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

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, 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 288 SNPs across 129 GBS loci, which were functionally annotated and analyzed separately from the remaining neutral loci. While strong population structure was observed on a regional scale in both neutral and outlier markers, neutral markers had greater power to detect fine-scale structure. 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). Anthropogenic movement of individuals between populations, either for aquaculture or restoration purposes, can confound signals of population structure and should be evaluated when drawing conclusions about genetic connectivity (David, 2018). For the numerous coastal marine species with planktonic dispersal, high connectivity can result in subtle 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 strength of selection overcomes the homogenizing effect of gene flow (Hellberg, 2009). Advances in genomic and computational techniques, such as genotype-by-sequencing (GBS), 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. Outlier loci have provided increased spatial resolution of population structure compared to neutral loci alone for some marine species (Drinan et al., 2018; Van Wyngaarden et al., 2016; Milano et al., 2014), but not all (Moore et al., 2014). In addition to potentially resolving fine-scale genetic differentiation, outlier

loci may be useful for 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 a 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). This study is 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 to go from raw data to results and figures, using Jupyter notebooks. Jupyter notebooks are interactive documents that integrate text, code, and analysis results (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 organisms.

MATERIALS AND METHODS

Sample collection

20-25 adult *Ostrea lurida* over 2cm in length were collected primarily by hand from the intertidal (approx. 0m to -1m tidal height) 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). When possible, oysters were sampled randomly along 10m transects. Each site represents a separate bay, except for Willapa Bay, WA and San Francisco Bay, CA which had two sampling sites each. In the case of one site from Willapa Bay, WA, oysters were collected subtidally (approx. 8m depth) through dredge harvesting of the Pacific oyster, *Crassostrea gigas*. 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 pool 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 pools with a maximum of 48 libraries each were sequenced on eight 100bp single-end lanes (246 different samples, 86 technical replicates, 332 libraries total). 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 GBS loci after final filtering (Step 7) was retained. Samples were removed if they had fewer than 200,000 raw sequencing reads, fewer than 15,000 assembled clusters of at least 10x read depth, and were missing data for over 55% of loci assembled across at least 75% of samples, with Steps 4-7 rerun using the remaining individuals.

The final assembly was then filtered for excess heterozygosity based on deviations from Hardy-Weinberg equilibrium (HWE) in at least two populations, sample coverage of 75%, and an overall minor allele frequency (MAF) of 2.5%, retaining only GBS loci found in at least one individual from all populations. Preliminary analyses conducted on datasets allowing more or less missing data 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 GBS loci. From these GBS 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" SNP dataset. Subsequent analyses of population structure were conducted on three SNP datasets: all SNPs (combined), outlier SNPs, and neutral SNPs—which excluded any SNP found on a putative adaptive GBS 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 10%, 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 0.1 were categorized as outliers.

Putative adaptive GBS 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 the neutral dataset 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). This method uses the covariance of allele frequencies between populations to build a maximum likelihood graph relating populations to their common ancestor, taking admixture events (”migration”) into account to improve the fit to the inferred tree. 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

Almost all replicate pairs showed > 88% similarity in heterozygous calls except for one, in which case both replicates were dropped. 117 samples remained after removal of 14 samples with < 200,000 raw sequencing reads, 49 samples with < 15,000 clusters, and 65 samples missing data for over 55% of loci assembled across at least 75% of samples. 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 (4–9 individuals per population, mean = 6.2). 42,081 biallelic SNPs across 9,836 GBS loci were genotyped in greater than 75% of these individuals (2.8% of prefiltered loci assembled by *ipyrad*). 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 GBS 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 D1. *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. 29 SNPs found across 15 GBS loci were identified as outliers by all three programs. 129 97 GBS loci contained SNPs identified as outliers by at least two approaches, with 288 168 SNPs included in the outlier dataset for subsequent population structure analyses. The neutral dataset, with 13,093 SNPs across 6,063 GBS loci, excluded any SNP found on a GBS locus with an outlier SNP.

Summary statistics, population differentiation, and spatial structure

Summary statistics

Global F_{ST} for outliers (F_{ST} = 0.399) was almost four five times greater than for the combined and neutral SNPs (F_{ST}= 0.105 (combined), 0.097 (neutral)). The outlier dataset had the lowest H_o, but the highest H_e (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 B1. 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 2). 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 2). The separation of *Willapa* sites from *Oregon* was not observed using the outlier dataset until $K = 8$.

PCA on both SNP datasets demonstrated a strong relationship between latitude and the first principal component (PC) (neutral: $R^2 = 0.858$, outlier: $R^2 = 0.806$) (Figure 2). 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 2). 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 similar regional spatial structure for PCs 2-4, but only showed slight separation of *Oregon* from *Willapa* (Figure 2). PC1 of the outlier dataset represented 21.4% of the total variance, and PCs 2-5 represented 9.2%-3.1% of the total variance.

Population differentiation and isolation-by-distance

Pairwise population-specific F_{ST} was higher for outlier SNPs, 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 3, Tables C1 & C2). The three comparisons with the lowest pairwise F_{ST} using the neutral dataset were Willapa Bay, WA/Coos Bay, OR (0.0014), Mugu Lagoon, CA/San Diego Bay, CA (0.0016), and North/South San Francisco Bay, CA (0.0039). Victoria, BC showed higher pairwise F_{ST} with the other British Columbia sites than with sites from Puget Sound, WA. 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 3).

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 4). 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 5). 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 = 0.86)(Figure 6), as did the outlier dataset (R^2 = 0.871).

Functional annotation of outliers

The 129 GBS loci containing outlier SNPs were functionally annotated using Blast2GO. Eighteen 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, vesicular glutamate transporter), immune or stress response (*nad5*, E3 ubiquitin-protein ligase, Ty3-G Gag-Pol, helicase domino), 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); Shiel et al. (2017); Pan et al. (2015); de Lorgeril et al. (2005). 21 additional outlier GBS loci had positive matches to InterPro signatures without any BLASTx hits or gene ontology annotation. Plotting minor allele frequency against latitude for outlier SNPs 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 D2).

DISCUSSION

Reduced-representation genomic methods, such as GBS, can greatly inform reintroduction efforts for threatened and exploited species by resolving fine-scaled population structure, providing estimates of genetic connectivity, and identifying informative markers for characterizing adaptive variation (Allendorf et al., 2010; Gagnaire et al., 2015). Using 13,444 GBS-derived SNPs, I characterized the rangewide population structure of the Olympia oyster from southern California to British Columbia and further identified 288 SNPs across 129 GBS loci potentially associated with local adaptation. Contrary to studies in some other marine species, neutral markers had greater power to detect fine-scale population structure compared to outliers. However, outlier loci did provide evidence for adaptive divergence among some populations with high inferred admixture, and are informative as candidate loci involved in local adaptation. This study highlights the importance of using both neutral and outlier markers for conservation and management applications.

Regional population structure and gene flow

Significant population structure was observed across the range of *O. lurida* in both the neutral and outlier markers, with sampling locations structured into six distinct regions. Notably, most of these regions fit well within previously described biogeographical provinces based on marine species distributions(Hall, 1964; Valentine, 1966; Fenberg et al., 2015). In addition to describing the rangewide population structure of *O. lurida*, the large geographic sampling of this study can facilitate the identification of oceanographic features along the eastern Pacific coast that may be important for structuring populations of marine species with similar life histories. Most of the inferred phylogeographic regions are bounded by areas of reduced gene flow, many of which align to oceanographic features that may be acting as barriers to dispersal. Below I discuss these phylogeographic regions and potential barriers in more detail, as well as provide some recommendations for management at local scales.

Southern California (SoCal)

The *SoCal* region, containing San Diego Bay, CA, Mugu Lagoon, CA, and Elkhorn Slough, CA, extends across both the Southern Californian and the Montereyan biogeographic provinces as defined by Hall (1964), with Monterey Bay as the northern 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). 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). This finding is consistent with meta-analyses demonstrating that strong population structure across Point Conception is the exception rather than the rule for many marine invertebrates(Kelly and Palumbi, 2010; Dawson, 2001) Finer-scaled sampling of populations on either side of Point Conception may provide evidence for slight genetic clines undetected by the current study.

SoCal exhibits the highest genetic diversity of any region, for which I propose three nonexclusive mechanisms. 1) The southward direction of the California Current 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 4). 2) Northern populations exhibit lower genetic diversity due to repeated extirpation or population bottlenecks from glaciation cycles (see *Puget+BC*) (Marko, 2004). 3) 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. The low F_{ST} between Mugu Lagoon and San Diego suggests either a recent transplantation between sites or high gene flow. If the former, I hypothesize that Mugu Lagoon is the recent transplant due to a high inbreeding coefficient (F_{IS}). Nevertheless, three outlier loci exhibited allele frequency shifts of at least 50% between these two populations, suggesting some potential adaptive population divergence.

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, located 46 km south of Humboldt Bay, is an established phylogeographic break for multiple marine species (Kelly and Palumbi, 2010). EEMS identifies an area of significantly reduced gene flow surrounding San Francisco Bay, which may correspond to Monterey Bay (Dawson, 2001), or anthropogenic introductions (see *Anthropogenic influences on population structure*). The two sites within San Francisco Bay (Candlestick Park and Point Orient), exhibit potential adaptive

divergence at some outlier GBS loci despite high potential for gene flow. This result supports evidence for local adaptation from reciprocal transplant studies within San Francisco Bay (Bible and Sanford, 2016), and highlights the importance of taking individual pairwise F_{ST} values (Tables C1, C2) into account when making reintroduction decisions. TreeMix inferred significant migration between San Francisco Bay and Elkhorn Slough, however migration is likely not consistent between these populations based on synchrony of recruitment dynamics(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. 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. This result may be 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*, or it may be an artifact of the model. To my knowledge, this is the first application of EEMS to a marine invertebrate—simulations and additional empirical studies are necessary to evaluate the behavior of EEMS in linear habitats. Currently the protections against importing shellfish from outside of the state are higher than moving shellfish within the state. The strong phylogeographic divide between Willapa Bay, WA and Puget Sound, WA presented here indicates that transfer of Olympia oysters or *Crassostrea* shells between the outer coast of WA and Puget Sound should be considered equivalent to importing oysters from out of state.

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

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 decreased migration out of the region. The *Puget+BC* region is comprised of Ladysmith Harbour, BC, Victoria Gorge, BC, and all four sites in Puget Sound, WA. Strong evidence suggests that Victoria Gorge, BC has a shared evolutionary history with Puget Sound, WA, although EEMS indicates that migration is reduced between these sites. Ladysmith Harbour may

belong to a separate phylogeographic region all together, as this site was intermediate between *NWBC* and *Puget+BC* regions in the STRUCTURE, PCA, and TreeMix analyses. Genetic sampling from additional sites on the central coast of British Columbia and eastern coast of Vancouver Island could test this hypothesis.

The separation of these two regions from those to the south corroborates previous evidence from mitochondrial loci of a strong phylogeographic divide (Polson et al., 2009). 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. Those that do focus on species with much longer dispersal times than *O. lurida* (Buonaccorsi et al., 2002; Cunningham et al., 2009; Iwamoto et al., 2015; Siegle et al., 2013; Jackson and O'Malley, 2017). 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 markers, with the exception of the northernmost site, Discovery Bay. 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 high 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. 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.

Beginning in 1850, oysters were shipped from Willapa Bay to northern California by the millions, including shipments of juvenile or "seed" oysters to be raised in local waters until reaching commercial size. The inferred migration events from North Bay, WA to California sites may be reflecting the historical transplantation of seed oysters from Oakland Bay, WA, about 30 km from North Bay(Woelke, 1959; Baker, 1995). After the crash of the Olympia oyster industry, the non-native oysters *Crassostrea virginica* and *C. gigas* were brought to the west coast for commercial aquaculture. The shells of these species are excellent substrate for 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)) may have resulted in the accidental transfer of *O. lurida* (Townsend, 1895). The low F_{ST} of Willapa Bay, WA and Coos Bay, OR, despite being separated by 415 km, corroborates the theory of an accidental introduction of Olympia oysters from Willapa Bay on *C. gigas* shells in the 1980s (Baker et al., 1999). The low inbreeding coefficient for Coos Bay suggests a potentially large founding population. Future movement of *Crassostrea* for aquaculture purposes 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.

Local adaptation

Detection of outlier loci using three different methods conservatively identified 129 GBS loci as under putative selection. Only 18 GBS loci mapped to protein-coding regions, 3 of which were identified as outliers by all three approaches. Mapping of outlier GBS loci to the forthcoming *O. lurida* genome will aid in detecting 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. While these 129 GBS loci are likely only a fraction of all loci under divergent selection across the *O. lurida* genome (Lowry et al., 2017) and their functional associations are strictly hypotheses, they are nonetheless excellent candidates for future directed studies.

Plotting minor allele frequency against latitude demonstrates that the majority of outlier SNPs show a clinal pattern, where one allele is fixed north of Coos Bay, OR and the other increases in frequency towards the

south (Figure D2). Many clinal GBS loci with functional annotations are associated either directly or indirectly with development. Glyoxalase 3 expression has been linked to developmental competence in female oyster gametes (Pauletto et al., 2017), and DNA N6-methyl adenine demethylase 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, where 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 diverged in *Oregon* and Willapa Bay, WA compared to the other populations. An E3 uniquitin-protein ligase was recently identified as an important component of the neuroendicrine-immune response in *C. gigas* and is primarily expressed in the gonads. Heavy metal-binding protein (HIP) also exhibits this hump-shaped distribution of allele frequencies.

Potential Limitations

Although genomic methods such as GBS have proven useful for evolutionary biology and conservation genetic studies (Andrews et al., 2016), several potential limitations of GBS and the present study should be addressed. Non-random missing data due to polymorphisms in the restriction enzyme cut site (”allelic dropout”) can bias population genetic analyses by underestimating genomic diversity and overestimating F_{ST}, however the impact of these biases on F_{ST} may be limited if effective population size (N_e) is small and if loci with large amounts of missing data are removed from analyses(Gautier et al., 2013; Cariou et al., 2016). Due to large variation in reproductive success every generation, N_e is likely small for *Ostrea* species (Hedgecock, 1994; Lallias et al., 2010). Loci with $> 25\%$ missing data were removed from population genetic analyses, and preliminary analyses allowing 40%–10% missing data still resulted in the same regional population structure and relative values of pairwise F_{ST}, although absolute values of F_{ST} changed slightly. Two reasons may

underlie the large number of individuals (128) removed during filtering. First, too many individuals may have been pooled per sequencing lane given the number of loci targeted, resulting in low sequencing depth for some individuals (Andrews et al., 2016). Second, these libraries were made and sequenced in-house as opposed to a dedicated commercial GBS facility. The protocol learning curve may be why a disproportionate number of individuals failed or had low sequencing depth in the first few prepared libraries. This filtering resulted in 4–9 individuals per population in the final dataset, which is sufficient for estimating F_{ST} when $>$ 1,000 SNPs are used (Willing et al., 2012). While these small population sizes may limit the power to detect outlier loci (Foll and Gaggiotti, 2008), the probability of false positives is reduced by comparing across multiple outlier methods (Rellstab et al., 2015). Lastly, while methods like EEMS and PCA can characterize genetic differentiation, they cannot distinguish between the different demographic scenarios that may result in these patterns(Petkova et al., 2016).

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, and a future sustainable fishery. 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 from southern California to Vancouver Island. All analyses conducted for this study can be replicated using annotated Jupyter notebooks, allowing for clear dissemination of bioinformatics methods and 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

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

Allendorf, F. W., Hohenlohe, P. A., and Luikart, G. (2010). Genomics and the future of conservation genetics. *Nat. Rev. Genet.*, 11(10):697–709.

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.

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

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.

David, A. A. (2018). Reconsidering panmixia: The erosion of phylogeographic barriers due to anthropogenic transport and the incorporation of biophysical models as a solution. *Frontiers in Marine Science*, 5:280.

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

de Lorgeril, J., Saulnier, D., Janech, M. G., Gueguen, Y., and Bachère, E. (2005). Identification of genes that are differentially expressed in hemocytes of the pacific blue shrimp (*Litopenaeus stylirostris*) surviving an infection with vibrio penaeicida. *Physiol. Genomics*, 21(2):174–183.

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.

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.

Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., and Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol. Appl.*, 8(8):769–786.

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.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J.-M., and Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Mol. Ecol.*, 22(11):3165–3178.

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.

Hedgecock, D. (1994). Does variance in reproductive success limit effective population sizes of marine organisms. *Genetics and evolution of aquatic organisms*, 122.

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.

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.

Lallias, D., Taris, N., Boudry, P., Bonhomme, F., and Lapègue, S. (2010). Variance in the reproductive success of flat oyster ostrea edulis l. assessed by parentage analyses in natural and experimental conditions. *Genet. Res.*, 92(3):175–187.

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.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., and Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.*, 17(2):142–152.

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.

Marko, P. B. (2004). 'what's larvae got to do with it?' disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Mol. Ecol.*, 13(3):597–611.

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.

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A. J., Hansen, J. H., Helyar, S. J., Nielsen, E. E., Ogden, R., Patarnello, T., Stagioni, M., FishPopTrace Consortium, Tinti, F., and Bargelloni, L. (2014). Outlier SNP markers reveal fine-scale genetic structuring across european hake populations (merluccius merluccius). *Mol. Ecol.*, 23(1):118–135.

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.

Moore, J.-S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., Chaput, G., and Bernatchez, L. (2014). Conservation genomics of anadromous atlantic salmon across its north american range: outlier loci identify the same patterns of population structure as neutral loci. *Mol. Ecol.*, 23(23):5680–5697.

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.

Pan, P.-Y., Marrs, J., and Ryan, T. A. (2015). Vesicular glutamate transporter 1 orchestrates recruitment of other synaptic vesicle cargo proteins during synaptic vesicle recycling. *J. Biol. Chem.*, 290(37):22593–22601.

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.

Shiel, B. P., Hall, N. E., Cooke, I. R., Robinson, N. A., and Strugnell, J. M. (2017). Epipodial tentacle

gene expression and predetermined resilience to summer mortality in the commercially important greenlip abalone, haliotis laevigata. *Mar. Biotechnol.*, 19(2):191–205.

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.

Willing, E.-M., Dreyer, C., and van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, 7(8):e42649.

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	0.097	0.095 ($1.38e^{-3}$ - 0.177)	0.0425 (-0.0953 - 0.1288)
Outlier	0.162	0.280	0.039	0.399	0.358 (0.0106 - 0.6030)	0.0241 (-0.2649 - 0.3149)

Table 1. Overall summary statistics for the combined (13,444 SNPs), neutral (13,093 SNPs), and outlier (288 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)

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_98257	uncharacterized protein LOC111133343	None	*C. virginica*
locus_121489	E3 ubiquitin-protein ligase TRIM9	F:zinc ion binding	*C. virginica*
locus_123004	Transposon Ty3-G Gag-Pol polyprotein	None	*Mizuhopecten yessoensis*
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_252560	helicase domino-like	None	*C. virginica*
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*
locus_339916	vesicular glutamate transporter 2.1	P:transmembrane transport	*C. gigas*

Table 2. BLASTx and gene ontology (GO) annotation results for outlier loci. Only the 18~~14~~ loci with positive BLAST hits are shown. F: molecular function, C: cellular component, P: biological process.

.

FIGURE LEGENDS

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

Figure 2: 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 inferred from PCA ($K = 6$ neutral, $K = 5$ outlier). (B, D) Principal components analysis plots for PCs 1-5. PC1 is plotted against latitude of sampling site, then PC2 vs. PC3 and PC4 vs. PC5. Large transparent circles indicate the centroid of populations. Colors refer to the phylogeographic regions of each population.

Figure 3: A) Heatmap of pairwise F_{ST} values for 19 populations of *O. lurida* using 13,093 neutral SNPs. Populations are ordered from north to south, starting with Klaskino, BC. B) 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 4: 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, and ordered by latitude where possible. Populations are colored by their inferred phylogeographic region.

Figure 5: 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 6: 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.

FIGURES

Figure 1

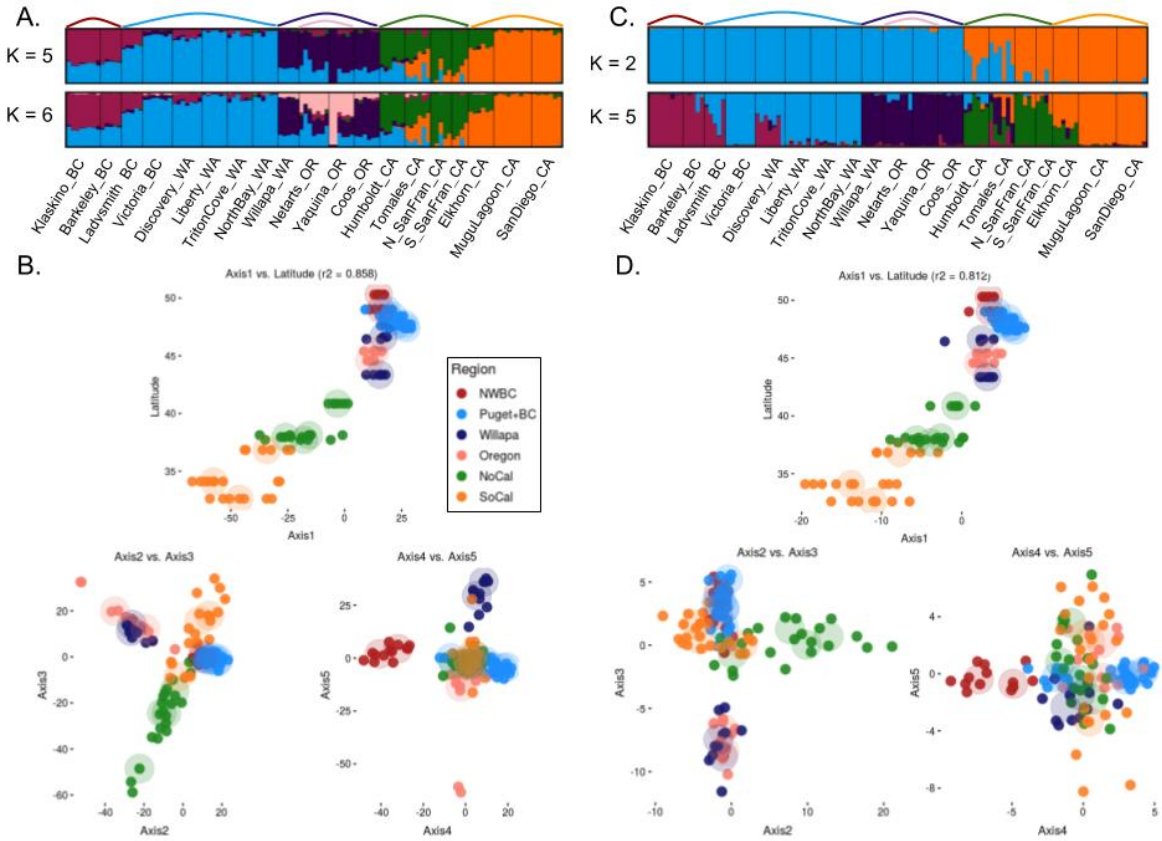

Figure 2

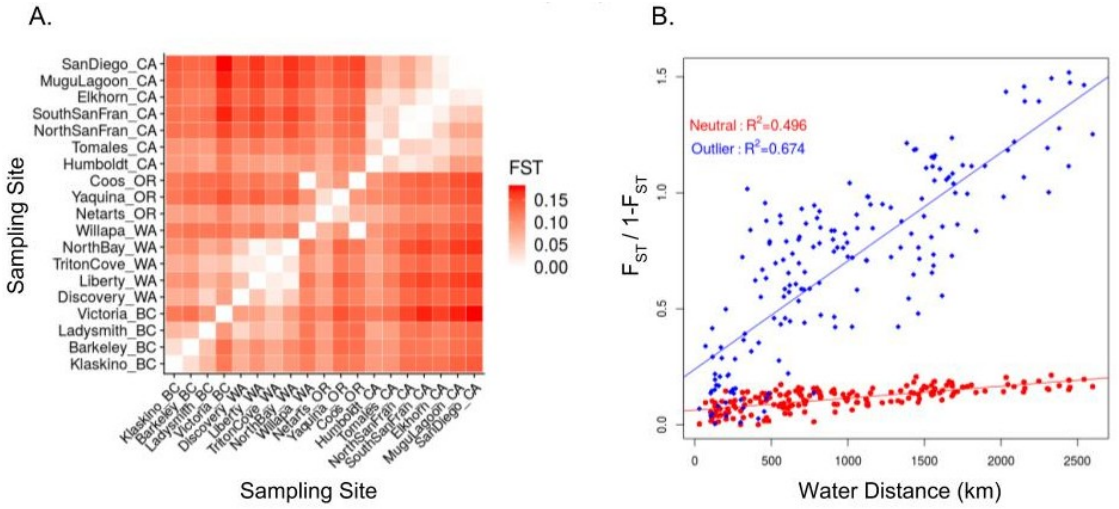

Figure 3

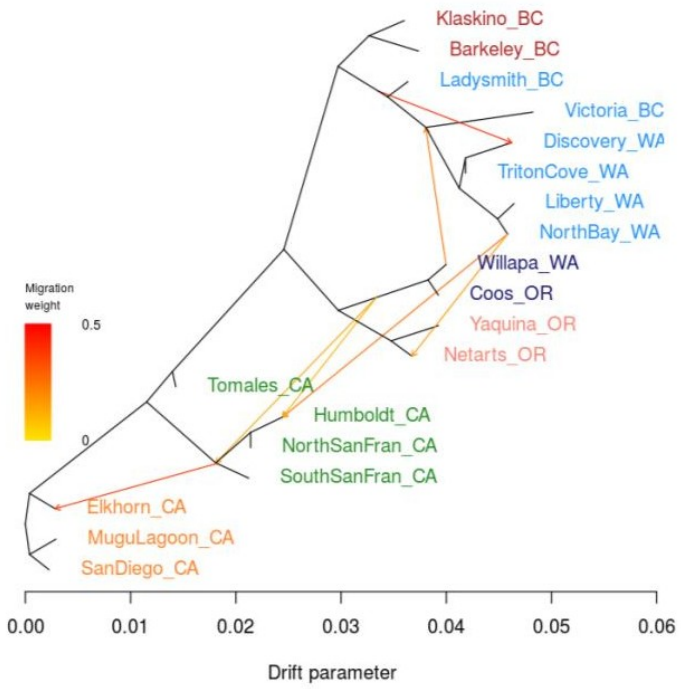

Figure 4

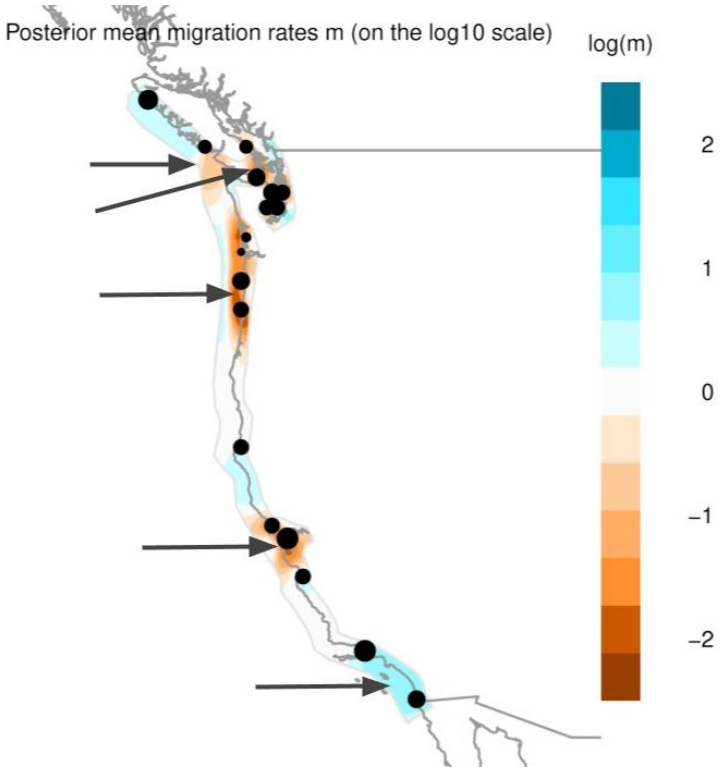

Figure 5

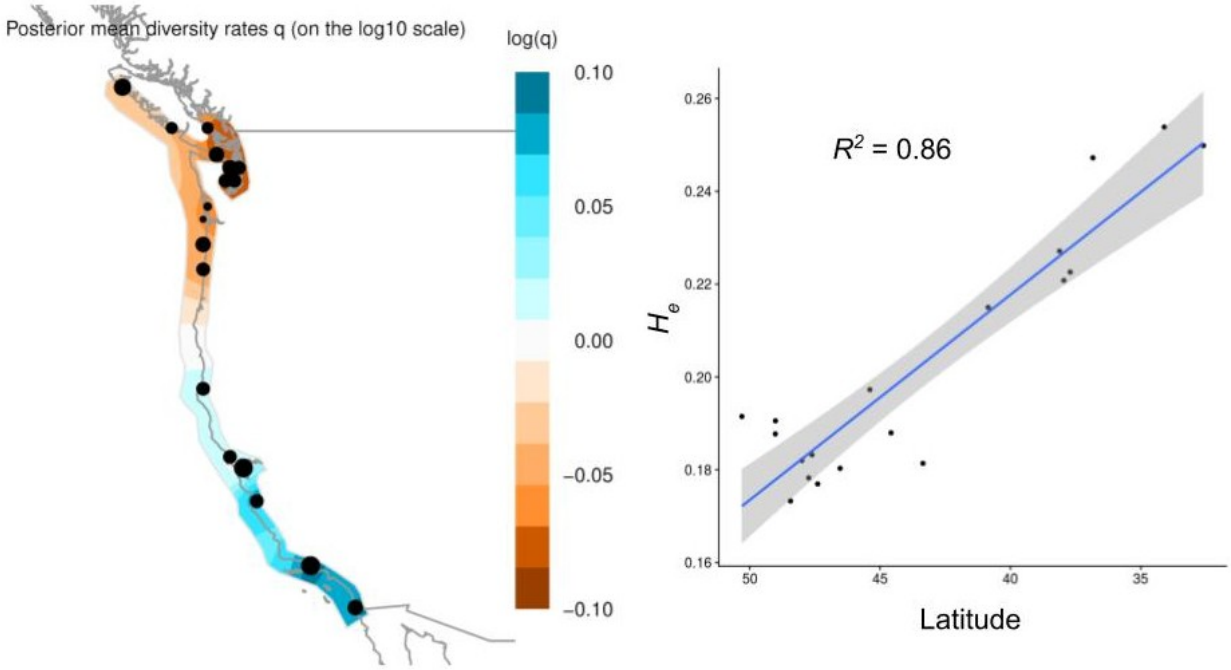
Posterior mean diversity rates q (on the log10 scale)
log(q)
0.10
0.05
0.00
−0.05
−0.10
R^2 = 0.86
H_e
0.26
0.24
0.22
0.20
0.18
0.16
50
45
40
35
Latitude

Figure 6

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: Summary statistics for phylogeographic regions

Region	H_o	H_e	F_{IS}	F_{ST}
NWBC	0.177	0.193	0.0821 (0.0738 - 0.0897)	0.016
Puget+BC	0.174	0.189	0.0814 (0.0758 - 0.0862)	0.046
Willapa	0.171	0.182	0.0583 (0.0495 - 0.0666)	0.001
Oregon	0.185	0.196	0.0556 (0.0474 - 0.0645)	0.016
NoCal	0.215	0.227	0.0536 (0.0472 - 0.0592)	0.022
SoCal	0.224	0.253	0.115 (0.1097 - 0.1209)	0.007

Table B.1. Overall summary statistics for each phylogeographic region using the neutral dataset of 13,093 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). Note that F_{ST} may be skewed by variation in sampling strategy across regions.

845

Appendix C: 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																	
Barkeley_BC	0.0306	0																
Ladysmith_BC	0.0663	0.0564	0															
Victoria_BC	0.1193	0.1270	0.0931	0														
Discovery_WA	0.0742	0.0768	0.0381	0.0873	0													
Liberty_WA	0.0965	0.1021	0.0650	0.0864	0.0526	0												
TritonCove_WA	0.0751	0.0800	0.0516	0.0575	0.0176	0.0166	0											
NorthBay_WA	0.0965	0.1050	0.0750	0.0904	0.0476	0.0075	0.0246	0										
Willapa_WA	0.1136	0.1228	0.1197	0.1247	0.1056	0.1184	0.0946	0.1141	0									
Netarts_OR	0.0825	0.0888	0.0824	0.1097	0.0804	0.0854	0.0703	0.0777	0.0642	0								
Yaquina_OR	0.1190	0.1304	0.1276	0.1480	0.1220	0.1379	0.1184	0.1359	0.0948	0.0310	0							
Coos_OR	0.1159	0.1245	0.1265	0.1338	0.1120	0.1217	0.1048	0.1189	0.00139	0.0687	0.0918	0						
Humboldt_CA	0.0915	0.0936	0.0789	0.1172	0.0884	0.0877	0.0767	0.0870	0.0927	0.0684	0.1045	0.0927	0					
Tomales_CA	0.0970	0.0940	0.0886	0.1282	0.1033	0.1165	0.0998	0.1171	0.1122	0.0821	0.1116	0.1117	0.0397	0				
NorthSanFran_CA	0.1248	0.1222	0.1198	0.1555	0.1309	0.1452	0.1252	0.1458	0.1207	0.1023	0.1244	0.1280	0.0176	0.0418	0			
SouthSanFran_CA	0.1256	0.1204	0.1258	0.1717	0.1424	0.1566	0.1347	0.1574	0.1364	0.1025	0.1318	0.1364	0.0295	0.0399	0.0039	0		
Elkhorn_CA	0.1228	0.1135	0.1108	0.1578	0.1292	0.1431	0.1182	0.1417	0.1231	0.1004	0.1247	0.1284	0.0470	0.0261	0.0398	0.0135	0	
MuguLagoon_CA	0.1351	0.1250	0.1241	0.1698	0.1414	0.1559	0.1366	0.1588	0.1402	0.1212	0.1386	0.1448	0.0880	0.0523	0.0779	0.0453	0.0164	0
SanDiego_CA	0.1417	0.1322	0.1320	0.1764	0.1445	0.1619	0.1392	0.1641	0.1481	0.1270	0.1467	0.1534	0.0892	0.0604	0.0811	0.0520	0.0132	0.0016

Table C.1. Pairwise F_{ST} values for all pairs of populations, using the neutral dataset of 13,093 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																	
Barkeley_BC	0.0675	0																
Ladysmith_BC	0.1818	0.0970	0															
Victoria_BC	0.4416	0.3321	0.2936	0														
Discovery_WA	0.2168	0.1194	0.0751	0.2534	0													
Liberty_WA	0.3640	0.2678	0.1744	0.1632	0.1199	0												
TritonCove_WA	0.3753	0.2819	0.1774	0.1460	0.1268	0.0376	0											
NorthBay_WA	0.3685	0.2609	0.1796	0.1240	0.0878	0.0050	0.0380	0										
Willapa_WA	0.4417	0.4195	0.4008	0.5042	0.3835	0.4887	0.4679	0.4744	0									
Netarts_OR	0.3970	0.3556	0.3021	0.3874	0.3152	0.3556	0.3692	0.3484	0.1720	0								
Yaquina_OR	0.4392	0.4105	0.3772	0.4653	0.3955	0.4443	0.4474	0.4341	0.2208	0.0462	0							
Coos_OR	0.4326	0.4106	0.3797	0.4822	0.3694	0.4791	0.4594	0.4553	0.0253	0.1354	0.2230	0						
Humboldt_CA	0.4835	0.4399	0.4147	0.5104	0.4588	0.4960	0.4957	0.4873	0.4646	0.4157	0.4387	0.4563	0					
Tomales_CA	0.3575	0.2969	0.3241	0.4331	0.3527	0.4330	0.4170	0.4072	0.3675	0.3171	0.3216	0.3696	0.2408	0				
NorthSanFran_CA	0.4997	0.4515	0.4470	0.5423	0.4934	0.5359	0.5269	0.5138	0.4870	0.4517	0.4557	0.4861	0.0539	0.2264	0			
SouthSanFran_CA	0.4633	0.3930	0.4231	0.5428	0.4732	0.5350	0.5372	0.5160	0.4583	0.4200	0.4274	0.4712	0.1025	0.1309	0.0646	0		
Elkhorn_CA	0.4551	0.3962	0.4212	0.5128	0.4606	0.5198	0.5096	0.4989	0.4197	0.4145	0.4187	0.4436	0.3149	0.1339	0.2500	0.1440	0	
MuguLagoon_CA	0.5437	0.4955	0.5163	0.5892	0.5494	0.5931	0.5823	0.5822	0.5282	0.5244	0.5252	0.5483	0.4751	0.3055	0.4247	0.3085	0.1186	0
SanDiego_CA	0.5558	0.5005	0.5271	0.5988	0.5610	0.6030	0.5960	0.5943	0.5419	0.5274	0.5270	0.5529	0.4670	0.2972	0.4330	0.2966	0.1198	0.0106

Table C.2. Pairwise F_{ST} values for all pairs of populations, using the outlier dataset of 288 SNPs.

Appendix D: Additional results of outlier analyses

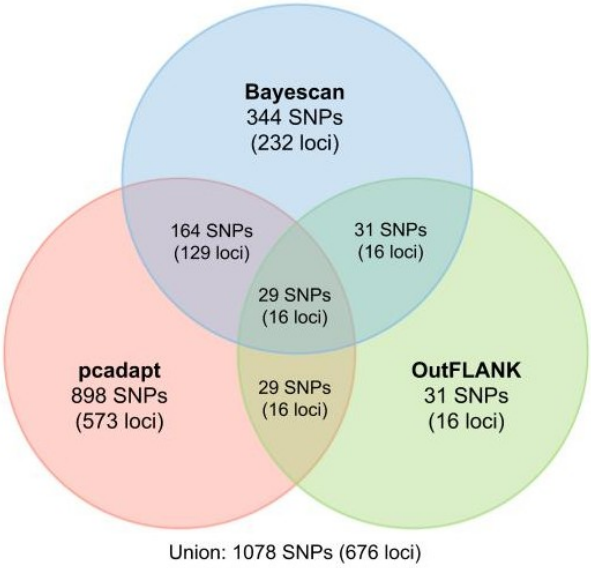

Figure D.1. Venn diagram with number of SNPs and Genotype-by-Sequencing loci identified as outliers by three methods: *pcadapt*, *OutFLANK*, and BayeScan

Figure D.2. Outlier loci predominantly show clinal patterns in allele frequency. Allele frequency in 129 individual outlier loci plotted against latitude for 19 populations of *O. lurida*. One SNP is represented 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.