Evidence for divergent selection and spatial differentiation in a putative zona pellucida gene is indicative of local adaptation in Pacific cod

Ingrid Spies[1*], Daniel P. Drinan[2], Eleni L. Petrou[2], Rory Spurr[2], Carolyn Tarpey[2], Theodore Hartinger[2], and Lorenz Hauser[2]

Author Affiliations:

[1]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA

[2]University of Washington, School of Aquatic and Fishery Sciences, 1122 NE Boat St, Box 355020, Seattle, USA 98195, USA

Divergent selection of ZP3 in Pacific cod

*Corresponding author.

Ingrid.Spies@noaa.gov

Abstract

Genetic differentiation has been observed in marine species even when no obvious barriers to gene flow exist. The study of highly differentiated outlier loci can provide information on how genetic variation might contribute to local adaptation. A locus which aligned to a predicted zona pellucida sperm-binding protein 3 gene (ZP3) in Atlantic cod was previously identified in Pacific cod populations as a high differentiation outlier. In other marine species, ZP3 is involved in reproductive isolation, local adaptation, and has neofunctionalized as an antifreeze gene, but the function of this genomic region is not understood in Pacific cod. We sequenced a 544 bp region of ZP3 in 230 Pacific cod collected from throughout their geographic range. Here we show that ZP3 haplotypes exhibit strong spatial structure and there is evidence for divergent selection at this locus in samples collected from the Bering Sea region. The potential for adaptation to different thermal regimes is particularly relevant given that Pacific cod have demonstrated high natural mortality during recent ocean warming events.

Introduction

Genetic differentiation across the genome is the result of complex interactions between genetic drift, selection, migration and recombination (Nosil et al., 2009). Natural selection may override the homogenizing effects of gene flow, and is one of the fundamental forces driving divergence, not only at loci under selection but also at linked loci, often leading to ‘genomic islands of divergence’, i.e. sections of the genome where differentiation is much higher than elsewhere (Nosil, Funk, & Ortiz-Barrientos, 2009; Via

2012). Genomic islands of divergence caused by hitchhiking selection and chromosomal inversions may be particularly common in marine species, as large population sizes lead to little genetic drift and few physical barriers result in relatively high connectivity between populations (Hauser & Carvalho 2008). Chromosomal inversions reduce the rate of recombination and have recently been found to be a major factor facilitating divergence despite ongoing gene flow (Wellenreuther & Bernatchez 2018). In Atlantic cod, for example, islands of genomic divergence were observed on several chromosomes that contained large blocks of linkage disequilibrium and likely represent inversions (Sodeland et al 2016; Berg et al. 2017). Outlier loci with particularly high F_{ST} values not only provide the opportunity to evaluate selection, but may also allow population assignment in high gene flow marine species, which can then be used to trace fish products (Nielsen et al. 2012), estimate population contributions to mixed fisheries (Bekkevold et al. 2015; Dahle, Johansen, Westgaard, & Aglen, 2018) identify migrating individuals (Fisher et al, in review) and examine selective forces that maintain population diversity (Petrou et al. 2021).

Pacific cod provide an interesting case study to investigate the effects of selective differentiation in a species characterized by low levels of genome-wide genetic drift. Studies of neutrally evolving loci have identified genetically differentiated populations of Pacific cod ($F_{ST} < 0.02$) throughout the North Pacific that display a general isolation-by-distance pattern indicating limited dispersal (e.g. Cunningham, Canino, Spies, & Hauser, 2009; Spies et al., 2012; Drinan et al., 2018; Smirnova et al., 2018). Populations at the southern edges of both the eastern and western Pacific show deep genetic divergence from northern populations. This pattern likely originated from secondary contact between

cod lineages occupying distinct glacial refugia but is maintained by differential selection between heterogenous environments (Canino, Spies, Cunningham, Hauser, & Grant, 2010; Drinan et al., 2018; Fisher et al., in review).

On annual scales, Pacific cod are believed to return to their natal spawning areas during winter months, punctuated by feeding migrations that are not well understood (Neidetcher, Hurst, Ciannelli, & Logerwell, 2014; Rand, Munro, Neidetcher, & Nichol, 2014). While many of the genetic differences between Pacific cod populations can be explained by isolation mediated by strong homing behavior, the effects of selection have not been specifically explored. Furthermore, recent ocean warming events have resulted in steep declines in Gulf of Alaska populations (Barbeaux, Holsman, & Zador, 2020a), and anomalous northward summer feeding migrations into the northern Bering Sea (Spies et al., 2020). These observations suggest that populations of Pacific cod are locally adapted to specific temperature regimes.

A previous population genetic study of Pacific cod using restriction site-associated DNA sequencing (RADseq) identified a locus with high differentiation that aligned to the gene encoding for a predicted zona pellucida sperm-binding protein 3 (ZP3) in Atlantic cod (Table S1, Drinan et al., 2018). Although we do not yet understand the function of the putative zona pellucida gene, hereafter referred to as ZP3, egg coat proteins include the family of zona pellucida genes, which are a central to reproductive isolation (Shu et al., 2015). In general, genes coding for proteins on gamete surfaces appear to evolve quickly and be subject to positive selection (Palumbi, 2009). Zona pellucida genes code for glycoproteins that are part of the chorion, enveloping and protecting vertebrate oocytes (Conner, Lefievre, Hughes, & Barratt, 2005; Sano et al.,

2017). Zona pellucida egg coat proteins function in a range of capacities such as protecting the embryo and mediating fertilization. Furthermore, copy number variants of zona pellucida genes have neofunctionalized to produce antifreeze proteins in the eggs of Antarctic notothenioid fishes (Gupta et al., 2012; Cao et al., 2016).

Egg coat proteins have been shown to drive intraspecific diversification and provide potential for populations to adapt to environmental change over relatively short time scales (Shu et al., 2015). Zona pellucida comprises a multi-gene family that prevents polyspermy in mammals, but may have more complicated functions in fish (Conner et al., 2005; Wassarman, 2008). Experiments in Japanese medaka (*Oryzias latipes*) show that ZP3 acts to guide sperm to micropyles, narrow canals that permit entry for fertilization (Iwamatsu, Yoshizaki, & Shibata, 1997). Following fertilization, the chorion hardens and glycoproteins are inactivated, resulting in no further sperm entry (Iwamatsu et al., 1997). ZP3 may also be functionally important for species-specific sperm-egg interactions (Swanson, Yang, Wolfner, & Aquadro, 2001; Conner et al., 2005); matching sperm and egg proteins have been identified in abalone (Swanson & Vacquier, 1997; Aagaard, Vacquier, MacCoss, & Swanson, 2010) and sea urchin (Kamei & Glabe, 2003). Reproductive incompatibility due to a sperm defect in penetration of the zona pellucida of eggs has also been observed among mice strains (Oka et al., 2007). Phenotypic changes in the zona pellucida have been shown to arise as a result of sequence variation in genes coding for ZP2 and ZP3 in humans; therefore, differences observed in zona pellucida coding sequences may have functional significance (Pökkylä, Lakkakorpi, Nuojua-Huttunen, & Tapanainen, 2011). ZP3 has been found to evolve faster than proteins in the inner portion of the zona pellucida matrix (Jagadeeshan & Singh, 2007),

and divergent selection may act to increase variation in that region (Swanson & Vaquier, 2002).

In this study, we used DNA sequence data to assess geographic variation in the ZP3 coding gene throughout the range of Pacific cod. Our study had two specific aims: (i) to assess spatial patterns of differentiation to identify isolated populations, and (ii) to confirm elevated differentiation among Pacific cod populations and test for signals of selection at this gene. We hypothesized that patterns of diversification in this putative ZP3 gene would provide clues to its functional significance.

Materials and Methods

Fin clips from Pacific cod were collected from 16 spawning locations across the species range and preserved in 95% ethanol (Table 1). Samples were taken during the winter spawning season (December through May) to take advantage of spawning site fidelity and avoid sampling of population mixtures. DNA was extracted using Qiagen Blood and Tissue kits following manufacturers' instructions (Qiagen, Inc., Valencia, CA).

Four sets of DNA primers were designed to screen for variation in exon regions aligning with the ZP3 coding region in Pacific cod (Table S2). Five individuals of Pacific cod from the Hecate Strait, Prince William Sound, Washington (WA) Coast, and Kodiak Island collections were sequenced for initial screening (Table 1, Table S2). Following this initial screening, a single set of primers was designed to focus on variable regions in the Pacific cod ZP3 coding region (primer ZP_GM, Table S2). All primers were designed using Primer3 v. 0.4.0 (Koressaar & Remm, 2007).

134 Polymerase chain reactions (PCRs) for the first four sets of primers were carried
135 out in 40µl volume with 10µM forward and reverse primers using the Qiagen Taq PCR
136 Master Mix kit and approximately 200ng DNA (Table S2). Thermal cycling conditions
137 were 95°C for 15 minutes, followed by 35 cycles of 94°C for 30 seconds, T_M for 90
138 seconds, and 72°C for 60 seconds, and a final elongation at 72°C for 30 seconds. PCRs
139 for the variable region (primer ZP_GM, Table S2) were carried out in a 25 µl volume,
140 with Phusion 5X buffer (New England Biolabs, Ipswich, MA), 10mM dNTPs, 10µM
141 forward and reverse primers, 0.2µL Phusion *Taq* polymerase and approximately 200ng
142 DNA. Thermal cycling conditions were 98°C for 30 sec, followed by 5 cycles of 98°C for
143 10 sec, 63-59°C touchdown for 30 seconds (-1°C each cycle for 5 cycles) and 72°C for
144 30 seconds, and then by 30 cycles of 98°C for 10 seconds, 58°C for 30 seconds and 72°C
145 for 30 seconds, and concluded at 72°C for 5 minutes.
146 Sanger sequencing was performed bidirectionally using forward and reverse
147 primers at MCLAB (320 Harbor Way, South San Francisco, CA). Contigs were aligned
148 using Sequencher v. 5.0 (Gene Codes Corporation, Ann Arbor, MI) and scores below
149 80% quality were discarded. Sequence calls were double checked and confirmed by two
150 readers when ambiguities were present. Consensus sequences were aligned in BioEdit v.
151 7.2 (Hall, 1999).
152 A total of 230 Pacific cod were sequenced to screen the variable region of ZP3
153 using primers ZP_GM (Table 1). Haplotypes for ambiguous (heterozygous) nucleotides
154 at ZP3 segregating sites were inferred using Bayesian methodology with a priori
155 expectations based on coalescent theory in PHASE v2.1.1 and SeqPHASE (Stephens,
156 Smith, & Donnelly, 2001; Flot, 2010). This software transformed unphased sequences to

157 fasta files with two haplotypes for all individuals. The most likely pair of haplotypes was
158 selected for each individual based on the highest posterior probability.
159 In order to ensure our sequenced region was indeed ZP3 and to correctly identify
160 synonymous and non-synonymous substitutions, sequences were aligned to the Atlantic
161 cod genome in GenBank using BLASTn (Sayers et al., 2019). The predicted zona
162 pellucida sperm-binding protein 3 (ZP3) spans a 3,407 bp DNA sequence of the
163 GadMor3 annotated genome. It is located on linkage group 9 (LG09) and spans
164 nucleotides 2,454,601-2,458,007 (GenBank Accession ID: GCF_902167405.1,
165 NC_044056.1). The zona pellucida sperm-binding protein 3-like of Atlantic cod based on
166 the GadMor3 (GCF 902167405.1) assembly on chromosome 9 was used to identify
167 intron and exon regions [Gene ID: 115551355, updated Nov. 22, 2020]. DNA sequence
168 from the variable region obtained during the second round of sequencing were aligned to
169 the reverse complemented GadMor3.0 assembly on chromosome 9 using BioEdit to
170 identify reading frame, introns, and exons.
171 Haplotype networks were generated in the R package *pegas* (Paradis, 2010), using
172 R version 4.03 (R Core Team, 2019) based on haplotypes observed in each collection.
173 Pairwise distances were computed among the 14 unique DNA haplotypes and the
174 complementary 544 bp sequence in Atlantic cod, based on GadMor3, to understand
175 which haplotypes appeared most similar to Atlantic cod, using the R package *ape* v.5.4.1
176 (Paradis & Schliep, 2019) and *phanghorn* (Schliep, Potts, Morrison, & Grimm, 2017).
177 The model of nuclear evolution that best fit the data was determined based on the lowest
178 Akaike information criterion score (AIC, Akaike, 1978).
179 We quantified the d_N/d_S ratios for all segregating sites in exons, and calculated

180 expected heterozygosity (H_e) per population using the estimator θ_H, which corrects for
181 potential overestimation when few loci are present, in Arlequin v. 3.5 (Excoffier &
182 Lischer, 2010). Observed heterozygosity (H_o) was calculated as the proportion of
183 heterozygous individuals (individuals with different ZP3 haplotypes), out of the total
184 number sequenced. Haplotype (h) and nucleotide diversity (π) were calculated in in the R
185 package *pegas*. We also performed tests for selective neutrality in Arlequin v. 3.5 with
186 16,000 simulated samples: the Ewens-Watterson homozygosity test and an exact test
187 based on Ewens' sampling theory. The Ewens-Watterson homozygosity test (Watterson,
188 1986) is based on Ewens sampling theory (Ewens, 1972), while the exact test based on
189 Ewens' sampling theory (Slatkin, 1996) compares the probability of the observed sample
190 with that of a random neutral sample with the same number of alleles and identical size.
191 For the exact test and the Ewens-Watterson homozygosity test, small p-values indicate
192 balancing selection ($p \leq 0.05$ considered significant and $p \leq 0.10$ marginally significant),
193 while large values indicate directional or divergent selection ($p \geq 0.95$ considered
194 significant and $p \geq 0.9$ marginally significant). Tajima's D test was calculated in *pegas,*
195 and p-values were calculated based on the assumption that D follows a beta distribution
196 scaled from 0 to 1. This test was used to identify whether sequences fit the neutral theory
197 model at equilibrium between mutation and genetic drift. A negative value of Tajima's D
198 indicates population size expansion, such as after a bottleneck or a selective sweep and/or
199 purifying selection, while a positive value indicates balancing selection or decrease in
200 population size. F_{IS} was calculated by haplotype as F_{IS}=1-H_o/H_e (Nei, 1986). Fisher's
201 exact test was used to test for significant deviations from Hardy-Weinberg equilibrium
202 for each collection using the R package *stats* (fisher.exact) and 10^7 Markov Chain

replicates.

Global and pairwise G_{ST} (Nei, 1973) were calculated using the R package *mmod*, and confidence intervals were generated with 1,000 bootstrap replicates. Dendrograms of hierarchical clusters of sample collections were generated based on pairwise G_{ST} using the *pvclust* package, and support for each cluster was based on 10,000 bootstrap replicates (Suzuki & Simodaira, 2006). Dendrogram probabilities were presented as bootstrap probabilities (BP, green) and approximately unbiased values (AU, red), which is considered more accurate than boostrap values (Suzuki & Simodaira, 2006).

Analysis of molecular variance (AMOVA) was performed on SNP data downloaded from https://datadryad.org/stash/dataset/doi:10.5061/dryad.402sb71 previously published in Drinan et al. (2018) using the R package *poppr* (Kamvar, Tabima, & Grünwald, 2014). The SNP found within ZP3 was removed, leaving 6,424 SNPs and data from 8 collection locations of interest: Adak, Hecate St., Strait of Juan de Fuca, Kodiak Is., Prince William Sound, Salish Sea, Unimak Pass, and WA Coast. AMOVA was performed with collection locations grouped into two regions, changing only the position of the Prince William Sound (PWS) collection. The first grouping was (Region 1=Adak, Kodiak, Unimak, PWS; Region 2 = Hecate St., Juan de Fuca, Salish Sea, WA Coast); and the second grouping was (Region 1=Adak, Kodiak, Unimak, Region 2 = PWS, Hecate Juan de Fuca, Salish Sea, WA Coast). The results were compared to assess whether neutral loci indicated a distinct pattern from ZP3.

We used RAD sequence data from Drinan et al. (2018) to visualize per-locus F_{ST} and estimate linkage disequilibrium along LG09. First, RAD loci reported in Drinan et al. (2018) were aligned GadMor3 using software bowtie2 v. 2.2.6 (--sensitive alignment)

(Langmead and Salzberg 2012). Loci were retained in the data set if they had mapping quality ≥ 20 and contained SNPs (i.e., indels were removed). Subsequently, linkage disequilibrium along LG09 was estimated using the R package *genetics* v. 1.3.8 (Warnes et al. 2019), and per-locus F_{ST} was measured in the R package *genepop* (Rousset, 2008).

Results

Our initial screening efforts revealed no differences in DNA sequence within Exons 1, 5, 6, or 8, which were sequenced using primers ZP_L1, ZP_L3, or ZP_L4. However, variation was observed in Exons 2, 3, and 4 in sequences generated using primers ZP_L2 (GenBank Accession numbers MW468336-MW468402). New primers (ZP_GM) were designed to sequence this region because primer ZP_L2 did not amplify in all collections (Table 1, Table S2). Data for the 230 Pacific cod sequenced using primers ZP_GM are available at Genbank (Accession numbers MW468106-MW468335). The number of individuals sequenced per collection ranged from 1 to 26 (Table 1). Samples from Puget Sound collected in 2003 and 2012 were combined for all analyses due small sample sizes.

Sequences were trimmed to 544 bp, which included 40 amino acid residues from Exon 2, 35 amino acid residues from Exon 3, and 4 amino acid residues from Exon 4. There were nine variable sites; three were located in exons and six in introns (Table 2). Variable sites at location 313 and 339 bp were located in Exon 3 and the variable site at location 542 was in Exon 4 (Table 2). No synonymous substitutions were observed within variable amino acid site so the d_N/d_S ratio could not be calculated (it was infinity).

All Bayesian probabilities for inference of heterozygote phasing were ≥ 0.9, with the exception of one individual. There were three possible haplotypes for a sample from Prince William Sound (Genbank Accession MW468263), with probabilities of 0.272, 0.256, and 0.469. This was a unique case with two heterozygous sites and probabilities were constrained to sum to 1. While 0.469 appears low, it was twice as likely as the other haplotypes; therefore the haplotype with the highest probability was selected.

There were 14 unique DNA haplotypes present among all sample collections (Table 3, Table S3). The most frequent, Haplotype 1, was found in over 85% of individuals collected from the Bering Sea, Aleutian Islands, Shumagin Islands, and Kodiak Island (Table 3, Table S3, Figure 1). Haplotype 1 was also found in 54% of the samples from Japan and a single individual from the Strait of Georgia, but in no other collections. Haplotype 13 was the second most commonly observed, found in 99 haploid sequences, and was only found in samples from Korea, Japan, Prince William Sound, Hecate Strait, Washington Coast, Strait of Georgia, and Puget Sound (Figure 1). Haplotype 2 was unique to the Kiska collection, and was found in 11% of those sequences (Figure 1, Figure 2, Figure 3a).

Variation in coding regions resulted in eight segregating polypeptide sequences (Table 4), with the most common (Variant I) present exclusively in Bering Sea, Kodiak Island, Japan, Aleutian Islands, and a single individual from the Strait of Georgia (Table S4, Figure 2). The second most common polypeptide (Variant II) was the only variant in samples from Korea, and was also observed in samples from the Strait of Georgia, Hecate Strait, Japan, Puget Sound, Prince William Sound, and Washington Coast, as well as a

single individual from Kiska (Figure 3b, Table S4). Genotypes for each individual are listed in Table S5.

The model of nuclear evolution with the lowest AIC score was the generalized time reversible model, with invariant sites and gamma distributed rate variation among sites (GTR+G+I). We selected the Jukes Cantor model (Jukes and Cantor 1969) with equal base and mutation rates to calculate the distance matrix (Table 3). Haplotype 13 was most similar to Atlantic cod, followed by haplotypes 5, 9, 11, and 12, while Haplotype 6 was the most different, followed by Haplotypes 5 and 1. Haplotype 13 was identical to Atlantic cod at the nine segregating sites but did differ at other sites within the 544 bp sequence (Table 3).

Global G_{ST} was 0.551, with 95% confidence intervals 0.514-0.588. The 120 pairwise G_{ST} comparisons summarized in the dendrogram showed 100% support for clustering of Bering Sea, Aleutian Islands, and Kodiak collections (Table S6, Figure 4). Collections from Korea clustered with 100% support, and eastern North Pacific populations also clustered with 100% support. The collections from Korea, Japan, eastern North Pacific and the Gulf of Alaska clustered but with lower support, and the placement of Japan within the Korean/eastern North Pacific cluster also appeared uncertain.

There were regional differences in heterozygosity and haplotypic diversity. The region with the lowest haplotypic diversity was Korea, with only a single haplotype identified in 21 sequenced individuals from three collections. Samples from the Shumagin Islands, Kodiak Island, and Aleutian Islands and Bering Sea also had relatively few haplotypes ($N_H = 6$). These collections were also all in HWE, with the exception of the Kiska collection. The highest number of haplotypes (10) were observed in Prince

William Sound and the eastern North Pacific collections. All collections except the Washington Coast sample deviated significantly from HWE (Table 5). In these collections that deviated from HWE, there was no pattern of deviations due to either excess or deficit of heterozygotes (Table 5). The collection from Japan also deviated from HWE and contained high haplotype diversity (N_H = 5).

Patterns of selection varied by region. The combined Bering Sea collection showed strong indication of divergent selection (>0.95). Within the Bering Sea collections, Kodiak had the highest signal of divergent selection, followed by the Shumagin collection, while Near and Kiska also had high but non-significant values. All of these collections presented negative values of Tajima's D, and Tajima's D over all Bering collections was marginally significant (p = 0.061). Conversely, Tajima's D for the Sea of Okhotsk sample was significantly positive.

AMOVA results based on 6,424 SNP loci (excluding ZP3) indicated stronger support for the Prince William Sound as part of the Bering Sea (grouping with Adak, Kodiak, and Unimak), rather than part of the Eastern Pacific, grouping with samples further south from Hecate Strait, Strait of Juan de Fuca, the Salish Sea, and Washington Coast (Table S7).

In total, we retained 177 SNPs from the RAD data on LG09 after aligning to GadMor3 as described. The only SNP with F_{ST} > 0.01 was found in the short read sequence identified in Drinan et al. that aligned to the predicted ZP3 gene (2018; Table S1). The F_{ST} associated with that SNP was 0.71. Linkage disequilibrium decayed quickly along LG09, and only 7 SNPs that were <0.025 Mb apart had R^2 values > 0.20.

Discussion

Multiple studies have documented neutral genetic diversity and limited gene flow in Pacific cod, but mechanisms for local adaptation are unknown. Our goal was to assess spatial variation in the putative ZP3 gene in Pacific cod to understand whether there was evidence for selection across a large geographic range. Significant recent declines have occurred in this species due to anomalously warm conditions in the Gulf of Alaska, and results were intended to assist in conservation and management of these stocks. We hypothesized that the ZP3 gene may show evidence for selection and patterns of population structure that differ from those of neutrally evolving markers and provide insight into its functional diversity.

Our first significant finding, based on DNA sequence data of the ZP3 variable region, was that sequence variation clustered into two broad geographic groups: samples within and adjacent to the Bering Sea, and collections further south (Fig. 3). Expressed sequences found in the Bering Sea group (collections from the Aleutian Islands, Bering Sea, Shumagins, and Kodiak Island collections) in the north were replaced by an almost completely different and diverse set of expressed sequences from Prince William Sound and southward. The inclusion of the Kodiak Island and Shumagin Islands collections in the Bering Sea group is notable from a management perspective, as the United States federal fishery for Pacific cod currently includes Kodiak Island and the Shumagin Islands region in the Gulf of Alaska management area (Figure 4; NPFMC 2020). Moving southward from the Bering Sea on the western side of the Pacific Ocean, Japan appeared to be a transition region, sharing expressed sequences with the Bering Sea, and it could

339 be considered part of the Bering Sea group. Samples from Korea exhibited only a single
340 expressed sequence type, and could be considered a third sub-group (Polypeptide II)
341 (Figure 3). This work provides new insight into adaptation in Pacific cod, as the
342 geographical pattern observed in ZP3 differed from patterns observed with SNPs (Table
343 S7).

344 Our second main finding was high differentiation in ZP3 (global G_{ST} = 0.551) and
345 evidence for strong divergent selection in the Bering Sea collection. The signal of
346 divergent selection was strongest in collections at the southern end of this region that may
347 be more likely to be subjected to gene flow from the Eastern Pacific region (Table 5).
348 Negative and marginally significant Tajima's D is also consistent with a population
349 expansion in the Bering Sea region following a selective sweep. Selection was not
350 apparent in other regions, although Tajima's D was positive and significant in the sample
351 from the Sea of Okhotsk. While a positive Tajima's D can be an indicator of balancing
352 selection, it can also represent a signal of recent population decline. Overfishing has been
353 reported in Pacific cod off Japan (Seafoodwatch.org) and Korean cod appear to have
354 suffered recent declines (Fisher et al., 2021).

355 The geographic pattern of expressed haplotypes coupled with high levels of
356 differentiation indicates that ZP3 may be driving local adaptation. While our experiment
357 was not designed to examine functional significance directly, geographic patterns and
358 known gene function in other species allow us to speculate on environmental factors
359 leading to observed patterns. ZP3 appears to have undergone neofunctionalization in fish
360 species as antifreeze glycoproteins for cold-water adaptation (Conner & Hughes, 2003;
361 Cao et al., 2016). Experiments have shown that zona pellucida proteins can non-

362 colligatively lower the freezing and melting point of a solution (Cao et al., 2016). Zona
363 pellucida paralogues as antifreeze glycoproteins are expressed in the exocrine pancreas in
364 adult fishes, circulated throughout the body, and may aid in reducing body ice load (Cao
365 et al., 2016). The collections in the Bering Sea group were situated within or adjacent to
366 the southeastern Bering Sea, an ecosystem structured by cold pool water <2°C that
367 remains following retreat of the Bering Sea ice sheet (Stevenson & Lauth, 2019).
368 Residence within colder water may be consistent with selective adaptation to colder
369 temperatures via ZP3 as antifreeze glycoproteins, but further testing is needed to
370 substantiate this hypothesis.
371 Other studies have shown a relationship between temperature and adaptation in
372 cod from the Korean peninsula (Fisher et al., in review). Evidence that cod from Kodiak
373 Island have been shown to hatch within a narrow thermal window (3-6°C) has been
374 attributed to high mortality in the Gulf of Alaska during recent marine heat waves (Laurel
375 & Rogers, 2020). Given the differences in ZP3 throughout the Gulf of Alaska, further
376 investigation of the thermal window tolerated for hatching among cod from the eastern
377 Gulf of Alaska may be of interest for management and conservation under climate
378 change. Population dynamics trajectories may also indicate a difference in adaptation to
379 temperature within the Gulf of Alaska. During and after the warm blob in the Gulf of
380 Alaska (2014-2016), the population size of the Eastern Gulf of Alaska cod stock
381 remained low but stable. However, the cod stock in the Central Gulf of Alaska dropped
382 by about 30% per year, and by 23% per year in the Western Gulf of Alaska. Since
383 2017, the central region has come back faster with a 35% increase per year since 2018
384 while the Western GOA has continued to drop about 17% per year in the same time

385 frame. These population trajectories are consistent with different thermal tolerances
386 related to ZP3 haplotypes, but this remains to be tested (Barbeaux et al., 2020a; Barbeaux
387 et al., 2020b).

388 Our data suggest that as cod recolonized the north Pacific and Bering Sea, a
389 selective sweep occurred in the putative ZP3 gene that was advantageous for northern
390 climates. Bering Sea collections presented negative values of Tajima's D, suggestive of a
391 population expansion following a selective sweep. Results support previous studies
392 indicating that Korea and the eastern North Pacific were once located within glacial
393 refugia for Pacific cod, and that more recent colonization occurred into the Bering Sea
394 and Aleutian Islands and Western Gulf of Alaska (Canino et al., 2010). This is evidenced
395 by the prevalence of Haplotype 13 in those regions, which is the most similar to Atlantic
396 cod (Table 3). In the Western Pacific Ocean, low diversity is consistent with cyclical
397 patterns of abundance or historical population bottlenecks (Nei, Maruyama, &
398 Chakraborty, 1975).

399 While one premise of this paper was to examine the putative ZP3 gene for
400 evidence of selective diversification, the actual gene function and whether it represents a
401 reproductive gene cannot be tested without laboratory experimentation. DNA regions that
402 code for a reproductive protein have been found to act as a barrier to reproduction in
403 other species, but we do not know whether observed differences in ZP3 act as a barrier
404 without controlled laboratory experiments. These studies would help provide insight into
405 the implications of diversification of ZP3 in Pacific cod. Also, we do not understand fine-
406 scale patterns of ZP3 within the Gulf of Alaska and the western Pacific, which may have
407 important management implications. The implications of this research would benefit

408 from more research on RNA transcriptomics or gene expression of ZP3 in the ovary and
409 other organs in Pacific cod to illuminate whether or how ZP3 provides protection from
410 cold temperatures. In addition, the unique expressed variant (Polypeptide VIII) found in
411 the Kiska Island region merits further exploration, which may be related to the unique
412 environment of the Eastern Aleutian Islands (Ladd, Hunt, Mordy, Salo, & Stabeno,
413 2005).

414 In Atlantic cod, genetic divergence among ecotypes has been shown occur within
415 large genomic inversions on linkage groups 1, 2, 7, and 12, each of which span several
416 Mb (Berg et al. 2017; Kirubakaran et al. 2017). The bi-allelic gene pantophysin, which
417 appears to be under such divergent selection in Atlantic cod and has been used to
418 discriminate among ecotypes, has been identified as residing within a large inversion on
419 linkage group 1 (Dahle et al., 2018; Otterå et al., 2020). While RADseq data is not
420 considered reliable for identifying small genomic inversions, we would expect to see
421 additional F_{ST} outliers and more linkage disequilibrium extending over greater genomic
422 distances if a large genomic inversion existed on linkage group 9 (Huang, Andrew,
423 Owens, Ostevik, & Rieseberg, 2020). Only one SNP out of 177 SNPs on linkage group 9
424 had an F_{ST} over 0.01, and linkage disequilibrium was minimal. ZP3 was identified as the
425 largest F_{ST} outlier, and further work is needed to resolve whether it resides within a
426 genomic island of divergence or a small inversion. Nevertheless, our results provide
427 insight into a distinction between adaptive differentiation in Pacific cod and Atlantic cod.
428 The specific mechanism driving selection and diversification in Pacific certainly merits
429 further research.

Overall this work provides evidence for the maintenance of different expressed haplotypes of the putative ZP3 gene through selection in Pacific cod. We found evidence for strong directional selection in the Bering Sea (Bering Sea, Aleutian Islands, Shumagin Islands, and Kodiak Island), and large regional differences among ZP3 haplotype frequencies between the Bering Sea group and regions further south. Results are also indicative of a selective sweep and selection currently acting on northern collections, as well as local adaptation driven by differences in ZP3. This work supports the idea that Korea and the eastern North Pacific were once glacial refugia for Pacific cod, based on the low diversity in Korean samples and the presence of variants that appear most similar to Atlantic cod. While further work is needed to understand the functional advantage of selected types of ZP3 in Pacific cod, we hypothesize that it may act as an antifreeze glycoprotein and provide protection from cold conditions at the egg or adult stages, possibly benefitting populations that occupy the cold pool of the Bering Sea.

References

Aagaard, J.E., Vacquier, V.D., MacCoss, M.J., & Swanson, W.J. (2010). ZP domain proteins in the abalone egg coat include a paralog of VERL under positive selection that binds lysin and 18-kDa sperm proteins. *Molecular Biology and Evolution*, *27*(1), 193–203.

Akaike, H. (1978). On newer statistical approaches to parameter estimation and structure determination. *IFAC Proceedings Volumes*, *11*(1), 1877–1884.

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

Barbeaux, S.J., Ferriss, B., Palsson, W., Shotwell, K,. Spies, I., … Zador, S. (2020b). Assessment of the Pacific cod stock in the Gulf of Alaska. Stock assessment and fishery evaluation report for the North Pacific Fishery Management Council. https://www.fisheries.noaa.gov/resource/data/2020-assessment-pacific-cod-stock-gulf-alaska.

Bekkevold, D., Helyar, S.J., Limborg, M.T., Nielsen, E.E., Hemmer-Hansen, J., … Carvalho, G.R., 2015. Gene-associated markers can assign origin in a weakly structured fish, Atlantic herring. *ICES Journal of Marine Science*, 72(6) 1790–1801.

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

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

Cao, L., Huang, Q., Wu, Z., Cao, D.D., Ma, Z., Xu, Q., … Zhou, C.Z., (2016). Neofunctionalization of zona pellucida proteins enhances freeze-prevention in the eggs of Antarctic notothenioids. *Nature communications*, 7(1), 1–11.

Conner, S.J. & Hughes, D.C., 2003. Analysis of fish ZP1/ZPB homologous genes-evidence for both genome duplication and species-specific amplification models of evolution. *REPRODUCTION-CAMBRIDGE-*, *126*(3), pp.347–352.

Conner, S.J., Lefievre, L., Hughes, D.C., & Barratt, C.L.R. (2005). Cracking the egg: increased complexity in the zona pellucida. *Human Reproduction*, *20*(5), 1148–1152.

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.

Dahle, G., Johansen, T., Westgaard, J-I., Aglen, A., Glover, K.A. (2018). Genetic management of mixed-stock fisheries "real-time": The case of the largest remaining cod fishery operating in the Atlantic in 2007–2017. *Fisheries Research* 205, 77–85.

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

Excoffier, L. & Lischer, H.E. (2010). Arlequin suite version 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, 10(3), 564-567.

Ewens, W.J., (1972). The sampling theory of selectively neutral alleles. *Theoretical population biology*, 3(1), 87–112.

Fisher, M., Helser, T., Kang, S., Gwak, W. S., Canino, M., & Hauser, L. (2021 Submitted). Individual assignment in a high gene flow marine fish: the power of peripheral populations. Molecular Ecology.

Flot, F.J. (2010). SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments, *Molecular Ecology Resources*, 10 (1), 162–166.

Gupta, S.K., Bhandari, B., Shrestha, A., Biswal, B.K., Palaniappan, C., Malhotra, S.S., & Gupta, N. (2012). Mammalian zona pellucida glycoproteins: structure and function during fertilization. *Cell and tissue research*, 349(3), 665–678.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.

Hauser, L. and Carvalho, G.R., 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4), 333–362.

Huang, K., Andrew, R.L., Owens, G.L., Ostevik, K.L. & Rieseberg, L.H., 2020. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular ecology*, 29(14), 2535–2549.

Iwamatsu, T., Yoshizaki, N., & Shibata, Y. (1997). Changes in the chorion and sperm entry into the micropyle during fertilization in the teleostean fish, *Oryzias latipes*. *Dev. Growth Differ.* 39, 33–41.

Jagadeeshan, S., & Singh, R. S. (2007). Rapid Evolution of outer egg membrane proteins in the *Drosophila melanogaster* subgroup: a case of ecologically driven evolution of female reproductive traits. *Molecular Biology and Evolution*, 24(4), 929–938.

Jukes, T.H., & Cantor, C.R. (1969). Evolution of protein molecules. *Mammalian Protein Metabolism*, *3*, 21–132.

Kamei, N., & Glabe, C.G. (2003). The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes & Development*, *17*(20), 2502–2507.

Kamvar, Z.N., Tabima, J.F., & Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2: e281.

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

Koressaar, T., & Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10), 1289–1291.

Ladd, C., Hunt Jr, G.L., Mordy, C.W., Salo, S.A., & Stabeno, P.J. (2005). Marine environment of the eastern and central Aleutian Islands. *Fisheries oceanography*, *14*, 22–38.

Langmead B., & Salzberg S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9, 357–359.

Laurel, B.J., & Rogers, L.A. (2020). Loss of spawning habitat and prerecruits of Pacific cod during a Gulf of Alaska heatwave. *Canadian Journal of Fisheries and Aquatic Sciences*, 77(4), 644–650.

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

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 29(1), 1–10.

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

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

Nielsen, E.E., Cariani, A., Mac Aoidh, E., Maes, G.E., Milano, I., …Bekkevold, D., 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature communications*, 3(1), 1–7.

583 Nosil, P., Funk, D.J., & Ortiz-Barrientos, D. (2009). Divergent selection and
584 heterogeneous genomic divergence. *Molecular ecology*, *18*(3), 375–402.
585
586 NPFMC 2020. Fishery Management Plan for Groundfish of the Gulf of Alaska
587 Management area. North Pacific Fishery Management Council (NPFMC). 1007 West
588 Third, Ste. 400, Anchorage, Alaska 99501.
589 https://www.npfmc.org/wp-content/PDFdocuments/fmp/GOA/GOAfmp.pdf
590
591 Oka, A., Aoto, T., Totsuka, Y., Takahashi, R., Ueda, M., Mita, A., … Moriwaki, K.
592 (2007). Disruption of genetic interaction between two autosomal regions and the X
593 chromosome causes reproductive isolation between mouse strains derived from different
594 subspecies. *Genetics*, *175*(1), 185–197.
595
596 Otterå, H.M., Johansen, T., Folkvord, A., Dahle, G., Bingh, M.K.S., … Glover, K.,
597 (2020). The pantophysin gene and its relationship with survival in early life stages of
598 Atlantic cod: Pantophysin in Atlantic cod.
599
600 Palumbi, S.R., (2009). Speciation and the evolution of gamete recognition genes: pattern
601 and process. *Heredity*, 102(1), pp.66–76.
602
603 Paradis, E. (2010). pegas: an R package for population genetics with an integrated–
604 modular approach. *Bioinformatics*, *26*(3), 419–420.
605
606 Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and
607 evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528.
608
609 R Core Team (2019). *R: A language and environment for statistical computing*. Vienna,
610 Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-
611 project.org/.
612
613 Petrou, E.L., Fuentes-Pardo, A.P., Rogers, L.A., Orobko, M., Tarpey, C., … Lowry, D.,
614 2021. Functional genetic diversity in an exploited marine species and its relevance to
615 fisheries management. *Proceedings of the Royal Society B*, 288(1945), 20202398.
616
617 Pökkylä, R.M., Lakkakorpi, J.T., Nuojua-Huttunen, S.H., & Tapanainen, J.S. (2011).
618 Sequence variations in human ZP genes as potential modifiers of zona pellucida
619 architecture. *Fertility and sterility*, 95(8), 2669-2672.
620
621 Rand, K.M., Munro, P., Neidetcher, S.K., & Nichol, D.G. (2014). Observations of
622 seasonal movement from a single tag release group of Pacific cod in the eastern Bering
623 Sea. *Marine and Coastal Fisheries*, 6(1), 287–296.
624
625 Rousset, F. (2008) Genepop'007: a complete reimplementation of the Genepop software
626 for Windows and Linux. *Molecular Ecology Resources,* 8:103–106
627

628 Sano, K., Kawaguchi, M., Katano, K., Tomita, K., Inokuchi, M., Nagasawa, T., … Arai,
629 K., (2017). Comparison of egg envelope thickness in teleosts and its relationship to the
630 sites of ZP protein synthesis. *Journal of Experimental Zoology Part B: Molecular and*
631 *Developmental Evolution*, 328(3), 240–258.
632
633 Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., & Karsch-Mizrachi, I.
634 (2019). GenBank. *Nucleic Acids Research*, 47(D1), D94–D99.
635
636 Schliep, K., Potts, A. J., Morrison, D. A., & Grimm, G. W. (2017). Intertwining
637 phylogenetic trees and networks. *Methods in Ecology and Evolution*, 8, 1212–-1220.
638 doi:10.1111/2041-210X.12760
639
640 Shimada, A., & Kimura, D. (1994). Seasonal movements of Pacific cod *Gadus*
641 *macrocephalus* in the eastern Bering Sea and adjacent waters based on tag-recapture data.
642 *Fishery Bulletin*, 92, 800–816.
643
644 Slatkin, M., (1996). A correction to the exact test based on the Ewens sampling
645 distribution. Genetics Research, 68: 259–260.
646
647 Smirnova, M.A., Orlova, S.Y., Kalchugin, P.V., Bojko, M.I., Park, J.H., & Orlov, A.M.
648 (2018). Population structure of Pacific cod *Gadus macrocephalus* in the southern part of
649 the range based on the microsatellite analyses. *Russian Journal of Genetics*, 54(6), 670–
650 679.
651
652 Sodeland, M., Jorde, P.E., Lien, S., Jentoft, S., Berg, P.R., …, Knutsen, H., 2016.
653 "Islands of Divergence" in the Atlantic cod genome represent polymorphic chromosomal
654 rearrangements. *Genome biology and evolution*, 8(4), 1012–1022.
655
656 Spies, I., Gruenthal, K.M., Drinan, D.P., Hollowed, A.B., Stevenson, D.E., Tarpey, C.M.,
657 & Hauser, L. (2020). Genetic evidence of a northward range expansion in the eastern
658 Bering Sea stock of Pacific cod. *Evolutionary applications*, *13*(2), 362–375.
659
660 Stephens, M., Smith, N.J., & Donnelly, P. (2001). A new statistical method for haplotype
661 reconstruction from population data. *The American Journal of Human Genetics*, 68(4),
662 978–989.
663
664 Stevenson, D.E., & Lauth, R.R. (2019). Bottom trawl surveys in the northern Bering Sea
665 indicate recent shifts in the distribution of marine species. *Polar Biology*, 42(2), 407–421.
666
667 Suzuki, R., & Shimodaira, H. (2006). *PVCLUST*: an R package for assessing the
668 uncertainty in hierarchical clustering. *Bioinformatics*, 22, 1540–1542
669
670 Swanson. W.J., & Vacquier, V.D. (1997). The abalone egg vitelline envelope receptor for
671 sperm lysin is a giant multivalent molecule. *Proceedings of the National Academy of*
672 *Sciences USA*, 94, 6724–6729.
673

674 Swanson, W.J., Yang, Z., Wolfner, M.F., & Aquadro, C.F. (2001). Positive Darwinian
675 selection drives the evolution of several female reproductive proteins in
676 mammals. *Proceedings of the National Academy of Sciences USA*, 98(5), 2509–2514.
677
678 Swanson, W.J., & Vacquier, V.D. (2002). The rapid evolution of reproductive
679 proteins. *Nature Reviews Genetics 3*(2), 137–144.
680
681 Teixeira, J.C., & Huber, C.D., (2020). Dismantling a dogma: the inflated significance of
682 neutral genetic diversity in conservation genetics. *arXiv preprint arXiv:2007.02569*.
683
684 Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during
685 ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B:*
686 *Biological Sciences*, 367(1587), 451–460.
687
688 Warnes, G., Gorjanc, G., Friedrich, L., & Man, M. (2019). Genetics: population genetics.
689 R package version 1.3.8.1.2 edn. See https://CRAN.R-project.org/package=genetics.
690
691 Wassarman, P.M., (2008). Zona pellucida glycoproteins. *Journal of Biological*
692 *Chemistry*, 283(36), 24285–24289.
693
694 Watterson, G.A., 1986. The homozygosity test after a change in population
695 size. *Genetics*, 112(4), 899–907.
696
697 Wellenreuther, M. & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal
698 inversions. *Trends in ecology & evolution*, 33(6), 427–440.
699
700 Winter, D.J., (2012). MMOD: an R library for the calculation of population
701 differentiation statistics. *Molecular ecology resources*, 12(6), 1158–1160.

Table 1. Pacific cod (*Gadus macrocephalus*) samples used in this study, listed in clockwise order around the Pacific Rim, including location name, region, and date sampled. Sample names refer to a nearby landmark. Sampling locations are considered within three broad regions: Western Pacific (including Korea and Japan), Bering Sea (including Aleutian Islands, Bering Sea, and Western Gulf of Alaska), and Eastern Pacific (eastern Gulf of Alaska southward to Washington State).

Location	Region	Date	Latitude	Longitude
Geoje, Korea	Western Pacific	Jan. 2015	34° 51’ N	128° 35’ E
Yellow Sea, Korea	Western Pacific	Feb. 2015	35° 12’N	124° 40’ E
Jukbeon, Korea	Western Pacific	Dec. 2007	37° 04’ N	129° 25’ E
Sea of Okhotsk, Japan	Western Pacific	May 2005	44° 20’ N	145° 52’ E
Near Islands	Bering	Feb. 2005	52° 34‘ N	174° 17’ E
Kiska Island	Bering	Mar. 2005	51° 48’ N	177° 47’ E
Pervenets Canyon	Bering	Mar. 2016	59° 21’ N	177° 13’ W
Adak Island	Bering	Mar. 2006	51° 40’ N	176° 36’ W
Unimak Pass	Bering	Feb. 2018	54° 35’ N	165° 15’ W
Shumagin Is.	Bering	Mar. 2019	55° 15’ N	159° 30’ W
Kodiak Island	Bering	Mar. 2003	57° 48’ N	152° 31’ W
Prince William S.	Eastern Pacific	Mar. 2012	60° 32’ N	147° 04’ W
Hecate Strait	Eastern Pacific	Mar. 2004	53° 13′ N	130° 57′ W
Strait of Georgia	Eastern Pacific	Apr. 2003	48° 54’ N	123° 06’ W
Puget Sound	Eastern Pacific	Mar. 2003	47° 35’ N	122° 30’ W
Puget Sound	Eastern Pacific	May 2012	48° 14’ N	122° 40’ W
Washington Coast	Eastern Pacific	Feb. 2005	47° 55’ N	125° 33’ W

711 Table 2. The position of each segregating site in the 544 bp sequence of Pacific cod
712 (*Gadus macrocephalus*), and its position on the zona pellucida sperm-binding protein 3
713 DNA coding region of linkage group 9 in the GadMor3 alignment (NC_044056.1). Also
714 shown are its codon position (1, 2, or 3), location in an intron or exon, and resulting
715 amino acid change.

Position in sequence	Position on Gadmor3	Reading frame	Intron or Exon	Amino acid change
17	2,455,229	-	Intron	
313	2,455,525	2	Exon 3	AGG (Arg) or ATG (Met)
339	2,455,551	1	Exon 3	GAC (Asp) or AAC (Asn)
447	2,455,649	-	Intron	
448	2,455,650	-	Intron	
449	2,455,651	-	Intron	
451	2,455,653	-	Intron	
452	2,455,654	-	Intron	
542	2,455,754	2	Exon 4	TAC (Tyr) or TTC (Phe)

716
717
718

Table 3. There were 14 inferred haplotypes of Pacific cod (*Gadus macrocephalus*), at nine segregating sites at 17, 313, 339, 447, 448, 449, 451, 452, 542 bp along the 544 bp sequence. *N* refers to the number of each haplotype observed, for 230 diploid individuals (460 sequences), and the pairwise distance calculated between each haplotype and the ZP3 sequence from Atlantic cod. The haplotype designations listed here are used throughout. *The haplotype from A. cod was taken from the GadMor3 alignment.

Designation	Haplotype	*N*	Pairwise Distance with A. cod
1	GTATATGCT	265	0.0436
2	GTATATGCA	5	0.0417
3	GGATATGCT	3	0.0417
4	GGGTATGCT	10	0.0397
5	GGGTATGCA	3	0.0378
6	GGGCGATTT	1	0.0495
7	GGGCGATTA	1	0.0475
8	ATGTATGCT	11	0.0397
9	ATGTATGCA	1	0.0378
10	AGATATGCT	13	0.0397
11	AGATATGCA	5	0.0378
12	AGGTATGCT	6	0.0378
13	AGGTATGCA	99	0.0358
14	AGGCGATTA	37	0.0456
A. cod*	AGGTATGCA	-	-

730 Table 4. Segregating sites located in ZP3 exons of Pacific cod (*Gadus macrocephalus*)
731 resulting in non-synonymous amino acid changes at 313, 339, and 542 bp. Designations
732 for each polypeptide are listed and are consistent throughout.
733

Designation	1	2	3	*N*
I	Met	Asn	Phe	265
II	Arg	Asp	Tyr	140
III	Arg	Asp	Phe	17
IV	Met	Asp	Phe	11
V	Arg	Asn	Phe	16
VI	Met	Asp	Tyr	1
VII	Arg	Asn	Tyr	5
VIII	Met	Asn	Tyr	5

734 Table 5. The number of individuals (*N*), number of distinct haplotypes present (N_H), haplotype diversity (*h*), nucleotide diversity (π),
735 observed (H_o) and expected (H_e) heterozygosity, F_{IS}, and probability that collections conform to Hardy Weinberg equilibrium (P_{HWE})
736 for each collection. Collections are listed by region, and the Bering and E. Pacific collections are summarized (shaded in grey).
737 Ewens-Watterson and Tajima's D results are shown: observed *F* (Obs. *F*), expected *F* (Exp. *F*), Ewens-Watterson *p*-value (E-W *p*),
738 and Slatkin's test *p*-value, the number of segregating sites in the sample (S), Tajima's D, and the associated *p*-value under a beta
739 distribution. Dash (-) indicates not applicable.
740

Location	*N*	N_H	*h*	π	*He*	*Ho*	F_{IS}	P_{HWE}	Obs. F	Exp. F	E-W *p*	Slatkin's *p*	S	Tajima's D (*p*)
Geoje	10	1	0	0	0	0	-	-	-	-	-	-	0	-
Yellow Sea	2	1	0	0	0	0	-	-	-	-	-	-	0	-
Jukbeon	9	1	0	0	0	0	-	-	-	-	-	-	0	-
Sea of Okhotsk	12	5	0.6522	0.0036	0.625	0.167	0.733	0.000	0.375	0.376	0.609	0.631	4	2.232 (0.027)
Near	22	2	0.0888	0.0003	0.087	0.091	-0.046	1.000	0.913	0.768	0.791	0.791	2	-1.130 (0.268)
Kiska	22	3	0.2463	0.0006	0.137	0.045	0.672	0.002	0.759	0.610	0.794	0.814	3	-1.105 (0.280)
Pervenets	21	1	0	0	0	0	-	-	-	-	-	-	0	-
Adak	7	1	0	0	0	0	-	-	-	-	-	-	0	-
Unimak	14	2	0.1984	0.0007	0.191	0.214	-0.120	1.000	0.809	0.755	0.578	0.578	2	-0.477 (0.675)
Shumagin	23	4	0.2773	0.0012	0.271	0.304	-0.122	1.000	0.729	0.519	0.886	0.855	7	-1.543 (0.108)
Kodiak	26	2	0.0385	0.0001	0.038	0.038	0.000	1.000	0.962	0.785	1.000	1.000	2	-1.460 (0.133)
(Bering Combined)	135	6	0.0004	0.1345	0.134	0.104	0.220	0.000	0.866	0.489	0.974	0.969	8	-1.692 (0.061)
PWS	12	6	0.7246	0.0040	0.694	0.250	0.640	0.040	0.306	0.311	0.579	0.326	9	-0.301 (0.803)
Hecate	13	5	0.6831	0.0055	0.657	0.846	-0.288	0.001	0.343	0.387	0.419	0.566	7	1.873 (0.070)
St. of Georgia	6	5	0.7273	0.0028	0.667	0.333	0.501	0.050	0.333	0.304	0.810	0.810	4	0.466 (0.656)
PugSnd	9	5	0.7516	0.0043	0.710	0.111	0.844	0.002	0.290	0.345	0.351	0.292	8	0.011 (0.966)
WA Coast	22	6	0.5973	0.0052	0.584	0.727	-0.245	0.627	0.416	0.362	0.741	0.953	9	1.047 (0.313)
(E. Pacific Combined)	62	10	0.7260	0.0055	0.726	0.525	0.277	0.000	0.280	0.286	0.578	0.781	9	-1.908 (0.072)

741 Acknowledgements

742

743 We thank Sonya Elmejjati, Mary Fisher, Sue Hoffman, Qtae Jo, Elisa Russ, Olav
744 Ormseth, Taka Yanagimoto for providing samples for this work. Jean-François Flot and
745 Robin Waples provided insight into data analysis. We also thank Wes Larson, Jay Orr,
746 Kimberly Rand, and two anonymous reviewers for providing helpful suggestions. This
747 work was funded by the Alaska Fisheries Science Center.

748

749 Data Availability Statement

750 The data that support the findings of this study are openly available in GenBank at
751 https://www.ncbi.nlm.nih.gov/genbank/, accession numbers MW468336-MW468402.

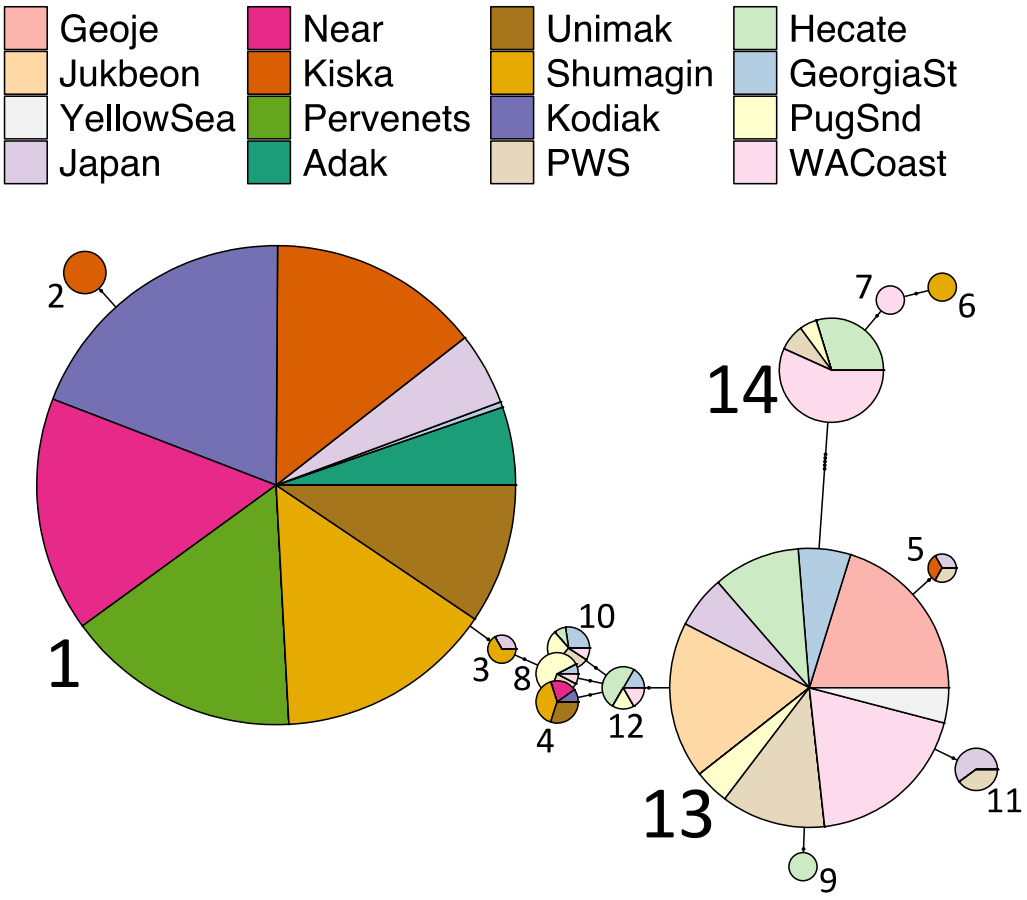
752

753 Author Contributions

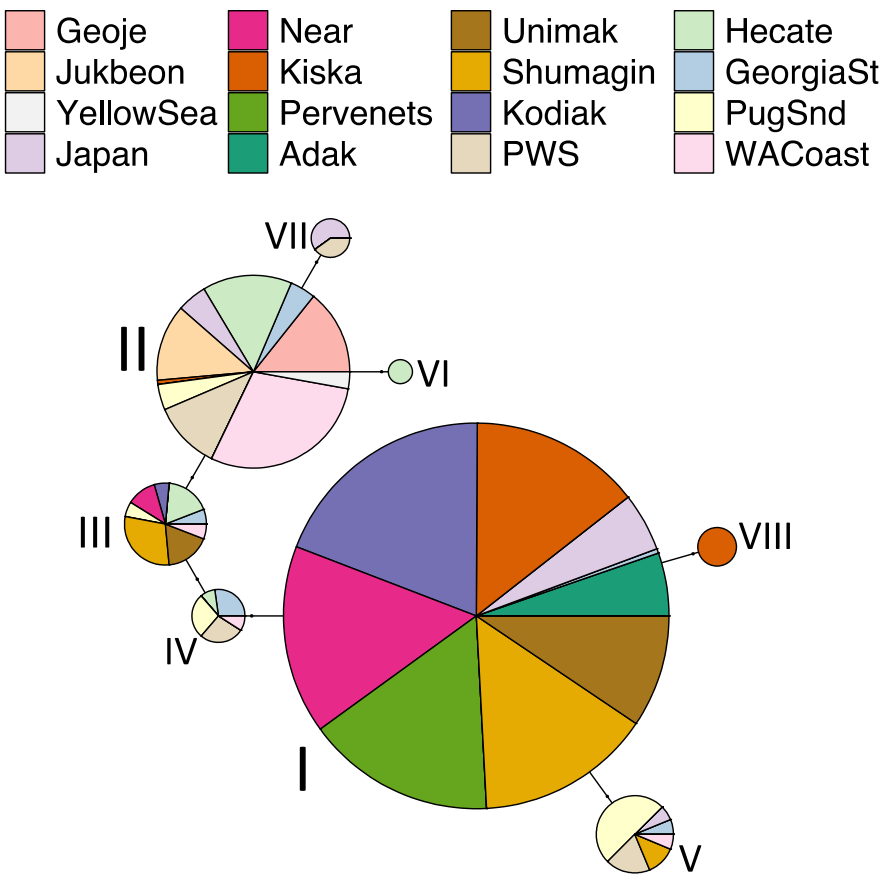
754 D.D. identified the outlier locus in a previous published paper and initiated the research
755 project, R.S. and T.H. performed DNA sequencing and analysis, E.P. contributed
756 analytical tools and writing, C.T. assisted with laboratory analyses, L.H. designed
757 research and contributed to writing the manuscript, and I.S. analyzed the data, prepared
758 figures and tables, and wrote the manuscript.

759

760 Figure captions
761



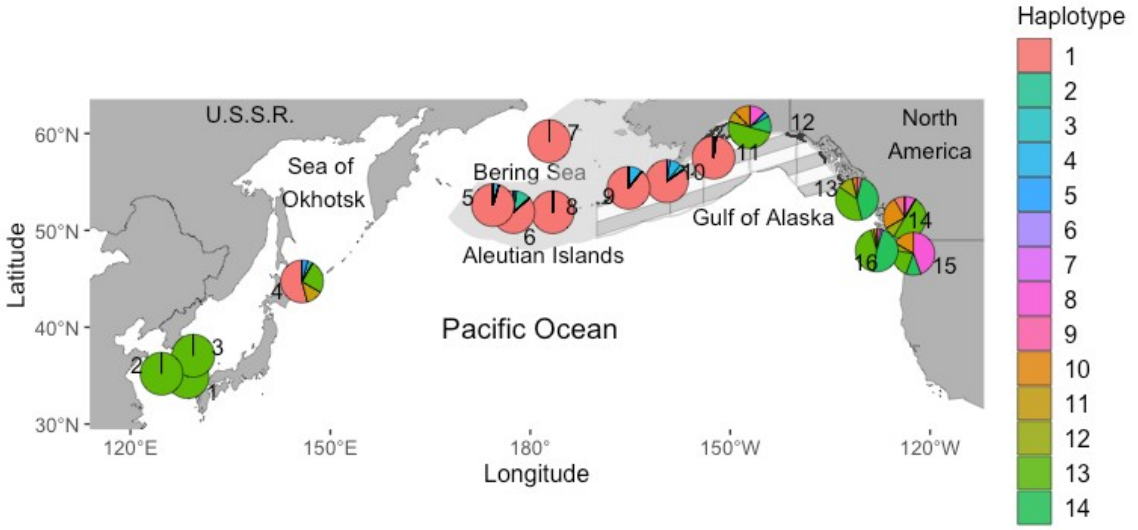
762
763 Figure 1. Haplotype network of phased ZP3 sequences of Pacific cod (*Gadus*
764 *macrocephalus*), circles represent haplotypes and pie slices represent contribution by
765 sample collection. The size of each circle is relative to the haplotype frequency, although
766 the scale ratio was adjusted for clarity, and the true proportions are shown in Table 3.
767 Haplotype designations are listed adjacent to each circle and are consistent with
768 designations in Table 3.
769



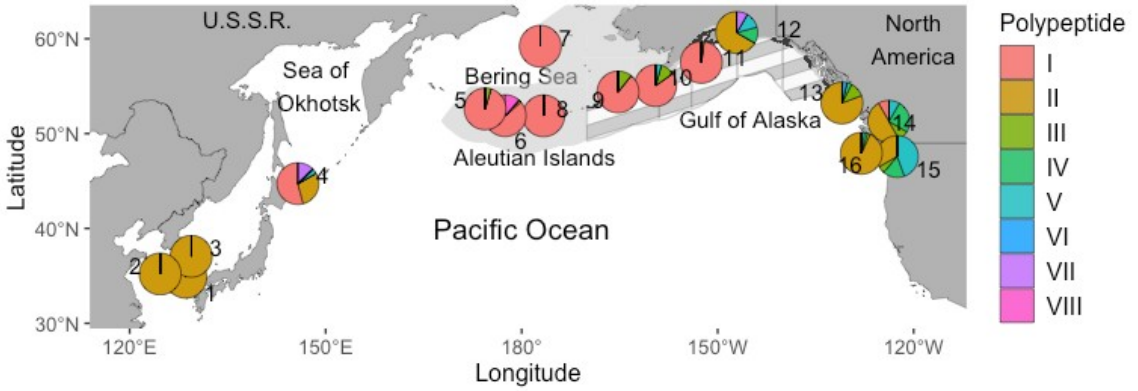
770
771 Figure 2. Network of ZP3 amino acid compositions (polypeptides) of Pacific cod (*Gadus*
772 *macrocephalus*). Circles represent amino acid sequences and pie slices represent
773 contribution by sample collection. There were eight unique protein coding amino acid
774 combinations among three sites, which are specified along with frequency in Table 4.
775 The size of each circle is relative to the number observed, although the scale ratio was
776 adjusted for clarity. Polypeptide designations are listed adjacent to each circle in roman
777 numerals.
778
779

780

a.



b.



781 Figure 3. Map with pie charts showing a. relative haplotype frequencies between all
782 collections, and b. amino acid sequences based on non-synonymous changes in the ZP3
783 DNA coding region. The haplotype numbers correspond to Table 3, and roman numerals
784 indicate polypeptide designation, as shown in Table 4. Numbers on map represent the
785 name of the collection, west to east: 1. Geoje, Korea, 2. Yellow Sea, Korea, 3. Jukbyeon,
786 Korea, 4. Sea of Okhotsk, Japan, 5. Near Islands, 6. Kiska Island, 7. Pervenets Canyon, 8.
787 Adak Island, 9. Unimak Island, 10. Shumagin Islands, 11. Kodiak Island, 12. Prince
788 William Sound, 13. Hecate Strait, 14. Strait of Georgia, 15. Puget Sound, 16. Washington
789 Coast. The spatial area comprising the Gulf of Alaska Fishery Management Plan is
790 shown with grey stripes and the area comprising the Bering Sea and Aleutian Islands
791 Fishery Management Plan is light grey.
792

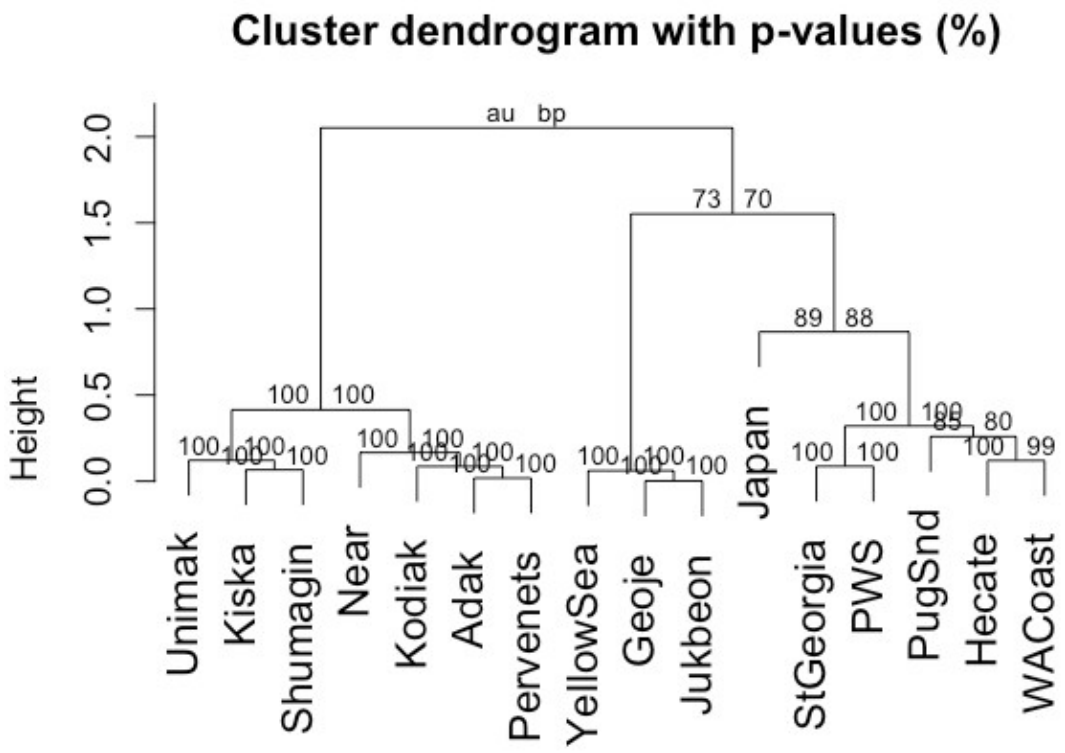


Figure 4. Dendrogram of hierarchical clusters of Pacific cod (*Gadus macrocephalus*) based on pairwise G_{ST} for phased ZP3 haplotypes. Numbers represent probabilities for cluster support based on 10,000 bootstrap replicates, and are presented as AU (left, approximately unbiased) p-values and BP (right, bootstrap probability) values.

Appendix

Table S1. Short read sequence (95bp) associated with the highest F_{ST} based on RAD sequencing (Drinan et al., 2018).

5’ TGCAGGAGGCATAACGTGAGCAGTCTGGCCCTGGACCCGATCTTCA CCCCGTTCTCGGCCATCAAAGTGTCCGAGGAGCTTTTGCACTTCAGTTT 3’

Table S2. Primer sets ZP_L1, ZP_L2, ZP_L3, and ZP_L4 were used to screen for variation across the DNA sequence coding for the zona pellucida sperm-binding protein 3 in samples of Pacific cod (*Gadus macrocephalus*). Primer set GP_GM was used to amplify a variable region. The annealing temperature, T_M (°C), used in polymerase chain reaction is shown for each primer set. The position of the forward and reverse primer on the GadMor3 sequence is shown, as well as which exons are included in the intervening sequence.

Primer name	Forward	Reverse	T_M (°C)	F pos.	R pos.	Exon
ZP_L1	gagcccgctagatcacttgt	tgtaatctggtgggcagcta	56.5	1	592	1
ZP_L2	agctgcccaccagattacat	gattgcgtgcgctatactga	62.5	574	1372	2,3,4
ZP_L3	aatgtttgtttccgcactca	ccgcatacaaacacacacat	57	1416	1766	5,6
ZP_L4	ttgcaggcttcaaaaagtca	catcatagcaagggcatgaa	56	3223	3242	8
ZP_GM	gcaatctgagggtaggacca	aacgcagtgatccacaaaga	58	612	1153	2,3,4

Table S3. Frequency of observed nucleotide haplotides in sample collections of Pacific cod (*Gadus macrocephalus*). The Haplotype number (No.) from 1 to 14 is shown with the number of inferred haplotypes from each sample collection (Georgia = Strait of Georgia, Hecate = Hecate Strait, Perv. = Pervenets, PS = Puget Sound, Shum. = Shumagins, WA = Washington Coast, YS = Yellow Sea, Japan = Sea of Okhotsk).

No.	Geoje	Jukbeon	YS	Japan	Near	Kiska	Perv.	Adak	Unimak	Shum.	Kodiak	PWS	Hecate	Georgia	PS	WA	Tot.
1	0	0	0	13	42	38	42	14	25	39	51	0	0	1	0	0	265
2	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5
3	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	3
4	0	0	0	0	2	0	0	0	3	4	1	0	0	0	0	0	10
5	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	3
6	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
8	0	0	0	0	0	0	0	0	0	0	0	3	0	1	8	1	13
9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
10	0	0	0	0	0	0	0	0	0	0	0	3	1	3	3	1	11
11	0	0	0	3	0	0	0	0	0	0	0	2	0	0	0	0	5
12	0	0	0	0	0	0	0	0	0	0	0	0	3	1	1	1	6
13	20	18	4	6	0	0	0	0	0	0	0	12	10	6	4	19	99
14	0	0	0	0	0	0	0	0	0	0	0	3	11	0	2	21	37

Table S4. Frequency of segregating polypeptide sequences in sample collections of Pacific cod (*Gadus macrocephalus*). The Variant number (No.) from I through VIII is shown with the number of inferred polypeptide sequences from each sample collection (Georgia = Strait of Georgia, Hecate = Hecate Strait, Perv. = Pervenets, PS = Puget Sound, Shum. = Shumagins, WA = Washington Coast, YS = Yellow Sea, Japan = Sea of Okhotsk).

No.	Geoje	YS	Jukbeon	Japan	Near	Kiska	Perv.	Adak	Unimak	Shum.	Kodiak	PWS	Hecate	Georgia	PS	WA	Tot.
I	0	0	0	13	42	38	42	14	25	39	51	0	0	1	0	0	265
II	20	4	18	7	0	1	0	0	0	0	0	16	21	6	6	41	140
III	0	0	0	0	2	0	0	0	3	5	1	0	3	1	1	1	17
IV	0	0	0	0	0	0	0	0	0	0	0	3	1	3	3	1	11
V	0	0	0	1	0	0	0	0	0	2	0	3	0	1	8	1	16
VI	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
VII	0	0	0	3	0	0	0	0	0	0	0	2	0	0	0	0	5
VIII	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5

831 Table S5. Diploid genotypes by collection of Pacific cod (*Gadus macrocephalus*), consisting of Haplotype A/Haplotype B, followed
832 by the number of individuals (in parentheses) in which it was observed.
833

Collection	Haploid genotypes (frequency)
Geoje	13/13 (10)
Yellow Sea	13/13 (2)
Jukbeon	13/13 (9)
Sea of Okhotsk	1/1 (6), 1/5 (1), 3/11 (1), 11/11 (1), 13/13 (3)
Near Islands	1/1 (20), 1/4 (2)
Kiska Island	1/1 (19), 2/2 (2), 2/5 (1)
Pervenets Canyon	1/1 (21)
Hecate Strait	8/12 (1), 9/13 (1) 12/13 (2), 13/14 (7), 14/14 (2)
Adak Island	1/1 (7)
Unimak Pass	1/1 (11), 1/4 (3)
Shumagin Islands	1/1 (16), 1/3 (2), 1/4 (4), 1/6 (1)
Kodiak Island	1/1 (25), 1/4 (1)
Prince William S.	5/8 (1), 8/8 (1), 10/10 (1), 11/11 (1), 10/13 (1), 13/13 (5), 13/14 (1), 14/14 (1)
Puget Sound	8/8 (1), 10/10 (4), 8/12 (1), 13/13 (2), 14/14 (1)
Washington Coast	7/14 (1), 8/14 (1), 10/14 (1), 12/13 (1), 13/13 (3), 13/14 (12), 14/14 (3)

834
835

836 Table S6. Pairwise G_{ST} among collections of Pacific cod (*Gadus macrocephalus*): Georgia = Strait of Georgia, PWS = Prince William
837 Sound, PugSnd = Puget Sound, WACoast = Washington Coast.

	Adak	Geoje	Georgia	Hecate	Japan	Jukbeon	Kiska	Kodiak	Near	PWS	Pervenets	PugSnd	Shumagin	Unimak	WACoast
Geoje	1.0000														
Georgia	0.4511	0.1837													
Hecate	0.4951	0.2939	0.0720												
Japan	0.1753	0.4021	0.1090	0.1604											
Jukbeon	1.0000	0.0000	0.1830	0.2932	0.4015										
Kiska	0.0505	0.7816	0.3314	0.3739	0.0940	0.7813									
Kodiak	-0.0018	0.9623	0.4351	0.4789	0.1643	0.9623	0.0418								
Near	0.0113	0.9154	0.4080	0.4518	0.1435	0.9152	0.0293	0.0002							
PWS	0.4735	0.1691	0.0041	0.0421	0.1231	0.1684	0.3553	0.4580	0.4316						
Pervenets	0.0000	1.0000	0.4560	0.4997	0.1812	1.0000	0.0565	0.0043	0.0174	0.4782					
PugSnd	0.4635	0.3617	0.0609	0.0997	0.1599	0.3611	0.3470	0.4483	0.4222	0.0520	0.4683				
Shumagin	0.0457	0.7573	0.3173	0.3595	0.0841	0.7570	0.0169	0.0332	0.0156	0.3417	0.0518	0.3331			
Unimak	0.0431	0.8219	0.3535	0.3966	0.1074	0.8217	0.0216	0.0257	0.0061	0.3778	0.0491	0.3688	-0.0021		
WACoast	0.5398	0.3131	0.0935	-0.0011	0.1844	0.3124	0.4115	0.5221	0.4936	0.0506	0.5441	0.1178	0.3964	0.4355	
YellowSea	1.0000	0.0000	0.1579	0.2697	0.3798	0.0000	0.7712	0.9604	0.9110	0.1435	1.0000	0.3386	0.7459	0.8132	0.2893

838
839

Table S7. Results of AMOVA among 6,424 SNPs previously published in Drinan et al. (2018), with the SNP for ZP3 removed. AMOVA was performed such that the Prince William Sound (PWS) collection was arranged two ways; "PWS with Bering Sea" (Group 1=Adak, Kodiak, Unimak, PWS; Group 2 = Hecate St., Juan de Fuca, Salish Sea, WA Coast); "PWS with Eastern Pacific" (Group 1=Adak, Kodiak, Unimak, Group 2 = PWS, Hecate Juan de Fuca, Salish Sea, WA Coast). Sigma (σ) represents the variance for each hierarchical level, and the percent of the total, *Phi* Φ represents population differentiation (higher *Phi* represents higher differentiation).

Test	PWS with Bering Sea				PWS with Eastern Pacific			
	d.f.		Sum Sq.	Mean Sq.	d.f.		Sum Sq.	Mean Sq.
Between Regions	1		3809.958	3809.9578	1		2907.828	2907.8277
Between Collections within Regions	6		5996.927	999.4878	6		6899.057	1149.8428
	Phi Φ				Phi Φ			
Phi-Collection-Region	0.0065				0.0097			
Phi-Region-total	0.0148				0.0092			
Test	Sigma σ	%	Std.Obs	*p*-value	Sigma σ	%	Std.Obs	*p*-value
Variations within samples	673.0099		-3.3786	<0.000999	673.0099		-3.3899	<0.001998
Variations between samples	11.1099		1.7388	>0.052947	11.1099		1.8297	>0.035964
Variations between Regions	10.3613	1.48	4.5071	>0.026973	6.4321	0.92	2.6496	>0.043956
Variations between collections within Regions	4.4674	0.64			6.7269	0.96		