

Korean stock structure and transoceanic divergence in Pacific cod (*Gadus macrocephalus*)

Mary Fisher

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Lorenz Hauser

Kerry Naish

Michael Canino

Thomas Helser

Program Authorized to Offer Degree:

Aquatic and Fishery Sciences

© Copyright 2018

Mary Fisher

University of Washington

Abstract

Korean stock structure and transoceanic divergence in Pacific cod (*Gadus macrocephalus*)

Mary Fisher

Chair of the Supervisory Committee:
Dr. Lorenz Hauser
Aquatic and Fishery Sciences

Large marine populations, while inhabiting an environment with few clear barriers to gene flow, have been shown to display complex genetic structure on varying spatial scales. Although next-generation sequencing has improved our ability to discern marine population structure and the underlying genomic architecture, we still have much to learn about the evolution of genetic divergence in the marine environment, and its implications for management of exploited marine species. Pacific cod (*Gadus macrocephalus*) is a marine finfish targeted by economically valuable fisheries throughout its range in the northern Pacific Ocean. The species displays diverse population structure, which arises from a combination of different demographic histories and oceanographic barriers to gene flow. Such genetic structure allows for the study of adaptive divergence across large spatial scales, and carries implications for fisheries management. We first analyzed fine-scale genetic stock structure of Pacific cod around the Korean peninsula, where it is the target of a commercial fishery and is supplemented by stock enhancement programs. Our RAD sequencing results suggest that the Korean Pacific cod fishery

should be managed as three genetically distinct stocks on the western, eastern, and southern Korean coasts. Although these three stocks were highly differentiated from each other, we observed a high rate of migration between different coasts during the spawning season. We then compared this Korean data set from the western Pacific with existing RAD sequencing data from eastern Pacific cod, sampled along the coasts of Alaska, British Columbia, and Washington State. We found no evidence of parallel adaptation between the highly diverged eastern and western lineages of Pacific cod, which is in contrast to reports on Atlantic cod. The number of genomic regions of elevated divergence was lowest among subpopulations in the eastern Pacific, higher in subpopulations around the Korean peninsula, and highest in the comparison between the eastern and western Pacific lineages. The majority of these differences, as well as the opposite trend in the width of genomic regions of elevated divergence, were non-significant. As the first comparison of eastern and western Pacific cod using next-generation sequencing, our study provided key information regarding the genetic characteristics and structure of Korean cod stocks, uncovered unexpected spawning migration behavior around the Korean peninsula, and provided insights into the diversity and dynamics of adaptive divergence in a marine species of considerable commercial and ecological importance.

Table of Contents

List of Figures	iv
List of Tables	vi
Chapter 1.....	1
1.1 Abstract	1
1.2 Introduction	2
1.3 Methods.....	7
1.4 Results	15
1.4.1 Population structure	17
1.4.2 Identification of potential migrants.....	18
1.4.3 Re-calculation of effective population size.....	19
1.4.4 Individual assignment tests	20
1.4.5 Outlier loci	21
1.5 Discussion	23
1.5.1 Potential influence of currents and glacial history on stock structure	24
1.5.2 Discrepancies in stock structure among studies.....	27
1.5.3 Are migrants individuals that skip spawning?	29
1.5.4 Low effective population size.....	31
1.5.5 Evidence for selection.....	34

1.5.6	Management applications and future directions	36
1.6	Figures	39
1.7	Tables	51
1.8	References	56
Chapter 2		65
2.1	Abstract	65
2.2	Introduction	66
2.3	Methods	71
2.4	Results	80
2.4.1	Identification of genomic regions of elevated divergence	81
2.4.2	Changes in size and frequency of genomic regions of elevated divergence.....	82
2.4.3	Overlap in genomic regions of elevated divergence	83
2.4.4	Overlap in outlier loci and annotations	84
2.4.5	Outlier loci between eastern and western Pacific cod	85
2.4.6	Population structure according to neutral and putatively selected loci.....	85
2.5	Discussion	87
2.5.1	Differentiation between eastern and western populations of Pacific cod	87
2.5.2	Genomic regions of elevated divergence across divergence scenarios	88
2.5.3	Differing signatures of adaptation in eastern and western Pacific cod	91

2.5.4	Limitations in identifying outlier loci and regions	94
2.5.5	Function of diverged loci	96
2.5.6	Conclusions	98
2.6	Figures	100
2.7	Tables	112
2.8	References	118

List of Figures

Figure 1.1. Pacific cod catch in Korean coastal waters, from 1926 to 2015.....	39
Figure 1.2. Sampling sites and their putative population membership from prior microsatellite studies around the Korean peninsula.	40
Figure 1.3. Pacific cod sampling sites around the Korean peninsula used for this study.	41
Figure 1.4. Distributions of total length versus body weight of sampled fish.	42
Figure 1.5. Boxplots of gonadosomatic index of sampled fish.....	43
Figure 1.6. Principal component analysis of all sampling sites and temporal replicates.....	44
Figure 1.7. PCA and DAPC of southern sampling sites and temporal replicates.....	45
Figure 1.8. Proportional membership to each population for every individual, estimated from STRUCTURE.	46
Figure 1.9. Naïve and corrected effective population size.....	47
Figure 1.10. Assignment success when assigning individuals to aggregate of origin.	48
Figure 1.11. Assignment success when assigning individuals to population of origin.	49
Figure 1.12. Pressure-gridded sea surface temperatures around the Korean peninsula during the month of February.....	50
Figure 2.1. Three scenarios of divergence through time.....	100
Figure 2.2. Eastern and western population sampling sites used in this study.	101
Figure 2.3. Manhattan plots displaying per-locus F_{st}	102
Figure 2.4. Number of loci used to calculate each weighted average of F_{st} for sliding window analyses.	103

Figure 2.5. Kernel-smoothing weighted average of F_{st} and H_o , with regions of elevated divergence highlighted.....	104
Figure 2.6. Number of genomic regions of elevated divergence per linkage group.....	105
Figure 2.7. Kernel-smoothing weighted average F_{st} along linkage group 19.	106
Figure 2.8. Kernel-smoothing weighted average F_{st} along linkage groups 16 and 20.....	107
Figure 2.9. Bayescan output with a conservative neutral prior odds of 1000.....	108
Figure 2.10. Number of putative outlier loci identified per linkage group.	109
Figure 2.11. Principal component analysis on neutral versus outlier loci, for the full data set.	110
Figure 2.12. Principal component analysis on neutral versus outlier loci, for the western and the eastern populations.....	111

List of Tables

Table 1.1. Pacific cod tissue or fin clip samples from the Korean peninsula.	51
Table 1.2. Population genetic metrics for each sampling site.	51
Table 1.3. Uncorrected effective population sizes for each sampling site.	52
Table 1.4. Pairwise F_{st} between sampling sites.	53
Table 1.5. Inferred proportional ancestry for putative migrants identified in the PCA.	54
Table 1.6. Uncorrected effective population sizes for the southern population.	54
Table 1.7. Outlier loci identified in both Bayescan and OutFLANK.	55
Table 1.8. Outlier loci that aligned to annotated, protein-coding regions.	56
Table 2.1. Divergence scenarios represented by populations of Pacific cod.	112
Table 2.2. Pacific cod tissue or fin clip samples used in this study.	113
Table 2.3. Observed heterozygosity and F_{is} for each putative subpopulation across both eastern and western populations of Pacific cod.	114
Table 2.4. Results from the analysis of molecular variance.	114
Table 2.5. Frequency and width of genomic regions of elevated divergence.	115
Table 2.6. Results from the Kruskal-Wallis rank sum test across divergence scenarios.	115
Table 2.7. Results from the Welch's t-tests between divergence scenarios.	116
Table 2.8. Gene annotations for loci that were present in the genomic region of elevated divergence on linkage group 19.	116
Table 2.9. Counts of putative outlier loci, listed by detection method(s).	117
Table 2.10. Gene annotations for loci identified as outliers in the western population.	117
Table 2.11. Gene annotations for loci identified as outliers in the eastern population.	118

Acknowledgements

I must first thank the University of Washington's College of the Environment and School of Aquatic and Fishery Sciences, and the Saltonstall-Kennedy Grant Program, for providing the funding for my degree. This research would not have been possible without funding from the Alaska Fisheries Science Center and Korea's National Institute of Fisheries Science, as well as samples from Gyeongsang National University and the National Institute of Fisheries Science.

I am profoundly grateful to the many people who have lent their time and resources both to this research and to my development as a scientist. I would like to thank Lorenz Hauser, for orchestrating this project and acting as my primary adviser. I also greatly appreciate the time, technical knowledge, and advice provided by my committee members, Kerry Naish, Mike Canino, and Tom Helser. The collaborative efforts of Tom Helser and the AFSC's Age & Growth Lab helped initiate and frame the first chapter of this work. I would also like to thank our collaborator Sukyung Kang, who provided tissue samples for this project as well as regional expertise, and who coordinated my travel to Korea through the Korea-U.S. Science Cooperation Center. Samples were also generously provided by Wooseok Gwak, another invaluable collaborator who shared his knowledge and expertise throughout this project. I am greatly indebted to current and past members of the UW's Molecular Ecology Research lab, who have provided me with academic and technical support throughout my degree. I would particularly like to thank Kristen Gruenthal, for her work on the eastern Pacific cod samples, Isadora Jimenez, for her training and advice in the lab, and Dan Drinan and Marine Briec, for their guidance on bioinformatics. Finally, the completion of my degree would not have been possible without unfailing support and encouragement from my parents.

Chapter 1.

RAD-sequencing of Pacific cod (*Gadus macrocephalus*) around the Korean peninsula reveals high migration rates during the spawning season despite strong population structure

1.1 ABSTRACT

Applied population genetic analysis can inform more sustainable management and conservation practices for exploited marine species. Pacific cod (*Gadus macrocephalus*) is a marine finfish targeted by commercial fisheries across its north Pacific range. In Korean coastal waters, the Pacific cod fishery has yet to recover from severe declines in the 1950s. Current management plans do not incorporate genetic population structure because fine-scale structure of Korean cod stocks, and the temporal stability of this structure, remain uncertain. We conducted restriction-site associated DNA (RAD) sequencing on 243 individuals sampled across four winter spawning seasons at seven known Pacific cod spawning grounds around the Korean peninsula. Our analyses show strong regional genetic breaks between Pacific cod sampled on the southern / southeastern, eastern, and western coasts (pairwise $F_{st} = 0.025 - 0.042$). As a result, we had 100% assignment success with as few as 100 markers when assigning individuals to their stock of origin. We did not find significant genetic differentiation within regional stocks, or between temporal replicates. Despite strong population structure, we detected an unexpectedly

high migration rate of mature individuals between stocks, raising the possibility of either skipped spawning in this species, or reduced reproductive success in between-region migrants around the Korean peninsula. We also estimated markedly low effective population sizes at sampling sites and for each stock ($N_e = 445 - 2051$), which is congruent with the history of depletion in this region and suggests that Korean cod stocks are at risk of reduced adaptive capacity from low genetic diversity. This research has clear applications to Pacific cod fishery management, and adds to a growing body of evidence which suggests the occurrence of migration that does not contribute to gene flow between highly diverged Pacific cod stocks around the Korean peninsula.

1.2 INTRODUCTION

Fishery management plans that include population genetic structure reduce the probability of overexploitation of local populations and allow for a more sustainable yield (Reiss et al., 2009; Spies & Punt, 2015; Waples, Punt, & Cope, 2008). Population genetic analyses of fisheries species can also establish baseline information to monitor change in harvested populations (Allendorf et al., 2008) and provide managers with useful tools, including the assignment of individuals to their population of origin. As our ability to discern genetic structure in large marine populations improves with next-generation sequencing, the use of population genetics to study marine fisheries species is of increasing importance in their conservation and management.

One such species, Pacific cod (*Gadus macrocephalus*), is a demersal marine finfish with a pan-oceanic range. Usually widely dispersed along the continental shelf of the northern Pacific Ocean (Rand et al., 2014; Shimada & Kimura, 1994), Pacific cod migrate to coastal spawning grounds each winter and form large aggregates to breed (Gustafson et al., 2000). The exact

timing of this spawning period, in addition to optimal spawning temperature, varies latitudinally (Gustafson et al., 2000).

Observations of temporally stable aggregates during the spawning season (Gustafson et al., 2000), and the results of a mark-recapture study in the northeastern Pacific (Rand et al., 2014), suggest that Pacific cod display annual site fidelity to spawning aggregates. The mark-recapture study by Rand et al. (2014) involved the release of tagged individuals at spawning grounds during the spring spawning season, and found that the majority of cod recaptured during the following spring were within 100km of their release location. In contrast, the majority recaptured during the subsequent fall and winter were much further (403-471 km on average) from their release location (Rand et al., 2014).

In combination with microsatellite analysis that showed limited effective dispersal between aggregates (Cunningham et al., 2009), these studies describe a system where genetic differences can accumulate between populations of Pacific cod during the spawning season, as a result of restricted gene flow. Population structure between spawning aggregates has been observed using microsatellite loci across the Pacific cod's range (Cunningham et al., 2009; Gwak & Nakayama, 2011; Kim, An, & Choi, 2010), although the delineation of fine-scale structure with next-generation sequencing is currently limited to stocks in the eastern Pacific (Drinan et al., 2018). Given the complex stock composition observed in the more intensely studied Atlantic cod, such as sympatrically spawning migratory and stationary ecotypes (Hemmer-Hansen et al., 2013; Kirubakaran et al., 2016), as well as the variation in Pacific cod population structure among existing studies, there is likely additional, unknown complexity in spawning migration and structure between Pacific cod populations.

With highly productive Pacific cod fisheries operating throughout the year, such population structure is an important consideration for fisheries management. In 2010, microsatellite analyses of Pacific cod spawning sites in the Bering Sea and along the west coast of the United States and Canada (Cunningham et al., 2009; Spies, 2012) contributed to the North Pacific Fishery Management Council's decision to split the Bering Sea / Aleutian Islands management unit into two separate stocks (McCracken, DiCosimo, & Fey, 2013). One of these studies (Cunningham et al., 2009) also found a sharp genetic break between northeastern coastal spawning sites and a depleted population of Pacific cod in the Salish Sea. This finding contributed to the National Marine Fisheries Service's decision to list Salish Sea Pacific cod as a Species of Concern in 2010 (NOAA NMFS, 2010). However, we still have much to learn about the genetic stock structure of Pacific cod fisheries in the western part of its range, particularly around the Korean peninsula where it is the subject of a rebounding fishery and hatchery-based stock supplementation.

The Pacific cod fishery was historically important in South Korea, where catch consistently reached above 20,000mt in the 1930s (**Figure 1.1**). The fishery has since faced sharp declines that began in the early 1950s, with historic lows in catch in the 1990s. Yearly catch along the southeastern coast remained below 1000mt until the early 2000s (Chung, Kim, & Kang, 2013). Although decadal flux in seawater temperature is believed to contribute to some variation in catch abundance (Chung, Kim, & Kang, 2013), the historical decline of the Pacific cod fishery is largely attributed to overfishing (Gwak & Nakayama, 2011; Kim et al., 2010; Lee & Midani, 2009).

Recent management efforts to rebuild the fishery have included the release of hatchery-reared larvae and juveniles (Kim et al., 2010) and the development of an ecosystem-based

Fisheries Stock Rebuilding Plan (Lee & Midani, 2009). Establishing the population genetic structure of these stocks has been identified as a concern for current fishery management and is one of the objectives of the Fisheries Panel formed under the Korea – NOAA Joint Project Agreement (Low, 2014). A better understanding of Korean stock structure would also improve ongoing hatchery supplementation programs, by allowing managers to monitor the genetic impacts of these programs, and by supporting a program design that minimizes genetic risk to wild populations (Naish et al., 2008).

There have been two previous studies of Pacific cod population structure in Korean coastal waters using microsatellite loci, but they provide conflicting characterizations (**Figure 1.2**). Kim et al. (2010) found that Pacific cod around the Korean peninsula are divided into an eastern and a western population. Their results align with an earlier distinction between eastern and western Korean cod based on morphological differences (Zhang, 1984). In contrast, Gwak & Nakayama (2011) described three distinct populations along the western, southern, and eastern coasts. The two studies also differ on the location of the southern edge of the eastern population, with the earlier study suggesting it may extend well into to the southeastern coastline (Kim et al., 2010) as opposed to a more northern boundary proposed by Gwak & Nakayama (2011) (**Figure 1.2**). Neither study explored the temporal stability of spatial population structure, or investigated the presence of distinct groups of early and late spawning fish at the Jinhae Bay spawning grounds, which is suggested based on morphological studies and field observations (W. Gwak, Gyeongsang National University, personal communication). This phenomenon has been observed in other fish species (Beacham et al., 2008; Bekkevold et al., 2016; Quinn, Unwin, & Kinnison, 2000), but not explored through genetic analysis in Pacific cod.

With these somewhat contradictory results from prior studies, and a need for analysis of smaller-scale differentiation and temporal stability, it is difficult to incorporate existing information on genetic stock structure into Korean fisheries management. Additionally, the next-generation sequencing techniques which have evolved since research by Gwak & Nakayama (2011) and Kim et al. (2010) facilitate more comprehensive investigations of seasonal migration and dispersal through the assignment of individuals to a population of origin.

Further genetic analyses of Korean cod stocks would also contribute to our understanding of how climate change will impact Pacific cod and the fisheries that rely upon it. Populations of Pacific cod around the Korean peninsula exist at the southern edge of the species' range. Korean stocks spawn at the highest temperatures observed in Pacific cod (5 to 9°C, in contrast to 1 to 5°C in the Bering Sea (Gustafson et al., 2000)), particularly along the southern coastline where they inhabit the warmer, more saline waters of the Tsushima Warm Current (Chang et al., 2004). Establishing a baseline of genetic stock structure and diversity can help monitor early climate change impacts in this species. Genomic comparisons of individuals at spawning grounds in the cooler waters of the eastern and western coasts with those in the warmer waters of the southern coast may also help identify local adaptations that are key to thermal tolerance during spawning and early life history stages. Such temperature-associated, highly diverged loci have been observed in Atlantic cod (Bradbury et al., 2010).

The aim of this study was to use next-generation sequencing and a more expansive set of samples than prior studies, including within- and between-year temporal replicates, to conduct a fine-scale analysis of Pacific cod population structure around the Korean peninsula. Once we established the genetic stock structure, we tested our ability to assign individual fish to their population of origin and calculated effective population sizes for each population. Finally, we

explored local adaptation at southern spawning grounds that may be linked to spawning and survival during early life history stages in different thermal regimes.

1.3 METHODS

1.3.1 *Sample collection*

A total of 322 fin clips and tissue samples were collected from aggregates at known Pacific cod spawning sites during the spawning season, from December to March, in coastal waters around the Korean peninsula (**Figure 1.3; Table 1.1**). Three sites were sampled during the 2007 – 2008 winter spawning season, three during the 2014-2015 season, and one during the 2015 – 2016 season. Two of the sampling sites had either between- or within- year temporal replicates: Geoje, which was sampled during both the 2013-2014 and 2014-2015 spawning seasons, and Jinhae Bay, which was sampled toward the beginning (December) and end (February) of the 2007-2008 spawning season to collect early- and late-spawning fish. Fin clips and tissues were preserved in 95-100% non-denatured ethanol and stored at 4°C.

1.3.2 *Verifying maturity of sampled fish*

It can be difficult to assess Pacific cod maturity and spawning condition in the field, so we analyzed morphometric data collected for each individual during sampling. Total length, standard length, weight, and sex data were available from all sampled aggregates except for the temporal replicate at Jinhae Bay (Feb. 2007-2008). The distribution of total length measurements were analyzed by sampling site and compared to estimates of 50% length-at-maturity from the literature. We applied 50% length-at-maturity estimates for males and females that were specific to the Yellow Sea and the East Sea (Lee et al., 2016) for the sampling sites in those regions; we

used the smaller Yellow Sea estimates for the sampling sites along the southern / southeastern coast, with the assumption that cod spawning in warmer waters, such as the Tsushima Current along the southern Korean coastline, mature more quickly than those in colder waters (Gustafson et al., 2000). For each individual with available gonad weight, which included approximately 50% of all samples, we also calculated gonadosomatic index (GSI). The GSI distribution at sampling sites was compared to mean GSI per maturity phase in Japanese waters, according to Hattori et al. (1992). GSI was also cross-referenced with length data for individuals which were classified as immature according to 50% length-at-maturity estimates.

1.3.3 *RAD-sequencing*

DNA was extracted from fin clip and tissue samples using DNeasy 96-well Blood & Tissue Kits (Qiagen Inc., Valencia, CA). DNA was then quantified using Quant-iT PicoGreen dsDNA Reagent (Invitrogen, Carlsbad, CA) and visualized on an agarose gel. Samples with DNA concentrations below 3 ng/uL ($n = 45$) were subsampled a second time and re-extracted with the same DNeasy 96-well Kit, using a slightly altered protocol to maximize DNA concentration (Appendix A). Ten samples were discarded due to insufficient DNA concentrations after re-extraction. This left a final 283 samples for restriction-site associated (RAD) sequencing.

RAD libraries were prepared according to Baird et al. (2008) and Etter et al. (2011), with modifications to incorporate Agencourt AMPure XP beads (Beckman Coulter, Inc., Pasadena, CA) for purification (Drinan et al., 2018). Both 300ng and 500ng DNA libraries were prepared, according to the maximum amount of DNA in each sample. Several samples were sequenced under both 500ng and 300ng protocols, to assess resulting genotype discrepancies (Appendix B).

Fragment sizes in RAD libraries were estimated with 1% E-Gel EX agarose gels (Invitrogen, Carlsbad, CA). DNA concentrations and quality were assessed using Quant-iT PicoGreen dsDNA Reagent (Invitrogen, Carlsbad, CA). Five libraries were prepared with 60-72 individuals pooled per library, for a total concentration of 5nM or 10nM. Sequencing of 150bp single end (n = 157) and paired end (n = 126) reads was completed on an Illumina HiSeq 4000 (Illumina Inc., San Diego, CA) at the University of Oregon's Genomics and Cell Characterization Core Facility.

1.3.4 *SNP discovery, genotyping, and filtering*

Quality filtering and demultiplexing of raw RAD sequencing data, *de novo* construction of a reference database of RAD loci, SNP discovery, and genotyping, was completed using a combination of the Stacks v1.44 pipeline (Catchen et al., 2011; Catchen et al., 2013), Bowtie (Langmead et al., 2009), and NCBI's Basic Local Alignment Search Tool, BLAST (Altschul et al., 1990), according to the procedures outlined in Drinan et al. (2018) and Brieuc et al. (2014). In summary, filtered and demultiplexed sequencing reads were sorted into loci in each individual *de novo*, and then ten individuals with the greatest sequencing depth (defined as the greatest number of *de novo* loci in samples that possessed at least the average number of total raw reads) from each aggregate were used to generate a reference catalog of loci. This catalog was filtered by aligning loci against themselves in BLAST and Bowtie, with any locus that aligned best to another locus removed (Brieuc et al., 2014). This filtered catalog of loci was then used as a reference database in a new run through the Stacks pipeline. Stacks flags (pstacks m = 5, populations m = 10, M = 3, N = 4, n = 3, max_locus_stacks = 3) were set according to Mastretta-Yanes et al. (2015) and Brieuc et al. (2014), to maximize the number of loci while reducing SNP and allele calling error rates. Loci were filtered to include one SNP per RAD tag (flag:

write_random_SNP), and then genotyped in each aggregate where $\geq 80\%$ of fish possessed that SNP (flag: r = 0.80). Stacks was used to call initial genotypes with a maximum likelihood approach.

Genotypes for biallelic SNPs were then re-called to account for undercalling of heterozygotes by Stacks, using methods and code described in Briec et al. (2014). Briec et al. (2014) call a heterozygote if two or more reads are present with the minor allele. Final filtering was then used to remove loci with a minor allele frequency (MAF) < 0.05 in every aggregate, and individuals with more than 30% of missing genotypes. Loci were also filtered for missing data across all aggregates at the same 30% cutoff, and were then removed if they did not conform to Hardy – Weinberg equilibrium (HWE). Genepop v4.2 was used to estimate expected heterozygosity and F_{is} and to perform exact tests, which were then used to identify loci significantly out of HWE according to Fisher's combination of probabilities from independent tests of significance (Sokal & Rohlf, 1995).

Potentially contaminated individuals were identified by their higher heterozygosity and putative relatedness to other individuals, and then removed from the data set. Observed heterozygosity was calculated in Genepop v4.2, and putative genetic relationships were estimated using the maximum likelihood method in ML Relate (Kalinowski, Wagner, & Taper, 2006). Individuals with an observed heterozygosity greater than 0.25 were listed for potential removal. ML Relate was then run in several iterations. Since ML Relate can only process a maximum of two populations but estimates relatedness using population-specific allele frequencies, the data set was first entered with all aggregates as a single population, then aggregates in each region were entered separately.

One full-sibling pair ($R = 0.861$) was detected outside sample replicates; these individuals were not flagged for high heterozygosity but were numbered consecutively at the same site, so were assumed to be the same fish sampled twice. The sample with the larger proportion of missing data was removed. Each half-sibling pair identified by ML Relate was assessed for individual observed heterozygosity and proximity to each other on the DNA extraction plate or RAD library preparation plate. If half-sibling pairs included an individual that had been flagged for high heterozygosity, that individual was removed. If half-sibling pairs were located next to each other during RAD library preparation, those individuals were removed. In total, 21 individuals were removed for displaying both high heterozygosity, and proximity to a half-sibling pair during RAD library preparation (Table S1.1).

1.3.5 *Population structure analyses*

Genepop v4.2 was used to estimate locus-specific F-statistics (F_{st} , F_{is} , F_{it}), F_{is} and observed heterozygosity per aggregate, and pairwise F_{st} (Weir & Cockerham, 1984). Significance of pairwise F_{st} was tested using Fisher's exact probability test on the distribution of diploid genotypes (Rousset, 2008), as well as a permutation test ($n=1000$) in the R package strataG (Archer, Adams, & Schneiders, 2016). To test whether distributions of locus-specific F_{is} and observed heterozygosity changed significantly over time within genetically distinct populations, we conducted a non-parametric Wilcoxon rank sum test in R, with alpha set at 0.05 (Mann & Whitney, 1947; Wilcoxon, 1945).

Principal component analysis (PCA) was used to visualize genetic differences between samples, conducted with the R package Adegenet (Jombart, 2008). For the southern sampling sites alone, a discriminant analysis of principal components (DAPC) was conducted as a follow-

up analysis when the sites did not separate across the first and second principal components in the PCA. Effective population size was estimated for each sampled aggregate and for each season within the genetically distinct southern population (per PCA and STRUCTURE analysis) using the linkage disequilibrium method (Waples & Do, 2010) in NeEstimator v2.0 (Do et al., 2014).

This “naïve” or uncorrected effective population size was then corrected for downward bias from linkage between loci using the least-squares regression of $\ln(chr)$, per Waples et al. (2016). We applied the least-squares regression equation under the assumption that Pacific cod have the same number of chromosomes as Atlantic cod (23) (Tørresen et al., 2017) and that the regression equation itself is robust across fish species.

The number of genetically distinct populations in the data was determined using STRUCTURE v2.3.4 (Porras-Hurtado et al., 2013; Pritchard, Stephens, & Donnelly, 2000). STRUCTURE places samples into population groupings according to Hardy-Weinberg and linkage equilibrium using a Bayesian clustering algorithm with Markov Chain Monte Carlo (MCMC) estimation (Pritchard et al., 2000). This allows for a progressive convergence toward reliable allele frequency estimates in each population, and membership probabilities of sampled individuals (Porras-Hurtado et al., 2013). While certain sampling sites in this data set consist of a relatively low sample size (lowest $n=11$), the lower variability of SNPs mean that sample sizes as low as four should still provide reliable data when using Bayesian population analysis methods (Shi et al., 2010).

STRUCTURE was run with a burn-in period of 50,000 followed by 100,000 MCMC replicates. The estimated number of populations (K) was varied from 1 – 9, with 3 iterations per

value of K . The likelihood of the data was calculated at each K , and then the mean log-likelihoods plotted to determine the optimum value of K (the smallest “stable” K value that maximizes the global likelihood of the data) (Evanno, Regnaut, & Goudet, 2005).

1.3.6 *Assignment testing*

Individuals were assigned to populations according to Benestan et al. (2015), using the training, holdout, leave-one-out method described in Anderson et al. (2008) and Anderson (2010) and implemented in *gsi_sim*. Analyses were conducted through the R package assigner v0.5.0 (Gosselin, Anderson, & Bradbury, 2016). Assigner identifies highly discriminatory loci using a training dataset, and then employs a leave-one-out cross validation method with a test data set (Anderson, Waples, & Kalinowski, 2008) to provide less biased estimates of genetic stock identification (Anderson, 2010). Assignment was conducted using 10, 50, 100, 200, 500, 1000, 2000, 5000, and all 5804 loci, with ten iterations. Half of all individuals were used as a training data set, and the remaining half used for the test data set. We tested our ability to assign individuals to both their sampling site and to their population (established by PCA and STRUCTURE) of origin.

1.3.7 *Potential migrants*

Any individuals that clustered with a different, genetically distinct population than the one in which they were caught were identified as potential migrants. We classified migrant maturity according to total length and gonadosomatic index, when available (per methods described in *Verifying maturity of sampled fish* above). We also re-assessed relevant metrics and laboratory notes from genetic analysis: relationships identified by ML Relate, observed heterozygosity, proportion of missing data, quality of the DNA, and proximity of the sample to

individuals from the assigned population on 96-well plates used for DNA extraction or RAD library preparation.

1.3.8 *Outlier testing and alignment to Atlantic cod genome*

Candidate loci under selection were identified by performing outlier tests with BayeScan v2.1 (Foll & Gaggiotti, 2008) and OutFLANK (Whitlock & Lotterhos, 2015) to account for variation in the relative power of outlier tests among different models of population structure (Whitlock & Lotterhos, 2015). Bayescan employs a Bayesian framework to directly estimate the posterior probability of a given locus being under selection, using an MCMC approach (Foll & Gaggiotti, 2008). Alternatively, OutFLANK identifies putative outliers using an inferred distribution of F_{st} for loci not affected by balancing selection or spatially diversifying selection (Whitlock & Lotterhos, 2015), allowing for low false positive rates and relatively high power across different population structures.

Putative outlier loci were detected in Bayescan v2.1 using the default settings and 20,000 iterations. The prior odds for the neutral model were varied stepwise from 10, to 100, 1000, and 10000 to identify outliers over the full data set, and among aggregates along the southern coast. Any additional, exploratory outlier analyses were run using only prior odds of 100 and 1000 because they were reasonable models halfway between a prior of 10 (suggested for a few hundred markers), and a prior of 10000 (used in for genome-wide associated studies with millions of SNPs) (Foll, 2012). With a false discovery rate of 0.05, loci were considered outliers if they possessed a $\log_{10}q$ value smaller than -1.3. In OutFLANK, we trimmed five percent of F_{st} values from the right and left tails of the distribution. The threshold for detection of outlier loci was set at $q = 0.05$.

In order to explore whether the outlier loci had co-localized with annotated genes or regions under selection in the Atlantic cod genome, candidate loci under selection were then aligned to the most recent assembly of the Atlantic cod genome, *gadMor2* (Tørresen et al., 2017), using Bowtie2 (Langmead & Salzberg, 2012). Bowtie2 uses the same approach as Bowtie but allows for gapped alignment through a two-step algorithm (Langmead & Salzberg, 2012), which was more appropriate for alignment of sequences between two distinct species. Outlier loci were aligned to the Atlantic cod genome if they were detected in either OutFLANK or Bayescan (with prior odds of 100 or above). Alignments were then filtered for a mapping quality greater than ten (less than a 1 in 10 chance that the read truly originated elsewhere). The function `closestBed` from `bedtools` v2.24 (Quinlan & Hall, 2010) was used to match alignments to the annotation file for the Atlantic cod genome.

1.4 RESULTS

The final data set consisted of 5,804 RAD loci and 243 individuals (**Table 1.1**). The average genotyping discrepancy between 300ng and 500ng protocols was 6.31%, and within protocols was 4.37%; this suggests that genotyping discrepancy due to protocol differences alone was approximately 2% (Appendix B).

At all sampling sites except Jukbyeon, the majority of individuals were above the sex-specific 50% length-at-maturity estimates by Lee et al. (2016) (**Figure 1.4; Table 1.1**). All individuals at Jukbyeon were below the 50% length-at-maturity estimate for the East Sea. An additional seven females at Pohang, and one female at Boryeong, were below the 50% length-at-maturity estimate for the Yellow Sea. The mean and variance of gonadosomatic index (GSI)

varied greatly between the four sampling sites for which gonad weight was available (**Figure 1.5**), with the minimum GSI across sites 0.64% (immature individual at Pohang) and the maximum 36.17% (mature individual at Geoje). There was one individual sampled at Pohang classified as immature based on 50% length-at-age estimates which appeared to be mature based on the gonadosomatic index (**Figure 1.5**). Based on GSI, individuals classified as immature by length fell within the yolkless (I) and yolk vesicle phases (II) in Hattori et al. (1992), whereas mature individuals spanned nearly all gonad maturity phases, from the yolk vesicle phase (II) to the spawning (VIII) phase.

Average F_{is} over loci remained close to zero for every spawning aggregate, ranging from -0.092 at Geoje (2014-2015 spawning season), to 0.066 at Jukbyeon (**Table 1.2**). Global locus F_{is} was -0.009, F_{st} was 0.022, and F_{it} was 0.013. Per-locus F_{is} , F_{it} , and F_{st} ranged from -0.288 to 0.661, -0.019 to 0.303, and -0.264 to 0.667, respectively (Figure S1.1).

Uncorrected effective population size for each spawning aggregate ranged from $N_e = 236$ at Jinhae Bay (Dec. 2007-2008) to $N_e = 2,051$ at Boryeong (2007-2008) when the lowest minor allele frequency used was 0.010 (**Table 1.3**). All upper confidence limits were finite and below $N_e = 3,000$. The two sampling sites along the western coast had 2-4 times greater N_e than all other sites. Effective population size estimates at Geoje remained consistent across spawning seasons ($N_e = 437 - 447$) when the lowest allele frequency used was 0.010. However, at Jinhae Bay, effective population size differed three-fold between the sample of early spawning fish in December 2007 ($N_e = 236$) and the sample of late spawning fish in February 2008 ($N_e = 654$).

1.4.1 Population structure

Pairwise F_{st} between sampling sites ranged from nearly no differentiation ($F_{st} = 0.0007$) to genetic divergence quite high for a marine species ($F_{st} = 0.0423$, **Table 1.4**). Pairwise F_{st} between sites on different coasts was highly significant, with no significant genetic differentiation within each regions. There were two exceptions according to the strataG permutation test, which found significant pairwise F_{st} between Geoje, spawning season 2014-2015, and the remainder of the southern aggregates, as well as between the two western sampling sites. This effect was not detected in Fisher's exact test. The southeastern sampling site, Pohang, showed significant differentiation from the eastern site, Jukbyeon ($F_{st} = 0.0378$), but no significant differentiation from the three southern sampling sites ($F_{st} = 0.0008 - 0.0027$). The within-season temporal replicate at Jinhae Bay did not show significant differentiation, with pairwise $F_{st} = 0.0002$ between early and late spawning fish.

The first and second principal components of the principal component analysis (PCA) explained 3.09% and 1.65% of the variation in the data. There were three distinct clusters, composed of the western (Yellow Sea Block, Boryeong), eastern (Jukbyeon), and southern (Namhae, Geoje, Jinhae Bay) sampling sites (**Figure 1.6**). The southeastern sampling site (Pohang) clustered with the southern sites in the ordination space, as would be expected from pairwise F_{st} .

A second PCA without the eastern and western sampling sites, and without individuals from southern sites which had clustered with the western coast, showed no discernable difference among the southern samples on either principal component (**Figure 1.7a**), as might be expected based on pairwise F_{st} . However, some separation appeared in the DAPC between the Geoje

samples taken during the 2014-2015 spawning season and the remainder of the southern sampling sites, including the other set of Geoje samples collected in 2013-2014 (**Figure 1.7b**).

Bayesian clustering in STRUCTURE suggested that there are $K = 3$ (Figure S1.2) genetically distinct populations within this data set, delineated according to the three clusters of the PCA: the western coast, the south / southeastern coast, and the eastern coast (**Figure 1.8**).

1.4.2 *Identification of potential migrants*

In the PCA (**Figure 1.6**), ten individuals clustered with sampling sites from a different coastal region than the one in which they were caught. Three individuals caught at Jukbyeon clustered with the western sites, and one with the southern sites. Six individuals from the southern sites (two each from Pohang, Geoje, and Namhae) clustered with the western sites. STRUCTURE confirmed the genetic origin of the ten putative migrants detected in the PCA (**Figure 1.8; Table 1.5**).

It is unlikely that we are detecting a false-positive signal of migration due to contamination; only three putative migrants were next to samples from the coastal region to which they were assigned during RAD library preparation, and even these displayed heterozygosity within the normal range and lacked relationships with neighboring samples according to ML Relate (Table S1.2a). Of the six immigrants identified in the southern sites, all but one individual were above the 50% length-at-maturity estimate for the Yellow Sea, its source population (**Figure 1.4; Table S1.2b**). Although the four potential migrants sampled at Jukbyeon were below the 50% length-at-maturity estimate for the East Sea, only one had a total length smaller than the estimate for 50% length-at-maturity in the Yellow Sea, which was the appropriate measure of maturity for its source population. For the four southern samples with

gonad weights available, the calculated gonadosomatic index ranged from 4.23 to 20.11% (**Figure 1.5**; Table S1.2b).

1.4.3 *Re-calculation of effective population size*

We recalculated effective population size excluding migrants for the Pohang, Geoje and Jukbyeon sampling sites. Removal of migrant individuals resulted in a 3.4% (Geoje 2014-2015, $N_e = 911$ to 943) to a 461% (Pohang, $N_e = 410$ to 1640) increase in effective population size (**Table 1.3**). With the exception of the Jinhae Bay aggregates, which did not include migrants, recalculation of effective population size without immigrants removed the substantial difference in N_e estimates between the western and southern sampling sites.

Our data set also provides two time points to compare effective population size within and among the western and southern populations: the first point from the 2007-2008 season, and the second from the 2014-2015 (southern population) / 2015-2016 (western population) seasons. We combined the latter two spawning seasons for the purpose of this comparison because each was the most recent spawning season sampled in the respective population, and the difference of a single year was minimal relative to the 7-8 year difference between the first and second time points.

First, to reflect the lack of significant genetic differentiation between aggregates along the southern / southeastern coastline, effective population size was recalculated for the entire southern population, excluding migrants, with aggregates grouped by spawning season (**Table 1.6**). Effective population size of the entire southern population increased three-fold from the 2007-2008 to the 2014-2015 spawning season. We observed a corresponding increase in observed heterozygosity and decrease in F_{is} over time (**Table 1.2**), both of which were statistically significant (Table S1.3). The effective population size of the southern population

reached a maximum $N_e = 1580$ ($MAF = 0.010$) during the 2014-2015 spawning season. This 2014-2015 N_e surpassed the western population's effective population size in 2015-2016, which was $N_e = 943$ (at Yellow Sea Block). Yet just seven years earlier, during the 2007-2008 spawning season, the southern population's N_e was four times lower than that of the western population. Within the western coast, effective population size from the 2007-2008 season (at Boryeong) was twice as large as effective population size from the 2015-2016 season (at Yellow Sea Block, **Table 1.3**). The decrease in observed heterozygosity and the increase in F_{is} during this period (**Table 1.2**) were statistically significant (Table S1.3).

The least-squares regression equation from Waples et al. (2016) showed a downward bias in our N_e estimates of approximately 27.4% as a result of physical linkage between loci. When we adjusted effective population size estimates at all sites, all point estimates of N_e remained below 3000, and all upper confidence limits below 4000 (**Figure 1.9**).

1.4.4 *Individual assignment tests*

Overall, only 44.2% individuals were successfully assigned back to aggregates using all 5804 markers (**Figure 1.10**). Assignment success was greatest for the eastern and western aggregates, likely because there was only one sampling site in the eastern region, and only two sites in the western region. A maximum of 99.2% of samples from the Yellow Sea Block site, 65.0% of samples collected from Boryeong, and 85.8% of samples collected from Jukbyeon were assigned correctly, using all markers. Assignment success for the southern aggregates ranged from 39.4% of individuals successfully assigned (Geoje site, 2014-2015 season) to zero individuals successfully assigned (Namhae, 2014-2015 season). Assignment success at several southern aggregates actually declined when a larger proportion of markers was used.

We then assigned individuals to the three populations identified by the PCA and validated with the pairwise F_{st} and STRUCTURE results. Assignment success was much greater when assigning individuals to population, rather than to aggregate, of origin (**Figure 1.11a**). Overall, 94.5% of individuals were correctly assigned to their population of origin using only 200 out of 5804 markers. The western population had a maximum assignment success of 100%, the eastern population a maximum of 87.8%, and the southern population a maximum of 95.8%, all with only 200 markers.

Assignment testing at the population level was rerun excluding the potential migrants identified in the PCA and STRUCTURE analyses. Without migrants, the percent of individuals successfully assigned to their population of origin increased to 100% in every population, with overall assignment success at 100% using only 100 markers, and almost 90% with only 10 markers (**Figure 1.11b**).

1.4.5 *Outlier loci*

Outlier analyses with Bayescan and OutFLANK were first conducted using the full data set. Bayescan identified 100, 29, 12, and 6 outlier loci when prior odds of the neutral model were set to one in 10, 100, 1000, and 10000, respectively. In contrast, OutFLANK identified only ten loci as showing significant evidence for selection at $q = 0.05$. Of the ten loci categorized as outliers in OutFLANK, nine, seven, five, and four loci were also identified in Bayescan, depending on the prior odds of the neutral model (10, 100, 1000, and 10000, respectively) (**Table 1.7**).

Identification of candidate loci under selection was then conducted with only the sampling sites in the southern population (Namhae, Geoje, Jinhae Bay, and Pohang, with

temporal replicates). Far more outliers were identified among southern aggregates when using OutFLANK than Bayescan. At a threshold of $q = 0.05$, OutFLANK identified 27 loci as putatively under selection. When OutFLANK was used to identify outliers between temporal replicates, one candidate locus was detected between spawning years at Geoje, while no outliers were detected between early and late spawning fish at Jinhae Bay.

Only one outlier locus (locus 24927) was identified by Bayescan among all southern aggregates, when using prior odds of 10-1000 (**Table 1.7**). This locus was also one of the 27 outlier loci identified by OutFLANK among all southern aggregates. Interestingly, this locus was identified by OutFLANK when analyzing the between-year temporal sample at Geoje. In order to further explore this pattern, we ran OutFLANK ($q = 0.05$) and Bayescan (false discovery rate = 0.05; prior odds 100, 1000) on two subsets of the southern aggregates: the first excluding the Geoje samples from 2013-2014, and the second excluding the Geoje samples from 2014-2015.

Locus 24927 was only identified as an outlier when the southern data set included the Geoje samples from the 2014-2015 spawning season, for both OutFLANK and Bayescan. No outlier loci were identified by Bayescan among the southern sites when the Geoje 2014-2015 sample was excluded from the data set.

This locus of interest was one of 43 outlier loci that aligned to sequences in the Atlantic cod genome with a mapping quality greater than ten. While the majority of these were spread across 17 of the 23 Atlantic cod linkage groups, eight aligned to linkage group 16 alone. Twenty aligned outlier loci lay within annotated protein-coding regions, seventeen of which coded for proteins of known function (**Table 1.8**).

1.5 DISCUSSION

Pairwise F_{st} , PCA, and STRUCTURE analysis show three genetically distinct populations of Pacific cod around the Korean peninsula: spawning fish on the western coast in the Yellow Sea, spawning fish along the southern / southeastern coast, and spawning fish on the eastern coast in the East Sea. Spatial population structure overwhelmed any temporal differentiation between sampling sites within each population which may have arisen from sampling the majority of study sites across a decade (2007-2008; 2014-2015/2015-2016). We also failed to find significant genetic structure between early and late spawning fish within the same year. Further investigation into population structure along the western coast, and between years at Geoje, may be warranted; the strataG permutation test showed significant differentiation between the Geoje 2014-2015 samples and all other southern sites, as well as between the two western sampling sites, although this was not reflected in the results from Fisher's exact test, STRUCTURE, or the PCA. As a result of high genetic differentiation between coastal populations and little within-population structure, we were able to correctly assign every individual to their population of origin (100% assignment success). Despite sharp genetic breaks between the three coastal regions, our study suggests that individuals do migrate between populations during the spawning season, as we identified ten individuals assigned to a different region than the one in which they were sampled. All potential migrants traveled counterclockwise around the Korean peninsula, from a more western to a more eastern population. These migrants caused a strong downward bias in our estimates of effective population size, such that excluding migrants increased N_e . However, effective population sizes remained relatively low ($< 3,000$) across sampling sites and populations.

1.5.1 *Potential influence of currents and glacial history on stock structure*

The levels of genetic differentiation between regional populations of Korean cod are surprisingly high for a marine fish over such short distances. Pairwise F_{st} among Korean cod stocks is higher than that among eastern Pacific cod spawning sites, which are spread over a much larger geographic area. Recent RAD sequencing along the west coast of the United States and Canada found that eastern Pacific spawning sites show an isolation-by-distance pattern, with a pairwise F_{st} (0.0013 – 0.0267) that is up to ten-fold smaller than F_{st} between coastal regions on the Korean peninsula (Drinan et al., 2018). For example, the pairwise F_{st} between aggregates at Jukbyeon and Pohang, which are approximately 100 kilometers apart, was 0.038. In contrast, pairwise F_{st} was only 0.0044 between the Kodiak and Adak sampling sites in the eastern Pacific, despite a geographic distance of approximately 1700 kilometers. This large genetic divergence over short geographic distances may result from a combination of oceanographic barriers to gene flow and lasting signatures of the species' phylogeographic history in the region.

Korean Pacific cod stocks likely underwent historical isolations and expansions from Pleistocene glacial-interglacial cycles. A combination of phylogeographic and palaeoceanographic research suggest that marine organisms inhabiting the two marginal seas around the Korean peninsula, the Yellow Sea and the East Sea, may have been isolated in separate glacial refugia. When sea levels fell 120-140m during late Pleistocene glaciations, the Yellow Sea and the East China Sea were reduced to the Okinawa Trough, with the shallow continental shelf of the Yellow Sea completely exposed (Xu et al., 2009; Park et al., 2000; Wang, 1999). The East Sea became a semi-isolated inshore basin, as the southern and northern entrances (Korea and Tsugaru Straits, respectively) experienced narrowing and/or shallowing (Oba et al., 1991). Recent studies propose that the deeper western channel of the Korea Strait,

along the southern coastline of the Korean peninsula, remained open during the last glacial maxima 15,000 – 18,000 years before present (Morley, Heusser, & Sarro, 1986; Park et al., 2000; Tada, 1999). However, inflow of the paleo-Tsushima Current was likely reduced by 95% of present day levels, and sill depth in the channel is believed to have been approximately 10 meters (Park et al., 2000). As a result, it is probable that this area was not inhabited by Pacific cod during the last glacial maxima. After the last glacial maxima, rising sea levels re-submerged the Yellow Sea, with the coastline migrating 1200km landward within 8000 years (Wang, 1999), and flooded the eastern and western channels of the Korea Strait to establish the present-day circulation patterns of the Tsushima Current. Thus the eastern and western populations of Korean Pacific cod may have been isolated in refugia in the East Sea and the Okinawa Trough, subsequently expanding after the last ice age to their current range. Such isolation during the last glacial maxima could have been pre-dated by several periods of population contraction and expansion during previous periods of glaciation. Two to four Pleistocene glacial maxima associated with equally dramatic falls in sea level (Waelbroeck et al., 2002) occurred after Pacific cod's expansion into the northwest Pacific, which was placed at either 102-339 kyr or ~500 kyr, based on nucleotide mismatch distributions and Bayescan skyline plots, respectively (Canino et al., 2010).

The phylogeography of other marine fauna in the region also reflect this glacial history. In a review of phylogenetic studies on marine fauna inhabiting the northwest Pacific's network of marginal seas, Ni et al. (2013) identified seven out of eleven marine species (two out of four finfish species) which displayed spatial structure reflective of ice-age separation in the independent glacial refugia of the East Sea (Sea of Japan) and the East China Sea. In studies

where time since divergence was calculated between population pairs, divergence times ranged from as early as 2 mya to as recently as 70 kya (Ni et al., 2013).

We can therefore hypothesize that periods of isolation and population contraction in Pacific cod around the Korean peninsula were driven primarily by falling sea levels, which narrowed the Korea Strait and exposed the Yellow Sea, in addition to the glaciation of much of the northern East Sea. Patterns of extirpation and recolonization in eastern Pacific cod were driven by similar forces; the exposure of large areas of the continental shelf in the Bering Sea and the northeast Pacific, in addition to low temperatures and the expansion of terrestrial glaciers into coastal areas, forced cod out of the Bering Sea, Aleutian Islands, and Gulf of Alaska, into southern refugia (Canino et al., 2010). However, temperature-mediated ice age range shifts played a greater role in the eastern Pacific than is likely around the Korean peninsula.

Present-day oceanographic currents may be reinforcing historical differentiation between Korean cod populations, and driving adaptive divergence in the southern population, by creating large differentials in sea surface temperature. The Tsushima Warm Current carries warmer, more saline waters along the southern coast before branching into the East Korean Warm Current, which continues up the southeast coast until it separates from the coastline and moves eastward (Chang et al., 2004; **Figure 1.12**). While the northern extent of this current oscillates throughout the year, sea surface temperature data from 2006 compiled by Chung et al. (2013) shows that during the winter months, the northern edge of the current can reach as far north as our sampling site at Pohang. In contrast, the southward-flowing North Korean Cold Current (East Sea) and Korean Coastal Current (Yellow Sea) are associated with colder temperatures on the eastern and western coasts, respectively (Chang et al., 2004; Hwang et al., 2014; **Figure 1.12**). For example, seawater temperature data extracted from the Korea Oceanographic Data Center by Chung et al.

(2013) showed February, 2006, near-coast temperatures ranging from 0-4°C at 100m depth along the eastern coast, as opposed to 13-16.9°C at 75m depth in Jinhae Bay. Since sea surface temperature during and after the spawning period are important to reproductive success in Pacific cod, and spawning temperature ranges vary across their distribution (Chung, Kim, & Kang, 2013; Gustafson et al., 2000), the warmer waters of the Tsushima and East Korean Warm Currents may be maintaining high genetic divergence along the southern coast (Gwak & Nakayama, 2011).

Similar isolation of Pacific cod spawning aggregates that form within the warm waters of the Tsushima Current was found around the coast of Japan, using microsatellite and mitochondrial DNA (Suda et al., 2017). Suda et al. (2017) sampled 16 spawning sites along the Japanese coastline, and found that the only two sites influenced by the Tsushima Warm Current were genetically distinct from the remaining sites. Pairwise F_{st} between the spawning aggregates in the Tsushima Warm Current and the remaining aggregates on the Japanese coastline was similar to the between-region F_{st} observed along the Korean coastline by Gwak & Nakayama (2011). More extensive population structure analysis with samples from both Korean and Japanese waters within the Tsushima Current would provide further insight into the roles of local adaptive evolution and demographic history – namely, the potential expansion of the same distinct, historical lineage – in genetic divergence between Pacific cod spawning sites within the Korea Strait.

1.5.2 *Discrepancies in stock structure among studies*

Our finding of three regional stocks confirmed the results of Gwak & Nakayama (2011), who discovered a similar regional population structure using five microsatellite loci, while challenging the conclusion by Kim et al. (2010) that there are only two genetically distinct

populations around the Korean peninsula. The degree of differentiation between southern, eastern, and western coast populations detected by RAD loci (pairwise F_{st} between sites = 0.025 – 0.042), and by microsatellites by Gwak & Nakayama (2011), makes it highly unlikely that spawning fish from these three coasts belong to only two populations.

This discrepancy in regional scale stock structure between studies is probably due to sampling bias – that is, the stock structure found is dependent on the group of fish that were actually sampled, and sampling Pacific cod aggregates during the spawning season is not enough to ensure that we are collecting mature, spawning fish on their resident spawning ground. It can be difficult to assess Pacific cod maturity and spawning condition in the field, and to exactly time sampling dates to collect ripe individuals. If the sampled aggregate is a pre-spawning aggregate of mixed-composition, or if there is migration of mature individuals between sites during the spawning season, measured stock structure may not reflect true genetic divergence between populations. Both our study as well as Gwak & Nakayama (2011) found strong evidence that Pacific cod may migrate between regions during the spawning season.

For example, Kim et al. (2010) did not find differentiation between the eastern and southern coasts. The data collected by Gwak & Nakayama (2011) showed a mixed-composition aggregate containing both eastern and southern coast individuals at the only sampling site on the northeast coast, Jumunjin, in the Kim et al. (2010) study. Kim et al. (2010) likely sampled southern coast fish from this aggregate ($n = 13$), which would not have been genetically different from fish sampled at their two southeastern sites (**Figure 1.2**), if the southern / eastern population border is indeed near Pohang as our data suggest. Kim et al. (2010) also did not find a significant difference between the southern and western coasts. Our data suggest that a surprising proportion of Pacific cod move from the western to the southern coast during the spawning

season (discussed in further detail below). It is possible that Kim et al. (2010) sampled a group containing at least some individuals which originated on the western coast, and so their samples were genetically indistinguishable from fish collected at their west coast sampling site.

1.5.3 *Are migrants individuals that skip spawning?*

A total of ten individuals were identified as potential between-population migrants, all moving counterclockwise (west to east) around the Korean peninsula. Migration distance along the coastline between any of these populations, even from the eastern sampling site at Jukbyeon to the western coast, is well within the more than 1000 kilometers that mature Pacific cod have been found to travel during the non-spawning season (Gustafson et al., 2000; Shimada & Kimura, 1994). It is therefore reasonable to expect Pacific cod to travel around the coastline of the Korean peninsula. However, such migration of mature individuals during the spawning season conflicts with the level of genetic divergence between the three Korean stocks of Pacific cod.

According to the Wright island model, assuming that the populations are at migration-drift equilibrium, $F_{st} \approx \frac{1}{4N_e m + 1}$ (Wright, 1965). With an average $F_{st} = 0.036$ between western and southern sampling sites, and the harmonic mean $N_e = 1161$ between the southern and western populations during the 2014-2015/2015-2016 spawning seasons, this implies a migration rate of 0.0058, or 6.7 migrants arriving per generation in the southern population. Applying this migration rate to our 65 individuals collected from southern sites during the 2014-2015 spawning season, we should have encountered 0.37 immigrants. We instead found five mature immigrants that season (**Table 1.5**), or a migration rate of 0.077.

This implies that the observed migrants do not contribute to gene flow, either by skipping spawning, or as a result of strong selection against immigrants and their offspring. Skipped spawning, in which members of an annually reproducing species do not reproduce for a given year, was recently reported to be far more ubiquitous in marine fish species than previously believed (Rideout & Tomkiewicz, 2011). The behavior has been observed in populations of Arctic cod (Skjaeraasen et al., 2012) and Atlantic cod (Rideout, Morgan, & Lilly, 2006; Rideout, Burton, & Rose, 2000); up to 18% of potential female spawners in northwestern populations of Atlantic cod were found to be non-reproductive, with as much as 33% of mature females inactive in one survey area (Rideout, Morgan, & Lilly, 2006). The alternative is that migrants are spawning, but that local adaptation within coasts limits their contribution to the next generation. Reduced reproductive success due to local adaptation has been observed in sockeye salmon, where dispersers from stream- and beach-spawning ecotypes produced half as many offspring in the alternative habitat (Peterson, Hilborn, & Hauser, 2014). Local adaptation between Pacific cod populations in this study is discussed further below, in relation to outlier analysis.

These competing theories can be preliminarily explored through the gonadosomatic indices (GSI) calculated for four of the mature migrant individuals, which suggest that two fish may be in spawning condition and the remaining two are not. Mean GSI for female cod in Japanese waters was 11.47 - 24.79% at the maturation phase and 24.46 - 33.81% at the spawning phase when GSI was calculated using total body weight (Hattori, Sakurai, & Shimazaki, 1992). According to the Hattori et al. (1992) maturity phases, with GSIs of 5.03% and 4.23%, two migrants can be classified as either developing, phases III – IV, or spent, having already released their eggs during spawning. According to histological criteria developed for Atlantic cod in the Baltic Sea, individuals which are skipping spawning, or “resting,” contain oocytes similar to

early development phases (Tomkiewicz, Tybjerg, & Jespersen, 2003). The third and fourth individuals had GSIs within and just below the maturation phase, respectively. If the two migrants with low GSI are spent, then it appears that migrants are spawning, but not contributing to the next generation. If the two migrants with low GSI have skipped spawning, there may be a mix of spawning and non-spawning migrants, and so true migration rate is lower than would be suggested by 50% length-at-maturity cutoff alone.

To determine which of these scenarios best describes Korean cod spawning behavior, additional research should include the collection of histological samples for more direct assessment of spawning condition. Further investigation into ontogenetic shifts in winter migration behavior, from juveniles to older adults, could also provide insight into whether a certain subset of the population is moving between coasts during the spawning season.

1.5.4 *Low effective population size*

Since the linkage disequilibrium (LD) method provides an estimate of effective population size based on the nonrandom associations of alleles at different loci (Waples & Do, 2010), the introduction of a migrant from a highly diverged subpopulation can downwardly bias the estimate of contemporary N_e . While simulations indicate that the LD method is relatively robust in the face of migration between subpopulations, N_e was affected when equilibrium migration rates were greater than 0.05-0.10 (Waples & England, 2011). Migration rates at our sampling sites range from 0.018 to 0.125, so we removed migrants and recalculated effective population size where appropriate.

Estimates of effective population sizes were remarkably low for a marine species that supports a commercial fishery. Even the larger uncorrected estimates of N_e were smaller than the uncorrected effective population sizes in the northeastern Pacific, including the Salish Sea

population segment deemed a NOAA Species of Concern. In the eastern Pacific, N_e ranged from approximately 2800 to 30,750, with finite upper confidence limits reaching 23,000 and one infinite upper limit (Drinan et al., 2018). Interestingly, our effective population size estimates, especially at the aggregate level, are more similar to recent estimates of N_e in the Gilbert Bay stock of Atlantic cod, a highly isolated subpopulation which has experienced a recent decline in biomass (Sinclair-Waters et al., 2018). The effective population sizes observed in Korean cod stocks are well within the range where genetic drift can reduce standing diversity within the population, threatening the population's adaptive potential in the face of environmental change.

Effective population size two to five orders of magnitude smaller than census size is more common than expected in marine species with high fecundity (Hauser & Carvalho, 2008). While relevant demographic factors like increasing fecundity with age and age-at-maturity are unlikely to have large effects on N_e estimates on their own (Waples, 2016), a high ratio of variance to mean reproductive success may be one of the most important ecological factors in reducing N_e in long-lived, highly fecund marine species (Hauser & Carvalho, 2008; Hedgecock & Pudovkin, 2011). Theoretical estimates show that having only a few successful breeders in a single generation will produce the “tiny” N_e/N observed in wild populations (Hedrick, 2007), and simulations of 100 loci have specified a ratio of 1 in 1000 adults successfully reproducing as the minimum needed to observe such an effect (Waples, 2016).

It is also possible to estimate small effective population sizes even if true N_e is large and the N_e/N ratio is close to 1.0 (Waples, 2016). Simulations of 100 loci show that when true effective population size is large ($N_e \sim 10^4 - 10^6$), the distribution of point estimates of N_e move toward a bimodal distribution where nearly all estimates are either infinity or very small (Waples, 2016). Our estimates of effective population size fall within the “sweet spot” of low N_e

estimates when true N_e is large, which ranges from the hundreds to low thousands (Waples, 2016). While these simulated point estimates of N_e did have upper bounds inclusive of the true effective population size, it is unlikely that the 95% confidence intervals in our data encompass the true effective population size of Korean Pacific cod stocks. A study by Waples et al. (2016) found that when estimating effective population size with thousands of loci, as in RAD sequencing, elevated linkage disequilibrium from physically linked loci will cause gross overestimation of the precision of point estimates by artificially narrowing 95% confidence intervals; with 4096 loci on eight chromosomes, the empirical 95% confidence interval was 20 times wider than the calculated 95% confidence interval. In our study we had 5804 loci on 23 chromosomes, assuming that Atlantic and Pacific cod share the same chromosome number. Although we applied the least-squares regression correction described by Waples et al. (2016) as an effective way to counteract downward bias in point estimates of N_e , this bias correction did not widen our 95% confidence intervals.

We also must take into account the sensitivity of N_e estimates to sampling bias. For example, although the Jinhae Bay samples were not significantly differentiated, effective population size changes by a factor of three between December and February of the same spawning season (**Table 1.3**). Similarly, we see an unexpectedly large increase between N_e estimates for the entire southern population within just one year, from the 2013-2014 to the 2014-2015 spawning season (**Table 1.6**). It is possible that we sampled slightly different groups of Pacific cod at Jinhae Bay and between the two spawning seasons for the southern sites which, while not significantly diverged, could have led to different point estimates of N_e . Alternatively, there may be undetected population structure within these groups, which has been shown to

reduce effective population size when subpopulations have very different levels of productivity (Turner, Wares, & Gold, 2002).

With this in mind, we can draw some tentative conclusions regarding change in effective population size between aggregates and years. The southern population of Korean cod experienced a four-fold increase in effective population size between the 2007-2008 and 2014-2015 spawning seasons. As contemporary N_e is sensitive to sudden changes in population abundance and has been shown to decline over time in exploited populations (Hauser et al., 2002), this increase may be due to stock rebuilding efforts after overfishing of Pacific cod in the region. In contrast, effective population size decreased by half in the western population between 2007-2008 and 2015-2016. While this may in turn suggest that fishing effort was causing declines in the western population, the same effect could be due to other factors described above, including sampling bias.

1.5.5 *Evidence for selection*

Overall, Bayescan identified more outlier loci than OutFLANK when the full data set was used. The discrepancy in the number of outlier loci identified could be a result of the hierarchical population structure in this region, which has been shown to increase the rate of false positives in Bayescan (Lotterhos & Whitlock, 2015). In contrast, OutFLANK identified a greater number of outlier loci (27) than Bayescan when both programs were run on a subset of the data only including southern sampling sites, which consisted of six aggregates. Bayescan only identified a single outlier locus between southern sites, Locus 24927.

Through additional iterations of Bayescan and OutFLANK focusing on this locus, we found that it showed significant evidence for selective differentiation between Geoje, spawning season 2014-2015, and all other southern sites – including the Geoje temporal replicate taken

during the 2013-2014 spawning season. This locus then aligned within a protein-coding sequence in the Atlantic cod genome which is similar to the human gene PRKAB1, a 5'-AMP-activated protein kinase (subunit beta-1) that monitors cellular energy status, and is activated by cellular metabolic stress (Monteverde et al., 2015; Sanchez-Cespedes, 2008). It is possible that the sample at Geoje during the 2014-2015 spawning season simply had a different composition of individuals than the sample at Geoje during the 2013-2014 spawning season. Such sampling bias could have resulted from low sample size in a mixed-composition aggregate, or the collection of samples during different months of the spawning season (January during the 2014-2015 season, February during the 2013-2014 season). This would align with some of our population structure analyses, which show significant differentiation between this sample and all other southern sites across years in strataG permutation tests, as well as separation of this sample from all other southern sites in a DAPC plot. It is also possible, although much less likely, that a strong selection event occurred in spawning fish at Geoje between the two spawning seasons that were sampled, which provided a selective advantage to individuals with a certain Locus 24927 allele.

The only two outlier loci in the full data set which aligned within protein-coding regions, and were identified by both OutFLANK and Bayescan, encoded for proteins related to reproductive activities. The first aligned within KMT2b, a methyltransferase that plays a role in oocyte growth and is required during a period of global transcriptional silencing that precedes oocyte survival and normal zygotic genome activation (Andreu-Vieyra et al., 2010; Dahl et al., 2016; Xu & Xie, 2018). The second aligned within DHX38, an RNA helicase involved in pre-mRNA splicing that affects embryogenesis, spermatogenesis, and cellular growth and division (Ajmal et al., 2014; Bourgeois, Mortreux, & Auboeuf, 2016; Schwer & Guthrie, 1991). A third

outlier locus, identified only by Bayescan, also aligned within the KMT2b gene region. These outliers suggest local adaptations between coastal regions may be linked to reproductive success and survival of early life history stages. Adaptations which impact fitness and survival during early life history stages would confer a selective advantage during a “critical period” (Hjort, 1914) of high mortality rates.

Some of the significantly diverged loci linked to reproduction and early life history survival could be in response to selection pressures from different thermal and salinity regimes. As described above, the Tsushima Warm Current and its northern branch, the East Korean Warm Current, carry warmer and more saline waters along the southern / southeastern coast throughout the spawning season and during early development. For example, post-spawning, larval Pacific cod in nursery areas of Jinhae Bay are believed to face temperatures of up to 16°C (W. Gwak, Gyeongsang National University, personal communication). Populations of Atlantic cod distributed along a salinity gradient have been found to possess outlier SNPs which lay within or nearby genes associated with osmoregulation and the hydration and development of oocytes (Berg et al., 2015). In addition, experiments in Atlantic cod determined that sperm swimming characteristics under different thermal regimes were genetically based, and that individuals from southern populations exhibited more phenotypic plasticity in sperm swimming characteristics under warmer temperatures (Beirão et al., 2014).

1.5.6 *Management applications and future directions*

The results of this study are highly informative for management of Pacific cod stocks around the Korean peninsula, particularly in light of declining catch. Through the use of next-generation sequencing to obtain thousands of genetic markers, as well as greater geographic and temporal breadth of sampling sites, we provided a more complete characterization of the

population structure of Korean Pacific cod stocks. Our results suggests that Pacific cod around the Korean peninsula should be managed as three separate stocks along the western, southern, and eastern coasts. Future hatchery efforts could maintain this baseline population structure by sourcing any broodstock from spawning sites on the coast where juveniles will be released, to avoid disruption of population structure and the resulting negative impacts on genetic diversity. However, further exploration of subpopulation structure within the western and eastern coasts is warranted; although our study included the most sampling sites to date, only one site was sampled from the eastern population, and only two from the western population. Additional sampling in the western population would clarify whether aggregates at spawning grounds show undetected within-population differentiation, as suggested by discordant statistical results.

Fishery managers should also closely monitor effective population size, which our study found to be quite small even when corrected for linkage between loci. While a number of other factors may bias estimates of effective population size and lead to deceptively narrow confidence limits, our results suggest a relatively low N_e in Korean Pacific cod stocks because of the consistency of small point estimates. Given the history of overfishing in the region and the trends in effective population size that we observed in the western and southern population, research into long-term effective population size to establish a pre-harvest baseline may be warranted to better understand the impacts of fishing on these stocks.

Another critical direction for further research is between-population migration during the spawning season. We observed ten migrant individuals in this study; in addition, Gwak & Nakayama (2011) have reported mixed-composition aggregations during the spawning season, and we believe that the different stock structure found by Kim et al. (2010) may be a function of between-region migration. A more thorough understanding of Pacific cod migration behavior in

this region would contribute not only to our understanding of the species as a whole, but also to the management of these Korean cod stocks during the spawning season.

One tool to make such research more feasible would be the development of SNP panels for Genotyping-in-Thousands by sequencing (GT-seq) (Campbell, Harmon, & Narum, 2015). Our study suggests that a panel of 100 RAD loci would be sufficient for highly accurate individual assignment to populations of origin using GT-seq. By reducing the cost of genotyping, GT-seq makes it feasible to conduct individual assignment for a wider range of management and research applications. Aside from additional research on migration behavior of Pacific cod, this panel could be applied to mixed-stock analysis for Korean cod fisheries (McKinney, Seeb, & Seeb, 2017; Ruzzante et al., 2000) and improved traceability of Pacific cod fishery products (Ogden, 2008). A similar approach could identify hatchery-reared Pacific cod in the wild and in the fishery, to assess the impacts and effectiveness of stock-enhancement programs (Anderson & Garza, 2006).

1.6 FIGURES

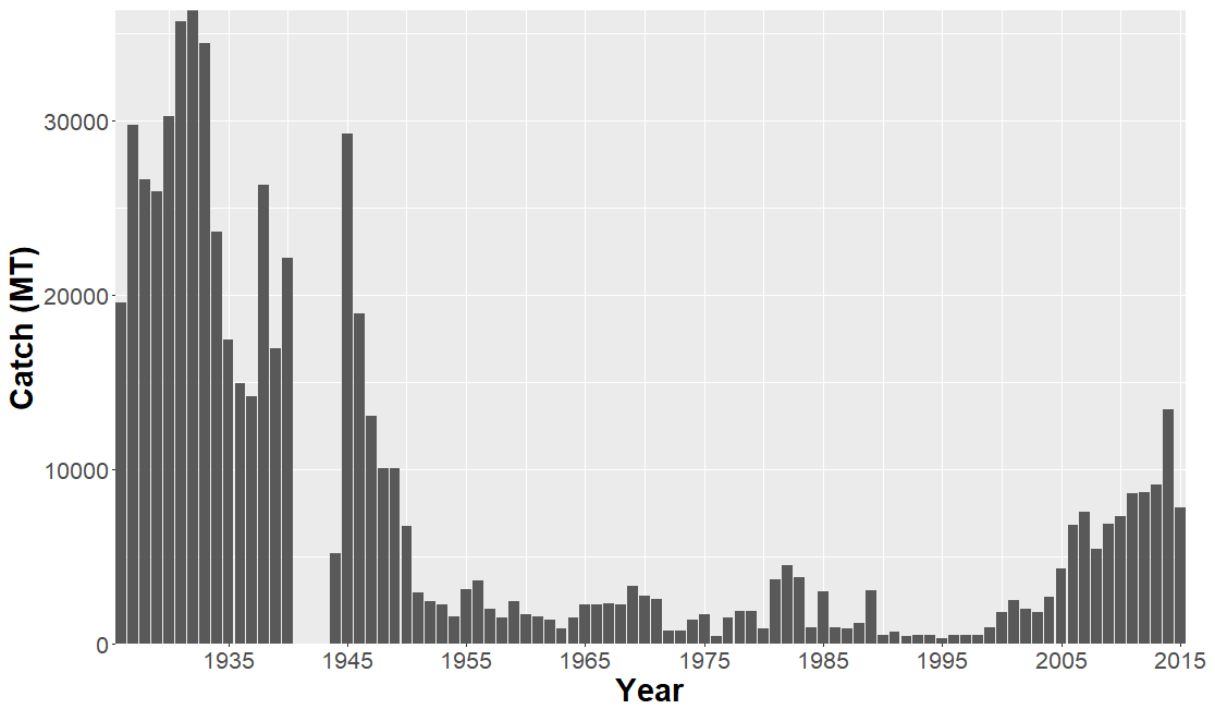


Figure 1.1. Pacific cod catch in Korean coastal waters, from 1926 to 2015, in metric tons. *Data provided by Dr. Sukyung Kang, National Institute of Fisheries Science.*

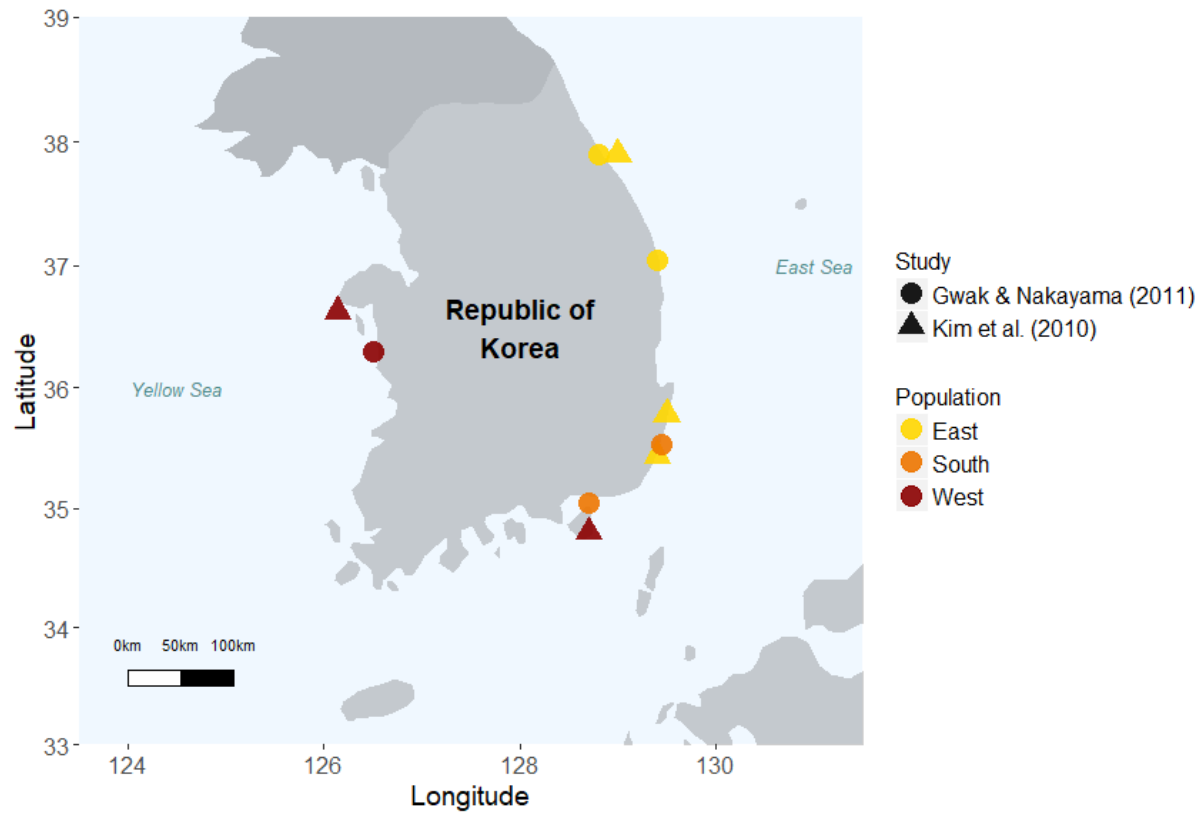


Figure 1.2. Sampling sites and their putative population membership from prior microsatellite studies around the Korean peninsula. Shape of point indicates study, color indicates the population to which the sampling site was assigned.

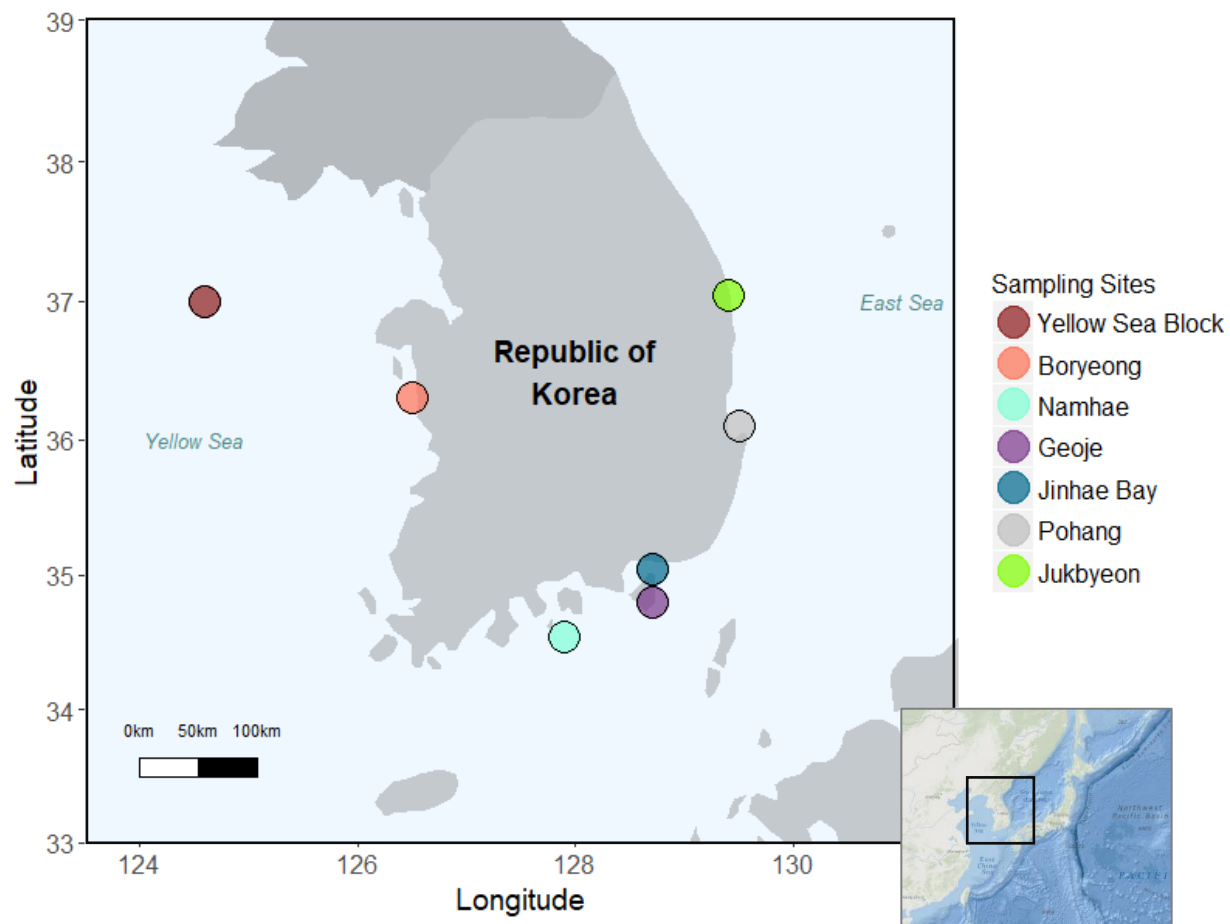


Figure 1.3. Pacific cod sampling sites around the Korean peninsula used for this study. Note that Geoje and Jinhae Bay were sampled twice for temporal replicates (see **Table 1.1**). Inset identifies location of Korean peninsula in the northwestern Pacific.

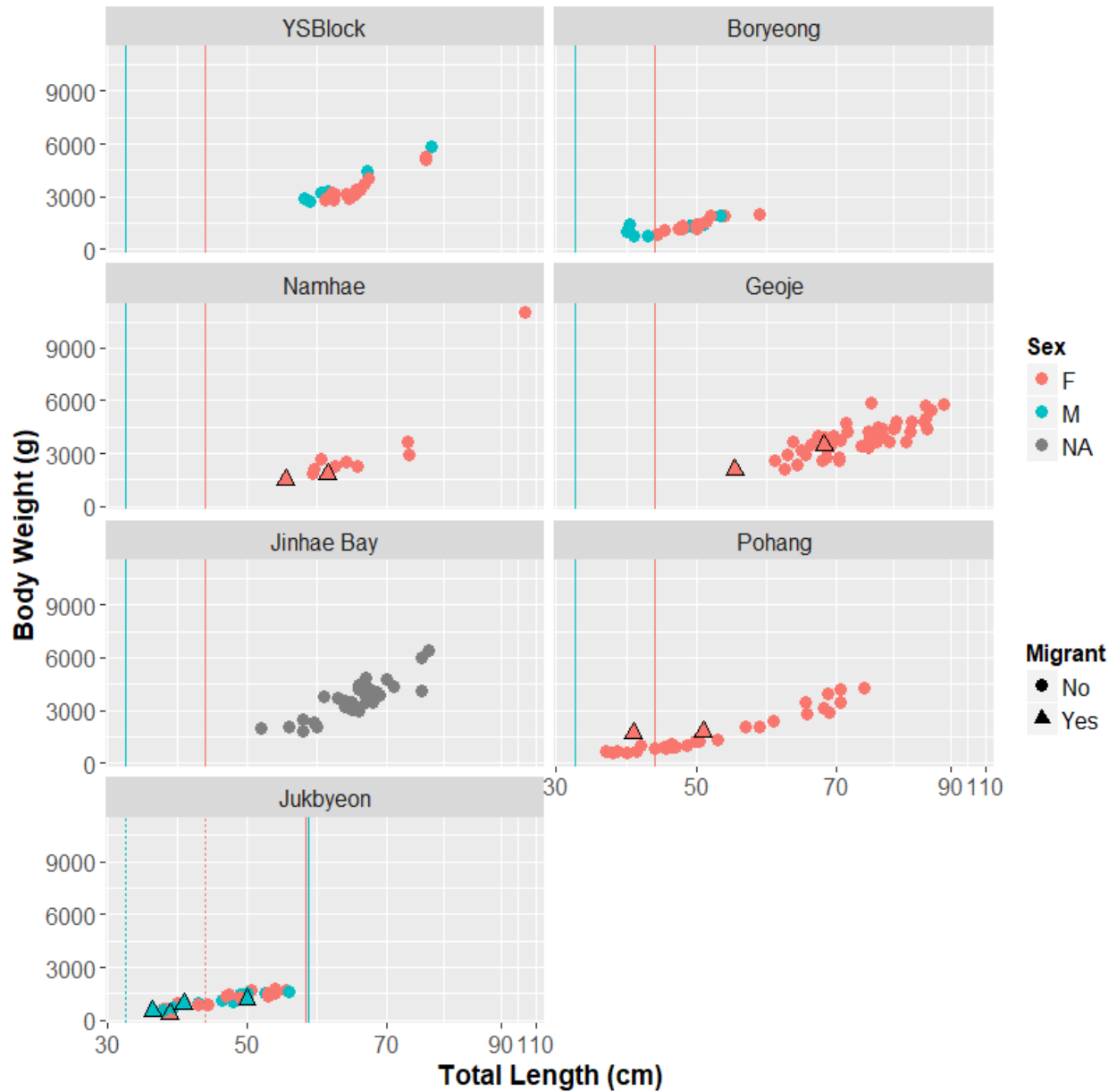


Figure 1.4. Distributions of sex-specific total length versus body weight at each sampling site. Both sampling years for Geoje are included, but only the early spawning fish from Jinhae Bay (Dec. 2007-2008) were measured. Sex data was not available for Jinhae Bay samples. Solid vertical lines indicate the sex-specific 50% length-at-maturity estimate for that region. Putative immigrant individuals are identified by triangle points. At Jukbyeon, the 50% length-at-maturity estimate for the immigrants' source population (western and southern coasts) differed from the estimate at that site, and so the 50% length-at-maturity estimate for the putative source population is depicted with dashed vertical lines. Note that the x axis from 90-110cm is collapsed for better visualization.

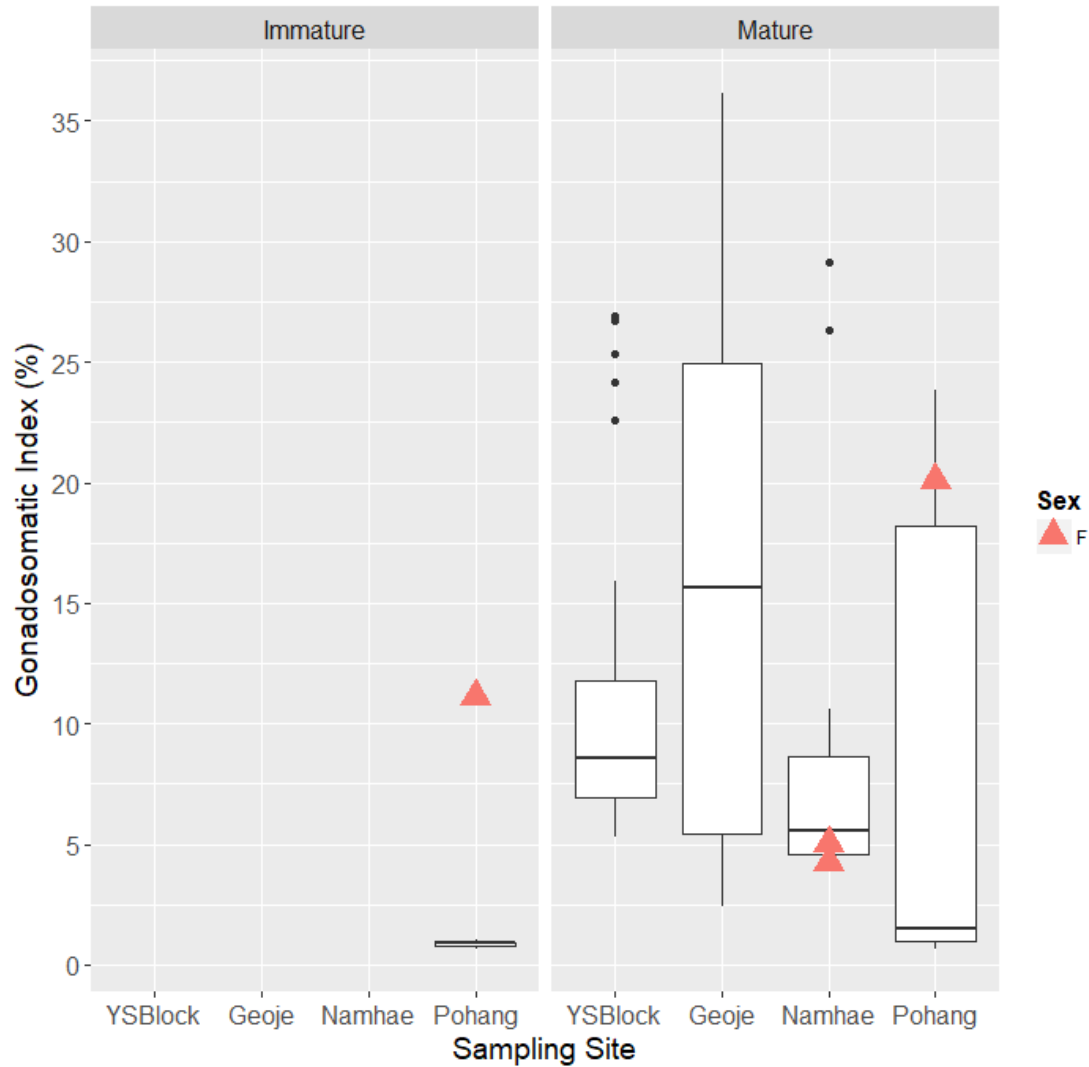


Figure 1.5. Boxplots of gonadosomatic index (GSI) at the four sites for which gonad weight was recorded for each individual. Maturity classification for each individual was determined according to sex-specific 50% length-at-maturity (see **Figure 1.4**). Overlaid triangles represent the gonadosomatic indices of putative immigrant individuals for which gonad weight was available.

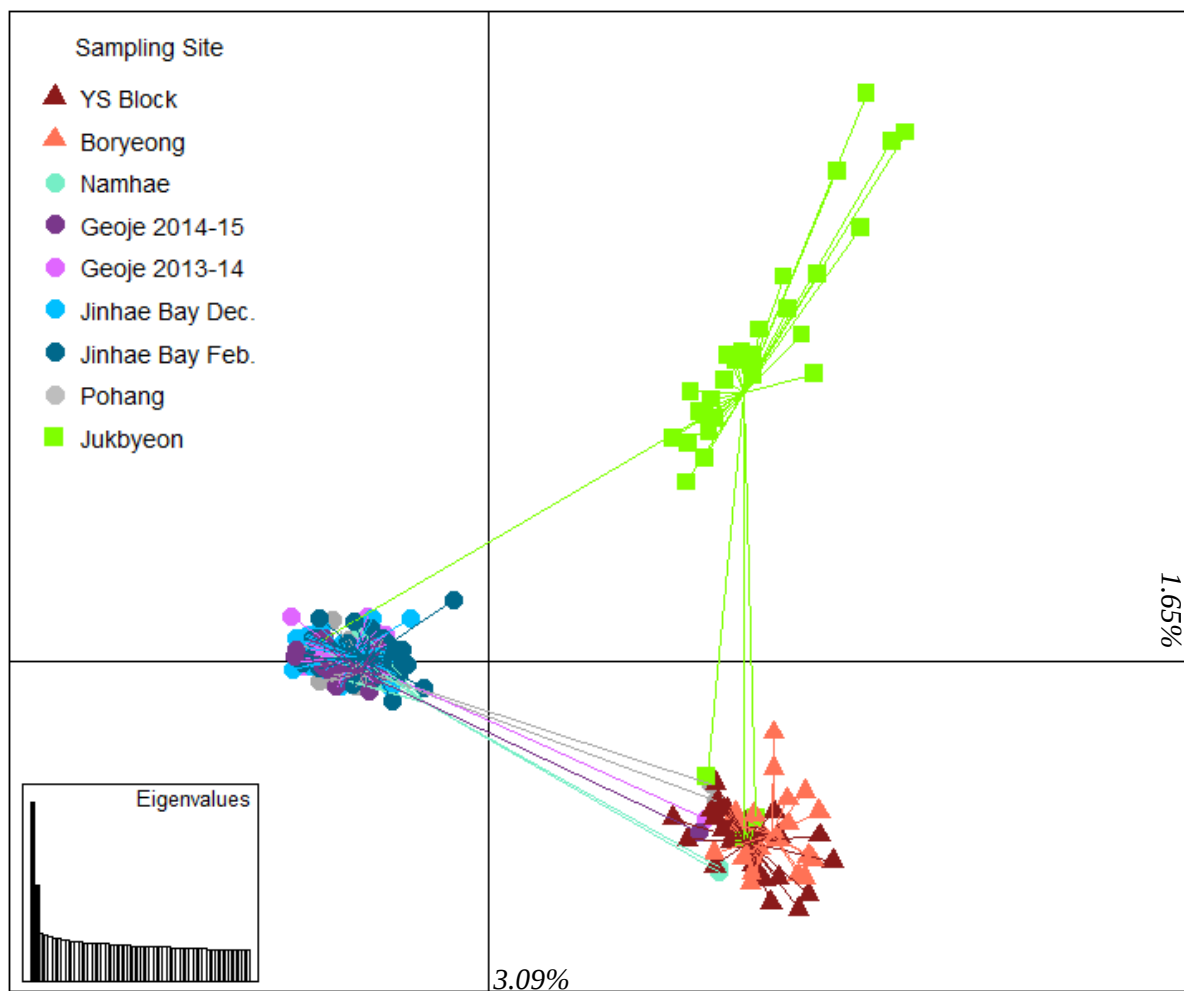


Figure 1.6. Principal component analysis of all sampling sites and temporal replicates. Color of point indicates sampling site, and matches colors used in **Figure 1.3**. Shape of point indicates region of coastline (west, south/southeast, east). Percentages given along the x and y axes describe the proportion of variation explained by the first and second principal component.

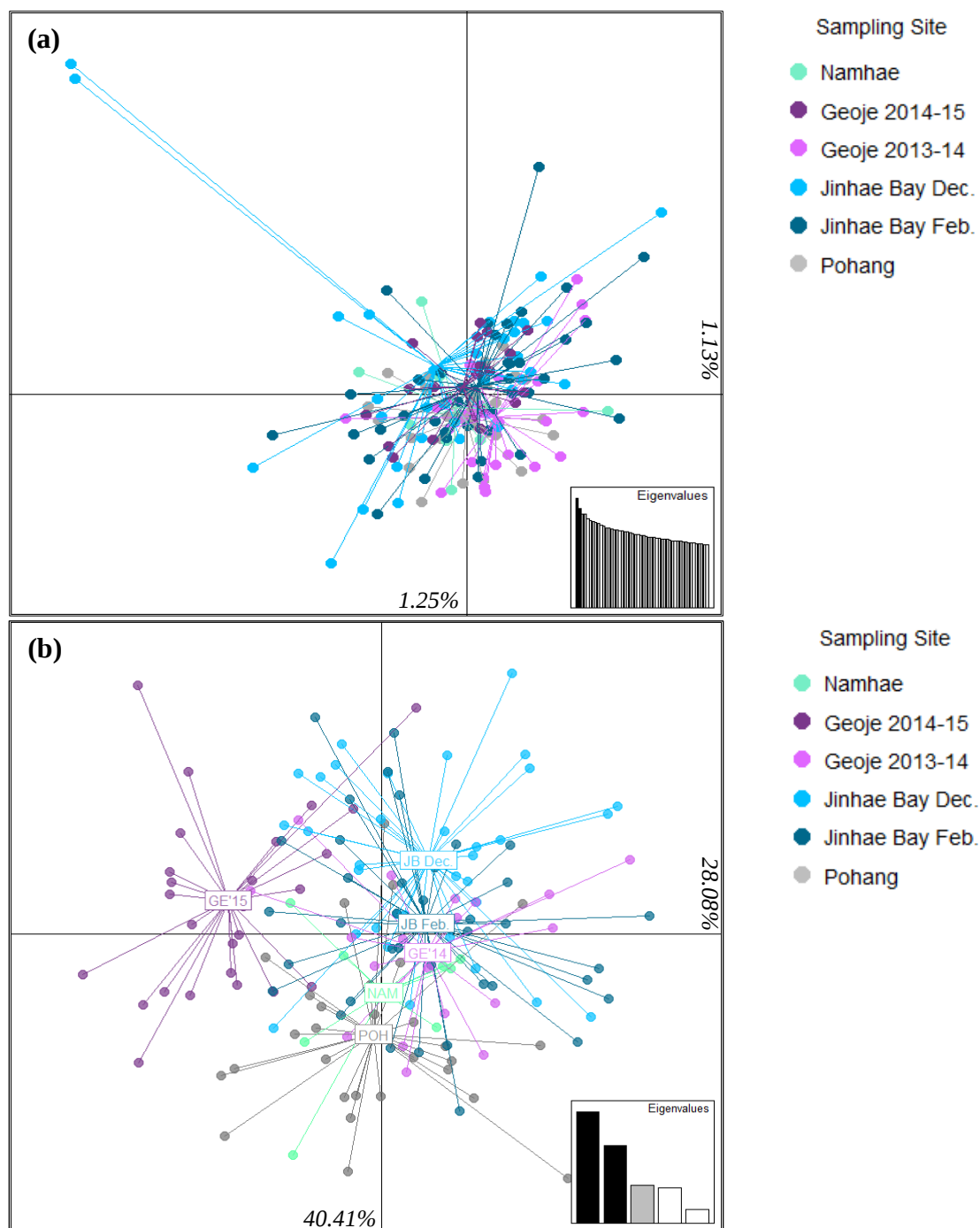


Figure 1.7. Southern sampling sites and temporal replicates analyzed using (a) principal component analysis, and (b) discriminant analysis of principal components (DAPC). The DAPC was conducted with 25 principal components. Individuals identified as immigrants in **Figure 1.6** were excluded from this analysis. Color of point indicates sampling site, and matches colors used in **Figure 1.3**. Percentages given along the x and y axes describe the proportion of variation explained by the first and second principal component.

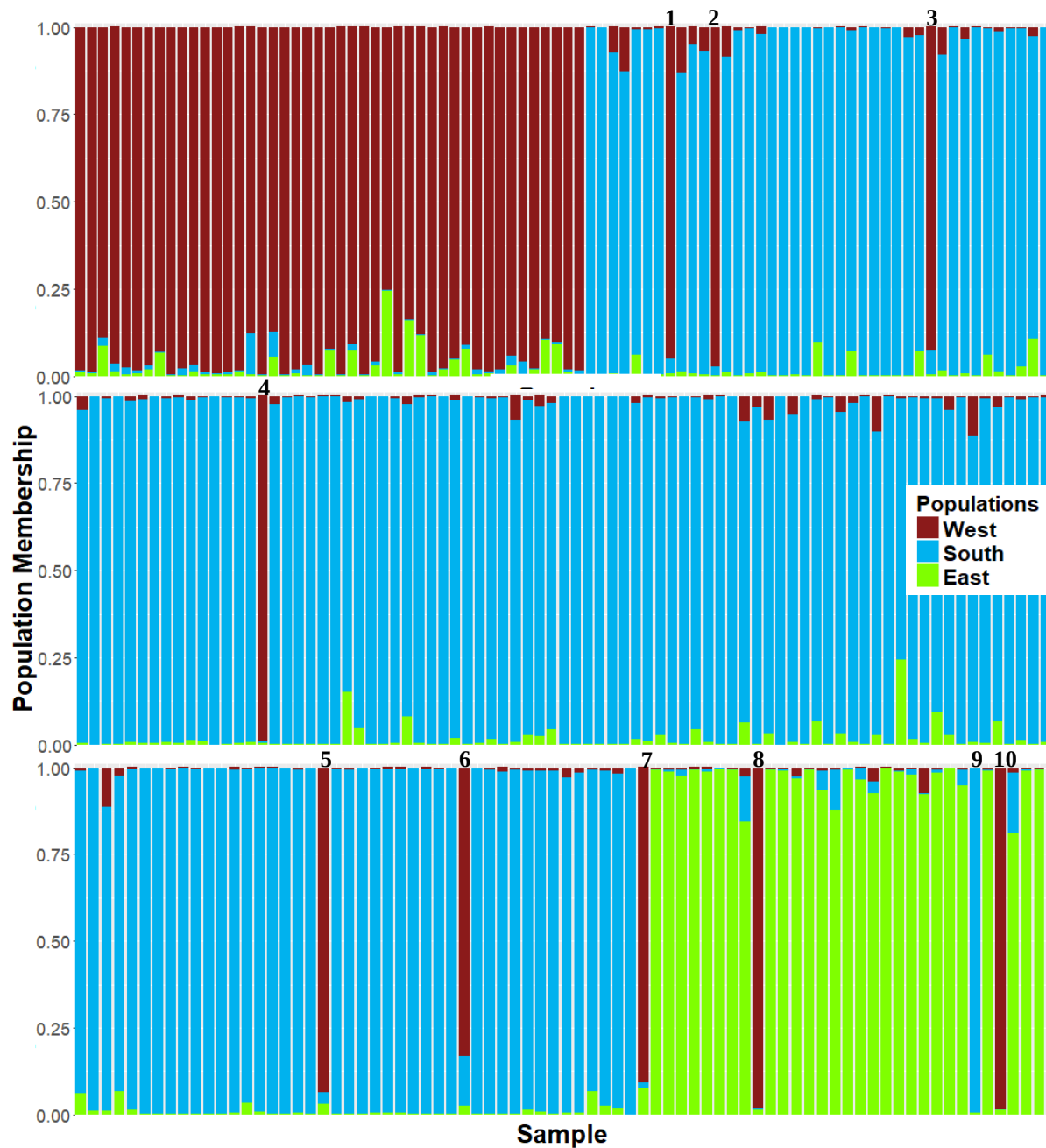


Figure 1.8. Proportional membership to each of the three populations (west, south, east) for every individual, estimated from STRUCTURE. The chart is continuous across the three separate rows, for easier visualization of individual results. Individuals with a majority population membership that is different than neighboring individuals (numbered at top) are the potential migrants listed in **Table 1.5**.

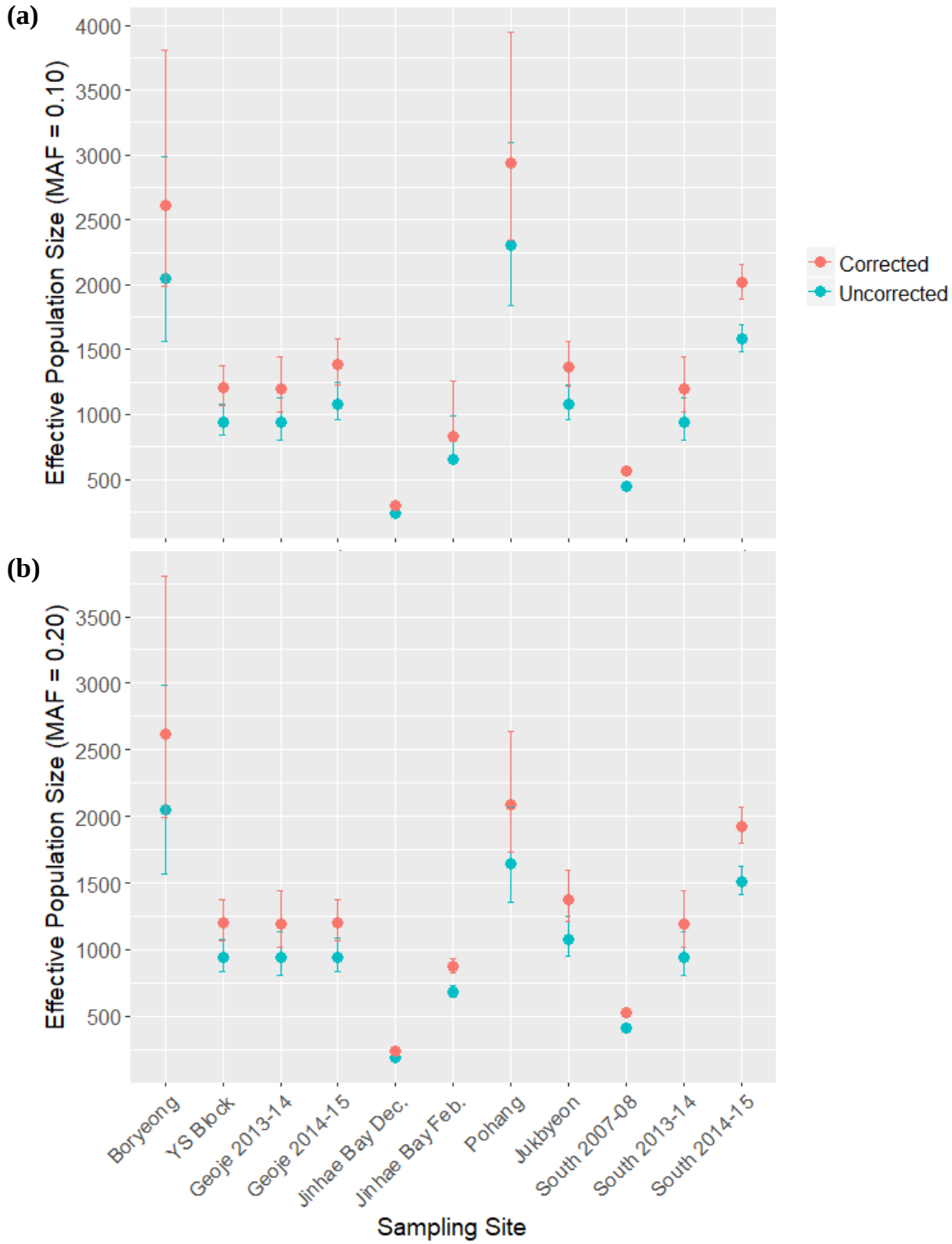


Figure 1.9. Effective population size after migrants were removed, with both naïve N_e and when N_e was corrected for downward bias from physical linkage. Estimates using the lowest minor allele frequency of **(a)** 0.010 and **(b)** 0.020 are shown with 95% confidence limits.

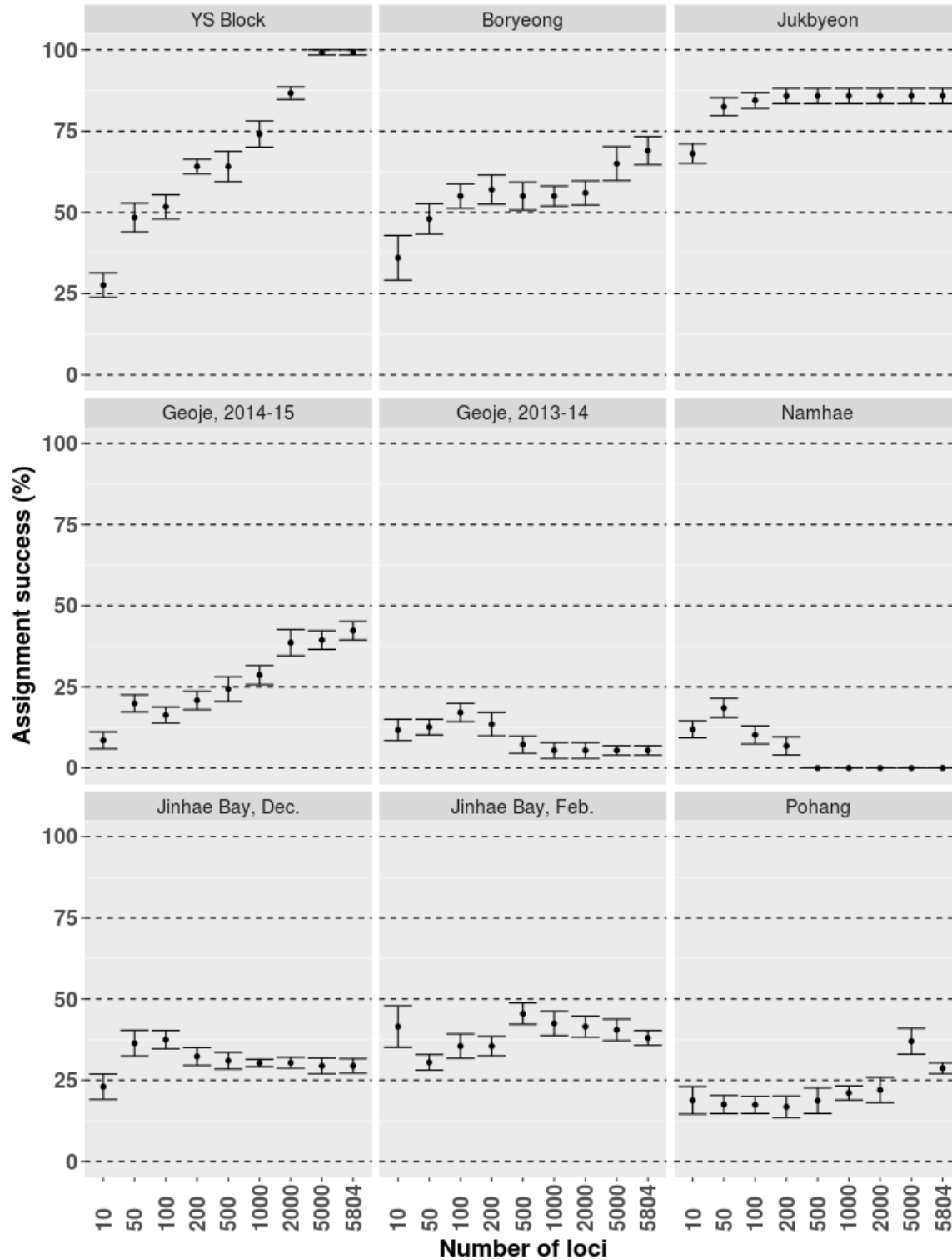


Figure 1.10. Assignment success (with standard error) when assigning individuals to aggregate of origin. Assignment success measured as the proportion of individuals sampled at each aggregate which were assigned back to that aggregate through genetic stock identification.

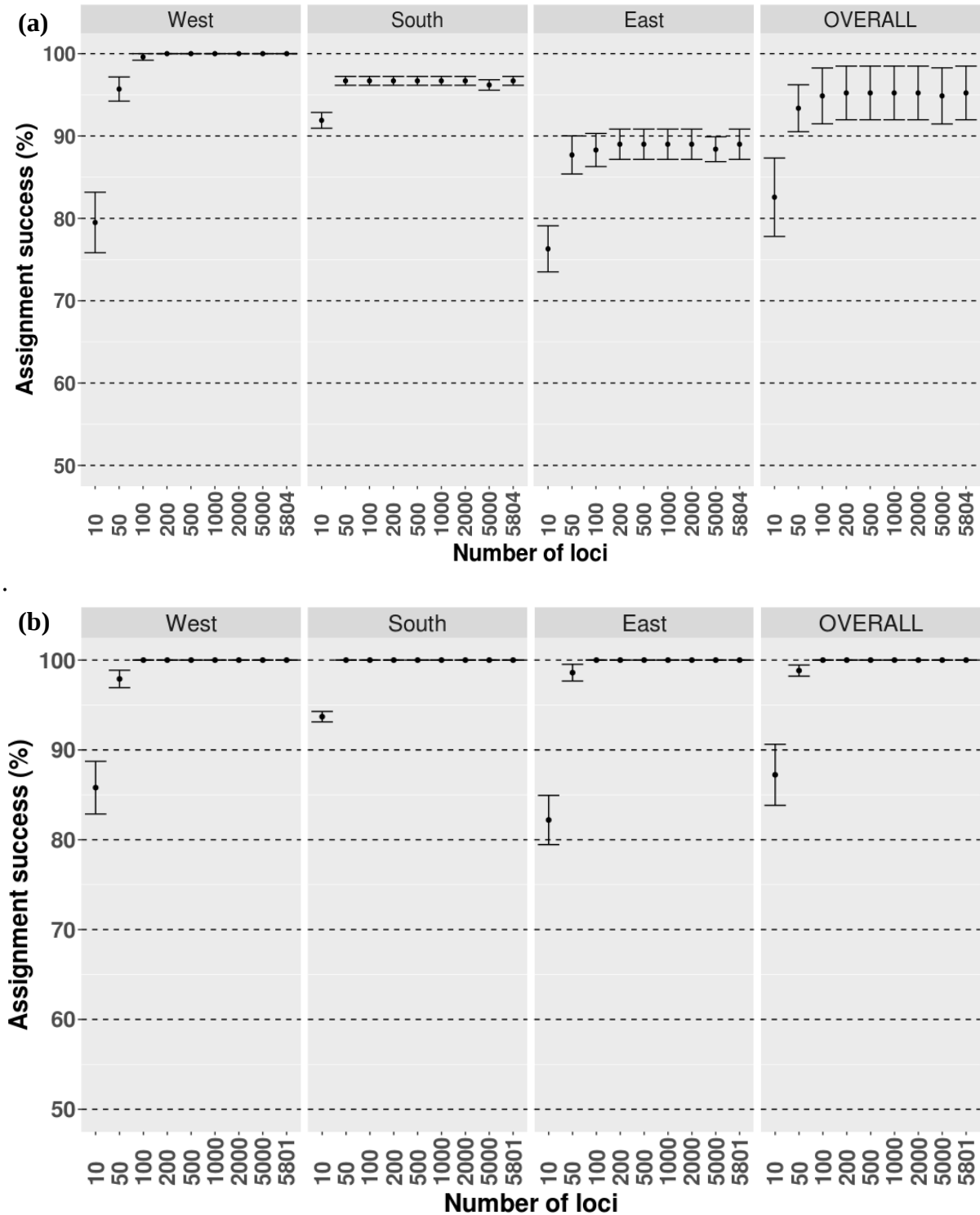


Figure 1.11. Assignment success (with standard error) when assigning individuals to population of origin, **(a)** including all samples and **(b)** excluding potential migrants. Assignment success measured as the proportion of individuals in each population which were successfully assigned back to that population through genetic stock identification.

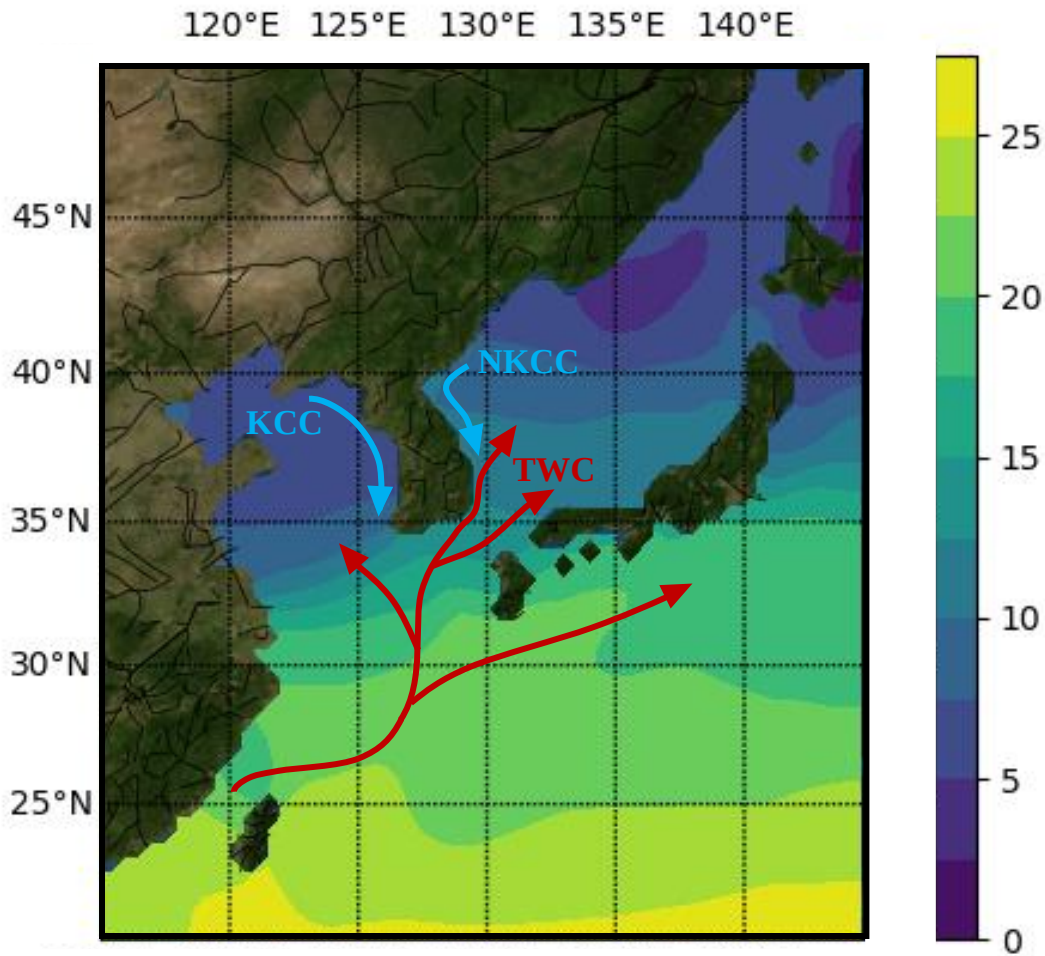


Figure 1.12. Pressure-gridded sea surface temperatures around the Korean peninsula during the month of February. Color scale indicates conservative temperature (TEOS-10) in degrees Celsius. Blue arrows mark the approximate pathways of the two cold water currents, the Korean Coastal Current (**KCC**) and the North Korean Cold Current (**NKCC**); red arrows mark the approximate pathway of the Tsushima Warm Current (**TWC**). Current locations were based on Park et al. (2015). Temperature data was extracted from the Monthly Isopycnal / Mixed-layer Ocean Climatology (MIMOC) database, v2.2, at a 0.5° lateral resolution (Schmidtke, Johnson, & Lyman, 2013).

1.7 TABLES

Table 1.1. Pacific cod tissue or fin clip samples from the Korean peninsula. Number of samples collected indicates total number of tissue or fin clip samples collected from each sampling site, whereas the number of samples retained includes only those samples used in final analyses. Samples removed from ‘collected’ to ‘retained’ include those with poor DNA quality, possible contamination, and too much missing data. Number of mature samples out of total samples retained estimated by comparison of individual total length to 50% length at maturity from Lee et al. (2016). *Samples provided by 1. Dr. Sukyung Kang, Korean National Institute of Fisheries Science (fin clip), and 2. Dr. Wooseok Gwak, Gyeongsang National University (tissue).*

Sampling Site	Collection Date(s)	Spawning Season	N. Samples Collected	N. Samples Retained	N. Samples Mature by Total Length
Yellow Sea (YS) Block ¹	12/13/2015	2015-16	30	25	25
Boryeong ²	1/24/2007	2007-08	22	21	20
Namhae ¹	2/10/2015	2014-15	16	11	11
Geoje ¹	1/12,23/2015	2014-15	34	28	28
	2/4/2014	2013-14	22	22	22
Jinhae Bay ²	2/11/2008	2007-08	48	39	39
	12/18/2007	2007-08	48	34	n/a
Pohang ¹	1/7/2015	2014-15	31	31	24
Jukbyeon ²	12/10/2007	2007-08	35	32	0

Table 1.2. Population genetic metrics for each sampling site, including temporal replicates.

Sampling Site	Spawning Season	Avg. Genotypes	F _{is}	H _o
Yellow Sea (YS) Block	2015-16	5578	-0.006	0.188
Boryeong	2007-08	5735	-0.020	0.194
Namhae	2014-15	5756	-0.020	0.185
Geoje	2014-15	5742	-0.092	0.204
	2013-14	5589	0.018	0.174
Jinhae Bay	2007-08, Dec.	5558	0.009	0.172
	2007-08, Feb.	5512	0.010	0.174
Pohang	2014-15	5751	-0.060	0.190
Jukbyeon	2007-08	5533	0.066	0.176

Table 1.3. Uncorrected effective population sizes for each sampling site, including temporal replicates. N_e reported for the two lowest minor allele frequencies used (0.01, 0.02), with 95% confidence intervals given in parentheses. For sites where putative immigrant individuals were identified in STRUCTURE and the PCA, migration rate (m) for that site was calculated as number of immigrants divided by sample size. Effective population size was then recalculated with immigrants removed. There were too few samples from Namhae to provide an accurate estimate of N_e ($n=11$).

Sampling Site	Spawning Season	N_e (0.010)	N_e (0.020)	m	Recalc. N_e (0.010)	Recalc. N_e (0.020)
Yellow Sea (YS) Block	2015-16	943 (838-1078)	943 (838- 1078)			
Boryeong	2007-08	2051 (1564-2987)	2051 (1564-2987)			
Namhae	2014-15	-	-	0.018	-	-
Geoje	2014-15	437 (415-459)	911 (814-1033)	0.036	1083 (959-1243)	942 (834-1081)
	2013-14	447 (415-484)	447 (415-484)	0.046	938 (802-1130)	939 (802-1130)
Jinhae Bay	2007-08, Dec.	236 (231-241)	186 (183-190)			
	2007-08, Feb.	654 (623-689)	685 (646-729)			
Pohang	2014-15	410 (395-427)	934 (849-1038)	0.065	2307 (1837-3099)	1640 (1357-2071)
Jukbyeon	2007-08	653 (615-696)	789 (727-861)	0.125	1074 (958-1223)	1078 (947-1249)

Table 1.4. Pairwise F_{st} between sampling sites, including temporal replicates. Lines segregate sites from different coasts (west, south/southeast, east). Statistically significant differentiation is indicated by (*) for Fisher's exact test, and (+) for strataG's permutation test.

	YS Block	Boryeong	Namhae	Geoje 2014-15	Geoje 2013-14	Jin. Bay Dec.	Jin. Bay Feb.	Pohang
Boryeong	0.0028 ⁺							
Namhae	0.0253 ^{*+}	0.0257 ⁺						
Geoje, 2014-15	0.0369 ^{*+}	0.0390 ^{*+}	0.0045					
Geoje, 2013-14	0.0365 ^{*+}	0.0382 ^{*+}	0.0029	0.0033 ⁺				
Jin. Bay, Dec.	0.0405 ^{*+}	0.0423 ^{*+}	0.0036	0.0035 ⁺	0.0010			
Jin. Bay, Feb.	0.0372 ^{*+}	0.0391 ^{*+}	0.0018	0.0030 ⁺	0.0007	0.0002		
Pohang	0.0348 ^{*+}	0.0364 ^{*+}	0.0021	0.0027 ⁺	0.0012	0.0016	0.0008	
Jukbyeon	0.0277 ^{*+}	0.0266 ^{*+}	0.0306 ^{*+}	0.0394 ^{*+}	0.0395 ^{*+}	0.0420 ^{*+}	0.0394 ^{*+}	0.0378 ^{*+}

Table 1.5. Inferred proportional ancestry for putative migrants identified in the PCA. Cluster 1 consisted primarily of western coast samples, Cluster 2 of southern/southeastern coast samples, and Cluster 3 of eastern coast samples. Proportions are averages of three replicate runs in STRUCTURE. (*) Indicates cluster to which the majority of individuals at that sampling site were assigned.

	Sampling Site	Collection Date	Sample ID	Inferred Ancestry		
				Cluster 1	Cluster 2	Cluster 3
1	Namhae	2/10/2015	22	0.950	0.043*	0.009
2	Namhae	2/10/2015	30	0.971	0.026*	0.002
3	Geoje	1/23/2015	05	0.924	0.072*	0.004
4	Geoje	2/4/2014	06	0.991	0.006*	0.004
5	Pohang	1/7/2015	11	0.937	0.032*	0.031
6	Pohang	2/5/2015	06	0.833	0.142*	0.025
7	Jukbyeon	12/10/2007	01	0.907	0.018	0.075*
8	Jukbyeon	12/10/2007	13	0.980	0.006	0.015*
9	Jukbyeon	12/10/2007	31	0.001	0.994	0.004*
10	Jukbyeon	12/10/2007	33	0.983	0.004	0.013*

Table 1.6. Uncorrected effective population sizes for the southern population with contributing aggregates grouped by spawning season. The southern population was defined according to STRUCTURE assignments and PCA clusters. Potential immigrant individuals were removed prior to analysis. Values are reported for two lowest minor allele frequencies used (0.01, 0.02), with 95% confidence intervals given in parentheses.

Spawning Season	N. Aggregates	N. Samples	F_{is}	H_o	N_e (0.010)	N_e (0.020)
2007-08	1	73	0.0095	0.1731	445 (437-454)	411 (403-419)
2013-14	1	22	0.0183	0.1737	939 (802-1130)	939 (802-1130)
2014-15	3	65	-0.0639	0.1945	1580 (1480-1696)	1509 (1412-1621)

Table 1.7. Outlier loci identified in both Bayescan (with given priors “p.”) and OutFLANK. Loci listed above the line were detected in full dataset; the locus listed below the line was detected only among southern sites.

Data set	Locus ID	OutFLANK	<i>q value</i>			
			Bayescan p. 10	Bayescan p. 100	Bayescan p. 1K	Bayescan p. 10K
All Data	10203	0.034	0	0	1.5×10^{-5}	0.0070
	14546	0.0057	0	0	5×10^{-5}	0.0008
	19221	0.0008	0	0	0	0
	2694	0.0123	0	0	7.5×10^{-5}	0.0039
	1904	0.0319	0	0.0002	0.0055	
	18723	0.0123	0.0004	0.0064		
	3699	0.0183	0.0001	0.0013		
	2606	0.0340	0.0029			
	3405	0.0144	0.0041			
Southern Population	24927	0.466526	5×10^{-5}	0.0005	0.0214	

Table 1.8. Outlier loci that aligned to annotated, protein-coding regions in the Atlantic cod genome. Loci listed above the line were identified in full data set, loci below the line identified between southern sites only. For each locus, the table lists the programs in which that locus was identified as an outlier, the reference linkage group to which the locus aligned in the Atlantic cod genome, the identification number for the gene corresponding to the alignment sequence in Atlantic cod, the similar protein listed in the annotation, and the species in which that protein function is known. All loci aligned within protein-coding sequences except locus 3313, which was 2.9 kb downstream of the listed gene.

Locus ID	Programs	Reference	Gene ID	Similar Protein; Species
10203	OutFLANK, Bayescan p10,000	Linkage Group 16	GAMO_00009696	Kmt2b; <i>M. musculus</i>
3699	OutFLANK, Bayescan p100	Linkage Group 14	GAMO_00017719	DHX38; <i>B. taurus</i>
8290	Bayescan p10,000	Linkage Group 10	GAMO_00049207	FRMPD3; <i>H. sapiens</i>
74	Bayescan p1000	Linkage Group 14	GAMO_00017638	PLEKHA7; <i>H. sapiens</i>
16976	Bayescan p1000	Linkage Group 11	GAMO_00019717	TMEM245; <i>B. taurus</i>
18636	Bayescan p1000	Linkage Group 21	GAMO_00066962	Fndc1; <i>R. norvegicus</i>
453	Bayescan p100	Linkage Group 10	GAMO_00050082	slu7; <i>D. rerio</i>
2705	Bayescan p100	Linkage Group 1	GAMO_00024607	Rbsn-5; <i>M. musculus</i>
15585	Bayescan p100	Linkage Group 11	GAMO_00021759	Kmt2b; <i>M. musculus</i>
16378	Bayescan p100	Linkage Group 16	GAMO_00009513	GOLIM4; <i>H. sapiens</i>
22902	Bayescan p100	Linkage Group 6	GAMO_00042274	ADD2; <i>H. sapiens</i>
3313	Bayescan p10	Linkage Group 16	GAMO_00010294	Vmn2r26; <i>M. musculus</i>
24927	OutFLANK, Bayescan p1000	Linkage Group 6	GAMO_00040950	PRKAB1; <i>P. abelii</i>
3583	OutFLANK	Linkage Group 18	GAMO_00065636	Unc13d; <i>M. musculus</i>
3884	OutFLANK	Linkage Group 1	GAMO_00026224	KLHDC8B; <i>H. sapiens</i>
10315	OutFLANK	Linkage Group 16	GAMO_00011302	Tiam1; <i>M. musculus</i>
22785	OutFLANK	Linkage Group 16	GAMO_00009424	MIPEP; <i>H. sapiens</i>

1.8 REFERENCES

- Ajmal, M., Khan, M. I., Neveling, K., Khan, Y. M., Azam, M., Waheed, N. K., . . . Cremers, F. P. M. (2014). A missense mutation in the splicing factor gene *DHX38* is associated with early-onset retinitis pigmentosa with macular coloboma. *Journal of Medical Genetics*, 51(7), 444-448. doi: 10.1136/jmedgenet-2014-102316
- Allendorf, F. W., England, P. R., Luikart, G., Ritchie, P. A., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*, 23(6), 327-337. doi:10.1016/j.tree.2008.02.008

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410.
- Anderson, E. C. (2010). Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, 10(4), 701-710. doi:10.1111/j.1755-0998.2010.02846.x
- Anderson, E. C., & Garza, J. C. (2006). The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172(4), 2567-2582. doi:10.1534/genetics.105.048074
- Anderson, E. C., Waples, R. S., & Kalinowski, S. T. (2008). An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(7), 1475-1486. doi:10.1139/F08-049
- Andreu-Vieyra, C. V., Chen, R., Agno, J. E., Glaser, S., Anastassiadis, K., Stewart, A. F., & Matzuk, M. M. (2010). MLL2 is required in oocytes for bulk Histone 3 Lysine 4 trimethylation and transcriptional silencing. *PLOS Biology*, 8(8), e1000453. doi:10.1371/journal.pbio.1000453
- Archer, F., Adams, P., & Schneiders, B. (2016). strataG: An R package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources*, 17(1), 5-11. doi:10.1111/1755-0998.12559
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., . . . Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3(10), e3376. doi:10.1371/journal.pone.0003376
- Beacham, T. D., Schweigert, J. F., MacConnachie, C., Le, K. D., & Flostrand, L. (2008). Use of microsatellites to determine population structure and migration of Pacific herring in British Columbia and adjacent regions. *Transactions of the American Fisheries Society*, 137(6), 1795-1811. doi:10.1577/T08-033.1
- Beirão, J., Purchase, C., Wringe, B., & Fleming, I. (2014). Sperm plasticity to seawater temperatures in Atlantic cod (*Gadus morhua*) is affected more by population origin than individual environmental exposure. *Marine Ecology Progress Series*, 495, 263-274. doi:10.3354/meps10552
- Bekkevold, D., Gross, R., Arula, T., Helyar, S. J., & Ojaveer, H. (2016). Outlier loci detect intraspecific biodiversity among spring and autumn spawning herring across local scales. *PLoS ONE*, 11(4), e0148499.
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, 24(13), 3299-3315. doi:10.1111/mec.13245
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2016). Erratum- Benestan et al . 2015. *Molecular Ecology*, 25(7), 1626-1629. doi:10.1111/mec.13600
- Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., . . . André, C. (2015). Adaptation to low salinity promotes genomic divergence in Atlantic Cod (*Gadus morhua* L.). *Genome Biology and Evolution*, 7(6), 1644-1663. doi:10.1093/gbe/evv093
- Bourgeois, C. F., Mortreux, F., & Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nature Reviews Molecular Cell Biology*, 17, 426-438. doi:10.1038/nrm.2016.50

- Bradbury, I.R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., . . . Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701), 3725-3734. doi:10.1098/rspb.2010.0985
- Brieuc, M. S., Waters, C. D., Seeb, J. E., & Naish, K. A. (2014). A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3: Genes, Genomes, Genetics*, 4(3), 447-460. doi:10.1534/g3.113.009316
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4), 855-867. doi:10.1111/1755-0998.12357
- 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. doi:10.1111/j.1365-294X.2010.04815.x
- Catchen, J. C., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. (2011). Stacks: building and genotyping loci de novo for short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171-182. doi:10.1534/g3.111.000240
- Catchen, J. C., Hohenlohe, P., Bassham, S., Amores, A., & Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-2140. doi:10.1111/mec.12354
- Chang, K. I., Teague, W. J., Lyu, S. J., Perkins, H. T., Lee, D. K., Watts, D. R., . . . Kim, K. (2004). Circulation and currents in the southwestern East/Japan Sea: overview and review. *Progress in Oceanography*, 61(2-4), 105-156. doi:10.1016/j.pocean.2004.06.005
- Chung, S., Kim, S., & Kang, S. (2013). Ecological relationship between environmental factors and Pacific cod (*Gadus macrocephalus*) catch in the southern East/Japan Sea. *Animal Cells and Systems*, 17(5), 374-382. doi:10.1080/19768354.2013.853693
- 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. doi:10.1139/F08-199
- Dahl, J. A., Jung, I., Aanes, H., Greggains, G. D., Manaf, A., Lerdrup, M., . . . Klungland, A. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*, 537, 548. doi:10.1038/nature19360
- Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14(1), 209-214. doi:10.1111/1755-0998.12157
- 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*, 1-17. doi:10.1111/eva.12639
- Etter, P. D., Preston, J. L., Bassham, S., Cresko, W. A., & Johnson, E. A. (2011). Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE*, 6(4), e18561. doi:10.1371/journal.pone.0018561

- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Foll, M. (2012). BayeScan v2.1 User Manual. Retrieved from: www.cmpg.unibe.ch/software/BayeScan
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), 977-993. doi:10.1534/genetics.108.092221
- Gosselin, T., Anderson, E. C., & Bradbury, I. R. (2016). assigner: assignment analysis with GBS/RAD Data using R. R package version 0.1.9. doi:10.5281/zenodo.51453
- Gustafson, R. G., Lenarz, W. H., McCain, B. B., Schmitt, C. C., Grant, W. S., Builder, T. L., & Methot, R. D. (2000). *Status Review of Pacific Hake, Pacific cod, and Walleye Pollock from Puget Sound, Washington* (NMFS-NWFSC-44). Retrieved from: <https://www.nwfsc.noaa.gov/publications/scipubs/techmemos/tm44/tm44.htm>
- Gwak, W., & Nakayama, K. (2011). Genetic variation and population structure of the Pacific cod *Gadus macrocephalus* in Korean waters revealed by mtDNA and msDNA markers. *Fisheries Science*, 77(6), 945-952. doi:10.1007/s12562-011-0403-2
- Hattori, T., Sakurai, Y., & Shimazaki, K. (1992). Maturation and reproductive cycle of female Pacific cod in waters adjacent to the southern coast of Hokkaido, Japan. *Nippon Suisan Gakkaishi*, 58(12), 2245-2252.
- Hauser, L., Adcock, G. J., Smith, P. J., Bernal Ramírez, J. H., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences*, 99(18), 11742. doi: 10.1073/pnas.172242899
- Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4), 333-362. doi:10.1111/j.1467-2979.2008.00299.x
- Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, 87(4), 971-1002. doi:10.5343/bms.2010.1051
- Hedrick, P. (2007). Large variance in reproductive success and the Ne/N ratio. *Evolution*, 59(7), 1596-1599. doi:10.1111/j.0014-3820.2005.tb01809.x
- Hemmer-Hansen, J., Nielsen Einar, E., Therkildsen Nina, O., Taylor Martin, I., Ogden, R., Geffen Audrey, J., . . . Carvalho Gary, R. (2013). A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, 22(10), 2653-2667. doi:10.1111/mec.12284
- Hjort, J. (1914). Fluctuations in the great fisheries of northern Europe, viewed in the light of biological research. *Rapports et Procès-Verbaux des Réunions du Conseil Permanent International Pour L'Exploration de la Mer*, 20, 1-228.
- Hwang, J. H., Van, S. P., Choi, B. J., Chang, Y. S., & Kim, Y. H. (2014). The physical processes in the Yellow Sea. *Ocean & Coastal Management*, 102, 449-457. doi:10.1016/j.ocecoaman.2014.03.026
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405. doi:10.1093/bioinformatics/btn129

- Kalinowski, S. T., Wagner, A. P., & Taper, M. L. (2006). ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, 6, 576-579. doi:10.1111/j.1471-8286.2006.01256.x
- Kim, M.J., An, H. S., & Choi, K.H. (2010). Genetic characteristics of Pacific cod populations in Korea based on microsatellite markers. *Fisheries Science*, 76(4), 595-603. doi:10.1007/s12562-010-0249-z
- Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., . . . Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, 25(10), 2130-2143. doi:10.1111/mec.13592
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359. doi:10.1038/nmeth.1923
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), R25. doi:10.1186/gb-2009-10-3-r25
- Lee, K., Cha, H. K., Kim, Y., Lee, J.Y., & Jung, S. (2016). Comparison of biological characteristics of Pacific cod (*Gadus macrocephalus*) between the East and Yellow Seas, Korea. *Korean Journal of Fisheries and Aquatic Sciences* 49(4), 499-508. doi:10.5657/KFAS.2016.0499
- Lee, S.G. & Midani, A.R. (2014). National comprehensive approaches for rebuilding fisheries in South Korea. *Marine Policy* 45, 156-162. doi:10.1016/j.marpol.2013.12.010
- Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24(5), 1031-1046. doi:10.1111/mec.13100
- Low, L. (2014). International Affairs and Research Collaboration: NOAA-Korea Joint Project Agreement. Retrieved from: <https://www.afsc.noaa.gov/Quarterly/jas2014/divrptsREFM10.htm>
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50-60. doi:10.1214/aoms/1177730491
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28-41. doi:10.1111/1755-0998.12291
- McCracken, J., DiCosimo, J., & Fey, M. (2013). *Overview of Apportionment of BSAI Pacific cod Sector Allocations Between BS and AI Areas and AI Pacific cod Processing Sideboards* Item C-3(a). Retrieved from https://www.npfmc.org/wp-content/PDFdocuments/catch_shares/Pcod/BS-AIpcodABC-TACsplit413.pdf
- McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2017). Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(4), 429-434. doi:10.1139/cjfas-2016-0443
- Monteverde, T., Muthalagu, N., Port, J., & Murphy, D. J. (2015). Evidence of cancer-promoting roles for AMPK and related kinases. *The FEBS Journal*, 282(24), 4658-4671. doi:10.1111/febs.13534

- Morley, J., Heusser, L., & Sarro, T. (1986). Latest Pleistocene and Holocene palaeoenvironment of Japan and its marginal sea. *Palaeogeography, Palaeoclimatology, Palaeoecology* 53(2-4), 349-358. doi:10.1016/0031-0182(86)90068-4
- Naish, K. A., Taylor, J. E., Levin, P. S., Quinn, T. P., Winton, J. R., Huppert, D., & Hilborn, R. (2008). An evaluation of the effects of conservation and fishery enhancement hatcheries on wild populations of salmon. In *Advances in Marine Biology*. San Diego CA: Elsevier Academic Press Inc.
- National Marine Fisheries Service. (2010). NOAA Species of Concern: Salish Sea Pacific cod. Retrieved from www.nmfs.noaa.gov/pr/pdfs/species/pacificcod_detailed.pdf
- Ni, G., Li, Q. I., Kong, L., & Yu, H. (2013). Comparative phylogeography in marginal seas of the northwestern Pacific. *Molecular Ecology*, 23(3), 534-548. doi:10.1111/mec.12620
- Oba, T., Kato, M., Kitazato, H., Koizumi, I., Omura, A., Sakai, T., & Takayama, T. (1991). Paleoenvironmental changes in the Japan Sea during the last 85,000 years. *Paleoceanography*, 6(4), 499-518. doi:10.1029/91PA00560
- Ogden, R. (2008). Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries*, 9(4), 462-472. doi:10.1111/j.1467-2979.2008.00305.x
- Park, K., Lee, E., Chang, E., & Hong, S. (2015). Spatial and temporal variability of sea surface temperature and warming trends in the Yellow Sea. *Journal of Marine Systems*, 143, 24-38. doi:10.1016/j.jmarsys.2014.10.013
- Park, S. C., Yoo, D. G., Lee, C. W., & Lee, E. I. (2000). Last glacial sea-level changes and paleogeography of the Korea (Tsushima) Strait. *Geo-Marine Letters*, 20(2), 64-71. doi:10.1007/s003670000039
- Peterson, D. A., Hilborn, R., & Hauser, L. (2014). Local adaptation limits lifetime reproductive success of dispersers in a wild salmon metapopulation. *Nature Communications*, 5, 3696. doi:10.1038/ncomms4696
- Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics* 4. doi:10.3389/fgene.2013.00098
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842. doi:10.1093/bioinformatics/btq033
- Quinn, T. P., Unwin, M. J., & Kinnison, M. T. (2000). Evolution of Temporal Isolation in the wild: Genetic divergence in timing of migration and breeding by introduced Chinook salmon populations. *Evolution*, 54(4), 1372-1385. doi:10.1111/j.0014-3820.2000.tb00569.x
- 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. doi:10.1080/19425120.2014.976680
- Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, 10(4), 361-395. doi:10.1111/j.1467-2979.2008.00324.x

- Rideout, R., Morgan, M., & Lilly, G. (2006). Variation in the frequency of skipped spawning in Atlantic cod (*Gadus morhua*) off Newfoundland and Labrador. *ICES Journal of Marine Science*, 63, 1101-1110. doi:10.1016/j.icesjms.2006.04.014
- Rideout, R. M., Burton, M. P. M., & Rose, G. A. (2000). Observations on mass atresia and skipped spawning in northern Atlantic cod, from Smith Sound, Newfoundland. *Journal of Fish Biology*, 57(6), 1429-1440. doi:10.1111/j.1095-8649.2000.tb02222.x
- Rideout, R. M., & Tomkiewicz, J. (2011). Skipped spawning in fishes: more common than you might think. *Marine and Coastal Fisheries*, 3(1), 176-189. doi:10.1080/19425120.2011.556943
- Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103-106. doi:10.1111/j.1471-8286.2007.01931.x
- Ruzzante, D. E., Taggart, C. T., Lang, S., & Cook, D. (2000). Mixed stock analysis of Atlantic cod near the Gulf of St. Lawrence based on microsatellite DNA. *Ecological Applications*, 10(4), 1090-1109. doi:10.1890/1051-0761(2000)010[1090:MSAOAC]2.0.CO;2
- Sanchez-Cespedes, M. (2008). PRKAB1 (protein kinase, AMP-activated, beta1 non-catalytic subunit). *Atlas of Genetics and Cytogenetics in Oncology and Haematology*, 12(2), 151-152. doi:10.4267/2042/38502
- Schmidtko, S., Johnson, G. C., & Lyman, J. M. (2013). MIMOC: A global monthly isopycnal upper-ocean climatology with mixed layers. *Journal of Geophysical Research: Oceans*, 118(4), 1658-1672. doi:10.1002/jgrc.20122
- Schwer, B., & Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, 349, 494. doi:10.1038/349494a0
- Shi, W., Ayub, Q., Vermeulen, M., Shao, R.-g., Zuniga, S., van der Gaag, K., . . . Tyler-Smith, C. (2010). A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Molecular Biology and Evolution*, 27(2), 385-393. doi:10.1093/molbev/msp243
- Shimada, A. M., & Kimura, D. K. (1994). Seasonal movements of Pacific cod, *Gadus macrocephalus*, in the eastern Bering Sea and adjacent waters based on tag-recapture data. *Fishery Bulletin*, 92(4), 800-816.
- Sinclair-Waters, M., Bentzen, P., Morris, C. J., Ruzzante, D. E., Kent, M. P., Lien, S., & Bradbury, I. R. (2018). Genomic tools for management and conservation of Atlantic Cod in a coastal marine protected area. *Canadian Journal of Fisheries and Aquatic Sciences*. doi:10.1139/cjfas-2017-0254
- Skjaeraasen, J. E., Nash, R. D. M., Korsbrekke, K., Fonn, M., Nilsen, T., Kennedy, J., . . . Kjesbu, O. S. (2012). Frequent skipped spawning in the world's largest cod population. *Proceedings of the National Academy of Sciences*, 109(23), 8995-8999. doi:10.1073/pnas.1200223109
- Sokal, R. R., & Rohlf, F. J. (1995). 18.1: Combining Probabilities from Independent Tests of Significance. In *Biometry: The Principles and Practice of Statistics in Biological Research* (3rd ed., pp. 794-797). New York, NY: W.H. Freeman and Co.
- 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. doi:10.1080/00028487.2012.711265

- 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. doi:10.1139/cjfas-2014-0050
- Suda, A., Nagata, N., Sato, A., Narimatsu, Y., Nadiatul, H. H., & Kawata, M. (2017). Genetic variation and local differences in Pacific cod *Gadus macrocephalus* around Japan. *Journal of Fish Biology*, 90(1), 61-79. doi:10.1111/jfb.13154
- Tada, R. (1999). Late Quaternary paleoceanography of the Japan Sea: an update. *The Quaternary Research Journal*, 38(3), 216-222.
- Tomkiewicz, J., Tybjerg, L., & Jespersen, Å. (2003). Micro- and macroscopic characteristics to stage gonadal maturation of female Baltic cod. *Journal of Fish Biology*, 62(2), 253-275. doi:10.1046/j.1095-8649.2003.00001.x
- Tørresen, O. K., Star, B., Jentoft, S., Reinart, W. B., Grove, H., Miller, J. R., . . . Nederbragt, A. J. (2017). An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics*, 18(1):95. doi:10.1186/s12864-016-3448-x
- Turner, T.F., Wares, J.P., & Gold, J.R. (2002). Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). *Genetics* 162(3), 1329-1339. doi:
- Waelbroeck, C., Labeyrie, L., Michel, E., Duplessy, J.C., McManus, J.F., . . . Labracherie, M. (2002). Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quaternary Science Reviews* 21(1-3), 295-305. doi:10.1016/S0277-3791(01)00101-9
- Wang, P. (1999). Response of western Pacific marginal seas to glacial cycles: paleoceanographic and sedimentological features. *Marine Geology*, 156, 5-39.
- Waples, R. (2016). Tiny estimates of the Ne/N ratio in marine fishes: Are they real? *Journal of Fish Biology*, 89(6), 2479-2504. doi:10.1111/jfb.13143
- Waples, R. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*, 117(4), 233-240. doi:10.1038/hdy.2016.60
- Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 3(3), 244-262. doi:10.1111/j.1752-4571.2009.00104.x
- Waples, R. S., & England, P. R. (2011). Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics*, 189(2), 633-644. doi:10.1534/genetics.111.132233
- Waples, R. S., Punt, A. E., & Cope, J. M. (2008). Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries*, 9(4), 423-449. doi:10.1111/j.1467-2979.2008.00303.x
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370.
- Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of Fst. *The American Naturalist*, 186(S1), S24-S36. doi:10.1086/682949
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6), 80-83. doi:10.2307/3001968

- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395-420.
- Xu, J., Chan, T.Y., Tsang, L. M., & Chu, K. H. (2009). Phylogeography of the mitten crab *Eriocheir sensu stricto* in East Asia: Pleistocene isolation, population expansion and secondary contact. *Molecular Phylogenetics and Evolution*, 52(1), 45-56.
doi:10.1016/j.ympev.2009.02.007
- Xu, Q., & Xie, W. (2018). Epigenome in early mammalian development: inheritance, reprogramming and establishment. *Trends in Cell Biology*, 28(3), 237-253.
doi:10.1016/j.tcb.2017.10.008
- Zhang, C. I. (1984). Pacific cod of South Korean waters. *Bulletin of the International North Pacific Fisheries Community*, 42, 116-129.

Chapter 2.

The genomics of adaptive divergence in Pacific cod (*Gadus macrocephalus*)

2.1 ABSTRACT

Genome-wide patterns of divergence in large marine populations provide insight into the evolution and maintenance of adaptation in the marine environment. The two highly diverged populations of Pacific cod (*Gadus macrocephalus*) on either side of the north Pacific offer a natural experiment to explore parallel adaptation and genomic signatures of divergence. Our study combined RAD sequencing data from 479 samples of eastern and western Pacific cod to provide the first next-generation sequencing analysis of the species across its transoceanic range. After alignment of 4,286 RAD loci to the Atlantic cod genome, shared signatures of adaptive divergence were assessed using a combination of outlier tests at individual loci, and a genome-wide kernel-smoothing weighted average F_{st} calculated through a sliding window analysis for the two populations. With the exception of a single region of elevated divergence on linkage group 19, we found no genetic signatures of parallel adaptation with either the individual locus or genome-wide approaches. However, gene annotations suggest that parallel phenotypic adaptation in circadian clock genes may be occurring without parallel genetic changes. By conducting a third sliding window analysis on the full data set, we were then able to compare the width and frequency of genomic regions of elevated divergence across three divergence scenarios: the

primary and secondary models of divergence-with-gene-flow, represented by the eastern and the western populations, respectively, and divergence-without-gene-flow, represented by the full data set. We observed larger and fewer outlier regions in the primary model, compared to the secondary model, of divergence-with-gene-flow. The cross-Pacific comparison of divergence-without-gene-flow contained the most, and smallest, outlier regions. However, many of these trends were non-significant. Overall, this research provided key insights into adaptation and divergence in Pacific cod, and can be used to direct future genetic analysis on this species.

2.2 INTRODUCTION

Genome-wide scans of differentiation can provide insight into adaptive evolution and the processes that maintain adaptation in the face of gene flow. Recent studies have used these scans in a variety of species to identify genomic regions involved in local adaptation and ecological speciation (Wolf & Ellegren, 2016). Although there are many interacting factors that dictate the distribution, size, and persistence of such regions (Yeaman, Aeschbacher, & Bürger, 2016), different mechanisms of divergence can leave distinct patterns of differentiation across the genome (Feder, Egan, & Nosil, 2012; Tigano & Friesen, 2016). While genome scans alone cannot definitely ascribe the formation of these patterns to a specific mechanism (Wolf & Ellegren, 2016), they can narrow down competing hypotheses and direct further research.

Marine species composed of large, geographically extensive populations have particularly strong potential to display diverse genome-wide patterns of adaptation and divergence (Bradbury et al., 2010). Many species are distributed across heterogeneous selection regimes, which promote subpopulation divergence through local adaptation when gene flow is restricted or when selection is strong (Kawecki & Ebert, 2004). In addition, their relatively large

effective population sizes are associated with higher levels of standing genetic variation, the raw material for adaptation (Hemmer-Hansen, Therkildsen, & Pujolar, 2014).

Extensive studies of genome-wide adaptation in Atlantic cod have related large regions of elevated differentiation, or outlier regions, to adaptive divergence between migratory and stationary ecotypes (Hemmer-Hansen et al., 2013; Sinclair-Waters et al., 2018), inshore and offshore spawning populations (Barney et al., 2017), and fjord and oceanic populations that spawn sympatrically (Kirubakaran et al., 2016; Sodeland et al., 2016). These regions of elevated divergence in the Atlantic cod genome represent polymorphic chromosome inversions, which facilitate the formation of genomic outlier regions by reducing local combination rates (Berg et al., 2017; Sodeland et al., 2016).

Atlantic cod has also been shown to display parallel adaptation to temperature-associated clines on either side of the northern Atlantic (Bradbury et al., 2010). Eastern and western populations of Atlantic cod likely diverged significantly earlier than the last glacial maximum, and according to concordant ecological-niche models and genetic data, were isolated in separate glacial refugia along the Canadian and European coasts during previous ice ages (Bigg et al., 2008). Yet even in long-diverged populations, exposure to similar selection pressures can cause parallel phenotypic evolution with a basis in parallel genetic changes, a phenomenon found to be particularly common in oceanic populations with high standing genetic variation (Hohenlohe et al., 2010). Atlantic herring, another widespread marine fish species, has also displayed evidence of parallel adaptation across temperature clines (Lamichhaney et al., 2017).

Genomic patterns of adaptation and divergence have yet to be explored in Pacific cod, a related marine fish species distributed across the northern Pacific Ocean (Mecklenburg, Mecklenburg, & Thorsteinson, 2002). Similar to Atlantic cod, Pacific cod consists of an eastern

and a western population which were isolated in separate marine refugia during past glaciations before expanding southward along opposite coasts (Canino et al., 2010; Bigg et al., 2014). Their current range consists of an eastern and western population, which represent two distinct lineages (Canino et al., 2010). Pacific cod spawning aggregates within both populations display phenotypic differences in multiple life history traits, including growth rate and age-at-maturity (Gustafson et al., 2000; Ormseth & Norcross, 2009; Zhang, 1984), although the genetic basis and heritability of these traits is unknown. Many of these traits vary along latitudinal gradients in the eastern Pacific (Gustafson et al., 2000; Ormseth & Norcross, 2009), but have also been observed to differ over smaller geographical scales in the western Pacific (Zhang, 1984) where oceanographic currents create sharp boundaries in thermal and salinity regimes.

Since Pacific cod is targeted by valuable fisheries throughout its range, research on adaptive divergence in this species may assist fisheries management by identifying locally adapted populations. Such diversity within exploited species has been shown to buffer negative impacts of variability on fisheries (Schindler et al., 2010), and so is important to conserve. In addition, adaptations related to thermal tolerance can provide a better understanding of the potential impacts of climate change on marine species' survival and distribution (Bernatchez et al., 2017), and the resulting effects on associated fisheries. In Pacific cod, such adaptations may underlie the species' phenotypic diversity in traits such as the temperature sensitivity of early life history stages (Alderdice & Forrester, 1971; Laurel et al., 2008). Climate change impacts may greatly affect Pacific cod populations which inhabit the warmest waters at the southern edge of its range, near the Korean peninsula in the western Pacific and the Washington coastline in the eastern Pacific.

Prior studies on Pacific cod have established that divergence in the species has evolved both sympatrically, in the presence of gene flow (Cunningham et al., 2009; Drinan et al., 2018), and during periods of allopatry (Canino et al., 2010). In divergence with gene flow, adaptations can arise at loci resistant to introgression, such as those which form the basis to key isolating traits affecting survival and reproduction (Cruickshank & Hahn). In contrast, adaptive divergence in allopatry may also represent selection imposed by a new environment (Cruickshank & Hahn, 2014), and can occur at loci under weaker selection pressures.

Recent literature addressing theoretical considerations of speciation-with-gene-flow explicitly identifies three scenarios through which divergence develops in wild populations (Cruickshank & Hahn, 2014; **Figure 2.1**): the ‘primary’ model of speciation-with-gene flow, or purely sympatric/parapatric speciation; the ‘secondary’ model of speciation-with gene flow, where populations exchanging gene flow experience a substantial period of independent evolution, which is then followed by the reintroduction of gene flow through secondary contact; and finally, allopatric speciation, when recently diverged species exchange no gene flow. While this three-scenario framework is traditionally used to characterize population divergence which results in separate species (Cruickshank & Hahn, 2014), it is also applicable to the development of divergence and adaptation within a species. Within marine species, which possess complex evolutionary histories from cycles of sea level change during past glaciations, and higher and more variable levels of gene flow than in terrestrial species, it can be assumed that empirical evidence of adaptation and divergence results from a combination of these three scenarios.

Populations of Pacific cod encompass all three scenarios (**Table 2.1**). The two large populations on either side of the Pacific have accumulated divergence in allopatry. Gene flow between the two populations is hypothesized to be negligible, particularly toward the

southwestern and southeastern edge of the range, although there is likely a secondary contact zone north of Japan (Stroganov & Orlov, 2012). Around the Korean peninsula in the western population, subpopulations associated with the western and southern Korean coasts likely represent the secondary model of divergence with gene flow. Differentiation between the two coastal subpopulations, which currently exchange low levels of gene flow, accumulated through a combination of isolation in separate ice-age refugia and local adaptation to different thermal and salinity regimes (see Chapter 1). The eastern Pacific population, representative of the primary model of divergence with gene flow, displays relatively low divergence and high gene flow between coastal subpopulations, creating an isolation-by-distance pattern of genetic differentiation (Cunningham et al., 2009; Drinan et al., 2018). A previous study using mitochondrial and microsatellite DNA proposes that these eastern subpopulations were established through the northward expansion of Pacific cod out of a single southern refugia (Canino et al., 2010). Combined, these three groups of Pacific cod create a natural experiment to observe how the genomic landscape of differentiation changes across three scenarios of population divergence.

Past research on adaptation and divergence in wild populations has been complicated by the highly heterogeneous landscapes of genomic differentiation between species and subspecies (Wolf & Ellegren, 2016), as genomic regions of elevated divergence can arise from complex interactions between selective pressures and a species' demographic history (Yeaman, Aeschbacher, & Bürger, 2016), or be confounded by variability in recombination and mutation rates (Wolf & Ellegren, 2016). Yet demographic histories of Pacific cod populations have already been detailed in prior population genetic studies. By focusing on genome-wide differentiation within a single species, we can also assume similar patterns of recombination and

mutation rates between groups (with the exception of potential chromosome inversions). And by calculating both F_{st} and H_o across the genome, we can partially account for false signals of elevated differentiation caused by our use of a relative measure of divergence (Cruickshank & Hahn, 2014; Wolf & Ellegren, 2016).

In this study, we used genome scans in eastern and western Pacific cod to generate kernel-smoothing weighted averages of genetic differentiation across the genome, in order to achieve two objectives. The first was to characterize patterns of genome-wide differentiation across the three scenarios of divergence represented by different populations of Pacific cod. We were specifically interested in whether the width and frequency of genomic regions of elevated divergence changed across divergence scenarios. The second objective was to test for parallel adaptation on an individual locus and a genome-wide scale, by aligning the kernel-smoothing weighted averages of genetic differentiation from eastern and western Pacific cod, and then identifying outlier loci putatively under selection in each population.

2.3 METHODS

2.3.1 *Data collection*

Raw RAD-sequencing data for Pacific cod were acquired from Drinan et al. (2018) and the first chapter of this thesis. The eastern Pacific cod data set consisted of 288 individuals collected from six known spawning grounds along the western coastline of the United States and Canada during the spawning season (**Figure 2.2a**; **Table 2.2**). Data were generated through single-read, 100 base pair RAD-sequencing on an Illumina HiSeq 2500, with 48 individuals pooled per lane of sequencing (Drinan et al., 2018). The western Pacific cod data set consisted of

251 individuals collected from five known spawning grounds around the Korean peninsula during the spawning season (**Figure 2.2b**; **Table 2.2**). Single-read and paired-end, 150 base pair RAD-sequencing was completed on an Illumina HiSeq 4000, with 72-74 individuals pooled per lane of sequencing. Only forward reads were used from samples for which paired-end data were produced.

2.3.2 *SNP discovery, genotype, and filtering*

Raw data for both eastern and western Pacific cod data sets were demultiplexed, filtered, and trimmed using the `process_radtags` function in the Stacks v1.44 pipeline (Catchen et al., 2011; Catchen et al., 2013). Demultiplexed fastq files were then aligned to a reference database of loci that was assembled *de novo* using samples from the western Pacific cod data set. A full description of reference database assembly is available in the first chapter of this thesis.

Alignment was completed using Bowtie (Langmead et al., 2009). Aligned reads from the 150bp western Pacific cod data set were trimmed to 92bp to match the trimmed sequence length of the eastern Pacific cod data set. SNP discovery and genotyping was then completed using the remainder of the Stacks pipeline. Stacks flags (`pstacks m = 5`, `populations m = 3`, `M = 3`, `N = 4`, `n = 3`, `max_locus_stacks = 3`) were set according to Mastretta-Yanes et al. (2015) and Drinan et al. (2018), to maximize retention of loci in both data sets while reducing SNP and allele calling error rates (Appendix C). The catalog of consensus loci used for `cstacks` was built using the ten individuals with the most data (measured as number of raw reads and number of aligned loci) from each sampling site in both eastern and western populations. Stacks called genotypes with a maximum likelihood approach for biallelic SNPs present in $\geq 50\%$ of fish per sampling site, with one SNP per RAD tag (`flag:write_random_SNP`). Each RAD locus therefore consisted of one biallelic SNP.

Final filtering removed loci with minor allele frequencies (MAFs) < 0.05 at every sampling site, and individuals missing genotypes at more than 30% of loci. Loci were then further filtered for more than 30% missing data across all sampling sites, and for not conforming to Hardy-Weinberg equilibrium (HWE) expectations. Allele frequencies for MAF filtering, as well as heterozygosity and HWE p-values, were calculated using Genepop v4.2 (Rousset, 2008, 2014). HWE p-values from Genepop were used to identify loci significantly out of HWE using Fisher's combination of probabilities from independent tests of significance, as applied by Sokal & Rohlf (1995).

Individuals from the northeast Pacific that had been identified in previous analyses (Drinan et al., 2018) as potentially contaminated or as duplicate samples were removed. We then calculated individual observed heterozygosity and ran ML Relate (Kalinowski, Wagner, & Taper, 2006) on the remaining samples to verify that all potentially contaminated or duplicate individuals were removed from the data set. This filtering resulted in a final total of 219 individuals in the western population, and 260 individuals in the eastern population (**Table 2.2**), which were genotyped at 5683 RAD loci.

2.3.3 *Alignment to the Atlantic cod genome*

All RAD loci were aligned to the Atlantic cod genome, in the absence of a Pacific cod genome assembly, in order to obtain the base pair positions required for the kernel-smoothing weighted averages. Alignments were completed with Bowtie2 (Langmead & Salzberg, 2012) to the 23 linkage groups in the most recent assembly of the Atlantic cod genome, gadMor2 (Tørresen et al., 2017). The consensus sequence for each RAD locus was built by the Stacks pipeline. Of the 5683 loci discovered by Stacks, 5095 (89.65%) aligned to the Atlantic cod

genome. Any aligned loci with a mapping quality lower than 10 were removed, leaving 4286 aligned RAD loci for all further analyses.

2.3.4 *Population structure*

We first performed a hierarchical analysis of molecular variance (Excoffier, Smouse, & Quattro, 1992), or AMOVA, on all retained individuals at all aligned loci to test for significant population structure between and within the eastern and western populations of Pacific cod. The first stratum for the partitioning of variance was the eastern and western populations, and the second stratum was the genetically distinct subpopulations within each population (**Table 2.2**). Subpopulations were based on results from Drinan et al., (2018) and the first chapter of this thesis. Whereas each sampling site in the eastern population was considered a separate subpopulation, sampling sites in the western population were pooled into two genetically distinct subpopulations: the western Korean coast (Yellow Sea Block and Boryeong) and the southern Korean coast (Namhae, Geoje, Jinhae Bay, Pohang). The eastern Korean coast subpopulation was excluded from this analysis due to low sample size (see Chapter 1). Previous analysis on our sampling sites in the western population showed no significant differentiation between sites within the same coast, and pooling provided larger sample sizes to avoid upward bias of F_{st} .

We used the R packages poppr v2.7.1 (Kamvar, Brooks, & Grunwald, 2015; Kamvar, Tabima, & Grunwald, 2014) and ade4 v1.7-10 (Chessel, Dufour, & Thioulouse, 2004; Dray & Dufour, 2007), to conduct an AMOVA with 10,000 permutations, and then tested the significance of the observed population structure through a Monte-Carlo permutation test with 10,000 replicates.

2.3.5 *Sliding window analysis*

Genomic regions of elevated divergence, or outlier regions, were identified using a kernel-smoothing moving average of per-locus F_{st} across 23 Atlantic cod linkage groups, a technique first described by Hohenlohe et al. (2010). The eastern population, western population, and full data set were run in separate analyses to identify differences location, frequency, and width of outlier regions among these three groups.

We first calculated Weir & Cockerham (1984) F_{st} at each aligned locus in Genepop v4.2 (Rousset, 2008). Global F_{st} was calculated for the eastern population and the entire data set, which consisted of twelve sampling sites from both the eastern and western populations. We calculated pairwise F_{st} between the two western subpopulations on the southern and western Korean coasts.

A moving, weighted average F_{st} with 95% confidence intervals was calculated across polymorphic loci using an R function by Brieuc et al. (2015). In short, a Gaussian weighted average was taken across loci within a window of 3σ , where $\sigma = 250\text{kb}$, at 150 divisions across each linkage group. The step size varied between 128kb and 232kb depending on the linkage group. The final window size and number of divisions were selected by adjusting starting values ($\sigma = 150\text{kb}$, divisions = 200) obtained from Hohenlohe et al. (2010) in stepwise increases, to reduce unnecessary variance while still allowing for the detection of genomic regions of elevated divergence (Larson et al., 2014). The 95% confidence interval for each calculated average was estimated from 100,000 bootstrap samples drawn at random from the entire data set of aligned loci (Brieuc et al., 2015). The output from this sliding window analysis for genetic differentiation was then filtered to include only weighted averages calculated from F_{st} at two or more loci (Larson et al., 2014). A genomic region of elevated divergence was defined as any region

containing one or more weighted F_{st} averages that exceeded the upper bound of the 95% confidence interval.

2.3.6 *Accounting for reduced diversity within populations*

A moving, weighted average with 95% confidence intervals was also calculated for observed heterozygosity (H_o) within the eastern and the western populations, to differentiate between outlier regions in the full data set that may have arisen from lower diversity within populations, rather than increased differences between them (Cruikshank & Hahn 2014). Observed heterozygosity per locus for the eastern and the western populations was first calculated with the R package Adegenet (Jombart, 2008), and then the kernel-smoothing moving average of per-locus heterozygosity calculated using the methods described for F_{st} .

A region of significantly reduced heterozygosity was defined as one or more weighted H_o averages that exceeded the lower bound of the 95% confidence interval. All genomic regions of elevated divergence identified in the full data set were aligned with the respective weighted H_o averages to determine whether regions of elevated divergence may have arisen out of within-population reduced diversity. We also conducted a linear regression of weighted F_{st} averages in the full data set against weighted averages of H_o in the eastern and the western populations, to test whether the level of genetic differentiation between the east and west Pacific was significantly correlated with observed heterozygosity within either population.

2.3.7 *Size and frequency of genomic regions of elevated divergence*

The frequency of genomic regions of elevated divergence was calculated as the total number of outlier regions across linkage groups, and the average number of outlier regions per linkage group. Region width was measured by the number of weighted averages that exceeded the upper bound of the 95% confidence interval within that region. Size and frequency

calculations were conducted separately for the eastern population, western population, and full data set using the appropriate sliding window analysis of genetic differentiation.

A Kruskal-Wallis rank sum test was used to determine whether the number or width of outlier regions differed significantly across all three divergence scenarios ($\alpha = 0.05$), because it allows for non-normality of the data, although it assumes homoscedasticity (Zar, 2010). We were also interested in whether the frequency and width of outlier regions differed significantly between each pair of divergence scenarios, and so conducted Welch's t-tests on ancillary hypotheses. A Welch's t-test does not assume equal variances and is relatively robust to non-normality, especially when the test is two-tailed and sample sizes are equivalent (Zar, 2010).

To test whether our results from the Kruskal-Wallis and Welch's t-test were biased by marker density, we conducted several linear regression analyses. Linear regressions assessed the correlation between the width and frequency of outlier regions on each linkage group, and the number of markers on that linkage group. We also compared the number of markers per weighted average F_{st} between averages that exceeded the 95% confidence interval (i.e. were within outlier regions) and those that did not. We hypothesized that the absence of a significant difference, or the absence of a strong correlation, would indicate that marker density was not a confounding factor in the detection of genomic regions of elevated divergence.

2.3.8 *Identification of outlier loci*

We tested for parallel adaptation at individual loci by identifying outlier loci in the eastern and western populations. Loci with significant evidence for divergent selection were identified using three programs to account for the varying assumptions among them: Bayescan (Foll & Gaggiotti, 2008), OutFLANK (Whitlock & Lotterhos, 2015), and PCAdapt (Luu, Bazin,

& Blum, 2017). Outlier loci were detected separately in the eastern population, the western population, and the full data set using each of these programs.

The first outlier detection method, employed by Bayescan v2.1, identifies loci with signatures of either divergent or balancing selection using a Bayesian framework to directly estimate the posterior probability of a given locus being under selection (Foll & Gaggiotti, 2008). Bayescan v2.1 was run with default settings and 20,000 iterations. The prior odds for the neutral model were run at 100 and 1000 (Foll, 2012), and the false discovery rate was set at 0.05.

The Bayesian framework used in Bayescan v2.1 assumes that samples have evolved independently from a common ancestor, and so can suffer large false positive rates when the data set contains non-random correlations between pairs of populations, such as in an isolation-by-distance or hierarchical population structure (Lotterhos & Whitlock, 2014; Whitlock & Lotterhos, 2015). We therefore also identified outlier loci using OutFLANK. By fitting an inferred distribution of F_{st} to loci not affected by selection, and then assigning q values to each locus based on the inferred distribution, OutFLANK has much lower false positive rates and relatively high power across multiple population structures (Whitlock & Lotterhos, 2015). OutFLANK was run using a threshold of $q = 0.05$ after trimming the right and left tails of the F_{st} distribution at 0.05. OutFLANK assumes that the inferred distribution is representative of the true distribution of F_{st} at neutral loci in the population, and that loci most affected by selection are in the upper and lower 5% of F_{st} values. OutFLANK performs best when sample sizes are large, or when sample sizes across loci are roughly equivalent if sample sizes are small (Whitlock & Lotterhos, 2015).

The final method, PCAdapt, detects candidate markers with respect to population structure. The R package first calculates the extent to which each SNP is related to the first K

principal components using a distance metric, and then identifies loci which do not follow the distribution of the distance measures for the majority of points (Luu, Bazin, & Blum, 2017). We used the default Mahalanobis distance method, which is more powerful than the communality statistic or Bayes factor implemented in previous versions of the program (Luu, Bazin, & Blum, 2017), and set $\alpha = 0.05$ for detection of outlier loci. The value of K was varied by data set to reflect population structure, and verified as the appropriate number of principal components using a scree plot (Figure S2.1): for the eastern population, which displays isolation-by-distance (Drinan et al., 2018), K was first set at two according to the scree plot, which likely reflected the geographic gap in the sampling regime between the four northern and two southern subpopulations; for the western population, K was set at two to representing the significantly differentiated western and southern Korean coasts; for the full data set, K was also set at two to represent the eastern and western populations.

Principal component analysis (PCA) was conducted separately on neutral and putative outlier loci using the R package Adegnet v2.1.1 (Jombart, 2008) to compare patterns of differentiation that arose from neutral processes only (mutation, migration, genetic drift) as opposed to also being affected by divergent selection. A locus was considered neutral if it was not identified using any of the three outlier methods. Outlier loci were included in the outlier PCA if they were identified in at least two of the three outlier methods.

2.3.9 *Annotation of RAD loci*

We also explored co-localization of RAD loci to annotated genes, in order to assess parallel patterns of gene function between loci putatively under selection in the east and west Pacific. We extracted the consensus sequence produced by Stacks for all 4286 aligned RAD loci, and then applied the function `closestBed` from `bedtools` v2.24 (Quinlan & Hall, 2010) to report

overlapping annotations in the Atlantic cod genome. Annotations were then filtered for loci that (1) fell within genomic regions of elevated divergence present in both the eastern and western populations, (2) were identified as outliers in the western population by any of the three outlier methods, and (3) were identified as outliers in the eastern population by two or more outlier methods. Criteria for retention of outlier loci was less stringent in the western population because the lower number of subpopulations in the west Pacific data set was expected to reduce the power of the outlier tests.

2.4 RESULTS

Of the 4286 loci that were retained after filtering and alignment to the Atlantic cod genome, 77.2% were polymorphic in the western population, and 62.5% polymorphic in the eastern population. The overall F_{is} of subpopulations ranged from 0.008 (Adak Island, eastern population) to 0.101 (southern Korean coast 2013-14, western population), while observed heterozygosity ranged from 0.099 (Washington Coast, eastern population) to 0.129 (western Korean coast 2007-08, western population) (**Table 2.3**). The two western subpopulations, even when further subdivided by spawning season, had greater observed heterozygosity than the eastern subpopulations. However, average F_{is} values for the western subpopulations ranged from quite low, 0.020, to the highest detected in this data set, 0.101.

As expected, genetic differentiation across all loci was smaller in the eastern population ($F_{st} = 0.010$) than in the western population ($F_{st} = 0.037$), and was highest in the full data set ($F_{st} = 0.285$). The highest single locus F_{st} was also lower in the eastern population ($F_{st} = 0.240$) than the western population ($F_{st} = 0.437$). Five loci in the full data set displayed near-fixed allele frequencies ($F_{st} > 0.990$) between the eastern and the western populations (**Figure 2.3**).

Differentiation between the eastern and western populations accounted for over half of the total variation in the data set ($\Phi_{ST} = 0.577$) (**Table 2.4**). Population structure between the eastern and western populations, and between subpopulations within each population, was statistically significant (**Table 2.4**; Figure S2.2).

2.4.1 *Identification of genomic regions of elevated divergence*

Average marker density in the sliding window analysis used to identify outlier regions ranged from 10.83 markers per weighted average in the full data set, to 7.17 markers per weighted average in the eastern population (**Figure 2.4**). Of the 3450 weighted averages calculated during sliding window analysis, 3405 had two or more markers in the full data set, 3378 in the western population, and 3345 in the eastern population. Sliding window analyses with 95% confidence intervals, plotted by chromosome and overlaid with per-locus F_{st} , can be found in the supplementary figures (Figures S2.8-10).

A per-locus kernel-smoothing average was also calculated for observed heterozygosity in the eastern and western populations to determine whether genomic regions of elevated divergence between the two populations were a result of reduced heterozygosity within either population. Regions of significantly elevated genetic differentiation between the eastern and western populations did not consistently co-occur with significant reductions in within-population heterozygosity; of the 44 outlier regions identified in the full data set, only one had significantly reduced heterozygosity in the western population (linkage group 16, **Figure 2.5a**), and none in the eastern population. We also observed that some outlier regions identified in the full data set corresponded to non-significant decreases in heterozygosity within the eastern or western population (e.g. **Figure 2.5b**). Linear regressions of weighted average F_{st} from the full

data set against weighted average H_o within the eastern and western populations showed very weak, but significant, correlations (Table S2.1; Figure S2.3).

2.4.2 *Changes in size and frequency of genomic regions of elevated divergence*

We observed larger and more frequent genomic regions of elevated divergence in the eastern population than in the western population, and in the western population than in the full data set (**Table 2.5**). However, these changes were not significant according to the Kruskal-Wallis rank sum test (**Table 2.6**), likely due to the high variation in outlier region frequency and size between linkage groups within each data set. This is evident in the large standard deviations around the mean frequency and width values (**Table 2.5**). Two-sample Welch's t-tests showed only one marginally significant comparison at the 0.05 alpha level: there were a significantly greater number of regions of elevated divergence in the full data set than in the eastern population (**Table 2.7**). The increases in outlier region width between the eastern and western populations and the full data set were just above our alpha level, with p values of 0.053 and 0.069, respectively.

When testing for the impact of marker density on outlier region frequency and width, we did not find a significant correlation between region frequency and marker density in any of the analyses, while only the western population showed a significant (moderately positive) correlation between outlier region width and marker density (Table S2.1). In addition, we failed to find evidence that the distribution of marker density per weighted average differs between weighted averages that were significantly differentiated, and those that were not (Figure S2.4).

The linkage group with the greatest number of genomic regions of elevated divergence varied by data set (**Figure 2.6**). In the western population, linkage group 16 contained the greatest number of outlier regions (5), whereas for the eastern population, the greatest number of

outlier regions was found on linkage group 4. Linkage group 5 contained the fewest number of outlier regions (one in the eastern population). The largest genomic region of elevated divergence in the eastern population was located at one edge of linkage group 19, with eleven weighted averages above the 95% confidence interval (spanning approximately 1.29 Mb). The largest genomic region of elevated divergence in the western population was located on linkage group 17, with nine significantly elevated weighted averages (spanning approximately 912 kb) (**Table 2.5**).

2.4.3 *Overlap in genomic regions of elevated divergence*

Only one genomic region of significantly elevated F_{st} overlapped between the eastern and western populations (**Figure 2.7**). This genomic region of elevated divergence on linkage group 19 was also found to be significantly differentiated over the full data set. The weighted averages of F_{st} within this region were calculated from 10 loci in the eastern population, two of which were identified as outlier loci; 14 loci in the western population; and 9 loci in the full data set, two of which were identified as outlier loci.

To better understand the composition and potential adaptive importance of the shared genomic region of elevated divergence, we assessed the functions of the genes to which loci within this outlier region aligned (**Table 2.8**). Five loci were present in the outlier regions of all three analyses, one of which aligned within an annotated gene, MDFIC. This locus was also identified as an outlier in both populations and the full data set. Of the six loci found only in the western population's outlier region, three aligned to the same protein-coding gene, glutamate metabotropic receptor 8, or Grm8. Of the five loci found only in the eastern population's outlier region, only two aligned within annotated protein-coding regions, both for the same gene (YY1-associated factor 2).

Although the genomic region of elevated divergence at the end of linkage group 19 was the only outlier region which directly aligned between the eastern and western populations, there were other areas of the genome which possessed multiple outlier regions in close proximity to each other on Atlantic cod linkage groups 2, 9, and 16 (Figure S2.5). Only one area of elevated divergence in the full data set aligned with an outlier region in the eastern population, and only one (different) area of elevated divergence in the full data set aligned with an outlier region in the western population (**Figure 2.8**).

2.4.4 *Overlap in outlier loci and annotations*

Putative loci under selection were identified using the three different methods employed by OutFLANK, Bayescan, and PCAdapt. These three outlier detection methods identified a total of 61 loci potentially under divergent selection in the eastern population, and 17 loci in the western population (**Table 2.9; Figure 2.9a&b**). No outlier loci were shared between the eastern and western populations.

All outlier loci identified in the western population, and outlier loci in the eastern population identified by two or more of the methods above, were checked for proximity to annotated genes in the Atlantic cod genome. Of the seventeen outlier loci in the western population, seven aligned to annotated gene regions, with four of these related to brain and neural development and function (**Table 2.10**). All seven annotated outlier loci were within their respective protein-coding gene sequence. Of the 20 outlier loci identified by two or more methods in the eastern population, 15 aligned to the Atlantic cod genome within annotated protein-coding sequences (**Table 2.11**). None of the annotated loci in the eastern and western populations aligned to the same genes as annotated loci in the opposite population.

2.4.5 *Outlier loci between eastern and western Pacific cod*

OutFLANK failed to find any putative loci under selection in the full data set, but a total of 519 outlier loci were identified using Bayescan and PCAdapt, 167 of which were identified as outliers by both programs (**Table 2.9**). Due to the long time since divergence and negligible gene flow between sampling sites in the eastern and western populations, Bayescan had the power to detect loci under either diverging or balancing selection in the full data set (**Figure 2.9c**); of the total number of outlier loci identified in Bayescan with neutral prior odds of 100 / 1000, 67 / 80% (437 / 356 loci) were putatively under diverging selection, and 32 / 20% (206 / 88 loci) under balancing selection.

Overall, outlier loci were evenly spread across the 23 Atlantic cod linkage groups, although no outlier locus identified in either the eastern or western population mapped to linkage group 18 (**Figure 2.10**). The full data set had notably more outlier loci which mapped to the first linkage group, a trend that was not replicated within the western or the eastern populations.

2.4.6 *Population structure according to neutral and putatively selected loci*

Principal component analyses (PCAs) were conducted on neutral loci and putative outlier loci separately, to determine if patterns of genetic differentiation would differ between loci primarily influenced by genetic drift and those potentially under selection. Loci identified as outliers were included in the analysis only if they had been identified in two or more of the programs listed above; since Bayescan was the only program which identified outlier loci under balancing selection, the PCA based on outlier loci for the full data set includes only loci under divergent selection.

The full data set comparing the eastern and western populations displayed the same patterns of separation across the two principal components with neutral and outlier loci; the

eastern and the western populations differentiated along the x axis, and the two western subpopulations differentiated along the y axis (**Figure 2.11**). However, the two western subpopulations were clustered more closely together with only outlier, as opposed to neutral, loci. This was likely a result of the greater differentiation between the eastern and western Pacific populations, compared to the divergence between the two subpopulations within the western Pacific. The western population also displayed the same separation across the two principal components with both neutral and outlier loci (**Figure 2.12a&b**).

In the eastern population, however, a different clustering pattern arose along the first principal component when the outlier and neutral loci were analyzed separately. With only neutral loci, the subpopulations segregated into two groups, delineated by the geographic gap in sampling between the Hecate Strait and Prince William Sound (**Figure 2.12c**). This pattern became more evident only after five outlier individuals from Unimak Pass were removed from the data set (Figure S2.6). However, when the PCA was conducted with putative outlier loci identified by two or more detection methods, the Kodiak Island individuals shifted position on the first principal component to span this gap, with the site's centroid closer to the Washington Coast and Hecate Strait sites than to the Aleutian Islands and the eastern Bering Sea sites (**Figure 2.12d**). In addition, the samples collected from Prince William Sound appear to consist of two different groups of fish; one with outlier allele frequencies more similar to the Aleutian Islands and Bering Sea samples, and one with outlier allele frequencies more similar to the Kodiak Island and the two southeastern sampling sites.

2.5 DISCUSSION

Our research is the first to observe differentiation across the geographic range of Pacific cod on a genome-wide scale, using thousands of single nucleotide polymorphisms. We produced sliding window analyses with adequate coverage after filtering for low marker density, and without a Pacific cod genome assembly. Comparisons of putatively adaptive loci and genome regions showed no overlap in signatures of adaptive divergence between eastern and western Pacific cod, with the exception of a single large outlier region that aligned to Atlantic cod linkage group 19. Therefore, this study provides no evidence that Pacific cod display parallel adaptation between the eastern and western Pacific. We also observed how the frequency and size of genomic regions of elevated divergence changed across three scenarios of divergence (**Figure 2.1; Table 2.1**). Although we found that the frequency of these outlier regions increased, and size decreased, from the first to the third divergence scenarios, only the comparison of outlier region frequency between the primary model of divergence with gene flow (eastern Pacific cod), and divergence without gene flow (full data set), was significant.

2.5.1 *Differentiation between eastern and western populations of Pacific cod*

Our study confirmed the presence of a very strong genetic break between the eastern and western populations (global $F_{st} = 0.29$; AMOVA $\Phi_{ST} = 0.58$), which adds to existing genetic evidence of long periods of isolation between the two populations in different glacial refugia (Canino et al., 2010).

While other population genetic studies of marine species distributed on both sides of the Atlantic or Pacific ocean have used microsatellites (McPherson, O'Reilly, & Taggart, 2011; O'Leary et al., 2007), and so direct quantitative comparisons of population-level Φ_{st} or F_{st} are

not appropriate, a more recent study using SNPs to compare east and west Atlantic cod found a maximum locus-specific $F_{st} = 0.60$ (Bradbury et al., 2010). Our research suggests that Pacific cod display greater pan-oceanic genetic differentiation than Atlantic cod, as we discovered several loci near fixation ($F_{st} > 0.99$) when comparing eastern and western Pacific populations.

Yet both eastern and western populations of Pacific and Atlantic cod were likely isolated in, and expanded from, separate glacial refugia during the Pleistocene (Bigg, 2014; Bigg et al., 2008; Canino et al., 2010). We can hypothesize that Atlantic cod population sizes in refugia may have been larger than those of Pacific cod, and so would have experienced less genetic drift during periods of allopatry. Alternatively, the higher trans-ocean differentiation that we observed in Pacific cod may be a function of geographic distance between sampling sites. Bradbury et al. (2010) sampled a latitudinal range spanning from approximately 50°N to 30°N on both sides of the Atlantic, whereas the northern extent of our sampling latitude was approximately 60°N in the east Pacific and approximately 38°N in the west Pacific. If we had sampled the eastern and western populations closer to the secondary contact zone between them, hypothesized to be north of Japan (Stroganov & Orlov, 2012), we may have observed a smaller per-locus maximum F_{st} and AMOVA Φ_{st} .

2.5.2 *Genomic regions of elevated divergence across divergence scenarios*

We observed a measurable increase in outlier region frequency from the eastern population, to the western population, to the full data set, despite having lower power to detect regions under selection in the western population and full data set due to elevated levels of neutral differentiation. This suggests that the correlation between outlier region frequency and divergence scenario may be stronger than is suggested here. A greater number of outlier regions in divergence without gene flow (the full data set) and the secondary model (west Pacific cod) is

unsurprising, because these diverging populations possess a combination of differences which accumulated prior to and during isolation, whereas in sympatric divergence (east Pacific cod), differences can only accumulate at loci that are resistant to introgression. According to simulations, the frequency of outlier regions should further increase when secondary contact is established after periods of allopatry, as erosion of genome-wide differentiation causes the rapid formation of new regions of elevated divergence (Yeaman, Aeschbacher, & Bürger, 2016). We therefore would have expected to see more outlier regions in the western population than in the full data set. Instead, we observed fewer regions, although the difference was non-significant.

Genomic regions of elevated divergence were larger within the eastern and western populations than in the full data set, although this difference was not statistically significant. This result may indicate that wider outlier regions form during divergence with gene flow, regardless of the model type, than in divergence without gene flow. However, exact region width was difficult to estimate consistently given differences in marker density between sliding window analyses; the increased marker density in the full data set may have allowed for more fine-scale delineation of outlier region boundaries, creating the false signal of narrower regions of elevated divergence.

Nevertheless, the observed heterogeneity in genome-wide differentiation across the three divergence scenarios is consistent with existing research. It is challenging to compare exact results across studies due to wide variation in study methodology and species' demographic history (Wolf & Ellegren, 2016), but we can make several comparisons between our findings and the existing literature. Similar to Gagnaire et al. (2013), who found that outlier regions were randomly distributed across linkage groups in lake whitefish, genomic regions of elevated divergence for all three scenarios of divergence in Pacific cod were widely distributed, without a

clear trend within or between groups. This is in contrast to genomic regions of elevated sequence divergence in *Ficedula* flycatchers, which Ellegren et al. (2012) found to be overrepresented at the end of chromosomes, the likely location of centromeres in that species.

Genome-wide analysis of Atlantic cod identified chromosomal inversions as the underlying architecture of large genomic regions of elevated divergence, which are key to ecotype divergence; polymorphic inversions spanning from 5 to 13Mb on linkage groups 2, 7, and 12 have been found between multiple populations across the species' range (Barney et al., 2017; Barth et al., 2017; Berg et al., 2017; Hemmer-Hansen et al., 2013; Kirubakaran et al., 2018; Sinclair-Waters et al., 2018; Sodeland et al., 2016). If large inversions were facilitating similar adaptive divergence in Pacific cod, we would have expected to see large outlier regions in the eastern and western populations, and potentially in the full data set, aligned to the same positions on the Atlantic cod genome. Yet we only observed one outlier region present in both eastern and western populations, and the largest outlier region was only ~1.29 Mb (in the eastern population). It is possible that smaller inversions are present in Pacific cod, but do not play as crucial a role in maintaining divergence between ecotypes on both sides of the Pacific Ocean, and so were not identified by our comparison of outlier region location between the eastern and western populations.

A better understanding of the processes leading to divergence with gene flow in this species could be achieved through analyses with higher marker density and a Pacific cod linkage map, which would allow for more fine-scale determination of the width of outlier regions and remove some of the uncertainties discussed above. In combination with the calculation of recombination rates and nucleotide diversity, further analyses may reveal regions of elevated divergence that have arisen due to suppressed recombination in inversions and centromeres, a

common finding in previous studies (Barth et al., 2017; Berg et al., 2017; Ellegren et al., 2012; Sodeland et al., 2016). Obtaining samples from the secondary contact zone between eastern and western populations may help to distinguish between outlier regions in the full data set that are resistant to introgression, and therefore play a key role in maintaining adaptive divergence, and those which are easily eroded when gene flow is re-established. Future studies could also work toward correlating the observed genomic regions of elevated divergence with ecological adaptations, as in Atlantic cod (Berg et al., 2017; Bradbury et al., 2010).

2.5.3 *Differing signatures of adaptation in eastern and western Pacific cod*

There was a marked lack of overlap in significantly diverged genomic regions and outlier loci between eastern and western Pacific cod, suggesting that these two populations are either under differential selective pressures, or they are adapting to similar local pressures through different genes. Investigation into the function of outlier loci, and loci within outlier regions, supports the hypothesis that, at least for certain traits, parallel phenotypic evolution is occurring without a parallel genetic basis (discussed more below, in *Function of Diverged Loci*). This lack of parallel patterns at the genetic level is in contrast to Atlantic cod, where parallel clines in response to bottom temperature were found on both sides of the Atlantic (Bradbury et al. 2010), and Atlantic herring, which showed parallel differentiation between groups in the eastern and western Atlantic (Lamichhaney et al., 2017).

It is plausible that certain latitudinal environmental clines which drive selection in the eastern Pacific are not as prevalent as other local selection pressures in Korean coastal waters, where samples were from a much smaller geographic area (**Figure 2.2**). However, we did expect to find some signature of parallel adaptation in response to thermal regimes. Pacific cod are sensitive to water temperature during early life history stages, and the oceanographic currents

around the Korean peninsula create a steep thermal and salinity gradient (Chang et al., 2004; Hwang et al., 2014) that may replicate the larger-scale latitudinal clines in the northeastern Pacific. Although parallel adaptation to thermal clines on either side of the Atlantic Ocean has been observed in Atlantic cod (Bradbury et al. 2010), eastern and western populations of Pacific cod appear to be more highly diverged from each other than eastern and western Atlantic cod. It is also possible that the two species respond differently to selective pressures. For example, studies have found evidence that the *pantophysin* gene is under significant positive selection in Atlantic cod (Pogson & Mesa, 2004) and walleye pollock (Canino & Bentzen, 2004), but not under significant selection in Pacific cod (M.F. Canino, NOAA NMFS, unpublished data). Our results are more similar to a recent study on three-spine stickleback, which found that highly diverged regions of the genome failed to align across five sets of lake-river population pairs despite similar selective pressures such as changes in salinity and parasite diversity (Feulner et al., 2015). Average pairwise divergence between the lake-river pairs was similar to the global differentiation between eastern and western Pacific cod populations ($F_{st} = 0.10 - 0.28$) (Feulner et al., 2015).

Before ruling out parallel adaptation in Pacific cod, future studies could collect samples from a wider geographic range in the western Pacific, and focus on the sequencing of expressed sequence tags to ensure data collection at loci within protein-coding regions (Bradbury et al., 2010; Parkinson & Blaxter, 2009) or whole-genome sequencing for greater genome-wide coverage. We have identified genomic locations where the eastern and western populations possessed regions of elevated divergence in close proximity to each other (Figure S2.5), which may be of interest to future studies with finer-scale discernment of outlier region boundaries.

Local adaptation also had a differential impact on population structure in the eastern and western Pacific, as clustering patterns from the principal component analysis (PCA) on outlier loci differed from the PCA on neutral loci in the eastern population, but not the western population. The similarity in outlier loci allele frequencies between Kodiak Island and the more southern Hecate Strait / Washington coast samples, rather than the neighboring Unimak Pass and latitudinally-proximate Adak Island, could be a result of stock-specific selection pressures. Pacific cod spawning at Kodiak Island are believed to belong to the Gulf of Alaska stock, which also includes the Hecate Strait and Washington Coast spawning grounds. Pacific cod spawning at Unimak Pass are believed to belong to the Bering Sea stock, due to the direction of oceanographic currents in the area (Stabeno et al., 1999) which drive advection of larvae from nearby northern rock sole spawning grounds into the eastern Bering Sea (Lanksbury et al., 2007). Whereas the Gulf of Alaska is a semi-enclosed basin where variability in currents and nutrient input are related to changes in fresh-water flux, water temperatures, and regional winds (Stabeno et al., 2004), the eastern Bering Sea covers an expanse of the continental shelf, and the ecosystem is strongly influenced by the seasonal advance and retreat of sea ice and correlated phytoplankton blooms (Alexander & Niebauer, 2003; Hunt & Stabeno, 2002; Stabeno et al., 1999). These dramatically different ecosystems and oceanographic regions likely put very different selection pressures on the Pacific cod populations that inhabit them.

We also observed the separation of individuals collected from the Prince William Sound into two groups. It is possible that samples were collected from a mixed aggregate of individuals from both the Gulf of Alaska stock and the Bering Sea / Aleutian Islands stock. Evidence from genetic studies around the Korean peninsula suggest that the formation of mixed-stock aggregates during the spawning season may be more prevalent than previously realized (Gwak &

Nakayama, 2011; see Chapter 1). Alternatively, the presence of two distinct groups of Pacific cod could be indicative of undetected subpopulation structure, such as the presence of a migratory and stationary stock at the same spawning grounds.

2.5.4 *Limitations in identifying outlier loci and regions*

The results of our outlier testing serve as an example of some of the limitations of outlier detection methods, many of which have been discussed in the literature (Hoban et al., 2016; Kelley et al., 2006; Lotterhos & Whitlock, 2014). The hierarchical population structure of the full data set deviates from Wright's island model (Wright, 1949), which is particularly important for the outlier detection method used by Bayescan. Hierarchical population structure causes a large proportion of false positives in Bayescan, as does the isolation-by-distance pattern found in the eastern population (Lotterhos & Whitlock, 2014; Whitlock & Lotterhos, 2015). Bayescan did identify a relatively high number of outlier loci in the full data set (14.98% and 10.36% of all loci, with prior odds of 100 and 1000, respectively). In the western population, the presence of only two genetically distinct subpopulations greatly reduces the power of Bayescan analysis while increasing the sampling variance of F_{st} (Foll, 2012). This may explain why Bayescan identified only two outlier loci in our western Pacific cod samples, in contrast with the 17 loci identified by OutFLANK.

Although OutFLANK has been shown to have higher power and fewer false positives under a range of population structures, including isolation-by-distance and one- and two-refugia expansion models (Whitlock & Lotterhos, 2015), the assumption that fewer than 5% of loci are under selection may be problematic in populations with very high divergence where per-locus F_{st} spans the entire range of 0 – 1. Simulations to test OutFLANK's performance were completed with an overall F_{st} of 0.05 (Whitlock & Lotterhos, 2015), which is nearly an order of magnitude

smaller than the global F_{st} of the full Pacific cod data set (0.29). Furthermore, visual assessment of the empirical versus the predicted distribution in OutFLANK, which is suggested by Whitlock & Lotterhos (2015) as a way to verify the fit of OutFLANK's null model, shows more of a mismatch between the predicted and the empirical distribution in the full data set than exists in the eastern or the western population alone (Figure S2.7). Our application of three different outlier tests was an effort to identify and account for these potentially unreliable results, although there may still be false positive (and false negative) signals of selection in our data. In addition, sliding window analyses like those conducted in this study can improve on results of locus-specific outlier tests (Hoban et al., 2016).

However, the calculation of a kernel-smoothing weighted average F_{st} for our sliding window analyses did face its own limitations, the greatest being marker density. Concern regarding our relatively low marker density led us to conduct several regression analyses, which showed a significant correlation between outlier region width and marker density in the western population. Although there was not a similar correlation in the eastern population or the full data set, or any significant correlation between outlier region frequency and marker density, these linear regressions were conducted on a relatively narrow range of marker densities. Our analysis used an average of 7.17 to 10.83 markers per weighted average, whereas the first implementation of the kernel-smoothing weighted average method by Hohenlohe et al. (2010) used an average of 81.6 markers per weighted average. It is therefore possible that a significant correlation exists that was not detected. Were outlier region width or frequency correlated with marker density, then the increasing marker density in this study from the eastern population, to the western population, to the full data set would be a confounding factor in the analysis relating the three scenarios of divergence to our Pacific cod data set.

Marker density may also have affected our ability to identify chromosomal rearrangements in Pacific cod. With low marker density, a large inversion could be detected as a series of smaller outlier regions; the random selection of loci within an inversion will invariably include some loci that have not differentiated significantly, which may lead to non-significant weighted averages in a sliding window analysis. In addition, prior studies that detected chromosomal inversions have benefitted from a species-specific linkage map (Berg et al., 2017; Hoffman & Rieseberg, 2008), whereas we assumed that locus position was identical in Atlantic and Pacific cod.

2.5.5 *Function of diverged loci*

The only annotated locus that was shared between the eastern and the western populations was within the linkage group 19 outlier region. This locus aligned within a gene similar to a section of the MyoD family inhibitor domain containing gene sequence (MDFIC). MDFIC is a transcriptional activator or repressor related to the *Zic* family of proteins (Aruga, 2004). Multiple studies have identified this family of zinc-finger proteins as playing several critical roles in the neural development of vertebrates, including processes such as neurogenesis, myogenesis, skeletal patterning, and left-right axis establishment (Aruga, 2004; Houtmeyers et al., 2013; Mizugishi et al., 2001).

Five outlier loci that aligned within annotated protein-coding sequences were functionally linked to circadian rhythms. Three loci in the western population's linkage group 19 outlier region, one of which was identified as an outlier locus in the full data set, aligned within a gene sequence involved in circadian entrainment. Entrainment is the process by which the intrinsic period of circadian clocks is adjusted to the environment and then stabilizes according to the environmental cycle (Zhdanova & Reeb, 2005). Light is the strongest synchronizer to entrain

circadian rhythms, although other factors, including temperature, food availability, and predation risk are also effective (Zhdanova & Reeb, 2005). A fourth locus in the western population, identified as an outlier in OutFLANK, aligned within the coding sequence for adenylyl cyclase type 1 (ADCY1). This protein regulates the circadian rhythm of daytime contrast sensitivity, or the relative ability to detect spatial variation in light intensity (Hwang et al., 2013). This regulation likely occurs by modulating the synthesis of cyclic AMP in the retina (Masada et al., 2008). The same locus was identified as an outlier in the full data set by Bayescan and PCAdapt.

Although no annotated outlier loci in the eastern population aligned to genes directly functioning in circadian clocks, one locus aligned to a gene which is expressed exclusively in photoreceptor cells in the retina (ABCA4), and is part of the pathway converting light to electric signals that travel on to the brain (Maugeri et al., 2000). This is one of the major pathways in light-dependent circadian systems. Since we only annotated a portion of outlier loci in the eastern population to avoid potential false positives, it is also possible that another non-annotated outlier locus was aligned within a circadian clock gene.

The divergence of these annotated outlier loci suggest that circadian rhythms linked to photoreception may play a role in the variation between Pacific cod subpopulations within and between the eastern and western populations. Photoperiod-based circadian rhythms drive a diverse range of behaviors in fish species, including locomotor activity, reproduction, feeding and sleep (Zhdanova & Reeb, 2005). Shortened photoperiods have been shown to significantly affect reproductive behavior of fish species such as three-spine stickleback (Yeates-Burghart et al., 2009) or masu salmon (Zhdanova & Reeb, 2005), by stimulating gonadal maturation. Recently published evidence for parallel adaptation in Atlantic herring suggests that a thyroid-stimulating hormone receptor involved in photoperiodic regulation differentiates between spring

and autumn spawners (Lamichhaney et al., 2017). In vertebrates, photoperiod response generally increases with increasing latitude, potentially due to a narrower window of opportunity for reproduction and development at more northern latitudes (Zhdanova & Reeb, 2005).

Photoreception can also be important for the direction of individual migration; for example, monarch butterflies may use the angle of polarized light for orientation during their long migrations to breeding grounds (Reppert, Gegear, & Merlin, 2010).

Given the role of circadian clock genes in reproductive timing and migration in other organisms, these genes may help regulate spawn timing and associated migrations in Pacific cod, which varies geographically. For example, in the northeastern subpopulations, cod move from shallow, middle-upper shelf feeding groups to aggregate in the deeper areas of the outer shelf (Gustafson et al., 2000), whereas cod along the southern Korean coast migrate to shallow coastal waters to spawn (Chung, Kim & Kang, 2013), and cod along the western Korean coast spawn in broader and less concentrated aggregates throughout the Yellow Sea (Zhang, 1984). Spawn timing differs by latitude in the eastern population; along the more southern British Columbian coastline, Pacific cod spawn from February to March, whereas more northern populations spawn from late December until April in the eastern Bering Sea, and until May in the western Bering Sea (Gustafson et al., 2000). In the western population, Korean stocks spawn from December to late February (Chung, Kim, & Kang 2013). Further research utilizing a greater number of genetic markers that specifically contrasts samples of Pacific cod which vary in spawn timing may help to reveal the genetic and physiological underpinnings of Pacific cod spawning behavior.

2.5.6 *Conclusions*

This study has provided key insight into adaptive divergence with gene flow in Pacific cod across its pan-oceanic range. Early genomic studies of non-model species like this one, while

constrained by marker density, have proven to be critical starting points for research on adaptive evolution and the underlying genomic architecture (Brieuc et al., 2015). Our research established marked differences in patterns of adaptation between eastern and western populations of Pacific cod, and identified and annotated highly differentiated loci of interest.

We also detailed new trends in the frequency, width and location of genomic regions of elevated divergence across three scenarios of divergence within the same species. Further research can build on this work, with promising directions including the development of Pacific cod linkage maps for more fine-scale observations of genomic regions of elevated divergence, and the collection of relevant environmental data to link such outlier regions to specific local adaptations.

2.6 FIGURES

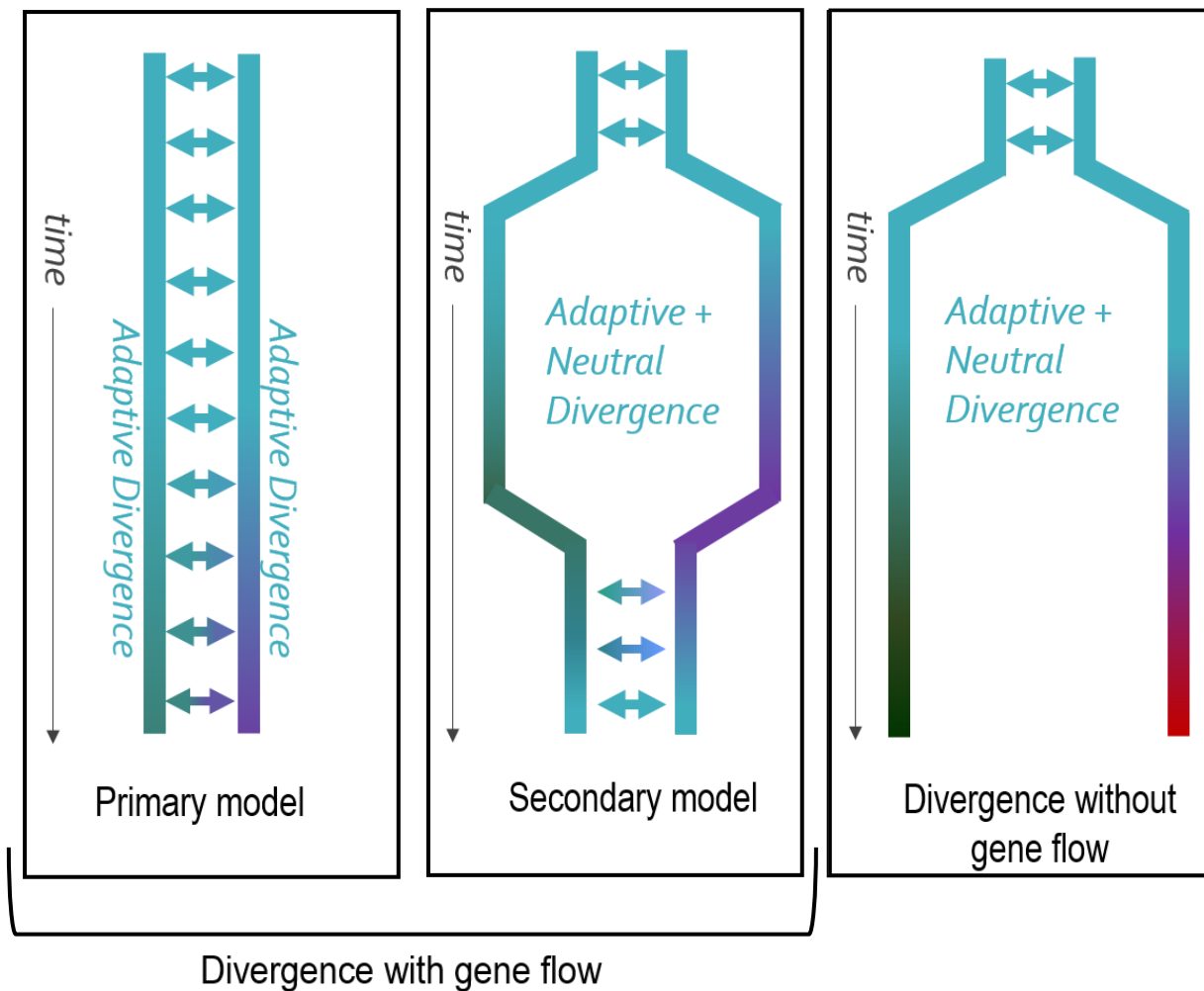


Figure 2.1. Three scenarios of divergence through time. The two bars in each diagram represent two lineages belonging to the same ancestral population. Arrows between lineages indicate gene flow. Accumulation of genetic differences is represented by the change in color of each lineage. The type of divergence accumulated – adaptive or neutral – is identified in italics for each scenario.

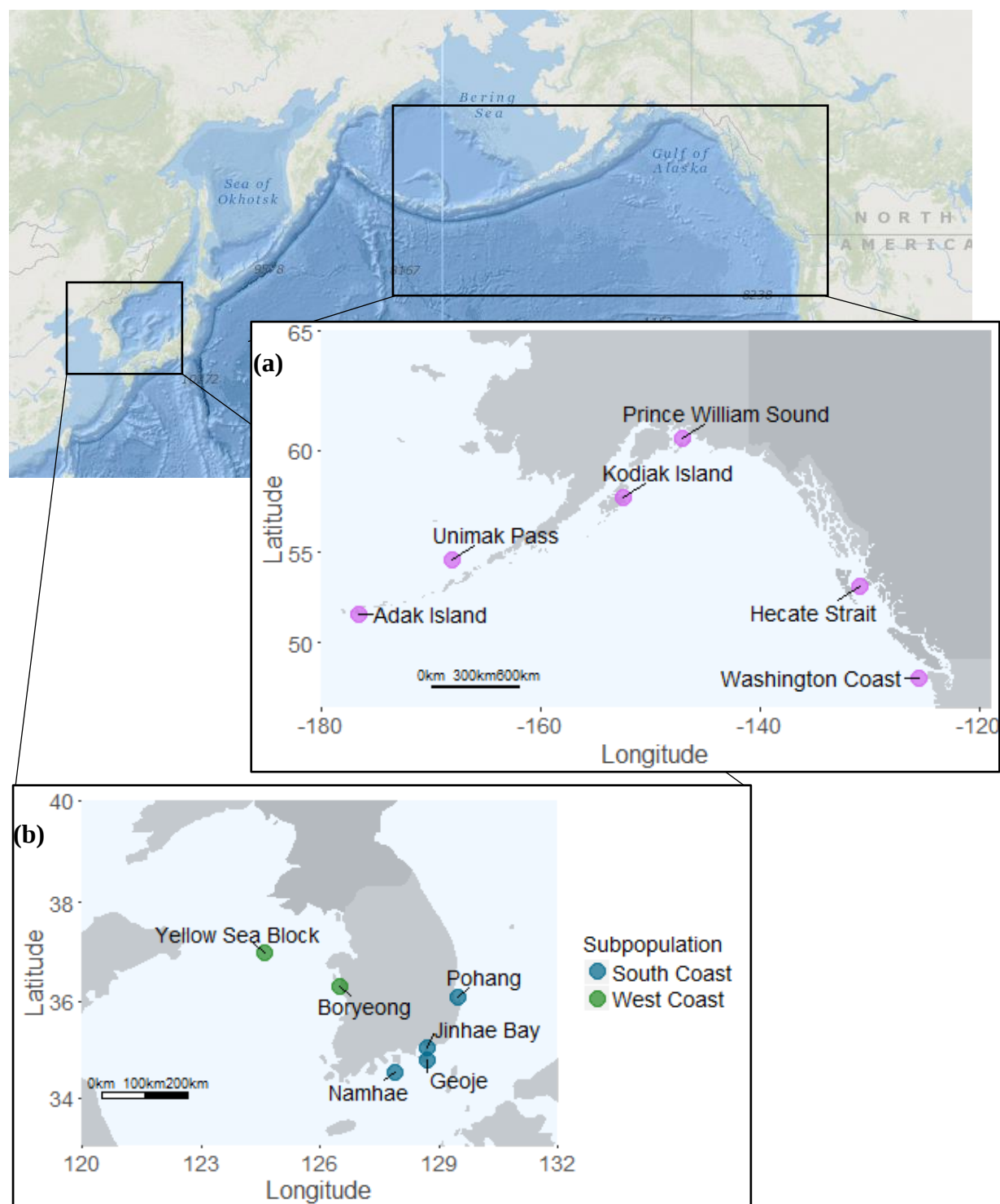


Figure 2.2. Sampling sites used in this study for the **(a)** western and **(b)** eastern population.

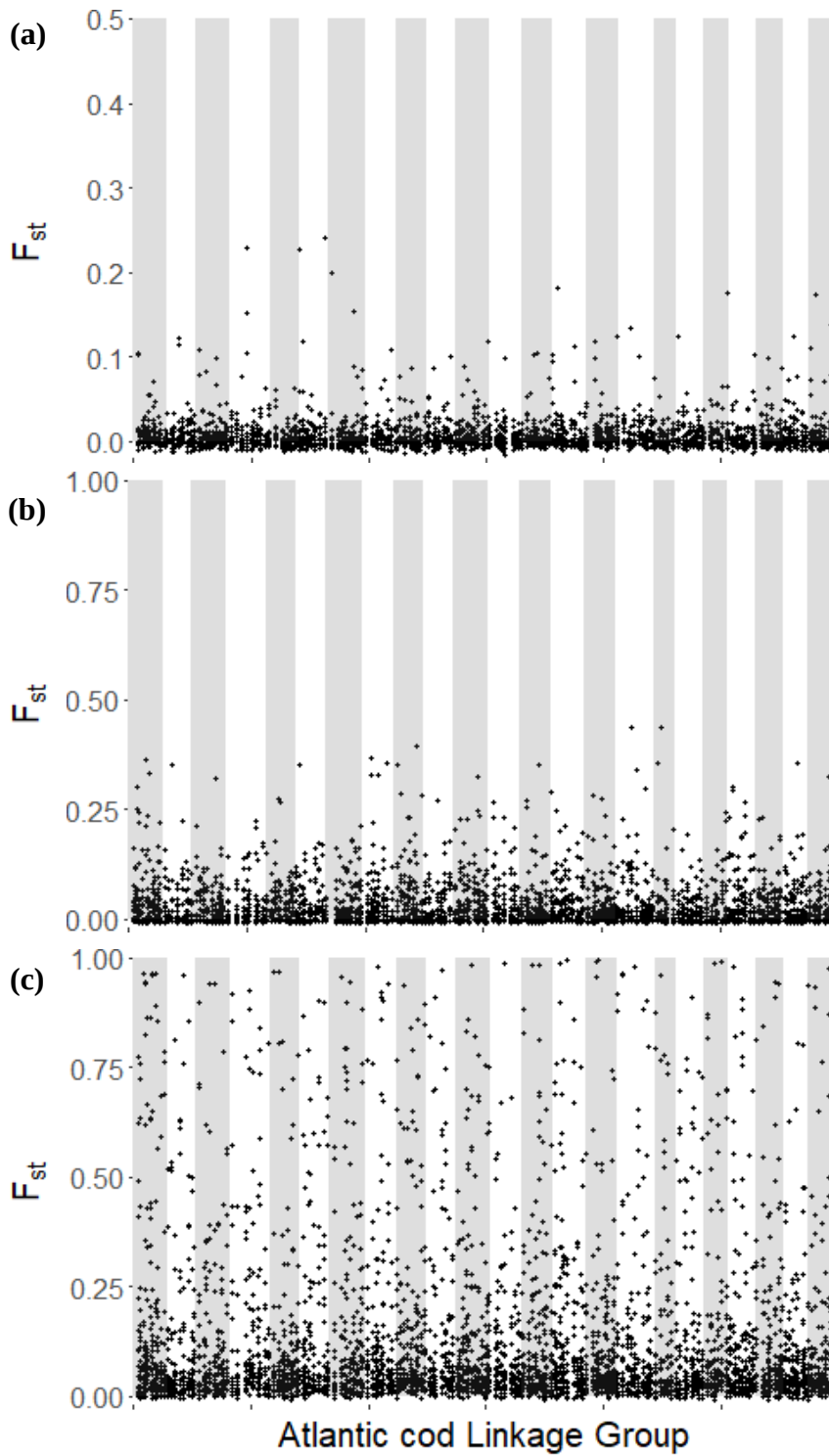


Figure 2.3. Manhattan plots displaying per-locus F_{st} . Per-locus F_{st} produced for **(a)** eastern population, **(b)** western population, and **(c)** full data set. Shading delineates Atlantic cod linkage groups. Loci positions were placed in 200kb bins.

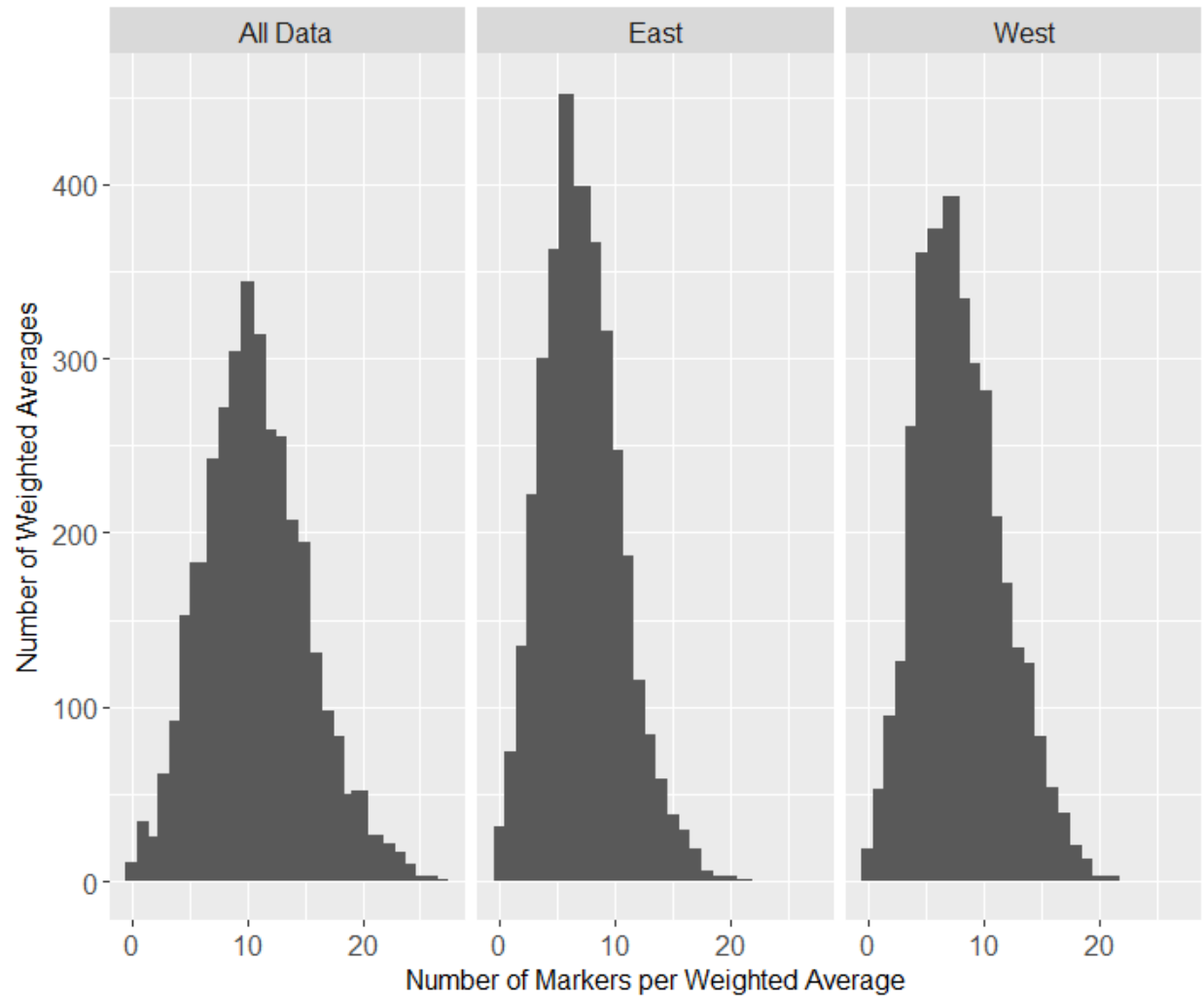


Figure 2.4. Number of markers (RAD loci) used to calculate each weighted average of F_{st} for sliding window analysis. Distributions for the full data set (mean = 10.83), the eastern population (mean = 7.17), and the western population (mean = 8.11) are shown separately.

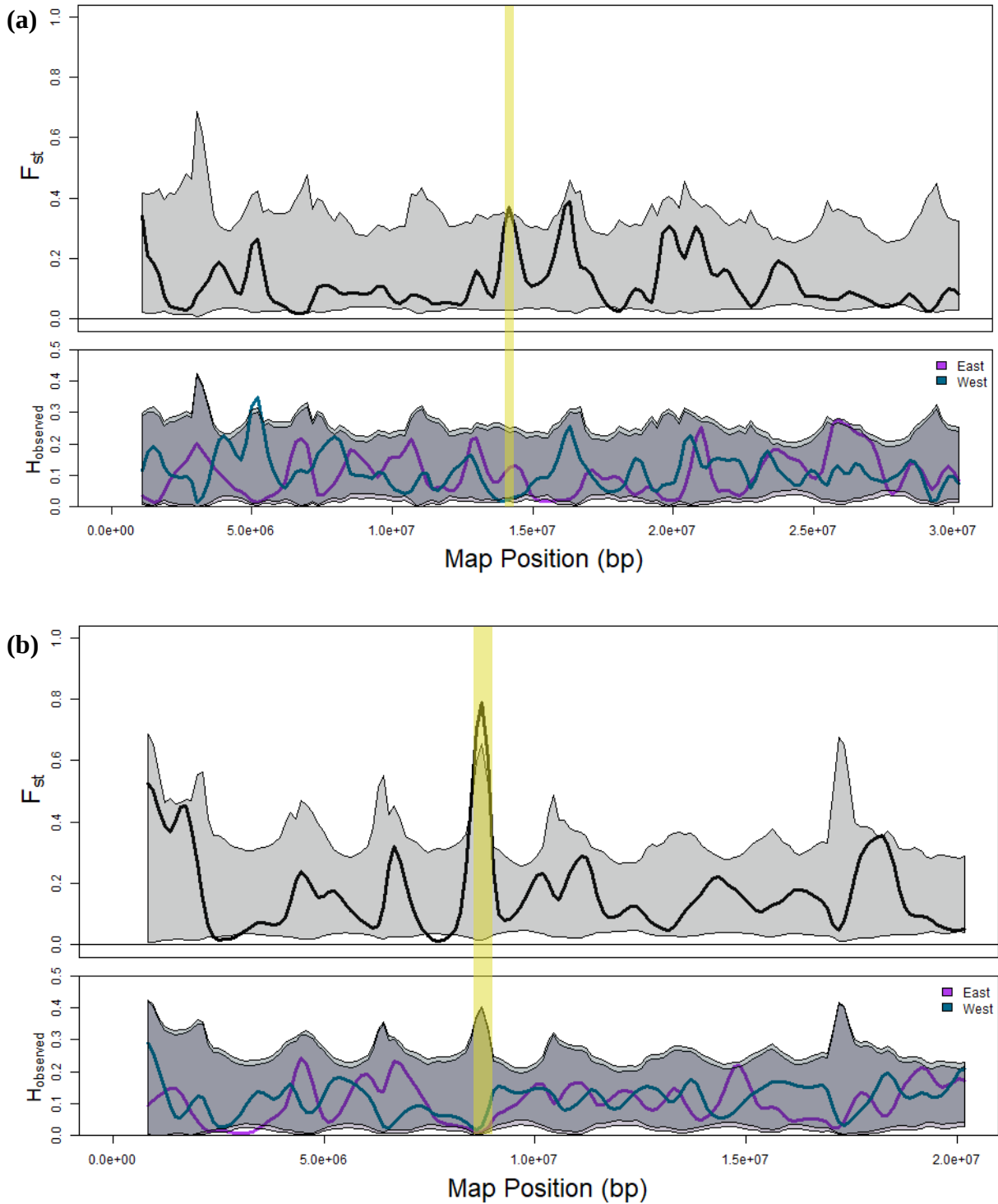


Figure 2.5. Kernel-smoothing weighted average of F_{st} (top) and H_o (bottom) with regions of elevated divergence highlighted. Graphs displayed for Atlantic cod linkage groups **(a)** 16, and **(b)** 19, which possess outlier regions associated with reduced within-population H_o that is **(a)** significant in the western population, and **(b)** non-significant in both populations

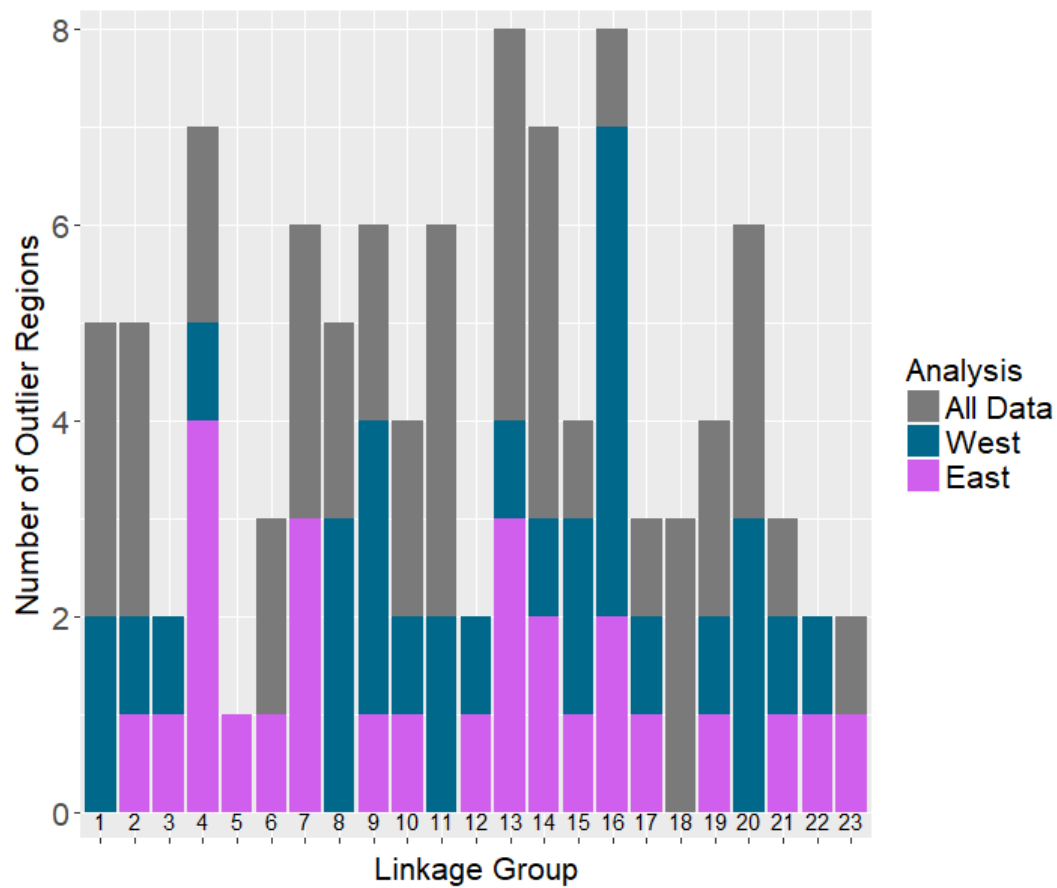


Figure 2.6. Number of genomic regions of elevated divergence per Atlantic cod linkage group. Color of bar indicates whether the outlier regions were identified in the full data set, the western population, or the eastern population.

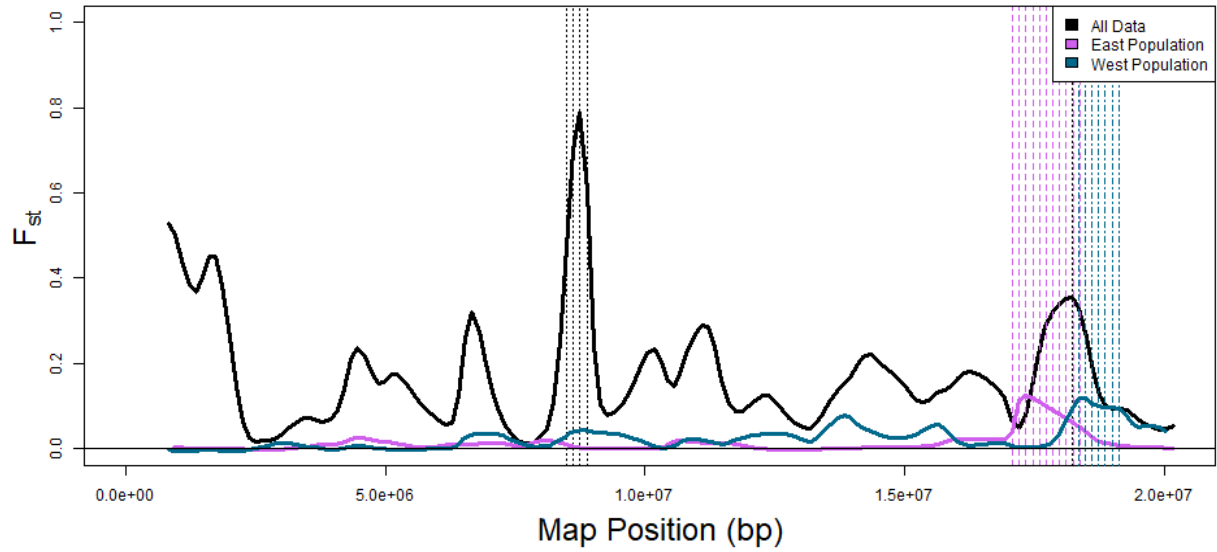


Figure 2.7. Kernel-smoothing weighted average F_{st} along Atlantic cod linkage group 19. Averages shown for the full data set (black), the eastern population (purple), and the western population (blue). Dashed lines mark the positions of weighted averages which exceeded the 95% confidence interval, indicative of genomic regions of elevated divergence.

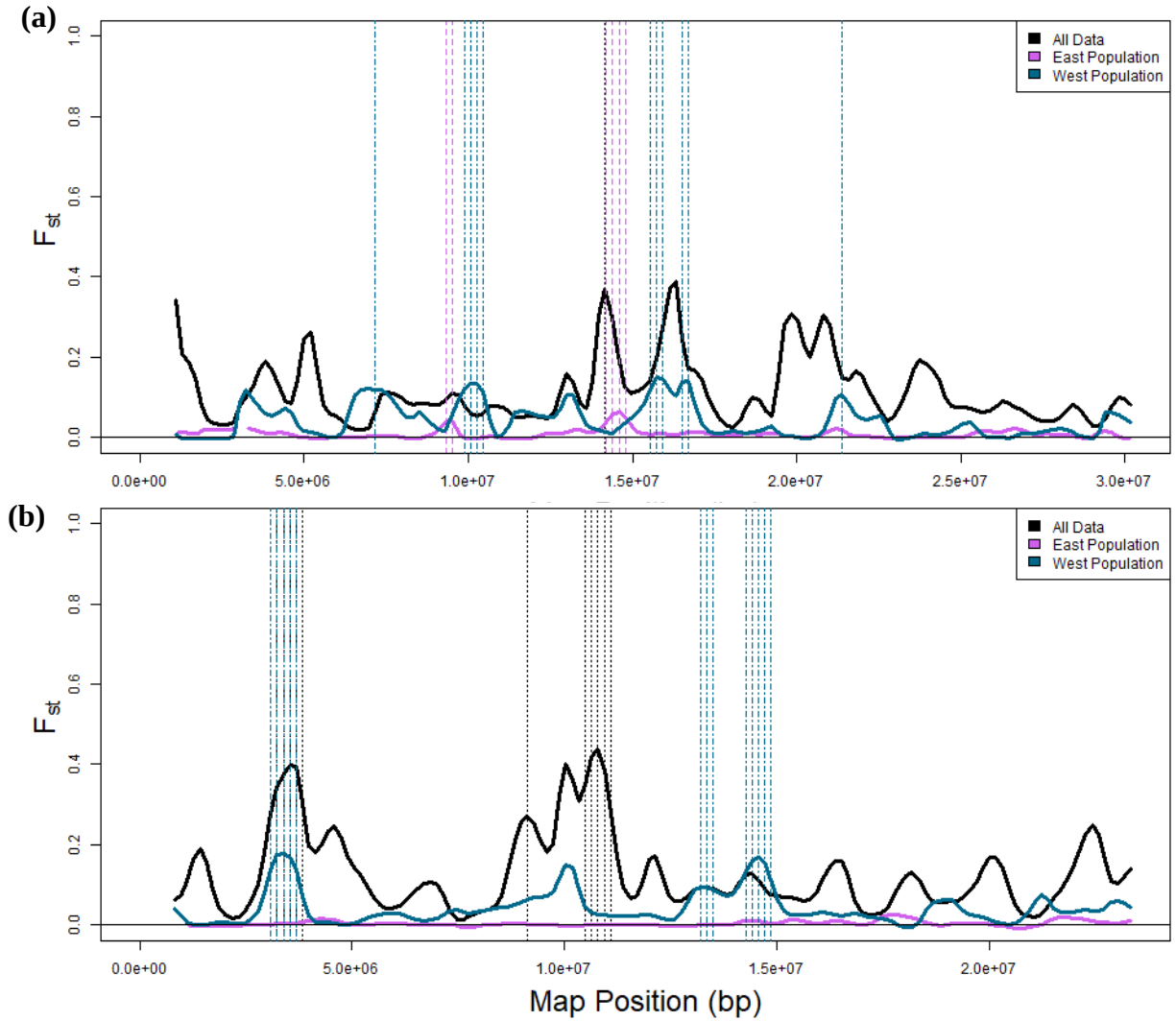


Figure 2.8. Kernel-smoothing weighted average F_{st} along Atlantic cod linkage groups **(a)** 16 and **(b)** 20. Averages shown for the full data set (black), the eastern population (purple), and the western population (blue). Dashed lines mark the positions of weighted averages which exceeded the 95% confidence interval, indicative of a genomic region of elevated divergence. These linkage groups included outlier regions in the full data set which aligned with outlier regions in the **(a)** eastern and **(b)** western populations.

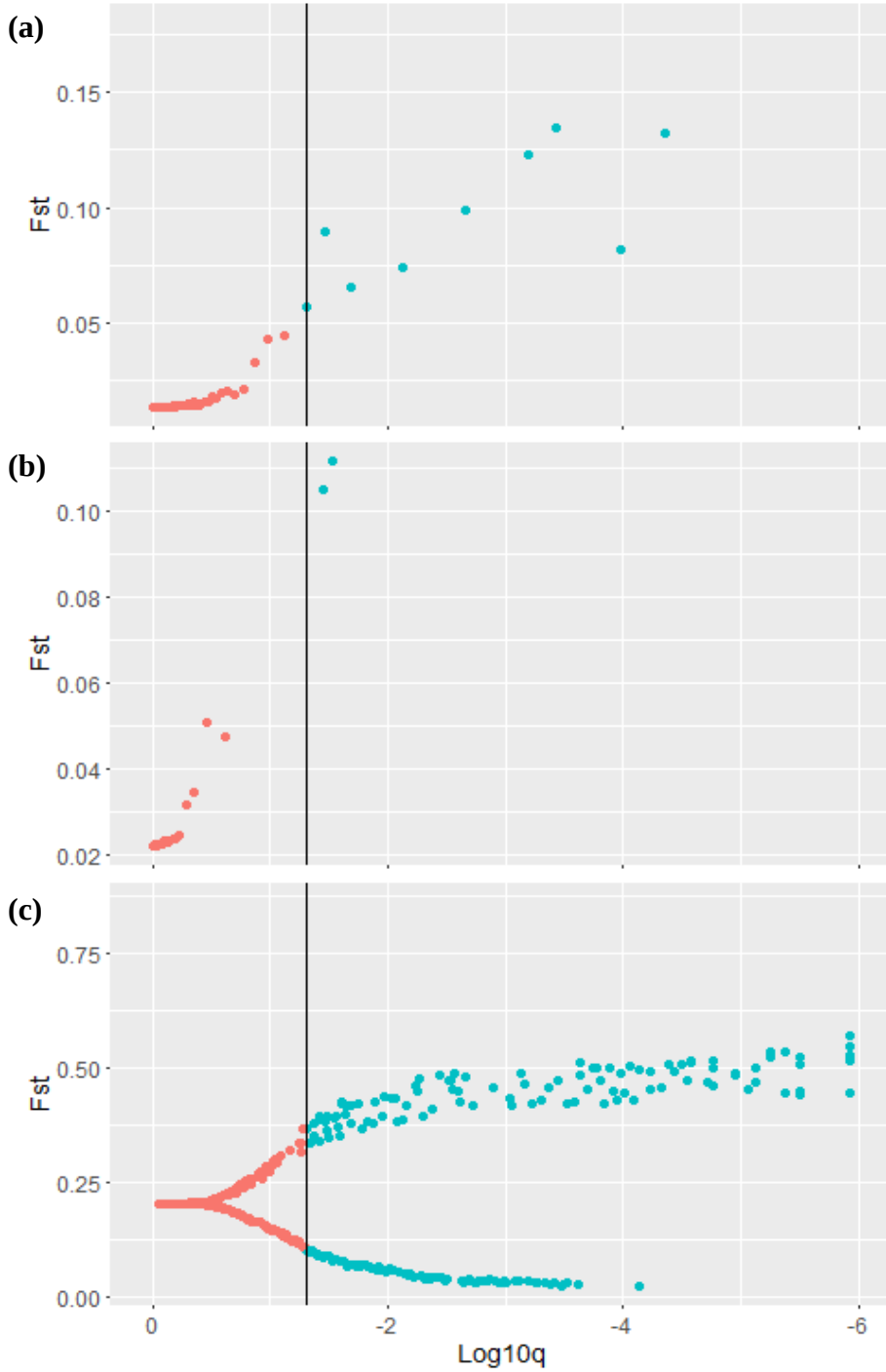


Figure 2.9. Bayescan output for the **(a)** eastern population, **(b)** western population, and **(c)** full data set, with a conservative neutral prior odds of 1000. Per-locus differentiation (F_{st}) is displayed on the y axis, and the $\log_{10}q$ value on the x axis. Vertical line indicates the false discovery rate used to identify putative outlier loci (0.05), which are highlighted blue in this plot. All points at the upper limit of the x axis had $\log_{10}q$ values of infinity.

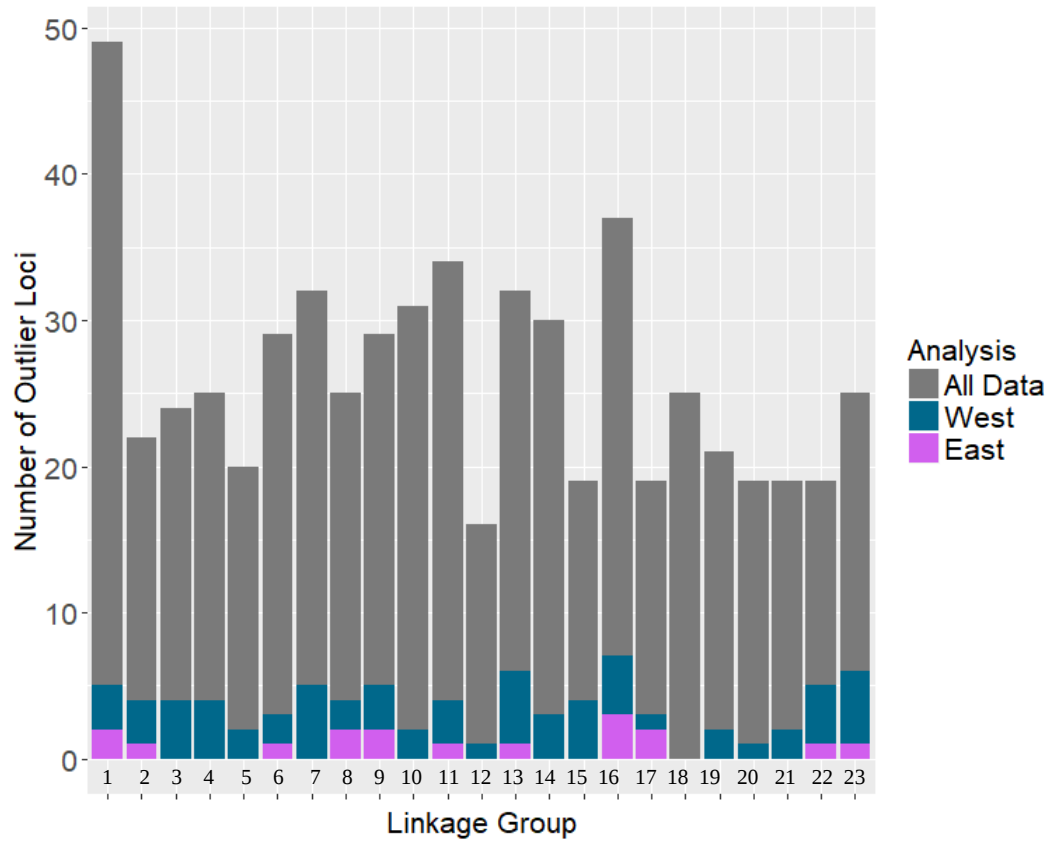


Figure 2.10. Number of putative outlier loci identified per Atlantic cod linkage group. Color of bar indicates whether the loci were identified in the full data set, the western population, or the eastern population.

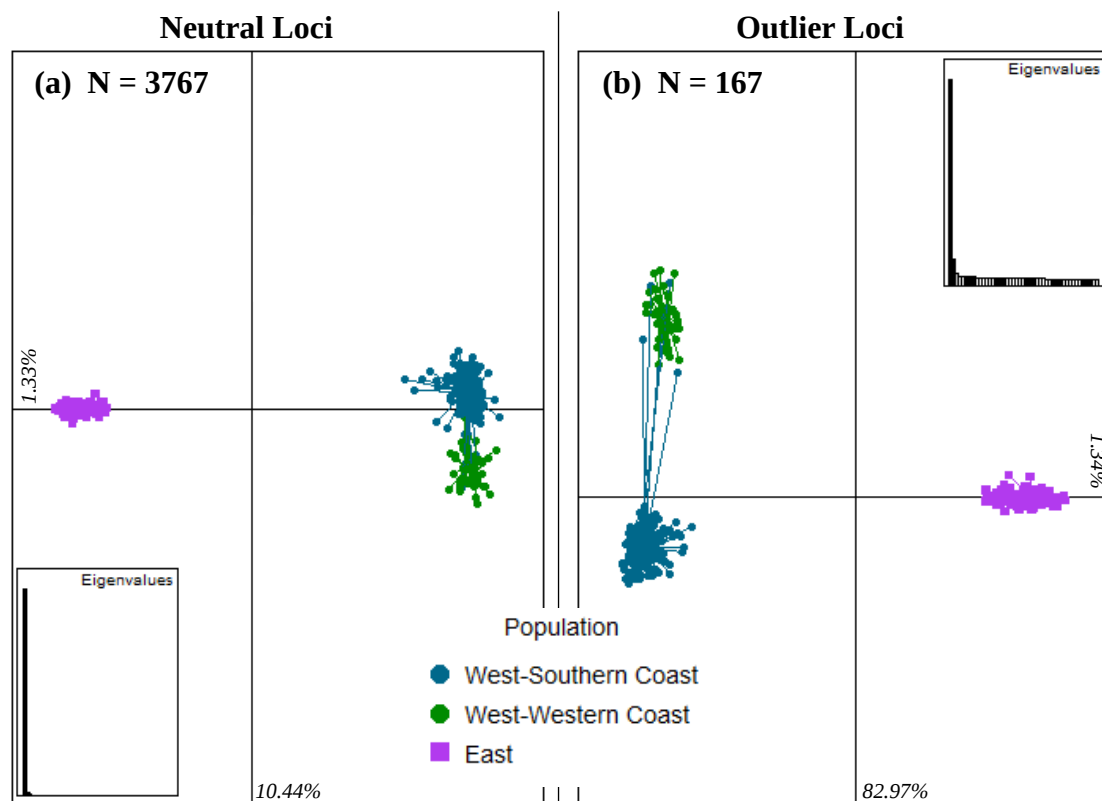


Figure 2.11. Principal component analysis on **(a)** neutral versus **(b)** outlier loci, for the full data set. Percent of variation explained by each principal component is labeled on the axes; the annotation at top of graph provides number of loci used in each analysis.

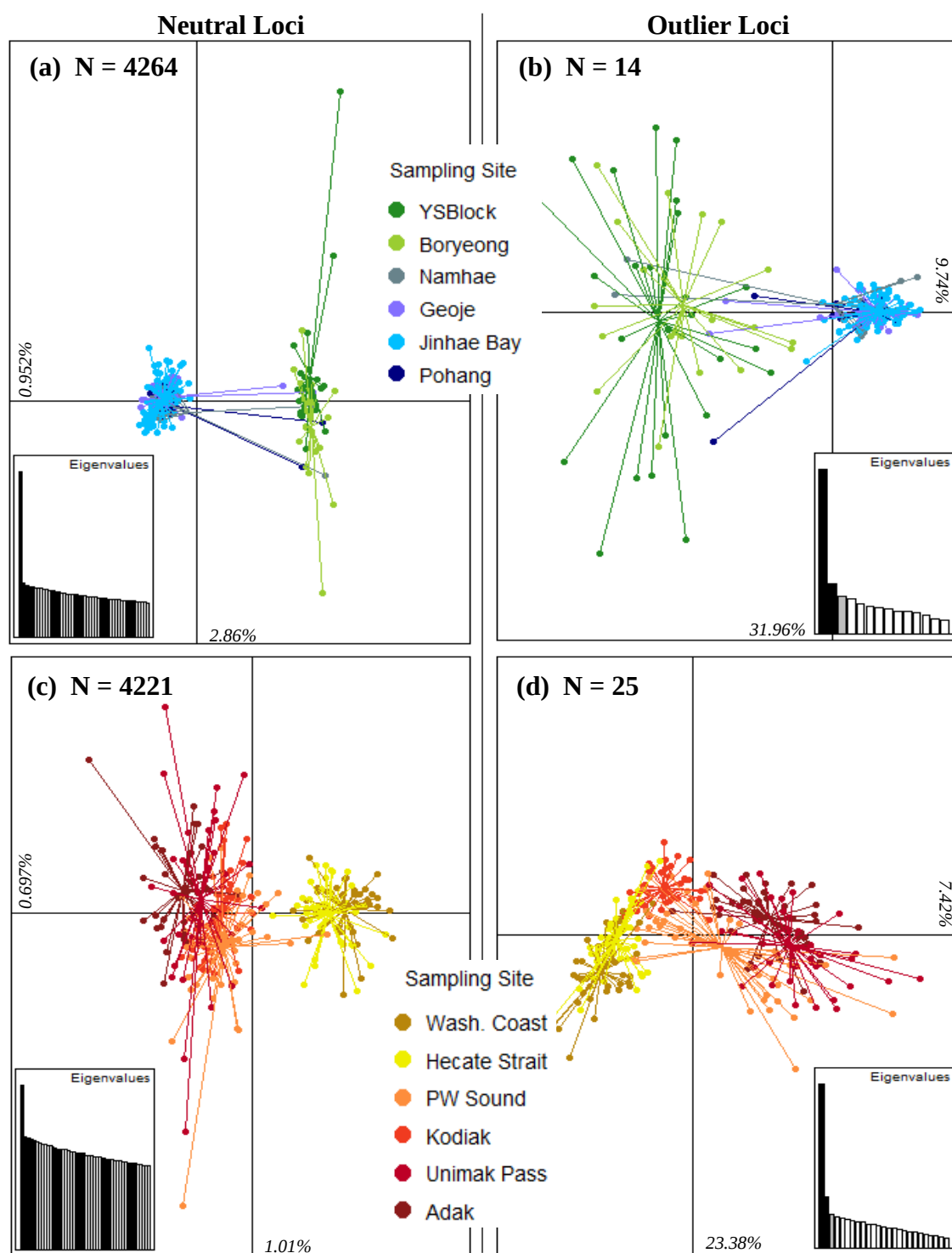


Figure 2.12. Principal component analysis on neutral versus outlier loci, for the western (**a & b**) and the eastern (**c & d**) populations. Percent of variation explained by each principal component is labeled on the axes; annotation at top of graph provides number of loci used in each analysis.

2.7 TABLES

Table 2.1. Divergence scenarios represented by populations of Pacific cod. Scenarios are ordered by increasing genetic divergence in the associated Pacific cod population.

Divergence Process	Pacific cod group	Description	
Scenario #1: Divergence with gene flow, primary model	Eastern population	+ Coastal subpopulations within the eastern population of Pacific cod, where each subpopulation is a single aggregate. + Isolation-by-distance gene flow pattern, no evidence of allopatric divergence.	<i>Lowest Divergence</i>
Scenario #2: Divergence with gene flow, secondary model	Western population	+ Subpopulations on the western and southern Korean coast, with each subpopulation consisting of pooled aggregates that are not significantly differentiated. + Hypothesized allopatric divergence in glacial refugia, currently display low levels of gene flow.	
Scenario #3: Divergence without gene flow	Eastern v. Western populations	+ A pan-oceanic comparison between the eastern and western populations of Pacific cod. + Distinct lineages with evidence for allopatric divergence in separate glacial refugia, currently display negligible levels gene flow.	<i>Highest Divergence</i>

Table 2.2. Pacific cod tissue or fin clip samples used in this study. ‘Number of samples collected’ lists total number of tissue or fin clip samples collected from each sampling site, whereas ‘number of samples retained’ includes only those samples used in final analyses. Samples removed from ‘collected’ to ‘retained’ include those with poor DNA quality, possible contamination, and missing data. *Samples provided by 1. Dr. Sukyung Kang, Korean National Institute of Fisheries Science (fin clip), 2. Dr. WooSeok Gwak, Gyeongsang National University (tissue), 3. National Marine Fisheries Service (fin clip), and 4. Prince William Sound Science Center (fin clip).*

Sampling Site	Collection Date(s)	N. Samples Collected	N. Samples Retained	Population, subpopulation
Yellow Sea Block ¹	12/2015	30	25	West, western Korean coast
Boryeong ²	01/2007	22	21	West, western Korean coast
Namhae ¹	02/2015	16	11	West, southern Korean coast
Geoje ¹	01/2015	34	28	West, southern Korean coast
	02/2014	22	22	West, southern Korean coast
Jinhae Bay ²	12/2007, 02/2008	96	81	West, southern Korean coast
Pohang ¹	01,02,03/2015	31	31	West, southern Korean coast
Washington Coast ³	05/2005	48	41	East, Washington Coast
Hecate Strait ³	03/2004	48	46	East, Hecate Strait
Prince William Sound ⁴	03/2012	48	47	East, Prince William Sound
Kodiak Island ³	03/2003	48	43	East, Kodiak Island
Unimak Pass ³	01/2003	48	42	East, Unimak Pass
Adak Island ³	03/2006	48	41	East, Adak Island

Table 2.3. Observed heterozygosity and F_{is} for each putative subpopulation across both eastern and western populations of Pacific cod. Note that spawning at the western sites occurs from December to February, so season is indicated by a year range, but spawning at the eastern sites occurs in the spring, so season is indicated by a single year.

Population	Subpopulation	Spawning Season	N. Samples	F_{is}	H_o
West	Western Korean coast	2007-2008	21	0.031	0.129
		2015-2016	25	0.047	0.125
West	Southern Korean coast	2007-2008	81	0.074	0.118
		2013-2014	22	0.101	0.115
		2014-2015	70	0.020	0.126
East	Washington Coast	2005	41	0.053	0.099
East	Hecate Strait	2004	46	0.041	0.101
East	Prince William Sound	2012	47	0.017	0.109
East	Kodiak Island	2003	43	0.031	0.104
East	Unimak Pass	2003	42	0.010	0.112
East	Adak Island	2006	41	0.008	0.111

Table 2.4. Results from analysis of molecular variance (AMOVA). This AMOVA was run as a two-level test, with subpopulation as the lowest stratum.

Source of Variation	d.f.	Sum of squares	% of Total Variation	Φ_{ST}	Alternate Hypothesis	p-value
Between east and west	1	65284.28	57.7	0.577	greater	1×10^{-4}
Between subpopulations within each population	6	1890.87	0.514	0.012	greater	1×10^{-4}
Within subpopulations	471	93196.18	41.8	0.585	less	1×10^{-4}
Total	478	160371.32	100		--	--

Table 2.5. Frequency and width of genomic regions of elevated divergence in the eastern population, western population, and full data set. The width of each outlier region is measured by the number of consecutive weighted averages that fell above the 95% confidence interval within that region of the genome. All means were recorded with standard deviation.

	Eastern population	Western population	Full Data Set
F_{st}			
Global / Pairwise genome-wide	0.010	0.037	0.285
Mean within outlier regions	0.062 ± 0.027	0.135 ± 0.037	0.403 ± 0.117
Frequency			
Total number of outlier regions	27	31	44
Mean outlier regions per linkage group	1.17 ± 1.03	1.34 ± 1.23	1.87 ± 1.29
Median outlier regions per linkage group	1.0	1.0	2.0
Maximum on one linkage group	4	5	4
Proportion of weighted averages within outlier regions	0.028	0.030	0.035
Width			
Mean size per outlier region	3.48 ± 2.15	3.23 ± 1.82	2.57 ± 1.26
Median size per outlier region	4.0	3.0	2.5
Maximum size	11	9	5
Minimum size	1	1	1

Table 2.6. Results from the Kruskal-Wallis rank sum test. The test was used determined whether the distributions of outlier region frequency and width were significantly different between divergence scenarios represented by the different Pacific cod data sets.

	Kruskal-Wallis X ²	d.f.	p-value
Total Number of Outlier Regions	4.51	2	0.105
Mean Width of Outlier Regions	4.07	2	0.131

Table 2.7. Results from Welch's t-tests. The test was used to determine if differences in **(a)** frequency or **(b)** width of genomic regions of elevated divergence were statistically significant between each of the three divergence scenarios represented by the different Pacific cod data sets.

(a)

First Data Set	Second Data Set	t-statistic	d.f.	p-value
Eastern population	Western population	-0.520	42.7	0.6055
Eastern population	Full data set	-2.02	41.9	0.0496
Western population	Full data set	-1.41	43.9	0.1672

(b)

First Data Set	Second Data Set	t-statistic	d.f.	p-value
Eastern population	Western population	0.484	51.2	0.6304
Eastern population	Full data set	2.00	37.2	0.0528
Western population	Full data set	1.85	73.0	0.0687

Table 2.8. Gene annotations for RAD loci that were present in the genomic region of elevated divergence on linkage group 19. Annotations are included for loci identified in any of the three data sets: eastern population, western population, or full data set. Only loci which aligned within annotated protein-coding Atlantic cod sequences are shown. If the locus was also identified as an outlier locus (e.g. 13693), an additional row describing outlier status was included.

Locus	Outlier Test	Analysis	Gene ID	Similar Protein; Species
13693	Outlier Region	East; West; All Data	GAMO_00075240	MDFIC; <i>Homo sapiens</i>
	Bayescan, PCAdapt	All Data		
	OutFLANK	East; West		
13773	Outlier Region	West; All Data	GAMO_00075183	IFRD1; <i>Sus scrofa</i>
21317	Outlier Region	West; All Data	GAMO_0007531	Yaf2; <i>Mus musculus</i>
21826	Outlier Region	West	GAMO_00075392	Grm8; <i>Rattus norvegicus</i>
12309	Outlier Region	West	GAMO_00075392	Grm8; <i>Rattus norvegicus</i>
18804	Outlier Region	West	GAMO_00075392	Grm8; <i>Rattus norvegicus</i>
8067	Outlier Region	East	GAMO_00074967	Pclo; <i>Rattus norvegicus</i>
23432	Outlier Region	East	GAMO_00074967	Pclo; <i>Rattus norvegicus</i>

Table 2.9. Counts of putative outlier loci detected in the eastern population, western population, and full data set, listed by detection method(s).

Outlier Detection Method(s)	Eastern Population	Western Population	Full Data Set
OutFLANK	64	17	0
Bayescan (prior 100)	26	2	642
Bayescan (prior 1000)	17	2	444
PCAdapt	22	1	242
All Methods	11	1	-
OutFLANK + Bayescan only	6	1	-
OutFLANK + PCAdapt only	3	0	-
Bayescan + PCAdapt only	0	0	167

Table 2.10. Gene annotations for loci identified as outliers in the western population. Only loci which aligned within annotated protein-coding Atlantic cod sequences are shown.

Locus	Outlier Test	Linkage Group	Gene ID	Similar Protein; Species
9545	OutFLANK, Bayescan	09	GAMO_00046432	ANKRD6; <i>Homo sapiens</i>
11759	OutFLANK	08	GAMO_00032731	ADCY1; <i>Bos taurus</i>
13333	OutFLANK	17	GAMO_00075820	Hecw2; <i>Mus musculus</i>
14955	OutFLANK	06	GAMO_00040879	Prdm12; <i>Mus musculus</i>
19784	OutFLANK	11	GAMO_00021589	CSMD3; <i>Homo sapiens</i>
21863	OutFLANK	23	GAMO_00062317	TPK1; <i>Bos taurus</i>
22068	OutFLANK	17	GAMO_00076021	mus81; <i>Danio rerio</i>

Table 2.11. Gene annotations for loci identified as outliers by two or more detection methods in the eastern population. Only loci which aligned within annotated protein-coding Atlantic cod sequences are shown.

Locus	Outlier Test	Linkage Group	Gene ID	Similar Protein; Species
4734	OutFLANK, Bayescan, PCAdapt	17	GAMO_00077701	ABCA4; <i>Homo sapiens</i>
4856	OutFLANK, PCAdapt	07	GAMO_00006692	SMTNL2; <i>Bos taurus</i>
7241	OutFLANK, PCAdapt	01	GAMO_00023174	Chst11; <i>Rattus norvegicus</i>
7509	OutFLANK, PCAdapt	03	GAMO_00012135	Prkca; <i>Rattus norvegicus</i>
8586	OutFLANK, Bayescan	21	GAMO_00067819	EXTL3; <i>Homo sapiens</i>
10913	OutFLANK, Bayescan	07	GAMO_00006664	HTR3A; <i>Homo sapiens</i>
11741	OutFLANK, Bayescan, PCAdapt	23	GAMO_00062862	esyt-2; <i>Xenopus laevis</i>
11990	OutFLANK, Bayescan	15	GAMO_00036879	Lrrc18; <i>Mus musculus</i>
12233	OutFLANK, Bayescan	15	GAMO_00034113	Cdh5; <i>Mus musculus</i>
13286	OutFLANK, Bayescan, PCAdapt	06	GAMO_00040799	fnbp11; <i>Xenopus laevis</i>
13411	OutFLANK, Bayescan, PCAdapt	14	GAMO_00016282	CHD7; <i>Homo sapiens</i>
14514	OutFLANK, Bayescan, PCAdapt	04	GAMO_00001498	ABL1; <i>Homo sapiens</i>
16650	OutFLANK, PCAdapt	13	GAMO_00038300	ITCH; <i>Homo sapiens</i>
20705	OutFLANK, Bayescan, PCAdapt	06	GAMO_00042946	ARID1A; <i>Homo sapiens</i>
22637	OutFLANK, PCAdapt	11	GAMO_00022322	HECTD4; <i>Homo sapiens</i>

2.8 REFERENCES

- Alderdice, D. F., & Forrester, C. R. (1971). Effects of salinity, temperature, and dissolved oxygen on early development of the Pacific cod (*Gadus macrocephalus*). *Journal of Fisheries Research Board of Canada*, 28(6), 883-902. doi:10.1139/f71-130
- Alexander, V., & Niebauer, H. J. (2003). Oceanography of the eastern Bering Sea ice-edge zone in spring. *Limnology and Oceanography*, 26(6), 1111-1125. doi:10.4319/lo.1981.26.6.1111
- Aruga, J. (2004). The role of *Zic* genes in neural development. *Molecular and Cellular Neuroscience*, 26(2), 205-221. doi:10.1016/j.mcn.2004.01.004

- Barney, B. T., Munkholm, C., Walt, D. R., & Palumbi, S. R. (2017). Highly localized divergence within supergenes in Atlantic cod (*Gadus morhua*) within the Gulf of Maine. *BMC Genomics*, 18(1), 271-285. doi:10.1186/s12864-017-3660-3
- Barth, J. M. I., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., . . . André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*, 26(17), 4452-4466. doi:10.1111/mec.14207
- Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., . . . Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity*, 119, 418-428. doi:10.1038/hdy.2017.54
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., . . . Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, 32(9), 665-680. doi:10.1016/j.tree.2017.06.010
- Bigg, G. R. (2014). Environmental confirmation of multiple ice age refugia for Pacific cod, *Gadus macrocephalus*. *Evolutionary Ecology*, 28(1), 177-191. doi:10.1007/s10682-013-9662-y
- Bigg, G. R., Cunningham, C. W., Ottersen, G., Pogson, G. H., Wadley, M. R., & Williamson, P. (2008). Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. *Proceedings of the Royal Society B: Biological Sciences*, 275(1631), 163-172. doi:10.1098/rspb.2007.1153
- Bradbury, I.R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., . . . Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701), 3725-3734. doi:10.1098/rspb.2010.0985
- Brieuc, M. S. O., Ono, K., Drinan, D. P., & Naish, K. A. (2015). Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, 24(11), 2729-2746. doi:10.1111/mec.13211
- Canino, M. F., & Bentzen, P. (2004). Evidence for positive selection at the pantophysin (Pan I) locus in walleye pollock, *Theragra chalcogramma*. *Molecular Biology and Evolution*, 21(7), 1391-1400. doi:10.1093/molbev/msh141
- 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. doi:10.1111/j.1365-294X.2010.04815.x
- Catchen, J. C., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. (2011). Stacks: building and genotyping loci de novo for short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171-182. doi:10.1534/g3.111.000240
- Catchen, J. C., Hohenlohe, P., Bassham, S., Amores, A., & Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-2140. doi:10.1111/mec.12354
- Chang, K. I., Teague, W. J., Lyu, S. J., Perkins, H. T., Lee, D. K., Watts, D. R., . . . Kim, K. (2004). Circulation and currents in the southwestern East/Japan Sea: Overview and review. *Progress in Oceanography*, 61(2-4), 105-156. doi:10.1016/j.pocean.2004.06.005
- Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The ade4 package - I: One-table methods. *R news*, 4, 5.

- Chung, S., Kim, S., & Kang, S. (2013). Ecological relationship between environmental factors and Pacific cod (*Gadus macrocephalus*) catch in the southern East/Japan Sea. *Animal Cells and Systems*, 17(5), 374-382. doi:10.1080/19768354.2013.853693
- 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. doi:10.1111/mec.12796
- 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. doi:10.1139/F08-199
- Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 20. doi:10.18637/jss.v022.i04
- 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*, 1-17. doi:10.1111/eva.12639
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., . . . Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491, 756-760. doi:10.1038/nature11584
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479/491.
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7), 342-350. doi:10.1016/j.tig.2012.03.009
- Feulner, P. G. D., Chain, F. J. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., . . . Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLOS Genetics*, 11(2), e1004966. doi:10.1371/journal.pgen.1004966
- Foll, M. (2012). BayeScan v2.1 User Manual. Retrieved from: www.cmpg.unibe.ch/software/BayeScan
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), 977-993. doi:10.1534/genetics.108.092221
- Gagnaire, P.A., Pavey, S.A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution* 67(9), 2483-2489. doi:10.1111/evo.12075
- Gustafson, R. G., Lenarz, W. H., McCain, B. B., Schmitt, C. C., Grant, W. S., Builder, T. L., & Methot, R. D. (2000). *Status Review of Pacific Hake, Pacific cod, and Walleye Pollock from Puget Sound, Washington* (NMFS-NWFSC-44). Retrieved from: <https://www.nwfsc.noaa.gov/publications/scipubs/techmemos/tm44/tm44.htm>
- Hemmer-Hansen, J., Therkildsen, N. O., & Pujolar, J. M. (2014). Population genomics of marine fishes: next-generation prospects and challenges. *The Biological Bulletin*, 227(2), 117-132. doi:10.1086/BBLv227n2p117
- Hemmer-Hansen, J., Nielsen Einar, E., Therkildsen Nina, O., Taylor Martin, I., Ogden, R., Geffen Audrey, J., . . . Carvalho Gary, R. (2013). A genomic island linked to ecotype

- divergence in Atlantic cod. *Molecular Ecology*, 22(10), 2653-2667.
doi:10.1111/mec.12284
- Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., . . . Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), 379-397.
doi:10.1086/688018
- Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, 39(1), 21-42.
doi:10.1146/annurev.ecolsys.39.110707.173532
- Hohenlohe, P., 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. doi:10.1371/journal.pgen.1000862
- Houtmeyers, R., Souopgui, J., Tejpar, S., & Arkell, R. (2013). The *Zic* gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cellular and Molecular Life Sciences*, 70(20), 3791-3811. doi:10.1007/s00018-013-1285-5
- Hunt, G. L., & Stabeno, P. J. (2002). Climate change and the control of energy flow in the southeastern Bering Sea. *Progress in Oceanography*, 55(1), 5-22. doi:10.1016/S0079-6611(02)00067-8
- Hwang, C. K., Chaurasia, S. S., Jackson, C. R., Chan, G. C. K., Storm, D. R., & Iuvone, P. M. (2013). Circadian rhythm of contrast sensitivity is regulated by a dopamine-neuronal PAS-domain protein 2-Adenylyl Cyclase 1 signaling pathway in retinal ganglion cells. *The Journal of Neuroscience*, 33(38), 14989-14997. doi:10.1523/jneurosci.2039-13.2013
- Hwang, J. H., Van, S. P., Choi, B. J., Chang, Y. S., & Kim, Y. H. (2014). The physical processes in the Yellow Sea. *Ocean & Coastal Management*, 102, 449-457.
doi:10.1016/j.ocecoaman.2014.03.026
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405. doi:10.1093/bioinformatics/btn129
- Kalinowski, S. T., Wagner, A. P., & Taper, M. L. (2006). ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, 6, 576-579. doi:10.1111/j.1471-8286.2006.01256.x
- Kamvar, Z. N., Brooks, J. C., & Grunwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6:e208. doi:10.3389/fgene.2015.00208
- Kamvar, Z. N., Tabima, J. F., & Grunwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2:e281. doi:10.7717/peerj.281
- Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters* 7, 1225-1241. doi:10.1111/j.1461-0248.2004.00684.x
- Kelley, J. L., Madeoy, J., Calhoun, J. C., Swanson, W., & Akey, J. M. (2006). Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Research*, 16(8), 980-989. doi:10.1101/gr.5157306
- Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., . . . Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between

- migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, 25(10), 2130-2143. doi:10.1111/mec.13592
- Lamichhaney, S., Fuentes-Pardo, A. P., Rafati, N., Ryman, N., McCracken, G. R., Bourne, C., . . . Andersson, L. (2017). Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proceedings of the National Academy of Sciences* 114(17), e3452-e3461. doi:10.1073/pnas.1617728114
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359. doi:10.1038/nmeth.1923
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), R25. doi:10.1186/gb-2009-10-3-r25
- Lankbury, J. A., Duffy-Anderson, J. T., Mier, K. L., Busby, M. S., & Stabeno, P. J. (2007). Distribution and transport patterns of northern rock sole, *Lepidopsetta polyxystra*, larvae in the southeastern Bering Sea. *Progress in Oceanography*, 72(1), 39-62. doi:10.1016/j.pocean.2006.09.001
- Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, 7(3), 355-369. doi:10.1111/eva.12128
- 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. doi:10.1093/plankt/fbn057
- Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of Fst outlier tests. *Molecular Ecology*, 23(9), 2178-2192. doi:10.1111/mec.12725
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). *Pcadapt*: an R package to perform genome scans. *Molecular Ecology Resources*, 17, 67-76. doi:10.1111/1755-0998.12592
- Masada, N., Ciruela, A., MacDougall, D. A., & Cooper, D. M. F. (2008). Distinct mechanisms of regulation by Ca²⁺/Calmodulin of Type 1 and 8 adenylyl cyclases support their different physiological roles. *The Journal of Biological Chemistry*, 283, 4451-4463. doi:10.1074/jbc.M807359200
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28-41. doi:10.1111/1755-0998.12291
- Maugeri, A., Klevering, B. J., Rohrschneider, K., Blankenagel, A., Brunner, H. G., Deutman, A. F., . . . Cremers, F. P. M. (2000). Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *The American Journal of Human Genetics*, 67(4), 960-966. doi:10.1086/303079
- McPherson, A. A., O'Reilly, P. T., & Taggart, C. T. (2011). Genetic Differentiation, Temporal Stability, and the Absence of Isolation by Distance among Atlantic Herring Populations. *Transactions of the American Fisheries Society*, 133(2), 434-446. doi:10.1577/02-106
- Mecklenburg, C. W., Mecklenburg, T. A., & Thorsteinson, L. K. (2002). *Fishes of Alaska*. Bethesda, Maryland: American Fisheries Society.

- Michel, A. P., Sim, S., Powell, T. H. Q., Taylor, M. S., Nosil, P., & Feder, J. L. (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, 107(21), 9724-9729. doi:10.1073/pnas.1000939107
- Mizugishi, K., Aruga, J., Nakata, K., & Mikoshiba, K. (2001). Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI protein. *Journal of Biological Chemistry*, 276, 2180-2188. doi:10.1074/jbc.M004430200
- Ormseth, O. A., & Norcross, B. L. (2009). Causes and consequences of life-history variation in North American stocks of Pacific cod. *ICES Journal of Marine Science*, 66(2), 349-357. doi:10.1093/icesjms/fsn156
- O'Leary, D. B., Coughlan, J., Dillane, E., McCarthy, T. V., & Cross, T. F. (2007). Microsatellite variation in cod *Gadus morhua* throughout its geographic range. *Journal of Fish Biology*, 70(sc), 310-335. doi:10.1111/j.1095-8649.2007.01451.x
- Parkinson, J., & Blaxter, M. (2009). Expressed Sequence Tags: an overview. In J. Parkinson (Ed.), *Expressed Sequence Tags (ESTs): Generation and Analysis* (pp. 1-12). Totowa, NJ: Humana Press.
- Pogson, G. H., & Mesa, K. A. (2004). Positive darwinian selection at the pantophysin (Pan I) locus in marine gadid fishes. *Molecular Biology and Evolution*, 21(1), 65-75. doi:10.1093/molbev/msg237
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842. doi:10.1093/bioinformatics/btq033
- Reppert, S. M., Gegear, R. J., & Merlin, C. (2010). Navigational mechanisms of migrating monarch butterflies. *Trends in Neurosciences*, 33(9), 399-406. doi:10.1016/j.tins.2010.04.004
- Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103-106. doi:10.1111/j.1471-8286.2007.01931.x
- Rousset, F. (2014). Genepop 4.3 for Windows/Linux/Mac OS x. Retrieved from: www.genepop.curtin.edu.au
- Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, 465, 609-612. doi:10.1038/nature09060
- Sinclair-Waters, M., Bradbury, I. R., Morris, C. J., Lien, S., Kent, M. P., & Bentzen, P. (2018). Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic Cod in the northwest Atlantic. *Molecular Ecology*, 27(2), 339-351. doi:10.1111/mec.14442
- Sodeland, M., Jorde, P. E., Lien, S., Jentoft, S., Berg, P. R., Grove, H., . . . Knutsen, H. (2016). "Islands of Divergence" in the Atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution*, 8(4), 1012-1022. doi:10.1093/gbe/evw057
- Stabeno, P. J., Bond, N., Hermann, A. J., Kachel, N., Mordy, C. W., & Overland, J. (2004). *Meteorology and oceanography of the Northern Gulf of Alaska* (Vol. 24).
- Stabeno, P. J., Schumacher, J. D., & Ohtani, K. (1999). The physical oceanography of the Bering Sea. In T. R. Loughlin & K. Ohtani (Eds.), *Dynamics of the Bering Sea: a summary of physical, chemical, and biological characteristics, and a synopsis of research on the Bering Sea* (pp. 1-28). Fairbanks, Alaska: University of Alaska Sea Grant.

- Stroganov, A. N., & Orlov, A. M. (2012). Special characteristics of the formation of population structure in Pacific cod. In O.P. Jenkins (Ed.) *Advances in Zoology Research* (Vol. 2, pp. 169-185). New York, New York: Nova Science Publishers, Inc.
- Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, 25(10), 2144-2164. doi:10.1111/mec.13606
- Tørresen, O. K., Star, B., Jentoft, S., Reinart, W. B., Grove, H., Miller, J. R., . . . Nederbragt, A. J. (2017). An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics*, 18(1):95. doi:10.1186/s12864-016-3448-x
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370.
- Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of Fst. *The American Naturalist*, 186(S1), S24-S36. doi:10.1086/682949
- Wolf, J. B. W., & Ellegren, H. (2016). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18, 87-100. doi:10.1038/nrg.2016.133
- Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, 15(1), 323-354. doi:10.1111/j.1469-1809.1949.tb02451.x
- Yeaman, S., Aeschbacher, S., & Bürger, R. (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, 25(11), 2542-2558. doi:10.1111/mec.13611
- Yeates-Burghart, Q. S., O'Brien, C., Cresko, W. A., Holzapfel, C. M., & Bradshaw, W. E. (2009). Latitudinal variation in photoperiodic response of the three-spined stickleback *Gasterosteus aculeatus* in western North America. *Journal of Fish Biology*, 75(8), 2075-2081. doi:10.1111/j.1095-8649.2009.02418.x
- Zar, J. H. (2010). *Biostatistical Analysis* (5 ed.). Upper Saddle River, New Jersey: Prentice Hall.
- Zhang, C. I. (1984). Pacific cod of South Korean waters. *Bulletin of the International North Pacific Fisheries Community*, 42, 116-129.
- Zhdanova, I. V., & Reeb, S. G. (2005). Circadian Rhythms in Fish. In K. Sloman, S. Balshine, & R. Wilson (Eds.), *Fish Physiology: Behavior and Physiology of Fish* (Vol. 24, pp. 31): Academic Press.