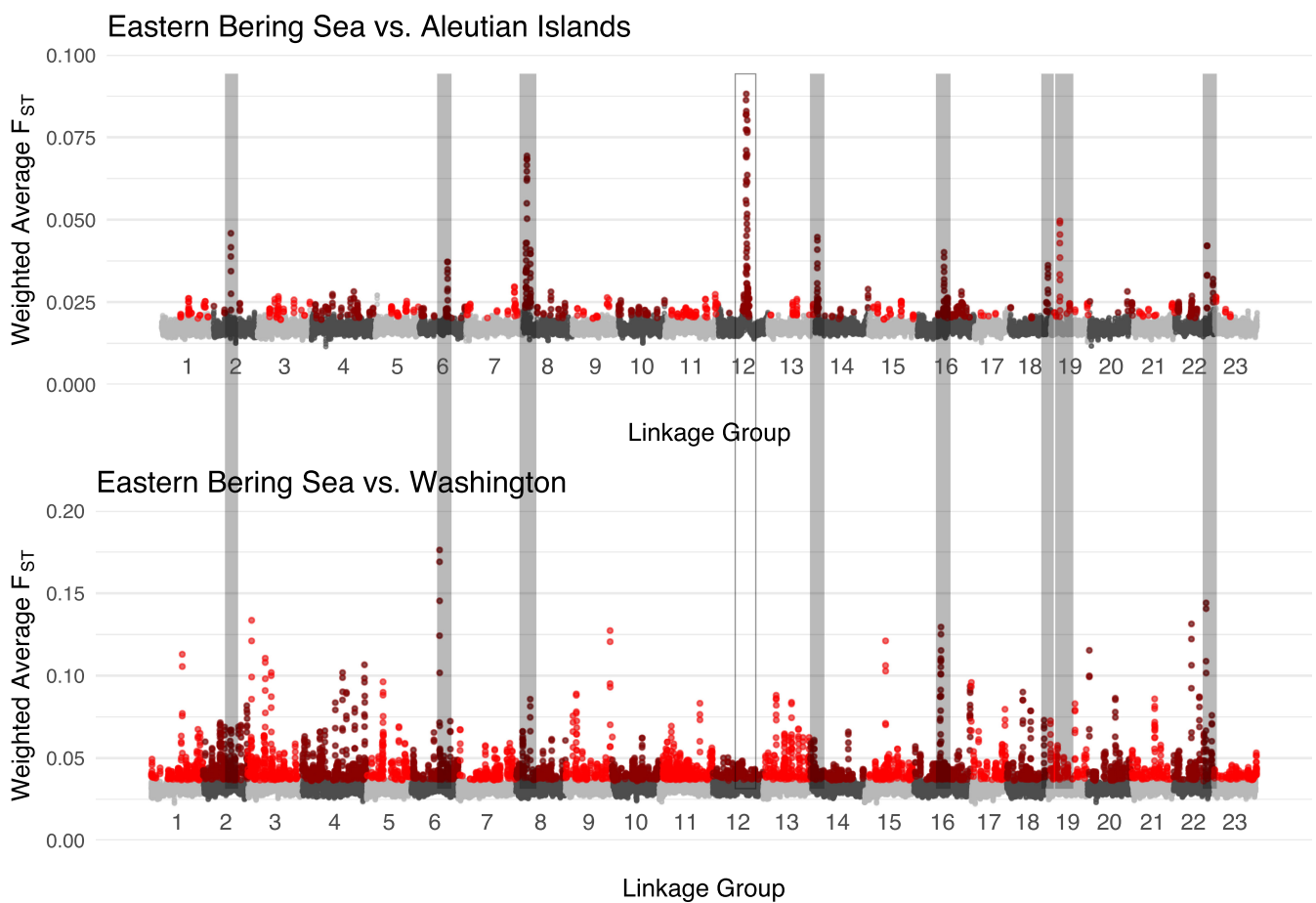
FIGURE 3 Weighted average F_{ST} for all SNPs aligned to GadMor2 in 180 kbp windows that overlap with a step size of 20,000 bp between the Bering Sea and Aleutian Islands superpools (upper panel), and the Bering Sea and Washington Coast superpools (lower panel). F_{ST} outlier windows are colored in red shades. Islands of differentiation identified in Table 2 are highlighted with gray bars. The island on linkage group 12 is not filled because it was not present in the EBS-WA comparison.

## 3.1 | Sliding window analysis and islands of differentiation

For the F_{ST} sliding window analysis, a step size of 20 kb and $\sigma = 30$ kb was selected for the Gaussian kernel smoothing moving weighted

average (KSMWA) because it optimized between SNPs per window, variation in F_{ST}, and reduced noise within windows (Figure S3 and Table A3). This resulted in 29,825 windows.

The pattern and extent of F_{ST} outlier windows differed between the EBS-AI and the EBS-WA superpool comparisons; the EBS-AI

comparison showed fewer outlier windows (829) than the EBS-WA comparison (4347), and the highest EBS-WA outlier F_{ST} windows peaked at $F_{ST} = 0.2$, approximately double that of the highest EBS-AI outlier windows (Figure 3). Furthermore, the coefficient of variation of the number of outlier windows per linkage group was higher for the EBS-AI comparison (CV = 0.026) than for EBS-WA (CV = 0.015), indicating more uneven distribution of outlier regions in the EBS-AI comparison. The distribution of weighted average F_{ST} values for all windows was higher in EBS-WA than in EBS-AI outlier and nonoutlier windows, indicative of a generally higher F_{ST} between allopatric than parapatric Pacific cod (Table 2, Figure 4). In contrast, mean d_{XY} was identical between EBS-WA and EBS-AI comparisons in nonoutlier regions, and slightly higher (0.00059 vs. 0.00053) in the EBS-WA comparisons than in the EBS-AI comparison (Table 2).

Mean Tajima's D was generally negative across all outlier and nonoutlier regions across all EBS-AI and EBS-WA superpool comparisons, but lower in outlier regions versus nonoutlier regions (Table 2). Tajima's D was similar between the EBS among outlier
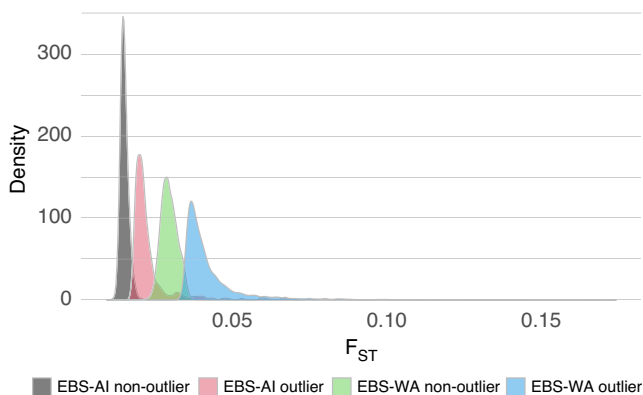
FIGURE 4 Density plot of normalized F_{ST} within and outside F_{ST} outlier windows for the Bering Sea vs. Aleutian Islands comparison (EBS-AI) and the Bering Sea vs. Washington comparison (EBS-WA).

versus nonoutlier regions (EBS-WA comparison), but remained negative. Nucleotide diversity, π, was lower in the EBS and Aleutian Islands outlier than nonoutlier regions, and EBS pools showed higher nucleotide diversity than Washington coast or Aleutian Island pools, particularly on LG06 and LG19 islands of differentiation (Table 2).

The largest EBS-AI F_{ST} outliers were on linkage groups (LG) 8 and 12, and smaller islands of differentiation were present on linkage groups 2, 6, 14, 16, 18, 19 and 22 (Table 2, Table A4, Figure 3). In one of these 11 islands of differentiation (LG16), d_{XY} exceeded the upper 5% quantile of observed over all EBS-AI in 2/10 windows, and in five islands of differentiation (LG02_1, LG12_1, LG16_1, LG18_1, and LG22_1), d_{XY} exceeded the genome-wide average (Table 2). Most islands of differentiation with elevated d_{XY} showed no evidence for significantly reduced Tajima's D, with the exception of LG12_01, the largest F_{ST} outlier window, in which three windows (position 15,300,000–15,520,000) contained reduced Tajima's D. This region of reduced Tajima's D occurred at one end of the window, whereas the region of high d_{XY} occurred at the other end. Reduced Tajima's D was observed at LG06_1 in the Aleutian Islands (in 4/10 windows) and at LG08_1 in the EBS samples (in 3/6 windows). Reduced nuclear diversity was present in both the EBS (5/6) and AI (1/6) at LG08_1, although the results do not rule out background selection in both regions. In LG14_1, there was reduced Tajima's D in 25 out of 27 EBS windows and 20 out of 27 AI windows.

3.2 | Gene annotations

Within the 11 islands of differentiation, we identified 68 regions with similarity to known genes (Table 4 and Table A5, Figure 5). The genes were responsible for a variety of functions, but notably there were five genes related to vision were identified within F_{ST} outlier regions (Table A5): CRB1 (LG8_1), OPN3 (LG08_3), rpe65c (LG12_1), PDE6G (LG18_1), and Gprc5c (LG18_1). PDE6G and Gprc5c were

TABLE 4 Annotated genes within which F_{ST} outlier SNPs regions were found

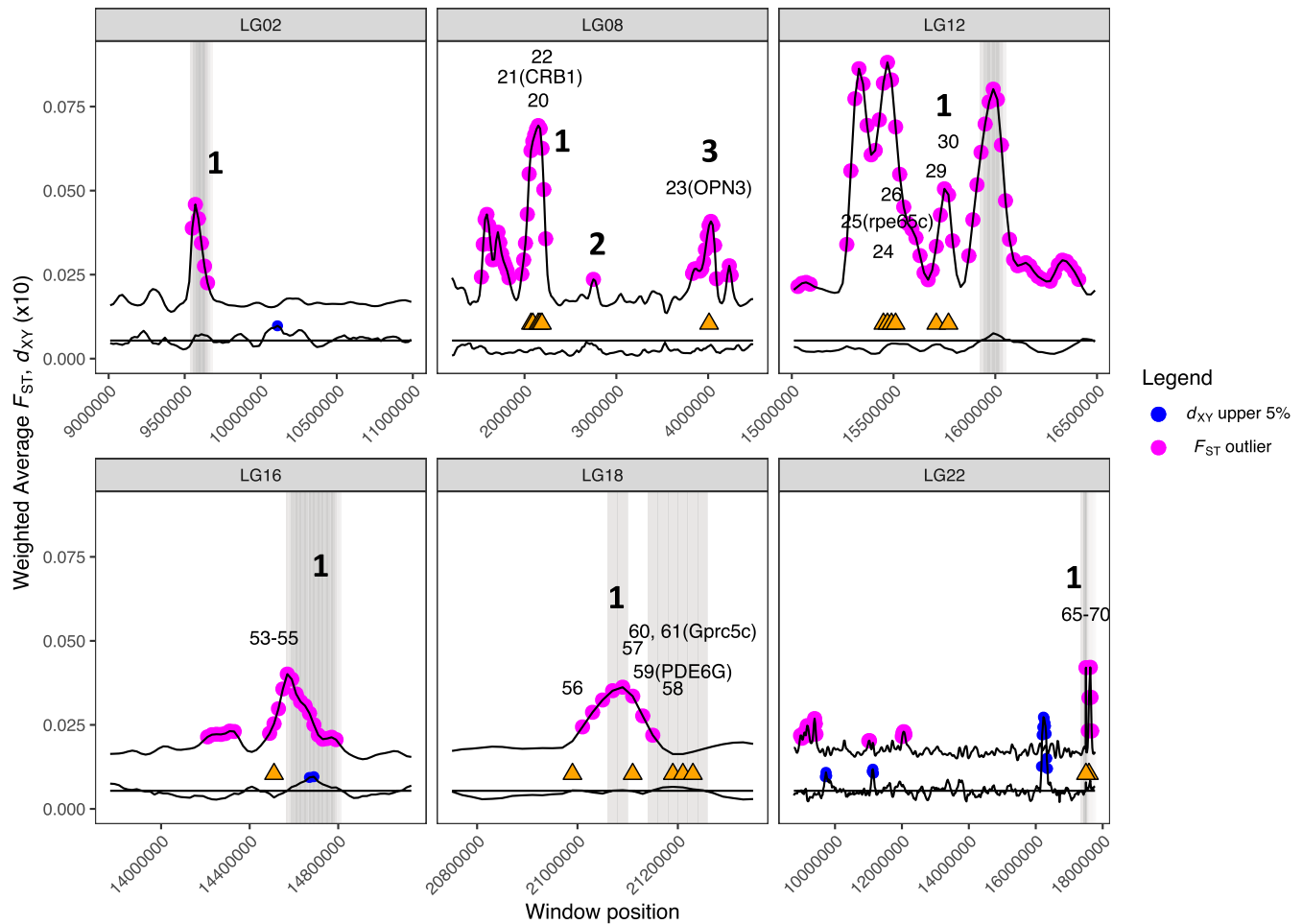| LG, number	Region (Mb)	Gene name	Description
8_01	2.18–2.19	CRB1	Plays a role in photoreceptor morphogenesis in the retina
8_03	4.03–4.04	OPN3	G-protein coupled receptor which selectively activates G proteins via ultraviolet A (UVA) light-mediated activation in the skin. Opsins are light-absorbing genes
12_1	15.47–15.48	rpe65c	Retinal Mueller cells isomerohydrolase. Catalyzes forms of retinoic acid to meet the high demand for the chromophore by cones
18_1	21.22–21.23	PDE6G	Retinal rod rhodopsin-sensitive cGMP participates in transmission, amplification of visual signal
18_1	21.24–21.25	Gprc5c	Retinoic acid-inducible G-protein coupled receptor; RA shapes the developing eye and is essential for normal optic vesicle and anterior segment formation

Note: This table represents a subset of the 68 gene regions; a complete list is in Table A5.

FIGURE 5 Weighted average F_{ST} and d_{XY} (the latter multiplied by 10 for visualization) in the six linkage groups containing F_{ST} outlier regions: linkage groups 2, 8, 12, 16, 18, and 22, labeled corresponding to F_{ST} outlier regions in Table 2. Shaded vertical lines represent outlier regions within which linkage groups with F_{ST} outliers are greater than 0.030 and pairwise d_{XY} is greater than the genome average (0.000539×10, horizontal black line). Blue points represent the upper 5% quantiles for d_{XY}, and pink dots indicate F_{ST} outlier windows. Orange triangles represent locations of annotated genes within F_{ST} outlier windows, which are labeled corresponding to the gene number in Table A5, and vision gene names are listed.

located adjacent to the F_{ST} outlier on LG18 in a region of above-average d_{XY} and were therefore considered potentially relevant.

3.3 | Environmental correlation

Salinity, temperature, current velocity, and chlorophyll patterns differed among regions (Figure S4 and Table A6). The lowest chlorophyll was typically found in Bering Sea sampling areas, Zhemchug and Pervenets. Salinity was highest at the Washington Coast site and lowest off Kodiak. Temperature was highest off the Washington Coast (~8°C) and lowest in Bering Sea sites, 2–3°C. The highest current velocities were found in the Aleutian Islands, with Adak the highest, followed by Near and Kiska. BayPass results were significantly correlated over each set of three independent runs (Table A7); therefore, results were averaged over each set of three runs. Among the EBS-AI comparisons, the most notable increase in Bayes factors in regions identified as islands

of differentiation was found in linkage group 12 (Figure S5). Tabulation of the correlates most strongly correlated with allele frequencies indicated that in LG12_1, velocity had the largest effect, with 135 SNPs with Bayes factors >30 (Table 2). Velocity was the biggest covariate in LG08_1, with 38 SNPs strongly correlated. The second most significant correlate after velocity was salinity, with 29 strongly correlated SNPs in LG08_1 (Table 2). In the EBS-WA BayPass analysis, results were tabulated over all data due to the increased number of high F_{ST} regions. In this comparison, the highest number of strongly correlated Bayes factor SNPs was correlated with salinity (36%) and velocity (32%), versus temperature (20%) and chlorophyll (12%, Table 2).

The EBS shelf has lower water clarity than the Aleutian Islands at ~20–160m bottom depths during summer, based on near-bottom optical depth, meaning that less surface-incident irradiance reaches the seafloor in the EBS than in the Aleutian Islands at those depths (Figure 6a). Based on GAM fitted means, the maximum difference in optical depth between the EBS and Aleutian Islands is at ~70m
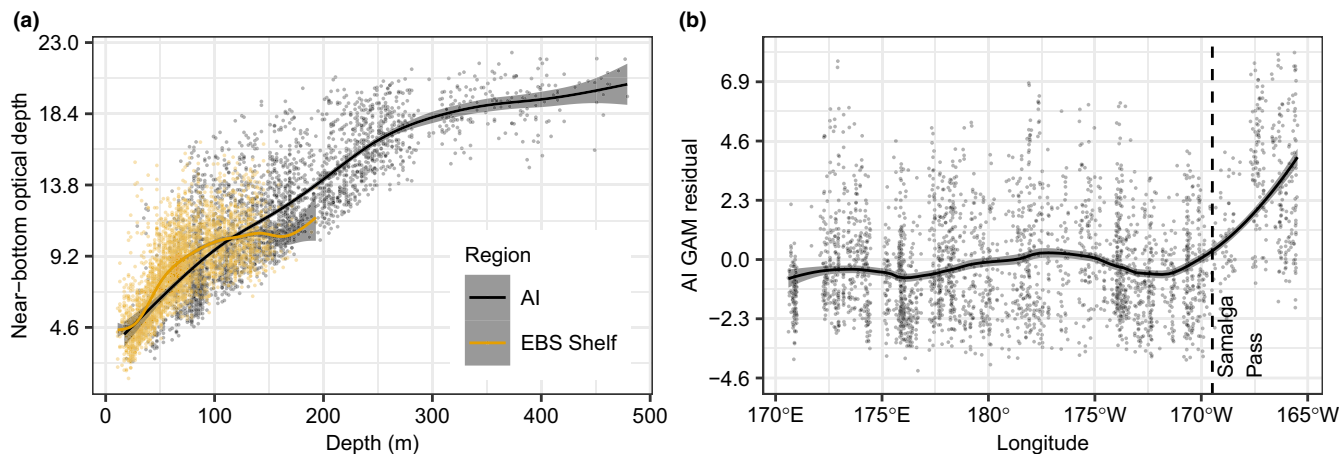
FIGURE 6 Generalized additive model (GAM) fitted between depth (meters) and near-bottom optical depth data in the Aleutian Islands (AI) and eastern Bering Sea from summer (June–July) 2006–2018. Panels show: (a) GAM fitted mean (line)±1 standard error (shading), (b) AI GAM residuals by longitude, mean residual based on LOESS regression (line)±1 standard error (shading), and position of Samalga Pass (~169°29′W). Point and line colors denote region (AI and EBS).

depth, where light transmission to the seafloor in the Aleutian Islands is slightly more than an order of magnitude higher than the EBS (difference in near-bottom optical depth: 2.45). In addition, there is an east–west break in GAM residuals in the Aleutian Islands around Samalga Pass (169°28′W), where bottom optical depth is higher ("darker") than predicted to the east of Samalga Pass than to the west (Figure 6b).

4 | DISCUSSION

While Pacific cod from the Aleutian Islands and EBS are sufficiently genetically differentiated to merit separate management units (Drinan et al., 2018; Spies, 2012; Spies & Punt, 2015), there is little information on the mechanisms driving these differences. We examined how genomic patterns of differentiation and environmental covariates differed among these proximate and distant comparisons, inferred processes by which differentiation occurs across the genome, and examined genetic factors that may play a role in divergence within islands of differentiation between Aleutian Island and Bering Sea cod. A primary finding was a pattern of heterogeneous genomic differentiation between Pacific cod from the Aleutian Islands and EBS. We identified 11 genomic islands of differentiation in nine (of 23) linkage groups between the proximate EBS and Aleutian Islands, and the presence of many more F_{ST} outlier regions between the EBS and Washington coast (Table 2; Figure 3). The EBS-AI islands of differentiation ranged in size from 60,000 to 940,000 bp, and weighted average F_{ST} over these 11 islands ranged from 0.023 to 0.067 (Figure 3; Table A4). The limited number of EBS-AI islands of differentiation interspersed with regions of low F_{ST} were consistent with emergent local adaptation among populations experiencing migration–selection balance, similar to our first hypothesis, although alternative explanations and limitations are discussed below.

High F_{ST} and elevated d_{XY} in five of the 11 genomic islands, particularly on LG12 and LG16 (Figure 5; Table 2), may indicate local adaptation or inversions (Lotterhos, 2019), but the Pool-Seq data preclude further examination of the nature of the outlier regions. There was evidence for selection in the Aleutian Islands region of LG06_1 and in the EBS region of LG08_1. In LG08_1, reduced nucleotide diversity in the EBS and Aleutian Islands could alternatively be a result of background selection or directional selection with hitchhiking in both regions. In LG14_1, reduced Tajima's D and reduced nucleotide diversity provided evidence for background selection, although it is not always possible to distinguish between background selection and directional selection with hitchhiking (Booker & Keightley, 2018). Higher nucleotide diversity in EBS pools could be a result of selection on Aleutian Islands or Washington Coast cod, or due to the disproportionately larger size of the EBS cod stock, as larger population sizes may lead to higher nucleotide diversity (Spies et al., 2020; Subramanian, 2019).

The pairwise F_{ST} between the EBS and Washington coast was roughly double that between the Aleutian Islands and Bering Sea spawning populations (Table 3; Figure 3). This may be consistent with a framework in which gene flow acts to homogenize allele frequencies between EBS and Aleutian Islands at most of the genome, except the few regions affected by selection, while lower gene flow or longer time since divergence between the EBS and Washington Coast allows for F_{ST} to increase. However, the similar absolute divergence (d_{XY}) in nonoutlier regions and slightly smaller d_{XY} in the EBS versus Aleutian Islands outlier regions compared with EBS versus Washington indicate a similar number of nucleotide differences and are consistent with a scenario in which EBS and Aleutian Islands populations began diverging soon after they colonized the region following the recession of the last glacial maximum (Canino et al., 2010).

Most EBS-AI F_{ST} outlier regions were also present in the EBS vs. Washington comparison (Figure 3), indicating that these EBS-AI

outlier regions may be important for EBS cod. Regions of elevated differentiation due to hitchhiking with genes under selection start as smaller regions that grow into wider regions over time and tend to be more pronounced in small populations where linkage is higher (Feder & Nosil, 2010). In contrast, islands of differentiation that are the result of genomic inversions are more likely to emerge in high gene flow species with strong selective pressure (Schaal et al., 2022). The divergent region on linkage group 12, the most predominant island of differentiation with the highest EBS-AI mean F_{ST}, was not present in the EBS-WA comparison, suggesting that this outlier region may have developed relatively recently if EBS and AI populations diverged more recently (Figure 3).

While there is no clear evidence for the mechanism leading to the island of differentiation observed on LG12 in Pacific cod, it is slightly less than 1 megabases in size, much smaller than four large inversions that have been identified in Atlantic cod between Northeast Arctic and Norwegian Coastal ecotypes (Berg et al., 2017; Sodeland et al., 2016). In Atlantic cod inversions on linkage groups 1, 2, 7, and 12 were 17, 5, 9.5, and 13 megabases long, respectively (Kirubakaran et al., 2016; Sodeland et al., 2016). While the island of differentiation in Pacific cod is on the same linkage group as one inversion in Atlantic cod (linkage group 12), it appears located between 15 and 16 mB and is outside the range suggested by previous studies (0.4–0.6 mB—14mB) which aligned to GadMor2 (Barth et al., 2017), and GadMorCeltic (Kirubakaran et al., 2016).

We hypothesized that temperature would be the most significant environmental correlate to genomic variation; however, current velocity and vision stood out as potential factors leading to islands of differentiation but temperature did not. Strong correlation with velocity based on BayPass results indicated that genes located within linkage group 12 outliers were associated with different environments of the EBS and AI, although the precise relationship is not clear (Table 2, Figure S4). Furthermore, the finding of five genes related to vision in outlier regions differentiating the Aleutian Islands and Bering Sea may be an indication of the genes underlying selection. However, identifying targets of selection within islands of differentiation is difficult because the lack of recombination, whether due to inversions or selection with hitchhiking, can link true targets selection with false positives (Berg et al., 2017). Visual systems have been shown to be under strong natural selection in several fish species including Atlantic cod in which differential expression of visual opsins and the rhodopsin *rhI* gene varies by ecotype (Berg et al., 2017; Hofmann & Carleton, 2009; Pampoulie et al., 2015; Valen et al., 2018). In other species, visual systems have evolved in response to environment; for example, the visual system of Midas cichlids (*Amphilophus* cf. *citrinellus*) has rapidly evolved to adapt to the clear water in crater lakes from more ancestral turbid lakes (Torres-Dowdall et al., 2017).

Our analysis of near-bottom optical depth suggests there are depth-associated differences in water clarity among cod habitat in the EBS and Aleutian Islands, as well as differences within the Aleutian Islands chain itself (Figure 6). The break in water clarity conditions at Samalga Pass is presumably due to higher concentrations

of chlorophyll (Mordy et al., 2005), chromophoric dissolved organic matter, and nonalgal particulate in the Alaska Coastal Current than in the more ocean-influenced waters to the west. Specific genes identified in outlier regions for Pacific cod included CRB1, which plays a role in photoreceptor morphogenesis in the retina, two genes related to retinoic acid (rpe65c and Gprc5), a gene related to retinal rod rhodopsin (PDE6G) and an opsin receptor (OPN3), and are consistent with life history strategies and adaptation associated with vision (Table 4). While the association between vision genes and islands of differentiation in Pacific cod is not evidence for causation, research in other teleosts and the commonality with Atlantic cod in which vision genes are differentially expressed among ecotypes provides a basis for further exploration of potential effects of foraging, predator avoidance, orientation, and social behavior (Hofmann & Carleton, 2009; Valen et al., 2018).

The data passed multiple checks, confirming the reliability Pool-Seq methodology for genotyping. Biological replicate pools have been used in other studies too as a quality check for accurate sample contribution and allele frequency calling (Dorant et al., 2019; Guirao-Rico & González, 2021; Hivert et al., 2018). Read depth (50–100) was within the range that is considered sufficient for resolution of the allele frequency spectrum, distinguished evolutionary patterns, and provided sufficient power for Tajima's D (Ferretti et al., 2013), while the number of samples per pool (43–48) was sufficient to reduce experimental bias (Dorant et al., 2019; Kofler et al., 2011). High relative agreement in allele frequencies among the Washington Coast pools was consistent with expectations that pools consisting of the same or similar individuals would produce highly correlated allele frequencies. The PCA also confirmed similarity among duplicate and proximate samples (Figure 2), although allele frequencies differed between the two Washington replicated pools by only adding five individuals (Figure 2). In the case of the Pervenets pools, high correlation indicated that pooling based on concentration of DNA within the observed ranges did not impact the results. F_{ST} estimates decreased with increasing concordance correlation coefficients, providing a second line of evidence for genotype similarities among pools (Table 3). The concordance correlation coefficient between the two Kodiak samples taken in different years was relatively high (6th highest out of 55 comparisons), but was lower than several comparisons between Kodiak and EBS spawning cod. The similarity between Kodiak and Pervenets samples is confirmed by pairwise F_{ST}; F_{ST} between temporal replicates from Kodiak were higher (indicating less similar) than any pairwise comparisons between Kodiak 2003/2005 and Pervenets Pool A or B. We do not completely understand the level of connectivity between western Gulf of Alaska spawning cod and those from the Bering Sea shelf, although the PCA is consistent with isolation by distance observed in previous studies (Figure 2; Cunningham et al., 2009). Similarly, several previous studies found similarity among Kodiak samples and Unimak spawning cod from the EBS (Table A2; Cunningham et al., 2009; Drinan et al., 2018).

While this was the first population genomics study of Pacific cod using whole genome sequencing, the pooled WGS platform posed some drawbacks. Without individual-level data, there was no insight

into linkage disequilibrium, which could provide evidence for genomic inversions. Further, ploidy was limited to 20 per pool, so differences among pools could only be observed in increments of 0.05, and subtle differences among pools may have been lost, although this is unlikely given the large number of SNPs. High replicability between pools provided confidence that the underlying measures of population allele frequencies were accurate. While the patterns observed in the PCA appear sound, the nature of Pool-Seq limited the number of principal components in the data to 10, and the resulting PCA was optimized with only a single principal component that explained 13.3% of the variance. We also acknowledge that our analyses were performed over large window sizes ($\sigma = 30$ kb, step size = 20 kb), which could reduce the possibility of finding the effects of single genes. Another concern was that while alignment to the genome of a congeneric species is routine among nonmodel species, chromosomal rearrangements may not have aligned and could have been missed. Therefore, we anticipate that our approach laid the groundwork for future studies to examine outlier regions in Pacific cod in more detail. Future work that includes low coverage whole genome sequencing (lcWGS) on an individual basis will provide further understanding of the outlier regions identified here, as well as identifying whether outlier regions represent inversions or regions of linked selection (Lou et al., 2021).

Overall, data provided new evidence for heterogeneous differentiation across the genome between spawning populations of cod from the EBS and Aleutian Islands, which we hypothesize may be the result of local adaptation despite some low level of gene flow. Furthermore, we found more extensive levels of relative differentiation but similar levels of absolute divergence among cod from the allopatric EBS vs. Washington coast. These results indicate that the EBS and Aleutian Islands may have started diverging soon after cod recolonized the North Pacific following the last glacial maximum. The presence of 11 islands of differentiation, five of which showed some level of elevated d_{XY}, and evidence for directional selection and background selection on two other islands of differentiation indicate that local adaptation has occurred between cod from these different environments. Results also indicate that these spawning groups are in migration–selection balance and that selection may be strong enough to balance the effects of gene flow. While our study was not designed to provide direct evidence of genes responsible for adaptation, annotated genes suggested that vision may play a role in adaptation to the distinct ocean environments of the Bering Sea and Aleutian Islands. Results build upon previous studies indicative of divergence between EBS and Aleutian Islands Pacific cod.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the Sequence Read Archive (SRA), which is accessible from National Center for Biotechnology Information (NCBI) at https://www.ncbi.nlm.nih.gov/sra/, reference number PRJNA675289 (University of Washington, 2021).

ORCID

Ingrid Spies https://orcid.org/0000-0002-5177-276X
Lorenz Hauser https://orcid.org/0000-0002-9238-654X

REFERENCES

Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., Corrado, L., Martinelli Boneschi, F., D'alfonso, S., & De Bellis, G. (2016). Next generation sequencing of pooled samples: Guideline for variants' filtering. *Scientific Reports*, *6*(1), 1–9.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92.

Árnason, E., & Halldórsdóttir, K. (2019). Codweb: Whole-genome sequencing uncovers extensive reticulations fueling adaptation among Atlantic, Arctic, and Pacific gadids. *Science Advances*, *5*(3), eaat8788.

Aydin, K., Gaichas, S., Ortiz, I., Kinzey, D., & Friday, N. (2007). A comparison of the Bering Sea, Gulf of Alaska, and Aleutian Islands large marine ecosystems through food web modeling. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-178, 298 p.

Baldwin-Brown, J. G., & Long, A. D. (2020). Genomic signatures of local adaptation in clam shrimp (*Eulimnadia texana*) from natural vernal pools. *Genome Biology and Evolution*, *12*(7), 1194–1206.

Barbeaux, S. J., Holsman, K., & Zador, S. (2020). Marine heatwave stress test of ecosystem-based fisheries management in the Gulf of Alaska Pacific cod fishery. *Frontiers in Marine Science*, *7*, 703.

Barth, J., Berg, P., Johsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., Jakobsen, K. S., Johannesson, K., Jorde, P. E., Knutsen, H., Moksnes, P.-O., Star, B., Stenseth, N. C., Svedång, H., Jentoft, S., & André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*, *26*, 4452–4466.

Batchelor, C. L., Margold, M., Krapp, M., Murton, D. K., Dalton, A. S., Gibbard, P. L., Stokes, C. R., Murton, J. B., & Manica, A. (2019). The configuration of northern hemisphere ice sheets through the quaternary. *Nature Communications*, *10*(1), 1–10.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, *57*(1), 289–300.

Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., Jentoft, S., & Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity*, *119*(6), 418–428.

Bœuf, G., & Payan, P. (2001). How should salinity influence fish growth? *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *130*(4), 411–423.

Booker, T. R., & Keightley, P. D. (2018). Understanding the factors that shape patterns of nucleotide diversity in the house mouse genome. *Molecular Biology and Evolution*, *35*(12), 2971–2988.

Boyce, D. G., Dowd, M., Lewis, M. R., & Worm, B. (2014). Estimating global chlorophyll changes over the past century. *Progress in Oceanography*, *122*, 163–173.

Bryan, D. R., McDermott, S. F., Nielsen, J. K., Fraser, D., & Rand, K. M. (2021). Seasonal migratory patterns of Pacific cod (*Gadus macrocephalus*) in the Aleutian Islands. *Animal Biotelemetry*, *9*(1), 1–18.

Canino, M. F., Spies, I. B., Cunningham, K. M., Hauser, L., & Grant, W. S. (2010). Multiple ice-age refugia in Pacific cod, *Gadus macrocephalus*. *Molecular Ecology*, *19*(19), 4339–4351.

Capotondi, A., Alexander, M. A., Bond, N. A., Curchitser, E. N., & Scott, J. D. (2012). Enhanced upper ocean stratification with climate change in the CMIP3 models. *Journal of Geophysical Research, Oceans*, *117*(C4), 1–23.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*(4), 1289–1303.

Cheng, Y., Miller, M. J., Zhang, D., Xiong, Y., Hao, Y., Jia, C., Cai, T., Li, S. H., Johansson, U. S., Liu, Y., & Chang, Y. (2021). Parallel genomic responses to historical climate change and high elevation in east Asian songbirds. *Proceedings of the National Academy of Sciences*, *118*(50), e2023918118.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157.

Cunningham, K. M., Canino, M. F., Spies, I. B., & Hauser, L. (2009). Genetic isolation by distance and localized fjord population structure in Pacific cod (*Gadus macrocephalus*): Limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(1), 153–166.

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, *209*(4), 1235–1278.

Deane, E. E., & Woo, N. (2009). Modulation of fish growth hormone levels by salinity, temperature, pollutants and aquaculture related stress: A review. *Reviews in Fish Biology and Fisheries*, *19*(1), 97–120.

Dennenmoser, S., Vamosi, S. M., Nolte, A. W., & Rogers, S. M. (2017). Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (*Cottus asper*) revealed by Pool-seq. *Molecular Ecology*, *26*(1), 25–42.

Dorant, Y., Benestan, L., Rougemont, Q., Normandeau, E., Boyle, B., Rochette, R., & Bernatchez, L. (2019). Comparing Pool-seq, rapture, and GBS genotyping for inferring weak population structure: The American lobster (*Homarus americanus*) as a case study. *Ecology and Evolution*, *9*(11), 6606–6623.

Drinan, D. P., Gruenthal, K. M., Canino, M. F., Lowry, D., Fisher, M. C., & Hauser, L. (2018). Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evolutionary Applications*, *11*(8), 1448–1464.

Elias, S. A., Short, S. K., Nelson, C. H., & Birks, H. H. (1996). Life and times of the Bering land bridge. *Nature*, *382*(6586), 60–63.

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, *64*, 1729–1747.

Ferretti, L., Ramos-Onsins, S. E., & Pérez-Enciso, M. (2013). Population genomics from pool sequencing. *Molecular Ecology*, *22*(22), 5561–5576.

Feulner, P. G., Chain, F. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., Lenz, T. L., Samonte, I. E., Stoll, M., Bornberg-Bauer, E., & Reusch, T. B. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genetics*, *11*(2), e1004966.

Fissel, B. E., Abelman, A., Dalton, M., Garber-Yonts, B. E., Haynie, A., Kasperski, S., Lee, J. T., Lew, D. K., Seung, C., Sparks, K., Szymkowiak, M., & Wise, S. (2020). Stock Assessment and Fishery Evaluation Report for the Groundfishes Fisheries of the Gulf of Alaska and Bering Sea/Aleutian Island Area: Economic Status of the Groundfish Fisheries off Alaska 2020. NPFMC. Nov. 2020. https://www.fisheries.noaa.gov/alaska/ecosystems/economic-status-reports-gulf-alaska-and-bering-sea-aleutian-islands

Gautier, M. (2015). BayPass genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, *201*(4), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Good, B. H., Walczak, A. M., Neher, R. A., & Desai, M. M. (2014). Genetic diversity in the interference selection limit. *PLoS Genetics*, *10*(3), e1004222.

Graham, A., Lavretsky, P., Muñoz-Fuentes, V., Green, A., Wilson, R., & McCracken, K. (2018). Migration-selection balance drives genetic differentiation in genes associated with high-altitude function in the speckled teal (*Anas flavirostris*) in the Andes. *Genome Biology and Evolution*, *10*(1), 14–32. https://doi.org/10.1093/gbe/evx253

Grebmeier, J. M., Overland, J. E., Moore, S. E., Farley, E. V., Carmack, E. C., Cooper, L. W., Frey, K. E., Helle, J. H., McLaughlin, F. A., & McNutt, S. L. (2006). A major ecosystem shift in the northern Bering Sea. *Science*, *311*(5766), 1461–1464.

Guirao-Rico, S., & González, J. (2021). Benchmarking the performance of Pool-seq SNP callers using simulated and real sequencing data. *Molecular Ecology Resources*, *21*(4), 1216–1229.

Han, F., Jamsandekar, M., Pettersson, M. E., Su, L., Fuentes-Pardo, A. P., Davis, B. W., Bekkevold, D., Berg, F., Casini, M., Dahle, G., & Farrell, E. D. (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *eLife*, *9*, e61076.

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, *65*(5), 910–924.

Hemmer-Hansen, J., Nielsen, E., Therkildsen, N. O., Taylor, M., Ogden, R., Geffen, A., Bekkevold, D., Helyar, S., Pampoulie, C., Johansen, T., & Carvalho, G. (2013). A genomic Island linked toecotype divergence in Atlantic cod. *Molecular Ecology*, *22*, 2653–2667.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from pool-seq data. *Genetics*, *210*(1), 315–330.

Hofmann, C. M., & Carleton, K. L. (2009). Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. *Integrative and Comparative Biology*, *49*(6), 630–643.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862.

Hughes, D. J., Campbell, D. A., Doblin, M. A., Kromkamp, J. C., Lawrenz, E., Moore, C. M., Oxborough, K., Prášil, O., Ralph, P. J., Alvarez, M. F., & Suggett, D. J. (2018). Roadmaps and detours: Active chlorophyll-a assessments of primary productivity across marine and freshwater systems. *Environmental Science & Technology*, *52*(21), 12039–12054.

Hunt, G. L., Jr., & Stabeno, P. J. (2005). Oceanography and ecology of the Aleutian archipelago: Spatial and temporal variation. *Fisheries Oceanography*, 14, 292–306.

Hurst, T. P., Laurel, B. J., & Ciannelli, L. (2010). Ontogenetic patterns and temperature-dependent growth rates in early life stages of Pacific cod (*Gadus macrocephalus*). *Fishery Bulletin*, 108, 382–392.

Kalinowski, S. T. (2005). Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity*, 94(1), 33–36.

Kirubakaran, T. G., Grove, H., Kent, M., Sandve, S., Baranski, M., Nome, T., De Rosa, M. C., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S., & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, 25, 2130–2143.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-seq). *Bioinformatics*, 27(24), 3435–3436.

Kristiansen, T., Drinkwater, K. F., Lough, R. G., & Sundby, S. (2011). Recruitment variability in North Atlantic cod and match-mismatch dynamics. *PLoS One*, 6(3), e17456.

Kurland, S., Wheat, C. W., de la Paz Celorio Mancera, M., Kutschera, V. E., Hill, J., Andersson, A., Rubin, C. J., Andersson, L., Ryman, N., & Laikre, L. (2019). Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. *Ecology and Evolution*, 9(19), 11448–11463.

Lang, M. G., & Livingston, P. A. (1996). Food habits of key groundfish species in the eastern Bering Sea slope region (Vol. 67). US Department of Commerce, National Oceanic and Atmoospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center.

Laurel, B. J., Hurst, T. P., Copeman, L. A., & Davis, M. W. (2008). The role of temperature on the growth and survival of early and late hatching Pacific cod larvae (*Gadus macrocephalus*). *Journal of Plankton Research*, 30(9), 1051–1060.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, 9(8), e1003118.

Lellouche, J.-M., Le Galloudec, O., Greiner, E., Garric, G., Regnier, C., Drevillon, M., & Le Traon, P. Y. (2018). The Copernicus Marine Environment Monitoring Service global ocean 1/12° physical reanalysis GLORYS12V1: description and quality assessment. *EGU General Assembly Conference Abstracts*, 20, 19806.

Lin, L. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45, 255–268.

Lin, L. (2000). A note on the concordance correlation coefficient. *Biometrics*, 56, 324–325.

Logerwell, E. A., Aydin, K., Barbeaux, S., Brown, E., Conners, M. E., Lowe, S., Orr, J. W., Ortiz, I., Reuter, R., & Spencer, P. (2005). Geographic patterns in the demersal ichthyofauna of the Aleutian Islands. *Fisheries Oceanography*, 14, 93–112.

Lotterhos, K. E. (2019). The effect of neutral recombination variation on genome scans for selection. *G3: Genes, Genomes, Genetics*, 9(6), 1851–1867.

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, 30(23), 5966–5993.

Luu, K., Bazin, E., & Blum, M. G. (2017). Pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67–77.

Mann, D. H., & Hamilton, T. D. (1995). Late Pleistocene and Holocene paleoenvironments of the North Pacific coast. *Quaternary Science Reviews*, 14(5), 449–471.

Menounos, B., Osborn, G., Clague, J. J., & Luckman, B. H. (2009). Latest Pleistocene and Holocene glacier fluctuations in western Canada. *Quaternary Science Reviews*, 28(21–22), 2049–2074.

Monnahan, P. J., Colicchio, J., & Kelly, J. K. (2015). A genomic selection component analysis characterizes migration-selection balance. *Evolution*, 69(7), 1713–1727.

Mordy, C. W., Stabeno, P. J., Ladd, C., Zeeman, S., Wisegarver, D. P., Salo, S. A., & Hunt, J. R. (2005). Nutrients and primary production along the eastern Aleutian Island archipelago. *Fisheries Oceanography*, 14, 55–76.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12), 3321–3323.

Nei, M. (1986). Definition and estimation of fixation indices. *Evolution*, 40(3), 643–645.

Neidetcher, S. K., Hurst, T. P., Ciannelli, L., & Logerwell, E. A. (2014). Spawning phenology and geography of Aleutian Islands and eastern Bering Sea Pacific cod (*Gadus macrocephalus*). *Deep Sea Research Part II: Topical Studies in Oceanography*, 109, 204–214.

Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103(6), 439–444.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402.

O'Leary, C. A., DeFilippo, L. B., Thorson, J. T., Kotwicki, S., Hoff, G. R., Kulik, V. V., Ianelli, J. N., & Punt, A. E. (2022). Understanding transboundary stocks' availability by combining multiple fisheries-independent surveys and oceanographic conditions in spatiotemporal models. *ICES Journal of Marine Science*, 79(4), 1063–1074.

Pampoulie, C., Skirnisdottir, S., Star, B., Jentoft, S., Jónsdóttir, I. G., Hjörleifsson, E., Thorsteinsson, V., Pálsson, O. K., Berg, P. R., Andersen, Ø., & Magnusdottir, S. (2015). Rhodopsin gene polymorphism associated with divergent light environments in Atlantic cod. *Behavior Genetics*, 45(2), 236–244.

Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., & Gross, S. S. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10), 983–987.

Rand, K. M., Munro, P., Neidetcher, S. K., & Nichol, D. G. (2014). Observations of seasonal movement from a single tag release group of Pacific cod in the eastern Bering Sea. *Marine and Coastal Fisheries*, 6(1), 287–296.

Rode, N. O., Holtz, Y., Loridon, K., Santoni, S., Ronfort, J., & Gay, L. (2018). How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Molecular Ecology Resources*, 18(2), 194–203.

Rohan, S. K., Kotwicki, S., Kearney, K. A., Schulien, J. A., Laman, E. A., Cokelet, E. D., Beauchamp, D. A., Britt, L. L., Aydin, K. Y., & Zador, S. G. (2021). Using bottom trawls to monitor subsurface water clarity in marine ecosystems. *Progress in Oceanography*, 194, 102554. https://doi.org/10.1016/j.pocean.2021.102554

Schaal, S. M., Haller, B. C., & Lotterhos, K. E. (2022). Inversion invasions: When the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Philosophical Transactions of the Royal Society B*, 377(1856), 20210200.

Schirrmann, M. K., Zoller, S., Croll, D., Stukenbrock, E. H., Leuchtmann, A., & Fior, S. (2018). Genomewide signatures of selection in Epichloë reveal candidate genes for host specialization. *Molecular Ecology*, 27(15), 3070–3086.

Shimada, A., & Kimura, D. (1994). Seasonal movements of Pacific cod *Gadus macrocephalus* in the eastern Bering Sea and adjacent waters based on tag-recapture data. *Fishery Bulletin*, 92, 800–816.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23–35.

Sodeland, M., Jorde, P. E., Lien, S., Jentoft, S., Berg, P. R., Grove, H., Kent, M. P., Arnyasi, M., Olsen, E. M., & Knutsen, H. (2016). "Islands of divergence" in the Atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution*, 8(4), 1012–1022.

Spies, I. (2012). Landscape genetics reveals population subdivision in Bering Sea and Aleutian Islands Pacific cod. *Transactions of the American Fisheries Society*, 141(6), 1557–1573.

Spies, I., Drinan, D. P., Petrou, E. L., Spurr, R., Tarpey, C., Hartinger, T., Larson, W., & Hauser, L. (2021). Evidence for selection and spatially distinct patterns found in a putative zona pellucida gene in Pacific cod, and implications for management. *Ecology and Evolution*, 11, 16661–16679.

Spies, I., Gruenthal, K. M., Drinan, D. P., Hollowed, A. B., Stevenson, D. E., Tarpey, C. M., & Hauser, L. (2020). Genetic evidence of a northward range expansion in the eastern Bering Sea stock of Pacific cod. *Evolutionary Applications*, 13(2), 362–375.

Spies, I., & Punt, A. E. (2015). The utility of genetics in marine fisheries management: A simulation study based on Pacific cod off Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(9), 1415–1432.

Stabeno, P. J., Farley, E. V., Jr., Kachel, N. B., Moore, S., Mordy, C. W., Napp, J. M., Overland, J. E., Pinchuk, A. I., & Sigler, M. F. (2012). A comparison of the physics of the northern and southern shelves of the eastern Bering Sea and some implications for the ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, 65, 14–30.

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*, 75(6), 1256–1273.

Stevenson, D. E., & Lauth, R. R. (2019). Bottom trawl surveys in the northern Bering Sea indicate recent shifts in the distribution of marine species. *Polar Biology*, 42(2), 407–421.

Stevenson, M. (2020). epiR: tools for the Analysis of Epidemiological Data. R package version 1.0-11. Available from: https://CRAN.R-project.org/package=epiR

Subramanian, S. (2019). Population size influences the type of nucleotide variations in humans. *BMC Genetics*, 20, 93. https://doi.org/10.1186/s12863-019-0798-9

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.

Torres-Dowdall, J., Pierotti, M. E., Härer, A., Karagic, N., Woltering, J. M., Henning, F., Elmer, K. R., & Meyer, A. (2017). Rapid and parallel adaptive evolution of the visual system of neotropical Midas cichlid fishes. *Molecular Biology and Evolution*, 34(10), 2469–2485.

Tørresen, O. K., Star, B., Jentoft, S., Reinar, W. B., Grove, H., Miller, J. R., Walenz, B. P., Knight, J., Ekholm, J. M., Peluso, P., & Edvardsen, R. B. (2017). An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics*, 18(1), 95.

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, 3(9), e285.

University of Washington. (2021). Pooled whole genome sequencing of Pacific cod spawning stocks. October 1, 2021, Bethesda, MD: National Center for Biotechnology Information (NCBI). PRJNA675289. Retrieved from https://www.ncbi.nlm.nih.gov/sra/PRJNA675289

Valen, R., Karlsen, R., & Helvik, J. V. (2018). Environmental, population and life-stage plasticity in the visual system of Atlantic cod. *Journal of Experimental Biology*, 221(1), jeb165191.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., & Banks, E. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1), 11–10.

Vendrami, D. L., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., Clark, M. S., Leese, F., McMinn, C., Moore, H., & Hoffman, J. I. (2017). RAD sequencing resolves fine-scale population structure in a benthic invertebrate: Implications for understanding phenotypic plasticity. *Royal Society Open Science*, 4(2), 160548.

Vestfals, C. D., Ciannelli, L., Duffy-Anderson, J. T., & Ladd, C. (2014). Effects of seasonal and interannual variability in along-shelf and cross-shelf transport on groundfish recruitment in the eastern Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 109, 190–203.

Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1587), 451–460.

Waples, R. S., & Gaggiotti, O. (2006). INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6), 1419–1439.

Waters, C. D., Hard, J. J., MSO, B., Fast, D. E., Warheit, K. I., Knudsen, C. M., Bosch, W. J., & Naish, K. A. (2018). Genomewide association analyses of fitness traits in captive-reared Chinook salmon: Applications in evaluating conservation strategies. *Evolutionary Applications*, 11(6), 853–868.

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97.

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, 110(19), E1743–E1751.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution*, 65(7), 1897–1911.

SUPPORTING INFORMATION

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

How to cite this article: Spies, I., Tarpey, C., Kristiansen, T., Fisher, M., Rohan, S., & Hauser, L. (2022). Genomic differentiation in Pacific cod using Pool-Seq. *Evolutionary Applications*, 00, 1–18. https://doi.org/10.1111/eva.13488