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, Chicago, Illinois

Correspondence
Katherine Silliman, Committee on Evolutionary Biology, University of Chicago, Chicago, IL. Email: ksilliman@uchicago.edu

Funding Information
National Shellfisheries Association; Office of Postsecondary Education, Grant/Award Number: #P200A150101; University of Chicago; American Fisheries Society; Directorate for Biological Sciences, Grant/Award Number: 1545870

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*, 13,424 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 235 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*. 

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd
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., overharvesting, increasing atmospheric CO₂, and nutrient runoff) 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 & Ayre, 2008; Palumbi, 2003). Anthropogenic movement of individuals between populations, either for aquaculture or for 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 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, Lwiza, Sponaugle, Paris, & Olson, 2000; Hauser & Carvalho, 2008; Sanford & Kelly, 2011; Weersing & 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, McKay, Hohenlohe, & Allendorf, 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 genotyping-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; Milano et al., 2014; Van Wyngaarden et al., 2016), 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 & 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 (Coen, Dumbauld, & Judge, 2011; zu Ermgassen, Gray, Langdon, Spalding, & Brumbaugh, 2013). 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 7 days to 8 weeks before settling on a hard substrate (Baker, 1995). The impact of maternal brooding on population structure in Ostertideae 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 & 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, Shanks, Rimler, Oates, & Rumrill, 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, Hewson, Eernisse, Baker, & Zacherl, 2009; Raith, Zacherl, Pilgrim, & Eernisse, 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 & 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 & Avise, 1996), or the null model of no significant genetic differentiation (Grosberg & 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, Herrera, Layden, Martindale, & Shank, 2013; Van Wyngaarden et al., 2016). Compared to the Atlantic coast of North America (Hoey & Pinsky, 2018), studies utilizing genomewide SNPs for marine taxa from the Pacific coast are far fewer in number and have been limited to regional spatial scales (De Wit & Palumbi, 2013; Drinan et al., 2018; Gleason & Burton, 2016; Larson et al., 2014; Martinez, Buonaccorsi, Hyde, & Aguilar, 2017) or in the number of sampling sites (Pespeni, Oliver, Manier, & Palumbi, 2010; Tepolt & 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, Khan, Lonie, & Sinnott, 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.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection

A total of 20–25 adult *Ostrea lurida* over 2 cm in length were collected primarily by hand from the intertidal (approx. 0 m to −1 m tidal height) at 20 sites ranging from Klagskin Inlet, Vancouver Island (50°17′55″), to San Diego Bay, CA (32°36′19″), in 2014 (Figure 1, Supporting Information Appendix S1: Table A1). When possible, oysters were sampled randomly along 10-m 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. 8 m 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 BioTek) with RNase A treatment following manufacturer instructions and limiting tissue digestion time to no more than 90 min. 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 μl extracted DNA on an 0.8% TAE gel.

### 2.2 | 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 https://github.com/ksil91/Ostrea PopStructure. All libraries were size-selected for fragment sizes between 200 bp and 450 bp 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 100 bp paired-end Illumina HiSeq 2500 lanes, with only the forward sequencing reads used for analysis. Seven other pools with a maximum of 48 libraries each were sequenced on eight 100 bp single-end lanes (246 different samples and 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 demultiplexed, 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 35 bp 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 10× read

**FIGURE 1** Map of 20 Olympia oyster (*Ostrea lurida*) collection sites from the west coast of North America
coverage. Replicate samples were assembled separately and then compared using custom Perl scripts by Mikhail Matz (Wang, Meyer, McKay, & Matz, 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 10× 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) (Carieu, Duret, & Charlat, 2016). Filtering steps were conducted using VCFtools (Danecek et al., 2011), custom Python code, and code adapted from Jon Puritz’s laboratory (Puritz, Hollenbeck, & Gold, 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.

### 2.3 Detection of loci under putative selection

Following recommendations to utilize multiple methods to detect loci under putative directional selection (Benestan et al., 2016; Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015), three approaches were used on the filtered SNP dataset: BayeScan v2.1, OutFLANK v0.2, and pcadapt v4.0.2. 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 & 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 & 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) and 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, Bazin, & Blum, 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.

### 2.4 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 & Jombart, 2015). Confidence intervals for population-specific $F_{IS}$ were determined using the boot.ppsis 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 v2.2.4 (Pickrell & 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, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). The $\Delta K$ method implemented in STRUCTURE HARVESTER was used to determine an optimal $K$ (Earl & vonHoldt, 2012). PCA was implemented in the R package adegenet (Jombart & 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 were 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, Williams, Pourreza, Wang, & Carbonetto, 2018). Results from STRUCTURE, PCA, and pairwise $F_{ST}$ were used to identify phylogeographic “regions.” Summary statistics, including $H_o$, $H_e$, $F_{ST}$, and $F_{ST}$, were calculated for each region using the basicstats function in the R package hierfstat (Goudet & Jombart, 2015).

2.5 | 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, Novembre, & Stephens, 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, and 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 & 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 and 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).

3 | RESULTS

3.1 | GBS and outlier detection

A total of 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). A total of 41,159 biallelic SNPs across 9,696 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 13,424 SNPs across 6,187 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 Supporting Information Appendix S1: Figure D1. OutFLANK, as the most conservative of the programs used (Whitlock & Lotterhos, 2015), had the lowest number of outlier markers detected with 31 SNPs across 16 GBS loci. Twenty-nine SNPs found across 16 GBS loci were identified as outliers by all three programs. A total of 129 GBS loci contained SNPs identified as outliers by at least two approaches, with 235 SNPs included in the outlier dataset for subsequent population structure analyses. The neutral dataset, with 13,073 SNPs across 6,057 GBS loci, excluded any SNP found on a GBS locus with an outlier SNP.

3.2 | Summary statistics, population differentiation, and spatial structure

3.2.1 | Summary statistics

Global $F_{ST}$ for outliers ($F_{ST} = 0.417$) was almost four to 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 $F_{St}$ 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 (Supporting Information Appendix S1: Table A1). Summary statistics for the six phylogeographic regions identified in the following section are...
shown in Supporting Information Appendix S1: 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.

### 3.2.2 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 neutral and outlier SNP datasets demonstrated a strong relationship between latitude and the first principal component (PC) (neutral: R^2 = 0.858, outlier: R^2 = 0.826) (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 23.6% of the total variance, and PCs 2–5 represented 10.3%–2.8% of the total variance.

### 3.2.3 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). The three comparisons with the lowest pairwise F_{ST} using the neutral dataset were Willapa Bay, WA/Coos Bay, OR (0.0015), Mugu Lagoon, CA/San Diego Bay, CA (0.0017), and North/South San Francisco Bay, CA (0.0035). 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 = 0.0011) (Figure 3).

### 3.3 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 0.994 for seven 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: (a) at the mouth of the Strait of Juan de Fuca; (b) around Victoria, BC; (c) extending from Willapa Bay, WA, to southern Oregon; and (d) around San Francisco Bay, CA (Figure 5). These inferred barriers further lend credence to the six 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 ($R^2 = 0.862$; Figure 6), as did the outlier dataset ($R^2 = 0.834$).

### 3.4 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, Wu, Li, & Yu, 2015). Annotated genes have potential roles in developmental regulation (glyoxalase 3, DNA N6-methyladenine demethylase-like, transcriptional regulator ERG, and serine/threonine-protein kinase), sensory information processing (serine/threonine-protein kinase, sodium-dependent phosphate transport, and vesicular glutamate transporter), immune or stress response (nad5, E3 ubiquitin-protein ligase, Ty3-G Gag-Pol, and helicase domain), energy metabolism (carnitine palmitoyltransferase and glucose dehydrogenase), heavy metal binding (heavy metal-binding protein HIP), and muscle contraction (myosin heavy chain-striated muscle, myosin-XVII) (Table 2) (Anderson et al., 2015; Cheng et al., 2016; de Lorgeril, Saulnier, Janech, Gueguen, & Bachère, 2005; Epelboin et al., 2016; Li, Song, Meng, Li, & Zhang, 2017; Pan, Marrs, & Ryan, 2015; Pauletto et al., 2017; Riviere et al., 2013; Shiel, Hall, Cooke, Robinson, & Strugnell, 2017; Szent-Györgyi, Kalabokis, & Perreault-Micale, 1999; Wang et al., 2018). Twenty-one 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

---

**FIGURE 2** Population structure results for 19 *Ostrea 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 $\Delta K$ 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 component analysis plots for PCs 1–5. PC1 is plotted against latitude of sampling site, then PC2 versus PC3 and PC4 versus PC5. Large transparent circles indicate the centroid of populations. Colors refer to the phylogeographic regions of each population.
either Coos Bay, OR, or San Francisco Bay, CA, to the north, and the other allele increases in frequency toward the south (Supporting Information Appendix S1: Figure D2).

4 | 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, Hohenlohe, & Luikart, 2010; Gagnaire et al., 2015). Using 13,424 GBS-derived SNPs, I characterized the rangewide population structure of the Olympia oyster from southern California to British Columbia and further identified 235 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.

4.1 | Regional population structure and gene flow

Significant population structure was observed across the range of Ostrea 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 biogeographic provinces based on marine species distributions (Fenberg, Menge, Raimondi, & Rivadeneira, 2015; Hall, 1964; Valentine, 1966). In addition to describing the rangewide population structure of Ostrea 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
SoCal exhibits the highest genetic diversity of any region, for which I propose three nonexclusive mechanisms. (a) 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). (b) Northern populations exhibit lower genetic diversity due to repeated extirpation or population bottlenecks from glaciation cycles (see Puget+BC) (Marko, 2004). (c) Ongoing or historical admixture from the southern sister species O. conchaphila has increased genetic diversity in these populations. Sampling and genotyping of O. conchaphila are 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, these outlier loci exhibited allele frequency shifts of at least 50% between these two populations, suggesting some potential adaptive population divergence.

### 4.1.2 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 & 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 Section 4.2). 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 & Sanford, 2016), and highlights the importance of taking individual pairwise $F_{ST}$ values (Supporting Information Appendix S1: Tables C1 and 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).

### 4.1.3 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 indicates 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, McDonald, & Armstrong, 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.

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

The NWBC region, comprised of Klaskino Inlet, BC, and Barkley 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 (Kelly & Palumbi, 2010; Valentine, 1966), 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, Kimbrell, Lynn, & Vetter, 2002; Cunningham, Canino, Spies, & Hauser, 2009; Iwamoto et al., 2015; Jackson & O’Malley, 2017; Siegle, Taylor, Miller, Witherl, & Yamanaka, 2013). To my knowledge, this is the first study in a marine mollusk 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.
4.1 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 are likely due to recent genetic bottlenecks and range expansion after the last glacial maximum, which reached just north of Willapa Bay, WA (49°N latitude), until 12–13 kya (Dyke & 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, Blake, Davis, Vadopalas, & Roberts, 2017; Heare, White, Vadopalas, & Roberts, 2018; Silliman, Bowyer, & Roberts, 2018). These results, coupled with experimental evidence for local adaptation to salinity among Northern California populations (Bible & 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, gray 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 (Baker, 1995; Woelke, 1959). 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 FST 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, Richmond, & Terwilliger, 2000). 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 non-native 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 to 2012, which utilized some broodstock from Willapa Bay.

4.3 | 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 toward the south (Supporting Information Appendix S1: 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 adenosine 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 (Epelboin et al., 2016; Gaitán-Espitia & Hofmann, 2017). 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 (Barber, Dexter, Grossman, Greiner, & Mocard, 2016; Silliman et al., 2018). 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 ubiquitin-protein ligase was recently identified as an important component of the neuron-endocrine-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.

4.4 | Potential limitations

Although genomic methods such as GBS have been proven useful for evolutionary biology and conservation genetic studies (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016), several potential limitations of GBS and the present study should be addressed. Nonrandom missing data due to polymorphisms in the restriction enzyme cut site (“allelic dropout”) can bias population genetic analyses by underestimating genomic diversity and overestimating FST; however, the impact of these biases on FST may be limited if effective population size (Ne) is small and if loci with large amounts of missing data are removed from analyses (Cariou et al., 2016; Gautier et al., 2013). Due to large variation in reproductive success every generation, Ne is likely small for Ostrea species (Hedgecock, 1994; Lallias, Taris, Boudry, Bonhomme, & Lapègue, 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, Dreyer, & van Oosterhout, 2012). While these small population sizes may limit the power to detect outlier loci (Foll & 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).

5 | 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.

ACKNOWLEDGEMENTS

I would like to thank the Washington Dept. of Fish and Wildlife, Jennifer Ruesink, Alan Trimble, Danielle Zacherl, Thomas McCormick, JoAnne Linnenbrink, Martin Ruane, 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 Grant No. P200A150101.

CONFICT OF INTEREST

None declared.

DATA ACCESSIBILITY

Genomic data (all filtered markers, putative neutral markers, and putative outliers) and sample metadata are available on Dryad (https://doi.org/10.5061/dryad.114j8m1). Raw demultiplexed DNA sequences for all sequenced individuals with >200,000 raw sequencing reads are available on NCBI SRA (Project Accession Number: SRP174167). Reproducible Jupyter Notebooks are available at https://github.com/ksil91/Ostrea_PopStructure.

ORCID

Katherine Silliman https://orcid.org/0000-0001-5964-3965

REFERENCES

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

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. Nature Reviews Genetics, 11(10), 697–709. https://doi.org/10.1038/nrg2844

Anderson, K., Taylor, D. A., Thompson, E. L., Melwani, A. R., Nair, S. V., & 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. https://doi.org/10.1371/journal.pone.0118839

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), 81–92. https://doi.org/10.1038/nrg.2015.28

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

Baker, P., Richmond, N., & Terwilliger, N. (2000). Reestablishment of a native oyster, Ostrea conchaphila, following a natural local extinction. In J. Pederson (Ed.), Marine bioinvasions (pp. 221–231). Proceedings of the First National Conference MA MIT Sea Grant Program.

Banas, N. S., McDonald, P. S., & Armstrong, D. A. (2009). Green crab larval retention in Willapa Bay, Washington: An intensive Lagrangian modeling approach. Estuaries and Coasts, 32(5), 893–905. https://doi.org/10.1007/s12237-009-9175-7

Barber, J. S., Dexter, J. E., Grossman, S. K., Greiner, C. M., & Mcardle, J. T. (2016). Low temperature brooding of Olympia oysters (Ostrea lurida) in northern Puget Sound. Journal of Shellfish Research, 35(2), 351–357. https://doi.org/10.2983/035.035.0209

Barrett, E. M. (1963). The California oyster industry. Fishery Bulletin, 123, 1–103.

Baums, I. B. (2008). A restoration genetics guide for coral reef conservation. Molecular Ecology, 17(12), 2796–2811. https://doi.org/10.1111/j.1365-294X.2008.03787.x
Gaitán-Espitia, J. D., & Hofmann, G. E. (2017). Gene expression profiling during the embryo-to-larva transition in the giant red sea urchin Mesocentrotus franciscanus. *Ecology and Evolution*, 7(8), 2798–2811. https://doi.org/10.1002/ece3.2850

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., ... Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, 22(11), 3165–3178. https://doi.org/10.1111/mec.12089

Gleon, L. U., & 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. https://doi.org/10.1111/mec.13703

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., Gaitán-Espitia, J. D., & Hofmann, G. E. (2017). Gene expression profiling of restriction site–associated DNA sequencing for genome scans of coastal Washington. *Marine Biology*, 164(6), 123. https://doi.org/10.1007/s00227-017-3152-7

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Kanwal, S., Khan, F. Z., Loni, A., & Sinnott, R. O. (2017). Investigating reproducibility and tracking provenance—A genomic workflow case study. *BMC Bioinformatics*, 18(1), 337. https://doi.org/10.1186/s12859-017-1747-0

Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS ONE*, 5(1), e8594. https://doi.org/10.1371/journal.pone.0008594

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... Team, Jupyter Development (2016). Jupyter notebooks – A publishing framework for reproducible computational workflows. In F. Loizides, & B. Schmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). Amsterdam, the Netherlands: IOS Press.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPACK: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Lallias, D., Taris, N., Boudry, P., Bonhomme, F., & Lapégue, S. (2010). Variance in the reproductive success of flat oyster *Ostrea edulis* assessed by parentage analyses in natural and experimental conditions. *Genetical Research*, 92(3), 175–187. https://doi.org/10.1017/S0016672310000248

Lawson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, 7(3), 355–369. https://doi.org/10.1111/eva.12128

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4), 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Li, B., Song, K., Meng, J., Li, L., & Zhang, G. (2017). Integrated application of transcriptomics and metabolomics provides insights into glycan content regulation in the Pacific oyster *Ostrea edulis*. *PLoS ONE*, 12(7), e0172946. https://doi.org/10.1371/journal.pone.0183361

Liu, B., Song, K., Meng, J., Li, L., & Zhang, G. (2017). Integrated application of transcriptomics and metabolomics provides insights into glycan content regulation in the Pacific oyster *Ostrea edulis*. *PLoS ONE*, 12(7), e0172946. https://doi.org/10.1371/journal.pone.0183361

Marko, P. B. (1998). Historical allopatry and the biogeography of speciation in the prosobranch snail genus *Nucella*. *Evolution*, 52(3), 757–774. https://doi.org/10.1111/j.1558-5646.1998.tb03700.x

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. *Molecular Ecology*, 13(3), 597–611. https://doi.org/10.1111/j.1365-294X.2004.02096.x

Martínez, E., Buonaccorsi, V., Hyde, J. R., & Aguilar, A. (2017). Population genomics reveals high gene flow in grass rockfish (Sebastes...
rastrelliger), Marine Genomics, 33, 57–63. https://doi.org/10.1016/j.margen.2017.01.004
Milano, I., Babbucci, M., Cariani, A., Atanassava, M., Bekkevold, D., Carvalho, G. R., ... Bargelloni, L. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (Merluccius merluccius), Molecular Ecology, 23(1), 118–135. https://doi.org/10.1111/mec.12568
Miller, K. J., & Ayre, D. J. (2008). Protection of genetic diversity and maintenance of connectivity among reef corals within marine protected areas. Conservation Biology, 22(5), 1245–1254. https://doi.org/10.1111/j.1523-1739.2008.00985.x
Moore, J. S., Bourret, V., Dionne, M., Bradbury, I., O’Reilly, P., Kent, M., ... 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. Molecular Ecology, 23(23), 5680–5697. https://doi.org/10.1111/mec.12972
Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. Annals of Human Genetics, 47(3), 253–259. https://doi.org/10.1111/j.1469-1809.1983.tb00993.x
Novembre, J., Williams, R., Pourreza, H., Wang, Y., & Carbonetto, P. (2010). Interference of population splits and fine-scale genetic structuring across European hake populations. Molecular Ecology, 29(17), 4348–4370. https://doi.org/10.1111/j.1365-294X.2010.04613.x
Riviere, G., Wu, G.-C., Fellous, A., Goux, D., Sourdaine, P., & Favrel, P. (2013). DNA methylation is crucial for the early development in the oyster C. gigas. Marine Biotechnology, 15(6), 739–753. https://doi.org/10.1111/j.1467-1556.2013.00923.x
Sanford, E., Gaylord, B., Hettinger, A., Lenz, E. A., Meyer, K., & Hill, T. M. (2014). Ocean acidification increases the vulnerability of native oysters to predation by invasive snails. Proceedings. Biological Sciences, 281(1778), 20132681. https://doi.org/10.1098/rspb.2013.2681
Sanford, E., & Kelly, M. W. (2011). Local adaptation in marine invertebrates. Annual Review of Marine Science, 3(1), 509–535. https://doi.org/10.1146/annurev-marine-120709-142756
Schoch, G. C., Menge, B. A., Allison, G., Kavanaugh, M., Thompson, S. A., & Wood, A. S. (2006). Fifteen degrees of separation: Latitudinal gradients of rocky intertidal biota along the California Current. Limnology and Oceanography, 51(6), 2564–2585. https://doi.org/10.4319/lo.2006.51.6.2564
Shiel, B. P., Hall, N. E., Cooke, I. R., Robinson, N. A., & Strugnell, J. M. (2017). Epipodal tentacle gene expression and predetermined resilience to summer mortality in the commercially important greenlip abalone, Haliotis laevigata. Marine Biotechnology, 19(2), 191–205. https://doi.org/10.1007/s10126-017-9742-z
Siegle, M. R., Taylor, E. B., Miller, K. M., Withler, R. E., & Yamanaka, K. L. (2013). Subtle population genetic structure in yelloweye rockfish (Sebastes ruberrimus) is consistent with a major oceanographic division in yelloweye rockfish (Sebastes ruberrimus) is consistent with a major oceanographic division in British Columbia, Canada. PLoS ONE, 8(8), e71083. https://doi.org/10.1371/journal.pone.0071083
Silliman, K. E., Bowyer, T. K., & Roberts, S. B. (2018). Consistent differences in fitness traits across multiple generations of Oyster populations. Scientific Reports, 8(1), 6080. https://doi.org/10.1038/s41598-018-24455-3
Sokal, R. R. (1979). Testing statistical significance of geographic variation patterns. Systematic Zoology, 28(2), 227–232. https://doi.org/10.2307/2412528
Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., ... Slate, J. (2010). Adaptation genomics: The next generation. Trends in Ecology & Evolution, 25(12), 705–712. https://doi.org/10.1016/j.tree.2010.09.002
Stick, D. A. (2011). Identification of optimal broodstock for Pacific Northwest oysters. PhD thesis, Oregon State University.
Szent-Györgyi, A. G., Kalabokis, V. N., & Perreault-Micale, C. L. (1999). Regulation by molluscan myosins. In S. Imai, I. Ohtsuki, & M. Endo (Eds.), Muscle physiology and biochemistry (pp. 55–62). Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-5543-8
Tepolt, C. K., & Palumbi, S. R. (2015). Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. Molecular Ecology, 24(16), 4145–4158. https://doi.org/10.1111/mec.13294
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. Limnology and Oceanography, 11, 198–211. https://doi.org/10.4319/lo.1966.11.2.0198
Van Wyngaarden, M., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodriguez-Ezepeleta, N., Jeffery, N. W., ... Bradford, I. R. (2016). Identifying patterns of dispersal, connectivity, and selection in the sea scallop, Placopecten magellanicus, using RAD-seq derived SNPs. Evolutionary Applications, 10, 102–117.
Wang, L. (2017). Population genomics scripts. Retrieved from https://github.com/wlz0726/Population Genomics Scripts/blob/master/03.treeMix/795
Wang, M., Liu, M., Wang, B., Jiang, K., Jia, Z., Wang, L., & Wang, L. (2018). Transcriptomic analysis of exosomal shuttle mRNA in Pacific oyster Crassostrea gigas during bacterial stimulation. Fish & Shellfish Immunology, 74, 540–550. https://doi.org/10.1016/j.fsi.2018.01.017

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods, 9(8), 808–810. https://doi.org/10.1038/nmeth.2023

Wares, J. P., Gaines, S. D., & Cunningham, C. W. (2001). A comparative study of asymmetric migration events across a marine biogeographic boundary. Evolution, 55(2), 295–306. https://doi.org/10.1111/j.0014-3820.2001.tb01294.x

Wasson, K., Hughes, B. B., Berriman, J. S., Chang, A. L., Deck, A., Dinnel, P. A., … Zacherl, D. C. (2016). Coast-wide recruitment dynamics of Olympia oysters reveal limited synchrony and multiple predictors of failure. Ecology, 97(12), 3503–3516. https://doi.org/10.1002/ecy.1602

Weersing, K., & Toonen, R. J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. Marine Ecology Progress Series, 393, 1–12. https://doi.org/10.3354/meps08287

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

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of $F_{ST}$. American Naturalist, 186, S24–S36. https://doi.org/10.1086/682949

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. New York, NY: Springer. https://doi.org/10.1007/978-3-319-24277-4

Willing, E.-M., Dreyer, C., & 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. https://doi.org/10.1371/journal.pone.0042649

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

Xiao, S., Wu, X., Li, L., & Yu, Z. (2015). Complete mitochondrial genome of the Olympia oyster Ostrea lurida (Bivalvia, Ostreidae). Mitochondrial DNA, 26(3), 471–472. https://doi.org/10.3109/19401736.2013.834428

zu Ermgassen, P. S. E., Gray, M. W., Langdon, C. J., Spalding, M. D., & Brumbaugh, R. D. (2013). Quantifying the historic contribution of Olympia oysters to filtration in Pacific Coast (USA) estuaries and the implications for restoration objectives. Aquatic Ecology, 47(2), 149–191. https://doi.org/10.1007/s10452-013-9431-6

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Silliman K. Population structure, genetic connectivity, and adaptation in the Olympia oyster (Ostrea lurida) along the west coast of North America. Evol Appl. 2019;12:923–939. https://doi.org/10.1111/eva.12766