Diversity and population structure of Pacific herring (*Clupea pallasi*) along the Northwest coast: an interdisciplinary investigation using high-throughput sequencing and ancient DNA

Eleni Leto Petrou

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Lorenz Hauser, Chair

Tim Essington

Ryan Kelly

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

University of Washington

Abstract

Diversity and population structure of Pacific herring (*Clupea pallasi*) along the Pacific Northwest coast: an interdisciplinary investigation using high-throughput sequencing and ancient DNA

Eleni Leto Petrou

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

Forage fish are key trophic links in coastal ecosystems, transferring energy from lower to higher trophic levels. In the Pacific Northwest, herring are one of the most abundant forage fish species and are prey for a wide variety or organisms. Additionally, human inhabitants of the region have used herring as a food resource for millennia, and archaeological research has demonstrated that herring bones are among the most abundant fish remains unearthed from ancient village sites. Given the importance of herring to the coastal ecosystem and human livelihoods, there is considerable interest in sustainably managing the fisheries which target this species. However, management is complicated by limited knowledge of herring population structure and a lack of long-term temporal data on population diversity. In this dissertation, I (i) develop a laboratory protocol to remove intraspecific contamination from tissues that is compatible with restriction site-

associated sequencing, (ii) investigate the mechanisms driving genetic differentiation between herring populations, using genomic data collected from wild spawning aggregations, and (iii) quantify the relative contributions of genetically distinct herring populations to food supplies over the last millennium, using ancient DNA extracted from herring bones.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FIGURES

LIST OF SUPPLEMENTAL TABLES

ACKNOWLEDGEMENTS

I would like to thank all of the people who taught and supported me while I was on this journey. First of all, I extend thanks to my adviser, Lorenz Hauser, who guided me intellectually throughout these last five years and who gave me the opportunity to conduct research on Pacific herring. I feel very lucky to work on a research system that is absolutely fascinating with such an enthusiastic and creative mentor. I also thank my wonderful and supportive PhD committee: Tim Essington, Phil Levin, and Ryan Kelly; their insights greatly strengthened this work. I acknowledge and thank the other herring enthusiasts who guided this research and inspired me: Robert Kopperl, Madonna Moss, Dana Lepofsky, Angela Fuentes-Pardo, Luke Rogers, Todd Sandell, Anna Kagley, and Dennis Lewarch. I am forever indebted to Antonia Rodrigues, Camilla Speller, and Dongya Yang for teaching me everything I know about ancient DNA and for being the most incredible hosts while I worked in their laboratories.

The research presented in this dissertation was made possible by the generosity and kindness of many people. I am extend my gratitude to Kelly Brown, Mike Reid, Davey Wilson, and Randy Carpenter (Heiltsuk Integrated Resource Management Division), and Russ Jones (Haida Fisheries Program) for guiding me in sampling design, teaching me about fisheries co-management, and helping organize sample collections. Many people generously donated their time and herring samples to this research: Paul Hershberger (United States Geological Survey); Lynn Lee, Leandre Veignault, and the Lavoies family; Kurt Stick, Dayv Lowry, Adam Linquist, and Todd Sandell (Washington Department of Fish and Wildlife); Terry Beacham (Department of Fish and Oceans Canada); Hans Daubenberger (Port Gamble S'Klallam Tribe); Wade Smith, Kira Krumhansl, Dan Okamoto, and Margot Hessing-Lewis (Hakai Research Institute); Sharon Wildes (NOAA Juneau); Chad Ormond (Q'ulhanumutsun Aquatic Resources Society); Kyle Rosendale, Jeff Feldpausch, and Mike Smith (Sitka Tribe of Alaska). I extend special thanks to the Heiltsuk Integrated Resource Management Division and the Hakai Research Institute for hosting and supporting field work- it was a very powerful learning experience to spend two herring seasons on the Central Coast and to witness firsthand the struggle for resource co-management. I also thank the Suquamish Tribe, the Puyallup Tribe, and the Burke Museum for generously sharing ancient herring samples with us.

Past and present members of the MerLab played a huge role in my graduate education. I thank Isadora Jiménez-Hidalgo, Carolyn Tarpey, and Molly Phillips for contributing their amazing laboratory skills to this research and for their unflagging emotional support. Kerry Naish, Marine Brieuc, Dan Drinan, Charles "Danger" Waters, Natalie "Lowe" Lowell, Mary Fisher, Sam May, Ramon "Jamon" Gallego, and Molly Jackson were an absolute delight to work with and taught me so much about genetics and bioinformatics. I couldn't have asked for a better group of colleagues and friends.

I also thank the funding sources that made this work possible: Washington Sea Grant, the Puget Sound Anglers Club, the School of Aquatic and Fishery Sciences, and the UW IGERT program on Ocean Change. I am eternally grateful to Terrie Klinger for letting me join this wonderful community of interdisciplinary researchers. The training I received through the IGERT program greatly expanded my horizons and I am so thankful for all of the research experiences and friendships that this program facilitated.

Everyone needs a helping hand to get them through graduate school. I thank my parents, Ethel and Athos Petrou, for being my scientific inspirations and calmly reminding me that certain things are "just [academic] rituals". Rachel Hovel and Casey Clark were unflagging science cheerleaders and my official qualifying exam support team. Finally, I give my heartfelt thanks to my partner in crime, Christian Esquilo Baldoz, for making every day fun and for doing all of the cooking during these last six months. Breh, I literally would have starved to death in front of my computer without you.

DEDICATION

This dissertation is dedicated to the herring, on whom so much depends.

Chapter 1. **Intraspecific DNA contamination distorts subtle population structure in a marine fish: decontamination of herring samples before restriction site-associated sequencing and its effects on population genetic statistics**

ABSTRACT

Wild specimens are often collected in challenging field conditions, where samples may be contaminated with the DNA of conspecific individuals. This contamination can result in false genotype calls, which are difficult to detect, but may also cause inaccurate estimates of heterozygosity, allele frequencies, and genetic differentiation. Marine broadcast spawners are especially problematic, because population genetic differentiation is low and samples are often collected in bulk and sometimes from active spawning aggregations. Here, we used contaminated and clean Pacific herring (*Clupea pallasi*) samples to test (i) the efficacy of bleach decontamination, (ii) the effect of decontamination on RAD genotypes, and (iii) the consequences of contaminated samples on population genetic analyses. We collected fin tissue samples from actively spawning (and thus contaminated) wild herring and non-spawning (uncontaminated) herring. Samples were soaked for 10 minutes in bleach or left untreated, and extracted DNA was used to prepare DNA libraries using a restriction-site associated DNA (RAD) approach. Our results demonstrate that intraspecific DNA contamination affects patterns of individual and population variability, causes an excess of heterozygotes, and biases estimates of population structure. Bleach decontamination was effective at removing intraspecific DNA contamination and compatible with RAD sequencing, producing high-quality sequences, reproducible genotypes, and low levels of missing data. Although sperm contamination may be specific to broadcast spawners, intraspecific contamination of samples may be common and difficult to detect from high-throughput sequencing data, and can impact downstream analyses.

INTRODUCTION

High-throughput DNA sequencing has advanced the field of molecular ecology by enabling comprehensive investigations of genetics and genomics in non-model species (Allendorf et al. 2010, Ekblom and Galindo 2011, Andrews et al. 2016). However, high-throughput sequencing is sensitive to the contamination of samples with exogenous (non-target) DNA. Errors introduced by interspecific DNA contamination have been identified in whole genome assemblies (Longo et al. 2011, Koutsovoulos et al. 2016), ancient DNA (Campana et al. 2014), and metagenomic datasets (Schmieder and Edwards 2011). To address the problem of interspecific contamination, bioinformatic tools have been developed to remove exogenous DNA from sequence data (Schmieder and Edwards 2011) before contaminated sequences are incorporated into downstream analyses. These methods typically identify non-target sequences by aligning them to databases of common contaminating species; as a result, they cannot be used to detect intraspecific contamination caused by the unintentional mixing of DNA between individual samples of the same species.

Intraspecific contamination may profoundly affect downstream analysis, even though it can be hard to detect in raw data. False heterozygotes inflate measures of observed heterozygosity (Jun et al. 2012) and genetic diversity, and can lead to biased estimates of allele frequencies and genetic differentiation. In species with weak population structure, contamination may either obscure true differentiation or, alternatively, suggest significant genetic differentiation where none exists.

Some bioinformatic tools have been developed to screen sequences for intraspecific DNA contamination (Jun et al. 2012, Flickinger et al. 2015), but these tools were primarily developed for human re-sequencing studies; as such, they require pre-existing baseline data on population

allele frequencies or high-coverage individual genotypes to identify contaminated individuals. These types of genomic resources are oftentimes unavailable for non-model species and consequently little attention has been given to the potential problem of intraspecific DNA contamination in most molecular ecology studies.

Intraspecific contamination can be particularly problematic in studies of wild populations of non-model organisms. First of all, samples are often collected in challenging or remote field conditions, where access to resources such as sterile water and clean tools is limited. In addition, field sampling can involve the bulk collection of multiple individuals. For example, animals such as fish or insects may be caught in nets where numerous individuals are in close contact with each other's tissues or bodily fluids, increasing the risk of intraspecific contamination (Mitchell et al. 2008, Greenstone et al. 2012). More generally, laboratory errors during sample handling or DNA library preparation can also result in intraspecific DNA contamination (Sehn et al. 2015), and the common use of Illumina adapters during high-throughput sequencing (such as restriction site-associated DNA (RAD) sequencing (Baird et al. 2008), means that any exogenous DNA present in a sample could be amplified during PCR.

One of the standard methods to decontaminate samples is treatment with bleach; this approach has been used to clean bone samples before sequencing of ancient DNA (Kemp and Smith 2005, Yang and Watt 2005), as well as fresh tissue samples for microsatellite (Mitchell et al. 2008) and mitochondrial analysis (Greenstone et al. 2012). However, traditional microsatellite and mitochondrial sequencing, as well as high-throughput sequencing of ancient DNA, can utilize short DNA fragments as template. In contrast, RAD sequencing requires very high-quality DNA with intact restriction sites, otherwise there is a dramatic reduction in the number of raw sequences produced (Graham et al. 2015). Given that bleach decontaminates samples by

degrading surface DNA (Kemp and Smith 2005), the effect of bleach on the quality and quantity of endogenous sequence reads produced by RAD sequencing is currently unknown. Therefore, bleach treatment may affect downstream analyses, even if decontamination were successful.

Here, we used contaminated and clean Pacific herring (*Clupea pallasi*) samples to test (i) the efficacy of bleach decontamination, (ii) the effect of decontamination on RAD genotypes, and (iii) the consequences of contaminated samples on population genetic analyses. By combining these results, we identified the impacts of contamination on population genetic analyses and empirically validated an approach aimed at minimizing contamination that is compatible with RAD sequencing.

MATERIALS AND METHODS

Sample collection

Sexually mature Pacific herring were collected immediately prior to or during active spawning events using seine nets or hook and line fishing gear (Table 1-1). Adult herring were sampled from genetically differentiated populations with different spawn timing (Small et al. 2005, Mitchell 2006, Beacham et al. 2008); our study included samples from the "primary-spawning" populations of Quilcene Bay (WA) and Spiller Channel (BC), and the "late-spawning" population from Cherry Point (WA). The sexual maturity of each individual was visually determined following the guidelines described in Bucholtz et al. (2008). During sampling, herring sperm was clearly visible in the water column and fish readily released gametes when slight pressure was applied to their abdomen. The density of sperm in the water column during a herring spawn may be as high as 80-210 sperm/mL (Hourston and Rosenthal 1976), resulting in considerable intraspecific DNA contamination (Mitchell et al. 2008). Thus, our samples were likely contaminated with the DNA of multiple herring. Fin or muscle tissue

samples were taken from each individual and immediately stored in 100% ethanol in individual vials.

Captive juvenile herring that were sexually immature were used as an uncontaminated control group. Juvenile herring were reared at the United States Geological Survey (USGS) Marrowstone Marine Field Station, WA, from fertilized eggs collected at Cherry Point, WA (Table 1-1). Herring were individually caught from aquaria and euthanized using tricaine methanesulfonate (MS-222). Fin tissue from each individual fish was sampled immediately and samples were preserved in 100% ethanol. To minimize the risk of cross-contamination during sampling, a new scalpel was used for each fish, and other sampling equipment (e.g., tweezers, cutting mats) was cleaned with 10% bleach solution followed by three rinses of distilled water and flame sterilization.

Experimental assessment of bleach treatment

Tissue samples taken from wild adults (N = 17) and captive juveniles (N = 20) were split into two pieces (approximately 2 mm^2) and exposed to the following experimental treatments:

1. Null treatment: samples were stored in 100% ethanol until DNA extraction.
2. Bleach treatment: Following a modified protocol of Mitchell et al. (2008), samples were placed in individual tubes and immersed in 180 μL of 0.12% sodium hypochlorite (bleach) (Sigma-Aldrich, St. Louis, MO, USA) for ten minutes. During bleach incubation, samples were vortexed at medium-high speed. Subsequently, we removed bleach from the tubes and added 200 μL of Milli-Q purified water (Millipore, Bedford, MA, USA). Samples were vortexed for one minute at medium-high speed, after which Milli-Q water was removed and fresh Milli-Q water was added to the tube. This water rinse was repeated five times, and samples were stored in 100% ethanol until DNA extraction.

To estimate genotyping error rates within and between treatment groups, five juvenile herring were subsampled in replicate, and both subsamples were subjected to both experimental treatments. In addition, we also created four "dirty cocktails" as reference positive controls for DNA contamination. Each dirty cocktail contained 25 ng/ μL of DNA from four different juvenile herring in equal proportions.

We tested the reproducibility of the bleach treatment by implementing it on a large number of spawning adult herring (N = 194). These fish were sampled from the same geographic location as the herring that were used in the null and bleached treatments (Table 1-1).

DNA library preparation and sequencing

Genomic DNA was extracted from each subsample using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA). DNA was visualized with agarose gel electrophoresis to assess DNA quality and quantified with the PicoGreen dsDNA Assay Kit (Invitrogen, Waltham, MA, USA). We standardized the DNA concentration of each sample to 25 ng/ μL.

As an initial check for contamination, six microsatellite loci (*Cpa-8, Cpa-104, Cpa-113* (Miller et al. 2001) and *Cpa-106, Cpa-107a*, *Cpa-111* (Olsen et al. 2002)) were used by the Washington Department of Fish and Wildlife Molecular Genetics Laboratory to screen every sample that was present in both the bleach and null treatment groups (N = 37), following the protocol of Olsen et al. (2002). Alleles were scored on Peak Scanner 2 (Life Technologies, Carlsbad, CA, USA). In the microsatellite data, we defined contaminated samples as those containing more than two alleles at any locus.

We followed the protocol of Etter at al. (2011) to prepare DNA libraries for restriction-site associated (RAD) sequencing. Depending on availability, 200 to 500 ng (depending on availability) of genomic DNA per individual was digested with the restriction enzyme *SbfI* (New

England Biolabs, Ipswich, MA). Samples were individually labeled using a custom set of 96 barcodes (Integrated DNA Technologies, San Diego, CA) and groups of 12 samples were pooled into libraries that were sheared to a length of approximately 500 base pairs (bp) using a Bioruptor sonicator (Diagenode, Denville, NJ). We modified the Etter et al. (2011) protocol by using AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) to size-select DNA fragments (300-500 bp) and purify DNA products. However, all other steps (blunt-end repair, 3'-dA overhang addition, P2 adapter ligation, PCR) were conducted as described in Etter at al. (2011). After PCR, the DNA concentration of each library was quantified using the PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA). We standardized the concentration of each library to 10 *nM* and pooled libraries such that 48 individuals were sequenced per lane of an Illumina HiSeq 2500 (Illumina Inc., San Diego, CA) at the University of Oregon Genomics Core Facility. The resulting sequences were single-end and 100 bp in length.

Bioinformatics analyses

We used the *process_radtags* script in *Stacks* version 1.39 (Catchen et al. 2013) to demultiplex individual samples, remove sequences with low quality scores (Phred score < 10), and trim sequences to a length of 90 base pairs. The quality of sequencing data was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk /projects/fastqc/).

Following the protocol of Brieuc et al. (2014), we created a reference database of herring RAD loci to facilitate sequence assembly and locus identification. The reference database was built using juvenile samples (null treatment) that had at least 1.5 million sequences (N = 19). First, we assembled sequences and identified loci in these samples using the *de novo* locus discovery pipeline in *Stacks*. Loci within each sample were allowed to have up to three nucleotide mismatches (*ustacks*, M = 3) and each allele had to be sequenced at a minimum depth

of 5X to be retained in the analysis (*ustacks*, m = 5). Subsequently, we removed loci with tandem repeat units using *Blast* version 2.2.25 (Altschul et al. 1990) and *bowtie* version 0.12.7 (Langmead et al. 2009) as described in Brieuc et al. (2014).

All sequenced samples (N = 280) were aligned to the reference database of RAD loci using *bowtie*, allowing up to three nucleotide mismatches between the reference and query sequences. Sequences that aligned to the database were subsequently processed with the *pstacks* script in *Stacks* to identify loci in each sample (minimum depth of coverage to report a stack = 10; SNP model, alpha = 0.05). We filtered out low-quality samples by only retaining those that contained at least 20,000 RAD loci after *pstacks*. To maximize the number of loci retained, a catalog of loci was constructed in *cstacks* using a subset of the ten most deeply sequenced individuals (bleach treatment) from each sampling location. All samples were genotyped using *sstacks* and we only retained loci that were present in 80% of samples from each treatment group.

We removed possible sequencing errors by filtering the SNPs discovered by *Stacks*. A custom python script published in Brieuc et al. (2014) were used to retain only loci with two haplotypes and to re-score genotypes. This method designates a heterozygote genotype if each allele is sequenced at least twice and the locus is sequenced to a depth of at least ten reads. Subsequently, we filtered out loci and individuals that had more than 20% missing data. Loci characterized by very low minor allele frequencies were filtered from the final dataset; a minor allele had to be present in at least one of the treatment groups at a frequency of 0.05 for that locus to be retained in downstream analyses. Finally, we tested for deviations from Hardy-Weinberg equilibrium (HWE) using the exact test based on 1,000 Monte Carlo permutations of alleles, as implemented in the R package *pegas* (Paradis 2010). Loci that were out of HWE in

every one of the population genetic samples (Cherry Point, Quilcene Bay, and Spiller Channel) were removed from the analysis. As a final assessment of locus assembly, we followed the recommendations of Paris et al. (2017) and aligned the filtered set of loci to the Atlantic herring genome using *bowtie2* version 2.2.6 (Langmead and Salzberg 2012). We also estimated per-locus F_{IS} at each sampling location using *Genepop* version 4 (Rousset 2008).

Individual multilocus heterozygosity (H_I), the number of heterozygous loci divided by the total number of loci genotyped, was calculated for each sample. Our expectation was that contaminated samples would be characterized by higher values of H_I than the uncontaminated control group (juvenile herring) because they would contain alleles from multiple individuals. Variation in multilocus heterozygosity among uncontaminated individuals and populations was expected to be small, as Pacific herring are characterized by large population sizes, low inbreeding and low genetic population differentiation (Small et al. 2005, Mitchell 2006, Beacham et al. 2008).

In addition, we tested whether bleach degraded target DNA and introduced error to the data by comparing the genotypes of identical juvenile herring in the null and bleach treatment groups (N = 20). This error was quantified as the number of genotype mismatches observed between replicate extractions from the same individual (N = 5). A Wilcoxon signed-rank test was used to assess whether the mean genotype mismatch rate differed between replicate samples and treatment groups ($\alpha = 0.05$).

Population structure

We investigated the effect of intraspecific DNA contamination on patterns of population structure by analyzing samples in the null and bleached treatment groups in combination with the larger number of bleached samples. First, we conducted a principal components analysis (PCA)

using the R-package *adegenet* (Jombart 2008). We also conducted an analysis with *Structure* version 2.3.4 (Pritchard et al. 2000) using two different subsets of the data: the first set included all samples, while the second included only bleached samples whose H_I was within the range observed in uncontaminated juvenile samples. We implemented the admixture model and allowed allele frequencies to be correlated among populations. Sampling location was used as prior information (LOCPRIOR model), which can help detect clusters when population structure is weak (Hubisz et al. 2009). Three repetitions of the model were run for each value of K (number of clusters) ranging from one to six. All runs consisted of 20,000 burn-in steps followed by 50,000 Markov chain Monte Carlo steps. We subsequently used *structure harvester* (Earl and vonHoldt 2012) to visualize likelihood values for different values of K and calculate the ad-hoc statistic ΔK to identify the highest hierarchical level of clustering in our data set (Evanno et al. 2005).

To further investigate the effects of contamination and bleach treatment on measures of population structure, populations were divided into 39 subsamples of approximately six individuals (range = 4 to 7 individuals), the sample size of the smallest collection of contaminated individuals from a single location. A recent study (Willing et al. 2012) showed that a small number of individuals (N= 4-6) can be used to obtain unbiased estimates of F_{ST} when large numbers of loci (N > 1,000) are genotyped. Pairwise F_{ST} (Weir and Cockerham 1984a) between subsamples was calculated in Genepop version 4 and used for non-metric multidimensional scaling (nMDS) in Primer 6 (Clarke and Gorley 2006). Observed and expected heterozygosity were calculated in GenAlEx version 6.5 (Peakall and Smouse 2012), and F_{IS} (Weir and Cockerham 1984b) was estimated in Genepop version 4 (Rousset 2008). To compare differentiation with and without contaminated individuals, hierarchical AMOVAs were

calculated in Arlequin version 3.52 (Excoffier and Lischer 2010), using two alternative groupings. In the first comparison, groups were defined by population (Cherry Point; Quilcene Bay; Spiller Channel) and subgroups consisted of the two different treatments (bleach, null). In the second comparison, groups were defined by population and subgroups consisted of subsamples of individuals (N= 4-7); different iterations of this AMOVA were conducted excluding untreated individuals and H_I outliers.

RESULTS

Sequencing and genotyping

We successfully genotyped 92% of individuals at three or more microsatellite loci. Six out of 17 adult herring in the null treatment group displayed more than two alleles per microsatellite locus, indicating that they were contaminated with the DNA of multiple herring. Treatment with bleach appeared to remove contamination from all but one of the samples. None of the 20 juvenile herring had more than two microsatellite alleles after either treatment, demonstrating lack of contamination and confirming our hypothesis that sample contamination was caused by the presence of sperm in the water column in wild spawning aggregations.

A reference database of RAD loci was built using sequences from 19 juvenile herring in the null treatment group; one individual was excluded from the database because it contained fewer than 1.5 million raw sequences. A total of 29,551 putative loci were initially identified, and 28,997 loci were retained in the reference database after filtering out loci with tandem repeats and highly repetitive sequences.

After removing loci that were out of HWE in every population, we identified 3,502 biallelic RAD loci that were sequenced at a minimum read depth of 10 sequences in more than 80% of individuals and had a minor allele frequency that exceeded 0.05 in at least one of the populations. We found that 93% of these loci aligned exactly once to the closely-

related Atlantic herring genome. Locus-specific estimates of F_{IS} were distributed around zero (
Supplemental Figure 1-1), which is concordant with expectations under HWE. A total of 240 herring had less than 20% missing genotypes and were retained in the final data set.

Sequencing quality was robust and genotyping error was low for juvenile samples in the null and bleached treatment groups. Juvenile samples treated with bleach were characterized by slightly more sequences containing the restriction site (RADtags), loci per sample, and average read depth (Figure 1-1). However, the genotype mismatch rate between treatments in the replicated juvenile individuals was very low (1.8 ± 1.4%, mean ± SD), and similar to repeated bleach treatments (1.4 ± 1.3%). The distribution of genotype mismatches did not differ statistically between replicate individuals in the same (bleached) or across (null vs. bleached) treatment groups (Wilcoxon sign rank test, $p = 0.55$), indicating that treatment with bleach does not alter the endogenous ("true") genotype of a sample.

Impacts of contamination on individual level variation

As expected, multilocus individual heterozygosity (H_I) was higher in the untreated adult samples than in any samples that were cleaned with bleach (Figure 1-2). Samples in the dirty cocktail group (N= 4) exhibited high H_I (median = 0.45) but low variation in H_I among individuals (25^{th} and 75^{th} quantiles = 0.44 - 0.46). In comparison, adult herring samples in the null treatment group (N =11) had slightly lower but more variable H_I (median = 0.41, 25^{th} and 75^{th} quantiles = 0.31-0.42), but the maximum H_I observed in this group was as high as 0.60. Adult herring samples treated with bleach (N = 174), were characterized by much lower H_I (median = 0.18, 25^{th} and 75^{th} quantiles = 0.17 - 0.20). These values were similar to that observed for non-spawning juvenile herring (N = 20), in the null (median H_I = 0.18, 25^{th} and 75^{th} quantiles = 0.17 – 0.19) and bleach (median H_I = 0.18, 25^{th} and 75^{th} quantiles = 0.18 – 0.20) treatments.

However, there was some evidence for residual contamination in cleaned adult samples, as 8% (14/174) of those samples had H_I that was above the range observed in juvenile samples (Figure 1-2).

Intraspecific contamination affected patterns of individual differentiation, as shown by PCA (Figure 1-3). When all samples were included in the same analysis, most of the variation was driven by contaminated adult samples (Figure 1-3). When these contaminated samples were removed from the analyses, less variation was explained by the first axis but outlier samples were still evident (Figure 1-3). These samples consisted of 14 adult herring that were treated with bleach but whose H_I was relatively high (between 0.25 and 0.34) and exceeded the maximum value observed in juvenile samples (0.23); we hereinafter refer to these samples as H_I outliers. Once these H_I outliers were removed from the analysis, Cherry Point adults and juveniles clustered separately from Quilcene Bay and Spiller Channel samples (Figure 1-3). Furthermore, cleaned adult samples collected from two different years at Cherry Point clustered together with juvenile samples originating from the Cherry Point population.

Multiple runs of *Structure* identified $K = 2$ as the most likely number of groups when only cleaned data were included in the analysis. This result was supported by estimates of the posterior probability of the data given K clusters ($LnP(D)$) and ΔK (Figure 1-4). Fish collected at Cherry Point (adults and juveniles) formed a distinct cluster, while fish collected at Quilcene Bay and Spiller Channel strongly assigned to a second cluster. In contrast, when all samples (including contaminated adults) were included in the same *Structure* analysis, $LnP(D)$ and ΔK did not converge on the same answer (Figure 1-4). The posterior probability of the data given K clusters was highest at $K = 4$, while the distribution of ΔK showed peaks at both $K = 2$ and $K = 4$ (Supplemental Figure 1-2). At $K = 2$, the estimated ancestry coefficient of bleached samples was

symmetric across all sampling locations (Q = 0.82 ± 0.02, mean ± SD), while it was quite different for contaminated samples (Figure 1-4). At K = 4 the same pattern was observed, although population differentiation was more apparent in both clean and contaminated samples (Figure 1-4). In all cases, however, all individuals appeared to be highly admixed, most likely because of low population differentiation.

Impacts of contamination on estimates of population structure

Similar and considerable effects of contamination were apparent for population parameters (H_e, F_{IS}, F_{ST}) estimated from subsamples of individuals drawn from each herring population (Cherry Point, Quilcene Bay, and Spiller Channel). All contaminated subsamples and the 'dirty cocktail' had a more negative F_{IS} (indicating an excess of heterozygotes) and higher expected heterozygosity values than bleached adult subsamples lacking H_I outliers (Supplemental Figure 1-3). In addition, subsamples of juvenile herring had similar values of heterozygosity and F_{IS} before and after bleaching. Most adult subsamples had similar heterozygosity and an F_{IS} close to zero after bleaching, especially when H_I outliers were removed.

Contamination also had a clear effect on genetic differentiation between subsamples of individuals selected from the same population. Subsamples containing highly contaminated individuals were outliers in the nMDS analysis (Figure 1-3). Both the 'dirty cocktail' and the unbleached adult subsamples exhibited high differentiation from bleached subsamples taken from the same population (Figure 1-3, F_{ST}= 0.015 - 0.070). After bleaching, adult herring subsamples taken from the same population were less differentiated from each other (Figure 1-3, F_{ST}= -0.009 - 0.019), although subsamples containing H_I outliers exhibited higher differentiation

(Fig 3E, F_{ST}= 0.016-0.028). The lowest F_{ST} values were observed between the bleached and unbleached replicate subsamples of the same juvenile individuals.

Hierarchical AMOVAs demonstrated that contamination can inflate underlying genetic population differentiation (Table 1-2). When contaminated individuals were included in comparisons of population and treatment (Table 1-2, AMOVA 1), the differentiation between treatment groups from the same population (F_{SC}) was greater than the differentiation observed between distinct populations (F_{CT}). When contaminated individuals were included in an AMOVA using subsamples of individuals (Table 1-2, AMOVA 2), contamination inflated the overall F_{ST}. Contamination also increased the differentiation between population groups (F_{CT}) as well as the differentiation among subsamples within a population (F_{SC}). Adding individual level analyses into the AMOVA did not change these trends, although the presence of contaminated samples was clearly indicated by more negative F_{IS} values.

DISCUSSION

Effects of contamination

Our results demonstrate that intraspecific DNA contamination affects patterns of individual and population variability, causes an excess of heterozygotes, and biases estimates of population structure. However, contamination could be easily removed, and treatment of tissues with bleach did not affect the quality of resulting sequencing results. Our results therefore highlight the importance of identifying and removing contamination in tissues intended for RAD sequencing.

Signals of intraspecific DNA contamination are more subtle in SNPs compared to microsatellite loci. In highly variable markers such as microsatellites, heavily contaminated individuals are easily identified by the presence of more than two alleles (in a diploid species) at

a single locus (Mitchell et al. 2008). In contrast, contaminated samples genotyped at biallelic SNPs simply exhibited higher individual heterozygosity (H_I) relative to uncontaminated sample. Nevertheless, SNP data appeared more sensitive to contamination than microsatellites: while only 35% of unbleached adult herring had three or more microsatellite alleles per locus, 82% of those same samples exhibited elevated H_I relative to juvenile herring.

These findings underscore the utility of using clean samples to estimate empirical distributions of H_I. A modest number of clean reference samples can be used to construct a baseline for comparison with potentially contaminated samples using the simple metric of H_I. Furthermore, H_I is a standard metric that is commonly reported in population genetic studies (Hoffman et al. 2014, Kjeldsen et al. 2016, Tarpey et al. 2017). To our knowledge, this is the one of the first studies of wild populations to examine patterns of H_I as a quality-control measure, even though a related metric (ratio of heterozygous/non-reference homozygous sites) is commonly used in the quality control of human genomic data (Wang et al. 2015). We recommend that researchers examine the distribution of H_I in their data across individuals and populations, and carefully consider whether outlier samples could be caused by intraspecific DNA contamination.

However, we recognize that interpreting H_I outliers in species with very small effective population sizes or inbreeding could be more complicated. Individual heterozygosity and inbreeding are strongly correlated with each other when population sizes are very small and mating systems are highly skewed (e.g. polygyny, selfing) (Balloux et al. 2004, Hoffman et al. 2014). Therefore, if individual heterozygosities were highly variable between individuals and/or populations, higher values of H_I in outbred individuals, immigrant individuals, or highly diverse populations could be mistaken for a signal of contamination. If those individuals were removed

from a dataset because they were mistaken for contamination, it would lead to be a reduction in the average heterozygosity of that population and bias sampling. For species with large populations and potentially high gene flow, such as herring (Beacham et al. 2008, Limborg et al. 2012, Lamichhaney et al. 2017) and many other marine fishes (Reiss et al. 2009, Knutsen et al. 2011), variability in individual heterozygosity should be low. Our results suggest that F_{IS} estimated even in relatively small subsamples of individuals (N= 4-7) is a sensitive indicator of contamination, that may be useful when H_I is variable.

Marine species are characterized by weak population differentiation that is sensitive to sampling errors (Waples 1998). A possible consequence of contamination would be that "noise" introduced into a dataset through contaminating alleles would overwhelm faint signals of genetic differentiation between populations. Indeed, this hypothesis was confirmed by our results; contaminated samples appeared as outliers in every analysis, and led to inflated estimates of population differentiation (F_{ST}) and differentiation among subsamples within a population (F_{SC}) in an AMOVA framework. Clustering approaches were also strongly affected by contamination: heavily contaminated individuals and population subsamples were outliers in PCA and nMDS analyses, and may thus impact the interpretation from such approaches. *Structure* results were also dominated by contaminated samples, and *LnP(D)* and ΔK did not converge on the same value of K when these contaminated samples were included in the data. Without contaminated samples, *Structure* detected subtle but clear population structure. Contamination can therefore distort true population structure, which is especially problematic in the context of conservation genetics and resource management, as genetic data are often used to help delineate conservation or management units (Palsbøll et al. 2007, Funk et al. 2012, Scribner et al. 2016). Thus, it is

possible that contaminated genotypes could lead to the erroneous designation of management units and the accidental overexploitation of harvested populations.

Efficacy of bleach treatment

Our research also confirms the efficacy of bleach treatment as a method to decontaminate tissue samples collected for RAD sequencing in challenging field conditions. Bleach removed the majority of contaminant DNA on samples collected from spawning adult herring; using this method, we were able to salvage 92% of adult samples collected during active spawn events and discover 3,502 polymorphic RAD loci in Pacific herring. After decontamination with bleach, only one sample was identified by microsatellites as being contaminated. However, a modest number of bleached adult samples (8%) were characterized by elevated values of H_I, which could be indicative of small amounts of residual contamination. It is possible that the concentration and/or duration of the bleach treatment was insufficient to remove all traces of contamination, and that low levels of residual contamination were still detectable in RAD sequences generated from these samples.

Once contaminated individuals were removed from the data set, subsamples of individuals taken from the same location produced very concordant estimates of F_{ST}, even though subsample sizes were tiny (N= 4- 7). However, it has been shown that reliable F_{ST} estimates can be obtained from very few individuals if loci can be sampled without bias (Willing et al. 2012). Furthermore, while the separation between Cherry Point herring and samples from Quilcene Bay and Spiller Channel in both individual and subsample clustering approaches confirms previous studies of Pacific herring (Small et al. 2005, Mitchell 2006, Beacham et al. 2008), which found that Cherry Point herring were reproductively isolated from other populations due to differences in their spawn timing, an nMDS and AMOVA based on

subsamples detected subtle but significant differentiation between Quilcene Bay and Spiller Channel, which were previously both considered part of the same population (Small et al. 2005, Mitchell 2006). This result indicates that analyses based on small subsamples of individuals may be more powerful than those based on full samples, as suggested by (Nielsen et al. 2012).

Previous research has shown that RAD sequencing requires very high-quality DNA as input, otherwise there is a significant reduction in the number of raw sequences produced (Graham et al. 2015). Treating tissue samples in a dilute solution of bleach did not hinder the construction of RAD sequencing libraries, reduce the number of loci discovered in each sample, or affect the quality of sequence reads. Instead, juvenile samples treated with bleach yielded slightly more loci and were characterized by greater read depth per locus when compared to the same sample in the null treatment. This is most likely due to batch effects caused by slight differences in the amplification success of pooled DNA libraries, which exclusively contained either samples from the null or bleached treatment group. Importantly, we found that bleach did not degrade the endogenous DNA of tissue samples; on average, 98% of loci had matching genotypes when we compared replicate extractions from the same juvenile herring (across and within treatment groups). This genotyping error rate is similar to rates observed in conventional RAD sequencing studies (Mastretta-Yanes et al. 2015, Fountain et al. 2016). In addition, the fact that juvenile samples (from either treatment) and cleaned adult samples (from both sampling years) from Cherry Point clustered together lends further support that bleach treatment did not degrade endogenous DNA and cause false patterns of genetic differentiation.

Although the problem of sperm contamination may be specific to broadcast spawners, intraspecific DNA contamination remains a possible source of error for wild-caught specimens of most species. Therefore, researchers will have to evaluate the risk of contamination on a case-

by-case basis. While treatment with bleach is a relatively simple and cost-effective way to clean adult tissue samples, it might only be appropriate for studies where robust pieces of tissue are available. For example, when we applied this method to delicate one-day-old herring larvae, almost no DNA could be recovered (data not shown). Thus the concentration and/or duration of bleach treatment might have to be adjusted for studies targeting very delicate samples. In addition, special consideration should be given to sampling conditions, such as the bulk collection (Greenstone et al. 2011, King et al. 2011) or storage of specimens that could result in the accidental mixing of bodily fluids or cells. For example, in forensic science, considerable attention has been given to the potential of intraspecific contamination during sample collection (Cale et al. 2016) and sample processing in the laboratory (Vandewoestyne et al. 2011), though such practices are less common in molecular ecology.

In conclusion, we show that intraspecific DNA contamination can affect subtle patterns of population structure that are characteristic of many marine fish. We verified that treatment with bleach is an appropriate method for removing surface contamination from tissue samples without degrading endogenous DNA, resulting in reproducible genotypes from RAD sequencing. Our approach is likely to be applicable to tissue samples from other species.

Table 1-1. Sampling locations and associated collection information for samples used in this study. Approximate GPS coordinates are provided for herring collected from Spiller Channel in 2001.

Sampling location	Latitude	Longitude	Sampling dates	Sexual maturity	Treatment groups	Sample size
Spiller Channel, BC	52.372	-128.188	3/14/2001, 4/4/2014	Spawning adult	Null, Bleach	11
Quilcene Bay, WA	47.808	-122.860	3/8/2012	Spawning adult	Null, Bleach	6
Cherry Point, WA	48.932	-122.798	9/21/2015	Juvenile	Null, Bleach	20
Spiller Channel, BC	52.372	-128.188	4/3/2015	Spawning adult	Bleach	48
Quilcene Bay, WA	47.808	-122.860	4/7/2014	Spawning adult	Bleach	48
Cherry Point, WA	48.932	-122.798	5/12/2014, 5/9/2016	Spawning adult	Bleach	98

Table 1-2. AMOVA results using two different hierarchical groupings. In AMOVA1, groups are defined by population (Cherry Point; Quilcene Bay; Spiller Channel) and subgroups consist of the two different treatments (bleach, null). In AMOVA 2, groups are defined by population and subgroups consist of subsamples of individuals (N = 4-6); different iterations of this AMOVA were conducted excluding untreated individuals and HI outliers. FSC is the differentiation among subsamples within a group, while FCT represents the differentiation among groups (i.e. among the three populations). **Bold formatting**: P< 0.001, no formatting: P> 0.05.

	Without individual level			*With individual level*		
AMOVA 1	F_{ST}	F_{SC}	F_{CT}	F_{IS}	F_{SC}	F_{CT}
All individuals	**0.0270**	**0.0414**	-0.0150	**-0.1034**	**0.0414**	-0.0139
AMOVA 2						
All individuals	**0.0255**	**0.0046**	**0.0209**	**-0.1100**	**0.0145**	**0.0210**
Bleached individuals	**0.0204**	**0.0010**	**0.0194**	**-0.0604**	**0.0065**	**0.0194**
Bleached individuals, no H_I outliers	**0.0206**	**0.0007**	**0.0199**	**-0.0356**	**0.0041**	**0.0199**

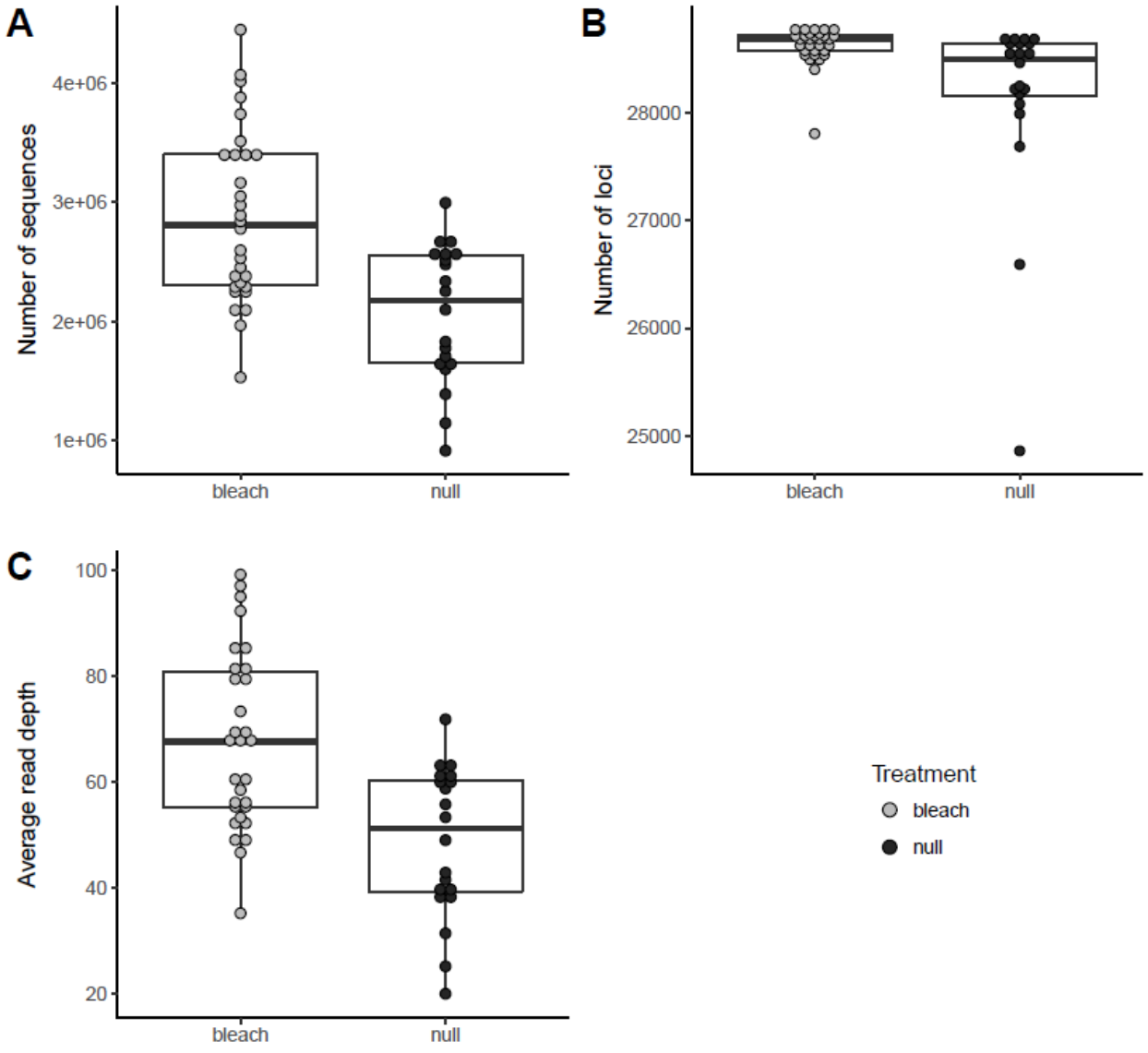


Figure 1-1. Sequencing quality data for juvenile herring in the null and bleach treatment groups. Each dot represents an individual herring sample. A) Number of raw sequences per sample containing a restriction site, B) number of RAD loci identified in each sample by *pstacks*, and C) average read depth per locus for each sample.

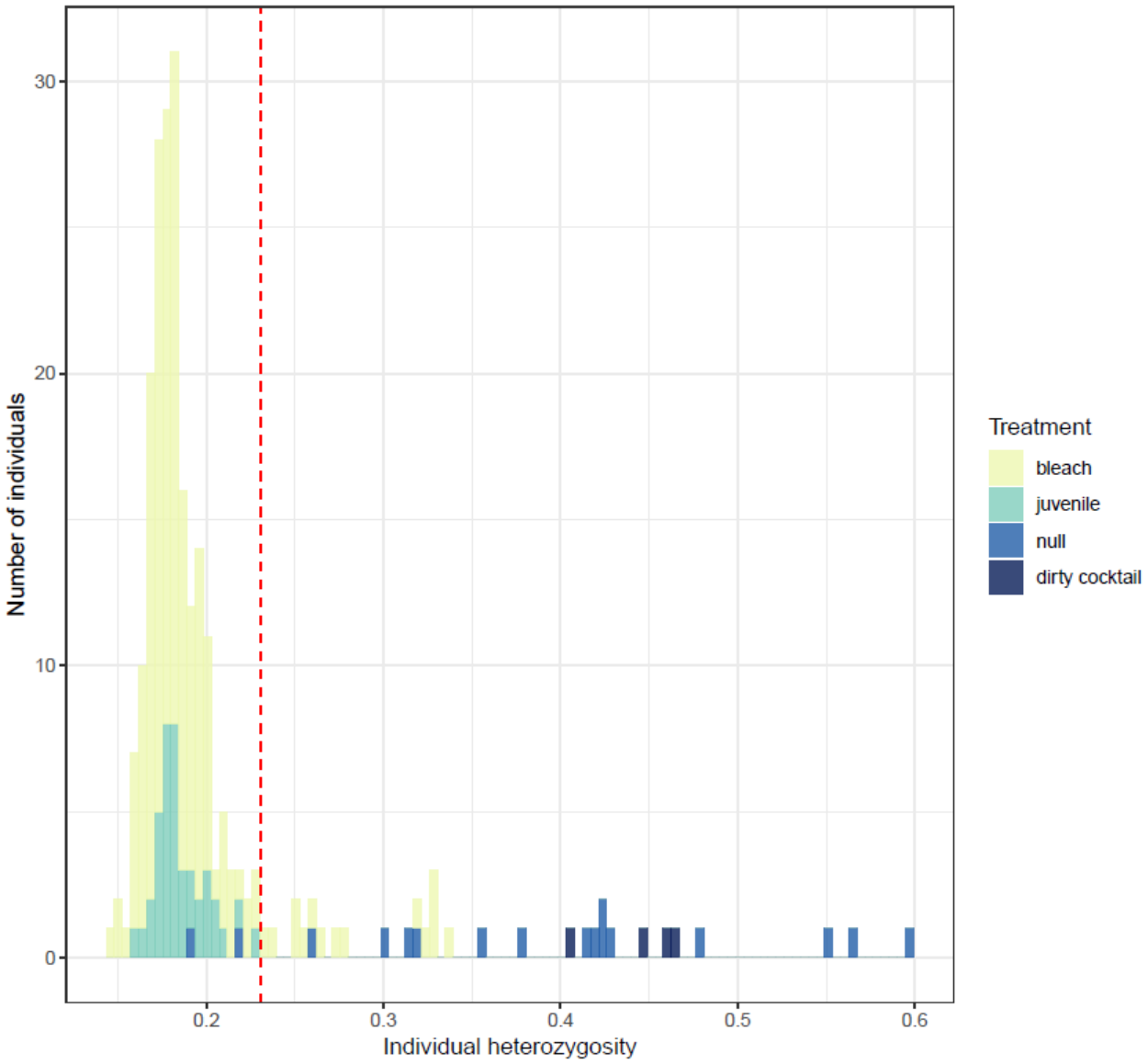


Figure 1-2. Distribution of H_I in each treatment group. Colors represent different treatments and the dashed line shows the upper limit of H_I observed in the juvenile samples. Bleached adult samples to the right of the dashed line are "H_I outliers" that likely contain residual contamination.

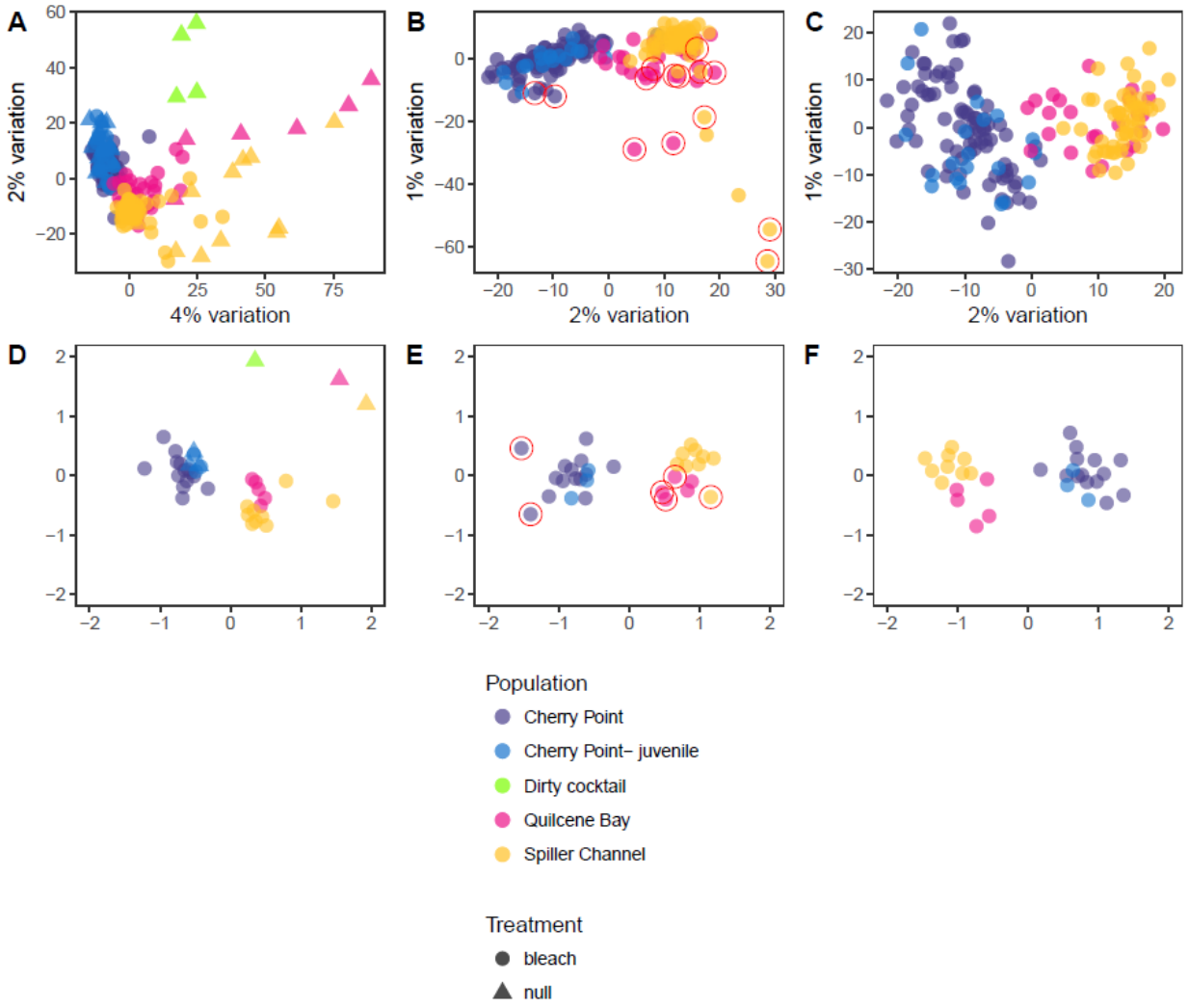


Figure 1-3. PCA (panels A, B, C) and nMDS (panels D, E, F) plots of herring genotyped at 3,502 RAD loci. In the PCA, each point represents an individual herring, while in the nMDS each point represents a subsample of multiple herring (N= 4-7). Different colors depict the population from which the samples were collected, while shapes (circle or triangle) are indicative of treatment group. Note that juvenile herring samples (in both null and bleach treatments) cluster together with adult samples collected from the same population (Cherry Point). A) PCA of all samples, B) PCA of bleached samples; H_I outliers are circled in red, C) PCA of bleached samples when H_I outliers are removed, D) nMDS of all samples, E) nMDS of bleached samples; H_I outliers are circled in red, F) nMDS of bleached samples when H_I outliers are removed.

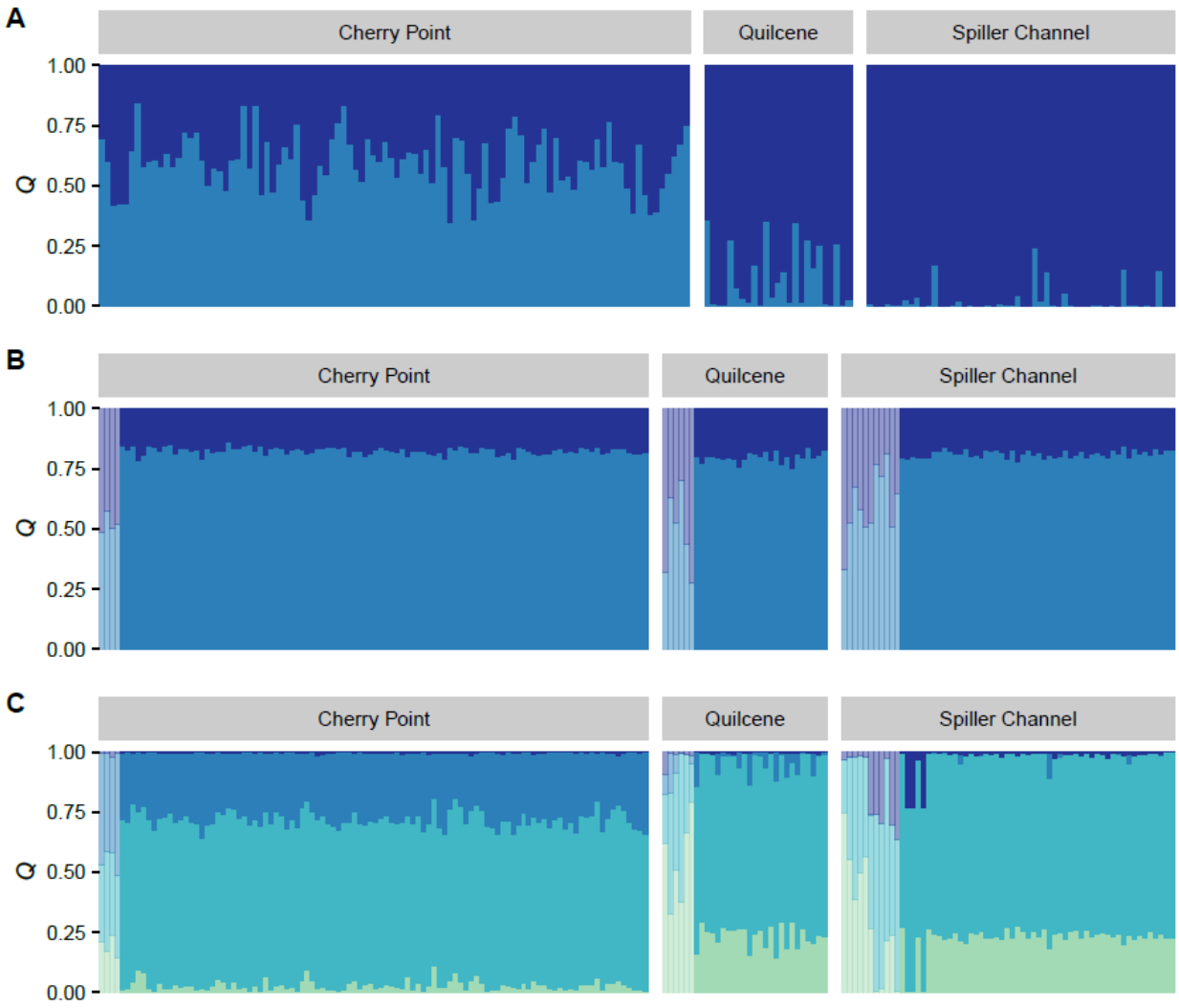
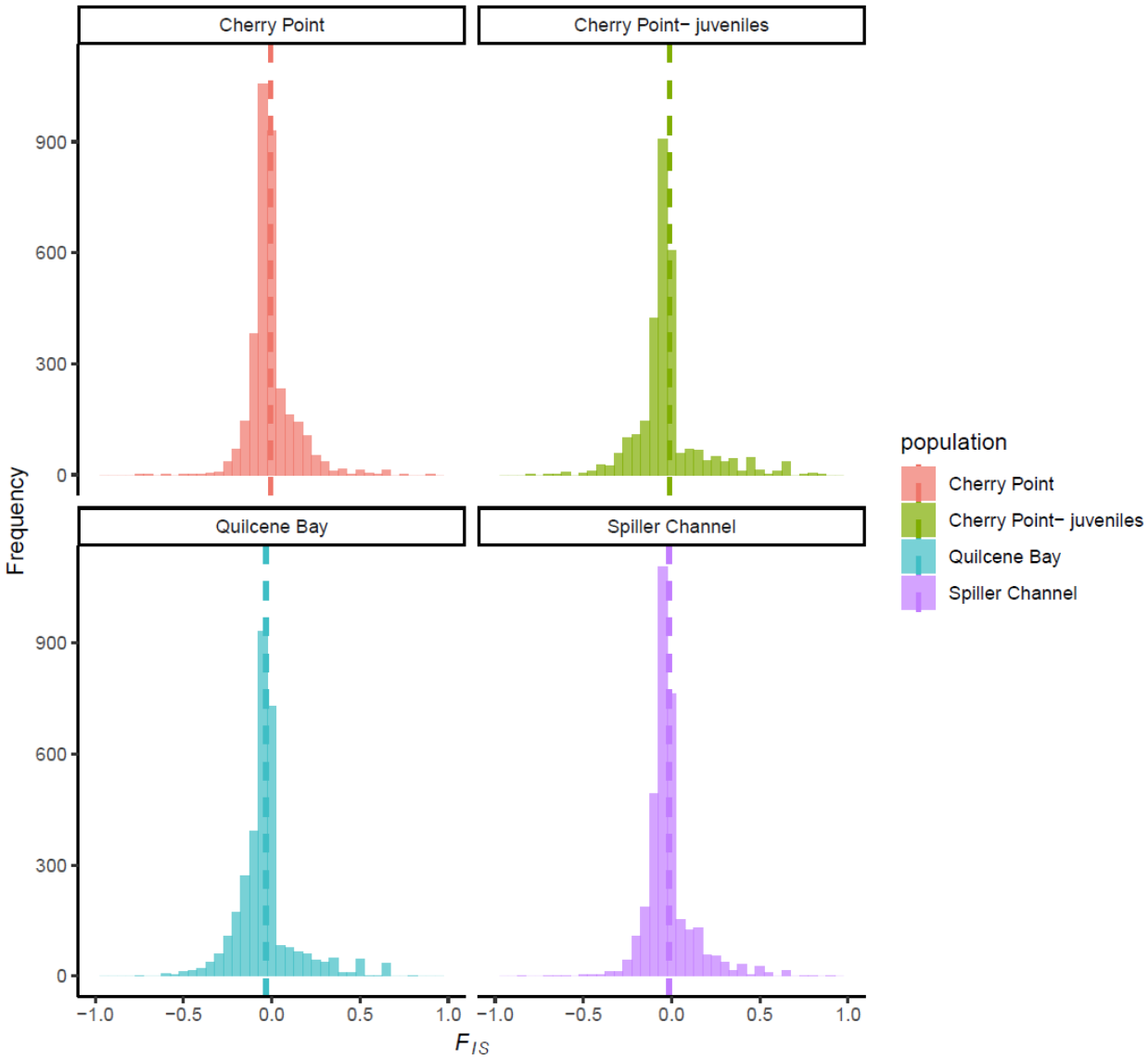
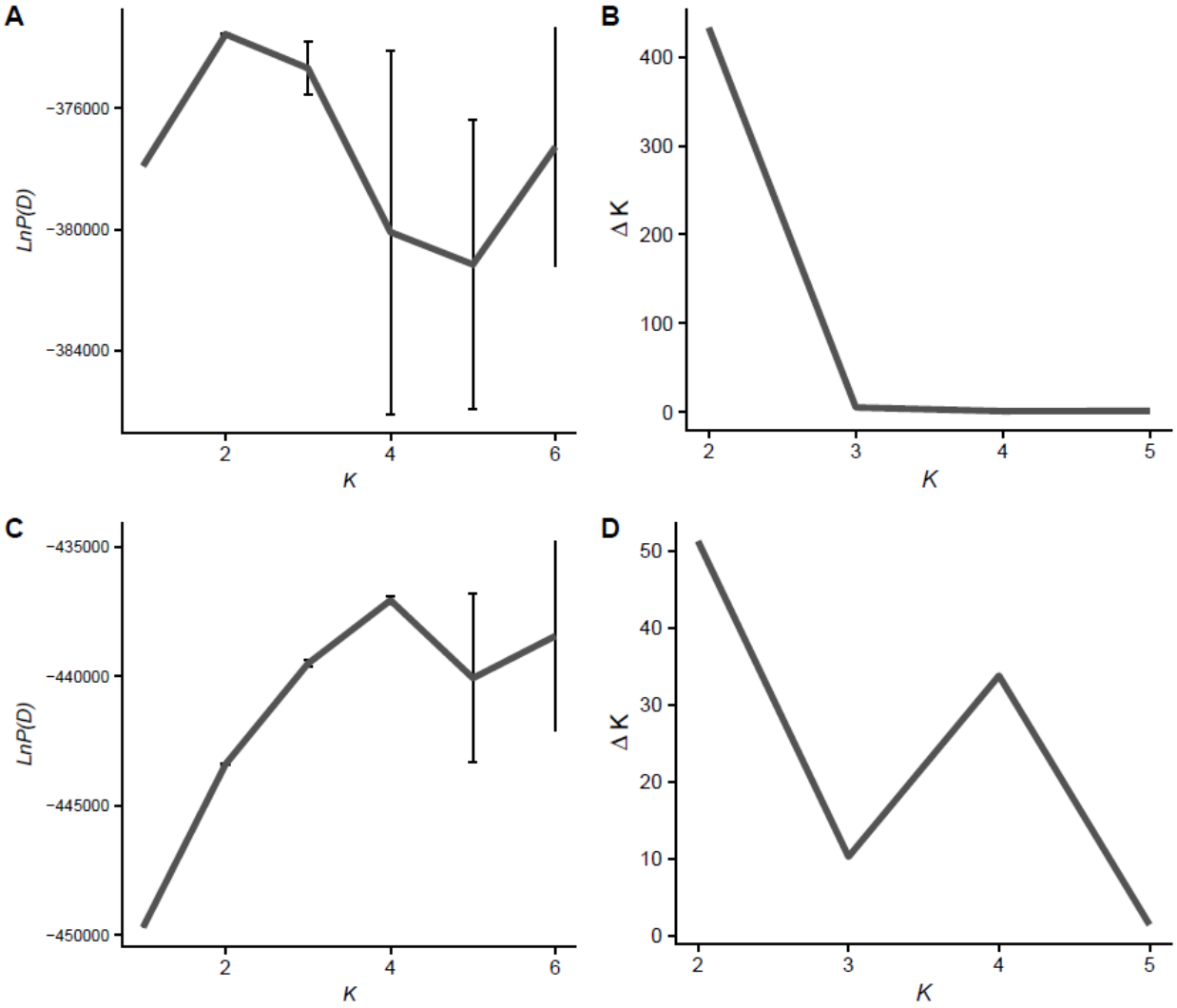


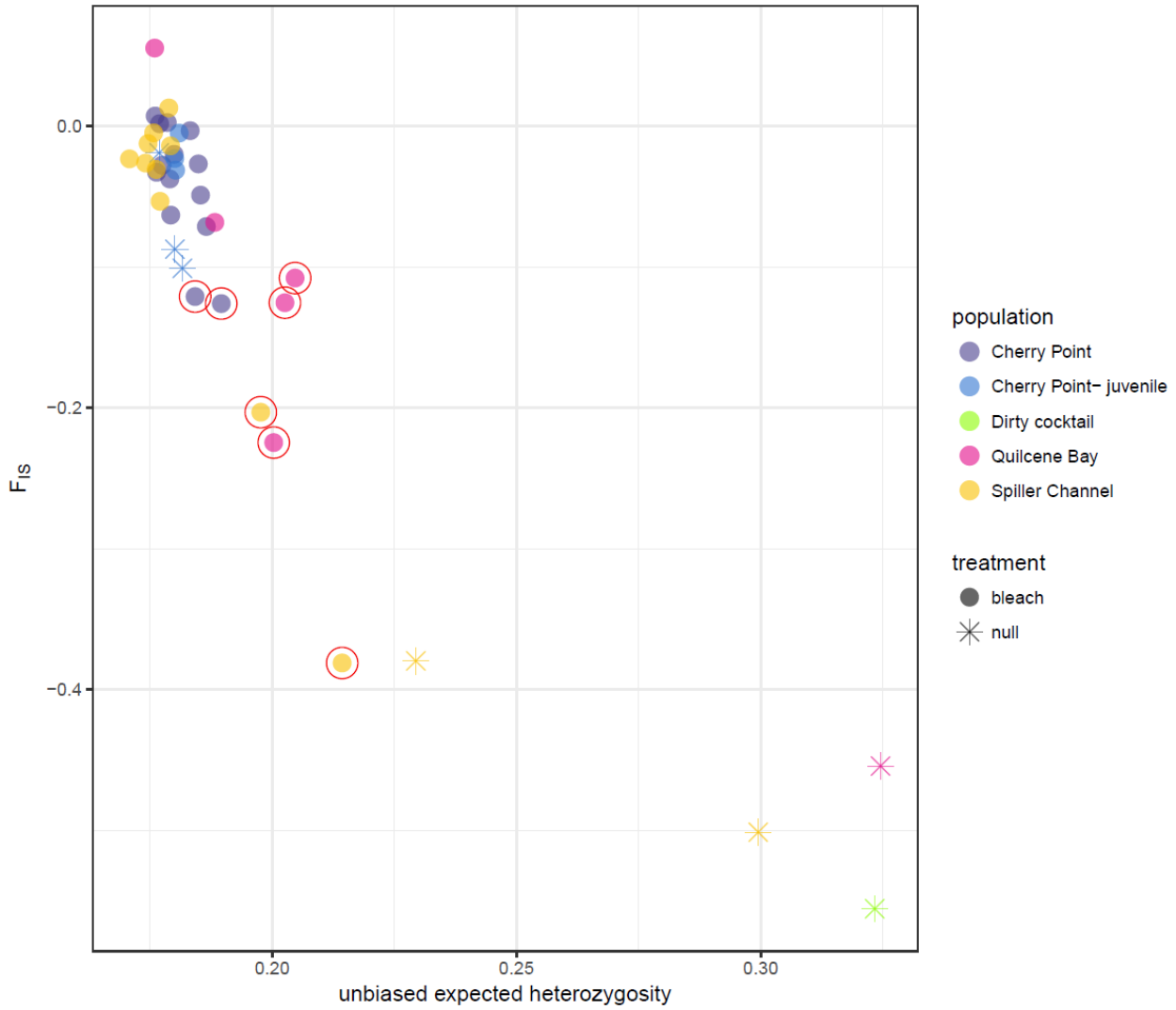
Figure 1-4. Population structure estimated using *Structure*. Each sample is portrayed by a vertical line which consists of colored segments, representing the estimated fraction of an individual's ancestry (*Q*) belonging to *k* clusters. Individuals represented by transparent bars are contaminated adult samples. A) *Structure* analysis using only bleached samples and no H_I outliers; *LnP(D)* and ΔK unambiguously identify $K = 2$ as the most likely number of clusters. These clusters correspond to the major known spawning phenotypes of Pacific herring ("late-spawners" and "primary-spawners"). B) *Structure* analysis using all samples and $K = 2$. The presence of contaminated samples alters the values of *LnP(D)* and ΔK, compared to the clean data set. C) *Structure* analysis using $K = 4$ and all samples.



Supplemental Figure 1-1. Distribution of per locus F_{IS} estimated from 3,502 RAD loci. Each panel represents a different sampling location. Dotted lines within each panel show the mean value of F_{IS} for that sampling site.



Supplemental Figure 1-2. Estimates of the posterior probability of the data given *K* clusters (*LnP(D)*) and *ΔK* using *Structure*. Results are shown for analyses that included bleached samples when H_I outliers were removed (Panels A and B) and analyses that included all samples (Panels C and D).



Supplemental Figure 1-3. F_{IS} and unbiased expected heterozygosity in population subsamples before and after bleaching. Each population is represented by a different color and subsamples consist of several individuals (N = 4 – 7). Bleached adult subsamples that contain H_I outliers are circled in red.

Chapter 2. **Parallel evolution of reproductive timing in Pacific and Atlantic herring**

SUMMARY

Variation in the timing of reproduction underlies key evolutionary and ecological processes in the wild. For example, temporal diversity in reproduction mediates gene flow (Hendry and Day 2005) and so can contribute to speciation (Yamamoto and Sota 2009, Taylor and Friesen 2017). At the ecosystem level, variation in reproductive behavior supports ecological portfolio effects (Schindler et al. 2010), and "resource waves" associated with spatial variation in reproductive timing prolong foraging opportunities for mobile predators (Armstrong et al. 2016). While the evolutionary and ecological significance of reproductive timing is widely recognized, little is known about its genetic basis, and what it is known stems largely from terrestrial plants and insects (Sakai and Ishida 2001, Buckler et al. 2009, Kaiser et al. 2016, Romero Navarro et al. 2017). Here we show that genetically determined reproductive timing drives population structure in a pelagic fish, the Pacific herring. We used genomic data from wild spawning aggregations to investigate the evolutionary basis of reproductive timing, and found genomic regions highly correlated with spawn timing in both Atlantic and Pacific herring. Loci differentiating populations with different spawn timing in both species were clustered around *SYNE2*, a gene which influences the development of retinal photoreceptors in vertebrates. However, none of the variable SNPs in *SYNE2* were shared between the species, suggesting parallel evolution in genes that may play a role in the photoperiodic regulation of reproduction. The observed genetic diversity likely underlies ecological resource waves provided by herring, and provides an example of how population-level genetic diversity may influence ecosystem processes.

MAIN TEXT

Reproductive allochrony may be an important evolutionary driver in the marine environment, where the high mobility of many species and few physical barriers to migration provide limited opportunities for spatial genetic divergence to arise. Furthermore, the genetic basis of reproductive timing has important ecological implications as temporal matching between larval emergence and plankton production determines recruitment success in many temperate marine species (Cushing 1969, Cushing 1990, Platt et al. 2003), and this trophic coupling may be disrupted by climate change (Asch 2015). Sister species Pacific and Atlantic herring (*Clupea pallasi*, *C. harengus*), estimated to have diverged 2.2 million years ago (Martinez Barrio et al. 2016), provide a robust system in which to study the genomic basis and evolution of reproductive phenology. Spawn timing varies across broad latitudinal gradients as well as within narrow geographic regions for both species (Haegele and Schweigert 1985), and populations with distinct spawning seasons (e.g., winter vs. spring spawners) are genetically differentiated (Small et al. 2005, Ruzzante et al. 2006, Beacham et al. 2008). Furthermore, the predictability of spawn timing within populations (Sinclair and Tremblay 1984) supports the long-standing hypothesis that reproduction and larval emergence is synchronized with cycles of marine productivity (Cushing 1969, Cushing 1990).

Recent studies have identified genomic regions associated with spawn season in Atlantic herring (Martinez Barrio et al. 2016, Lamichhaney et al. 2017), but it is unknown whether these same regions contribute to the fine-scale diversity in reproductive phenology of Pacific herring. If so, it would demonstrate a genetic basis of allochrony that either predates speciation or that arose in parallel between Atlantic and Pacific herring. Current fishery management practices (Tillotson and Quinn 2018) as well as global climate change (Edwards and Richardson 2004)

may irrevocably reduce this diversity before its significance is fully recognized, thereby reducing the ability of species to adapt to future environmental conditions.

We first tested whether subtle differences in spawn timing (in the range of weeks to months) influence the genetic population structure of Pacific herring. We sampled 1,104 individuals from 23 spawning aggregations on the Pacific Coast of North America whose spawn timing ranged from January to June (Figure 2-1). Individuals were sequenced to an average read depth of 75X and genotyped at 7,261 polymorphic restriction site-associated loci that aligned to the Atlantic herring genome (Supplementary Materials). A pattern of isolation by time (Hendry and Day 2005) was detected (Mantel $r = 0.51$, Mantel $p = 0.00099$, Figure 2-2) despite low overall levels of genetic differentiation (global $F_{ST} = 0.014$), indicating reduced gene flow among sites with different spawn times. However, within a specific spawning period there was also evidence of isolation by distance (Wright 1943) (Figure 2-2), and a discriminant analysis of principal components revealed differentiation amongst populations spawning in different geographic regions (Figure 2-2). We assessed the relative importance of reproductive timing and geographic distance in structuring Pacific herring populations by conducting a model-based Bayesian analysis of population structure (Foll and Gaggiotti 2006, see Supplemental Materials for more details). This analysis estimated the highest posterior probability for a model that included information on spawn timing (*posterior probability* = 0.96), followed by a model that incorporated both information on spawn timing and geographic distance between spawning sites (*posterior probability* = 0.04). Pacific herring population structure is therefore primarily driven by spawn timing, with limited spatial dispersal within a particular spawning period. This fine-scale genetic structuring is notable, given the large census population sizes and high dispersal capabilities that are characteristic of pelagic fish. Additionally, the observed patterns of

population structure were stable through time (Supplementary Materials), suggesting that spawning sites were used by populations that exhibited seasonal and geographic spawning fidelity across multiple generations.

Environmental cues such as temperature and photoperiod trigger seasonal reproduction in many temperate fishes (Migaud et al. 2010), with the latter being a very predictable and consistent environmental signal. We identified genomic regions associated with the timing of reproduction by finding loci whose allele frequencies were correlated with spawning photoperiod at each sampling location (hereinafter called correlation loci), using a Bayesian framework that accounts for the covariance of alleles due to hierarchical population structure and sampling error (Coop et al. 2010, Günther and Coop 2013b). There were 130 SNPs that were strongly correlated with spawning photoperiod ($\log_{10}$ Bayes factor > 2; Figure 2-3). These SNPs were distributed across 56 distinct genomic scaffolds, suggesting that spawn timing is a polygenic trait. Differentiation at these loci was an order of magnitude higher than average genetic differentiation (Figure 3C; mean F_{ST} of correlation loci = 0.1; mean F_{ST} of other loci = 0.01; p-value < 2.2×10^{-16}, Wilcoxon signed-rank test). SNPs with the strongest correlations ($\log_{10}$ Bayes factor > 10; Figure 2-3) mapped to the mRNA encoding *SYNE2*, a nuclear membrane protein that influences the position of retinal photoreceptors early in vertebrate development (Tsujikawa et al. 2007, Yu et al. 2010, Maddox et al. 2015). These associations provide support for a mechanistic link between the observed genetic variation and the photoperiodic regulation of spawn timing in Pacific herring.

Such mechanistic links provide an opportunity to test for parallel evolution, that is the evolution of similar or identical mutations in independent lineages that result in phenotypic similarities (Stern 2013). A recent study of Atlantic herring demonstrated that specific genomic

regions related to the photoperiodic regulation of reproduction were highly associated with spawn timing in populations from both sides of the Atlantic Ocean (Lamichhaney et al. 2017), but the presence of shared haplotypes led the authors to conclude that adaptive alleles likely emerged from standing variation in the ancestral population or had been spread by gene flow (collateral evolution *sensu* Stern 2013). We tested for parallel evolution in Pacific and Atlantic herring by investigating whether the same genomic regions were correlated with reproductive phenology in both species. We found that SNPs near or within four genes (*SYNE2, NRXN3B, CEP128, HK3*) were highly correlated with reproductive timing in both species (Figure 2-3), suggesting that similar genomic regions may underlie highly complex traits such as the timing of reproduction. Moreover, SNPs in the genomic region around *SYNE2* were among the most highly associated with spawn timing in both species and exhibited concordant peaks of correlation (Figure 2-3), a pattern that is unlikely to be the product of coincident demographic processes. Finally, although there was overlap in the genomic regions correlated with reproductive timing, SNPs identified in Pacific herring differed from those identified in Atlantic herring; for example, in Pacific herring we identified five SNPs within *SYNE2* but those nucleotides were monomorphic in Atlantic herring. This pattern suggests that these species-specific polymorphisms arose independently and may indicate parallel evolution in sister species. A recent study of migration timing in closely related salmonids (Prince et al. 2017) observed a similar phenomenon, leading the authors to conclude that intraspecific SNPs were the product of independent evolutionary events.

We propose that spawn timing is likely a trait with a genetic basis that also structures herring populations in time and space. Given the important trophic role of forage fish in marine ecosystems (Pikitch et al. 2014), it is worth considering how the observed patterns of genetic

variation and population structure interact with other ecological processes. For example, limited gene flow between temporally separated spawning aggregations likely contributes to independent population dynamics and portfolio effects observed in Pacific herring (Siple and Francis 2016). In addition, genetic variation in reproductive timing is the basis of the prolonged resource availability that spawning herring provide to coastal environments in spring (Lok et al. 2012b). It is currently unknown how photoperiod interacts with other environmental cues (such as sea surface temperature or tidal cycle) to influence the timing of reproduction in herring. Recent studies of marine phenology indicate that climate change is causing shifts towards earlier larval emergence in some pelagic fish (Poloczanska et al. 2013, Asch 2015), a phenomenon that could lead to trophic mismatches and increased recruitment variability. However, species whose reproductive timing has a strong genetic basis and is driven by environmental stimuli other than temperature may be vulnerable to climate change because they have reduced capacity to adapt to changing thermal conditions. Furthermore, increasing sea surface temperatures could influence metabolic processes at early life history stages (Hufnagl and Peck 2011), thus leading to emergent effects that may impact survival. Fishery management practices should strive to maintain both spatial and temporal population diversity and mitigate trends in spawn timing compression (Hay and McCarter 1999), as this diversity may enhance species-level resilience to environmental variation and climate change.

FIGURES

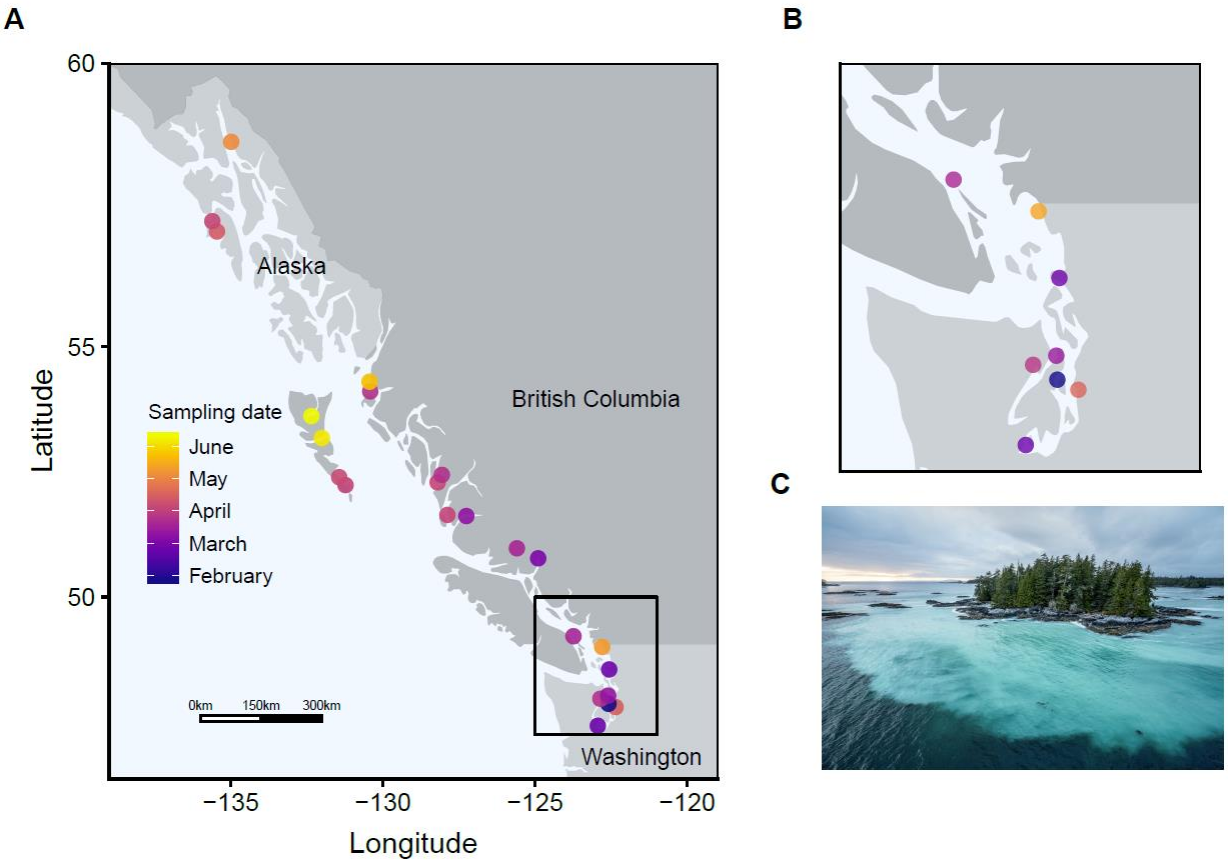


Figure 2-1. Map of all sampling locations. The color of each circle depicts the date of sampling an aggregation of spawning or sexually mature fish. B) Map of sampling sites in the Salish Sea. C) A herring spawn on the Central Coast of British Columbia; milky water indicates the presence of sperm in the water column and provides a visual cue that spawning is occurring. Photo by Ian McAllister/Pacific Wild.

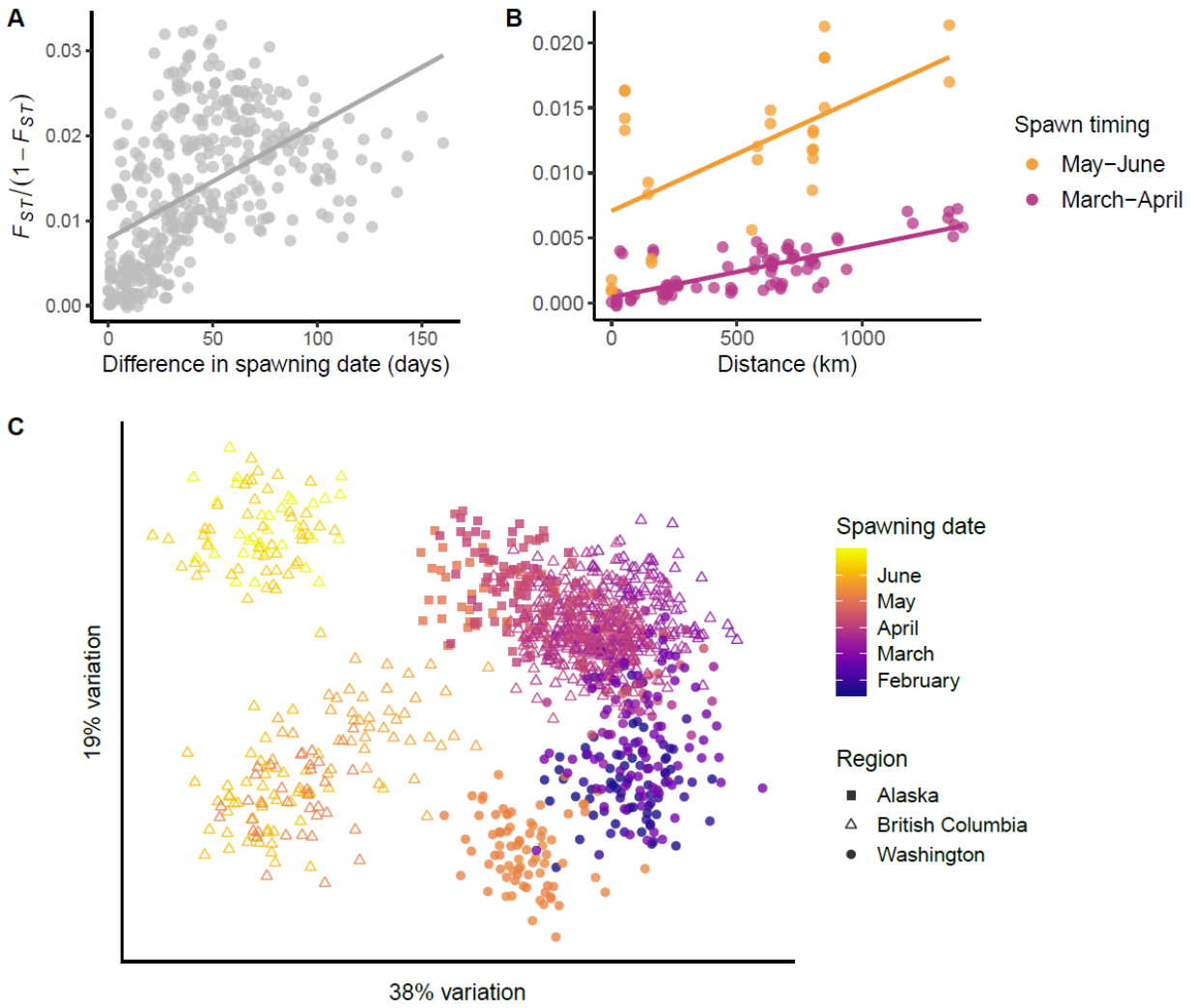


Figure 2-2. Genetic population structure in Pacific herring. A) Isolation by time in Pacific herring. Linear regression of pairwise F_{ST} to the number of days separating sampling events at each location (Mantel r = 0.51, Mantel p = 0.00099). Points represent pairwise comparisons between sampling locations. B) Isolation by distance in Pacific herring. Points represent pairwise comparisons among sampling locations within a specific spawning period (orange points: May-June spawners; purple points: March-April spawners). Linear regression of pairwise F_{ST} to the geographic distance separating sampling locations is estimated separately for each spawning period (May-June spawners: Mantel r = 0.61, Mantel p = 0.00099; March-April spawners: Mantel r = 0.77, Mantel p = 0.00099). C) Discriminant analysis of principal components. Different colors indicate the date of sampling while shapes represent the geographic region that a sample was collected from.

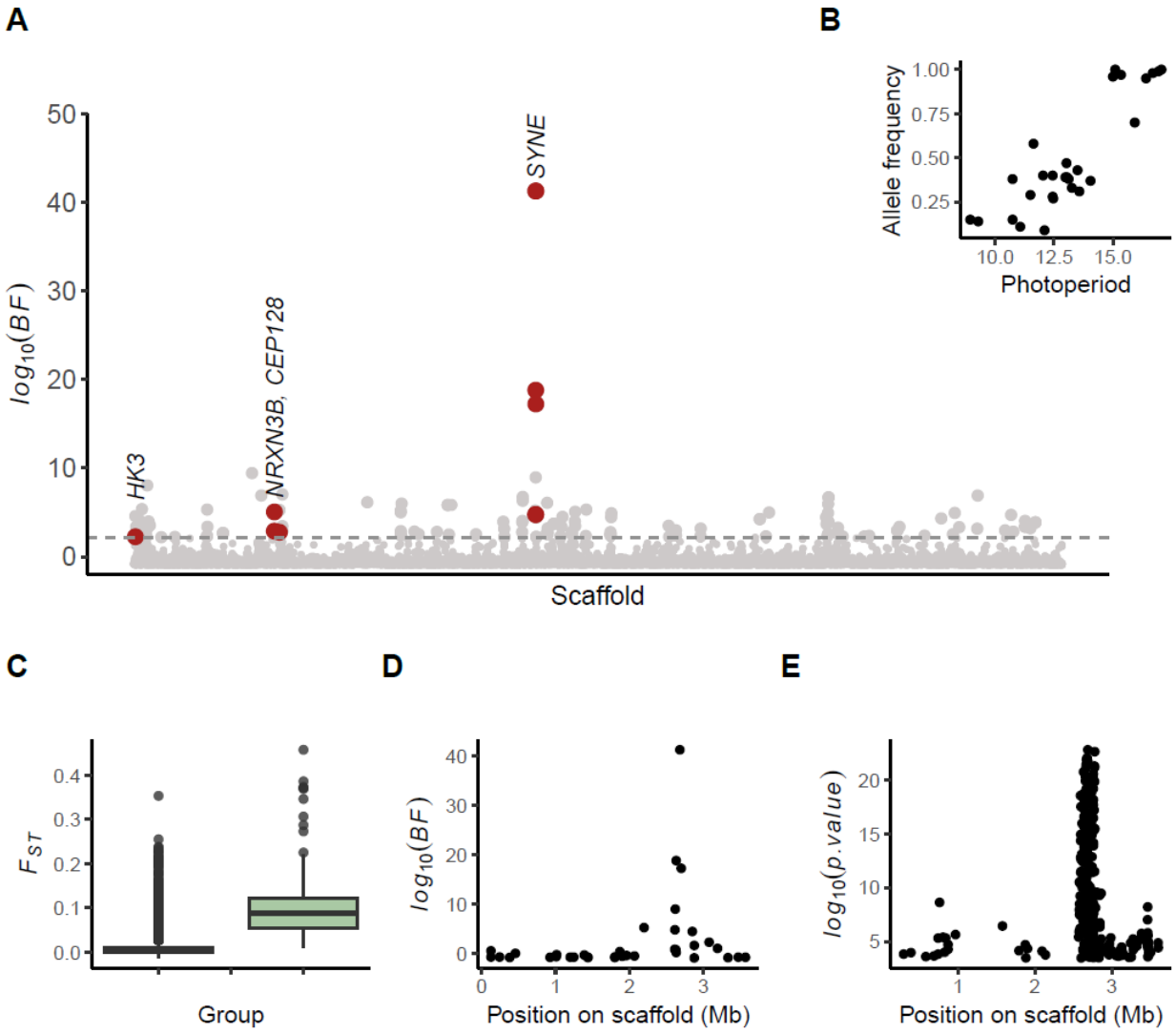


Figure 2-3. SNPs correlated with spawn timing in herring. A) SNPs correlated with the photoperiod of reproductive timing in Pacific herring. Grey points show strength of correlation (log_{10} Bayes factor) between allele frequency and photoperiod in Pacific herring. SNPs are ordered along the x-axis by their respective genomic scaffold. Red points highlight Pacific herring SNPs that are highly correlated with photoperiod (log_{10} Bayes factor >2) and are also within 5k bp of a SNP that is associated with spawn timing in Atlantic herring. These shared loci are labelled with the name of the gene they are found within. B) Regression of allele frequency to photoperiod for the SNP most highly correlated with photoperiod in Pacific herring (within *SYNE2;* F_{ST} = 0.36*;* R^2 = 0.74, p-value = 2.7 x 10^{-9}, log_{10} Bayes factor = 41). C) F_{ST} distribution of loci correlated with spawn timing in Pacific herring (green points: loci whose log_{10} Bayes factor >2; grey points: all other loci). D) Strength of correlation with photoperiod in Pacific herring on scaffold 312, the genomic scaffold containing *SYNE2*. The peak of association occurs around 2.68 Mb. E) Strength of correlation with spawn timing in Atlantic herring on scaffold 312. The peak of association also occurs around 2.68 Mb.

SUPPLEMENTARY MATERIALS

Methods

Sample collection

Approximately 48 herring were collected from each of 23 distinct spawning sites on the Pacific Northwest coast of North America (Supplemental Table 2-1). All but one of these sample collections consisted of sexually mature herring that were caught in pre-spawning aggregations (before 2013) or during active spawning events (after 2013) using nets or hook and line fishing gear. Late-stage (eyed) eggs were collected at one site (Elliot Bay, Washington). In addition, five different locations were sampled twice across different years (Supplemental Table 2-1); we hereinafter refer to these samples as "temporal replicates". All herring (N = 1,344) were collected by state, federal, or Tribal biologists or permitted subsistence fishers. Tissue was subsampled, individually preserved in 100% ethanol, and sent to the University of Washington for genetic analyses

DNA sequencing and genotyping

Prior to DNA extraction, each adult herring sample was treated with bleach to remove any potential DNA contamination, following the protocol of Petrou et al. (2019). Eyed eggs (from Elliot Bay) were not decontaminated with bleach because they were very delicate and had not been collected during an active spawning event.

Genomic DNA was extracted from each sample using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA). We followed the protocol of Etter at al. (2011) to prepare DNA libraries for restriction-site associated (RAD) sequencing. In brief, 500 ng of genomic DNA per sample was digested with the restriction enzyme *SbfI* (New England Biolabs

(NEB), Ipswich, MA). Samples were individually labeled using a custom set of 96 barcodes (Integrated DNA Technologies, San Diego, CA) and T4 DNA Ligase (NEB). Groups of 12 samples were pooled into libraries that were sheared to a length of approximately 500 base pairs (bp) using a Bioruptor sonicator (Diagenode, Denville, NJ). We subsequently selected DNA fragments that were between 300-500 bp long using a Blue Pippin (Sage Science, Beverly, MA). Blunt-end repair, 3'-dA overhang addition, and P2 adapter ligation, and PCR were conducted as described in Etter at al. (2011). PCR products were cleaned with AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) and the DNA concentration of each amplified library was quantified using the PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA). Libraries were standardized to a concentration of 10 *nM* and pooled such that 96 individuals were sequenced per lane of an Illumina HiSeq 4000 (Illumina Inc., San Diego, CA) at the University of Oregon Genomics Core Facility. The resulting sequences were single-end and had a length of 100 base pairs.

We used the computational pipeline in Stacks version 1.46 (Catchen et al. 2013) to analyze sequencing data and genotype samples. The *process_radtags* script was used to demultiplex individual samples, and remove barcodes and sequences with low quality scores (Phred score < 10). We visualized the quality of sequencing data using *FastQC* (http://www.bioinformatics.babraham.ac.uk /projects/fastqc/) and decided to trim sequences to a length of 90 base pairs to avoid sequencing errors near the end of the sequence. Subsequently, we aligned each sequence to a reference database of herring RAD loci to facilitate sequence assembly and locus identification, as described in Petrou et al. (2018). Sequences that aligned to the database were processed with the *pstacks* script to identify alleles in each sample (minimum depth of coverage to report a stack = 10). To maximize the number of loci retained, we filtered

out low-quality samples by only retaining those that contained at least 20,000 RAD loci after *pstacks* and a catalog of loci was constructed in *cstacks* using a subset of the ten most deeply sequenced individuals from each sampling location.

We filtered the data set by removing loci that had more than 20% missing data in any one sampling location. Subsequently, we applied the method of McKinney et al. (2017) to identify and remove potential paralogous loci. In brief, this approach identifies paralogous loci by combining information on the proportion of heterozygous individuals within each population, and the sequencing read depth of each allele into a z-score for each locus. The z-score describes the deviation between observed and expected allele-specific read counts from a binomial distribution. We visualized the distribution of z-scores in our data and removed loci that were more than one standard deviation from the mean (that is, they were characterized by both high heterozygosity and read depth across many populations).

In addition, single nucleotide polymorphisms (SNPs) whose minor allele frequency was less than 0.05 across all samples were removed from the data using *VCFtools* (Danecek et al. 2011). We chose to remove these low frequency polymorphisms because they are difficult to differentiate from sequencing errors and can bias tests for selection (Roesti et al. 2012). As a final quality control measure for locus assembly, we aligned loci to the Atlantic herring genome (ASM96633v1; Martinez Barrio et al. 2016) using *bowtie2* version 2.2.6 (Langmead and Salzberg 2012) and only retained loci that aligned with a mapping quality ≥ 20.

Individual herring samples were assessed for missing data and removed if they had more than 10% missing genotypes. We also screened individuals for intraspecific DNA contamination by calculating individual multilocus heterozygosity (H_I) and comparing it to a set of clean reference samples, as described in Petrou et al. (2019). Individual samples were filtered from the

data set if they were characterized by values of H_I that were above what was observed in clean reference samples ($H_I > 0.32$), as in Petrou et al. (2019).

Population structure

All analyses of population structure used RAD loci that were coded as multiallelic haplotypes rather than SNPs. Loci were tested for deviations from Hardy-Weinberg equilibrium (HWE) in each sampling location using the exact test based on 10,000 Monte Carlo permutations of alleles, as implemented in Genepop version 4 (Rousset 2008). We used the false discovery rate to correct for multiple testing of HWE with the R package *qvalue* (citation; Storey 2002). A locus was identified as being out of HWE when the *q*-value was less than 0.05. We then conducted all analyses of population structure on two different data sets: the first contained all loci, while the second only contained loci that were in HWE.

Pairwise F_{ST} (Weir and Cockerham 1984a) between sample collections and F_{IS} within each sample collection were calculated in *Genepop* (Raymond and Rousset 1995, Rousset 2008). Per-locus F_{ST}s were estimated with the R package *hierfstat* (Goudet 2005). We visualized population structure using a discriminant analysis of principal components (DAPC) with the R package *adegenet* (Jombart 2008). The *optim.a.score* function in *adegenet* was used to identify the number of principal components to retain in the analysis.

To test whether the timing of reproduction affected genetic differentiation (isolation by time; Hendry and Day 2005) we conducted a linear regression of linearized pairwise F_{ST} ($F_{ST}/(1-F_{ST})$ to the number of days separating sampling events. The statistical significance of isolation by time was evaluated using a Mantel test (Pearson's product moment correlation; 10,000 permutations) as implemented in the R package *vegan* (Oksanen et al. 2018). We also tested for isolation by distance (Wright 1943, Kimura and Weiss 1964) by conducting a linear regression

of linearized pairwise F_{ST} to the geographic distance separating sample collections. The shortest distance ("as the crow flies") separating two sampling locations was calculated using the R package *geosphere* (Hijmans 2017), and the statistical significance of isolation by distance was evaluated as described above for time.

We assessed the relative importance of reproductive timing and geographic distance in structuring Pacific herring populations by conducting a hierarchical Bayesian analysis of population structure with the program *GESTE* (Foll and Gaggiotti 2006). In brief, this analysis constructs a null model in which subpopulations evolve after splitting off from an ancestral population and experience genetic drift, resulting in population-specific F_{ST} values. Alternative models are built by incorporating environmental data through the prior distribution of population-specific F_{ST}*'s,* and the strength of the correlation between genetic and environmental data is evaluated using a generalized linear model. Finally, the relative influence of different environmental factors on the observed patterns of genetic structure is assessed using the posterior probabilities of the null and alternative models. We used geographic and temporal connectivity as environmental factors in our analysis. Temporal connectivity was defined as the difference (in days) between the spawning time of a particular sampling location and the average spawning time across all sampling sites (as in Gaggiotti et al. 2009). Geographic connectivity was defined as the mean distance between each sampling site and all other sampling sites (as in Gaggiotti et al. 2009). We used default parameters of the program (sample size = 10,000; thinning interval = 20; pilot runs = 10; pilot run length = 5,000; additional burn-in = 50,000) to estimate posterior probabilities of the different models (Supplemental Table 2-2).

Loci correlated with spawn timing

Reproductive timing in temperate fishes is primarily controlled by temperature and photoperiod (Migaud et al. 2010, Wang et al. 2010), with the latter being the most predictable and consistent environmental signal. We therefore identified genomic regions associated with the timing of reproduction by finding loci whose allele frequencies were correlated with the photoperiod of spawning at each sampling location (hereinafter referred to as "correlation loci"), using a Bayesian framework that accounts for the covariance of alleles due to hierarchical population structure and sampling error (Bayenv 2.0; Coop et al. 2010, Günther and Coop 2013a). This protocol uses biallelic SNPs as input, so we conducted this analysis using the first SNP in each RAD locus. With this method, a covariance matrix of allele frequencies between sampling locations is estimated using all SNPs and used as a null model, in which overall population structure reflects neutral processes (such as genetic drift and gene flow) and sampling error. SNPs that deviate strongly from patterns of neutral population structure are identified using an alternative model, in which an environmental variable has a linear effect on standardized allele frequency. This analytical approach has been shown to exhibit a low false positive rate under a variety of demographic histories (Lotterhos and Whitlock 2014).

Photoperiod was calculated at each sampling location and date using the *daylength* function from the *geosphere* package (Hijmans 2017) package in R, and was normalized to a mean of zero and a standard deviation of one. The covariance matrix of allele frequencies was estimated in *Bayenv2* using the first SNP in each RAD locus and 100,000 MCMC iterations. Bayes factors expressing support for the alternative model at each locus were also estimated with 100,000 MCMC iterations.

We tested for parallel evolution in two species inhabiting different oceans by investigating whether the same genomic regions and SNPs were correlated with reproductive phenology in both Atlantic and Pacific herring. Atlantic herring SNPs that are correlated with spawn timing were previously published in Lamichhaney et al. (2017). We assessed whether any Pacific herring correlation SNP was proximate to an Atlantic herring correlation SNP or gene using the annotated Atlantic herring genome (assembly ASM96633v1) and the *GenomicRanges* package (Lawrence et al. 2013) in *Bioconductor* version 3.8.

Results

DNA sequencing and genotyping

We identified 7,261 polymorphic RAD loci that were sequenced at a minimum read depth of 10 sequences, had a minor allele frequency that exceeded 0.05 over all samples, and mapped to the Atlantic herring genome. Average sequencing depth per RAD locus for each individual was 75X, and the majority of loci were biallelic (Supplemental Figure 2-1). A total of 1,104 herring were characterized by less than 10% missing data and were retained in the final data set. After filtering the data, sample sizes at each geographic location ranged from 16 to 48 individuals (mean $N = 40$ individuals) and the maximum percentage of missing data at any one locus was 2%.

Population structure

Locus-specific estimates of F_{IS} at each sampling location were generally distributed around zero (Supplemental Figure 2-2). We identified 632 loci that deviated from HWE in any one sample collection, and fewer than 2% of loci were out of HWE in any single collection. We subsequently conducted analyses of population structure on two different data sets: the first contained all loci, while the second only contained loci that were in HWE. Patterns of population

structure were consistent across both data sets (for example, see Supplemental Figure 2-5). Thus, we conducted *bayenv2* analyses using SNPs from the full subset of RAD loci, in case deviations from HWE were caused by evolutionary forces rather than genotyping error. For the sake of brevity we report results below using all RAD loci.

We estimated a global F_{ST} of 0.014 over all populations, while pairwise population F_{ST} estimates ranged from 0 to 0.032 (Supplemental Figure 2-3). Locus-specific estimates of overall F_{ST} ranged from -0.013 to 0.458 (mean = 0.012; Supplemental Figure 2-4). We conducted a discriminant analysis of principal components using all samples (Supplemental Figure 2-5) and identified 63 principal components that summarized 15% of the variance in the data. The first discriminant axis explained 38% of the retained variance and separated herring spawning in May or June (at Cherry Point, Masset, Skidegate, and Metlakatla) from all other individuals (Supplemental Figure 2-5). In addition, May/June spawners formed distinct clusters in the discriminant analysis of principal components, based on their geographic location. The second discriminant axis summarized 19% of the retained variance and separated fish sampled along a latitudinal gradient. Subsequently, we conducted a second discriminant analysis of principal components that included only individuals spawning in March and April, as these samples represent the predominant spawning phenotype and clustered together in the initial analysis. In this second analysis, the first discriminant axis separated individuals spawning in coastal/estuarine habitats from those spawning in inlet habitats (Supplemental Figure 2-5), a pattern of genetic differentiation that was also observed in a previous study using microsatellites (Beacham et al. 2008).

We assessed the temporal stability of allele frequencies at a given spawning site across different years. Pairwise population F_{ST} estimates between temporal replicates ranged from 0 to

0.002, and were among the smallest pairwise F_{ST} values observed in the entire dataset (Supplemental Figure 2-6). This indicates that allele frequencies within a particular spawning location and season remained relatively consistent over multiple years, suggesting that these sites were used by populations that exhibited seasonal and geographic spawning fidelity across multiple generations. Although the number of temporal replicates was small, it is possible that Pacific herring exhibit homing behavior to their natal spawning site, allowing for local adaptation to productivity regimes to arise. Indeed, Pacific herring spawning is closely correlated with the spring plankton bloom, and the amount of overlap between these phenologies has been shown to affect the survival of juveniles (Schweigert et al. 2013).

There was evidence of isolation by time in our data. When all samples were considered together, we found a statistically significant correlation between linearized F_{ST} and the difference in spawning date (Supplemental Figure 2-7; slope = 1.30×10^{-4}, Mantel $r = 0.51$, Mantel $p = 0.00099$). This relationship improved, after we removed highly divergent May-June spawners from the analysis (Supplemental Figure 2-7; slope = 2.78×10^{-4}, Mantel $r = 0.70$, Mantel $p = 0.00099$). There was no correlation between the geographic and temporal distance (number of days between spawning) separating sampling locations, and so we also investigated whether a pattern of isolation by distance were present in the data. When all sample collections were included in the same isolation by distance regression, we did not find a correlation between the genetic and geographic distance separating populations (Supplemental Figure 2-8). However, there was evidence of isolation by distance within a specific spawning season (Supplemental Figure 8B; March-April spawners: slope = 3.91×10^{-6}, Mantel $r = 0.77$, Mantel $p = 0.00099$; May-June spawners: slope = 8.59×10^{-6}, Mantel $r = 0.61$, Mantel $p = 0.005$), indicating that geographic connectivity also plays a role in structuring populations.

The Bayesian analysis of genetic structure assigned the highest posterior probability to a model that included prior information on temporal connectivity (Supplemental Table 2-2; *posterior probability* = 0.96, σ^2 = 0.27), followed by one that incorporated data on both temporal and geographic connectivity between spawning sites (*posterior probability* = 0.04). All other models (null model; model using only prior information on geographic connectivity) had posterior probabilities of zero.

Loci correlated with spawn timing

There were 130 SNPs that exhibited very strong correlations with photoperiod (Bayes factor > 100; Supplemental Figure 2-9) distributed across 56 distinct genomic scaffolds in pacific herring. SNPs with the strongest associations ($\log_{10}$ Bayes factor > 10) were within 18,000 bp of each other and mapped to *SYNE2*, a gene coding for a nuclear membrane protein that influences the position of retinal photoreceptors early in vertebrate development (Tsujikawa et al. 2007, Yu et al. 2010). Additional genomic regions that were highly associated with spawn timing in Pacific herring and may play a role in reproduction include *SH2D4A*, a locus with estrogen receptor 1 binding activity (Tingting et al. 2009) and *SPATA4*, a locus associated with spermatogenesis (Xie et al. 2007). Only 4 out of the 130 SNPs correlated with photoperiod in Pacific herring were out of HWE in any sample collection, and none of them mapped to the aforementioned genes.

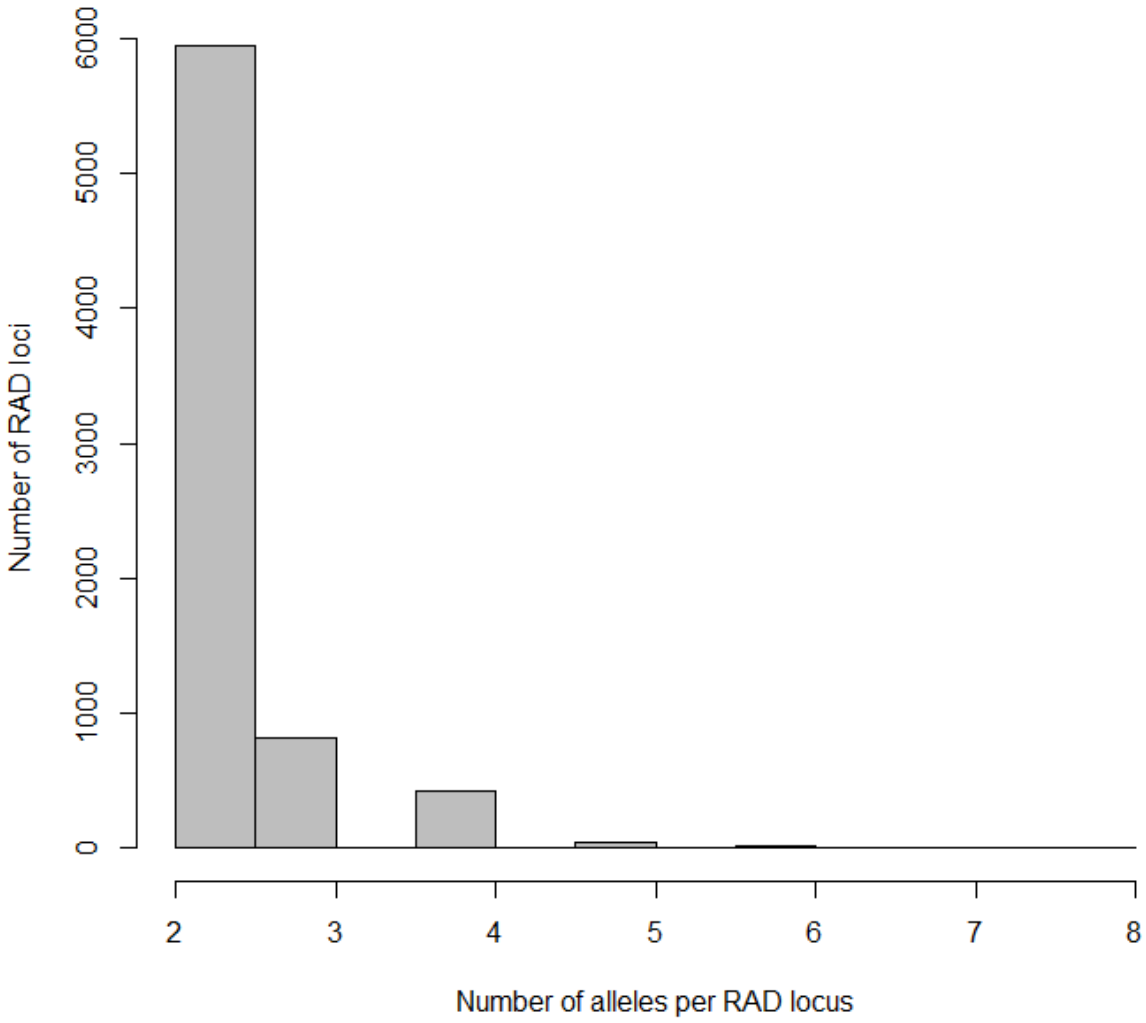
Supplemental tables and figures

Supplemental Table 2-1. Geographic location and date of each sample collection.

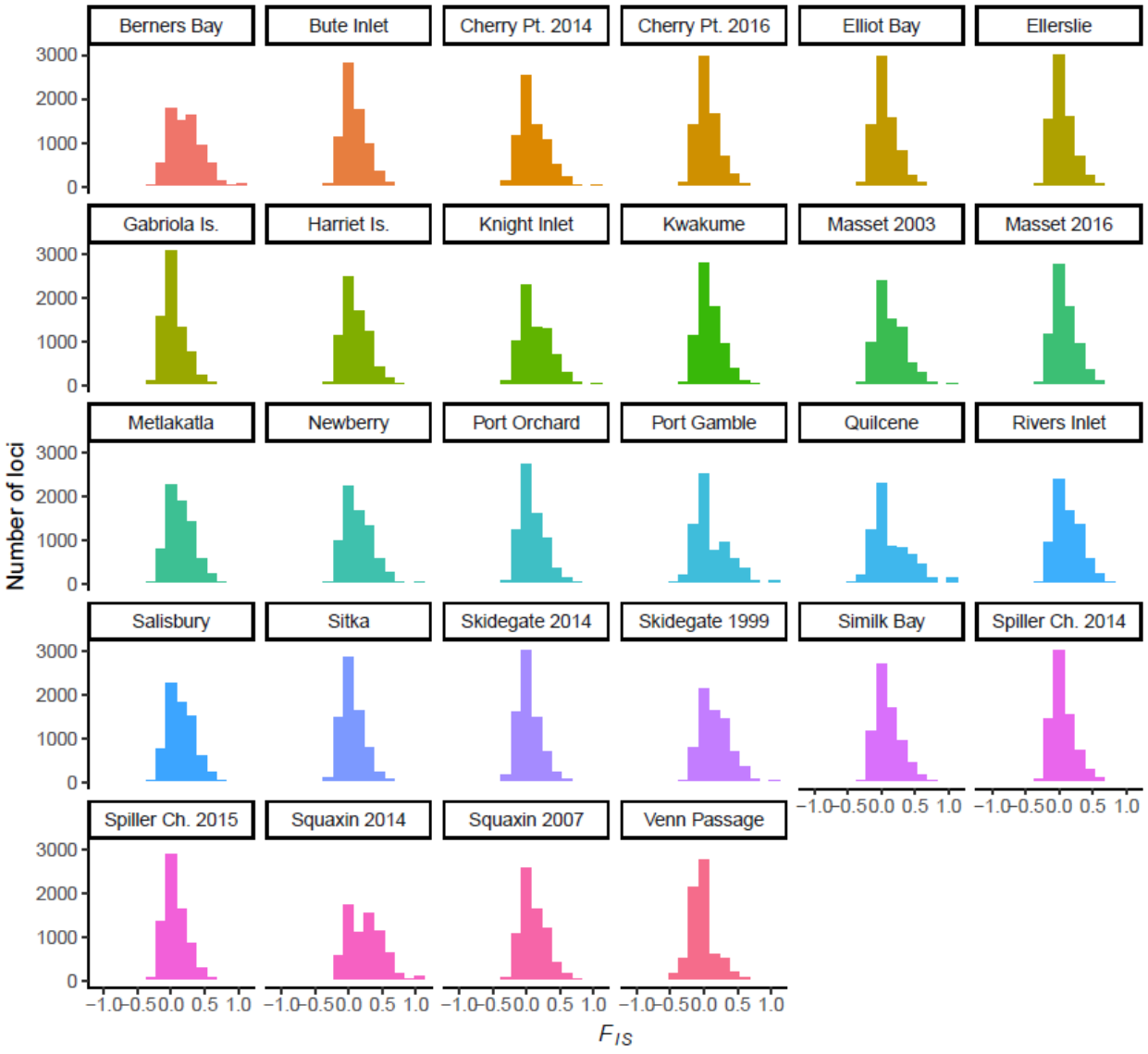
Location name	State/Province	Subregion	Latitude	Longitude	Sampling date	*N*
Squaxin 2007	Washington	Salish Sea	47.20	-122.94	1/17/2007	41
Squaxin 2014	Washington	Salish Sea	47.20	-122.94	2/24/2014	27
Elliott Bay	Washington	Salish Sea	47.62	-122.36	4/15/2015	45
Port Orchard	Washington	Salish Sea	47.69	-122.59	1/27/2014	42
Quilcene Bay	Washington	Salish Sea	47.80	-122.85	3/26/2014	16
Port Gamble	Washington	Salish Sea	47.87	-122.60	3/12/2014	18
Similk Bay	Washington	Salish Sea	48.44	-122.57	2/25/2015	48
Cherry Point 2014	Washington	Salish Sea	48.93	-122.80	5/12/2014	30
Cherry Point 2016	Washington	Salish Sea	48.93	-122.80	5/9/2016	48
Gabriola Island	British Columbia	Salish Sea	49.16	-123.74	3/19/2015	45
Bute Inlet	British Columbia	BC inlet	50.81	-124.90	3/3/2004	47
Knight Inlet	British Columbia	BC inlet	51.02	-125.60	3/20/2003	28
Rivers Inlet	British Columbia	BC inlet	51.68	-127.26	3/11/2001	43
Kwakume	British Columbia	Central Coast	51.70	-127.88	4/3/2015	47
Harriet Island	British Columbia	Haida Gwaii	52.31	-131.23	4/2/2014	44
Spiller Channel 2014	British Columbia	Central Coast	52.36	-128.20	4/4/2014	47
Spiller Channel 2015	British Columbia	Central Coast	52.36	-128.20	4/3/2015	47
Newberry Cove	British Columbia	Haida Gwaii	52.47	-131.44	4/6/2014	36
Ellerslie Lagoon	British Columbia	Central Coast	52.51	-128.06	3/25/2015	48
Skidegate 2014	British Columbia	Haida Gwaii	53.24	-132.01	6/4/2014	47
Skidegate 1999	British Columbia	Haida Gwaii	53.24	-132.01	5/7/1999	36
Masset 2003	British Columbia	Haida Gwaii	53.67	-132.35	6/26/2003	34
Masset 2016	British Columbia	Haida Gwaii	53.67	-132.35	6/9/2016	47
Venn Passage	British Columbia	Prince Rupert	54.33	-130.42	3/25/1999	23
Metlakatla	British Columbia	Prince Rupert	54.34	-130.44	5/23/2002	45
Sitka	Alaska	SE Alaska	57.12	-135.45	4/13/2017	46
Salisbury Sound	Alaska	SE Alaska	57.30	-135.61	4/5/2016	46
Berners Bay	Alaska	SE Alaska	58.68	-134.99	5/3/2016	33

Supplemental Table 2-2. Estimated posterior model probabilities from a hierarchical Bayesian analysis of population structure.

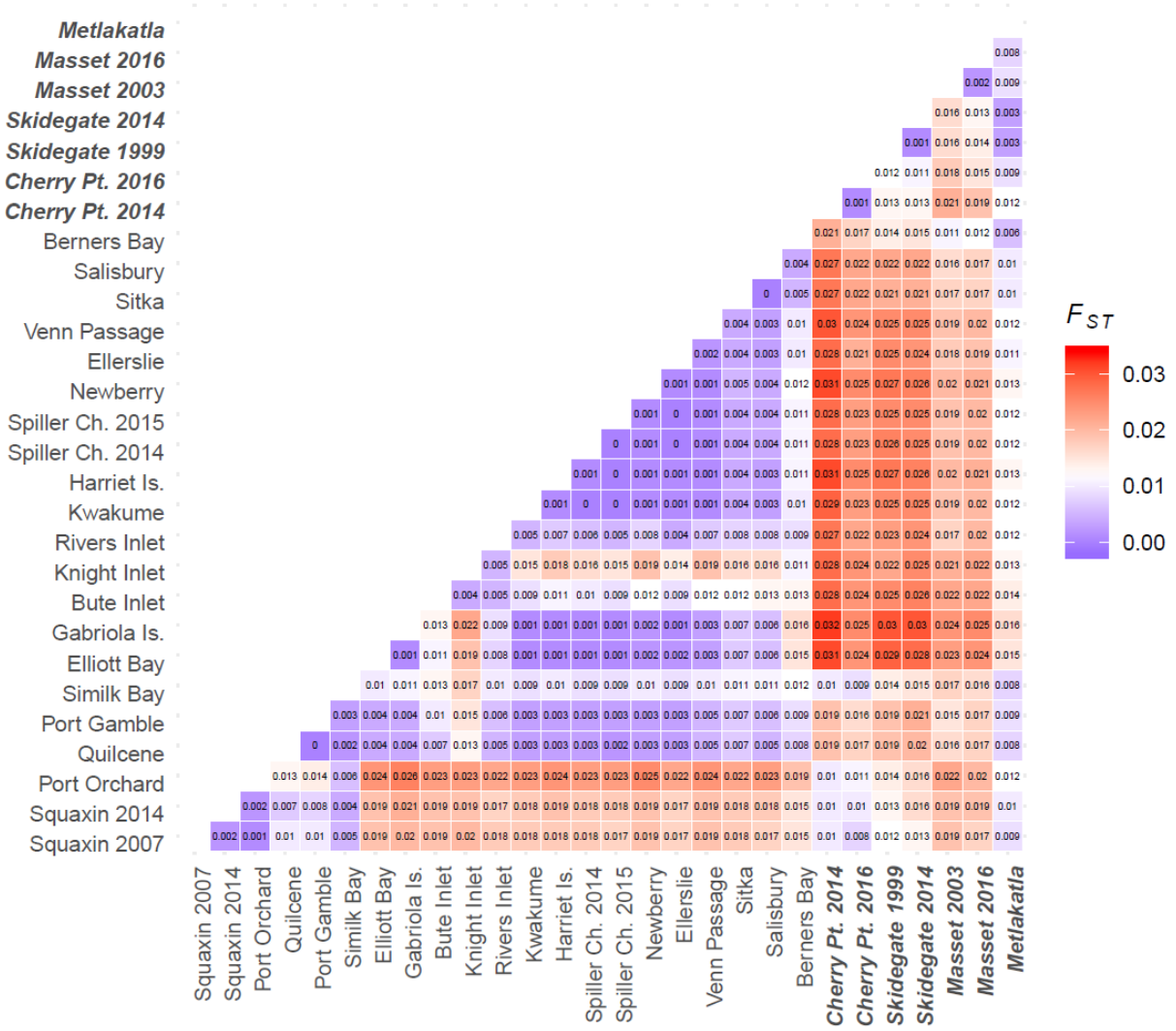
Model	Factors included	Posterior probability
1	Constant	0
2	Constant, geographic distance	0
3	Constant, temporal distance	0.96
4	Constant, geographic distance, temporal distance	0.04
5	Constant, geographic distance, temporal distance, interaction between temporal and geographic distances	0



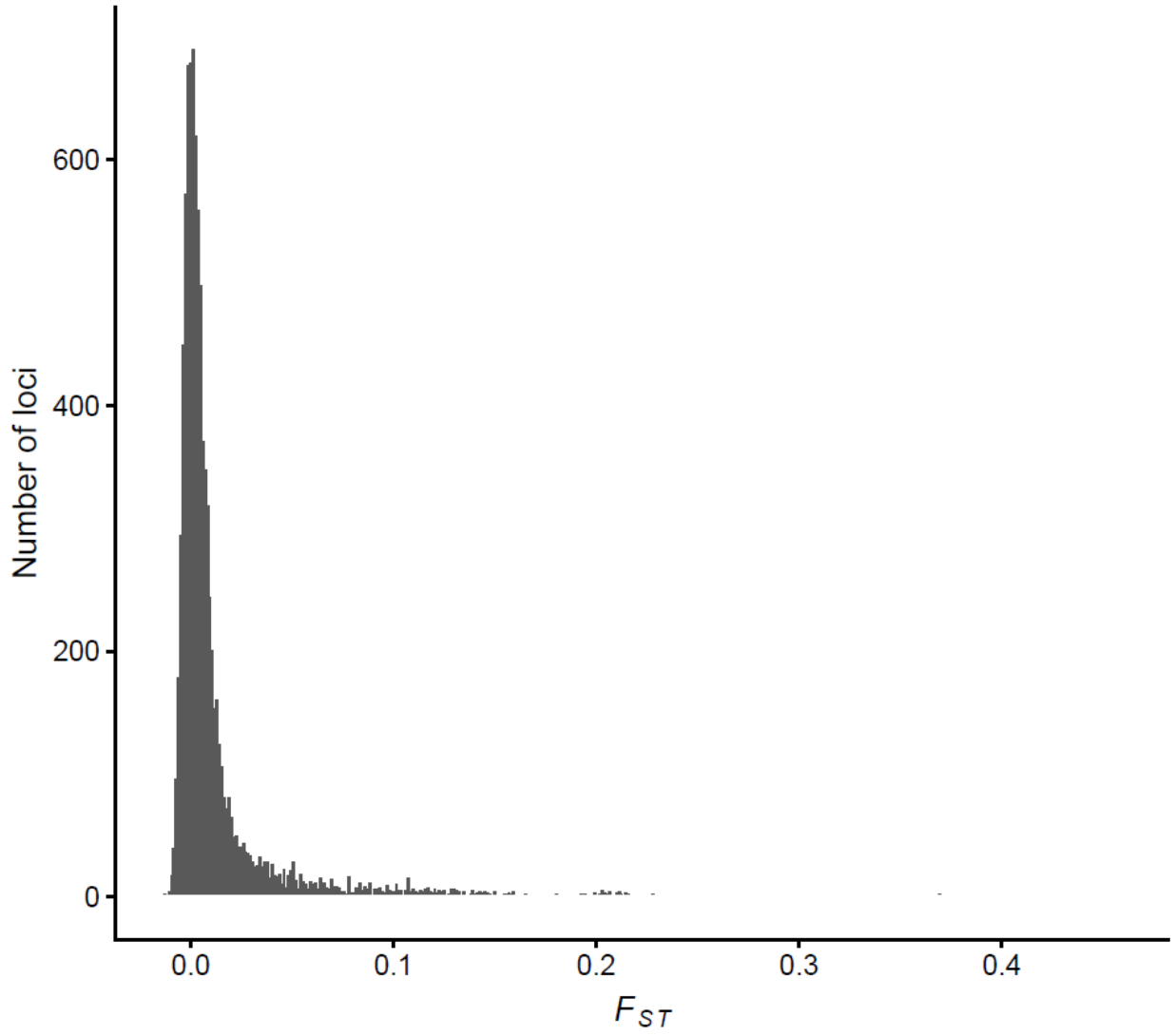
Supplemental Figure 2-1. Distribution of the number of alleles per RAD locus.



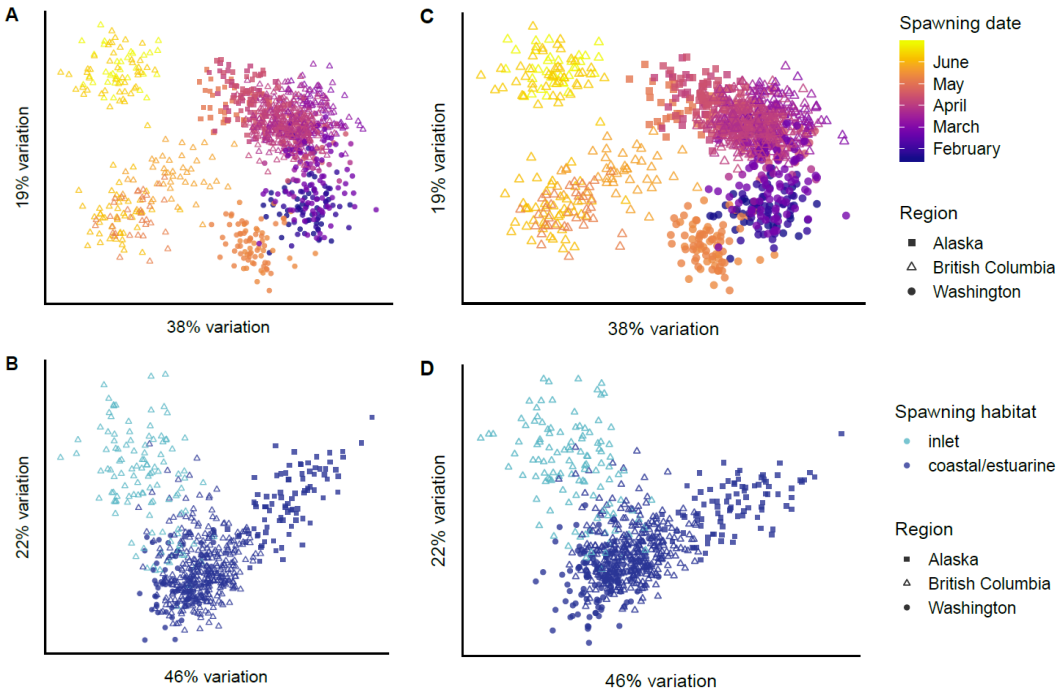
Supplemental Figure 2-2. Distribution of per-locus F_{IS} estimated for each sampling location.



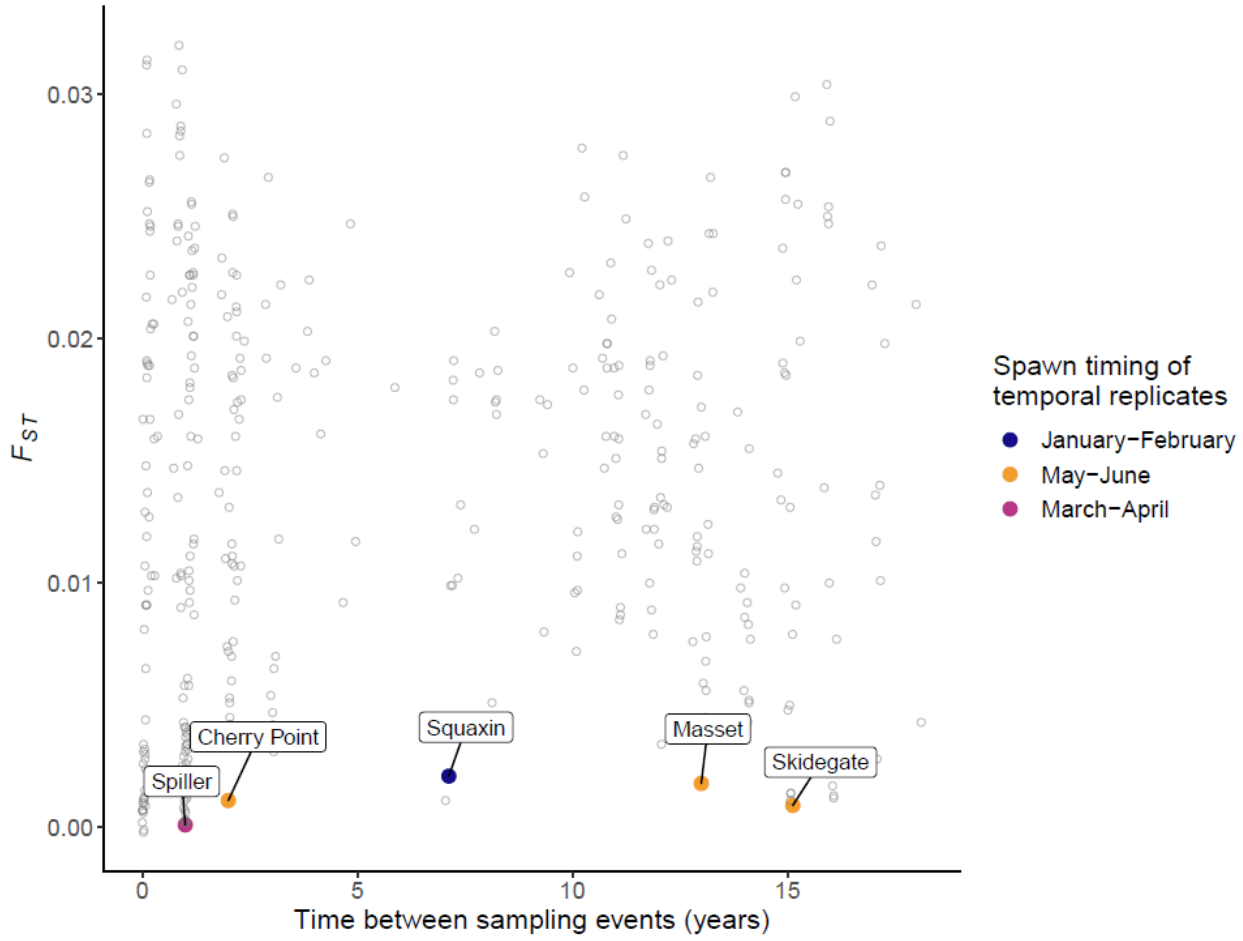
Supplemental Figure 2-3. Pairwise F_{ST} estimated between sampling locations.



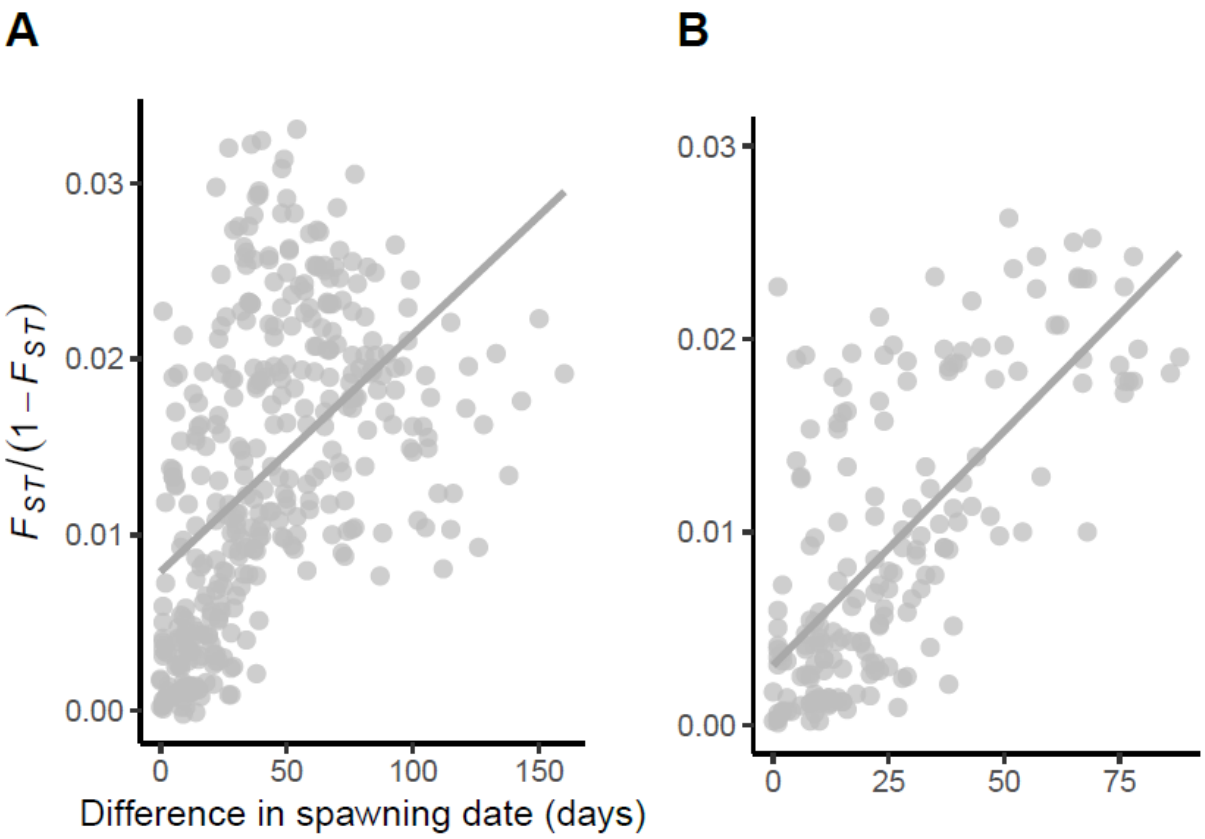
Supplemental Figure 2-4. Histogram showing the distribution of per-locus overall F_{ST} over all samples.



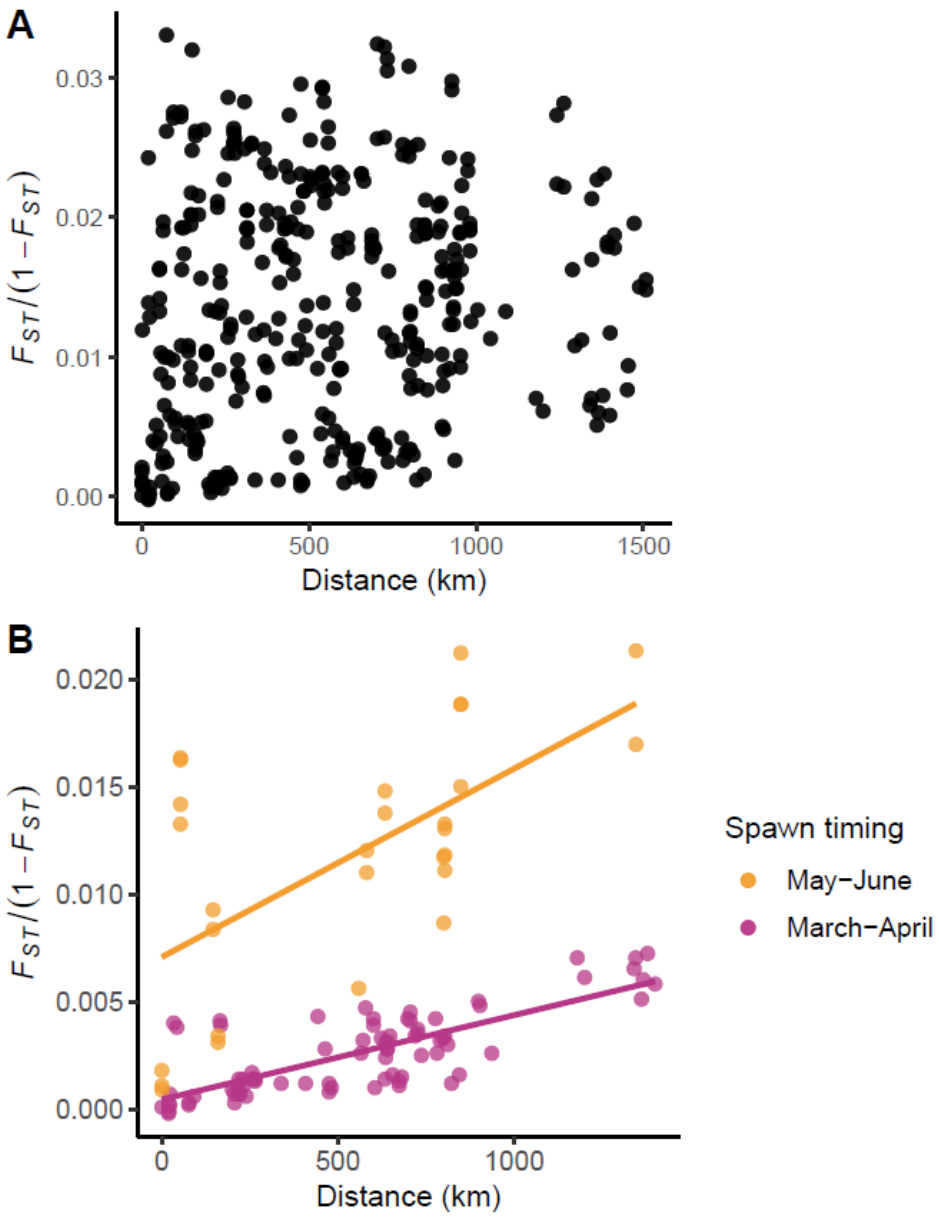
Supplemental Figure 2-5. Discriminant analysis of principal components (DAPC). A) DAPC using all samples and all loci. Different colors indicate the Julian day of sampling while shapes represent the geographic region that a sample was collected from. B) DAPC using only individuals that spawned in March and April and all loci. Inlet spawners (from Bute Inlet, Knight Inlet, or Rivers Inlet) are in light blue while other samples are in dark blue. C) DAPC using all samples and loci in HWE. D) DAPC using only individuals that spawned in March and April and loci in HWE.



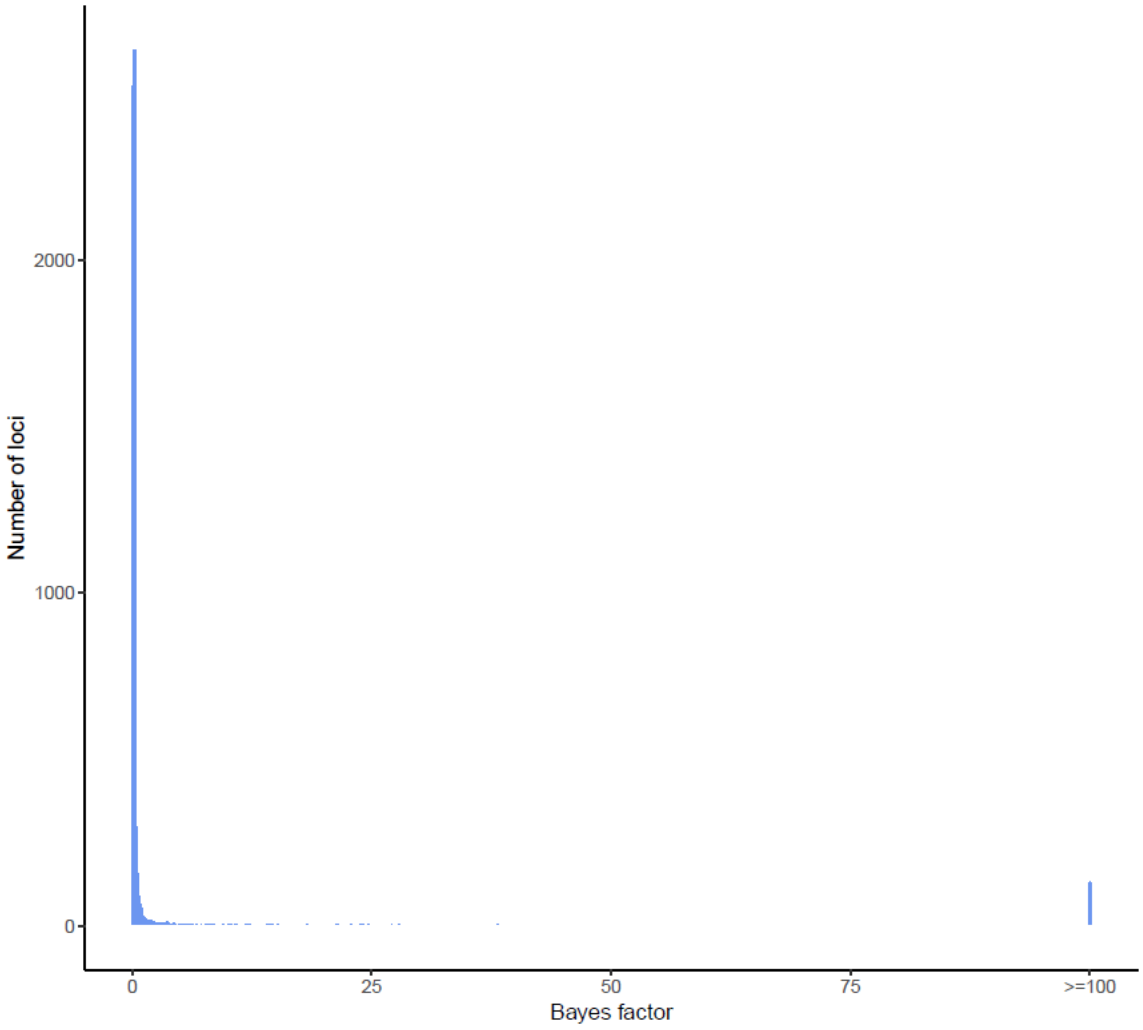
Supplemental Figure 2-6. Pairwise F_{ST} among sample collections. Temporal replicates are highlighted in color and labeled with their respective geographic location. Pairwise comparisons among all other sample collections are represented by open circles.



Supplemental Figure 2-7. The influence of spawn timing on genetic differentiation (isolation by time). Each circle represents an estimate of pairwise F_{ST} between two sampling locations. Panels A and B show linear regressions of pairwise F_{ST} to the number of days separating sampling events at each location, using different subsets of the data: A) Isolation by time using all samples (slope = 1.30×10^{-4}, Mantel $r = 0.51$, Mantel $p = 0.00099$); B) Isolation by time when highly differentiated May-June spawners are removed (slope = 2.78×10^{-4}, Mantel $r = 0.70$, Mantel $p = 0.00099$).



Supplemental Figure 2-8. The influence of geographic distance on genetic differentiation (isolation by distance). Analysis using A) pairwise comparisons among all sample collections, and B) pairwise comparisons among sample collections within a particular spawning season: March-April spawners (pink) and May-June spawners (orange).



Supplemental Figure 2-9. Distribution of Bayes factors, indicating strength of association between allele frequency and photoperiod in Pacific herring.

Chapter 3. **Ancient DNA reveals harvest of winter spawning herring populations by Coast Salish fisheries over 900 years**

INTRODUCTION

There is growing recognition that phenotypic and genetic diversity contribute to the productivity and resilience of wild populations, and the ecosystem goods that they provide (Luck et al. 2003). For example, diverse life-histories in salmonids are associated with asynchronous population dynamics which stabilize overall abundance (Hilborn et al. 2003) through portfolio effects (Schindler et al. 2010, Moore et al. 2014). Meanwhile, genetic variation between and within wild populations has been shown to contribute to disease resistance (Hofinger et al. 2011, Savage and Zamudio 2011), and represents the evolutionary potential of populations to adapt to new environmental conditions (Colosimo et al. 2005, Jones et al. 2012, Barrett et al. 2019). Despite the recognized value of intraspecific variation, population diversity is eroding at a rapid rate globally due to human activities (Hughes et al. 1997, Ceballos et al. 2017).

The extent to which different populations or phenotypes contribute to ecosystem goods over long time scales is mostly unknown, as ecological studies typically collect data over the course of years or decades. This limited temporal data can result in the collective memory loss of past biodiversity and ecosystem states, a phenomenon known as "shifting baselines syndrome" (Pauly 1995). One potential method to gather information over extended temporal scales is to quantify faunal remains in natural or human-associated (archaeological) deposits. Although coarser in temporal resolution than modern ecological studies (Erlandson and Rick 2009), these types of data can be compared to contemporary datasets and supply information about long-term ecological conditions (Hadly and Barnosky 2009, Wolverton and Lyman 2012) and spatially-explicit patterns of variability (Wright et al. 2005, Rogers et al. 2013, McClatchie et al. 2017).

Additional layers of information can be provided by analyzing the ancient DNA (aDNA) preserved in faunal remains; when different populations or phenotypes are genetically differentiated from each other, it may be possible to use aDNA to identify their presence and relative abundance. Additionally, aDNA preserved in archaeological sites may archive the genetic and phenotypic diversity of populations that were harvested by human communities over long periods of time. For example, recent studies of Pacific salmon have revealed the migration phenotypes that supported ancient fisheries (Halffman et al. 2015, Thompson et al. 2019) and quantified reductions in genetic diversity associated with the large-scale habitat degradation of riverine systems in the 20th century (Johnson et al. 2018, Thompson et al. 2019). In marine systems, ancient DNA has been used to identify cod populations harvested during the Viking Age and demonstrated the early development of trade routes for marine resources from the North Atlantic Ocean (Star et al. 2017, Star et al. 2018).

Temporal studies of population diversity are particularly important in forage fish, as they are foundational to coastal food webs and fisheries. Pacific herring (*Clupea pallasii*) are an abundant forage fish in coastal waters of the Pacific Northwest and are important prey for a variety of organisms, including marine mammals (Thomas and Thorne 2001, Womble et al. 2005), salmon (Chamberlin et al. 2017), and sea birds (Thomas and Thorne 2003, Schrimpf et al. 2012, Hatch 2013). The reproductive migrations of herring provide an ephemeral pulse of food for coastal marine organisms (Willson and Womble 2006), as these fish aggregate to nearshore environments to spawn. The timing of herring reproduction varies across broad latitudinal gradients, such that populations in the southern part of the species range typically spawn in winter, while northerly populations reproduce in spring (Haegele and Schweigert 1985). Populations with different spawn timing are connected by limited gene flow (this dissertation,

second chapter). Additionally, scattered across the species range there are also reproductively isolated populations with unusually late spawning in May or June (Small et al. 2005, Beacham et al. 2008; Chapter 2, this dissertation); we hereinafter refer to these populations as "late spawners". This genetic and phenotypic diversity in reproductive timing gives mobile consumers the opportunity to access herring in the nearshore environment over an extended period of time, a phenomenon known as a resource wave (Lok et al. 2012a, Armstrong et al. 2016).

Resource waves may have also been important to the livelihoods of human communities. On the Pacific coast of North America, humans have harvested herring for millennia, as demonstrated by specimens found in archaeological sites that are dated to ~10,000 years before present (ybp) (Moss et al. 2015). Archaeological studies show that herring bones are among the most abundant fish remains found in ancient settlements (Kopperl 2001, McKechnie et al. 2014), indicating that herring were a critical food resource. Artifacts such as herring rakes and fish traps (Caldwell 2011) attest to past harvests, and ethnographic studies from the Puget Sound in the early 20th century contain lively descriptions of their use: *"A man and his wife usually worked together in getting herring with this implement. The woman sat in the stern of the canoe and paddled, while the man sat in the bow facing her. When they got in the midst of a school of herring, he took the "rake" and swept the surface of the water, moving it in a wide semi-circle toward the stern of the boat and then giving it a jerk or fling. Herring which became impaled in the points were thus thrown into the canoe. Informants have told me that by these methods it was sometimes possible to fill a boat with herring in an hour."* (Waterman 1973). Accounts from the late 1800's indicate that herring were eaten fresh or cured (e.g. by the Twana people; Eells and Castile 1985), while their eggs were harvested using evergreen branches or kelp fronds that were submerged during spawning events (e.g. by the Puyallup and Nisqually people; Smith 1940).

Today, herring eggs harvested in this way are a beloved food item in some coastal communities (e.g. Thornton 2015) and herring harvest is considered an expression of indigenous culture and sovereignty (Powell 2012, Gauvreau et al. 2017), while income generated from herring fisheries supports coastal livelihoods (Powell 2012).

Industrial commercial fisheries for herring began with the colonization of the Pacific Northwest coast in the late 19th century by European and Euro-American settlers, soon followed by reports of declining catches. A letter published in the American Angler in 1886 states, *"From 1869 to 1877 it was not an uncommon occurrence for us to catch from 200 to 300 barrels of herring in a night, but since 1877 they have been growing less in number, until now the largest night's work is about 20 barrels. This is a great falling off, and it is much the same way with all other fish in the [Puget] Sound*" (Hammond 1886). In the 20th century, large reduction fisheries harvested tens of thousands of tons of herring yearly throughout the Pacific Northwest (Carrothers 1941), but collapsing populations in British Columbia (Pearse 1982) and Washington State (Bargmann 1998) led to their closure in the latter part of the century. Today, Alaska (Hebert 2017) and British Columbia (Martell et al. 2012) support significant sac roe fisheries for herring, but Washington State only allows a relatively small bait fishery to operate in its waters (Stick et al. 2014).

It has been approximately 30 years since the last large commercial herring fisheries operated in Puget Sound, Washington (Stick and Lindquist 2009). Despite this, an approximately 50-year time series of spawning biomass surveys indicates that certain populations of herring have not recovered to their former population abundance and remain severely depressed (Stick et al. 2014). For example, the late spawning herring population reproducing at Cherry Point was once the largest stock in Puget Sound, but has declined to a small fraction of its former

abundance (Figure 2). It is unclear if the dramatic decline of this population is representative of natural variations in population productivity or if this pattern is unique to recent decades. Furthermore, it is unknown whether different spawning phenotypes display fidelity to specific geographic locations or if their geographic range fluctuates over long time scales.

Ancient DNA preserved in herring bones contains information about the genetic and phenotypic diversity of herring populations over long time scales. Recent studies (Speller et al. 2012, Star et al. 2017, Rodrigues et al. 2018, Thompson et al. 2019) have demonstrated that aDNA can be extracted from fish bones, even when they are several thousand years old and weigh less than 10 mg. Fish remains are common in archaeological assemblages but there are few population genetic studies of marine fish using ancient specimens because these organisms are generally characterized by subtle population structure that is only resolved through the genotyping of nuclear DNA (Speller et al. 2012). An added complication is that ancient samples typically contain very low quantities of endogenous nuclear DNA that is highly fragmented (Hofreiter et al. 2001).

In this study, we use ancient nuclear DNA from herring bones deposited in archaeological sites in the Puget Sound to investigate the relative contributions of genetically distinct populations of Pacific herring to food supplies over the last millennium. These genetically distinct herring populations spawn at different times of year, and we hereinafter refer to them as winter spawners and late spawners. Specifically, we ask the following questions: (i) Did Coast Salish people in the central Puget Sound harvest winter spawning or late spawning herring? (ii) Did the relative abundance of these herring populations in the archaeological record change through time? iii) Did the relative abundance of winter spawners and late spawners vary based on the geographic location of an archaeological assemblage?

METHODS

Description of archaeological sites

Ancient herring bones were previously excavated from two archaeological sites in the Puget Sound (Lewarch et al. 2002, Stein 2002) and housed at the Burke Museum at the University of Washington. We received permission from the Puyallup Tribe of Indians to use herring bones retrieved from the Burton Acres Shell Midden (45KI437), and from the Suquamish Tribe to use bones excavated from the Bay Street Shell Midden (45KP115).

The Burton Acres Shell Midden is located on Vashon Island, WA (Figure 3-1). Archaeological evidence suggests that the site was first inhabited approximately 1,000 years before present (ybp) and its "deposits of shell, charred plants, modified rocks, animal bones, and broken artifacts indicate that people caught and preserved fish and shellfish at this site" (Stein 2002). Pacific herring are the most abundant fish in the assemblage, contributing up to 90% of fish remains in certain layers (Figure 3-3; Kopperl 2001), while salmonids are the next most abundant group. The midden's upper layers contain metal, demonstrating that the site was used during the last two centuries (Stein 2002), a finding also corroborated by historic documents noting that Vashon Island was inhabited by Southern Coast Salish people (De Danaan 2002).

The Bay Street Shell Midden, first occupied between 800 and 500 ybp, is located in Port Orchard, WA (Figure 3-1). Archaeological evidence supports that the site was used periodically over an 800-year period, while more recent ethnographic records indicate that it was used seasonally by Suquamish people (Lewarch et al. 2002). The discovery of bone and antler tools at the site demonstrates that its inhabitants hunted and fished. Pacific herring account for one-third of identifiable fish remains in the total assemblage, indicating that they were prominent in the

diet (Figure 3-3; Lewarch et al. 2002). Other common fish in the assemblage include flatfish, salmonids, and spiny dogfish.

DNA extraction of archaeological samples

Herring bones (*N* = 192) were morphologically identified by zooarchaeologist Robert Kopperl and selected from two different temporal layers at each archaeological assemblage (Table 3-2). These bones were vertebrae or prootic elements whose mass ranged from 3 to 26 mg. When possible, bones were collected from a variety of different archaeological subsamples within each layer, to minimize the chance of sampling the same individual multiple times. To establish the approximate age of these samples, a small piece of wood was extracted from each temporal layer and radiocarbon dated by a commercial laboratory (Direct AMS, Washington State).

All sample preparations and aDNA extractions were conducted in a dedicated aDNA laboratory at Simon Fraser University, Canada, following strict protocols to minimize contamination with modern DNA (Yang and Watt 2005). Surface contamination was removed from each bone by immersing it in a 6% sodium hypochlorite solution for five minutes. Samples were subsequently rinsed twice with ultra-pure water and each side of bone was exposed to UV light for 15 minutes. We extracted DNA from each sample following the protocol described in Yang et al. (2008). In brief, samples were incubated overnight at 50 °C in 3 mL of proteinase K lysis buffer, and were subsequently concentrated to approximately 100 µL using spin columns (Amicon 15kDA Centrifugal Filter Units). DNA extracts were purified using Qiagen MinElute columns following the manufacturer's instructions, and DNA was eluted in 50 µL of warmed (56 °C) Buffer EB. To test for contamination during DNA extraction, blank controls were included in each extraction and processed during all subsequent steps.

Development and testing of SNP assays on modern samples

Samples were taken from spawning adult herring or eyed eggs (N = 347) collected at eight different spawning sites in the Salish Sea (Figure 3-1), and prepared for restriction-site associated (RAD) sequencing as described in Petrou et al. (2018). The collection and processing of these RAD sequences are described in the second chapter of this dissertation, and resulted in the identification of 7,261 polymorphic RAD loci which mapped to the Atlantic herring genome (Martinez Barrio et al. 2016).

We identified SNPs showing high differentiation between populations with distinct spawn timing by estimating per-locus F_{ST}s with the R package *hierfstat* (Goudet 2005) and conducting a discriminant analysis of principal components (DAPC) with the R package *adegenet* (Jombart 2008). Loci were ranked by their loadings along the first and second discriminant axes of the DAPC, as these axes separate winter spawners from late spawners. Loci were subsequently retained in the analysis if they had a global $F_{ST} > 0.1$, mapped to distinct scaffolds on the Atlantic herring genome, contained a maximum of 2 SNPs per RAD sequence, and displayed stable allele frequencies across years at the same sampling location. As the amount of DNA template available in ancient samples is limited, we selected small numbers of high F_{ST} loci (N= 5-8) and tested them for deviations from Hardy-Weinberg equilibrium (HWE) in each population using the exact test with 10,000 Monte Carlo permutations of alleles, as implemented in *Genepop* version 4 (Rousset 2008). We also examined patterns of linkage disequilibria between each locus pair and within populations using exact tests in *Genepop*.

Subsequently, we assessed the ability of these loci to assign individual herring to a spawning group (winter spawners vs. late spawners).We simulated individual assignment using the “leave-one-out” approach and the Bayesian method of Rannala and Mountain (1997)

implemented in *Geneclass2* (Piry et al. 2004). An individual was assigned to a spawning group if the sum of assignment probabilities across populations in a spawning group exceeded an 85% probability threshold, otherwise that individual was considered unassigned. This analysis was applied to modern herring samples using the genotypes generated by RAD sequencing, and a set of seven loci (Table 3-3, Figure 3-4) was identified that could assign individuals to their spawning group (winter spawners vs. late spawners) with a high success rate (99%).

We also evaluated whether these seven loci could be used to estimate the proportion of populations in a mixed-stock fishery, using the Bayesian method of Moran and Anderson (2018) that is implemented in the R package *rubias*. In brief, this approach uses Markov chain Monte Carlo (MCMC) to estimate the proportion of individuals in a mixture that originate from different reference populations or aggregates of reference populations (reporting groups), given genotypic data at those reference populations. We assessed the predicted accuracy of mixed stock analysis by simulating multiple mixture data sets of known mixture proportions using modern herring samples (simulated mixture size = 48 individuals; number of repetitions of mixture simulation and MCMC = 50) and comparing mixture estimates with simulated mixture proportions. Individual genotypes in the simulated mixtures were generated by sampling from the allele frequency distribution of a reference population, following the "leave one out" method (Moran and Anderson 2018). We also conducted "100% simulations", where all of the simulated individuals were generated from the allele frequency distribution of a single reference population. We analyzed the simulated data using two reporting groups (winter spawners vs. late spawners) and three reporting groups (January-February spawners vs. March-April spawners vs. May-June spawners). The modern herring populations belonging to each reporting group are summarized in Table 3-1.

Additionally, we tested for population structure between and within spawning groups (winter spawners vs. late spawners) by conducting hierarchical analyses of molecular variance (AMOVAs) using all 7,261 RAD loci and the subset of seven loci. AMOVAs were done using the R packages *ade4* (Dray and Dufour 2007) and *poppr* (Kamvar et al. 2014), and statistical significance was estimated using 1,000 permutations of sample collections across spawning groups. Pairwise F_{ST} (Weir and Cockerham 1984a) between modern sample collections was calculated in *hierfstat*, and its significance was estimated using 1,000 permutations in the R package *strataG* (Archer et al. 2017).

Custom TaqMan™ assays (Thermo Fisher Scientific, Waltham, MA) were developed to genotype SNPs at these seven loci. In brief, we submitted approximately 100 bp of DNA sequence around each SNP to the Custom TaqMan Assay Design Tool (Applied Biosystems) to design primers and probes (Table 3-3).

SNP genotyping of ancient samples

In order to increase the amount of template aDNA available for TaqMan genotyping reactions, we first conducted a preamplification PCR with all primers following the protocol of Smith et al. (2011). These reactions were set up in a dedicated clean room at the University of Washington. Preamplification reactions were conducted in 24 µL volumes containing Qiagen Multiplex PCR Master Mix, 0.2 µM of each forward and reverse SNP primer, ultra-pure water, and 4 µL of template aDNA. Thermal cycling was performed on a Bio-Rad C1000 Touch (Hercules, CA), using these conditions: initial denaturation at 95 °C for 15 minutes, followed by 14 cycles of 94 °C for 30 seconds, 57 °C for 90 seconds, 72 °C for 60 seconds, and a final extension of 72 °C for 10 minutes. Negative controls were included with each preamplification reaction.

We diluted these preamplification PCR products 1:3 for use in subsequent TaqMan genotyping reactions. All genotyping reactions took place in 12 µL volumes containing 1X TaqMan Universal PCR Master Mix, 1X TaqMan assay, nuclease-free water, and 2 uL of template DNA. Thermal cycling was performed on an Applied Biosystems7900HT Fast Real-Time PCR system (Foster City, CA) as follows: initial denaturation at 95 ℃ for 10 minutes, followed by 60 cycles of 95 °C for 15 seconds and 60 ℃ for 60 seconds. Following quality control procedures established in Speller et al. (2012), each sample was genotyped twice and negative controls were included with every genotyping reaction. We estimated the genotyping discrepancy rate by dividing the total number of genotype mismatches by the total number of samples. Ancient samples that were missing genotypes at more than one locus were filtered from the data set.

Genetic stock identification of ancient samples

Pairwise F_{ST} (Weir and Cockerham 1984a) between ancient and modern sample collections genotyped at seven nuclear loci was calculated in *hierfstat*, and its significance was estimated using 1,000 permutations in the R package *strataG* (Archer et al. 2017). Patterns of genetic differentiation in the ancient and modern samples were visualized using a PCA conducted with the R package *adegenet*.

We used two different genetic stock identification approaches to estimate the contribution of distinct herring populations to ancient food supplies. First, we assigned individual ancient herring samples to a modern spawning group (winter spawners vs. late spawners) using the method of Rannala and Mountain (1997) that is implemented in *Geneclass2.* Secondly, we conducted a mixed stock analysis to estimate the relative proportions of different spawning groups in the archaeological samples using the Bayesian method described in Moran and

Anderson (2018) and implemented in the R package *rubias*. Modern herring samples were designated as the reference populations, and each temporal layer at an archaeological site was analyzed as a separate fisheries mixture. We conducted the mixed stock analysis (number of MCMC iterations = 10,000, burn-in steps = 1,000) using two reporting groups (winter spawners vs. late spawners) and three reporting groups (January-February spawners vs. March-April spawners vs. May-June spawners).

RESULTS

Modern samples

We identified seven loci that could assign individual modern samples to their spawning group (winter vs. late spawners) with a high success rate (99%). Five of these loci were within genes (*THSb, CADPS, SYNE2, NFATC2*, and *GRB2*) and two were in intergenic regions (Supplemental Table 3). At these seven loci, the number of genotyped modern samples ranged from 14 to 48 individuals in each population, while observed heterozygosities in each population ranged from 0.00 to 0.62. Per-locus estimates of F_{ST} at the seven loci ranged from 0.11 to 0.37, which is much higher than the average per-locus F_{ST} using all 7,261 RAD loci (mean per-locus F_{ST} = 0.009). AMOVAs using all 7,261 loci and the subset of seven loci both detected significant population structure between winter spawning and late spawning herring, and the estimated variance between spawning groups was much larger when the AMOVA was conducted using the subset of seven loci (Table 3-4).

Using the subset of seven loci, we found that four out of 64 (6%) tests for HWE across populations were statistically significant ($\alpha = 0.05$) due to heterozygote deficiency but these deviations were not concentrated within particular loci or samples. Additionally, we found that

12 out of 189 (5%) tests for linkage disequilibrium were statistically significant ($\alpha = 0.05$). Pairwise F_{ST} between modern sample collections ranged from -0.003 to 0.48.

The seven loci successfully assigned modern samples to their known spawning group of origin (winter vs. late spawners) using an individual assignment approach. Out of 347 modern herring samples, 11 (3%) failed to assign to a spawning group using an 85% probability threshold (Table 3-5). Of the remaining modern samples (N = 334), only three assigned to a spawning group that did not match their known group of origin, indicating a very high assignment success rate (>99%).

We also evaluated whether these seven loci could be used to estimate the proportion of distinct populations in a mixed stock fishery using mixed stock analysis. With two reporting groups (winter spawners vs. late spawners), the predicted accuracy of the genetic stock identification was very high (Figure 3-6); simulated mixtures originating from a single reporting group (100% simulations) resulted in mean estimate of mixture proportions for late spawners of 98%, and for winter spawners of 99% (Figure 3-6) and correlations between simulated mixtures and resulting estimates were high (r^2= 0.92). With three reporting groups (January-February spawners vs. March-April spawners vs. May-June spawners), the predicted accuracy of the genetic stock identification was lowest for early winter spawners and highest for spring spawners (Figure 3-6). In this case, the predicted accuracy of mixture analysis in 100% simulations ranged from 83% to 98% and correlations between simulated mixture proportion and estimates from r^2=0.64 to r^2=0.87 (Figure 3-6).

Ancient samples

We were able to successfully genotype 83% of archaeological samples (N = 159) at six or more SNP loci. Negative controls did not amplify in any genotyping reaction and we did not

observe any genotype mismatches between repeated genotyping of ancient samples. The number of genotyped ancient samples at each locus, archaeological assemblage, and temporal layer ranged from 46 to 30 individuals, while observed heterozygosities ranged from 0 to 0.50. The number of tests with statistically significant deviations from HWE ($\alpha = 0.05$) ranged from one to four per archaeological site and layer, and the most deviations from HWE ($N = 4$) were observed in samples collected from the older Burton Acres temporal layer (915-680 ybp). Overall, six out of 42 (14%) tests for linkage disequilibrium were statistically significant ($\alpha = 0.05$), and these only occurred in the Burton Acres assemblage (three significant tests per temporal layer).

We visualized patterns of genetic differentiation in the ancient and modern samples using PCA and this analysis showed that most ancient samples clustered with contemporary herring populations that spawn in winter. Pairwise F_{ST} between ancient and modern sample collections ranged from -0.002 to 0.416, while pairwise F_{ST} between the two ancient sites ranged from 0.003 to 0.046.

Individual assignment of the ancient samples to contemporary herring populations showed that most ancient samples assigned to the winter spawning group ($N = 140$), while a small number ($N = 4$) were assigned to late spawners (Table 5) using an 85% probability threshold. The Burton Acres assemblage dated to 915-680 ybp contained the largest number of late spawning herring ($N = 3$). Fifteen samples failed to assign to either spawning group using an 85% probability threshold.

A mixed stock analysis using two reporting groups (winter vs. late spawners) yielded similar patterns; late spawners constituted $\leq 1\%$ of total samples across most sites and layers, except for the Burton Acres assemblage dated to 915-680 ybp (16%, Figure 3-7). Using three reporting groups for mixed stock analysis (January-February spawners vs. March-April spawners

vs. May-June spawners) allowed us to identify additional variability in the ancient samples. March-April spawners were abundant at the Bay Street assemblages, contributing ~90% of total samples in both temporal layers (Figure 3-7, Table 3-7). In contrast, the Burton Acres assemblage was more diverse; the more recent layer (post-contact) had much higher proportions of early spawners (~30% January-February spawners) than the other samples (Table 3-7), while May-June spawners accounted for 16% of samples in the older layer (915-680 ybp; Figure 3-7; Table 3-7).

DISCUSSION

Herring populations supporting ancient fisheries

By investigating patterns of past population diversity with aDNA, we were able to identify herring populations harvested by Coast Salish fishers over a period of approximately 900 years. Our genetic data reveal that catches in central Puget Sound were dominated by winter spawning herring (Figure 3-7). However, several late spawning herring were detected in the older Burton Acres assemblage (915-680 ybp), using both individual assignment and mixed stock analysis methods. Additionally, mixed stock analysis indicated that the Burton Acres herring catches were more diverse and likely consisted of mixtures of populations (Figure 3-7). Given that the estimated assignment accuracy to late spawners was between 98-99%, it is possible that some individuals were erroneously assigned to the late spawning population. However, if these ancient samples are truly late spawners, then our results may be an indication that people at the Burton Acres site used a portfolio of herring populations and benefited from the resource wave created by different spawning phenotypes.

The ancient faunal assemblage is a product of both human foraging behaviors and resource availability, and also the result of taphonomic processes. Thus, there are a variety of

scenarios regarding the spatial scale of fishing, the relative abundance of herring populations, and/or the seasonality of herring harvest that may explain why winter spawning populations were highly abundant in the sampled archeological sites. We discuss each scenario below and place it in the context of available ecological and ethnographic information.

One possible explanation for the predominance of winter spawners in the archaeological assemblages is that people primarily harvested herring that were available locally. Regional analyses of Northwest archeological sites (Kopperl 2001, Moss 2012) indicate that species encountered in faunal assemblages are correlated with local habitats; for example, riverine sites typically contain abundant salmon bones while coastal sites are characterized by marine species such as flatfish, cod, and sculpin. Pacific herring are relatively abundant throughout Puget Sound year-round (Rice et al. 2012, Greene et al. 2015) but their reproductive behavior makes them easily accessible to coastal fishers and predators in the weeks and months prior to spawning, as adult fish aggregate to the nearshore environment to reproduce (Haegele and Schweigert 1985). A recent synthesis of archaeological data from the Northwest coast of North America (McKechnie et al. 2014) demonstrated that archaeological sites with a very high abundance (> 80%) of herring bones are spatially correlated with modern spawning locations, often co-occurring at distances less than 1 km, while sites with lower abundances of herring bones are more variable in regards to their distance to contemporary spawning locations.

Both archaeological sites analyzed here are proximate (within 10 km) to contemporary winter spawning grounds (Figure 3-1), and the Burton Acres site is located in a bay with documented yearly spawning in winter ("Quartermaster Harbor" in Figure 3-2). Furthermore, multidecadal spawn surveys (Stick et al. 2014) show that most herring populations in central Puget Sound reproduce in winter and early spring, while the only late spawning population

reproduces at Cherry Point, a location that is approximately 170 km to the north of the sampled archaeological sites (Figure 3-1). Genetic data (this dissertation, second chapter) suggest that populations exhibit seasonal and geographic fidelity to spawning sites across multiple generations, thus it is possible that the great abundance of winter spawners in central Puget Sound faunal assemblages is the result of both spatiotemporal stability in the distribution of spawning sites and spatially localized fishing during the spawning season. To rigorously test this hypothesis, future research efforts could focus sampling such that ancient specimens are collected close to locations used by contemporary late spawning populations.

A second possible explanation for the high abundance of winter-spawning herring in the archaeological record is that this resource may have been most desirable during the winter months. Pacific herring and its roe are calorie-rich and nutritious (Moss 2016), and these items may have been an especially welcome food source in the winter, when other foods were not easily accessible or in short supply: "*Herring eggs were, also, eaten fresh with smoked salmon or, if the supply of smoked salmon were exhausted by the time herring eggs were available, they were eaten with sprouts*" (Smith 1940). Furthermore, the ethnographic literature contains a few descriptions that explicitly mention harvest during the spawning season. For example, here is an account of fishing by Puyallup and Nisqually people in central Puget Sound: "*During the spawning season when large schools of herring and smelt crowded in to shore, the fish were dipped out with a loosely twined piece of matting. This matting was six feet long, about two and a half feet wide in the center and tapered to points at each end. It was made from green cedar boughs split once through the heart...The fisherman grasped the handles, one in each hand, the flat surface of the split poles toward him, bent the matting into a crescent-shaped curve, waded*

in among the spawning fish, scooped them up and dumped them on the beach beyond the water line." (Smith 1940).

A third scenario that may explain the large number of winter-spawners in the faunal assemblage is that people harvested populations that were locally abundant and/or available for extended periods of time in the nearshore environment. It is important to consider this scenario, as ethnographies document that Coast Salish people also fished for herring outside of the spawning season (Smith 1940). Unfortunately, there is limited contemporary information on the population-specific movements and distribution of Puget Sound herring outside of their reproductive season. However, studies analyzing herring otolith microchemistry (Gao et al. 2001) and persistent organochlorine pollutants (West et al. 2008) demonstrate that contemporary late spawning herring from Cherry Point have different chemical signatures from winter spawning populations from central Puget Sound. These results indicate that the contemporary late spawning population spends relatively little time in the estuary. Therefore it is possible that Coast Salish fishing communities in central Puget Sound depended on winter spawning herring populations because of their longer residence times in the estuary and their localized abundance.

Although the temporal resolution of the archaeological record is limited to centennial scales, our study demonstrates that human communities in central Puget Sound predominantly harvested winter and early spring spawning herring over a period of 900 years. This stability in resource use is noteworthy given that our sampling encompassed the climatic oscillations of the Medieval Warm Period (900-1200 CE, 800-1100 ybp) and Little Ice Age (1300 – 1850 CE, 150-700 ybp), which are associated with large changes in resource use in other coastal communities such as the Norse Greenlanders (Dugmore et al. 2012) and New Zealand Maori (Nunn et al. 2007). Bioenergetic models indicate that warm oceanic regimes are associated with low

recruitment and biomass in Pacific herring (Rose et al. 2008), and water temperatures are estimated to have been ~ 0.9 °C warmer during the Medieval Warm Period than during the Little Ice Age in the Pacific Ocean (Rosenthal et al. 2013). Despite this climatic variability, our data suggest that winter spawning herring populations remained abundant enough that fishing communities in Puget Sound could access them. Stability in marine resource use across this time period is also observed in other archaeological sites on the Northwest Coast (Hopt and Grier 2018, Butler et al. 2019).

Ancient population genetics and the quantification of past diversity

In this study, we demonstrated the utility of preamplification PCR and SNP assays to analyze nuclear DNA in ancient fish samples. By genotyping a small number of SNPs with divergent allele frequencies in modern herring populations (winter spawners vs. late spawners), we were able to assign ancient individuals to contemporary populations and estimate their reproductive timing. This approach allowed us to cost-effectively genotype large numbers of ancient individuals and obtain sample sizes similar to those used in genetic studies of modern populations. The ability to analyze robust sample sizes reduces random sampling error and may lead to more accurate estimates of genetic diversity in ancient assemblages.

However, we likely overestimated the predicted accuracy of our locus panel for individual assignment to spawning group because we did not conduct a double cross-validation of assignment using an independent set of modern samples with known spawn timing (Anderson 2010). Sample sizes of the modern herring populations were relatively modest (mean $N = 38$) and we did not want to further subdivide those samples into smaller "training" and "holdout" data sets, as the estimated population allele frequencies would be subject to greater sampling error. Furthermore, we decided to use temporally replicated samples to verify that allele

frequencies at the seven loci were stable through time. Another limitation of our study is that we could only distinguish between winter and late spawning herring; because we used a small number of loci, we may have missed additional layers of genetic diversity, such as the presence of distinct winter spawning populations or extinct populations that are not present in our contemporary genetic baseline. Finally, our analyses are based on the assumption that the loci differentiating contemporary populations have temporally stable allele frequencies, and were thus also correlated with spawn timing in ancient populations. One of the loci used in our analyses (Locus_8468, Figure 4) is within a gene (*SYNE2*) strongly correlated with spawn timing in both Atlantic and Pacific herring (this dissertation, second chapter), suggesting that this genomic region is involved in the regulation of reproduction. Furthermore, allele frequencies at Locus_8468 are very divergent when comparing late and winter spawning populations across the Pacific Northwest coast (this dissertation, second chapter). These lines of evidence suggest that allele frequency differences at this locus are maintained by natural selection or have been present for a long time.

At present, there is a trade-off between analyzing whole genomes for small numbers of samples and analyzing many samples at few loci. Although working with aDNA is technically challenging and ancient samples may be limited or scarce, new developments in laboratory techniques, such as hybridization capture (Carpenter et al. 2013), and the declining cost of sequencing may make the analysis of large numbers of ancient individuals and loci more common in the future (Shapiro and Hofreiter 2014). This will not only facilitate evolutionary studies in charismatic species such as large mammals (Lynch et al. 2015, Skoglund et al. 2015, Barlow et al. 2018) and hominids (Meyer et al. 2012, Prüfer et al. 2013, Nielsen et al. 2017) but

may also allow for population genetic and ecological investigations in species that are currently important to human livelihoods and ecosystem processes.

In conclusion, we identified genetic diversity in ancient herring catches that is associated with phenotypic diversity and ecological resource waves in contemporary populations. Using aDNA, we demonstrated that ancient catches in central Puget Sound were predominantly composed of populations spawning in winter and early spring, a pattern consistent with current spawning distributions. At one of the archaeological sites, there was also evidence that Coast Salish people targeted a variety of herring populations and benefited from the ecological resource wave created by different spawning phenotypes.

TABLES AND FIGURES

Table 3-1. The geographic location and date of each modern herring sample collection. Reporting group (for individual assignment and genetic stock identification) are also noted.

Location name	Sample	Sample size	Latitude	Longitude	Sampling date	Two reporting groups	Three reporting groups
Squaxin Pass	modern	42	47.20	-122.94	1/17/2007	winter	Jan-Feb
Port Orchard	modern	46	47.69	-122.59	1/27/2014	winter	Jan-Feb
Skagit Bay	modern	48	48.44	-122.57	2/25/2015	winter	Jan-Feb
Port Gamble	modern	22	47.87	-122.60	3/12/2014	winter	Mar-Apr
Quilcene Bay	modern	16	47.80	-122.85	3/26/2014	winter	Mar-Apr
Gabriola Island	modern	45	49.16	-123.74	3/19/2015	winter	Mar-Apr
Elliott Bay	modern	47	47.62	-122.36	4/15/2015	winter	Mar-Apr
Cherry Point A	modern	34	48.93	-122.80	5/12/2014	late	May-Jun
Cherry Point B	modern	47	48.93	-122.80	5/9/2016	late	May-Jun

Table 3-2. The geographic location and relative age of each archaeological sample collection. Radiocarbon dates were obtained from pieces of charcoal within each unit, with the exception of the "post-contact" layer from Burton Acres. This layer was likely deposited after CE ~1850 because it contained metal artifacts such as a dime minted between 1860 and 1890, and a button likely manufactured in the 1890s (Stein and Phillips, 2002).

Location name	Sample	Sample size	Latitude	Longitude	Age (years before present)
Bay Street Shell Midden	ancient	48	47.54	-122.64	400-100 ybp
Bay Street Shell Midden	ancient	48	47.54	-122.64	800-550 ybp
Burton Acres Shell Midden	ancient	48	47.39	-122.45	post-contact
Burton Acres Shell Midden	ancient	48	47.39	-122.45	915-680 ybp

Table 3-3. Locus metadata for TaqMan assays used in this study.

Assay name	Forward primer	Reverse primer	Genomic scaffold (Atlantic herring)	metadata	Gene Annotation	SNP	mutation	global F_{ST}
Locus_24510	CCTTCCAGGTGACCTACAATCG	GGGTTATCGGCTGGTTCCT	NW_012222569.1	mRNA	thyrotopin subunit beta (THSb)	A/G	transition	0.24
Locus_30660	GAAGCTATGGTTCAATAAAACCTGCAG	GGTCCTTTATAACCCACCCTCTCT	NW_012224044.1	mRNA	calcium dependent secretion activator (CADPS)	G/T	transversion	0.17
Locus_6354	TGCGGCCGCCTTACC	CACACACACACACACACTACTC	NW_012220857.1	intergenic		G/T	transversion	0.11
Locus_8468	GTGTGTGTGCACATCTGTACCT	GGAGCGCTGCAGAGACAA	NW_012220751.1	mRNA	nesprin-2 (SYNE2)	TG/CT	transition, transversion	0.37
Locus_519	GGTGTCCTACCGTTCCATAGAGAA	CTTGCTGCTGCCCATCTG	NW_012220167.1	intergenic		A/G	transition	0.18
Locus_31014	AAGCCACACGCTTTCTACCA	CTGGACGTGGTGACGGT	NW_012218584.1	mRNA	nuclear factor of activated T-cells, cytoplasmic 2-like (NFATC2)	C/T	transition	0.25
Locus_725	ACAGGACGTACCATTTCTGTATCTGT	GATGGAAATGTGGAATGGGAGAGAA	NW_012217786.1	mRNA	growth factor receptor-bound protein 2 (GRB2)	G/T	transversion	0.25

Table 3-4. Hierarchical AMOVA results demonstrating significant population structure between spawning groups (winter spawners vs. late spawners) in modern samples, using a set of 7,261 RAD loci and a subset of seven loci.

Variance component	Number of loci	Variance	Percent of total variance (%)	*P-value*	Φ-statistic
Between spawning groups	*N* = 7	1.26	44.7	0.022	Φ_{CT} = 0.51
Between populations within spawning groups	*N* = 7	0.13	4.6	0.001	Φ_{SC} = 0.09
Within populations	*N* = 7	1.43	50.7	0.001	Φ_{ST} = 0.55
Between spawning groups	*N* = 7,261	19.58	1.93	0.035	Φ_{CT} = 0.02
Between populations within spawning groups	*N* = 7,261	16.45	1.62	0.001	Φ_{SC} = 0.02
Within populations	*N* = 7,261	980.93	96.46	0.001	Φ_{ST} = 0.04

Table 3-5. Results of individual assignment for modern herring samples.

Site	Collection date	Samples genotyped at 6+ loci	Unassigned samples	Samples assigned to winter-spawners	Samples assigned to late-spawners
Squaxin Pass	1/17/2007	42	1	40	1
Port Orchard	1/27/2014	45	1	44	0
Skagit Bay	2/25/2015	48	3	44	1
Port Gamble	3/12/2014	22	1	21	0
Quilcene Bay	3/26/2014	16	1	15	0
Gabriola Island	3/19/2015	45	1	44	0
Elliott Bay	4/15/2015	47	0	47	0
Cherry Point A	5/12/2014	32	2	0	30
Cherry Point B	5/9/2016	47	2	1	44

Table 3-6. Results of individual assignment for ancient samples.

Site	Radiocarbon date of layer	Samples genotyped at 6+ loci	Unassigned samples	Samples assigned to winter-spawners	Samples assigned to late-spawners
Bay Street	400-100 ybp	41	0	41	0
Bay Street	800-550 ybp	46	5	40	1
Burton Acres	post-contact	40	6	34	0
Burton Acres	915-680 ybp	32	4	25	3

Table 3-7. Results of mixed stock analysis for ancient samples.

Site	Radiocarbon date of layer	Reporting group	Mean estimated proportion	5th-95th credible interval
Two reporting groups				
Bay Street	400-100 ybp	late	< 0.01	0.00-0.03
Bay Street	400-100 ybp	winter	0.99	0.97-1.00
Bay Street	800-550 ybp	late	0.01	0.00-0.06
Bay Street	800-550 ybp	winter	0.99	0.94-1.00
Burton Acres	post-contact	late	0.01	0.00-0.04
Burton Acres	post-contact	winter	0.99	0.96-1.00
Burton Acres	915-680 ybp	late	0.16	0.06-0.29
Burton Acres	915-680 ybp	winter	0.84	0.71-0.94
Three reporting groups				
Bay Street	400-100 ybp	Jan-Feb	0.08	0.00-0.26
Bay Street	400-100 ybp	Mar-Apr	0.91	0.73-1.00
Bay Street	400-100 ybp	May-Jun	0.01	0.00-0.03
Bay Street	800-550 ybp	Jan-Feb	0.09	0.01-0.21
Bay Street	800-550 ybp	Mar-Apr	0.90	0.77-0.98
Bay Street	800-550 ybp	May-Jun	0.01	0.00-0.07
Burton Acres	post-contact	Jan-Feb	0.36	0.21-0.52
Burton Acres	post-contact	Mar-Apr	0.63	0.47-0.79
Burton Acres	post-contact	May-Jun	0.01	0.00-0.04
Burton Acres	915-680 ybp	Jan-Feb	0.15	0.01-0.35
Burton Acres	915-680 ybp	Mar-Apr	0.69	0.51-0.85
Burton Acres	915-680 ybp	May-Jun	0.16	0.06-0.29

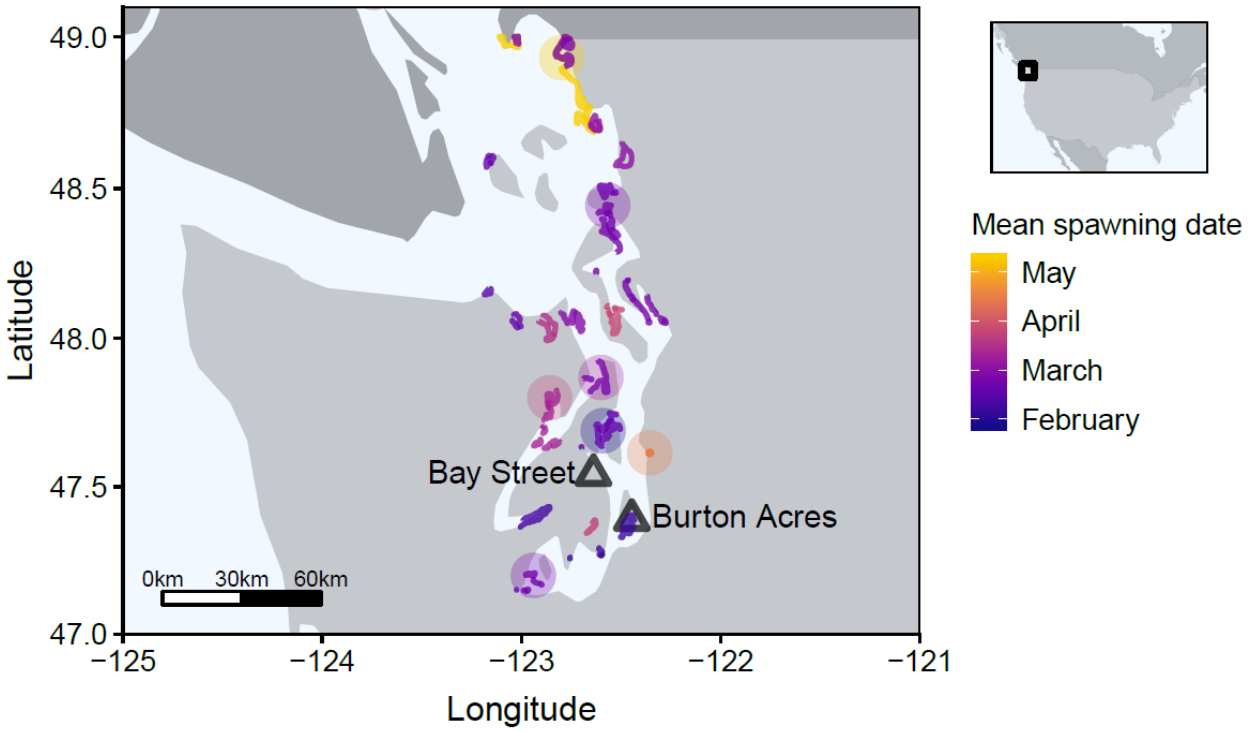


Figure 3-1. Map of archaeological sites (black triangles) in relation to contemporary herring spawning locations in Puget Sound. Spawning locations in Washington State are depicted by colored outlines that represent the average reproductive date, estimated using data from multiple decades of spawn surveys (Stick et al. 2014). We do not show spawning locations in British Columbia. Genetic samples from contemporary spawning herring were collected at the locations marked with translucent circles.

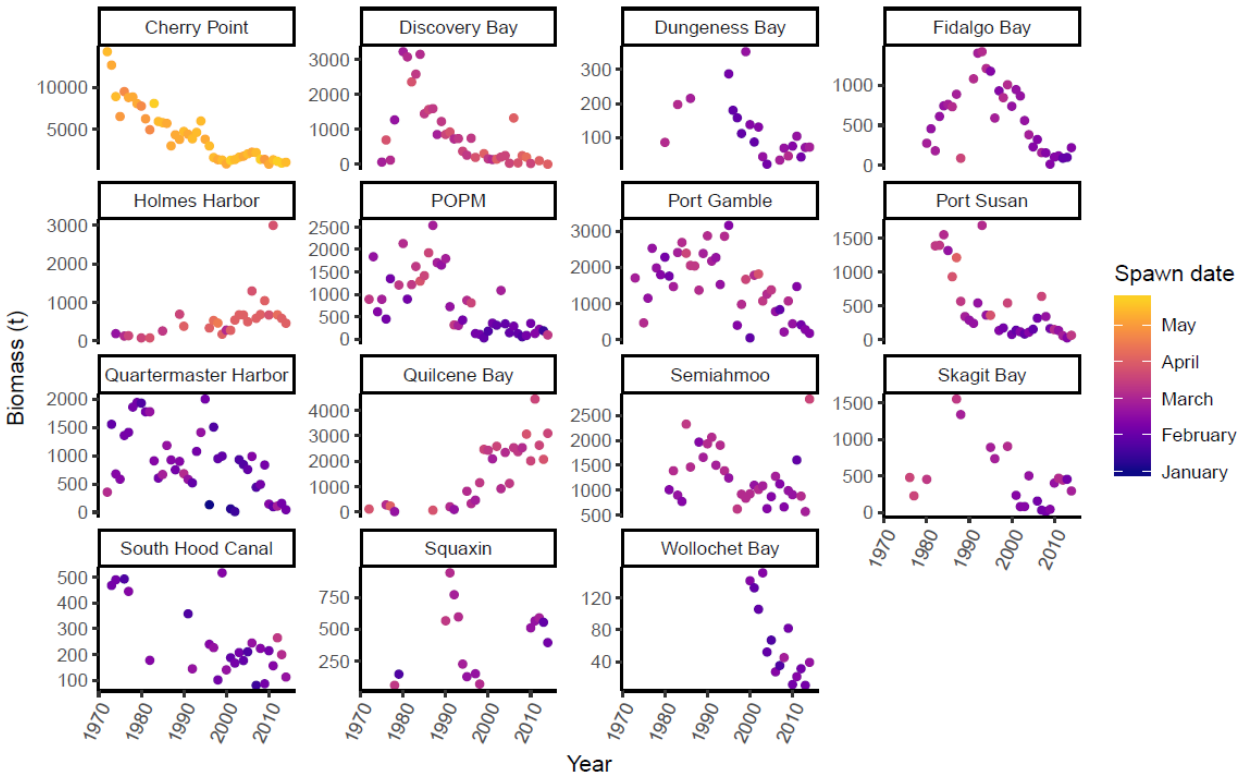


Figure 3-2. Biomass of adult herring estimated for different spawning sites in Puget Sound. Data are from the Washington Department of Fish and Wildlife (Stick et al. 2014). The color of each point represents the date of peak spawning activity in that year. We only show sites that have been surveyed for more than 10 years. The site "POPM" refers to Port Orchard-Port Madison. Note that the y-axes are different for each plot, and that the late-spawning population once had the largest biomass.

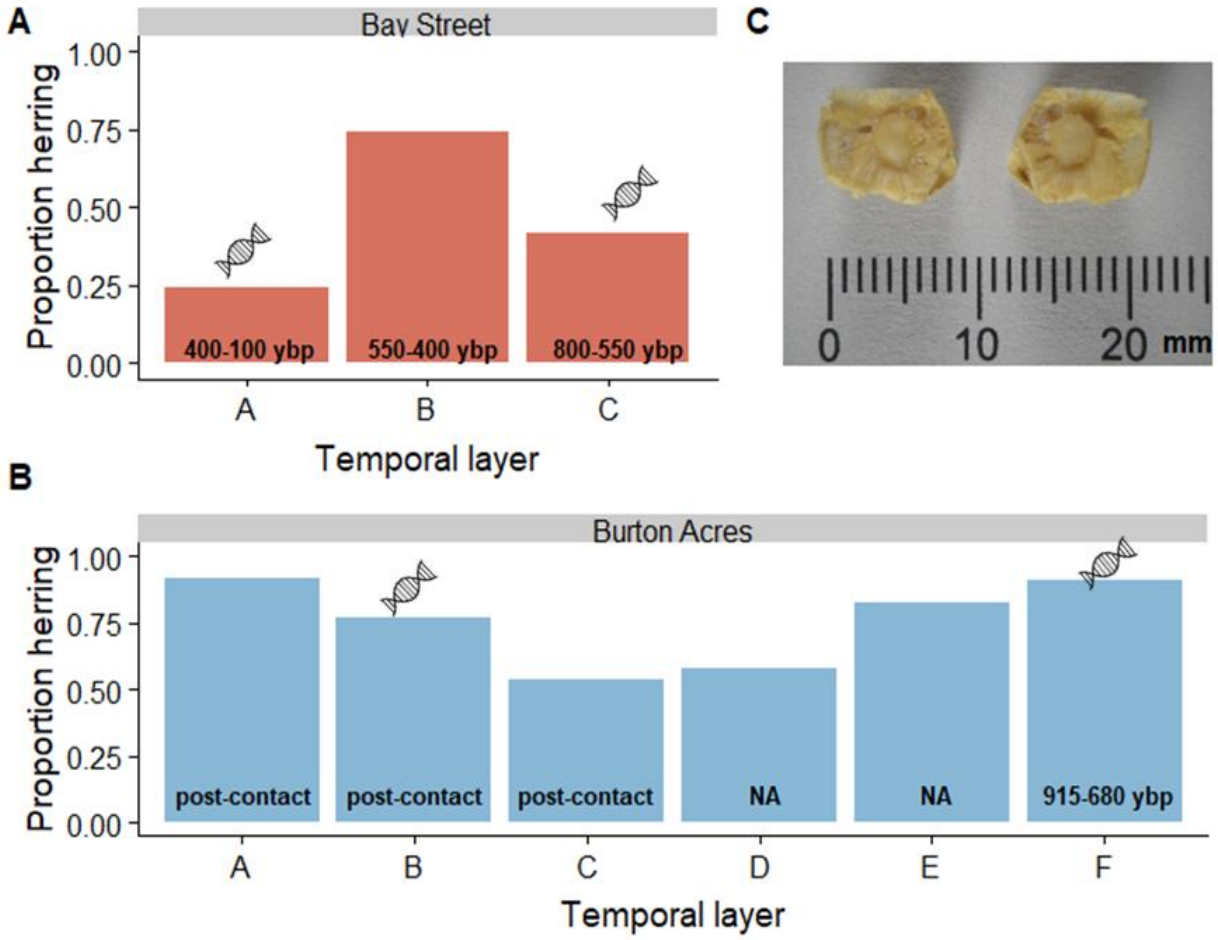


Figure 3-3. Proportions of herring in relation to other fish bones at each archaeological site. Temporal layers are labelled with their estimated age, when that information is available. Layers sampled in this study for aDNA are marked with a double helix. A) Bay Street shell midden (data from Lewarch et al. 2002), B) Burton Acres (data from Kopperl 2001), and C) examples of prootic bones used in this study.

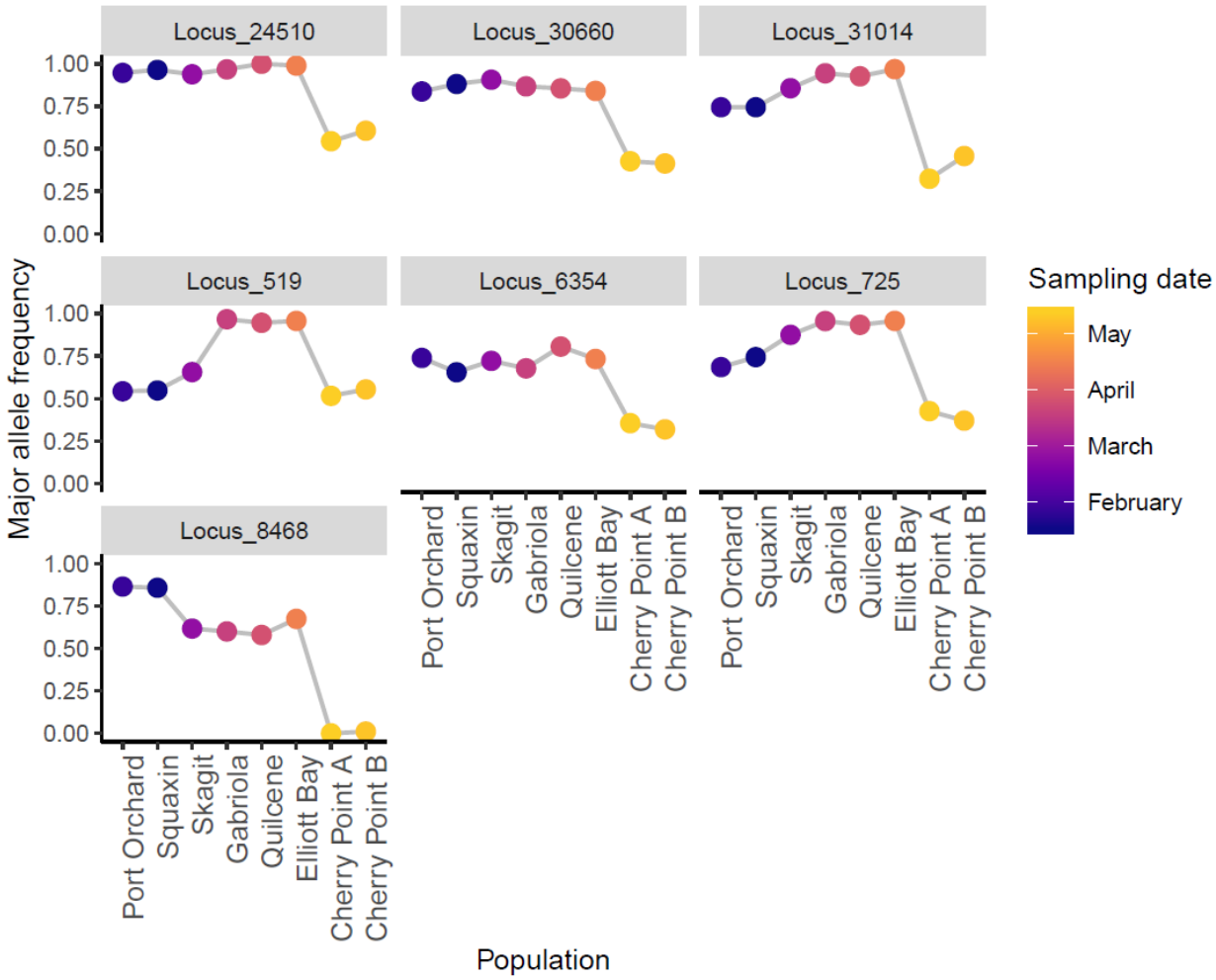


Figure 3-4. Major allele frequency patterns of SNPs in contemporary herring populations, based on genotypes generated by RAD sequencing. The color of each point represents the date of sampling a spawning event. Note the relatively large differences in allele frequency when comparing late-spawners to other herring populations in Puget Sound.

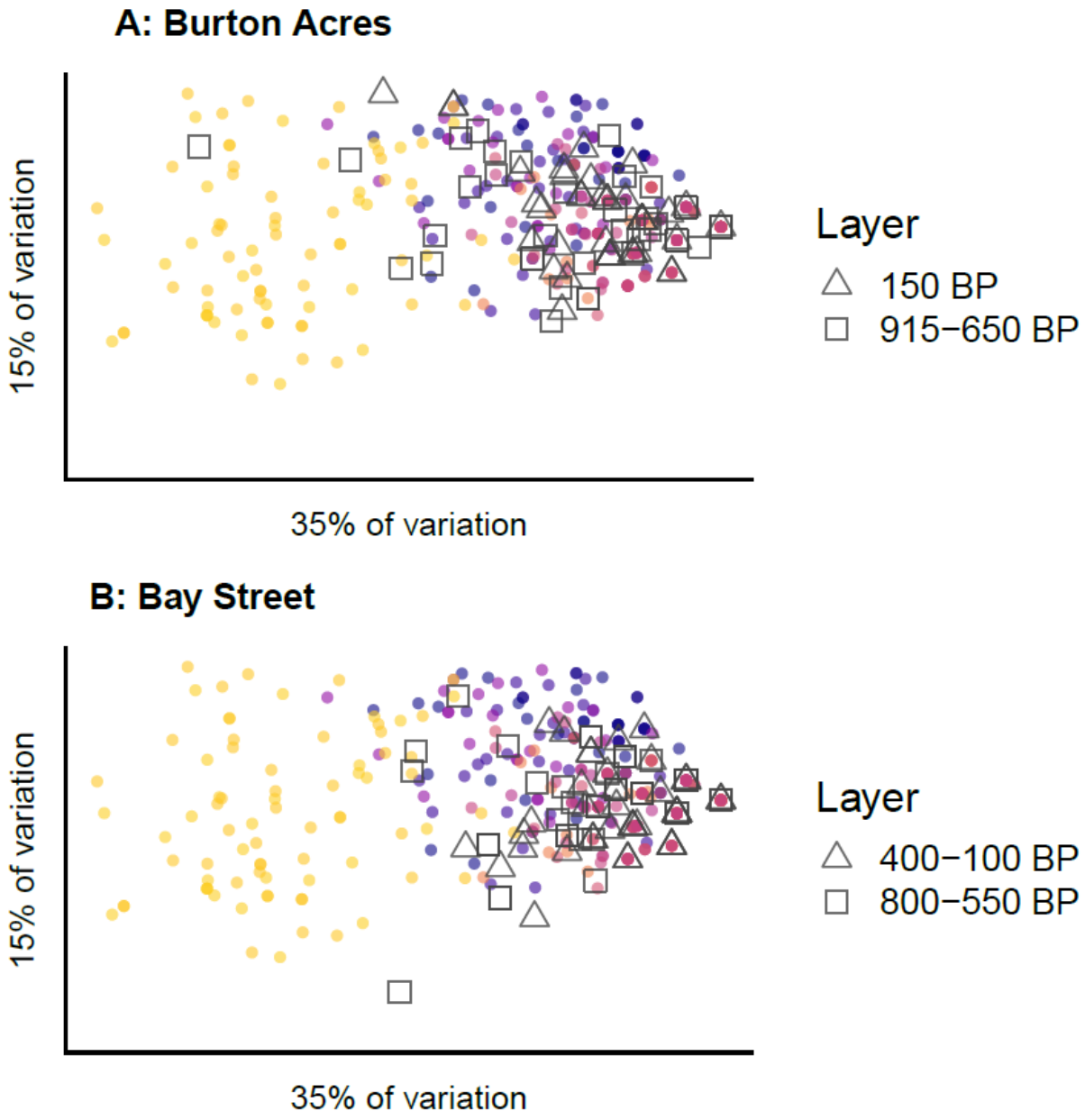


Figure 3-5. PCA of modern and ancient herring samples using 7 nuclear SNPs. Points represent the modern samples genotyped using RAD sequencing, and color indicates the date of sample collection (as in Figure 1). The ancient herring samples, genotyped using SNP-based assays, are represented by black triangles or squares, according to their estimated age. A) Comparison of modern herring and ancient samples collected from Burton Acres; B) Comparison of modern herring and ancient samples collected from the Bay Street shell midden.

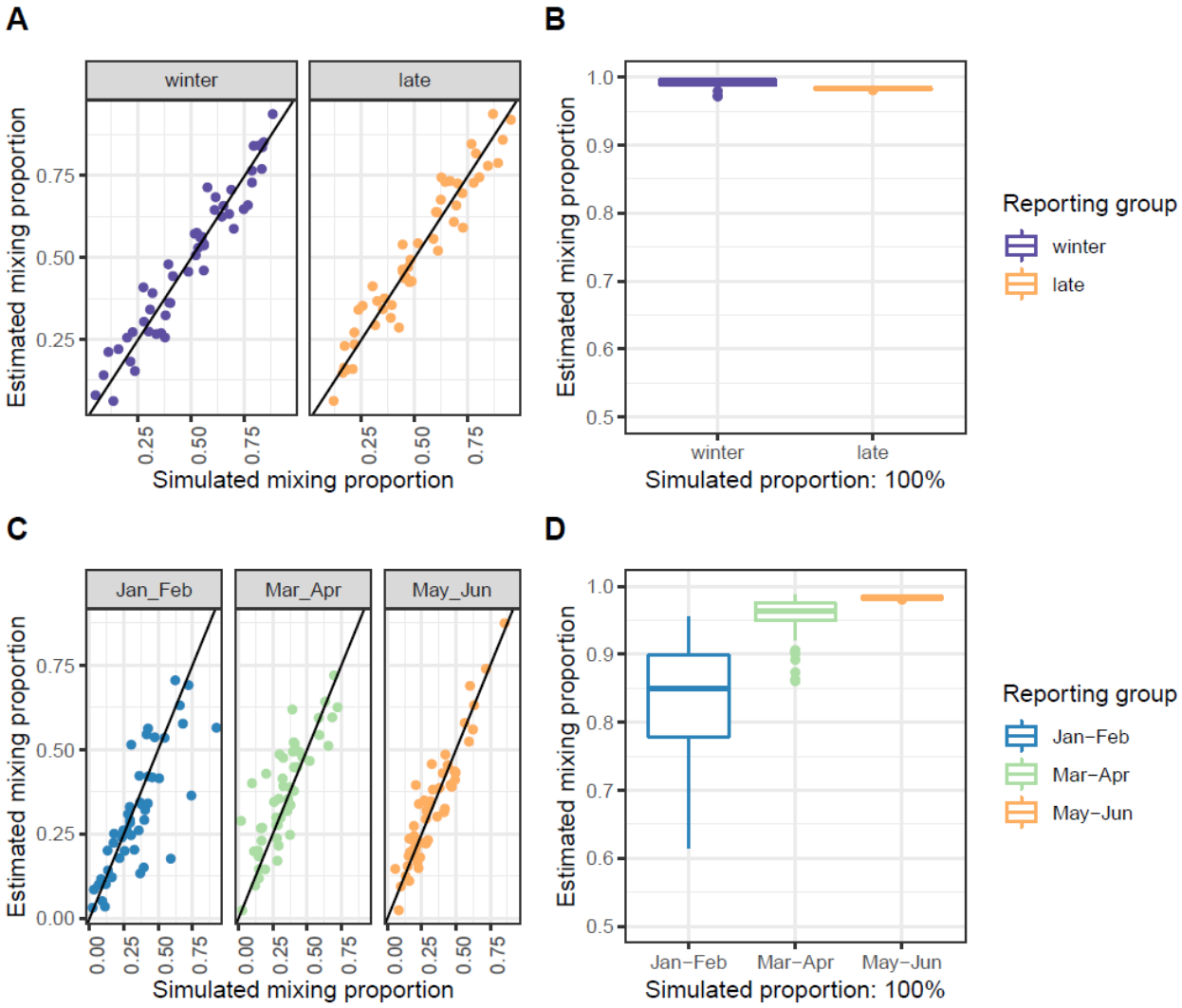


Figure 3-6. Predicted accuracy of mixed stock analysis using simulated mixtures with known mixture proportions (x-axis), generated from modern herring samples genotyped at seven nuclear loci. A) Estimates of simulated mixture proportions using two reporting groups. Diagonal line indicates expectations for perfect assignment. B) Results of 100% simulations using two reporting groups. C) Estimates of simulated mixture proportions using three reporting groups. Diagonal line indicates expectations for perfect assignment. D) Results of 100% simulations using three reporting groups.

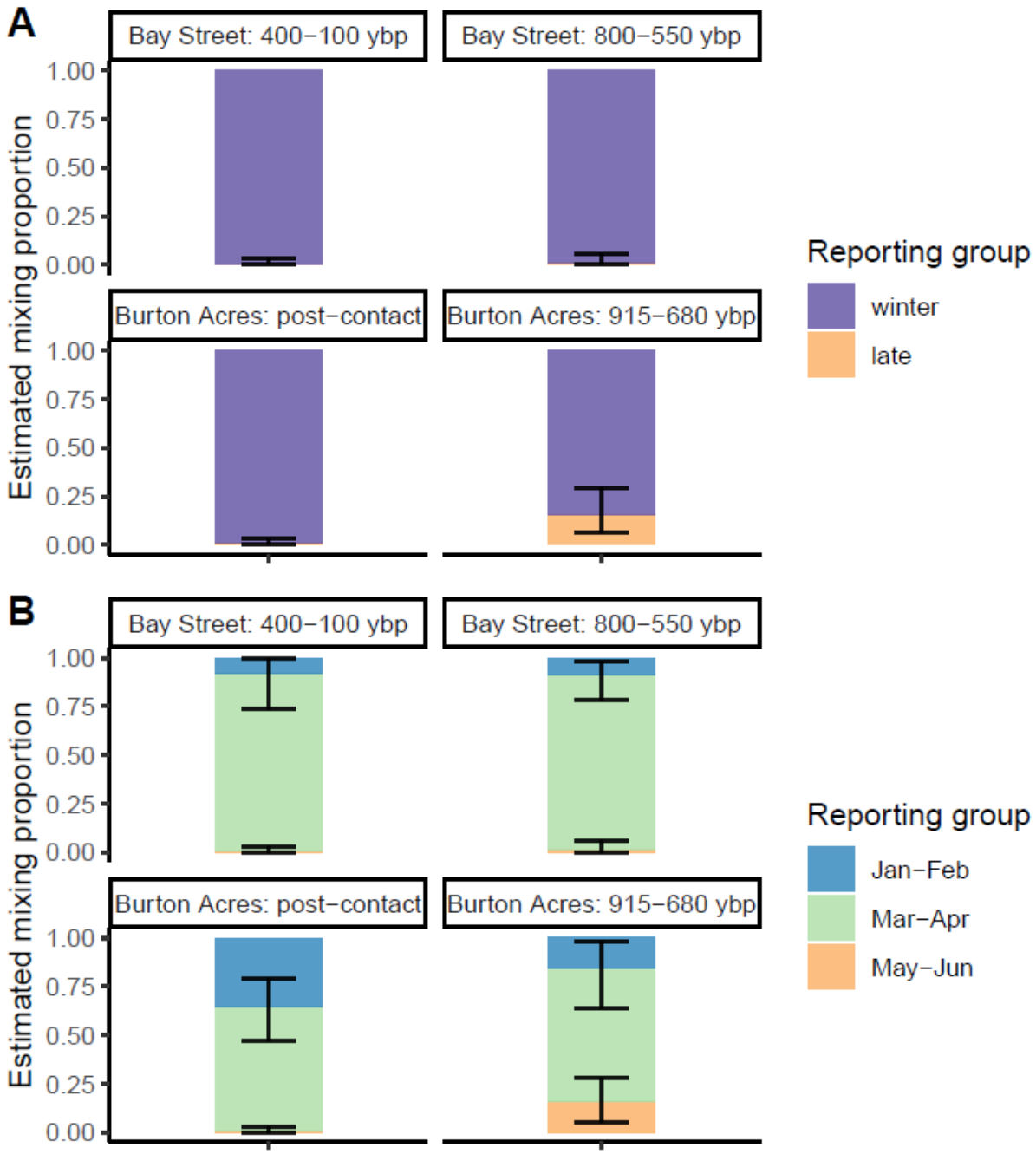


Figure 3-7. Results of mixed stock analysis for ancient herring samples. Estimated mixture proportions are displayed on y-axis and error bars indicate the 95% credible intervals. Different panels represent archaeological sites and temporal layers, while colors correspond to a reporting group. A) Estimated mixture proportions using two reporting groups; B) Estimated mixture proportions using three reporting groups

REFERENCES

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics **11**:697-709.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology **215**: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**:701-710.

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

Archer, F. I., P. E. Adams, and B. B. Schneiders. 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources **17**:5-11.

Armstrong, J. B., G. Takimoto, D. E. Schindler, M. M. Hayes, and M. J. Kauffman. 2016. Resource waves: phenological diversity enhances foraging opportunities for mobile consumers. Ecology **97**:1099-1112.

Asch, R. G. 2015. Climate change and decadal shifts in the phenology of larval fishes in the California Current ecosystem. Proceedings of the National Academy of Sciences **112**:E4065.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE **3**:e3376.

Balloux, F., W. Amos, and T. Coulson. 2004. Does heterozygosity estimate inbreeding in real populations? Molecular Ecology **13**:3021-3031.

Bargmann, G. 1998. Forage fish management plan. Washington Department of Fish and Wildlife.

Barlow, A., J. A. Cahill, S. Hartmann, C. Theunert, G. Xenikoudakis, G. G. Fortes, J. L. A. Paijmans, G. Rabeder, C. Frischauf, A. Grandal-d'Anglade, A. García-Vázquez, M. Murtskhvaladze, U. Saarma, P. Anijalg, T. Skrbinšek, G. Bertorelle, B. Gasparian, G. Bar-Oz, R. Pinhasi, M. Slatkin, L. Dalén, B. Shapiro, and M. Hofreiter. 2018. Partial genomic survival of cave bears in living brown bears. Nature Ecology & Evolution **2**:1563-1570.

Barrett, R. D. H., S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, J. S. Duke-Cohan, J. D. Jensen, and H. E. Hoekstra. 2019. Linking a mutation to survival in wild mice. Science **363**:499.

Beacham, T. D., J. F. Schweigert, C. MacConnachie, K. D. Le, and L. Flostrand. 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**:1795-1811.

Brieuc, M. S. O., C. D. Waters, J. E. Seeb, and K. A. Naish. 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**:447-460.

Bucholtz, R. H., J. Tomkiewicz, and J. Dalskov. 2008. Manual to determine gonadal maturity of herring (*Clupea harengus L.*). Pages 1-45 *in* N. I. o. A. R. DTU Aqua, editor. DTU Aqua-report. DTU Aqua, National Institute of Aquatic Resources, Charlottenlund, Denmark.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J. C. Glaubitz, M. M. Goodman, C. Harjes, K. Guill, D. E. Kroon, S. Larsson, N. K. Lepak, H. Li, S. E. Mitchell, G. Pressoir, J. A. Peiffer, M. O. Rosas, T. R. Rocheford, M. C. Romay, S. Romero, S. Salvo, H. S. Villeda, H. Sofia da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M. D. McMullen. 2009. The genetic architecture of maize flowering time. Science **325**:714.

Butler, V. L., S. K. Campbell, K. M. Bovy, and M. A. Etnier. 2019. Exploring ecodynamics of coastal foragers using integrated faunal records from Číxwicən village (Strait of Juan de Fuca, Washington, U.S.A.). Journal of Archaeological Science: Reports **23**:1143-1167.

Caldwell, M. 2011. Fish traps and shell middens at Comox Harbour, British Columbia. Pages 235-246 *in* M. L. Moss and A. Cannon, editors. The Archaeology of North Pacific Fisheries. University of Alaska Press.

Cale, C. M., M. E. Earll, K. E. Latham, and G. L. Bush. 2016. Could secondary DNA transfer falsely place someone at the scene of a crime? Journal of Forensic Sciences **61**:196-203.

Campana, M. G., N. Robles García, F. J. Rühli, and N. Tuross. 2014. False positives complicate ancient pathogen identifications using high-throughput shotgun sequencing. BMC Research Notes **7**:111.

Carpenter, Meredith L., Jason D. Buenrostro, C. Valdiosera, H. Schroeder, Morten E. Allentoft, M. Sikora, M. Rasmussen, S. Gravel, S. Guillén, G. Nekhrizov, K. Leshtakov, D. Dimitrova, N. Theodossiev, D. Pettener, D. Luiselli, K. Sandoval, A. Moreno-Estrada, Y. Li, J. Wang, M. Thomas P. Gilbert, E. Willerslev, William J. Greenleaf, and Carlos D. Bustamante. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. The American Journal of Human Genetics **93**:852-864.

Carrothers, W. A. 1941. The British Columbia Fisheries. The University of Toronto Press, Toronto, Canada.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology **22**:3124-3140.

Ceballos, G., P. R. Ehrlich, and R. Dirzo. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences **114**:E6089.

Chamberlin, J. W., B. R. Beckman, C. M. Greene, C. A. Rice, and J. E. Hall. 2017. How relative size and abundance structures the relationship between size and individual growth in an ontogenetically piscivorous fish. Ecology and evolution **7**:6981-6995.

Clarke, K. R., and R. N. Gorley. 2006. PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science **307**:1928.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics **185**:1411.

Cushing, D. H. 1969. The regularity of the spawning season of some fishes. Journal du Conseil **33**:81-92.

Cushing, D. H. 1990. Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. Pages 249-293 *in* J. H. S. B. a. A. J. Southward, editor. Advances in Marine Biology. Academic Press.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and G. Genomes Project Analysis. 2011. The variant call format and VCFtools. Bioinformatics **27**:2156-2158.

De Danaan, L. 2002. Ethnographic background. Pages 17-31 *in* J. K. Stein and L. S. Phillips, editors. Vashon Island archaeology: a view from Burton Acres Shell Midden. Burke Museum, Seattle, Washington.

Dray, S., and A.-B. Dufour. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software; Vol 1, Issue 4 (2007).

Dugmore, A. J., T. H. McGovern, O. Vésteinsson, J. Arneborg, R. Streeter, and C. Keller. 2012. Cultural adaptation, compounding vulnerabilities and conjunctures in Norse Greenland. Proceedings of the National Academy of Sciences **109**:3658.

Earl, D. A., and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources **4**:359-361.

Edwards, M., and A. J. Richardson. 2004. Impact of climate change on marine pelagic phenology and trophic mismatch. Nature **430**:881.

Eells, M., and G. P. Castile. 1985. *The Indians of Puget Sound : the Notebooks of Myron Eells*. University of Washington Press, Whitman College

Ekblom, R., and J. Galindo. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity **107**:1-15.
Erlandson, J. M., and T. C. Rick. 2009. Archaeology Meets Marine Ecology: The Antiquity of Maritime Cultures and Human Impacts on Marine Fisheries and Ecosystems. Annual Review of Marine Science **2**:231-251.
Etter, P., S. Bassham, P. Hohenlohe, E. Johnson, and W. Cresko. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Pages 157-178 Molecular Methods for Evolutionary Genetics. Humana Press.
Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology **14**:2611-2620.
Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources **10**:564-567.
Flickinger, M., G. Jun, Gonçalo R. Abecasis, M. Boehnke, and Hyun M. Kang. 2015. Correcting for sample contamination in genotype calling of DNA sequence data. The American Journal of Human Genetics **97**:284-290.
Foll, M., and O. Gaggiotti. 2006. Identifying the environmental factors that determine the genetic structure of populations. Genetics **174**:875.
Fountain, E. D., J. N. Pauli, B. N. Reid, P. J. Palsbøll, and M. Z. Peery. 2016. Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. Molecular Ecology Resources **16**:966-978.
Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution **27**:489-496.
Gaggiotti, O. E., D. Bekkevold, H. B. H. Jørgensen, M. Foll, G. R. Carvalho, C. Andre, and D. E. Ruzzante. 2009. Disentangling the effects of evolutionary, demographic, and environmental factors influencing the genetic structure of natural populations: Atlantic herring as a case study. Evolution **63**:2939-2951.
Gao, Y. W., S. H. Joner, and G. G. Bargmann. 2001. Stable isotopic composition of otoliths in identification of spawning stocks of Pacific herring (*Clupea pallasi*) in Puget Sound. Canadian Journal of Fisheries and Aquatic Sciences **58**:2113-2120.
Gauvreau, A. M., D. Lepofsky, M. Rutherford, and M. Reid. 2017. "Everything revolves around the herring": the Heiltsuk–herring relationship through time. Ecology and Society **22**.
Goudet, J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes **5**:184-186.
Graham, C. F., T. C. Glenn, A. G. McArthur, D. R. Boreham, T. Kieran, S. Lance, R. G. Manzon, J. A. Martino, T. Pierson, S. M. Rogers, J. Y. Wilson, and C. M. Somers. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Molecular Ecology Resources **15**:1304-1315.
Greene, C., L. Kuehne, C. Rice, K. Fresh, and D. Penttila. 2015. Forty years of change in forage fish and jellyfish abundance across greater Puget Sound, Washington (USA): anthropogenic and climate associations. Marine Ecology Progress Series **525**:153-170.
Greenstone, M. H., D. C. Weber, T. A. Coudron, M. E. Payton, and J. S. Hu. 2012. Removing external DNA contamination from arthropod predators destined for molecular gut-content analysis. Molecular Ecology Resources **12**:464-469.
Greenstone, M. H., D. C. Weber, T. C. Coudron, and M. E. Payton. 2011. Unnecessary roughness? Testing the hypothesis that predators destined for molecular gut-content analysis must be hand-collected to avoid cross-contamination. Molecular Ecology Resources **11**:286-293.
Günther, T., and G. Coop. 2013a. Robust identification of local adaptation from allele frequencies. Genetics **195**:205.
Günther, T., and G. Coop. 2013b. Robust Identification of Local Adaptation from Allele Frequencies. Genetics **195**:205-220.

Hadly, E. A., and A. D. Barnosky. 2009. Vertebrate fossils and the future of conservation b iology. Pages 39-59 *in* G. P. Dietl and K. W. Flessa, editors. Conservation Paleobiology: using the past to manage for the future Paleontological Society Papers, Lubbock, TX.

Haegele, C. W., and J. F. Schweigert. 1985. Distribution and Characteristics of Herring Spawning Grounds and Description of Spawning Behavior. Canadian Journal of Fisheries and Aquatic Sciences **42**:s39-s55.

Halffman, C. M., B. A. Potter, H. J. McKinney, B. P. Finney, A. T. Rodrigues, D. Y. Yang, and B. M. Kemp. 2015. Early human use of anadromous salmon in North America at 11,500 y ago. Proceedings of the National Academy of Sciences **112**:12344-12348.

Hammond, J. 1886. Fish in Puget Sound. The American Angler **25**:392-393.

Hatch, S. A. 2013. Kittiwake diets and chick production signal a 2008 regime shift in the Northeast Pacific. Marine Ecology Progress Series **477**:271-284.

Hay, D. E., and P. B. McCarter. 1999. Distribution and timing of herring spawning in British Columbia. Fisheries and Oceans Canada.

Hebert, K. 2017. Southeast Alaska 2017 herring stock assessment s urveys Alaska Department of Fish and Game, Divisions of Sport Fish and Commercial Fisheries.

Hendry, A. P., and T. Day. 2005. Population structure attributable to reproductive time: isolation by time and adaptation by time. Molecular Ecology **14**:901-916.

Hijmans, R. J. 2017. geosphere: Spherical Trigonometry.

Hilborn, R., T. P. Quinn, D. E. Schindler, and D. E. Rogers. 2003. Biocomplexity and fisheries sustainability. Proceedings of the National Academy of Sciences **100**:6564-6568.

Hoffman, J. I., F. Simpson, P. David, J. M. Rijks, T. Kuiken, M. A. S. Thorne, R. C. Lacy, and K. K. Dasmahapatra. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences **111**:3775-3780.

Hofinger, B. J., J. R. Russell, C. G. Bass, T. Baldwin, M. Dos Reis, P. E. Hedley, Y. Li, M. Macaulay, R. Waugh, K. E. Hammond-Kosack, and K. Kanyuka. 2011. An exceptionally high nucleotide and haplotype diversity and a signature of positive selection for the eIF4E resistance gene in barley are revealed by allele mining and phylogenetic analyses of natural populations. Molecular Ecology **20**:3653-3668.

Hofreiter, M., D. Serre, H. N. Poinar, M. Kuch, and S. Paabo. 2001. Ancient DNA. Nat Rev Genet **2**:353-359.

Hopt, J., and C. Grier. 2018. Continuity Amidst Change: Village Organization and Fishing Subsistence at the Dionisio Point Locality in Coastal Southern British Columbia. The Journal of Island and Coastal Archaeology **13**:21-42.

Hourston, A. S., and H. Rosenthal. 1976. Sperm density during active spawning of Pacific herring (*Clupea harengus pallasi*). Journal of the Fisheries Research Board of Canada **33**:1788-1790.

Hubisz, M., J., D. Falush, M. Stephens, and K. Pritchard Jonathan. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources **9**:1322-1332.

Hufnagl, M., and M. A. Peck. 2011. Physiological individual-based modelling of larval Atlantic herring (*Clupea harengus*) foraging and growth: insights on climate-driven life-history scheduling. ICES Journal of Marine Science **68**:1170-1188.

Hughes, J. B., G. C. Daily, and P. R. Ehrlich. 1997. Population diversity: its extent and extinction. Science **278**:689.

Johnson, B. M., B. M. Kemp, and G. H. Thorgaard. 2018. Increased mitochondrial DNA diversity in ancient Columbia River basin Chinook salmon *Oncorhynchus tshawytscha*. PLOS ONE **13**:e0190059.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403-1405.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers,

C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, P. Broad Institute Genome Sequencing, T. Whole Genome Assembly, J. Baldwin, T. Bloom, D. B. Jaffe, R. Nicol, J. Wilkinson, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature **484**:55.

Jun, G., M. Flickinger, Kurt N. Hetrick, Jane M. Romm, Kimberly F. Doheny, Gonçalo R. Abecasis, M. Boehnke, and Hyun M. Kang. 2012. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. The American Journal of Human Genetics **91**:839-848.

Kaiser, T. S., B. Poehn, D. Szkiba, M. Preussner, F. J. Sedlazeck, A. Zrim, T. Neumann, L.-T. Nguyen, A. J. Betancourt, T. Hummel, H. Vogel, S. Dorner, F. Heyd, A. von Haeseler, and K. Tessmar-Raible. 2016. The genomic basis of circadian and circalunar timing adaptations in a midge. Nature **540**:69.

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

Kemp, B. M., and D. G. Smith. 2005. Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. Forensic Science International **154**:53-61.

Kimura, M., and G. H. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49**:561-576.

King, R. A., J. S. Davey, J. R. Bell, D. S. Read, D. A. Bohan, and W. O. C. Symondson. 2011. Suction sampling as a significant source of error in molecular analysis of predator diets. Bulletin of Entomological Research **102**:261-266.

Kjeldsen, S. R., K. R. Zenger, K. Leigh, W. Ellis, J. Tobey, D. Phalen, A. Melzer, S. FitzGibbon, and H. W. Raadsma. 2016. Genome-wide SNP loci reveal novel insights into koala (*Phascolarctos cinereus*) population variability across its range. Conservation Genetics **17**:337-353.

Knutsen, H., E. M. Olsen, P. E. Jorde, S. H. Espeland, C. André , and N. C. Stenseth. 2011. Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. Molecular Ecology **20**:768-783.

Kopperl, R. E. 2001. Herring use in southern Puget Sound: analysis of fish remains at 45-KI-437 Northwest Anthropological Research Notes **35**:1-20.

Koutsovoulos, G., S. Kumar, D. R. Laetsch, L. Stevens, J. Daub, C. Conlon, H. Maroon, F. Thomas, A. A. Aboobaker, and M. Blaxter. 2016. No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proceedings of the National Academy of Sciences **113**:5053-5058.

Lamichhaney, S., A. P. Fuentes-Pardo, N. Rafati, N. Ryman, G. R. McCracken, C. Bourne, R. Singh, D. E. Ruzzante, and L. Andersson. 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**:E3452-E3461.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth **9**:357-359.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10**:R25.

Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey. 2013. Software for computing and annotating genomic ranges. PLOS Computational Biology **9**:e1003118.

Lewarch, D. E., L. A. Forsman, S. K. Kramer, L. R. Murphy, L. L. Larson, D. R. Iversen, and A. E. Dugas. 2002. Data recovery excavations at the Bay Street Shell Midden (45KP115), Kitsap County, Washington. Larson Anthropological Archaeological Services Limited.

Limborg, M. T., S. J. Helyar, M. De Bruyn, M. I. Taylor, E. E. Nielsen, R. O. B. Ogden, G. R. Carvalho, F. P. T. Consortium, and D. Bekkevold. 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). Molecular Ecology **21**:3686-3703.

Lok, E. K., D. Esler, J. Y. Takekawa, S. W. De La Cruz, W. S. Boyd, D. R. Nysewander, J. R. Evenson, and D. H. Ward. 2012a. Spatiotemporal associations between Pacific herring spawn and surf scoter spring migration: evaluating a 'silver wave' hypothesis. Marine Ecology Progress Series **457**:139-150.

Lok, E. K., D. Esler, J. Y. Takekawa, S. W. De La Cruz, W. S. Boyd, D. R. Nysewander, J. R. Evenson, and D. H. Ward. 2012b. Spatiotemporal associations between Pacific herring spawn and surf scoter spring migration: evaluating a 'silver wave' hypothesis. Marine Ecology Progress Series **457**:139-150.

Longo, M. S., M. J. O'Neill, and R. J. O'Neill. 2011. Abundant human DNA contamination identified in non-primate genome databases. PLoS ONE **6**:e16410.

Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Molecular Ecology **23**:2178-2192.

Luck, G. W., G. C. Daily, and P. R. Ehrlich. 2003. Population diversity and ecosystem services. Trends in Ecology & Evolution **18**:331-336.

Lynch, Vincent J., Oscar C. Bedoya-Reina, A. Ratan, M. Sulak, Daniela I. Drautz-Moses, George H. Perry, W. Miller, and Stephan C. Schuster. 2015. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. Cell Reports **12**:217-228.

Maddox, D. M., G. B. Collin, A. Ikeda, C. H. Pratt, S. Ikeda, B. A. Johnson, R. E. Hurd, L. S. Shopland, J. K. Naggert, B. Chang, M. P. Krebs, and P. M. Nishina. 2015. A mutation in Syne2 causes early retinal defects in photoreceptors, secondary neurons, and Müller Glia. Investigative Ophthalmology & Visual Science **56**:3776-3787.

Martell, S. J., J. F. Schweigert, V. Haist, and J. S. Cleary. 2012. Moving towards the sustainable fisheries framework for Pacific Pacific Herring: data, models, and alternative assumptions; Stock Assessment and Management Advice for the British Columbia Pacific Herring Stocks: 2011 Assessment and 2012 Forecasts.

Martinez Barrio, A., S. Lamichhaney, G. Fan, N. Rafati, M. Pettersson, H. Zhang, J. Dainat, D. Ekman, M. Höppner, P. Jern, M. Martin, B. Nystedt, X. Liu, W. Chen, X. Liang, C. Shi, Y. Fu, K. Ma, X. Zhan, C. Feng, U. Gustafson, C.-J. Rubin, M. Sällman Almén, M. Blass, M. Casini, A. Folkvord, L. Laikre, N. Ryman, S. Ming-Yuen Lee, X. Xu, and L. Andersson. 2016. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. eLife **5**:e12081.

Mastretta-Yanes, A., N. Arrigo, N. Alvarez, T. H. Jorgensen, D. Piñero, and B. C. Emerson. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular Ecology Resources **15**:28-41.

McClatchie, S., I. L. Hendy, A. R. Thompson, and W. Watson. 2017. Collapse and recovery of forage fish populations prior to commercial exploitation. Geophysical Research Letters:n/a-n/a.

McKechnie, I., D. Lepofsky, M. L. Moss, V. L. Butler, T. J. Orchard, G. Coupland, F. Foster, M. Caldwell, and K. Lertzman. 2014. Archaeological data provide alternative hypotheses on Pacific herring (Clupea pallasii) distribution, abundance, and variability. Proceedings of the National Academy of Sciences.

McKinney, G. J., R. K. Waples, L. W. Seeb, and J. E. Seeb. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Molecular Ecology Resources **17**:656-669.

Meyer, M., M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andrés, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, and S. Pääbo. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science **338**:222.

Migaud, H., A. Davie, and J. F. Taylor. 2010. Current knowledge on the photoneuroendocrine regulation of reproduction in temperate fish species. Journal of Fish Biology **76**:27-68.

Miller, K. M., K. Laberee, A. D. Schulze, and K. H. Kaukinen. 2001. Development of microsatellite loci in Pacific herring (*Clupea pallasi*). Molecular Ecology Notes **1**:131-132.
Mitchell, D., P. McAllister, K. Stick, and L. Hauser. 2008. Sperm contamination in archived and contemporary herring samples. Molecular Ecology Resources **8**:50-55.
Mitchell, D. M. 2006. Biocomplexity and metapopulation dynamics of Pacific herring (*Clupea pallasii*) in Puget Sound, Washington. University of Washington, Seattle, WA.
Moore, J. W., J. D. Yeakel, D. Peard, J. Lough, and M. Beere. 2014. Life-history diversity and its importance to population stability and persistence of a migratory fish: steelhead in two large North American watersheds. Journal of Animal Ecology **83**:1035-1046.
Moran, B. M., and E. C. Anderson. 2018. Bayesian inference from the conditional genetic stock identification model. Canadian Journal of Fisheries and Aquatic Sciences **76**:551-560.
Moss, M. L. 2012. Understanding variability in Northwest Coast faunal assemblages: beyond economic intensification and cultural complexity. The Journal of Island and Coastal Archaeology **7**:1-22.
Moss, M. L. 2016. The nutritional value of Pacific herring: An ancient cultural keystone species on the Northwest Coast of North America. Journal of Archaeological Science: Reports **5**:649-655.
Moss, M. L., A. T. Rodrigues, C. F. Speller, and D. Y. Yang. 2015. The historical ecology of Pacific herring: Tracing Alaska Native use of a forage fish. Journal of Archaeological Science: Reports.
Nielsen, E. E., A. Cariani, E. M. Aoidh, G. E. Maes, I. Milano, R. Ogden, M. Taylor, J. Hemmer-Hansen, M. Babbucci, L. Bargelloni, D. Bekkevold, E. Diopere, L. Grenfell, S. Helyar, M. T. Limborg, J. T. Martinsohn, R. McEwing, F. Panitz, T. Patarnello, F. Tinti, J. K. J. Van Houdt, F. A. M. Volckaert, R. S. Waples, and G. R. Carvalho. 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat Commun **3**:851.
Nielsen, R., J. M. Akey, M. Jakobsson, J. K. Pritchard, S. Tishkoff, and E. Willerslev. 2017. Tracing the peopling of the world through genomics. Nature **541**:302.
Nunn, P. D., R. Hunter-Anderson, M. T. Carson, F. Thomas, S. Ulm, and M. J. Rowland. 2007. Times of Plenty, Times of Less: Last-Millennium Societal Disruption in the Pacific Basin. Human Ecology **35**:385-401.
Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2018. vegan: Community Ecology Package.
Olsen, J. B., C. J. Lewis, E. J. Kretschmer, S. L. Wilson, and J. E. Seeb. 2002. Characterization of 14 tetranucleotide microsatellite loci derived from Pacific herring. Molecular Ecology Notes **2**:101-103.
Palsbøll, P. J., M. Bérubé, and F. W. Allendorf. 2007. Identification of management units using population genetic data. Trends in Ecology & Evolution **22**:11-16.
Paradis, E. 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics **26**:419-420.
Paris, J., R., J. Stevens, R., and J. Catchen, M. 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution **8**:1360-1373.
Pauly, D. 1995. Anecdotes and the shifting baseline syndrome of fisheries. Trends in Ecology and Evolution **10**:430-430.
Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics **28**:2537-2539.
Pearse, P. H. 1982. Turning the tide : a new policy for Canada's Pacific fisheries. The Commission on Pacific Fisheries Policy, Vancouver, B.C.
Petrou, E. L., D. P. Drinan, R. Kopperl, D. Lepofsky, D. Yang, M. L. Moss, and L. Hauser. 2018. Intraspecific DNA contamination distorts subtle population structure in a marine fish: Decontamination of herring samples before restriction-site associated sequencing and its effects on population genetic statistics. Molecular Ecology Resources **0**.
Petrou, E. L., D. P. Drinan, R. Kopperl, D. Lepofsky, D. Yang, M. L. Moss, and L. Hauser. 2019. Intraspecific DNA contamination distorts subtle population structure in a marine fish:

decontamination of herring samples before restriction site-associated (RAD) sequencing and its effects on population genetic statistics. Molecular Ecology Resources **Accepted Author Manuscript**.

Pikitch, E. K., K. J. Rountos, T. E. Essington, C. Santora, D. Pauly, R. Watson, U. R. Sumaila, P. D. Boersma, I. L. Boyd, D. O. Conover, P. Cury, S. S. Heppell, E. D. Houde, M. Mangel, É. Plagányi, K. Sainsbury, R. S. Steneck, T. M. Geers, N. Gownaris, and S. B. Munch. 2014. The global contribution of forage fish to marine fisheries and ecosystems. Fish and Fisheries **15**:43-64.

Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity **95**:536-539.

Platt, T., C. Fuentes-Yaco, and K. T. Frank. 2003. Spring algal bloom and larval fish survival. Nature **423**:398-399.

Poloczanska, E. S., C. J. Brown, W. J. Sydeman, W. Kiessling, D. S. Schoeman, P. J. Moore, K. Brander, J. F. Bruno, L. B. Buckley, M. T. Burrows, C. M. Duarte, B. S. Halpern, J. Holding, C. V. Kappel, M. I. O'Connor, J. M. Pandolfi, C. Parmesan, F. Schwing, S. A. Thompson, and A. J. Richardson. 2013. Global imprint of climate change on marine life. Nature Climate Change **3**:919.

Powell, M. 2012. Divided Waters: Heiltsuk Spatial Management of Herring Fisheries and the Politics of Native Sovereignty. The Western Historical Quarterly **43**:463-484.

Prince, D. J., S. M. O'Rourke, T. Q. Thompson, O. A. Ali, H. S. Lyman, I. K. Saglam, T. J. Hotaling, A. P. Spidle, and M. R. Miller. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Science Advances **3**:e1603198.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics **155**:945-959.

Prüfer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, and S. Pääbo. 2013. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature **505**:43.

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences **94**:9197-9201.

Raymond, M., and F. Rousset. 1995. Genepop version 1.2: population genetics software for exact tests and ecumenicism. Journal of Heredity **86**:248-249.

Reiss, H., G. Hoarau, M. Dickey-Collas, and W. J. Wolff. 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries **10**:361-395.

Rice, C. A., J. J. Duda, C. M. Greene, and J. R. Karr. 2012. Geographic patterns of fishes and jellyfish in Puget Sound surface waters. Marine and Coastal Fisheries **4**:117-128.

Rodrigues, A. T., I. McKechnie, and D. Y. Yang. 2018. Ancient DNA analysis of Indigenous rockfish use on the Pacific Coast: Implications for marine conservation areas and fisheries management. PLOS ONE **13**:e0192716.

Roesti, M., W. Salzburger, and D. Berner. 2012. Uninformative polymorphisms bias genome scans for signatures of selection. BMC Evolutionary Biology **12**:94.

Rogers, L. A., D. E. Schindler, P. J. Lisi, G. W. Holtgrieve, P. R. Leavitt, L. Bunting, B. P. Finney, D. T. Selbie, G. Chen, I. Gregory-Eaves, M. J. Lisac, and P. B. Walsh. 2013. Centennial-scale

fluctuations and regional complexity characterize Pacific salmon population dynamics over the past five centuries. Proceedings of the National Academy of Sciences **110**:1750.
Romero Navarro, J. A., M. Willcox, J. Burgueño, C. Romay, K. Swarts, S. Trachsel, E. Preciado, A. Terron, H. V. Delgado, V. Vidal, A. Ortega, A. E. Banda, N. O. G. Montiel, I. Ortiz-Monasterio, F. S. Vicente, A. G. Espinoza, G. Atlin, P. Wenzl, S. Hearne, and E. S. Buckler. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nature Genetics **49**:476.
Rose, K. A., B. A. Megrey, D. Hay, F. Werner, and J. Schweigert. 2008. Climate regime effects on Pacific herring growth using coupled nutrient-phytoplankton-zooplankton and bioenergetics models. Transactions of the American Fisheries Society **137**:278-297.
Rosenthal, Y., B. K. Linsley, and D. W. Oppo. 2013. Pacific Ocean Heat Content During the Past 10,000 Years. Science **342**:617.
Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources **8**:103-106.
Ruzzante, D., E., S. Mariani, D. Bekkevold, C. André, H. Mosegaard, A. W. Clausen Lotte, G. Dahlgren Thomas, F. Hutchinson William, M. C. Hatfield Emma, E. Torstensen, J. Brigham, E. J. Simmonds, L. Laikre, C. Larsson Lena, J. M. Stet René, N. Ryman, and R. Carvalho Gary. 2006. Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. Proceedings of the Royal Society B: Biological Sciences **273**:1459-1464.
Sakai, T., and N. Ishida. 2001. Circadian rhythms of female mating activity governed by clock genes in *Drosophila*. Proceedings of the National Academy of Sciences **98**:9221.
Savage, A. E., and K. R. Zamudio. 2011. MHC genotypes associate with resistance to a frog-killing fungus. Proceedings of the National Academy of Sciences **108**:16705.
Schindler, D. E., R. Hilborn, B. Chasco, C. P. Boatright, T. P. Quinn, L. A. Rogers, and M. S. Webster. 2010. Population diversity and the portfolio effect in an exploited species. Nature **465**:609-612.
Schmieder, R., and R. Edwards. 2011. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. PLoS ONE **6**:e17288.
Schrimpf, M. B., J. K. Parrish, and S. F. Pearson. 2012. Trade-offs in prey quality and quantity revealed through the behavioral compensation of breeding seabirds. Marine Ecology Progress Series:247-259.
Scribner, K. T., W. H. Lowe, E. Landguth, G. Luikart, D. M. Infante, G. E. Whelan, and C. C. Muhlfeld. 2016. Applications of genetic data to improve management and conservation of river fishes and their habitats. Fisheries **41**:174-188.
Sehn, J. K., D. H. Spencer, J. D. Pfeifer, A. J. Bredemeyer, C. E. Cottrell, H. J. Abel, and E. J. Duncavage. 2015. Occult specimen contamination in routine clinical next-generation sequencing testing. American Journal of Clinical Pathology **144**:667-674.
Shapiro, B., and M. Hofreiter. 2014. A Paleogenomic Perspective on Evolution and Gene Function: New Insights from Ancient DNA. Science **343**:1236573.
Sinclair, M., and M. J. Tremblay. 1984. Timing of Spawning of Atlantic Herring (Clupea harengus harengus) Populations and the Match–Mismatch Theory. Canadian Journal of Fisheries and Aquatic Sciences **41**:1055-1065.
Siple, M. C., and T. B. Francis. 2016. Population diversity in Pacific herring of the Puget Sound, USA. Oecologia **180**:111-125.
Skoglund, P., E. Ersmark, E. Palkopoulou, and L. Dalén. 2015. Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. Current Biology **25**:1515-1519.
Small, M. P., J. L. Loxterman, A. E. Frye, J. F. Von Bargen, C. Bowman, and S. F. Young. 2005. Temporal and Spatial Genetic Structure among Some Pacific Herring Populations in Puget Sound and the Southern Strait of Georgia. Transactions of the American Fisheries Society **134**:1329-1341.

Smith, M. J., C. E. Pascal, Z. A. C. Grauvogel, C. Habicht, J. E. Seeb, and L. W. Seeb. 2011. Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. Molecular Ecology Resources **11**:268-277.
Smith, M. W. 1940. The Puyallup-Nisqually. Columbia University Press, New York,.
Speller, C. F., L. Hauser, D. Lepofsky, J. Moore, A. T. Rodrigues, M. L. Moss, I. McKechnie, and D. Y. Yang. 2012. High Potential for Using DNA from Ancient Herring Bones to Inform Modern Fisheries Management and Conservation. PLoS ONE **7**:e51122.
Star, B., J. H. Barrett, A. T. Gondek, and S. Boessenkool. 2018. Ancient DNA reveals the chronology of walrus ivory trade from Norse Greenland. Proceedings of the Royal Society B: Biological Sciences **285**.
Star, B., S. Boessenkool, A. T. Gondek, E. A. Nikulina, A. K. Hufthammer, C. Pampoulie, H. Knutsen, C. André, H. M. Nistelberger, J. Dierking, C. Petereit, D. Heinrich, K. S. Jakobsen, N. C. Stenseth, S. Jentoft, and J. H. Barrett. 2017. Ancient DNA reveals the Arctic origin of Viking Age cod from Haithabu, Germany. Proceedings of the National Academy of Sciences **114**:9152.
Stein, J. K. 2002. The project: an archaeological perspective. Pages 5-16 *in* J. K. Stein and L. S. Phillips, editors. Vashon Island archaeology: a view from Burton Acres Shell Midden. Burke Musseum, Seattle, Washington.
Stern, D. L. 2013. The genetic causes of convergent evolution. Nature Reviews Genetics **14**:751.
Stick, K. C., and A. Lindquist. 2009. 2008 Washington State herring stock status report. Washington Department of Fish and Wildlife.
Stick, K. C., A. Lindquist, and D. Lowry. 2014. 2012 Washington State herring stock status report. Washington Department of Fish and Wildlife.
Tarpey, C. M., J. E. Seeb, G. J. McKinney, W. D. Templin, A. V. Bugaev, S. Sato, and L. W. Seeb. 2017. SNP data describe contemporary population structure and diversity in allochronic lineages of pink salmon (Oncorhynchus gorbuscha). Canadian Journal of Fisheries and Aquatic Sciences.
Taylor, R. S., and V. L. Friesen. 2017. The role of allochrony in speciation. Molecular Ecology **26**:3330-3342.
Thomas, G. L., and R. E. Thorne. 2001. Night-time predation by Steller sea lions. Nature **411**:1013-1013.
Thomas, G. L., and R. E. Thorne. 2003. Acoustical-optical assessment of Pacific herring and their predator assemblage in Prince William Sound, Alaska. Aquatic Living Resources **16**:247-253.
Thompson, T. Q., M. R. Bellinger, S. M. O'Rourke, D. J. Prince, A. E. Stevenson, A. T. Rodrigues, M. R. Sloat, C. F. Speller, D. Y. Yang, V. L. Butler, M. A. Banks, and M. R. Miller. 2019. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. Proceedings of the National Academy of Sciences **116**:177.
Thornton, T. F. 2015. The ideology and practice of Pacific herring cultivation among the Tlingit and Haida. Human Ecology **43**:213-223.
Tillotson, M. D., and T. P. Quinn. 2018. Selection on the timing of migration and breeding: A neglected aspect of fishing-induced evolution and trait change. Fish and Fisheries **19**:170-181.
Tingting, L., L. Wei, L. Jingyu, L. Hong, L. Yinghui, amp, and Z. Yanyan. 2009. SH2D4A regulates cell proliferation via the ERα/PLC-γ/PKC pathway. BMB Rep. **42**:516-522.
Tsujikawa, M., Y. Omori, J. Biyanwila, and J. Malicki. 2007. Mechanism of positioning the cell nucleus in vertebrate photoreceptors. Proceedings of the National Academy of Sciences **104**:14819.
Vandewoestyne, M., D. Van Hoofstat, S. De Groote, N. Van Thuyne, S. Haerinck, Haerinck S, F. Van Nieuwerburgh, and D. Deforce. 2011. Sources of DNA contamination and decontamination procedures in the forensic laboratory. Journal of Forensic Research **S2**.
Wang, J., L. Raskin, D. C. Samuels, Y. Shyr, and Y. Guo. 2015. Genome measures used for quality control are dependent on gene function and ancestry. Bioinformatics **31**:318-323.
Wang, N., F. Teletchea, P. Kestemont, S. Milla, and P. Fontaine. 2010. Photothermal control of the reproductive cycle in temperate fishes. Reviews in Aquaculture **2**:209-222.
Waples, R. S. 1998. Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. Journal of Heredity **89**:438-450.

Waterman, T. T. 1973. Notes on the ethnology of the Indians of Puget Sound Museum of the American Indian, Heye Foundation, New York.
Weir, B., and C. Cockerham. 1984a. Estimating F-Statistics for the analysis of population structure. Evolution **38**:1358-1370.
Weir, B. S., and C. C. Cockerham. 1984b. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.
West, J. E., S. M. O'Neill, and G. M. Ylitalo. 2008. Spatial extent, magnitude, and patterns of persistent organochlorine pollutants in Pacific herring (Clupea pallasi) populations in the Puget Sound (USA) and Strait of Georgia (Canada). Science of The Total Environment **394**:369-378.
Willing, E. M., C. Dreyer, and C. van Oosterhout. 2012. Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. PLOS ONE **7**:e42649.
Willson, M. F., and J. N. Womble. 2006. Vertebrate exploitation of pulsed marine prey: a review and the example of spawning herring. Reviews in Fish Biology and Fisheries **16**:183-200.
Wolverton, S., and R. L. Lyman. 2012. Introduction to Applied Zooarchaeology. Pages 1-22 *in* S. Wolverton and R. L. Lyman, editors. Conservation Biology and Applied Zooarchaeology. University of Arizona Press, Tuscon.
Womble, J. N., M. Willson, F., M. F. Sigler, B. P. Kelly, and G. VanBlaricom, R. 2005. Distribution of Steller sea lions (*Eumetopias jubatus*) in relation to spring-spawning fish in SE Alaska. Marine Ecology Progress Series **294**:271-282.
Wright, C. A., A. Dallimore, R. E. Thomson, R. T. Patterson, and D. M. Ware. 2005. Late Holocene paleofish populations in Effingham Inlet, British Columbia, Canada. Palaeogeography, Palaeoclimatology, Palaeoecology **224**:367-384.
Wright, S. 1943. Isolation by distance. Genetics **28**:114-138.
Xie, M.-C., C. Ai, X.-M. Jin, S.-F. Liu, S.-X. Tao, Z.-D. Li, and Z. Wang. 2007. Cloning and characterization of chicken SPATA4 gene and analysis of its specific expression. Molecular and Cellular Biochemistry **306**:79-85.
Yamamoto, S., and T. Sota. 2009. Incipient allochronic speciation by climatic disruption of the reproductive period. Proceedings of the Royal Society B: Biological Sciences **276**:2711-2719.
Yang, D. Y., L. Liu, X. Chen, and C. F. Speller. 2008. Wild or domesticated: DNA analysis of ancient water buffalo remains from north China. Journal of Archaeological Science **35**:2778-2785.
Yang, D. Y., and K. Watt. 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. Journal of Archaeological Science **32**:331-336.
Yu, J., K. Lei, R. Xu, T. Xu, Y. Zhuang, M. Han, G. Xu, M. Zhou, and C. M. Craft. 2010. KASH protein Syne-2/Nesprin-2 and SUN proteins SUN1/2 mediate nuclear migration during mammalian retinal development. Human Molecular Genetics **20**:1061-1073.