Population structure, genetic connectivity, and adaptation in the Olympia oyster (*Ostrea lurida*) along the west coast of North America

Katherine Silliman

Committee on Evolutionary Biology, University of Chicago

Email address: ksilliman@uchicago.edu

ABSTRACT

Effective management of threatened and exploited species requires an understanding of both the genetic connectivity among populations and local adaptation. The Olympia oyster (*Ostrea lurida*), patchily distributed from Baja California to the central coast of Canada, has a long history of population declines due to anthropogenic stressors. For such coastal marine species, demographic population structure could follow a continuous isolation-by-distance model, contain regional blocks of genetic similarity separated by barriers to gene flow, or be consistent with a null model of no population structure. To distinguish between these hypotheses in *O. lurida*, 13,444 single-nucleotide polymorphisms (SNPs) were used to characterize rangewide population structure, genetic connectivity, and adaptive divergence. Samples were collected across the species range on the west coast of North America, from southern California to Vancouver Island. A conservative approach for detecting putative loci under selection identified 168 SNPs across 97 loci, which were functionally annotated and analyzed separately from the remaining neutral loci. Strong population structure was observed on a regional scale in both neutral and outlier markers. Geographic regions of reduced gene flow aligned with known marine biogeographic barriers, such as Cape Mendocino, Monterey Bay, and the currents around Cape Flattery. The outlier loci identified as under putative selection included genes involved in developmental regulation, sensory information processing, energy metabolism, immune response, and muscle contraction. These loci are excellent candidates for future research and may provide targets for genetic monitoring programs. Beyond specific applications for restoration and management of the Olympia oyster, this study lends to the growing body of evidence for both population structure and adaptive differentiation across a range of marine species exhibiting the potential for panmixia. Computational notebooks are available to facilitate reproducibility and future open-sourced research on the population structure of *O. lurida*.

Keywords: population genomics, connectivity, outlier loci, Olympia oyster, genotype-by-sequencing, management recommendations

INTRODUCTION

Coastal marine ecosystems provide important services such as carbon sequestration, food production, and recreation (Luisetti et al., 2014), yet contain some of the most exploited and threatened species on earth. As evidence for the direct impacts of human activities (e.g., over-harvesting, increasing atmospheric CO_2, and nutrient run-off) on these species grows, there has been increased focus on restoring depleted abundances, recovering ecosystem services, and determining which species are capable of adapting to environmental change (Granek et al., 2010). Effective management of threatened and exploited species requires an understanding of both the genetic connectivity among populations and adaptation across environmental gradients (Baums, 2008; Miller and Ayre, 2008; Palumbi, 2003). For the numerous coastal marine species with planktonic dispersal, high connectivity among populations can obscure population boundaries and oppose the diversifying effects of natural selection through gene flow (Lenormand, 2002). A growing body of evidence indicates that both limited effective dispersal and local adaptation may be more common in marine species than previously hypothesized (Cowen et al., 2000; Hauser and Carvalho, 2008; Sanford and Kelly, 2011; Weersing and Toonen, 2009).

Neutral molecular markers (e.g., microsatellites) have traditionally been used exclusively to identify the geographic structure of subpopulations and estimate the genetic connectivity between them (Funk et al., 2012); however, these do not give insight into the scale or magnitude of adaptive divergence. For populations connected by dispersal, adaptation to local conditions can still occur if the selective strength overcomes the homogenizing effect of gene flow (Hellberg, 2009). Advances in genomic and computational techniques have facilitated the detection of genomic regions that may be influenced by natural selection ("outlier loci") (Stapley et al., 2010). Although often referred to as adaptive markers, these outliers may only be linked to loci that are under natural selection rather than confer an adaptive advantage themselves. Adaptive markers can be useful for delineating populations when neutral markers show a strict isolation-by-distance (IBD) pattern without clear population boundaries (Drinan et al., 2018), as well as characterizing the adaptive potential of a species or population to future environmental conditions (Eizaguirre and Baltazar-Soares, 2014). Inferences from both neutral and adaptive markers should be combined when making management recommendations (Funk et al., 2012).

The Olympia oyster (*Ostrea lurida*, Carpenter 1864) is an native estuarine bivalve found from Baja California to the central coast of Canada, patchily distributed over strong environmental gradients (Chan et al., 2017; Schoch et al., 2006). Oysters are ecosystem engineers in estuaries, providing structured habitat and removing suspended sediments (zu Ermgassen et al., 2013; Coen et al., 2011). Unlike other oysters where both males and females spawn gametes (e.g., *Crassostrea*), the females fertilize eggs with sperm from the water column and initially brood larvae in the mantle cavity. After release, the larvae have been reported to be planktonic from seven days to eight weeks before settling on a hard substrate (Baker, 1995). The impact of maternal brooding on population structure in Osterideae has not been examined.

Following devastating commercial exploitation in the 19th and early 20th centuries, recovery of Olympia oyster populations has been stifled by other anthropogenic threats (e.g., water quality issues, habitat loss, and possibly ocean acidification (Blake and Bradbury, 2012; Hettinger et al., 2013; Sanford et al., 2014). The last 15 years has seen increased interest in the Olympia oyster, with restoration projects underway by both government and nongovernment agencies across its range (Pritchard et al., 2015). Current knowledge about the population genetic structure of *O. lurida* comes primarily from an unpublished 2011 dissertation, which sampled from San Francisco, CA to Vancouver Island, BC and found regional population structure using microsatellites (Stick, 2011). Two phylogeographic studies using two mitochondrial loci identified a phylogeographic break north of Willapa Bay, WA and established the southern boundary divide between *O. lurida* and its sister species *Ostrea conchaphila* (Polson et al., 2009; Raith et al., 2016). Future and ongoing management plans would benefit greatly from thorough analysis of the fine-scale genetic structure using modern genomic techniques and rangewide sampling (Camara and Vadopalas, 2009).

The objective of this study was to characterize the spatial population structure of the Olympia oyster across the majority of its range using both neutral and adaptive markers derived from genome-wide single nucleotide polymorphisms (SNPs). I specifically tested whether patterns of genetic variation suggest a smooth continuum of allele frequency shifts consistent with isolation-by-distance (IBD) (Malécot, 1968), regional blocks of genetic similarity that correspond to physical barriers (Hare and Avise, 1996), or the null model of no significant genetic differentiation (Grosberg and Cunningham, 2001). SNPs produced from high-throughput sequencing have led to the identification of previously undetected population structure in a number of marine and terrestrial species (Everett et al., 2016; Reitzel et al., 2013; Van Wyngaarden et al., 2016). Compared to the Atlantic coast of North America (Hoey and Pinsky, 2018), studies utilizing genome-wide SNPs for marine taxa from the Pacific coast are far fewer in number and have been limited to regional spatial scales (De Wit

and Palumbi, 2013; Drinan et al., 2018; Gleason and Burton, 2016; Larson et al., 2014; Martinez et al., 2017) or in the number of sampling sites (Pespeni et al., 2010; Tepolt and Palumbi, 2015), making this study the first of my knowledge to utilize thousands of SNPs to extensively survey the rangewide population structure of a marine species along this coast.

A secondary aim of this study was to produce a reproducible computational pipeline based on Jupyter notebooks to go from raw data to results and figures(Kluyver et al., 2016). A major issue for genomic analyses today is how to clearly explain the computational methods used in order to allow for reproducibility (Kanwal et al., 2017). This open access pipeline is intended to provide an example template to improve reproducibility in future studies and function as an instructional tool for biologists and early career scientists who wish to apply these methods to their own study organism.

MATERIALS AND METHODS

Sample collection

20-25 adult *Ostrea lurida* were collected primarily from the intertidal by hand at 20 sites ranging from Klaskino Inlet, Vancouver Island (50°17' 55") to San Diego Bay, CA (32°361' 9") in 2014 (Figure 1, Table A1). Each site represents a separate bay, except for Willapa Bay, WA and San Francisco Bay, CA which had two sampling sites each. Adductor tissue samples were preserved in RNALater, followed by storage in -80°C. DNA was isolated using DNeasy Blood & Tissue kits (Qiagen) and E.Z.N.A. Mollusc DNA kits (Omega Bio-Tek) with RNAse-A treatment following manufacturer instructions and limiting tissue digestion time to no more than 90 minutes. All DNA samples were quantified using the Qubit dsDNA BR Assay kit (Life Technologies) on a Qubit v2.0 (Life Technologies). DNA quality was verified by agarose gel electrophoresis of 1-2 uL extracted DNA on an 0.8% TAE gel.

Genotype-by-sequencing analysis

Library preparation for genotype-by-sequencing (GBS) followed the protocol by Elshire et al. (2011) using the *ApeKI* restriction enzyme, with an additional size selection step and slight modifications to PCR amplification. A detailed protocol can be found at [redacted]. All libraries were size-selected for fragment sizes between 200bp-450bp on a Blue PippinPrep (Sage Science) to reduce the number of loci sequenced and ensure adequate sequencing coverage. One library of 90 samples was sequenced across two 100bp paired-end Illumina HiSeq 2500 lanes, with only the forward sequencing read used for analysis. Seven other libraries with a maximum of 48 samples each (332 total, including 63 replicates) were sequenced on eight 100bp

single-end lanes. Sampling sites were spread out across all libraries in order to minimize batch effects from library preparation and sequencing.

Raw sequencing reads were de-multiplexed, quality filtered, and de novo clustered using the API version of the seven step computational pipeline *ipyrad* v.0.7.24 (Eaton, 2014) and implemented in Python via a Jupyter Notebook running on a large computational cluster. Demultiplexing (Step 1) used sample-specific barcode sequences, allowing one mismatch in the barcode sequence. Base calls with a Phred quality score under 20 were converted to Ns, and reads containing more than 5 Ns were discarded. Adapter sequences, barcodes, and the cutsite sequences were trimmed from filtered reads, with only reads greater than 35bp retained (Step 2). Reads were then clustered using a sequence similarity threshold of 85% both within (Step 3) and between samples to genotype polymorphisms (Steps 4, 5) and identify orthologous loci (Step 6) with a minimum of 10x read coverage. Replicate samples were assembled separately and then compared using custom Perl scripts by Mikhail Matz (Wang et al., 2012). The replicate with the largest number of loci after final filtering (Step 7) was retained. Samples were iteratively pruned throughout the analysis to remove those with low numbers of sequencing reads or assembled loci, with Steps 4-7 rerun using the remaining individuals. One of the sampling sites for Willapa Bay, WA had a low number of individuals after filtering, so individuals from these two sites were combined into one population, for 19 total populations.

The final assembly was then filtered for excess heterozygosity based on deviations from Hardy-Weinberg equilibrium in at least two populations, sample coverage of 75%, and an overall minor allele frequency of 2.5%, retaining only loci found in at least one individual from all populations. Preliminary analyses conducted on datasets with more or less missing data allowed showed that the inferred population structure was robust to missing data up to 40%.Population genetic summary statistics, with the exception of F_{ST}, did change quantitatively due to missing data but not qualitatively (not shown) (Cariou et al., 2016). Filtering steps were conducted using VCFtools (Danecek et al., 2011), custom Python code, and code adapted from Jon Puritz's lab (Puritz et al., 2014). Input files and formats for subsequent analysis of population structure were created using a combination of custom Python code, custom R code, and the *radiator* R package (Gosselin, 2017). Every step of the assembly, filtering process, and creation of input files can be reproduced through Jupyter notebooks.

Detection of loci under putative selection

Following recommendations to utilize multiple methods to detect loci under putative directional selection (Benestan et al., 2016; Rellstab et al., 2015), three approaches were used on the filtered SNP dataset: BayeScan

v.2.1, *OutFLANK* v.0.2, and *pcadapt* v.4.0.2. For BayeScan and *OutFLANK*, individuals were grouped into populations by sampling site. GBS loci which had SNPs identified as outliers in at least two of the approaches were classified as putative adaptive loci. From these loci, any SNP that had been identified as an outlier by at least one approach was separated from the full SNP dataset to create an "outlier" dataset. Subsequent analyses of population structure were conducted on three SNP datasets: all SNPs (combined), outliers, and neutral SNPs- which excluded any SNP found on a putative adaptive locus.

BayeScan uses a Bayesian approach to apply linear regression to decompose F_{ST} coefficients into population- and locus-specific components and estimates the posterior probability of a locus showing deviation from Hardy–Weinberg proportions (Foll and Gaggiotti, 2008). BayeScan analysis was based on 1:100 prior odds, with 100,000 iterations, a burn-in length of 50,000, a false discovery rate (FDR) of 0.05, and default parameters. Results were visualized in R. *OutFLANK* is an R package that identifies F_{ST} outliers by inferring a distribution of neutral F_{ST} using likelihood on a trimmed distribution of F_{ST} values. Because of its likelihood method, *OutFLANK* calculates F_{ST} without sample size correction when inferring the neutral distribution. Simulation studies have shown that this approach has lower false positive rates compared to other F_{ST} outlier methods (Whitlock and Lotterhos, 2015). *OutFLANK* was run using default parameters and a *q*-value threshold of 0.1, which can be considered a false discovery rate (FDR) of 10%. For the R package *pcadapt*, individuals are not sorted into predefined populations. Instead, *pcadapt* ascertains population structure using principal component analysis (PCA), then identifies markers under putative selection as those that are excessively correlated with population structure. When compared to BayeScan, *pcadapt* was shown to have greater power in the presence of admixed individuals and when population structure is continuous (Luu et al., 2017) — both scenarios which are likely in *O. lurida*. A scree plot representing the percentage of variance explained by each PC was used to choose the number of principal components (K) for *pcadapt*, and SNPs with a *q*-value threshold of 0.05 were categorized as outliers.

Outlier loci were functionally annotated through Blast2GO. Sequences were compared against molluscan sequences in GenBank's *nr* database using the BLASTx algorithm with default parameters and a *e*-value hit filter of 10^{-3}, and against EMBL-EBI InterPro signatures. Gene ontology terms were mapped to annotations with default parameters except for an *e*-value hit filter of 10^{-3} (Götz et al., 2008). Minor allele frequency was plotted against latitude individually for each outlier SNP in order to identify clinal patterns of allele frequency shifts.

Summary statistics, population differentiation, and spatial structure

Population genetic summary statistics were calculated on the combined, neutral, and outlier datasets to describe and compare overall and population-specific genetic diversity. Observed heterozygosity (H_o), expected heterozygosity (H_e), overall F_{ST}, and F_{IS} were calculated using the *basic.stats* function in the R package *hierfstat* (Goudet and Jombart, 2015). Confidence intervals for population-specific F_{IS} were determined using the *boot.ppfis* function in *hierfstat* with 1,000 bootstrap replicates. Pairwise F_{ST} following Weir and Cockerham (1984) was calculated using the *genet.dist* function in *hierfstat*. Heatmaps of pairwise F_{ST} values were created using *ggplot2* (Wickham, 2016). A Mantel test of coastal water distance (calculated by drawing routes between all sites on Google Earth) and $F_{ST}/1 - F_{ST}$ as implemented in *adegenet* tested for evidence of isolation-by-distance (Sokal, 1979).

Rangewide population structure of *O. lurida* was characterized using a combination of Bayesian clustering and multivariate ordination approaches. These methods were applied to both the outlier and neutral datasets. The model-based Bayesian clustering method STRUCTURE v.2.2.4 (Pickrell and Pritchard, 2012) as implemented in the *ipyrad* API was used to determine the number of distinct genetic clusters (K) with a burn-in period of 50,000 repetitions followed by 200,000 repetitions. Five replicate analyses were performed for each dataset with values of K = 1-10, with each replicate independently subsampling one SNP per GBS locus and using a different random seed. Replicates were summarized and visualized using the CLUMPAK server (Kopelman et al., 2015). The δK method implemented in STRUCTURE HARVESTER was used to determine an optimal K (Earl, Dent A. and vonHoldt, Bridgett M, 2012). PCA was implemented in the R package *adegenet* (Jombart and Ahmed, 2011) using "unlinked" datasets, where a single SNP with the least missing data across samples was chosen for each GBS locus (or the first SNP in the locus in the case of a tie). Missing data was filled by randomly drawing an allele based on the overall allele frequency across all individuals. The R package *PCAviz* was used to visualize PCA results and correlate PC loadings with latitude (Novembre et al., 2018). Results from STRUCTURE, PCA, and pairwise F_{ST} were used to identify phylogeographic "regions". Summary statistics, including H_o, H_e, F_{IS}, and F_{ST} were calculated for each region using using the *basic.stats* function in the R package *hierfstat* (Goudet and Jombart, 2015).

Estimating connectivity and historical relationships

Spatial variation in gene flow and genetic diversity was calculated and visualized using the program EEMS (Estimated Effective Migration Surfaces) (Petkova et al., 2016). This method identifies geographic regions

where genetic similarity decays more quickly than expected under isolation-by-distance based on sampling localities and a pairwise genetic dissimilarity matrix derived from SNP data. These regions may be interpreted as having reduced gene flow. A dissimilarity matrix was calculated for each of the three datasets using a variant of the *bed2diffs* R code included in the EEMS package that takes input from a *genind* R object. An outer coordinate file for defining the potential habitat of *O. lurida* was produced using the polyline method in the Google Maps API v3 tool (http://www.birdtheme.org/useful/v3tool.html). The habitat shape followed the shape of the coastline and excluded land regions that *O. lurida* larvae would not naturally be able to cross (e.g., the Olympic peninsula separating outer coast populations and those in Puget Sound, WA). The EEMS model is recommended to be run for various numbers of demes, which establishes the geographic grid size and resulting potential migration routes. Three independent analyses were run for each deme size (200, 250, 300, 350, 400, 500, 600, 700) for a total of 24 runs, with a burn-in of 1,000,000 and MCMC length of 5,000,000 iterations. The convergence of runs was visually assessed and results were combined across all analyses and visualized using the *Reemsplots* R package — producing maps of the effective diversity (q) and effective migration rate (m). To infer the evolutionary relationship among sampling sites, including population splits and migration events, I reconstructed population graph trees using the software TreeMix (Pickrell and Pritchard, 2012). The population graph was rooted with the two southernmost *O. lurida* population (San Diego, CA and Mugu Lagoon, CA), then run allowing between 0-10 migration events. For each value of migration events, I calculated the proportion of variance in relatedness between populations that is explained by the population graph to evaluate model fit (Wang, 2017).

RESULTS

GBS and outlier detection

After iterative pruning of individuals with low sequencing depth and a low proportion of shared loci, 42,081 biallelic SNPs were retained across 9,824 GBS loci genotyped in greater than 75% of 117 individuals (2.8% of prefiltered loci assembled by *ipyrad*). Almost all replicate pairs showed $> 88\%$ similarity in heterozygous calls except for one, in which case both replicates were dropped. Average read depth per individual per GBS locus ranged from 21 to 120 (mean = 32 ± 14). Further filtering by HWE and MAF $> 2.5\%$ reduced the dataset to 13,444 SNPs across 6,192 loci (the ”combined” dataset).

Three different methods were employed to identify putative SNPs under selection. The number of outliers detected by each program and the overlap between programs is illustrated in Figure 2. *OutFLANK*, as the most

conservative of the programs used (Whitlock and Lotterhos, 2015), had the lowest number of outlier markers detected with 31 SNPs across 16 GBS loci. Twenty-four SNPs found across 15 loci were identified as outliers by all three programs. 97 GBS loci contained SNPs identified as outliers by at least two approaches, with 168 SNPs included in the outlier dataset for subsequent population structure analyses. The neutral dataset, with 13,192 SNPs across 6,094 loci, excluded any SNP found on a locus with an outlier SNP.

Summary statistics, population differentiation, and spatial structure

Summary statistics

Global F_{ST} for outliers (F_{ST} = 0.448) was almost five times greater than for the combined and neutral SNPs (F_{ST}= 0.1051 (combined), 0.099 (neutral)). H_o was lowest in the outlier dataset, while H_e was the highest in the outliers (Table 1). Average F_{IS} within populations for the combined dataset was 0.0424, with all populations having a significantly positive F_{IS} value except Ladysmith, BC, Tomales Bay, CA, and South San Francisco Bay, CA which had small, yet significantly negative F_{IS} values. Mugu Lagoon had the highest F_{IS} value (Table A1). Summary statistics for the six phylogeographic regions identified in the following section are show in Table 2. Summary statistics were quantitatively very similar for the combined and neutral datasets, so that only the results for the outlier and neutral datasets are reported for all subsequent analyses.

Spatial structure

Bayesian population structure analysis from STRUCTURE differed slightly between the neutral and outlier datasets. For neutral SNPs, K = 5 had the strongest support based on the Evanno method. Visual inspection of STRUCTURE admixture plots for K = 6 included an additional population grouping of Willapa Bay, WA and Coos Bay, WA that was further supported by PCA and F_{ST} results (Figure 3). I refer to these six groupings as phylogeographic "regions": Northwest Vancouver Island, BC (*NWBC*), Puget Sound, WA plus Victoria, BC and Ladysmith Harbour, BC (*Puget+BC*), Willapa Bay, WA plus Coos Bay, OR (*Willapa*), the other two Oregon sites (*Oregon*), Northern California (*NoCal*) from Humboldt Bay, CA to San Francisco Bay, CA, and Southern California (*SoCal*) from Elkhorn Slough, CA to San Diego Bay, CA. STRUCTURE results for the outlier SNPs supported K = 2, but visually results were similar between the outlier and neutral SNPs at K = 5 with the exception of Discovery Bay, WA in Puget Sound showing higher admixture with *NWBC* (Figure 3). The separation of *Willapa* sites from *Oregon* was not observed with the outlier dataset until K = 8.

PCA on both SNP datasets demonstrated a strong relationship between latitude and the first principal component (PC) (neutral: R^2 = 0.857, outlier: R^2 = 0.827) (Figure 4). PCs 2-5 in the neutral dataset separated

out individuals by phylogeographic region, with PC2 separating (*Puget+BC*, *NWBC*) and (*Willapa*, *Oregon*), PC3 separating *NoCal* and *SoCal*, PC4 separating *NWBC* and *Puget+BC*, and PC5 separating *Oregon* from *Willapa* (Figure 3). PC1 of the neutral dataset represented 5.8% of the total variance and PCs 2-5 represented 2.5%-1.5% of the variance. The outlier dataset showed regional spatial structure for PCs 2-4 (Figure 3). PC1 of the outlier dataset represented 22.8% of the total variance, and PCs 2-4 represented 11.9%-4.8% of the total variance.

Population differentiation and isolation-by-distance

Pairwise population-specific F_{ST} was higher for outlier loci, but both datasets qualitatively illustrated roughly six geographic "regions" of genetically similar populations and an overall trend of isolation-by-distance, where F_{ST} values were higher between sites that were farther away from each other (Figure 5). The three comparisons with the lowest pairwise F_{ST} using the neutral dataset were Willapa Bay, WA/Coos Bay, OR (0.00159), Mugu Lagoon, CA/San Diego Bay, CA (0.00126), and North/South San Francisco Bay, CA (0.0038). Victoria, BC showed higher pairwise F_{ST} with the other British Columbia sites than with sites from Puget Sound, WA, particularly in the outlier dataset. Pairwise F_{ST} values from the outlier dataset identified stronger relative population differentiation for some sites than was observed with the neutral dataset. For example, the two sites from San Francisco had one of the lowest F_{ST} values using neutral loci, but had an F_{ST} of 0.084 using outlier loci. The two southernmost sites in California, San Diego Bay and Mugu Lagoon, showed much higher relative pairwise F_{ST} values with sites north of California in the outlier dataset compared to the neutral dataset. Pairwise F_{ST} is similarly visualized for the six phylogeographic regions and Ladysmith Harbour, BC (Figure 6). Mantel tests showed a significant correlation between pairwise F_{ST} and coastal water distance for both datasets, indicating a strong trend of isolation-by-distance (p-value = 0.001) (Figure 7).

Connectivity and historical relationships

TreeMix produced a population graph that supported a major phylogeographic split between the outer coast of Washington and Puget Sound. When allowing for an increasing number of migration events, the proportion of variance in relatedness between populations explained by the model began to asymptote at 0.991 for 7 migration edges (Figure 8). All of these migration events involve either the *Puget+BC* or *Willapa* regions, except for one from a *NoCal* population to a *SoCal* population. Because the Coos Bay, OR population is likely a recent anthropogenic introduction from Willapa Bay, Coos Bay oysters were excluded from the EEMS analysis. The combined EEMS map for all runs identified four significant (posterior probability $> 95\%$)

barriers to gene flow: 1) at the mouth of the Strait of Juan de Fuca, 2) around Victoria, BC, 3) extending from Willapa Bay, WA to southern Oregon, and 4) around San Francisco Bay, CA (Figure 9). These inferred barriers further lend credence to the 6 phylogeographic regions identified through other means. An area of significantly increased gene flow was inferred between Mugu Lagoon and San Diego. The EEMS method also estimated and mapped the genetic diversity parameter q, which is an estimate of the expected within-deme coalescent time and is proportional to average heterozygosity (H_e). Populations from Oregon northwards had much lower genetic diversity than those in California. A linear regression of population-specific H_e and latitude using the neutral dataset shows a strong relationship between genetic diversity and latitude (R^2 = 0.859)(Figure 10), while the outlier dataset showed a similar, yet weaker relationship (R^2 = 0.759).

Functional annotation of outliers

The 97 loci containing outlier SNPs were functionally annotated using Blast2GO. Fourteen of these mapped to protein coding genes in the GenBank database, primarily from *Crassostrea virginica* and *Crassostrea gigas*. One mapped to the *O. lurida* mitochondrial NADH dehydrogenase subunit 5 gene (*nad5*), which exhibits high variability in oyster species and is commonly used for metazoan phylogenetics (Xiao et al., 2015). Annotated genes have potential roles in developmental regulation (glyoxalase 3, DNA N6-methyl adenine demethylase-like, transcriptional regulator ERG, serine/threonine-protein kinase), sensory information processing (serine/threonine-protein kinase, sodium-dependent phosphate transport), osmoregulation (sodium-dependent phosphate transport), immune or stress response (*nad5*, E3 ubiquitin-protein ligase), energy metabolism (carnitine palmitoyltransferase, glucose dehydrogenase), heavy metal binding (Heavy metal-binding HIP), and muscle contraction (myosin heavy chain-striated muscle, myosin-XVIIIa) (Table 3) Epelboin et al. (2016); Szent-Györgyi et al. (1999); Li et al. (2017); Cheng et al. (2016); Anderson et al. (2015); Riviere et al. (2013); Pauletto et al. (2017); Wang et al. (2018). Sixteen additional outlier loci had positive matches to InterPro signatures without any BLASTx hits or gene ontology annotation. Plotting minor allele frequency against latitude for outlier loci demonstrates that the majority of outliers show a clinal pattern, where one allele is fixed from either Coos Bay, OR or San Francisco Bay, CA to the north, and the other alelle increases in frequency towards the south (Figure 11). In 29 of these cases, different alleles are fixed in at least one *SoCal* population and at least one *Puget+BC* population.

DISCUSSION

Effective management of threatened and exploited species requires an understanding of both the genetic connectivity between populations and the adaptive population structure. Using 13,444 SNPs genotyped in 117 Olympia oysters from California to British Columbia, I reject both a null hypothesis of panmixia and a strict pattern of isolation-by-distance. Instead, I detect barriers to effective migration resulting in six distinct phylogeographic regions. 97 outlier loci were identified and functionally annotated, indicating candidate loci involved in local adaptation. Comparisons of combined and outlier markers for population structure analyses demonstrates that, while neutral markers may have greater power to detect fine scale population structure, outlier loci can provide evidence for adaptive divergence among otherwise homogeneous populations.

Regional population structure and gene flow

Significant population structure was observed across the range of *O. lurida* in both the neutral and outlier loci, separating sampling locations into six distinct regions. Notably, these regions fit well within previously described biogeographical provinces built around marine species distributions (Hall, 1964; Valentine, 1966; Fenberg et al., 2015), with some exceptions. As region-specific results are important for local management decisions, I describe the evidence for and genetic characteristics of each region separately below. See the section on *Anthropogenic influences on population structure* for a discussion on how this inferred population structure may have been influenced by human activities.

Southern California (SoCal)

The *SoCal* region, containing Elkhorn Slough, CA, Mugu Lagoon, CA, and San Diego Bay, CA, extends across both the Southern Californian and the Montereyan biogeographic provinces as defined by Hall (1964), with Monterey Bay as the northern boundary and Punta Eugenia, Baja California as the southern boundary. Monterey Bay is a known biogeographic barrier for some marine algae (Abbott and Hollenberg, 1976), and has been proposed as a potentially important barrier to gene flow in marine invertebrates as well (Dawson, 2001). Mitochondrial markers have established Punta Eugenia as the boundary between *O. lurida* and its sister species *O. conchaphila* to the south (Raith et al., 2016), suggesting that populations between San Diego Bay and Punta Eugenia may also belong to the *SoCal* region. This region extends across Point Conception, which is a well-known site of species turnover (Valentine, 1966) and a phylogeographic barrier for some taxa (Marko, 1998; Wares et al., 2001). Meta-analyses of numerous marine invertebrates have demonstrated that strong population structure across Point Conception is the exception rather than the rule(Kelly and Palumbi,

2010; Dawson, 2001), which may be the case for *O. lurida* as well. Finer-scaled population genomic sampling of populations on either side of Point Conception may be required to detect slight genetic clines.

The high inferred migration rate between Mugu Lagoon and San Diego, coupled with low F_{ST} at neutral loci, suggests that there is either significant gene flow between these two populations, or that one of them is a recent transplant from the other. In the latter case, I hypothesize that Mugu Lagoon is the recent transplant as it had the highest inbreeding coefficient (F_{IS}) of any population, which may suggest that the oysters sampled from this site had shared ancestry. Nevertheless, a pairwise F_{ST} of 0.045 at outlier loci suggests that some adaptive population divergence has potentially occurred between these populations. The *SoCal* region exhibited the highest genetic diversity of any region, for which I propose three potential, nonexclusive mechanisms. 1) The southward direction of the California beginning at Cape Flattery, WA results in asymmetric gene flow and an accumulation of genotypes in the south (Wares et al., 2001). This hypothesis is supported by the inferred directionality of migration events in TreeMix (Figure 8). 2) Ongoing or historical admixture from the southern sister species *O. conchaphila* has increased genetic diversity in these populations. Sampling and genotyping of *O. conchaphila* is underway to test this hypothesis. 3) Northern populations demonstrate lower genetic diversity due to repeated extirpation or population bottlenecks from glaciation cycles (see *Puget+BC*).

Northern California (NoCal)

San Francisco Bay, Tomales Bay, and Humboldt Bay constitute the *NoCal* region, which is encompassed by the northern half of the Montereyan biogeographic province as identified by (Fenberg et al., 2015) and delineated by Cape Mendocino to the north. Cape Mendocino is 46 km south of Humboldt Bay and an established phylogeographic break for multiple marine species (Kelly and Palumbi, 2010). Overall F_{ST} for this region (neutral: 0.022, outlier: 0.140) is the highest after *Puget+BC*, indicating that there is still significant population structure within the *NoCa*l region. The two sites within San Francisco Bay (Candlestick Park and Point Orient), exhibit potential adaptive divergence at some outlier loci despite high potential for gene glow (e.g., locus_200167, locus_219420, locus_270303). These results support evidence for local adaptation from reciprocal transplant studies within San Francisco Bay (Bible and Sanford, 2016). EEMS identifies an area of significantly reduced gene flow surrounding San Francisco Bay, which may correspond to Monterey Bay, or anthropogenic influences (see *Anthropogenic influences on population structure*). Due to the elevated population structure within this region, individual pairwise F_{ST} values (Table B1, B2) should be taken into account when making restoration decisions (e.g., avoid transplantation projects between Humboldt and Tomales Bay (F_{ST} = 0.042 netutral, 0.285 outlier). Results from STRUCTURE and PCA with the

neutral dataset show Elkhorn Slough, CA as an intermediate population between *NoCal* and *SoCal*. Inferred evolutionary relationships with TreeMix and lower population differentiation at outlier loci support the inclusion of Elkhorn Slough in *SoCal*, along with evidence of migration from San Francisco Bay into Elkhorn Slough. While a strong migration event is inferred from San Francisco Bay and Elkhorn Slough by TreeMix, it does not indicate the frequency or timing of this migration. Migration is not likely to be consistent between these populations, as a study of recruitment dynamics in Olympia oysters observed recruitment failure in Elkhorn Slough despite successful recruitment in San Francisco Bay (Wasson et al., 2016).

Oregon and Willapa

Both the *Oregon* region, comprised of Netarts Bay and Yaquina Bay, and the *Willapa* region with Willapa Bay, WA and Coos Bay, OR, fall within the Mendocinian biogeographic province, which is usually demarcated by either Cape Flattery (Blanchette et al., 2008) or Vancouver Island (Fenberg et al., 2015) to the north. Strong evidence from TreeMix and Structure indicate that these two regions have a shared phylogeographic history — likely a combination of evolutionary and anthropogenic processes. EEMS robustly infers an area of significantly reduced migration from Willapa Bay, WA to southern Oregon, which I hypothesize is partly due to the high retention of oyster larvae within Willapa Bay during the summer reproductive season (Banas et al., 2009). EEMS also infers an area of slightly increased migration to the west of the sampling sites. I interpret this result as the EEMS model attempting to incorporate evidence for long-range migration events, likely anthropogenic in nature, between *Puget+BC* sites and sites in *Oregon* and *NoCal* (e.g. North Bay, WA to Humboldt Bay, CA). The phylogeographic divide between Willapa Bay, WA and Puget Sound, WA has significant implications for management of Olympia oysters within Washington State. Although Washington Dept. of Fish and Wildlife maintains strict oversight of shellfish translocations in the state, currently the protections against importing shellfish from outside of the state are higher than moving shellfish within the state. The results presented here indicate that transfer between the outer coast of WA and Puget Sound of Olympia oysters or *Crassostrea* shells that may harbor Olympia oysters, should be considered equivalent to importing oysters from out of state.

Puget Sound, WA and British Columbia (Puget+BC and NWBC)

Due to the complex oceanographic setting of Puget Sound, WA and Vancouver Island, BC compared to the near linearity of the outer coast, it is not surprising that describing the population structure in these regions is less straightforward. The *NWBC* region, comprised of Klaskino Inlet, BC and Barkeley Sound, BC, is

significantly differentiated from other sites on Vancouver Island and shows evidence for increased migration within the region and sharply decreased migration out of the region. Management practices of Olympia oysters in British Columbia should recognize the distinctiveness of these northwestern sites, particularly for restoration. Results from TreeMix, Structure, and PCA clearly identify all sites from Puget Sound and Victoria Gorge, BC as having a shared evolutionary history, however Victoria Gorge has elevated levels of pairwise FST between Puget Sound sites at both neutral and outlier SNPs. EEMS confirms the likelihood of reduced migration between Victoria Gorge and Puget Sound. Ladysmith Harbour, BC is similiar to Elkhorn Slough, CA in that it appears intermediate between the *NWBC* and *Puget+BC* regions in the STRUCTURE, PCA, and TreeMix analyses. The decision to include Ladysmith Harbour, BC in *Puget+BC* is based on pairwise F_{ST} values at outlier loci, particularly between Ladysmith Harbour and Discovery Bay, WA (F_{ST} = 0.085). It is possible that Ladysmith Harbour belongs to a separate phylogeographic region all together. Genetic sampling from additional sites on the central coast of British Columbia and eastern coast of Vancouver Island would be required to test this hypothesis.

Although Cape Flattery and Puget Sound itself have both been classified as biogeographic barriers due to a bifurcation in ocean currents (Valentine, 1966; Kelly and Palumbi, 2010), there are surprisingly few studies evaluating the genetic structure of species found both within Puget Sound and on the outer coast of Washington. The majority of these focus on rockfish (Buonaccorsi et al., 2002; Cunningham et al., 2009; Iwamoto et al., 2015; Siegle et al., 2013), and recently Dungeness crab (Jackson and O'Malley, 2017), both of which have much longer dispersal stages relative to the Olympia oyster. To my knowledge, this is the first study in a marine mollusc to evaluate and identify significant population differentiation among Puget Sound populations and the outer coast. More studies are required to fully characterize the importance of this barrier across marine taxa.

Genetic differentiation within Puget Sound is relatively low at both neutral and outlier loci, with the exception of the northernmost site, Discovery Bay, which was more genetically similar to British Columbia populations. The weak population structure within Puget Sound and the overall low genetic diversity in northern sites is likely due to recent genetic bottlenecks and range expansion after the last glacial maximum, which reached just north of Willipa Bay, WA (49°N latitude) until 12-13 kya (Dyke and Prest, 1987). Despite such low genetic differentiation, experimental assessments of local adaptation for populations within Puget Sound have detected heritable differences in fitness traits such as reproductive timing, growth rate, and gene expression in response to stress (Heare et al., 2017, 2018; Silliman et al., 2018). These results, coupled with

experimental evidence for local adaptation to salinity among Northern California populations (Bible and Sanford, 2016), suggest that adaptive divergence in this species can occur in the face of ongoing gene flow.

Anthropogenic influences on population structure

The evidence for reduced effective migration, low differentiation within most of the phylogeographic regions, and external estimates of effective dispersal (Carson, 2010), suggests that long distance dispersal is not a significant force in shaping population structure in this species. However, TreeMix inferred a few such migration events that cross aforementioned barriers to gene flow and some pairwise comparisons of population differentiation are less than expected given the distance between sites. To explain this evidence, I investigated the history of Olympia oyster exploitation and aquaculture through literature reviews, technical reports, grey literature, historical first-person accounts, and discussions with current restoration practitioners. The historical impact of human take and transportation on the Olympia oyster is substantial.

Washington State has been the hub of the Olympia oyster fishery since the early 1800s, and currently contains some of the largest extant populations (although still considerably below pre-1850 levels). Oysters were shipped from Willapa Bay to northern California by the millions beginning in 1850, including shipments of juvenile or "seed" oysters to be outplanted in local waters until reaching commercial size. Oakland Bay in south Puget Sound, about 30 km coastal distance from the sampled North Bay, was used to produce seed oysters primarily for the rest of Puget Sound as well as northern California (Woelke, 1959; Baker, 1995). This direct movement of oysters between these Washington sites and northern California may explain the inferred migration events to Humboldt Bay, CA and San Francisco Bay, CA from North Bay, WA and the *Willapa* region.

After the decimation of the Olympia oyster industry, the non-native oysters *Crassostrea virginica* and *Crassostrea gigas* were brought to the west coast, with the latter establishing self-sustaining populations in Willapa Bay by the 1950s. The shells of these species provide excellent substrate for the establishment of Olympia oysters, and the movement of *Crassostrea* oysters between bays for culturing purposes (e.g., San Francisco to Humboldt in 1910, Willapa to Humboldt in 1950s (Barrett, 1963)) might have resulted in the accidental transfer of Olympia oysters, which were considered a "fouling" organism of the larger, more robust *Crassostrea* species (Townsend, 1895). The exceptionally low F_{ST} of Willapa Bay and Coos Bay, separated by 415 km, is one of the most striking cases of human-mediated migration and confirms the theory that Olympia oysters were imported accidentally from Willapa Bay on *C. gigas* shells in the 1980s (Baker et al., 1999). The inbreeding coefficient for Coos Bay is nevertheless low, suggesting that either the transplantation included a

large number of founders or significant time has passed for genetic diversity to be influenced by migration and mutation. Future movement of *Crassostrea* for aquaculture purposes among bodies of water along the west coast should be carefully monitored to prevent the accidental migration of nonnative *O. lurida* genotypes.

While it is encouraging that programs such as TreeMix can recover known human-mediated migration events, such artificial movement of individuals can complicate the determination of natural connectivity patterns. For example, the area of low effective migration inferred around San Francisco Bay may be due to the introgression of Washington genotypes rather than actual physical barriers to gene flow. Fortunately, intentional movement of Olympia oysters between regions ceased over 80 years ago with the exception of restoration efforts in Netarts Bay from 2005-2012 which utilized broodstock from Willapa Bay. Outlier loci due to spatially-dependent selection can be particularly useful for disentangling human-mediated introgression and "natural" population structure. If selection is strong enough and sufficient time has passed, then maladapted foreign genotypes will be removed or in low frequency in a population.

Local adaptation

Detection of outlier loci using three different methods conservatively identified 97 GBS loci as under putative selection, 15 of which were identified by all three methods. Only 15 loci mapped to protein-coding regions, 2 of which were identified as outliers by all three approaches. Mapping of all the outlier loci identified here to the forthcoming *O. lurida* genome will aid in detecting outlier loci that may be tightly linked to a gene or regulatory region. Direct information about the function of genes or proteins in oysters is sparse, but rapidly increasing with transcriptomic and physiological studies on the commercially important *Crassostrea* species. Plotting minor allele frequency against latitude for these loci demonstrates that the majority of loci show a clinal pattern, where one allele is fixed north of Coos Bay, OR and the other increases in frequency towards the south. Most of the loci with functional annotations are included in this clinal set, which suggests that some of these genes experience divergent selection in *SoCal* compared to more northern populations. All of the functional associations presented here are strictly hypotheses that merit further investigation, however the high divergence of SoCal populations at outlier loci strongly suggests that these populations are adaptively diverged from those to the north.

One trend among annotated clinal loci is variation in genes associated either directly or indirectly with development. Glyoxalase 3 expression has been linked to developmental competence in female oyster gametes (Pauletto et al., 2017). DNA N6-methyl adenine demethylase can regulate gene expression through demethylation of coding regions, which has been linked to developmental timing in oysters (Riviere et al.,

2013). ERG transcriptional regulator, kinase B-raf, and Fez family zinc finger protein are likely to be directly involved with developmental regulation (Gaitán-Espitia and Hofmann, 2017; Epelboin et al., 2016). Olympia oysters exhibit latitudinal variation in gonad development and spawning, with California oysters initiating spawning up to 6°C warmer than those from Puget Sound, WA (Coe, 1932; Hopkins, 1937). Recent evidence suggests that there is heritable, adaptive variation in reproductive timing, even among populations of oysters within the same phylogeographic region (Silliman et al., 2018; Barber et al., 2016). Another clinal locus of interest mapped to the mitochondrial gene carnitine Ol-palmitoyltransferase, which has been strongly associated with regulation of glycogen content (and therefore, tastiness) in *C. gigas* (Li et al., 2017). Other clinal genes have putative functions in sensory information processing and muscle contraction.

Some outlier loci exhibit the opposite of a clinal pattern, with populations in the middle of the range predominantly have a different allele than the northern and southern populations. E3 ubiquitin-protein ligase (locus_121489) is fixed for one allele in *Oregon* and Willapa Bay, WA, which is only present at low or intermediate frequencies in the other populations. One recent study identified that an E3 uniquitin-protein ligase is an important component of the neuroendicrine-immune response in *C. gigas* and is primarily expressed in the gonads. Oregon oyster populations have a history of increased disease incidence (Duan and Su, 2005; Kaysner et al., 1987), which proposes one hypothesis for the spatial variation observed at this gene. Heavy metal-binding protein (HIP) also exhibits this hump-shaped distribution of allele frequencies.

CONCLUSIONS

This study provides the first comprehensive characterization of both neutral and adaptive population structure in the Olympia oyster, an ecologically important coastal species in North America. These results have direct implications for management policies and ongoing restoration efforts to recover ecosystem services and potentially support a sustainable fishery in the future. Putative adaptive loci identified here are excellent candidates for future research and may provide targets for genetic monitoring programs. Beyond these specific applications, this study contributes to the growing body of evidence for both population structure and adaptive differentiation in marine species. In particular, it is one of the first to utilize thousands of SNPs to characterize population structure all the way from southern California to Vancouver Island, identifying potential barriers to gene flow that align with previously described biogeographic boundaries. All analyses conducted for this study can be replicated using annotated Jupyter notebooks, allowing for clear dissemination of bioinformatics methods. Furthermore, these notebooks can facilitate future open-sourced research on the population structure

of *O. lurida*.

DATA ARCHIVING STATEMENT

Reproducible Jupyter notebooks will be available at [redacted]. Raw sequencing data will be stored on the NCBI Sequence Read Archive. Sequence data after all filtering and assembly will be available on Dryad.

ACKNOWLEDGEMENTS

I would like to thank the Washington Dept. of Fish and Wildlife, Jennifer Ruesink, Alan Trimble, Danielle Zacherl, Thomas McCormick, JoAnne Linnenbrink, Daniel Silliman, Brian Kingzett, Steven Rumrill, and Dave Couch for help and advice with sample collections, as well as the Field Museum Pritzker DNA Lab for providing reagents, facilities, and funding for molecular work. Additional funding was provided by the University of Chicago Hinds Fund and student awards through the National Shellfisheries Association and the American Fisheries Society. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1545870 and the Department of Education Graduate Assistance in Areas of National Need Fellowship.

LITERATURE CITED

REFERENCES

Abbott, L. A. and Hollenberg, G. J. (1976). *Marine algae of California*. Stanford University Press, Stanford, CA.

Anderson, K., Taylor, D. A., Thompson, E. L., Melwani, A. R., Nair, S. V., and Raftos, D. A. (2015). Meta-analysis of studies using suppression subtractive hybridization and microarrays to investigate the effects of environmental stress on gene transcription in oysters. *PLoS One*, 10(3):e0118839.

Baker, P. (1995). Review of ecology and fishery of the Olympia oyster, *Ostrea lurida* with annotated bibliography. *J. Shellfish Res.*, 14(2):501–518.

Baker, P., Richmond, N., and Terwilliger, N. (1999). Reestablishment of a native oyster, *Ostrea conchaphila*, following a natural local extinction. In Pederson, J., editor, *Marine Bioinvasions*, pages 221–231.

Banas, N. S., McDonald, P. S., and Armstrong, D. A. (2009). Green crab larval retention in Willapa Bay, Washington: An intensive lagrangian modeling approach. *Estuaries Coasts*, 32(5):893–905.

Barber, J. S., Dexter, J. E., Grossman, S. K., Greiner, C. M., and Mcardle, J. T. (2016). Low temperature brooding of Olympia oysters (*Ostrea lurida*) in northern Puget Sound. *J. Shellfish Res.*, 35(2):351–357.

Barrett, E. M. (1963). The California oyster industry. *Fish. Bull.*, 123.

Baums, I. B. (2008). A restoration genetics guide for coral reef conservation. *Molecular Ecology*, 17(12):2796–2811.

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., Rochette, R., and Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol. Ecol.*, 25(20):5073–5092.

Bible, J. M. and Sanford, E. (2016). Local adaptation in an estuarine foundation species: Implications for restoration. *Biological Conservation*, 193:95–102.

Blake, B. and Bradbury, A. (2012). Washington Department of Fish and Wildlife plan for rebuilding Olympia oyster (*Ostrea lurida*) populations in Puget Sound with a historical and contemporary overview. Technical report, Washington Department of Fish and Wildlife, Brinnon, WA.

Blanchette, C. A., Melissa Miner, C., Raimondi, P. T., Lohse, D., Heady, K. E. K., and Broitman, B. R. (2008). Biogeographical patterns of rocky intertidal communities along the Pacific coast of North America. *J. Biogeogr.*, 35(9):1593–1607.

Buonaccorsi, V. P., Kimbrell, C. A., Lynn, E. A., and Vetter, R. D. (2002). Population structure of copper rockfish (*Sebastes caurinus*) reflects postglacial colonization and contemporary patterns of larval dispersal. *Can. J. Fish. Aquat. Sci.*, 59(8):1374–1384.

Camara, M. D. and Vadopalas, B. (2009). Genetic aspects of restoring Olympia oysters and other native bivalves: Balancing the need for action, good intentions, and the risks of making things worse. *Journal of Shellfish Research*, 28(1):121–145.

Cariou, M., Duret, L., and Charlat, S. (2016). How and how much does RAD-seq bias genetic diversity estimates? *BMC Evol. Biol.*, 16(1):240.

Carson, H. S. (2010). Population connectivity of the Olympia oyster in Southern California. 55(1):134–148.

Chan, F., Barth, J. A., Bl, C. A., Byrne, R. H., C, F., and Cheriton, O. (2017). Persistent spatial structuring of coastal ocean acidification in the California Current System. *Scientific Reports*, 7:e2526.

Cheng, Q., Wang, H., Jiang, S., Wang, L., Xin, L., Liu, C., Jia, Z., Song, L., and Zhu, B. (2016). A novel ubiquitin-protein ligase E3 functions as a modulator of immune response against lipopolysaccharide in Pacific oyster, *Crassostrea gigas*. *Dev. Comp. Immunol.*, 60:180–190.

Coe, W. R. (1932). Development of the gonads and the sequence of the sexual phases in the California oyster (*Ostrea lurida*). *Scripps Institution of Oceanography*.

Coen, L. D., Dumbauld, B. R., and Judge, M. L. (2011). Expanding shellfish aquaculture: A review of the ecological services provided by and impacts of native and cultured bivalves in shellfish-dominated ecosystems. In Shumway, S. E., editor, *Shellfish Aquaculture and the Environment*, chapter 9, pages 239–295. John Wiley & Sons, Inc.

Cowen, R. K., Lwiza, K. M., Sponaugle, S., Paris, C. B., and Olson, D. B. (2000). Connectivity of marine populations: Open or closed? *Science*, 287:857–859.

Cunningham, K. M., Canino, M. F., Spies, I. B., and Hauser, L. (2009). Genetic isolation by distance and localized fjord population structure in Pacific cod (*Gadus macrocephalus*): Limited effective dispersal in the northeastern Pacific Ocean. *Can. J. Fish. Aquat. Sci.*, 66(1):153–166.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., and 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158.

Dawson, M. N. (2001). Phylogeography in coastal marine animals: A solution from California? *Journal of Biogeography*, 28(6):723–736.

De Wit, P. and Palumbi, S. R. (2013). Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Mol. Ecol.*, 22(11):2884–2897.

Drinan, D. P., Gruenthal, K. M., Canino, M. F., Lowry, D., Fisher, M. C., and Hauser, L. (2018). Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evol. Appl.*, 103:17302.

Duan, J. and Su, Y.-C. (2005). Occurrence of *Vibrio parahaemolyticus* in two Oregon oyster -growing bays. *J. Food Sci.*, 70(1):M58–M63.

Dyke, A. S. and Prest, V. K. (1987). Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Gogr. Phys. Quat.*, 41(2):237.

Earl, Dent A. and vonHoldt, Bridgett M (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.*, 4(2):359–361.

Eaton, D. A. R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13):1844–1849.

Eizaguirre, C. and Baltazar-Soares, M. (2014). Evolutionary conservation—evaluating the adaptive potential of species. *Evol. Appl.*, 7(9):963–967.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5):e19379.

Epelboin, Y., Quintric, L., Guévélou, E., Boudry, P., Pichereau, V., and Corporeau, C. (2016). The kinome of Pacific oyster *Crassostrea gigas*, its expression during development and in response to environmental factors. *PLoS One*, 11(5):e0155435.

Everett, M. V., Park, L. K., Berntson, E. A., Elz, A. E., Whitmire, C. E., Keller, A. A., and Clarke, M. E. (2016). Large-scale genotyping-by-sequencing indicates high levels of gene flow in the deep-sea octocoral *Swiftia simplex* (nutting 1909) on the west coast of the united states. *PLoS One*, 11(10):e0165279.

Fenberg, P. B., Menge, B. A., Raimondi, P. T., and Rivadeneira, M. M. (2015). Biogeographic structure of the northeastern Pacific rocky intertidal: The role of upwelling and dispersal to drive patterns. *Ecography*, 38(1):83–95.

Foll, M. and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2):977–993.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., and Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends Ecol. Evol.*, 27(9):489–496.

Gaitán-Espitia, J. D. and Hofmann, G. E. (2017). Gene expression profiling during the embryo-to-larva transition in the giant red sea urchin *Mesocentrotus franciscanus*. *Ecol. Evol.*, 7(8):2798–2811.

Gleason, L. U. and Burton, R. S. (2016). Genomic evidence for ecological divergence against a background of population homogeneity in the marine snail *Chlorostoma funebralis*. *Molecular Ecology*, 25:3557–3573.

Gosselin, T. (2017). radiator: RADseq data exploration, manipulation and visualization using R.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.*, 36(10):3420–3435.

Goudet, J. and Jombart, T. (2015). *hierfstat: Estimation and Tests of Hierarchical F-Statistics*.

Granek, E. F., Polasky, S., Kappel, C. V., Reed, D. J., Stoms, D. M., Koch, E. W., Kennedy, C. J., Cramer, L. A., Hacker, S. D., Barbier, E. B., Aswani, S., Ruckelshaus, M., Perillo, G. M. E., Silliman, B. R., Muthiga, N., Bael, D., and Wolanski, E. (2010). Ecosystem services as a common language for coastal

ecosystem-based management. *Conserv. Biol.*, 24(1):207–216.

Grosberg, R. and Cunningham, C. W. (2001). Genetic structure in the sea. In *Marine community ecology*, pages 61–84. Sinauer, Sunderland, MA.

Hall, C. A. (1964). Shallow-water marine climates and molluscan provinces. *Ecology*, 45:226–234.

Hare, M. P. and Avise, J. C. (1996). Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). *Evolution*, 50(6):2305–2315.

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

Heare, J. E., Blake, B., Davis, J. P., Vadopalas, B., and Roberts, S. B. (2017). Evidence of *Ostrea lurida* Carpenter, 1864, population structure in Puget Sound, WA, USA. *Mar. Ecol.*, 38(5):e12458.

Heare, J. E., White, S. J., Vadopalas, B., and Roberts, S. B. (2018). Differential response to stress in *Ostrea lurida* as measured by gene expression. *PeerJ*, 6:e4261.

Hellberg, M. E. (2009). Gene flow and isolation among populations of marine animals. *Annu. Rev. Ecol. Evol. Syst.*, 40:291–310.

Hettinger, A., Sanford, E., Hill, T. M., Lenz, E. A., Russell, A. D., and Gaylord, B. (2013). Larval carry-over effects from ocean acidification persist in the natural environment. *Glob. Chang. Biol.*, 19(11):3317–3326.

Hoey, J. A. and Pinsky, M. L. (2018). Genomic signatures of environmental selection despite near-panmixia in summer flounder. *Evol. Appl.*

Hopkins, A. E. (1937). Experimental observations on spawning, larval development, and setting in the Olympia oyster, *Ostrea lurida*. *Bulletin of the Bureau of Fisheries*, 68:437–502.

Iwamoto, E. M., Elz, A. E., García-De León, F. J., Silva-Segundo, C. A., Ford, M. J., Palsson, W. A., and Gustafson, R. G. (2015). Microsatellite DNA analysis of Pacific hake *Merluccius productus* population structure in the Salish Sea. *ICES J. Mar. Sci.*, 72(9):2720–2731.

Jackson, T. M. and O'Malley, K. G. (2017). Comparing genetic connectivity among dungeness crab (*Cancer magister*) inhabiting puget sound and coastal washington. *Mar. Biol.*, 164(6):123.

Jombart, T. and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21):3070–3071.

Kanwal, S., Khan, F. Z., Lonie, A., and Sinnott, R. O. (2017). Investigating reproducibility and tracking provenance - a genomic workflow case study. *BMC Bioinformatics*, 18(1):337.

Kaysner, C. A., Abeyta, Jr, C., Wekell, M. M., DePaola, Jr, A., Stott, R. F., and Leitch, J. M. (1987).

Virulent strains of *Vibrio vulnificus* isolated from estuaries of the United States West Coast. *Appl. Environ. Microbiol.*, 53(6):1349–1351.

Kelly, R. P. and Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One*, 5(1):e8594.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., and Jupyter Development Team (2016). Jupyter notebooks — a publishing framework for reproducible computational workflows. In Loizides, F. and Schmidt, B., editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90. IOS Press.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.*, 15(5):1179–1191.

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., and Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evol. Appl.*, 7(3):355–369.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends Ecol. Evol.*, 17(4):183–189.

Li, B., Song, K., Meng, J., Li, L., and Zhang, G. (2017). Integrated application of transcriptomics and metabolomics provides insights into glycogen content regulation in the Pacific oyster *Crassostrea gigas*. *BMC Genomics*, 18(1):713.

Luisetti, T., Turner, R. K., Jickells, T., Andrews, J., Elliott, M., Schaafsma, M., Beaumont, N., Malcolm, S., Burdon, D., Adams, C., and Watts, W. (2014). Coastal Zone Ecosystem Services: From science to values and decision making; a case study. *Sci. Total Environ.*, 493:682–693.

Luu, K., Bazin, E., and Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.*, 1:67–77.

Malécot, G. (1968). *The Mathematics of Heredity*. San Francisco.

Marko, P. B. (1998). Historical allopatry and the biogeography of speciation in the prosobranch snail genus *Nucella*. *Evolution*, 52(3):757–774.

Martinez, E., Buonaccorsi, V., Hyde, J. R., and Aguilar, A. (2017). Population genomics reveals high gene flow in grass rockfish (*Sebastes rastrelliger*). *Mar. Genomics*, 33:57–63.

Miller, K. J. and Ayre, D. J. (2008). Protection of genetic diversity and maintenance of connectivity among reef corals within marine protected areas. *Conserv. Biol.*, 22(5):1245–1254.

Nei, M. and Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.*, 47(3):253–259.

Novembre, J., Williams, R., Pourreza, H., Wang, Y., and Carbonetto, P. (2018). *PCAviz: Visualizing Principal Components Analysis*.

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecol. Appl.*, 13(1):146–158.

Pauletto, M., Milan, M., Huvet, A., Corporeau, C., Suquet, M., Planas, J. V., Moreira, R., Figueras, A., Novoa, B., Patarnello, T., and Bargelloni, L. (2017). Transcriptomic features of *Pecten maximus* oocyte quality and maturation. *PLoS One*, 12(3):e0172805.

Pespeni, M. H., Oliver, T. A., Manier, M. K., and Palumbi, S. R. (2010). Restriction Site Tiling Analysis: accurate discovery and quantitative genotyping of genome-wide polymorphisms using nucleotide arrays. *Genome Biol.*, 11(4):R44.

Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.*, 48(1):94–100.

Pickrell, J. K. and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.*, 8(11):e1002967.

Polson, M. P., Hewson, W. E., Eernisse, D. J., Baker, P. K., and Zacherl, D. C. (2009). You say *Conchaphila*, I say *Lurida*: Molecular evidence for restricting the Olympia oyster (*Ostrea lurida* Carpenter 1864) to temperate western north america. *J. Shellfish Res.*, 28(1, SI):11–21.

Pritchard, C., Shanks, A., Rimler, R., Oates, M., and Rumrill, S. (2015). The Olympia oyster *Ostrea lurida*: Recent advances in natural history, ecology, and restoration. *J. Shellfish Res.*, 34(2).

Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. (2014). ddocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2:e431.

Raith, M., Zacherl, D. C., Pilgrim, E. M., and Eernisse, D. J. (2016). Phylogeny and species diversity of Gulf of California oysters (Ostreidae) inferred from mitochondrial DNA. *Am. Malacol. Bull.*, 33(2):263–283.

Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., and Shank, T. M. (2013). Going where traditional markers have not gone before: Utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Mol. Ecol.*, 22(11):2953–2970.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., and Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.*, 24(17):4348–4370.

Riviere, G., Wu, G.-C., Fellous, A., Goux, D., Sourdaine, P., and Favrel, P. (2013). DNA methylation is crucial for the early development in the oyster *C. gigas*. *Mar. Biotechnol.*, 15(6):739–753.

Sanford, E., Gaylord, B., Hettinger, A., Lenz, E. A., Meyer, K., and Hill, T. M. (2014). Ocean acidification increases the vulnerability of native oysters to predation by invasive snails. *Proc. Biol. Sci.*, 281(1778):20132681.

Sanford, E. and Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Ann. Rev. Mar. Sci.*, 3(1):509–535.

Schoch, G. C., Menge, B. A., Allison, G., Kavanaugh, M., Thompson, S. A., and A. Wood, S. (2006). Fifteen degrees of separation: Latitudinal gradients of rocky intertidal biota along the California Current. *Limnol. Oceanogr.*, 51(6):2564–2585.

Siegle, M. R., Taylor, E. B., Miller, K. M., Withler, R. E., and Yamanaka, K. L. (2013). Subtle population genetic structure in yelloweye rockfish (*Sebastes ruberrimus*) is consistent with a major oceanographic division in British Columbia, Canada. *PLoS One*, 8(8):e71083.

Silliman, K. E., Bowyer, T. K., and Roberts, S. B. (2018). Consistent differences in fitness traits across multiple generations of Olympia oysters. *Sci. Rep.*, 8(1):6080.

Sokal, R. R. (1979). Testing statistical significance of geographic variation patterns. *Syst. Zool.*, 28(2):227–232.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D., Beckerman, A. P., and Slate, J. (2010). Adaptation genomics: The next generation. *Trends Ecol. Evol.*, 25(12):705–712.

Stick, D. A. (2011). *Identification of optimal broodstock for Pacific Northwest oysters*. PhD thesis, Oregon State University.

Szent-Györgyi, A. G., Kalabokis, V. N., and Perreault-Micale, C. L. (1999). Regulation by molluscan myosins. In Imai, S., Ohtsuki, I., and Endo, M., editors, *Muscle Physiology and Biochemistry*, pages 55–62. Springer US, Boston, MA.

Tepolt, C. K. and Palumbi, S. R. (2015). Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Mol. Ecol.*, 24(16):4145–4158.

Townsend, C. H. (1895). The transplanting of eastern oysters to Willapa Bay, Washington, with notes on the native oyster industry. Technical Report 334, U.S. Bureau of Fisheries.

Valentine, J. W. (1966). Numerical analysis of marine molluscan ranges on extratropical northeastern Pacific

shelf. *Limnol. Oceanogr.*, 11:198–211.

Van Wyngaarden, M., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E., and Bradbury, I. R. (2016). Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus,* using RAD-seq derived SNPs. *Evol. Appl.*, (August).

Wang, L. (2017). Population_genomics_scripts. https://github.com/wlz0726/Population_Genomics_Scripts/blob/master/03.treen

Wang, M., Liu, M., Wang, B., Jiang, K., Jia, Z., Wang, L., and Wang, L. (2018). Transcriptomic analysis of exosomal shuttle mRNA in Pacific oyster *Crassostrea gigas* during bacterial stimulation. *Fish Shellfish Immunol.*, 74:540–550.

Wang, S., Meyer, E., McKay, J. K., and Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods*, 9(8):808–810.

Wares, J. P., Gaines, S. D., and Cunningham, C. W. (2001). A comparative study of asymmetric migration events across a marine biogeographic boundary. *Evolution*, 55(2):295–306.

Wasson, K., Hughes, B. B., Berriman, J. S., Chang, A. L., Deck, A., Dinnel, P. A., Endris, C., Espinoza, M., Dudas, S., Ferner, M. C., Grosholz, T., Kimbro, D., Ruesink, J. L., Trimble, A. C., Schaaf, D. V., Zabin, C., and Zacherl, D. C. (2016). Coast-wide recruitment dynamics of Olympia oysters reveal limited synchrony and multiple predictors of failure. *Ecology*, 97(12):3503–3516.

Weersing, K. and Toonen, R. J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Mar. Ecol. Prog. Ser.*, 393:1–12.

Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6):1358–1370.

Whitlock, M. C. and Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F(ST). *Am. Nat.*, (S1):S24–36.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Woelke, C. E. (1959). Pacific oyster *Crassostrea gigas* mortalities: With notes on common oyster predators in Washington waters. Technical report, Wash- ington Dept. of Fisheries, Quilcene Shellfish Laboratory.

Xiao, S., Wu, X., Li, L., and Yu, Z. (2015). Complete mitochondrial genome of the Olympia oyster *Ostrea lurida* (bivalvia, ostreidae). *Mitochondrial DNA*, 26(3):471–472.

zu Ermgassen, P. S. E., Gray, M. W., Langdon, C. J., Spalding, M. D., and Brumbaugh, R. D. (2013). Quantifying the historic contribution of Olympia oysters to filtration in Pacific Coast (USA) estuaries and the implications for restoration objectives. *Aquat. Ecol.*, 47(2).

TABLES

Dataset	H_o	H_e	F_{IS}	F_{ST}	Avg. pairwise F_{ST}	Avg. within-population F_{IS} (Min - Max)
Combined	0.191	0.225	0.051	0.105	0.1026 ($1.83e^{-3}$ - 0.1902)	0.0424 (-0.0960 - 0.1327)
Neutral	0.192	0.224	0.052	0.099	0.097 ($1.26e^{-3}$ - 0.179)	0.0424 (-0.0993 - 0.1315)
Outlier	0.164	0.304	0.024	0.448	0.398 ($5.68e^{-3}$ - 0.692)	0.0104 (-0.384 - 0.266)

Table 1. Overall summary statistics for the combined (13,444 SNPs), neutral (13,191 SNPs), and outlier (168 SNPs) datasets. H_o, observed heterozygosity averaged across loci; H_e, expected heterozygosity averaged across loci; F_{IS} & F_{ST}, Wright's F-statistics averaged across loci (Nei and Chesser, 1983); pairwise F_{ST}, average of all pairwise F_{ST} values (Weir and Cockerham, 1984)

Region	H_o	H_e	F_{IS}	F_{ST}
NWBC	0.177	0.193	0.0822 (0.0748 - 0.0899)	0.016
Puget+BC	0.173	0.189	0.0812 (0.0761 - 0.0865)	0.046
Willapa	0.171	0.181	0.0587 (0.0502 - 0.0674)	0.001
Oregon	0.185	0.196	0.0554 (0.0470 - 0.0632)	0.016
NoCal	0.215	0.227	0.0527 (0.0472 - 0.0583)	0.022
SoCal	0.224	0.253	0.116 (0.1097 - 0.1216)	0.007

Table 2. Overall summary statistics for each phylogeographic region using the neutral dataset of 13,191 SNPs. H_o, observed heterozygosity averaged across loci; H_e, expected heterozygosity averaged across loci; F_{IS} & F_{ST}, Wright's F-statistics averaged across loci (Nei and Chesser, 1983)

Locus ID	Gene description	Top GO IDs	Top hit species
locus_5648	DNA N6-methyl adenine demethylase	F:dioxygenase activity activity	*C. gigas*
locus_6412	glucose dehydrogenase [FAD, quinone]	None	*C. gigas*
locus_7299	transcriptional regulator ERG	None	*C. gigas*
locus_10670	Fez family zinc finger protein 1	F:nucleic acid binding	*C. gigas*
locus_44811	sodium-dependent phosphate transport protein 2B	F:sodium-dependent phosphate transmembrane transporter activity	*C. gigas*
locus_50945	glyoxalase 3-like	None	*C. virginica*
locus_57217	uncharacterized protein LOC111115623	None	*C. virginica*
locus_121489	E3 ubiquitin-protein ligase TRIM9	F:zinc ion binding	*C. virginica*
locus_170867	carnitine O-palmitoyltransferase 2, mitochondrial	F:calcium ion binding, F: transferase activity	*C. gigas*
locus_196263	myosin-XVIIIa	F:actin filament binding	*C. gigas*
locus_251628	myosin heavy chain, striated muscle	F:microtubule motor activity	*C. gigas*
locus_276278	heavy metal-binding protein HIP	None	*C. gigas*
locus_277490	NADH dehydrogenase subunit 5, mitochondrion	C:mitochondrion	*O. lurida*
locus_339584	serine/threonine-protein kinase B-raf	F:metal ion binding, F:kinase activity, P:intracellular signal transduction	*C. virginica*

Table 3. BLASTx and gene ontology (GO) annotation results for outlier loci. Only the 14 loci with positive BLAST hits are shown.

FIGURE LEGENDS

Figure 1: Map of 20 Olympia oyster (*Ostrea lurida*) collection sites from the west coast of North America.

Figure 2: Venn diagram with number of SNPs and Genotype-by-Sequencing loci identified as outliers by three methods: *pcadapt*, *OutFLANK*, and Bayescan.

Figure 3: Population structure results for 19 *O. lurida* populations using (A-B) neutral loci and (C-D) outlier loci. (A, C) Plots of individual admixture determined using the program STRUCTURE at the K recommended by the δK method ($K = 5$ neutral, $K = 2$ outlier), as well as at the value of K that best represents population structure as inferred by other analyses ($K = 6$ neutral, $K = 5$ outlier). (B, D) Principal components analysis plots for PCs 2-5, plotted against latitude of sampling site. Large circles indicate the centroid of populations. Colors refer to the assigned phylogeographic regions for each population.

Figure 4: Principal components analysis plot for neutral loci in *O. lurida*. PC1 represents 5.8% of the total variance and PC2 represents 2.5%. Large circles are the centroid of each population. Colors refer to latitude of sampling site.

Figure 5: Heatmap of pairwise F_{ST} values for 19 populations of *O. lurida* using (A) neutral loci and (B) outlier loci. Populations are ordered from north to south, starting with Klaskino, BC.

Figure 6: Figure 5: Heatmap of pairwise F_{ST} values for 6 phylogeographic regions inferred for *O. lurida* using (A) neutral loci and (B) outlier loci. Regions are ordered from north to south, starting with *NWBC*.

Figure 7: Isolation by distance plot of $F_{ST}/1 - F_{ST}$ versus population pairwise coastal water distance. Neutral loci in red circles (p ¡ 0.001) and outlier loci in blue diamonds (p ¡ 0.001).

Figure 8: TreeMix results for 19 *O. lurida* populations using 1 SNP per neutral locus.Seven migration events are modeled, as this was the best value inferred by evaluating model fit. The tree is rooted by the southernmost populations, San Diego Bay, CA and Mugu Lagoon, CA, are is ordered by latitude where possible. Populations are colored by their inferred phylogeographic region.

Figure 9: Model of effective migration rates as inferred by EEMS for neutral loci in *O. lurida*. Orange represents ares of low migration relative to the average and blue are areas of higher migration. Grey arrows indicate regions of significantly reduced or increased migration (posterior probability ¿ 95%).

Figure 10: Diversity increases from north to south in *O. lurida*. (A) Effective diversity rates as inferred by EEMS, with orange representing areas of lower diversity and blue representing high diversity. (B) Expected heterozygosity (H_e) within each population versus population latitude.

Figure 11: Outlier loci predominantly show clinal patterns in allele frequency. Allele frequency in 97

individual outlier loci plotted against latitude for 19 populations of *O. lurida*. One representative SNP was chosen for each locus, except in the case where two outlier SNPs from the same locus showed different spatial patterns (e.g., locus_277490). Populations are colored by inferred phylogeographic regions.

FIGURES

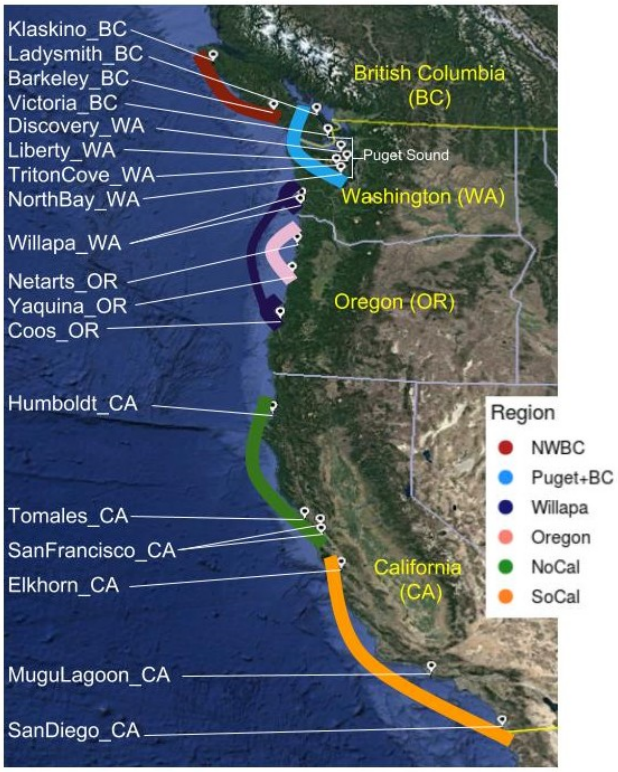

Figure 1

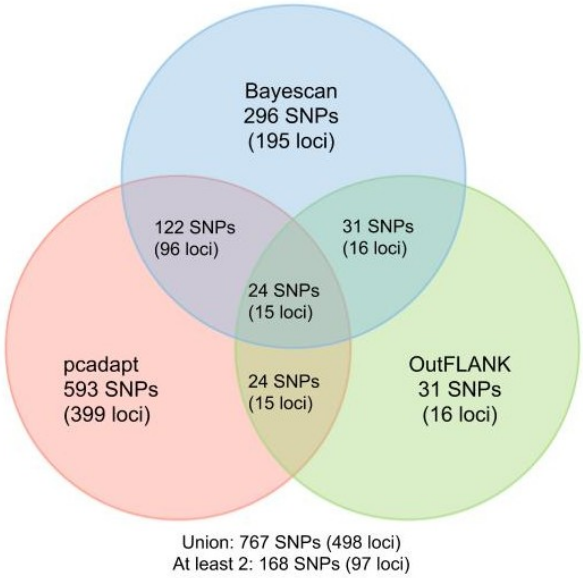

Figure 2

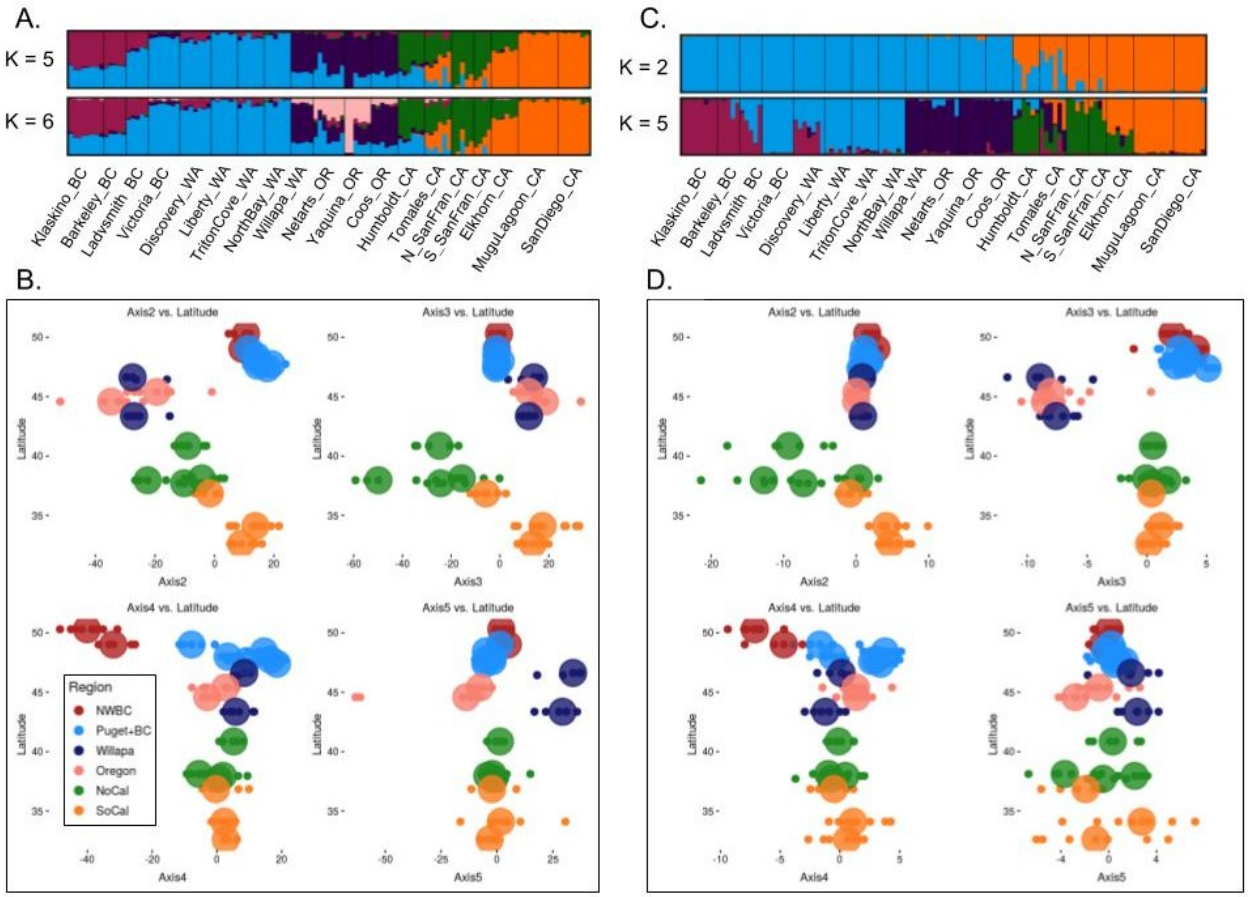

Figure 3

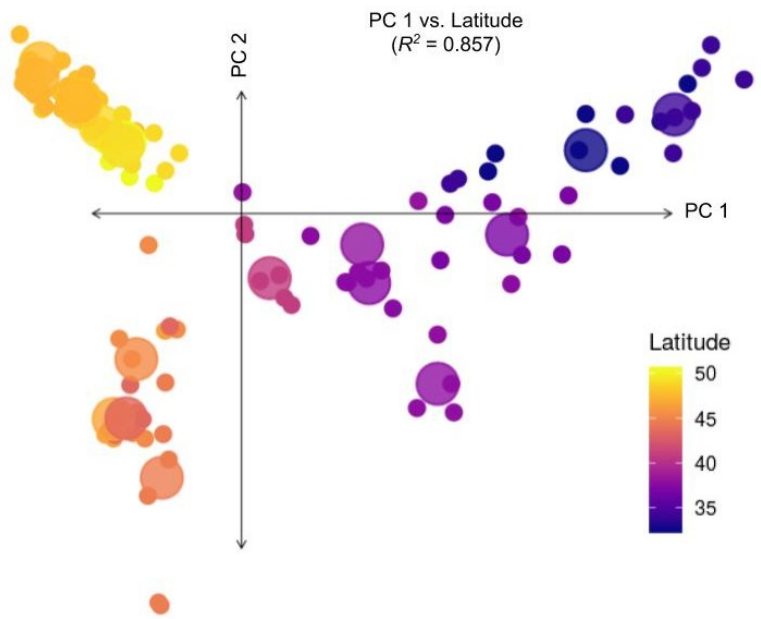

Figure 4

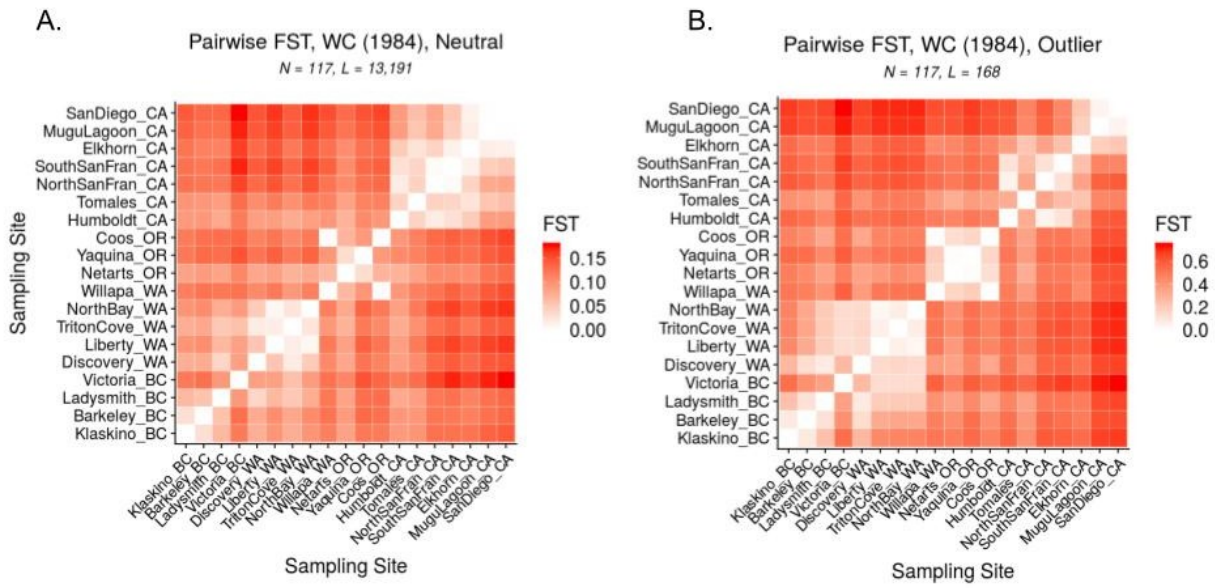

Figure 5

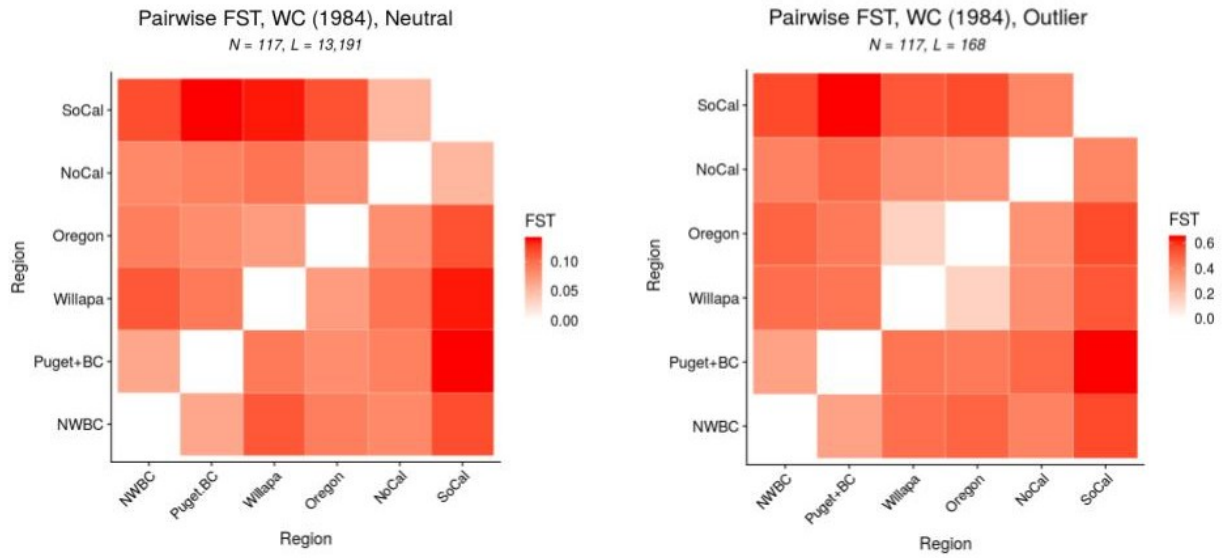

Figure 6

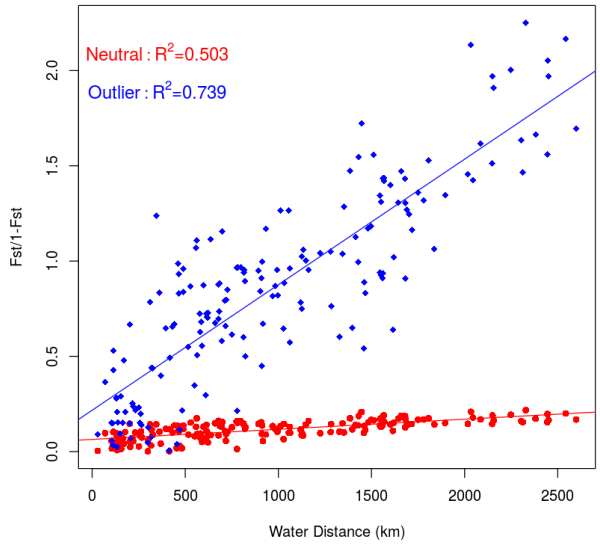

Figure 7

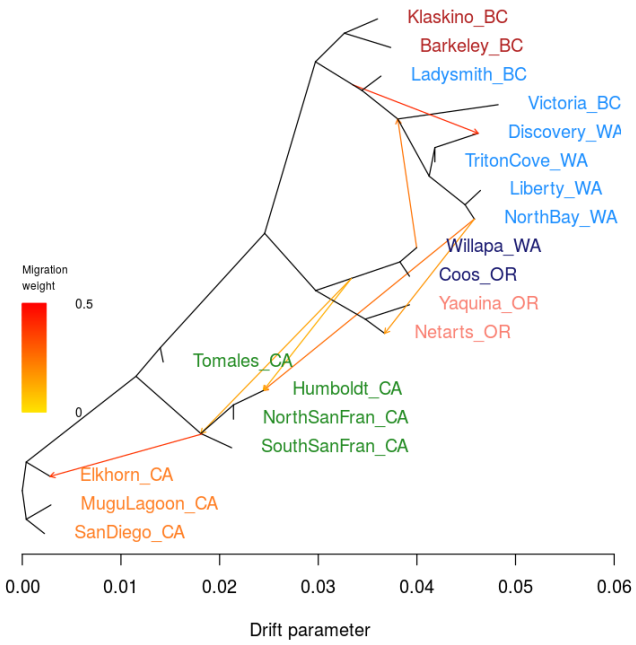

Figure 8

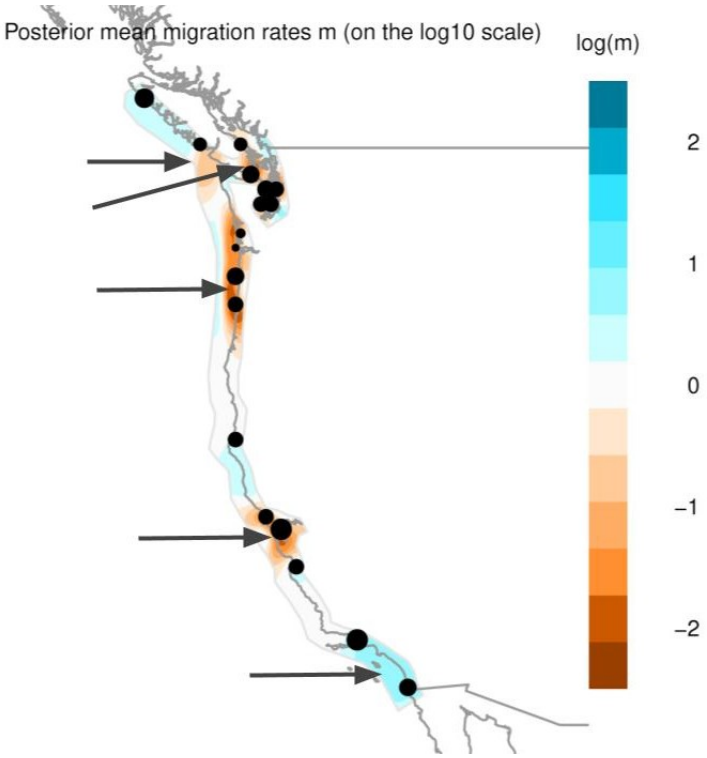

Figure 9

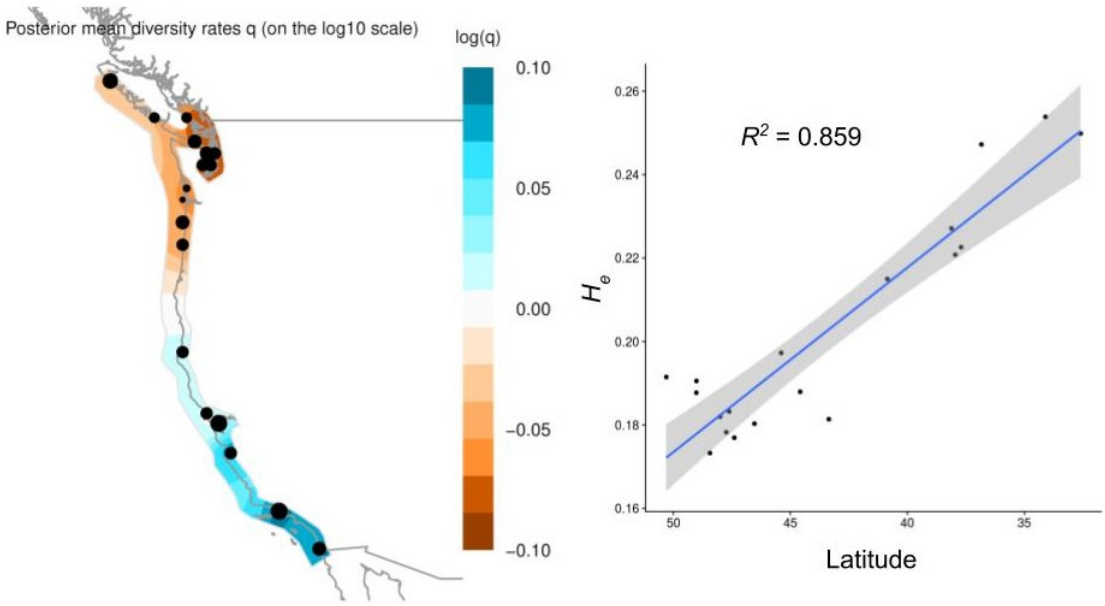

Figure 10

Figure 11

APPENDICES

Appendix A: Sampling locations and population-specific summary statistics

Sampling Site	Latitude	Longitude	# of individuals used in analysis	H_e	F_{IS} (C.I.)
Klaskino Inlet, BC	50.29867	-127.72363	8	0.1904	0.0693 (0.0601 - 0.0788)
Barkley Sound, BC	49.01585	-125.31417	5	0.1867	0.0.0672 (0.0544 - 0.0800)
Ladysmith Harbour, BC	49.01138	-123.8357	5	0.1900	−0.0538 (−0.0684 - −0.0391)
Victoria Gorge, BC	48.43567	-123.37791	7	0.1719	0.0525 (0.0419 - 0.0640)
Discovery Bay, Puget Sound, WA	47.9978	-122.8824	7	0.1812	0.0606 (0.0495 - 0.0724)
Liberty Bay, Puget Sound, WA	47.7375	-122.6507	6	0.1772	0.0320 (0.0218 - 0.0433)
Triton Cove, Puget Sound, WA	47.6131	-122.982	6	0.1820	0.0355 (0.0231 - 0.0465)
North Bay, Puget Sound, WA	47.3925	-122.8138	6	0.1759	0.0535 (0.0421 - 0.0652)
Willapa Bay, WA (North & South)	46.62477 46.4400	-123.98879 -124.004	3 2	0.1801	0.0578 (0.0465 - 0.0710)
Netarts Bay, OR	45.39116	-123.95590	7	0.1968	0.0584 (0.0470 - 0.0698)
Yaquina Bay, OR	44.57954	-123.99577	6	0.1876	0.0143 (0.0018 - 0.0274)
Coos Bay, OR	43.35599	-124.19316	6	0.1809	0.0531 (0.0411 - 0.0654)
Humboldt Bay, CA	40.85580	-124.09746	6	0.2146	0.0327 (0.0209 - 0.0451)
Tomales Bay, CA	38.11755	-122.87450	6	0.2270	−0.0023 (−0.0133 - 0.0077)
Point Orient, San Francisco Bay, CA	37.95507	-122.42180	5	0.2207	0.0560 (0.0450 - 0.0668)
Candlestick Park, San Francisco Bay, CA	37.70867	-122.37761	4	0.2232	−0.0960 (−0.1191 - −0.0755)
Elkhorn Slough, CA	36.83982	-121.74278	6	0.2477	0.0859 (0.0745 - 0.0978)
Mugu Lagoon, CA	34.10191	-119.10434	9	0.2535	0.1327 (0.1239 - 0.1411)
San Diego Bay, CA	32.60250	-117.11889	7	0.2500	0.0958 (0.0845 - 0.1055)

Table A.1. GPS coordinates of sampling sites and population-specific summary statistics averaged across markers using the combined dataset of 13,444 SNPs. H_e, expected heterozygosity; F_{IS}, inbreeding coefficient within the population, mean and 25%-75% confidence intervals (Nei and Chesser, 1983);

Appendix B: Pairwise FST values for outlier and neutral datasets

	Klaskino_BC	Barkeley_BC	Ladysmith_BC	Victoria_BC	Discovery_WA	Liberty_WA	TritonCove_WA	NorthBay_WA	Willapa_WA	Netarts_OR	Yaquina_OR	Coos_OR	Humboldt_CA	Tomales_CA	NorthSanFran_CA	SouthSanFran_CA	Elkhorn_CA	MuguLagoon_CA
Klaskino_BC	0.00000000																	
Barkeley_BC	0.03060334	0.00000000																
Ladysmith_BC	0.06596750	0.05632379	0.00000000															
Victoria_BC	0.12018407	0.12747581	0.09359197	0.00000000														
Discovery_WA	0.07464669	0.07684306	0.03844188	0.08834988	0.00000000													
Liberty_WA	0.09669808	0.10239134	0.06499154	0.08672007	0.05277780	0.000000000												
TritonCove_WA	0.07575397	0.08086087	0.05193009	0.05802171	0.01824506	0.016415713	0.00000000											
NorthBay_WA	0.09703220	0.10539988	0.07546962	0.09080939	0.04766346	0.007248794	0.02451076	0.000000000										
Willapa_WA	0.11534218	0.12472687	0.12109705	0.12541328	0.10649118	0.119501986	0.09575937	0.114854540	0.000000000									
Netarts_OR	0.08465183	0.09128231	0.08432521	0.11007895	0.08165150	0.086406739	0.07169278	0.078559030	0.065021717	0.00000000								
Yaquina_OR	0.12072701	0.13238189	0.12912052	0.14859452	0.12351923	0.138945773	0.11992852	0.136697061	0.095820841	0.03121995	0.00000000							
Coos_OR	0.11828506	0.12710696	0.12852303	0.13460237	0.11332560	0.123254197	0.10643871	0.120053560	0.001587836	0.06889381	0.09277906	0.000000000						
Humboldt_CA	0.09281375	0.09440997	0.07952773	0.11805891	0.08932435	0.088700071	0.07742667	0.088039469	0.094258976	0.07054527	0.10595386	0.094723494	0.00000000					
Tomales_CA	0.09819200	0.09442913	0.08966615	0.12936532	0.10461893	0.118264845	0.10140657	0.118781406	0.114486423	0.08496575	0.11344236	0.114468137	0.04002286	0.00000000				
NorthSanFran_CA	0.12613121	0.12286693	0.12076010	0.15705810	0.13264461	0.147145338	0.12705518	0.147887540	0.123937016	0.10588973	0.12696699	0.131602331	0.01774369	0.04191384	0.000000000			
SouthSanFran_CA	0.12657934	0.12083608	0.12625444	0.17218626	0.14268340	0.157520143	0.13588307	0.158393469	0.138494120	0.10498712	0.13329963	0.138698020	0.02903454	0.03918215	0.003815139	0.000000000		
Elkhorn_CA	0.12453296	0.11460893	0.11245797	0.15942105	0.13142779	0.145513342	0.12069515	0.144028955	0.126151421	0.10394088	0.12689935	0.131816427	0.04809204	0.02600998	0.040221628	0.013257806	0.00000000	
MuguLagoon_CA	0.13761081	0.12736511	0.12650053	0.17252943	0.14453011	0.159297402	0.14008200	0.162016764	0.143655680	0.12487981	0.14097128	0.148816539	0.09055632	0.05348669	0.079969886	0.046157948	0.01672973	0.000000000
SanDiego_CA	0.14365896	0.13394676	0.13365882	0.17867517	0.14717819	0.164628265	0.14217128	0.166942988	0.151473195	0.13066741	0.14896379	0.157494327	0.09106996	0.06106390	0.082524121	0.052472292	0.01325982	0.001256013

Table B.1. Pairwise F_{ST} values for all pairs of populations, using the neutral dataset of 13,191 SNPs.

	Klaskino_BC	Barkeley_BC	Ladysmith_BC	Victoria_BC	Discovery_WA	Liberty_WA	TritonCove_WA	NorthBay_WA	Willapa_WA	Netarts_OR	Yaquina_OR	Coos_OR	Humboldt_CA	Tomales_CA	NorthSanFran_CA	SouthSanFran_CA	Elkhorn_CA	MuguLagoon_CA
Klaskino_BC	0.00000000																	
Barkeley_BC	0.06950578	0.00000000																
Ladysmith_BC	0.22842268	0.11285477	0.00000000															
Victoria_BC	0.48939518	0.39998902	0.34652812	0.00000000														
Discovery_WA	0.25811262	0.13139648	0.08495661	0.2679350	0.00000000													
Liberty_WA	0.41393244	0.30486712	0.20194923	0.1725264	0.13070316	0.000000000												
TritonCove_WA	0.41265856	0.30400080	0.19137940	0.1326984	0.12873214	0.05673845	0.00000000											
NorthBay_WA	0.44195336	0.32963388	0.21862749	0.1790313	0.12991464	0.02308110	0.06572933	0.000000000										
Willapa_WA	0.46613613	0.43929118	0.40038117	0.5533337	0.39323270	0.49670966	0.48254880	0.51688959	0.000000000									
Netarts_OR	0.44344176	0.39538301	0.33625988	0.4531638	0.35457564	0.38592384	0.40492138	0.41075846	0.130399427	0.00000000								
Yaquina_OR	0.48821942	0.46435459	0.40323342	0.5261192	0.42050911	0.46699389	0.46899055	0.49134933	0.164864885	0.03300564	0.00000000							
Coos_OR	0.44960047	0.42343183	0.37488388	0.5362216	0.38008026	0.48447898	0.47203086	0.49963518	0.005679752	0.12202345	0.17214323	0.000000000						
Humboldt_CA	0.51225275	0.48839418	0.42819483	0.5585456	0.49002220	0.50579606	0.51441240	0.50997720	0.459027847	0.41966002	0.45596489	0.454664935	0.00000000					
Tomales_CA	0.38987657	0.37565988	0.35153295	0.5090688	0.39374278	0.47046994	0.45419423	0.47666239	0.364326819	0.31028923	0.33402770	0.367240647	0.28506286	0.00000000				
NorthSanFran_CA	0.55464443	0.52961192	0.48457568	0.6072214	0.53944795	0.57332786	0.56722068	0.56643185	0.500545366	0.46484688	0.48685790	0.491551733	0.03742984	0.2995336	0.000000000			
SouthSanFran_CA	0.53778821	0.49823446	0.48340883	0.6326408	0.54186478	0.58929429	0.58679939	0.59553076	0.487927460	0.45063522	0.47630336	0.491661706	0.10357620	0.2195258	0.08403621	0.000000000		
Elkhorn_CA	0.51550590	0.48095073	0.47579364	0.5892790	0.50473443	0.56611924	0.55915317	0.56848536	0.432748055	0.43938381	0.46952852	0.457214394	0.35743461	0.1881936	0.32377496	0.22497754	0.00000000	
MuguLagoon_CA	0.62024996	0.59254050	0.60221164	0.6810025	0.61772181	0.66315562	0.65613097	0.66721079	0.576511058	0.58309958	0.60918184	0.595900975	0.55861218	0.3973190	0.52729568	0.42145353	0.1774841	0.000000000
SanDiego_CA	0.62873406	0.59423528	0.60934529	0.6924158	0.62458407	0.67233610	0.66337205	0.68428358	0.587263483	0.57396264	0.60437037	0.589030169	0.56235166	0.3927001	0.53900391	0.40215131	0.1765813	0.04522866

Table B.2. Pairwise F_{ST} values for all pairs of populations, using the outlier dataset of 168 SNPs.