PLOS ONE

RESEARCH ARTICLE

# Population structure and adaptive differentiation in the sea cucumber *Apostichopus californicus* and implications for spatial resource management

**Natalie Lowell[1]*, Andy Suhrbier[2], Carolyn Tarpey[1], Samuel May[1], Henry Carson[3], Lorenz Hauser[1]**

**1** School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington, United States of America, **2** Pacific Shellfish Institute, Olympia, Washington, United States of America, **3** Washington Department of Fish and Wildlife, Olympia, Washington, United States of America

* nclowell@uw.edu





**Citation:** Lowell N, Suhrbier A, Tarpey C, May S, Carson H, Hauser L (2023) Population structure and adaptive differentiation in the sea cucumber *Apostichopus californicus* and implications for spatial resource management. PLoS ONE 18(3): e0280500. https://doi.org/10.1371/journal.pone.0280500

**Editor:** Roberta Cimmaruta, Universita degli Studi della Tuscia, ITALY

**Received:** August 8, 2022

**Accepted:** January 3, 2023

**Published:** March 16, 2023



**Data Availability Statement:** Raw fastq files will be available on GenBank. Environmental data, metadata, genotype data, and sequences of RAD loci with putatively adaptive SNPs will be available on Dryad. Additionally, our minimal data sets have been uploaded to publicly accessible data bases and will be made available upon publication. Our raw fastq files can be found on NCBI (accession number PRJNA899570). Our final genotype files (all SNPs, putatively neutral SNPs, and putatively adaptive SNPs), sequences of RAD loci with

## Abstract

A growing body of evidence suggests that spatial population structure can develop in marine species despite large population sizes and high gene flow. Characterizing population structure is important for the effective management of exploited species, as it can be used to identify appropriate scales of management in fishery and aquaculture contexts. The California sea cucumber, *Apostichopus californicus*, is one such exploited species whose management could benefit from further characterization of population structure. Using restriction site-associated DNA (RAD) sequencing, we developed 2075 single nucleotide polymorphisms (SNPs) to quantify genetic structure over a broad section of the species' range along the North American west coast and within the Salish Sea, a region supporting the Washington State *A. californicus* fishery and developing aquaculture production of the species. We found evidence for population structure (global fixation index ($F_{ST}$) = 0.0068) with limited dispersal driving two patterns of differentiation: isolation-by-distance and a latitudinal gradient of differentiation. Notably, we found detectable population differences among collection sites within the Salish Sea (pairwise $F_{ST}$ = 0.001–0.006). Using $F_{ST}$ outlier detection and gene-environment association, we identified 10.2% of total SNPs as putatively adaptive. Environmental variables (e.g., temperature, salinity) from the sea surface were more correlated with genetic variation than those same variables measured near the benthos, suggesting that selection on pelagic larvae may drive adaptive differentiation to a greater degree than selection on adults. Our results were consistent with previous estimates of and patterns in population structure for this species in other extents of the range. Additionally, we found that patterns of neutral and adaptive differentiation co-varied, suggesting that adaptive barriers may limit dispersal. Our study provides guidance to decision-makers regarding the designation of management units for *A. californicus* and adds to the growing body of literature identifying genetic population differentiation in marine species despite large, nominally connected populations.

putatively adaptive SNPs, environmental data, and collection site coordinates can be found on Dryad (https://doi.org/10.5061/dryad.3tx95x6jn).

**Funding:** This work was supported by the NOAA Saltonstall-Kennedy program (#NA15NMF4270322, awarded to LH; https://www.fisheries.noaa.gov/grant/saltonstall-kennedy-grant-competition), by a grant from Washington Sea Grant, University of Washington, pursuant to National Oceanic and Atmospheric Administration (# NA17OAR4170223, R/SFA/N-5, awarded to LH; https://wsg.washington.edu/research/26155-2-15/), and a National Marine Fisheries Service – Sea Grant Joint Fellowship (#NA17OAR4170241, awarded to NL; https://seagrant.noaa.gov/NMFS-SG-Fellowship). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Over the last two decades, genetic studies have provided evidence that some marine species can develop spatial population structure despite very large population sizes and planktonic larval dispersal [1–3]. The traditional view of marine populations as demographically panmictic followed from the hypothesis that an apparent lack of barriers to dispersal in the marine environment and long-distance planktonic larval dispersal common to many marine species would result in high gene flow across large geographic regions [1]. However, it is likely that dispersal in marine populations is more restricted than previously thought, likely due to larval behavior [4, 5] and the presence of oceanographic barriers [6]. Moreover, selection can drive population differentiation in marine populations characterized by negligible genetic drift and/or high gene flow [7, 8]. Lastly, intrinsic reproductive barriers such as chromosome inversions or genetic incompatibilities may further restrict gene flow [9]. Characterization of genetic population structure is particularly important to effective management of exploited species. Genetic population structure can be used to define the scale of management actions to best meet management objectives [10, 11]. Ignoring spatial structure in fishery management may result in overexploitation of less productive or more accessible populations [12]. In turn, overexploitation may reduce population diversity supporting adaptability, and thus sustainability, for current and future uses [13, 14]. Additionally, demographic parameters such as the scale and direction of migration can be inferred from population structure [15], and this information can be used to inform sustainable management practices aimed at maintaining and rebuilding wild populations, such as those in marine protected areas [16].

Estimates of genetic population structure are also useful in informing best practices for aquaculture. Sourcing wild broodstock is a common practice in aquaculture production of species that are outplanted within their native range [17, 18]. If spatial population structure exists, then collecting wild broodstock from one distinct population and outplanting their seed into another can erode spatial population genetic structure and may lead to negative fitness consequences such as loss of local adaptations [19]. Conversely, sourcing broodstock and outplanting their seed within a single population could help maintain both the differentiation among wild populations and fitness of released offspring in the local environment if populations are locally adapted [20]. In recognition of such processes, the Department of Fisheries and Oceans Canada restricts translocation to within "shellfish transfer zones," in part to minimize loss of population differentiation associated with aquaculture [21, 22].

The California sea cucumber, *Apostichopus californicus*, supports wild fisheries and is a novel aquaculture species [23], and thus would benefit from the characterization of spatial population structure. Demand for *A. californicus* has increased over the last several decades [24] as part of a global trend of growing demand for sea cucumbers [25]. Within Washington State, intense fishing pressure on this species peaked between 1988 and 1994, which reduced wild population sizes [25–27] and resulted in the closure of the Central Puget Sound fishery in 2014 [24]. Despite some management intervention, including reduced quotas and a closure during peak spawning season, stocks have been slow to recover [24]. Fishing pressure within Washington State is distributed differentially among six management areas [28], largely based on historic administrative boundaries and tribal fishing rights, but it is not known whether these areas represent biologically meaningful populations. Moreover, shellfish growers and scientists are developing methods for commercial aquaculture production of *A. californicus* [29] to supplement wild harvest, because demand remains high in overseas markets [24]. Defining management units relevant to fisheries would benefit the development of aquaculture as well, as it could help to develop guidelines for broodstock sourcing and stocking practices.

Population differentiation depends in part on life history. *A*. *californicus* reproduce through broadcast spawning, with females producing approximately 100,000–600,000 eggs per spawning event [30]. Larvae may remain in the planktonic stage for weeks to months until they settle, while adults are motile [31, 32]. Recent evidence suggests that adult sea cucumber dispersal can be facilitated through tumbling and floating behaviors, allowing them to take advantage of currents [33]. Formal estimates of life span are unavailable for *A*. *californicus*, but have been postulated to be about 12 years [34]. Other species of sea cucumbers live to be 5 to 10 years old [35]. Thus, dispersal potential is high in *A*. *californicus* because all life history stages are motile and *A*. *californicus* are potentially long-lived, which could lead to high population connectivity. The life history strategy of *A*. *californicus* is that of a periodic strategist with long generation time, moderate reproductive effort, high batch fecundity and low investment per offspring [36]. This strategy is an adaptation to situations where environmental variation affecting juvenile survival is unpredictable but occurs on a large geographic scale [37], which in turn may result in high interannual variation in recruitment [36], high variance in reproductive success and low $N_e/N$ ratios [38]. The life history of *A*. *californicus* may therefore suggest higher population differentiation than expected from census population sizes and high susceptibility to overfishing [36].

*A*. *californicus* is distributed from Baja California, Mexico, to Alaska, from the lower intertidal to depths of 250 *m* [31, 39], a region containing oceanographic barriers to dispersal that could shape population connectivity in *A*. *californicus*. For example, the Salish Sea, a large estuary divided into sub-basins by sills, is known to harbor genetically distinct populations in Pacific Cod, *Gadus macrocephalus* [40], and Brown Rockfish, *Sebastes auriculatus* [41]. The Victoria Sill and Admiralty Inlet are two potentially important oceanographic barriers to dispersal within the Salish Sea. The Victoria Sill is the outermost oceanographic barrier dividing the Salish Sea from the Strait of Juan de Fuca, which connects the Salish Sea to the outer coast [42]. The two sills at Admiralty Inlet separate the main Puget Sound basin from the San Juan Islands region and the Georgia Strait [43]. Another potential oceanographic barrier is the North Pacific Current (NPC), which bifurcates as it approaches the Pacific coast of North America into the northward Alaska and southward California currents. The NPC was already identified as an oceanographic barrier to dispersal in *A*. *californicus* [22] and the Bat Star, *Patiria miniata* [44].

Spatial population structure correlated with oceanographic processes was recently characterized in *A*. *californicus* in the region near British Columbia, Canada [22, 45]. These authors found evidence for local adaptation associated with mean bottom temperature, and to a lesser extent, surface salinity and bottom current velocity [45]. However, this study was restricted to British Columbia, and did not include the southern extent of the range, in particular the southern Salish Sea, where the majority of sea cucumbers are harvested in the contiguous United States [46] and where the fishery has experienced closures due to overfishing [24]. Thus, estimates of spatial population structure are particularly needed in this region. Sampling the northern and southern extents of the region, beyond British Columbia, is also needed, as fisheries for *A*. *californicus* extend from Alaska to California.

In this study, we quantified and described patterns of population differentiation for *A*. *californicus* to inform effective management of wild populations and facilitate the development of management guidelines for aquaculture. We sampled populations at small and large geographic scales to test for both fine- and broad-scale population structure, particularly in regions that allow us to test hypotheses of potential oceanographic barriers to dispersal. Using restriction site-associated DNA (RAD) sequencing, we quantified genetic structure and investigated its potential drivers. Specifically, we tested whether limited dispersal and local adaptation shape population structure. We compared our results with existing estimates of spatial

**Table 1. General information by collection site.**

| *Site, State* | *SS* | *PS* | *NPC* | *Code* | *Lat* | *Long* | $I_S$ | $I_R$ | $H_E$ | $H_O$ | $F_{IS}$ | $S_P$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chiniak Bay, AK | O | O | N | CB_AK | 57.712 | -152.357 | 27 | 19 | 0.278 | 0.251 | 0.073 | 0.931 |
| Yakutat Bay, AK | O | O | N | YB_AK | 59.716 | -139.846 | 41 | 39 | 0.279 | 0.250 | 0.091 | 0.976 |
| Auke Bay, AK | O | O | N | AB_AK | 58.367 | -134.668 | 12 | 10 | 0.275 | 0.240 | 0.093 | 0.839 |
| Bella Bella, BC | O | O | N | BB_BC | 52.152 | -128.139 | 41 | 32 | 0.286 | 0.232 | 0.120 | 0.950 |
| Sekiu, WA | O | O | S | SK_WA | 48.160 | -124.175 | 47 | 47 | 0.279 | 0.234 | 0.147 | 0.980 |
| James Island, WA | I | O | S | JI_WA | 48.513 | -122.774 | 53 | 50 | 0.277 | 0.233 | 0.151 | 0.971 |
| Keyport, WA | I | I | S | KP_WA | 47.701 | -122.634 | 44 | 42 | 0.278 | 0.233 | 0.157 | 0.966 |
| Eld Inlet, WA | I | I | S | EI_WA | 47.147 | -122.935 | 53 | 48 | 0.280 | 0.234 | 0.154 | 0.973 |
| Charleston, OR | O | O | S | CH_OR | 43.340 | -124.378 | 39 | 38 | 0.277 | 0.233 | 0.155 | 0.967 |

Column Site, State refers to the collection site name, followed by the state or province using two letter codes, SS refers to whether inside (I) or outside (O) the Salish Sea, NPC refers to whether north (N) or south (S) of the North Pacific Current (NPC), Code contains the collection site code used in tables and figures, Lat and Long refer to latitude and longitude of collection sites, $I_S$ refers to number of individuals sequenced (excluding replicates), $I_R$ refers to number of individuals retained in analyses (excluding replicates), $H_E$ refers to mean expected heterozygosity, $H_O$ refers to mean observed heterozygosity, $F_{IS}$ refers to mean locus $F_{IS}$, and $S_P$ refers to the proportion of SNPs that are polymorphic.

https://doi.org/10.1371/journal.pone.0280500.t001

population structure to identify common patterns and build hypotheses about drivers of differentiation in the species. To contextualize our results for managers, we developed a two-population model to evaluate which conditions of effective population size, migration rate, number of generations of drift resulted in pairwise population differentiation within the range observed in this study.

## Materials and methods

### Sample collection

Adult *A. californicus* were collected by scuba divers from nine collection sites along the Pacific Coast of North America, ranging from Alaska to Oregon, including four collection sites within the southern Salish Sea (Table 1, Fig 1). Sample size ranges from 12 to 53 samples per collection site. For each animal, a tissue sample was excised from a radial muscle band and stored in 100% ethanol.

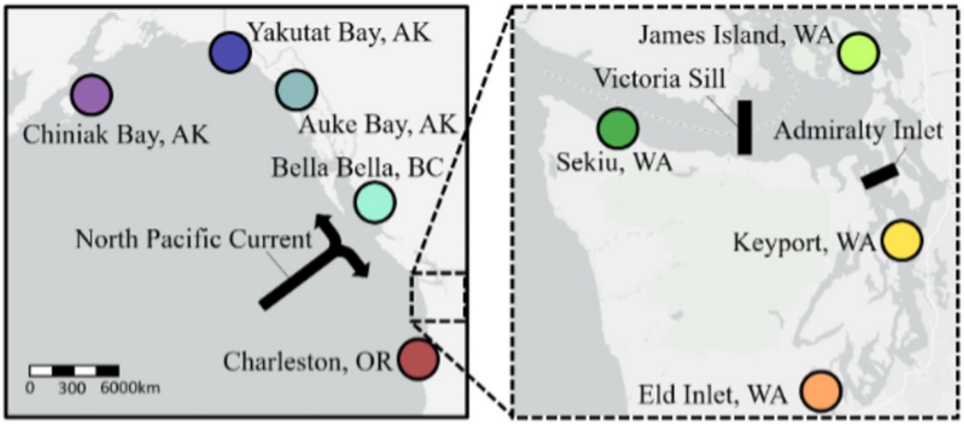


**Fig 1. A map of collection sites.** Collection sites are located in the Northeast Pacific Ocean, including an inset of the southern Salish Sea (right) and potential oceanographic barriers investigated in this study.

https://doi.org/10.1371/journal.pone.0280500.g001

## DNA library preparation

DNA was extracted from tissue samples using the EZNA Mollusc DNA Kit (OMEGA Bio-tek, Norcross, GA, USA) and the Qiagen DNeasy Kit (Qiagen, Germantown, MD, USA). DNA was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) and DNA quality was checked by gel electrophoresis. DNA concentration was normalized to 500 *ng* in 20 *μL* of PCR-grade water. We selected samples with high DNA quality for restriction site associated DNA (RAD) sequencing and RAD libraries were prepared following standard protocols [47]. Briefly, DNA samples were barcoded with an individual six-base identifier sequence attached to an Illumina P1 adapter. Samples were then pooled into sub-libraries, containing approximately 12 individuals. Sub-libraries were sheared using a Bioruptor sonicator and size selected to 200–400 *bp* using a MinElute Gel Extraction Kit (Qiagen, Germantown, MD, USA). P2 adapters were ligated to DNA in sub-libraries and amplified with PCR using 12–18 cycles as in Etter et al. [47]. Finally, amplified sub-libraries were combined into pools of approximately 72 individuals. Paired-end 2 x 150-base pair sequencing was performed on an Illumina HiSeq4000 (San Diego, California, USA) at the Beijing Genomics Institute and the University of Oregon Genomics and Cell Characterization Core Facility. Only forward reads were used for analysis. To estimate genotyping error, 14 individuals were sequenced twice.

## Genotyping individuals

Raw RAD sequencing data were demultiplexed and sequences were trimmed to 140 *bp* using the *process_radtags* module in the pipeline *STACKS* v.1.44 [48]. A threshold of 800,000 reads was used to exclude poorly sequenced individuals. Because a genome was not available for *A. californicus*, we aligned individual sequences to the genome of a closely related species, *A. parvimensis* (GenBank accession number = GCA_000934455.1). The *A. parvimensis* genome was 760,654,621 *bp*, with 21,559 scaffolds and an N50 size of 9,587. We retained reads with a minimum mapping quality score of 20. Then, we used *dDocent* v.2.7.8 to perform a reference-guided locus assembly using the filtered reads and default parameters [49]. Additionally, a parallel *de novo* assembly was performed, which produced nearly identical results for population structure (Table A and Fig A in S1 File) and 1.8–2.8% lower mean expected heterozygosity, 0.9–1.8% higher mean observed heterozygosity, and 1.2–3.3% higher proportions of polymorphic SNPs (Table B in S1 File) than in the with-reference assembly, although with similar patterns across collection sites. The reference-guided assembly was retained for further analyses due to decreased confidence in identifying genotyping errors in the *de novo* assembly [50].

We used *vcftools* v.0.1.16 [51] to remove indels and to retain only single nucleotide polymorphisms (SNPs) with a minimum quality score of 20, minimum minor allele frequency of 0.05, and maximum missing data per locus of 30% across collection sites. Individuals with more than 30% missing data across SNPs were removed. In cases of multiple SNPs per RAD tag, we retained the SNP with the highest minor allele frequency [52]. SNPs that were not in Hardy Weinberg Equilibrium (HWE) were considered sequencing errors or poorly assembled loci and were removed from our data set, as selection and inbreeding are unlikely to cause significant deviations from HWE equilibrium at biallelic loci [53]. We tested SNPs for deviations from HWE using the R package *genepop* v.1.1.4 [54]. SNPs were identified as being out of HWE if they had a $q$-value below 0.05 in at least 2 of the collection sites after correcting for false discovery rate, following Waples [53].

## Population genetic structure analyses

We used a suite of R packages, stand-alone software, and custom scripts in the programming language R v.3.5.0 [53, 55] to quantify genetic diversity and population structure. Mean

expected heterozygosity, observed heterozygosity, and the inbreeding coefficient ($F_{IS}$) per SNP were calculated using the R package *genepop*. The proportion of polymorphic SNPs per collection site was calculated using a custom R script.

To investigate population structure, we first calculated Weir-Cockerham fixation index ($F_{ST}$) [56] to quantify population differentiation using the R packages *genepop* and *hierfstat* v.0.5.7. Exact *G*-tests [57] were used to test for significant genic differentiation using the R package *genepop*. To investigate patterns of spatial differentiation among collection sites, the R package *adegenet* v.2.1.1 [58] was used to conduct discriminant analyses of principal components (DAPC), a multivariate method that summarizes the between-group variation (i.e., population structure), while minimizing within-group variation [59]. The built-in optimization algorithm was used to retain the number of principal components that minimized over-fitting and under-fitting of the model. To determine the potential number of underlying populations, the program *ADMIXTURE* v.1.3.0 was used to conduct a clustering analysis [60]. Specifically, *ADMIXTURE* uses a maximum likelihood-based approach to estimate individual ancestries across different assumed numbers of populations, with the best fit selected using cross-validation. To examine the presence of hierarchical population structure, we conducted analyses of molecular variance (AMOVA) using the *ade4* method of the R package *poppr* v.2.8.1 [61]. Significance of AMOVAs was determined using permutation tests with 1,000 iterations. Using AMOVA, we investigated whether the following oceanographic barriers limit dispersal: 1) the Victoria Sill (Victoria Sill grouping), 2) Admiralty Inlet (Admiralty Inlet grouping), and 3) the North Pacific Current (NPC grouping). Because AMOVAs for each oceanographic barrier include sites in an area with other potential oceanographic barriers, we added a fourth grouping of all three oceanographic barriers (All Barriers grouping) to investigate the relative role of oceanographic barriers compared to other factors. Additionally, we conducted an AMOVA by state or province (State grouping). Although not biologically meaningful, we included the State grouping to determine how much genetic variation is captured by regional management boundaries. Isolation-by-distance (IBD) was tested with Mantel tests [62] in R as correlation between linearized $F_{ST}$ [63] using all SNPs and shortest Euclidean distance through water (in-water distance hereafter) approximated in Google Maps [64], between all pairs of collection sites. Following Xuereb et al. [65], we also tested IBD in the northern and southern population section separately. Following Buonaccorsi et al. [41], we estimated mean dispersal distance from the slope of the regression of linearized $F_{ST}$ and in-water distance. We used this one-dimensional model because it is an appropriate approximation for coastal species with dispersal distances greater than the second dimension of the habitat [66] and because spatial patterns do not seem to affect empirical IBD patterns [67]. We estimated mean dispersal distance from a set of potential population density estimates as population density estimates were unavailable.

## Identifying putative adaptive differentiation

We used two approaches to investigate putatively adaptive SNPs: $F_{ST}$ outlier detection and gene-environment association. $F_{ST}$ outlier detection is used to identify loci potentially under spatial selection [68, 69], although this method does not identify the potential cause of selection. Although gene-environment association does not explicitly test whether such associations are adaptive, this method is used to identify locus-environment associations as evidence for potential local adaptation [70]. SNPs were classified as putatively adaptive if they were detected as $F_{ST}$ outliers or if they were significantly correlated to environmental predictors using gene-environment association.

For $F_{ST}$ outlier detection, two methods were used: the program *Bayescan* v.2.1 [68] and the R package *OutFLANK* v.0.2 [69]. *Bayescan* first applies linear regression to decompose $F_{ST}$ into a population- and a locus-specific component. Using these components as Bayesian priors, the program estimates the posterior probability that a locus is under selection [68]. *OutFLANK* detects $F_{ST}$ outliers using a maximum likelihood approach. The program first infers a distribution of neutral $F_{ST}$ from a trimmed distribution of empirically collected $F_{ST}$ values and uses this neutral distribution to identify outliers. *OutFLANK* advances earlier $F_{ST}$ outlier methods [71] by accounting for sampling error and non-independent sampling of populations, and has lower false positive rates compared to other $F_{ST}$ outlier methods [69]. We used default parameters in both programs, including a false discovery rate of 0.05. SNPs were classified as $F_{ST}$ outliers if they were detected with either program, to include SNPs under weak selection, which are likely the majority of SNPs under selection [72].

Prior to investigating gene-environment association, we gathered estimates for oceanographic variables at each collection site using the Bio-Oracle and Bio-Oracle 2 databases [73], which contain geophysical, biotic, and environmental data layers for marine realms, through the R package *sdmpredictors* v.0.2.8 [74]. We selected a broad suite of 29 oceanographic variables (including temperature, current velocity, salinity, and pH; complete list in Table C in S1 File), including those used by Xuereb et al. [45] for comparison. Where possible, oceanographic variables for both sea surface and mean bottom depth (near-bottom) were used to account for the conditions experienced by pelagic larvae and benthic adults respectively. The Eld Inlet, Washington collection site was removed from these analyses because environmental predictor data were not available. We also calculated Pearson's correlation coefficients among predictor variables (Table D in S1 File) using the R package *psych* [75] and interpreted results in light of these correlations.

To investigate genomic evidence for local adaptation, gene-environment associations were explored using a univariate association method, *Bayenv2* [70], and a multivariate method, redundancy analysis (RDA). We used both methods as each has an advantage: the interpretation of results for univariate methods like *Bayenv2* can be clearer than multivariate methods for gene-environment association, and multivariate methods such as RDA have lower false positive rates and greater sensitivity for detecting weak and multi-locus selection [75].

The program *Bayenv2* [76] uses a univariate Bayesian framework to test for significant correlation between allele frequencies and environmental predictor variables, accounting for population structure by first estimating covariance among loci. Correlations with a minimum Bayes Factor of 10, or minimum "strong" support [77], were retained in the analysis.

Redundancy analyses were performed in the R package *vegan* v.2.5–6 [78]. Redundancy analysis summarizes the variation in a set of response variables (here, the allele frequencies) due to a set of explanatory variables (here, the oceanographic variables), using an extension of multiple linear regression that allows regression of multiple response variables on multiple explanatory variables. Here, only biallelic SNPs were retained, and allele frequencies were Hellinger-transformed prior to RDA [79]. To avoid overdetermination of RDA models with many environmental predictors (Fig B in S1 File), we conducted multiple RDAs on subsets of environmental predictors and reduced the dimensionality of our environmental predictors within sets by first combining them into orthogonal principal components (Tables E–H in S1 File). Specifically, environmental predictors were grouped into four *a priori* sets: (1) all, (2) sea surface, (3) near-bottom and (4) current velocity and temperature predictors (measured at either sea surface or near-bottom). Sets 2 and 3 were chosen to investigate differences in putative adaptation at pelagic and benthic life history stages. Environmental predictor set 4 was chosen for comparison with the existing study on *A. californicus* in a different part of the species range, as these variables were strongly correlated with genetic population structure [45].

Predictors were standardized to a mean of 0 and standard deviation of 1 prior to PCA. PCA was performed on each set of predictors, and principal components were retained if their corresponding eigenvalue was above the mean eigenvalue across principal components.

To account for the potential confounding factor of neutral population structure in our RDA, we conducted a complementary partial RDA for each set of predictor variables, in which the effects of spatial variation were partialled out. We used Euclidean distances among collection sites to compute distance-based Morgan's eigenvector maps (MEMs) using the R package *codep* v.0.9–1 [80], to be used as conditioning variables in partial RDAs [81]. We report the variance inflation factors (VIF) to identify collinearity among spatial variables and environmental predictors, although we note that the effects of such collinearity are addressed in the removal of genetic variation explained by spatial variables in partial RDAs. ANOVA permutation tests for full models and per axis were used to assess the significance of RDA results. For significant models, we identified SNPs putatively involved in local adaption based on the loadings of SNPs in ordination space for significant axes. Specifically, SNPs were classified as putatively adaptive if their loading score was outside of 3 standard deviations of the mean [75].

## Comparing putatively neutral and adaptive differentiation

SNPs were classified as putatively adaptive if they were detected as $F_{ST}$ outliers using either *BayeScan* or *OutFLANK*, or if they were significantly correlated to environmental predictors using *Bayenv2* or RDA. Once putatively adaptive SNPs were identified, they were used to distinguish a putatively neutral SNP set and a putatively adaptive SNP set.

To address questions related to demographic and selective processes, we used the same methods on putatively neutral and putatively adaptive data sets separately in testing for IBD, estimating dispersal distance, and estimating effective population size per collection site. We estimated effective population size per collection site with *NeEstimator* v.2.1, using the linkage disequilibrium method and default settings [82].

To build hypotheses for the mechanisms underlying adaptive differentiation, potential biological processes associated with putatively adaptive SNPs were identified using *blastx* v.2.5.0 [83] and the UniProt Knowledge Base (Swiss-Prot, manually annotated) [84]. We queried the 2000 *bp* region flanking the SNP, following alignment against the reference genome of *A*. *parvimensis*. Matches with a maximum *e*-value score of $10^{-10}$ were retained. Gene ontology slim terms for biological processes were retrieved using an adaptation of the Mouse Genome Informatics database, as developed by Gavery and Roberts [85]. Gene ontology slim terms "other biological processes" and "other metabolic processes" were excluded. A particular locus could be associated with multiple gene ontologies, and a particular gene ontology could be associated with multiple gene ontology slim terms for biological processes. All matches were retained.

## Simulations

To contextualize genetic connectivity in terms of migration rate for management, we developed a simulation model using *simuPOP* v.1.1.10.9 [86] in Python v.3.7.3 to determine which population sizes, migration rates, and number of generations of drift reproduced our empirically derived pairwise $F_{ST}$ results. From an infinitely large ancestral population with global allele frequencies based on our empirical data, we sampled two populations of a specified size and carried those through forward-time simulations with discrete generations, random mating, and no selection. Within each population, two parents were selected at random with replacement to produce one offspring, leading to a random distribution of reproductive success, and allowing for census size to approximate effective population size. The model was parameterized using empirical global allele frequencies for putatively neutral SNPs and

simulations were run for each combination of effective population size (500, 2500, 10,000) and migration rate (0.01%, 0.03%, 0.1%, 0.3%, 1%, 3%, 10%, 30%). Additionally, simulations were run for 10 (short-term), 100 (medium-term), and 1,000 (long-term) generations of drift. The long-term option was chosen to approximate equilibrium conditions, although true equilibrium depends on population history and may not be reached in some wild populations within 1,000 generations. Pairwise $F_{ST}$ was calculated after the pre-determined number of generations of drift had elapsed and averaged across 5 replicates for each parameter combination.

## Results

### Sequencing

After removing 33 (9% of 358 sequenced individuals) poorly sequenced individuals, the average number of reads per individual was 1.76M (standard deviation (SD) = 0.787M). The *dDocent* assembly produced 6,738,423 variant sites, and 2,075 SNPs were ultimately retained after filtering (Table I in S1 File). All individuals included in the *dDocent* assembly were retained after filtering for missing data across SNPs, resulting in an average of 36 individuals per population (SD = 13.7) (Table 1). Genotyping error was estimated to be 1.4% among the 14 replicated individuals. All errors were mismatches of a single allele. Errors were distributed fairly evenly across SNPs, with one error at 246 SNPs, two errors at 19 SNPs, three errors at 7 SNPs, and four errors at one SNP, across replicated individuals. After all filtering steps were completed, the mean missing data per individual and per site was 4.9%, 0.5% of SNPs were multiallelic, and 194 SNPs were out of HWE in one collection site (and 0 SNPs in two or more collection sites) when applying the same HWE filtering method (i.e., FDR approach, *q*-value = 0.05).

### Patterns in genetic diversity

Global population differentiation was highly significant using all SNPs (global $F_{ST} = 0.0068$, 95% CI [0.005, 0.0105]; global genic differentiation test, $p < 0.001$), although clustering analyses with *ADMIXTURE* provided the strongest support for the model with a single underlying population (Fig C in S1 File). For the DAPC using all SNPs, we retained 53 PCs based on the optimization algorithm and observed clustering by collection site (Fig 2). The DAPC showed evidence for separation of Alaska, British Columbia, Washington, and Oregon sites, with some evidence for separation between sites inside and outside of the Salish Sea and between sites north and south of Admiralty Inlet within the Salish Sea. Permutation test results from AMOVAs demonstrated significant population structure for the NPC grouping using all SNPs (Table 2), suggesting that the North Pacific Current is an oceanographic barrier to dispersal in *A. californicus*. Significant population structure for the State grouping was also detected using permutation test results from AMOVAs using all SNPs. The State grouping also had the highest among-group differentiation compared to within-group differentiation ($\Phi_{CT} / \Phi_{SC}$).

A Mantel test revealed significant positive correlations (adjusted $R^2 = 0.5622$, $p < 0.001$, $y = (3.2^{*}10^{-6})^{*}x + 3.5^{*}10^{-3}$) between pairwise linearized $F_{ST}$ using all SNPs and in-water distance (Fig 3A). This IBD was also evident in the northern part of the range (i.e., Alaska and BC, Fig 3B; adjusted $R^2 = 0.8433$, $p < 0.05$, $y = (3.1^{*}10^{-6})^{*}x + 1.8^{*}10^{-3}$) with a remarkably similar slope, but it was not significant in the southern samples (i.e., Washington and Oregon, Fig 3C; adjusted $R^2 = 0.5009$, $p > 0.05$, $y = (4.3^{*}10^{-6})^{*}x + 3.2^{*}10^{-3}$). Assuming an effective population density between 100 and 10,000 adults per *km*, this slope suggested a mean dispersal distance per generation of 2 to 20 *km* (Table 3).

Of all 36 pairwise site comparisons using all SNPs (Table 4), 24 revealed significant genic differentiation ($p < 0.05$), with 8 of the 12 not statistically significant tests corresponding to

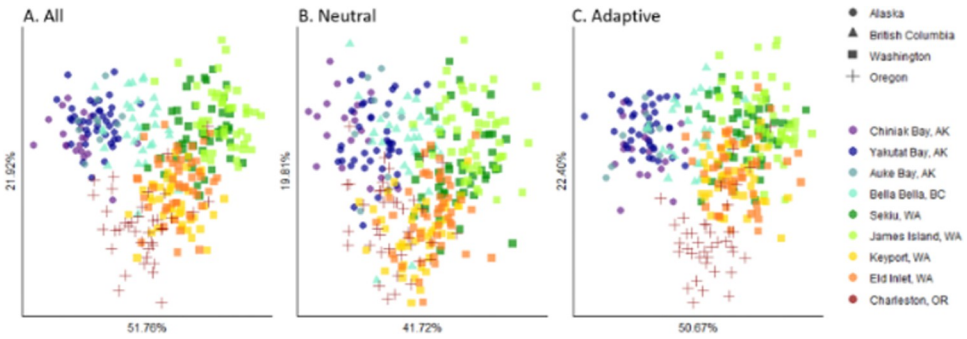


**Fig 2.** Discriminant analyses of principal components using all SNPs (A), putatively neutral SNPs (B), and putatively adaptive SNPs (C). Each plot represents the first two discriminant functions. For each plot, axes are labeled with the proportion of among-population variance explained by that discriminant function. Points represent individuals and are colored by collection site, matching the colors of collection sites in the map (Fig 1). Point shapes differ by state to highlight regional effects.

https://doi.org/10.1371/journal.pone.0280500.g002

collection site pairings with the small collection from Auke Bay, AK (Table 1). Only one pairwise site comparison (Keyport and Eld Inlet, WA) yielded a pairwise $F_{ST}$ value that contained 0 in the 95% confidence interval (Table J in S1 File). The largest pairwise $F_{ST}$ values were found among collection sites from the northern (Alaska and British Columbia) and southern (Washington and Oregon) parts of the species range ($F_{ST}$ = 0.005–0.015). Notably, pairwise differentiation between collection sites within Washington was broader in range and at times greater than that within Alaska, despite substantially smaller in-water distances (mean = 162 *km*, SD = 62 *km* in WA and mean = 731 *km*, SD = 339 in AK).

Mean expected heterozygosity and proportion of polymorphic SNPs did not vary substantially across collection sites, apart from a reduced proportion of polymorphic SNPs in Auke Bay and Chiniak Bay, AK (Table 1). Observed heterozygosity decreased and $F_{IS}$ increased from

**Table 2. Results of hierarchical population structure, based on AMOVA results.**

| *Grouping* | *SNPs* | $\Phi_{CT}$ | $\Phi_{SC}$ | $\Phi_{CT}$ / $\Phi_{SC}$ |
|---|---|---|---|---|
| NPC | All | 0.0040* | 0.0043* | 0.944 |
| | Neutral | 0.0016* | 0.0020* | 0.804 |
| | Adaptive | 0.0253* | 0.0247* | 1.024 |
| Victoria Sill | All | 0.0012 | 0.0056* | 0.207 |
| | Neutral | 0.0006 | 0.0025* | 0.234 |
| | Adaptive | 0.0063 | 0.0334* | 0.187 |
| Admiralty Inlet | All | 0.0007 | 0.0059 | 0.123 |
| | Neutral | 0.0005 | 0.0026 | 0.197 |
| | Adaptive | 0.0026 | 0.0357 | 0.074 |
| All barriers | All | 0.0016 | 0.0052* | 0.308 |
| | Neutral | 0.0012* | 0.0021* | 0.585 |
| | Adaptive | 0.0052* | 0.0335* | 0.155 |
| State | All | 0.0045* | 0.0032* | 1.424 |
| | Neutral | 0.0015* | 0.0018* | 0.833 |
| | Adaptive | 0.0313* | 0.0154* | 2.032 |

Column Grouping contains four groupings considered with AMOVA: North Pacific Current (NPC), Victoria Sill, Admiralty Inlet, State, and All Barriers. Column SNPs refers to which data set was used in the AMOVA: all SNPS (All), putatively neutral SNPs (Neutral), or putatively adaptive SNPs (Adaptive). Variations among groups ($\Phi_{CT}$) and among collection sites within groups ($\Phi_{SC}$) reported, and significance of permutation tests noted with *: $p < 0.05$. The ratio of $\Phi_{CT}$ / $\Phi_{SC}$ also reported.

https://doi.org/10.1371/journal.pone.0280500.t002

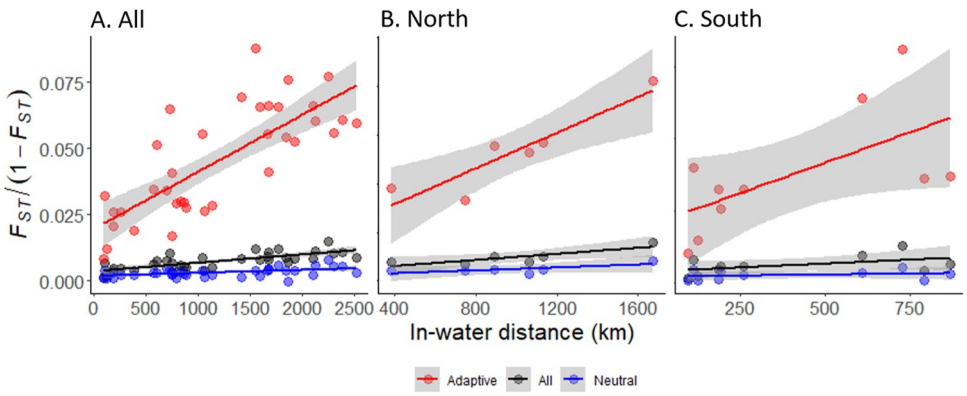


**Fig 3.** Test for isolation by distance for all collection sites (A), northern collection sites (B), and southern collection sites (C). Correlation between in-water distance and linearized $F_{ST}$ using all SNPs in black, putatively neutral SNPs in blue, and putatively adaptive SNPs in red.

https://doi.org/10.1371/journal.pone.0280500.g003

North to South, while expected heterozygosity was similar across the range (Table 1). All collection site $N_e$ estimates were large or infinite, with infinite upper confidence limits, using both all putatively neutral SNPs and only those with a minor allele frequency of at least 5% (Table 5).

## Adaptive differentiation

The $F_{ST}$ outlier SNP detection methods identified 60 (2.9% of 2,075) SNPs as putatively adaptive (Fig 4), with 50 identified by *BayeScan*, 46 identified by *OutFLANK*, and 36 SNPs overlapping across both methods.

Using *Bayenv2*, 181 SNPs (8.7% of 2,075) were identified as putatively adaptive. Each of these SNPs was correlated with an average of 2.5 environmental predictors and at least one SNP was correlated with each of the 29 environmental predictors (Table K in S1 File). The five environmental variables with the most associated SNPs included mean salinity, mean nitrate, temperature range, and pH, all at the sea surface, and mean temperature at near-bottom (Table 6). Over 60% more SNP-predictor correlations represented environmental variables measured at the surface than at near-bottom, for variables with both measurements: 397 and 615 SNP-predictor correlations with near-bottom and sea surface variables, respectively.

After removing 10 multiallelic SNPs, we retained 2,065 biallelic SNPs for use in redundancy analyses (RDA). We retained two RDA models (out of 8) for statistical significance: the model

**Table 3. Dispersal distance results.**

| | Dispersal distance (*km*) | | |
|---|---|---|---|
| **Density** | **All** | **Neutral** | **Adaptive** |
| 1 | 198 | 337 | 75 |
| 10 | 63 | 107 | 24 |
| 100 | 20 | 34 | 8 |
| 1,000 | 6 | 11 | 2 |
| 10,000 | 2 | 3 | 1 |
| 100,000 | 1 | 1 | 0 |

Mean dispersal distances per generation in km, estimated using all, putatively neutral, and putatively adaptive SNPs, depending on effective population density (adults / km).

https://doi.org/10.1371/journal.pone.0280500.t003

**Table 4. Pairwise $F_{ST}$ and genic differentiation test results.**

| | CB_AK | YB_AK | AB_AK | BB_BC | SK_WA | JI_WA | KP_WA | EI_WA | CH_OR |
|---|---|---|---|---|---|---|---|---|---|
| CB_AK | - | | | | * | * | * | * | * |
| YB_AK | 0.004 | - | | * | * | * | * | * | * |
| AB_AK | 0.004 | 0.004 | - | | | | | | |
| BB_BC | 0.008 | 0.005 | 0.005 | - | * | * | * | * | * |
| SK_WA | 0.011 | 0.009 | 0.008 | 0.006 | - | | * | * | * |
| JI_WA | 0.015 | 0.012 | 0.012 | 0.007 | 0.002 | - | * | * | * |
| KP_WA | 0.010 | 0.009 | 0.008 | 0.007 | 0.004 | 0.006 | - | | * |
| EI_WA | 0.011 | 0.008 | 0.011 | 0.006 | 0.005 | 0.005 | 0.001 | - | * |
| CH_OR | 0.009 | 0.008 | 0.007 | 0.009 | 0.008 | 0.010 | 0.004 | 0.005 | - |

Pairwise $F_{ST}$ in the lower left and pairwise genic differentiation test results in the upper right of the table (*: $p < 0.05$). Site abbreviations provided in Table 1.

https://doi.org/10.1371/journal.pone.0280500.t004

using sea surface predictors was marginally significant at the full model level (ANOVA, $p = 0.063$) and for the first axis (ANOVA, $p = 0.052$), and the model using temperature and current velocity predictors (at sea surface or near-bottom) was significant at the full model level (ANOVA, $p = 0.034$) and for the first axis (ANOVA, $p = 0.035$). For the RDA using sea surface predictors, the first RDA axis explained 24.8% of the variance, and PC 2 had the greatest loading on the first RDA axis (Table L in S1 File). The three predictors with the greatest loadings on PC 2 included pH, mean salinity, and temperature range (Table E in S1 File). For the RDA using temperature and current velocity predictors, the first RDA axis explained 26.8% of the variance, and PC 2 had the greatest loading on the first RDA axis (Table M in S1 File). The three predictors with the greatest loading on PC 2 included temperature range at the sea surface and near-bottom, and minimum temperature at the near-bottom (Table H in S1 File). Collection site clustering patterns were similar among RDA biplots (Fig 5), including the Salish Sea collection sites clustering tightly together and North-South separation driven by the first RDA axis in each model. Charleston, Oregon clustered more closely with the Salish Sea collection sites in the model using temperature and current velocity predictors (Fig 5). We identified 32 putatively adaptive SNPs based on loadings of significant (and marginally significant) RDA axes, with 24 SNPs identified from the RDA using sea surface predictors and 29 from the RDA using temperature and current velocity predictors. No partial RDAs were

**Table 5. Linkage desquilibrium $N_e$ estimates with 95% confidence intervals.**

| | *$Pcrit_{0.05}$* | | *$Pcrit_{0.0}$* | |
|---|---|---|---|---|
| ***Site*** | ***Estimate*** | ***95% CI*** | ***Estimate*** | ***95% CI*** |
| Chiniak Bay, AK | 1170.4 | 319.6—Infinity | 40375.1 | 389.6—Infinity |
| Yakutat Bay, AK | 6368.8 | 1088.9—Infinity | 21798.0 | 1223.1—Infinity |
| Auke Bay, AK | Infinity | 312.6—Infinity | Infinity | 312.6—Infinity |
| Bella Bella, BC | 1063.5 | 238.1—Infinity | 1453.5 | 248.2—Infinity |
| Sekiu, WA | 16057.9 | 2365.8—Infinity | 21545.6 | 2550.8—Infinity |
| James Island, WA | 181563.8 | 2407.7—Infinity | 25852.1 | 2949.4—Infinity |
| Keyport, WA | Infinity | 4596.7—Infinity | Infinity | 5719.8—Infinity |
| Eld Inlet, WA | 5117.3 | 1605.0—Infinity | 5683.6 | 2055.4—Infinity |
| Charleston, OR | Infinity | 2645.8—Infinity | Infinity | 3272.6—Infinity |

Confidence intervals estimated by jackknifing across samples. Pcrit is the minimal allele frequency to retain a locus in the analysis. Both $Pcrit_{0.05}$ (0.05%) and $Pcrit_{0.0}$ (0%) were included because rare alleles can affect $N_e$ estimates.

https://doi.org/10.1371/journal.pone.0280500.t005

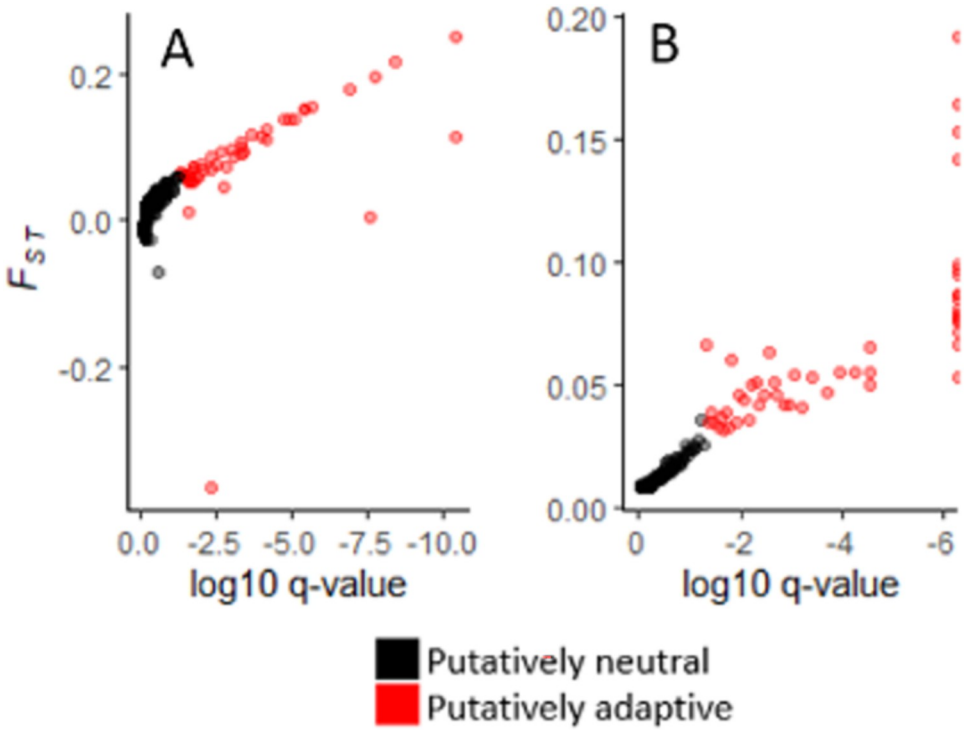


**Fig 4. Results of $F_{ST}$ outlier detection methods.** $Log_{10}$ q-value by $F_{ST}$ for outlier detection methods, using results generated by OutFLANK (n = 46) in panel A and BayeScan (n = 50) in panel B. Each point represents a single SNP, colored by whether it was detected as an outlier (red) or not (black).

https://doi.org/10.1371/journal.pone.0280500.g004

significant. Spatial variables used in partial RDAs covaried with predictor variables in most models, measured with variance inflation factors (Table 7). The environmental predictor loadings for retained PCs in all models and a correlation matrix of environmental predictors can be found in Tables D–H in S1 File.

In total, we identified 211 (10.2% of 2,075 total SNPs) putatively adaptive SNPs, of which 41 (19.4% of 211 putatively adaptive SNPs) were detected with either $F_{ST}$ outlier approaches and

**Table 6. Univariate gene-environment association results.**

| *Predictor* | *Depth* | $S_C$ | $S_P$ | *Biological processes* |
|---|---|---|---|---|
| Mean salinity | S | 38 | 4 | signal transduction (0.259), cell organization and biogenesis (0.207), protein metabolism (0.207), cell cycle and proliferation (0.138), RNA metabolism (0.086), transport (0.069), DNA metabolism (0.034) |
| Mean nitrate | S | 34 | 4 | signal transduction (0.294), cell organization and biogenesis (0.235), developmental processes (0.216), cell cycle and proliferation (0.078), transport (0.078), protein metabolism (0.059), DNA metabolism (0.039) |
| Temperature range | S | 31 | 2 | cell organization and biogenesis (0.387), signal transduction (0.29), RNA metabolism (0.097), developmental processes (0.065), DNA metabolism (0.065), protein metabolism (0.065), stress response (0.032) |
| pH | S | 30 | 3 | cell organization and biogenesis (0.353), developmental processes (0.324), signal transduction (0.265), DNA metabolism (0.059) |
| Mean temperature | B | 27 | 1 | signal transduction (0.5), transport (0.5) |

Putatively adaptive loci associated with environmental variables from Bayenv2 for five environmental predictors (column Predictor) with most correlated SNPs. Column Depth refers to whether the environmental variable is measured at sea surface (S) or mean bottom depth (B). Column $S_C$ contains the number of SNPs with evidence of correlation to the row's variable. Column $S_P$ contains the number of correlated SNPs that also had matching gene ontology SLIM terms for biological processes. Column Biological processes contains the associated biological processes from matching GO slim terms in order of decreasing frequency, with the proportion of matches per process noted in parentheses.

https://doi.org/10.1371/journal.pone.0280500.t006

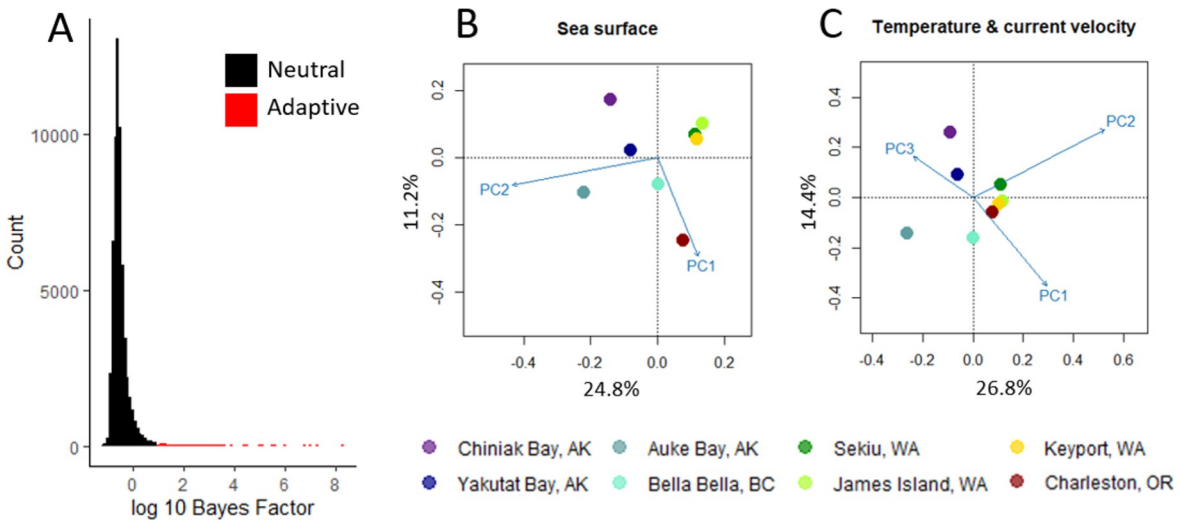


**Fig 5. Summary of the results of univariate (Bayenv2) and multivariate (RDA) gene-environment associations.** Panel A is a histogram of log10 Bayes Factors for each combination of SNP and environmental predictor. Panel B and C are RDA biplots of the first two RDA axes, with predictors as vectors and sites as points in ordination space, using Type 1 scaling, a scaling method appropriate for questions related to distance among objects. Panel B is of the retained model using sea surface predictors, and panel C is of the retained model using temperature and current velocity predictors. For each RDA biplot, the proportion of variance explained by each RDA axis is labeled on the axis.

https://doi.org/10.1371/journal.pone.0280500.g005

at least one gene-environment association approach and 134 (63.5% of 211) were only identified by *Bayenv2* (Fig D in S1 File).

Only 19 (9% of 211) putatively adaptive SNPs matched to gene ontologies, and 15 (79% of 19) matched to more than one gene ontology. Of the SNPs identified as putatively adaptive using *Bayenv2*, 15 SNPs matched to gene ontologies (Table 6, Table K in S1 File). Of these 15 SNPs, three were also identified using *Bayescan* and *OutFLANK* and one was also identified using *Bayescan*, *OutFLANK*, and RDA. An additional four SNPs were also identified as putatively adaptive using only $F_{ST}$ outlier methods, with the most represented biological processes including DNA metabolism, cell organization and biogenesis, and developmental processes. For comparison, the top five most frequent biological processes in the database in descending

**Table 7. Results of the redundancy analyses (RDA).**

| *RDA* | *Predictors* | *Partial* | $R^2_{adj}$ | $p_{full}$ | $p_{axis}$ | *VIF > 10* |
|---|---|---|---|---|---|---|
| 1 | all | Yes | 0.094 | >0.1 | >0.1 | 1 MEM & 2 PCs |
| 2 | all | No | 0.07 | >0.1 | >0.1 | - |
| 3 | surface | Yes | 0.16 | >0.1 | >0.1 | 2 MEMs & 2 PCs |
| 4 | surface | No | 0.11 | 0.063 | 0.052–0.606 | - |
| 5 | bottom | Yes | -0.081 | >0.1 | >0.1 | - |
| 6 | bottom | No | -0.0007 | >0.1 | >0.1 | - |
| 7 | cv & temp | Yes | 0.087 | >0.1 | >0.1 | 1 MEM & 2 PCs |
| 8 | cv & temp | No | 0.14 | 0.034 | 0.035–0.695 | - |

Column RDA assigns a number to each RDA for reference; Predictors refers to the set of environmental predictors used in the RDA, where all refers to all 29 predictors, surface refers to sea surface, bottom refers to mean bottom depth, and cv & temp refers to current velocity and temperature predictors; Partial refers to whether spatial variables were conditioned in a partial RDA; $R^2_{adj}$ refers to the adjusted $R^2$ value; $p_{full}$ and $p_{axis}$ refer to the p-value from the ANOVA on the full model and for axes, respectively; and VIF > 10 refers to whether any variance inflation factors (VIF) were over ten, suggesting substantial co-variance among explanatory variables, where MEM refers to spatial variables (Morgan's eigenvector maps) and PC refers to principal components.

https://doi.org/10.1371/journal.pone.0280500.t007

order are 1) developmental processes, 2) cell organization and biogenesis, 3) transport, 4) protein metabolism, and 5) stress response (Table N in S1 File).

## Neutral vs. putatively adaptive differentiation

DAPC and AMOVA revealed higher differentiation but similar spatial patterns for putatively adaptive SNPs compared to putatively neutral SNPs (Table 2, Fig 2). We retained 53 and 26 PCs for the DAPC using putatively neutral SNPs and putatively adaptive SNPs, respectively, based on the optimization algorithm. Clustering patterns using putatively neutral and putatively adaptive SNPs reflected those using all SNPs, with more distinct clustering using putatively adaptive SNPs (Fig 2). Permutation test results from AMOVAs demonstrate significant population structure for NPC and State groupings, using putatively neutral and adaptive SNPs (Table 2). The ratio of among-group variation to within-group variation ($\Phi_{CT} / \Phi_{SC}$) was higher using putatively adaptive SNPs for the State and NPC groupings, and higher using putatively neutral SNPs for the Victoria Sill and Admiralty Inlet groupings. Mantel tests for correlation among linearized $F_{ST}$ and in-water distance were significant using putatively neutral SNPs (adjusted $R^2 = 0.2217$, $p < 0.01$, $y = (1.1^*10^{-6})^*x + 1.7^*10^{-3}$) and putatively adaptive SNPs (adjusted $R^2 = 0.5786$, $p < 0.001$, $y = 2.2^*10^{-5} + 2.0^*10^{-2}$) (Fig 3A). Similar to analyses with all SNPs, slopes for adaptive SNPs were similar across the entire range and in Alaska and BC (adjusted $R^2 = 0.9187$, $p < 0.05$, $y = (1.7^*10^{-5})^*x + 8.8^*10^{-3}$), and not significant in samples from Oregon and Washington (adjusted $R^2 = 0.621$, $p > 0.05$, $y = (3.1^*10^{-5})x + 1.6^*10^{-2}$; Fig 3). Mantel tests using putatively neutral SNPs were not significant when performed only on northern collection sites (adjusted $R^2 = 0.6757$, $p > 0.05$, $y = (1.5^*10^{-6})^*x + 9.5^*10^{-4}$) or southern collection sites (adjusted $R^2 = 0.2631$, $p > 0.05$, $y = (1.1^*10^{-6})^*x + 1.8^*10^{-3}$). Correlation among linearized $F_{ST}$ and in-water distance was strongest and the slope of the linear regression was greatest using putatively adaptive SNPs: the slope of the linear regression using putatively adaptive SNPs was nearly 7 and 20 times greater than the slopes using all SNPs and putatively neutral SNPs, respectively. Assuming a population density between 100 and 10,000 adults per *km*, we estimated mean dispersal distance to be between 3 and 34 *km* using putatively neutral SNPs and between 1 and 8 *km* using putatively adaptive SNPs (Table 3).

## Simulations

Simulations revealed that in an idealized two-population system, simulated pairwise $F_{ST}$ within the range observed in this study occurred in the long-term (1,000 generations since divergence) with small effective population size ($N_e = 500$) and high migration ($m \geq 3\%$), with moderate effective population size ($N_e = 2{,}500$) and moderate migration ($1\% \geq m \geq 3\%$), and with large effective population size ($N_e = 10{,}000$) and small to moderate migration ($0.1\% \geq m \geq 1\%$). Most simulations that produced pairwise $F_{ST}$ estimates within the range observed in this study were cases with moderate to large effective population size ($N_e \geq 2{,}500$) and migration rates below 10%. Simulated pairwise $F_{ST}$ remained below the lowest empirically observed pairwise $F_{ST}$ in this study in the long-term (1,000 generations) only when as effective population size was at least moderate ($N_e \geq 2{,}500$) and migration rate was high ($m \geq 10\%$). Under realistic conditions (e.g., $N_e = 10{,}000$ and 1,000 generations since divergence; our $N_e$ estimates were large and unbounded), simulated pairwise $F_{ST}$ was within the range of empirical pairwise $F_{ST}$ when migration rates were 0.1–1.0% (Fig 6).

# Discussion

In this study, we quantified patterns of population differentiation in *A. californicus* and investigated potential drivers of differentiation from Alaska to Oregon, representing more than half

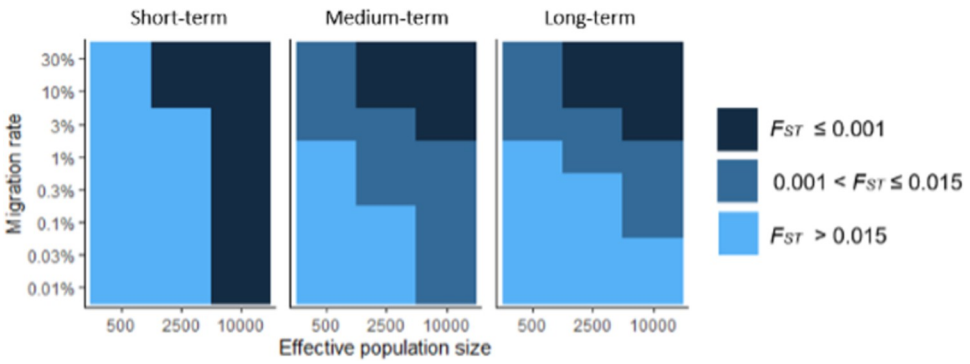


**Fig 6. Results of simulations that investigated demographic conditions that reproduced empirical $F_{ST}$ results.** Tile maps represent pairwise $F_{ST}$ for each combination of effective population size, migration rate, and number of generations of drift elapsed. Short-term (left) represents results from 10 generations of drift, Medium-term (middle) from 100 generations of drift, and Long-term (right) from 1,000 generations of drift. Tile color is divided into three groups based on the scale of $F_{ST}$: smaller than the observed $F_{ST}$ values in this study, within the range of observed pairwise $F_{ST}$, and greater than the observed pairwise $F_{ST}$.

https://doi.org/10.1371/journal.pone.0280500.g006

of the species' range and regions not yet surveyed. We observed population genetic structure at both fine- and broad-scales, likely driven by limited dispersal and local adaptation. Notably, we found detectable differentiation at small scales within the Salish Sea, co-varying signals of adaptive and neutral differentiation, and a latitudinal pattern in genetic diversity. Estimates of effective population size were large. Using simulations, we found that migration rates may be below 1% if populations of *A*. *californicus* are large ($N_e$ = 10,000) and have undergone many (1,000) generations of drift after splitting from a single ancestral population. These estimates corresponded well to average dispersal distances per generation from isolation-by-distance patterns, which were on the scale of dozens of kilometers despite the long pelagic larval duration.

## Broad- and fine-scale population genetic structure and potential drivers

Broad-scale population structure that we observed can be described by two patterns: differentiation across the bifurcation zone of the North Pacific Current and isolation-by-distance. Many species of marine invertebrate exhibit population differentiation along a latitudinal gradient, with the region between Alaska and Oregon as a known region of divergence [87, 88]. Xuereb et al. [45] posit that the genetic break observed in *A*. *californicus* in British Colombia is due to limited dispersal across the bifurcation zone of the North Pacific Current. The North Pacific Current has also been identified as an oceanographic barrier to gene flow in the Bat Star, *Patiria miniata* [44, 89], and the Rosethorn Rockfish, *Sebastes helvomaculatus* [90]. In addition to this break, Xuereb et al. [22] detected IBD in *A*. *californicus* along the coast of British Columbia [22], but only across the entire study range, not within the southern and northern population component, suggesting that IBD was an artefact of population subdivision rather than a signal of limited dispersal in a continuous population [91]. Here, IBD was found across the entire range from Alaska to Oregon as well as in the section north of the North Pacific current bifurcation, but not among more southern samples. IBD was also found in other species of sea cucumbers including *Holothuria edulis* [92], *H*. *scabra* [93], and *H*. *nobilis* [94]. Our results did not provide evidence for limited dispersal across the Victoria Sill or the sills at Admiralty Inlet, despite their potential roles in shaping population structure in other marine species [40, 41, 95, 96]. Although biologically not meaningful, the State grouping yielded the greatest among-group variation, likely because it captured the effects of isolation-by-distance

and the divergence across the NPC bifurcation zone. While this result may also be caused by our specific sampling design (e.g., only one sample from BC and Oregon, all Washington samples within the Salish Sea), it lends support to current management by individual state agencies.

Additionally, we found a subtle pattern of decreasing observed heterozygosity, constant expected heterozygosity and increasing $F_{IS}$ from north to south, a pattern that can also be observed in the data of Xuereb et al. [65]. In temperate species, genetic diversity usually declines polewards as a consequence of successive founder events during postglacial recolonization [97], opposite to the pattern observed here. Higher $F_{IS}$ values in the south could be an effect of selection against heterozygotes caused by high genetic load of deleterious mutations, as commonly observed in bivalves [98], but also other highly fecund marine species such as echinoderms [99]. Such selection may be more severe in warmer waters for cold adapted species, as environmental stress increases selection against deleterious mutations [100], thus explain higher $F_{IS}$ values in southern samples. Heterozygote deficiencies can also be caused by mixtures (Wahlund effect) of populations, cohorts or local demes [53]–although population mixture appears unlikely given generally weak population structure, mixtures of cohorts in mixed-age samples produced by sweepstake recruitment could produce heterozygote deficiencies [101] in the south. Population mixtures could also be caused by temporally varying recruitment sources or selection [102, 103]. Alternative explanations such as inbreeding or assortative mating appear less likely because estimates of effective population size all had infinite upper confidence limits, Alaska supports larger sea cucumber fisheries compared to southern regions [104], and expected heterozygosity was similar between northern and southern samples.

Notably, we found that geographic distance was more strongly correlated to genetic distance among collection sites using putatively adaptive SNPs than using putatively neutral SNPs, though neither relationship was significant in the southern samples. The higher slope of the relationship resulted in lower estimates of mean dispersal estimates using putatively adaptive SNPs than putatively neutral SNPs. This finding suggests that selection gradients may reduce effective dispersal distances, leading to stronger signals in isolation-by-distance from neutral processes alone [105]. One exception may be the Charleston, Oregon collection site, which clusters more distinctly from other collection sites using putatively adaptive SNPs and is positioned more closely to the Salish Sea collection sites in the RDA biplot for temperature and current velocity predictors than the biplot for sea surface predictors. The difference in RDA biplots suggests different patterns in adaptive differentiation, dependent on environmental predictors. Lastly, co-variation between neutral and adaptive differentiation could also represent a false positive result for gene-environment association. Isolation-by-distance, as we found in this study, can increase the chance of a false negative result for gene-environment association if the environmental predictor is spatially auto-correlated [106].

At finer scales, we found detectable differences among most collection site pairs and between collection sites close together as 110 *km* (in-water distance) apart, consistent with earlier findings on differentiation as close as 60 *km* in *A*. *californicus* [22] and 100 *km* apart in the Tar-spot Sea Cucumber, *Cucumaria pseudocurata* [107]. The sample from Auke Bay, Alaska was an outlier across multiple analyses, likely due to stochasticity associated with smaller sample size, and leading to most insignificant pairwise genic differentiation tests and a lower proportion of polymorphic SNPs. Pairwise differentiation was lower in Alaska despite larger in-water distances between collections, possibly because of longer larval periods in colder Alaskan waters resulting in more extensive dispersal. Colder temperatures led to slower larval growth in laboratory experiments in the sister species, the Japanese sea cucumber, *A*. *japonicus* [108], and meta-analyses in marine fishes showed longer larval duration, larger dispersal distances,

and higher $F_{ST}$ in high latitude species [109]. Alternatively, greater genetic differentiation among populations within Washington compared to Alaska may point to underlying oceanographic barriers that can shape dispersal in complex and asymmetrical ways. In future work, a denser sampling scheme paired with an oceanographic model could be used to illuminate the potential for seascape features to influence dispersal, particularly within the Salish Sea.

We found evidence for putatively adaptive differentiation, corroborated by multiple lines of investigation. The majority of SNPs identified as putatively adaptive using $F_{ST}$ outlier detection were identified using both methods and were also identified as putatively adaptive using at least one gene-environment association method. Using *Bayenv2*, we detected many SNPs that were not identified as putatively adaptive compared to redundancy analysis, consistent with expectations that *Bayenv2* may produce more false positives compared to redundancy analysis [75]. The five environmental predictors with the most significantly correlated SNPs were salinity, nitrate, temperature range, and pH at sea surface and mean temperature at near-bottom. Of these, salinity at the sea surface and temperature at the sea bottom were also identified as potential drivers of local adaptation in the same species in British Columbia [45]. Temperature and salinity are significant factors shaping growth and survival in many marine organisms [110] and have been shown to affect growth and survival in *A*. *japonicus* juveniles [111] and growth, survival, locomotory speed, and metamorphosis in *A*. *japonicus* larvae [108]. The greater number of SNPs significantly correlated to sea surface predictor variables using *Bayenv2* and the occurrence of temperature variables as top predictor variables is consistent with RDA results: the significant RDA models were those with environmental predictors at the sea surface and environmental predictors related to temperature and current velocity. Thus, we hypothesize that selection at early life history stages, when larvae are at the sea surface, drives adaptive differentiation in *A*. *californicus*, with salinity, nitrate, pH, and temperature as possible selection factors.

We note that we cannot exclude balanced polymorphism as an alternative explanation to observed putative adaptive differentiation: it is possible that selective mortality at early life stages leads to observed patterns of putative adaptive differentiation in adults (as were sampled here), but that genetic variation is maintained across time through high gene flow [3]. As mentioned above, heterozygote deficiency could be caused by mixtures of cohorts originating from temporally or spatially varying recruitment events undergoing different selection. However, increasing evidence suggests that selection for locally adapted alleles can occur despite high gene flow [7, 112], particularly in species with large effective population sizes.

In generating hypotheses about which environmental predictors are potential drivers of selection, it is prudent to be cautious and highlight 1) spatial auto-correlation of predictors, particularly in light of correlation between geographic distance and genetic distance using putatively adaptive SNPs as observed here, 2) correlation among environmental predictors and 3) the incomplete genome coverage of reduced representation approaches such as RAD sequencing. Almost all environmental predictor sets contained spatially auto-correlated variables, evidenced by high variance inflation factors for spatial variables and environmental predictor variables in partial RDAs (Table 7). Additionally, the correlation among environmental variables adds uncertainty about which predictors are driving patterns in adaptive differentiation. These correlations highlight the limits of these analyses. Finally, although the usefulness of RAD sequencing for identifying adaptive divergence has been questioned [113], RAD sequencing has been used successfully to reveal genomic signatures of selection in many non-model species [114]. Our results are fairly consistent with those of Xuereb et al. [45] and Xuereb et al. [115] even though they sampled genetic and environmental data from different collection sites. Confidence in gene-environment association results may increase if results are further corroborated in future studies, particularly if using alternative methods such as whole

genome sequencing. In any case, the use of adaptive genetic variation for practical management applications is currently an issue of debate and needs further research [115].

For the environmental variables with the most correlated SNPs, and for such SNPs with associated biological processes, signal transduction appeared among the most common biological processes. Future research investigating the connections between genotype, phenotype, and observed patterns in adaptive differentiation may start with the potential connection between signal transduction pathways and salinity, nitrate, temperature, and pH. Others have identified signal transduction genes as part of an adaptive response to salinity adaptation and stress in sea cucumbers [116]. In *A*. *japonicus*, high salinity conditions were correlated with downregulation of acetylcholinesterase, the enzyme that terminates signal transduction. Inhibition of this enzyme can lead to excessive stimulation of nerve and muscle tissue, ultimately leading to paralysis and even death in some species [117]. Transcriptomic approaches can be used in future studies of wild populations of *A*. *californicus* across environmental gradients or in common garden experiments with varying environmental treatments to build cases for which environmental factors shape adaptive differentiation and through which physiological pathways.

## Spatial considerations for sustainable management of *A*. *californicus*

Simulation results suggest limited dispersal among collection sites: if effective population sizes were at least 10,000 and separated at least 1,000 generations ago, genetic differentiation suggested migration rates of 0.1–1% (Fig 6). Here, our estimates of effective population size were large with infinite upper confidence limits, demonstrating that populations are likely not small [118]. This level of migration (0.1–1%) may be large enough to prevent genetic isolation, but small enough to lead to demographic independence [119]. This has implications for spatial management of wild *A*. *californicus*, though those implications may differ depending on the management aim. For example, marine protected areas for *A*. *californicus* may need to operate at smaller scales than expected from signals of genetic differentiation at broad scales, to maintain demographic connectivity and prevent overexploitation of local populations. Designation of management units for fisheries may need to occur at similar scales, because fishery management units are used to prevent overexploitation of local populations [12], largely a demographic concern. On the other hand, management units designed to protect wild populations as a genetic resource for aquaculture aim to prevent the loss of genetic diversity and fitness of wild populations due to broodstock collection and farm escapees, and may thus be better defined at a larger geographic scale than demographic units for fisheries management.

Gradual differentiation associated with the IBD pattern of population structure poses a unique challenge in determining the spatial scale of management units for fisheries and aquaculture. Without distinct boundaries between populations, it is difficult to delineate management units. However, in species with IBD, genetic variation can be preserved even if arbitrary boundaries are chosen [120]. In *A*. *californicus*, it is likely that IBD and oceanographic barriers such as the NPC shape population structure, such that variation may be gradual in many parts of the range at fine and broad scales, but that oceanographic barriers may create distinct boundaries at fine scales [22]. Assessing population structure at finer scales in areas of interest may illuminate oceanographic barriers to dispersal not yet detected and facilitate delineation of management units. In the absence of further sampling, it may be assumed from a precautionary standpoint that distances over which detectable variation can occur (~100 *km*) may be a useful starting point for defining boundaries in delineation of management units, particularly for fishery management units. As populations are further subdivided for management based on genetic data, decision-makers must also consider the risk of reduced precision

associated with the need to complete stock assessments for more stocks with less data [121]. From a less precautionary standpoint, state and provincial boundaries may also be used in the absence of further data, as clustering patterns and partitioning of variance using AMOVAs suggested that this hierarchical grouping explained the most genetic variation of considered groupings. Due to observed isolation-by-distance, it is likely that additional samples will uncover differentiation at smaller scales than states and provinces.

## Supporting information

**S1 File.**
(PDF)

## Acknowledgments

We thank Brent Vadopalas for contributions to design and implementation of the study. We thank W. Steward Grant and Wei Cheng for providing samples and Vassili Kalashnikov, Michael Ulrich, Larry LeClair, Viviane Barry, Elizabeth Unsell, and Mike McCorkle for collecting samples. We thank Katherine Silliman, Eleni Petrou, Steven Roberts, and Sam White for bioinformatics support and access to computational resources.

## Author Contributions

**Conceptualization:** Andy Suhrbier, Henry Carson, Lorenz Hauser.

**Data curation:** Natalie Lowell, Andy Suhrbier, Henry Carson.

**Formal analysis:** Natalie Lowell, Carolyn Tarpey.

**Funding acquisition:** Natalie Lowell, Andy Suhrbier, Lorenz Hauser.

**Investigation:** Natalie Lowell, Lorenz Hauser.

**Methodology:** Natalie Lowell, Carolyn Tarpey, Samuel May.

**Project administration:** Natalie Lowell, Lorenz Hauser.

**Resources:** Andy Suhrbier, Henry Carson.

**Software:** Carolyn Tarpey.

**Supervision:** Lorenz Hauser.

**Validation:** Natalie Lowell, Andy Suhrbier, Carolyn Tarpey.

**Visualization:** Natalie Lowell.

**Writing – original draft:** Natalie Lowell.

**Writing – review & editing:** Natalie Lowell, Andy Suhrbier, Carolyn Tarpey, Samuel May, Henry Carson, Lorenz Hauser.

## References

1. Hauser L, Carvalho GR. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish Fish. 2008; 9:333–62.
2. Palumbi SR. Marine reserves and ocean neighborhoods: the spatial scale of marine populations and their management. Annu Rev Env Resour. Annual Reviews; 2004; 29:31–68.
3. Sanford E, Kelly MW. Local adaptation in marine invertebrates. Annu Rev Mar Sci. 2011; 3:509–35. https://doi.org/10.1146/annurev-marine-120709-142756 PMID: 21329215

4. D'Aloia CC, Bogdanowicz SM, Francis RK, Majoris JE, Harrison RG, Buston PM. Patterns, causes, and consequences of marine larval dispersal. Proc Natl Acad Sci. National Acad Sciences; 2015; 112:13940–5. https://doi.org/10.1073/pnas.1513754112 PMID: 26508628

5. Gerlach G, Atema J, Kingsford MJ, Black KP, Miller-Sims V. Smelling home can prevent dispersal of reef fish larvae. Proc Natl Acad Sci. National Acad Sciences; 2007; 104:858–63. https://doi.org/10.1073/pnas.0606777104 PMID: 17213323

6. Costello MJ, Chaudhary C. Marine biodiversity, biogeography, deep-sea gradients, and conservation. Curr Biol. Elsevier; 2017; 27:R511–27.

7. Lamichhaney S, Barrio AM, Rafati N, Sundström G, Rubin C-J, Gilbert ER, et al. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. Proc Natl Acad Sci. National Academy of Sciences; 2012; 109:19345–50. https://doi.org/10.1073/pnas.1216128109 PMID: 23134729

8. Pettersson ME, Rochus CM, Han F, Chen J, Hill J, Wallerman O, et al. A chromosome-level assembly of the Atlantic herring genome—detection of a supergene and other signals of selection. Genome Res [Internet]. 2019 [cited 2020 Jun 24]; Available from: http://genome.cshlp.org/content/early/2019/10/24/gr.253435.119 PMID: 31649060

9. Barth JMI, Villegas-Ríos D, Freitas C, Moland E, Star B, André C, et al. Disentangling structural genomic and behavioural barriers in a sea of connectivity. Mol Ecol. 2019; 28:1394–411. https://doi.org/10.1111/mec.15010 PMID: 30633410

10. Carvalho GR, Hauser L. Molecular genetics and the stock concept in fisheries. Rev Fish Biol Fish. 1994; 4:326–50.

11. Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish Fish. 2009; 10:361–95.

12. Spies I, Punt AE. The utility of genetics in marine fisheries management: a simulation study based on Pacific cod off Alaska. Can J Fish Aquat Sci [Internet]. NRC Research Press; 2015 [cited 2021 Feb 27]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/cjfas-2014-0050

13. Kenchington E, Heino M, Nielsen EE. Managing marine genetic diversity: time for action? ICES J Mar Sci. Oxford University Press; 2003; 60:1172–6.

14. Schindler DE, Hilborn R, Chasco B, Boatright CP, Quinn TP, Rogers LA, et al. Population diversity and the portfolio effect in an exploited species. Nature. 2010; 465:609–12. https://doi.org/10.1038/nature09060 PMID: 20520713

15. Rousset F. Inferences from Spatial Population Genetics. Handb Stat Genet [Internet]. American Cancer Society; 2004 [cited 2020 Jun 16]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470022620.bbc24

16. Palumbi SR. Population Genetics, Demographic Connectivity, and the Design of Marine Reserves. Ecol Appl. 2003; 13:146–58.

17. Davis B, Allee B, Amend D, Bachen B, Davidson B, Gharrett T, et al. Alaska Department of Fish and Game Genetic Policy [Internet]. 1985. Available from: http://www.adfg.alaska.gov/static/fishing/PDFs/research/genetics_finfish_policy.pdf

18. Tringali MD, Bert TM, Cross F, Dodrill JW, Gregg LM, Halstead WG, et al. Genetic policy for the release of finfishes in Florida. Fla Fish Wildl Res Inst Publ. 2007; 33.

19. Waples RS, Hindar K, Hard JJ. Genetic Risks Associated with Marine Aquaculture [Internet]. 2012 Sep. Available from: http://www.westcoast.fisheries.noaa.gov/publications/aquaculture/geneticrisksaquaculturetm119.pdf

20. Ward RD. The importance of identifying spatial population structure in restocking and stock enhancement programmes. Fish Res. 2006; 80:9–18.

21. Miller KM, Supernault KJ, Li S, Withler RE. Population structure in two marine invertebrate species (*Panopea abrupta* and *Strongylocentrotus franciscanus*) targeted for aquaculture and enhancement in british columbia. J Shellfish Res. 2006; 25:33–42.

22. Xuereb A, Benestan L, Normandeau E, Daigle RM, Curtis JM, Bernatchez L, et al. Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). Mol Ecol. 2018; 27:2347–64.

23. Azad AK, McKinley RS, Forster IP, Pearce CM. The California sea cucumber-a potential candidate for aquaculture. World Aquac. World Aquaculture Society; 2014; 45:43–8.

24. Carson HS, Ulrich M, Lowry D, Pacunski RE, Sizemore R. Status of the California sea cucumber (*Parastichopus californicus*) and red sea urchin (*Mesocentrotus franciscanus*) commercial dive fisheries in the San Juan Islands, Washington State, USA. Fish Res. 2016; 179:179–90.

25. Anderson SC, Flemming JM, Watson R, Lotze HK. Serial exploitation of global sea cucumber fisheries. Fish Fish. 2011; 12:317–39.

26. Mueller K. Fishery biology of the sea cucumber *Parastichopus californicus* (Stimpson 1857) from the San Juan Islands, Washington. Bellingham, WA: Lummi Natural Resources Division; 2016.

27. Purcell SW, Mercier A, Conand C, Hamel J-F, Toral-Granda MV, Lovatelli A, et al. Sea cucumber fisheries: global analysis of stocks, management measures and drivers of overfishing. Fish Fish. 2013; 14:34–59.

28. WDFW. Commercial sea cucumber fishery [Internet]. 2020 [cited 2020 Jun 17]. Available from: https://wdfw.wa.gov/fishing/commercial/sea-cucumber

29. DeWeerdt S. Can aquaculture overcome its sustainability challenges? Nature. Nature Publishing Group; 2020; 588:S60–2.

30. Whitefield CR, Hardy SM. Estimates of Reproductive Potential and Timing in California Sea Cucumbers *Parastichopus californicus* (Stimpson, 1857) from Southeast Alaska Based on Natural Spawning. J Shellfish Res. National Shellfisheries Association; 2019; 38:191–9.

31. Cameron JL, Fankboner PV. Reproductive biology of the commercial sea cucumber *Parastichopus californicus* (Stimpson) (Echinodermata: Holothuroidea). II. Observations on the ecology of development, recruitment, and the juvenile life stage. J Exp Mar Biol Ecol. 1989; 127:43–67.

32. Strathmann MF. Reproduction and Development of Marine Invertebrates of the Northern Pacific Coast: Data and Methods for the Study of Eggs, Embryos, and Larvae. University of Washington Press; 2017.

33. Sheehan EV. Motion in the ocean—Paradigm shift in movement ecology requires "sedentary" organisms to be redefined. J Anim Ecol. Wiley Online Library; 2019; 88:816–9.

34. Phillips AC, Boutillier JA. Stock assessment and quota options for the sea cucumber fishery. Invertebr Work Pap Rev Pac Stock Assess Rev Comm PSARC In. 1995;147–67.

35. Keegan BF, O'Connor BDS. Echinodermata. CRC Press; 1985.

36. Winemiller KO. Life history strategies, population regulation, and implications for fisheries management. Can J Fish Aquat Sci. NRC Research Press Ottawa, Canada; 2005; 62:872–85.

37. Winemiller KO, Rose KA. Patterns of life-history diversification in North American fishes: implications for population regulation. Can J Fish Aquat Sci. NRC Research Press Ottawa, Canada; 1992; 49:2196–218.

38. Hedgecock D. 2.5: Does variance in reproductive success limit effective population sizes of marine organisms. Genet Evol Aquat Org [Internet]. 1994 [cited 2017 Jan 17]. p. 122. Available from: https://www.researchgate.net/profile/Dennis_Hedgecock/publication/245970620_Does_variance_in_reproductive_success_limit_effective_population_sizes_of_marine_organisms_In_A/links/5618340108ae044edbad2220.pdf

39. Abbott DP, Haderlie EC. Intertidal Invertebrates of California. Stanford University Press; 1980.

40. Cunningham KM, Canino MF, Spies IB, Hauser L. 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. NRC Research Press; 2009; 66:153–66.

41. Buonaccorsi VP, Kimbrell CA, Lynn EA, Vetter RD. Limited realized dispersal and introgressive hybridization influence genetic structure and conservation strategies for brown rockfish, *Sebastes auriculatus*. Conserv Genet. 2005; 6:697–713.

42. Davenne E, Masson D. Water Properties in the Straits of Georgia and Juan de Fuca [Internet]. 2001 Aug. Available from: https://waves-vagues.dfo-mpo.gc.ca/Library/40587976.pdf

43. Cannon GA, Holbrook JR, Pashinski DJ. Variations in the Onset of Bottom-Water Intrusions over the Entrance Sill of a Fjord. Estuaries. Coastal and Estuarine Research Federation; 1990; 13:31–42.

44. Keever CC, Sunday J, Puritz JB, Addison JA, Toonen RJ, Grosberg RK, et al. Discordant Distribution of Populations and Genetic Variation in a Sea Star with High Dispersal Potential. Evolution. 2009; 63:3214–27. https://doi.org/10.1111/j.1558-5646.2009.00801.x PMID: 19663996

45. Xuereb A, Kimber CM, Curtis JM, Bernatchez L, Fortin M-J. Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. Mol Ecol. 2018; 27:5035–48.

46. Washington Sea Grant. Shellfish Aquaculture in Washington State: Final Report to Washington State Legislature [Internet]. 2015 Dec p. 92. Available from: https://wsg.washington.edu/wordpress/wp-content/uploads/Shellfish-Aquaculture-Washington-State.pdf

47. Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA. Local de novo assembly of RAD paired-end contigs using short sequencing reads. PloS One. 2011; 6:e18561. https://doi.org/10.1371/journal.pone.0018561 PMID: 21541009

48. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013; 22:3124–40. https://doi.org/10.1111/mec.12354 PMID: 23701397

49. Puritz JB, Hollenbeck CM, Gold JR. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ. 2014; 2:e431. https://doi.org/10.7717/peerj.431 PMID: 24949246

50. Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, et al. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol. 2017; 8:907–17.

51. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156–8. https://doi.org/10.1093/bioinformatics/btr330 PMID: 21653522

52. Larson WA, Seeb LW, Everett MV, Waples RK, Templin WD, Seeb JE. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol Appl. 2014; 7:355–69.

53. Waples RS. Testing for Hardy–Weinberg proportions: have we lost the plot? J Hered. Oxford University Press US; 2015; 106:1–19. https://doi.org/10.1093/jhered/esu062 PMID: 25425676

54. Rousset F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour. 2008; 8:103–6. https://doi.org/10.1111/j.1471-8286.2007.01931.x PMID: 21585727

55. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.

56. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. evolution. 1984; 38:1358–70. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x PMID: 28563791

57. Goudet J, Raymond M, de Meeüs T, Rousset F. Testing differentiation in diploid populations. Genetics. Oxford University Press; 1996; 144:1933–40. https://doi.org/10.1093/genetics/144.4.1933 PMID: 8978076

58. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinforma Oxf Engl. 2008; 24:1403–5.

59. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010; 11:94. https://doi.org/10.1186/1471-2156-11-94 PMID: 20950446

60. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011; 12:246. https://doi.org/10.1186/1471-2105-12-246 PMID: 21682921

61. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet. Frontiers; 2015; 6:208.

62. Mantel N, Valand RS. A technique of nonparametric multivariate analysis. Biometrics. JSTOR; 1970;547–58. PMID: 5480664

63. Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics. Oxford University Press; 1997; 145:1219–28. https://doi.org/10.1093/genetics/145.4.1219 PMID: 9093870

64. Google. 'Measure distance" feature [Internet]. Google Maps. 2019 [cited 2019 Aug 15]. Available from: maps.google.com

65. Xuereb A, Benestan L, Normandeau É, Daigle RM, Curtis JMR, Bernatchez L, et al. Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). Mol Ecol. 2018; 27:2347–64.

66. Slatkin M. Isolation by distance in equilibrium and non-equilibrium populations. Evolution. Wiley Online Library; 1993; 47:264–79. https://doi.org/10.1111/j.1558-5646.1993.tb01215.x PMID: 28568097

67. Crispo E, Hendry AP. Does time since colonization influence isolation by distance? A meta-analysis. Conserv Genet. Springer; 2005; 6:665–82.

68. Foll M, Gaggiotti OE. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008. https://doi.org/10.1534/genetics.108.092221 PMID: 18780740

69. Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F ST. Am Nat. 2015; 186:S24–36. https://doi.org/10.1086/682949 PMID: 26656214

70. Günther T, Coop G. Robust identification of local adaptation from allele frequencies. Genetics. 2013; 195:205–20. https://doi.org/10.1534/genetics.113.152462 PMID: 23821598

71. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. 1973; 74:175–95. https://doi.org/10.1093/genetics/74.1.175 PMID: 4711903

72. Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 2015; 24:1031–46. https://doi.org/10.1111/mec.13100 PMID: 25648189

73. Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O. Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. Glob Ecol Biogeogr. 2018; 27:277–84.

74. Bosch S, Tyberghein L, De Clerck O. sdmpredictors: an R package for species distribution modelling predictor datasets. Mar Species Distrib Data Predict Models. 2017; 49.

75. Revelle W, Revelle MW. Package 'psych.' Compr R Arch Netw. 2015; 337:338.

76. Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Mol Ecol. Wiley Online Library; 2018; 27:2215–33. https://doi.org/10.1111/mec.14584 PMID: 29633402

77. Günther T, Coop G. Robust identification of local adaptation from allele frequencies. Genetics. Oxford University Press; 2013; 195:205–20. https://doi.org/10.1534/genetics.113.152462 PMID: 23821598

78. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. Taylor & Francis; 1995; 90:773–95.

79. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, et al. The vegan package. Community Ecol Package. 2007; 10:719.

80. Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of species data. Oecologia. Springer; 2001; 129:271–80. https://doi.org/10.1007/s004420100716 PMID: 28547606

81. Guénard G, Legendre P, Boisclair D, Bilodeau M. Multiscale codependence analysis: an integrated approach to analyze relationships across scales. Ecology. Wiley Online Library; 2010; 91:2952–64.

82. Dray S, Legendre P, Peres-Neto PR. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model. Elsevier; 2006; 196:483–93.

83. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour. Wiley Online Library; 2014; 14:209–14. https://doi.org/10.1111/1755-0998.12157 PMID: 23992227

84. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

85. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019; 47:D506–15. https://doi.org/10.1093/nar/gky1049 PMID: 30395287

86. Gavery MR, Roberts SB. Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*. Comp Biochem Physiol Part D Genomics Proteomics. 2012; 7:94–9.

87. Peng B, Kimmel M. simuPOP: a forward-time population genetics simulation environment. Bioinformatics. 2005; 21:3686–7. https://doi.org/10.1093/bioinformatics/bti584 PMID: 16020469

88. Kelly RP, Palumbi SR. Genetic Structure Among 50 Species of the Northeastern Pacific Rocky Intertidal Community. PLOS ONE. Public Library of Science; 2010; 5:e8594. https://doi.org/10.1371/journal.pone.0008594 PMID: 20062807

89. Kyle CJ, Boulding EG. Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal. Mar Biol. 2000; 137:835–45.

90. Sunday JM, Popovic I, Palen WJ, Foreman MGG, Hart MW. Ocean circulation model predicts high genetic structure observed in a long-lived pelagic developer. Mol Ecol. 2014; 23:5036–47. https://doi.org/10.1111/mec.12924 PMID: 25231198

91. Rocha-Olivares A, Vetter RD. Effects of oceanographic circulation on the gene flow, genetic structure, and phylogeography of the rosethorn rockfish. 1999; 56:11.

92. Meirmans PG. The trouble with isolation by distance. Mol Ecol. Wiley Online Library; 2012; 21:2839–46. https://doi.org/10.1111/j.1365-294X.2012.05578.x PMID: 22574758

93. Soliman T, Fernandez-Silva I, Reimer JD. Genetic population structure and low genetic diversity in the over-exploited sea cucumber *Holothuria edulis Lesson*, 1830 (Echinodermata: Holothuroidea) in Okinawa Island. Conserv Genet. 2016; 17:811–21.

94. Uthicke S, Purcell S. Preservation of genetic diversity in restocking of the sea cucumber *Holothuria scabra* investigated by allozyme electrophoresis. Can J Fish Aquat Sci. NRC Research Press; 2004; 61:519–28.

95. Uthicke S, Benzie J a H. Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. Mol Ecol. 2003; 12:2635–48.

96. Andrews KS, Nichols KM, Elz A, Tolimieri N, Harvey CJ, Pacunski R, et al. Cooperative research sheds light on population structure and listing status of threatened and endangered rockfish species. Conserv Genet. 2018; 19:865–78.

97. Iwamoto E, Ford MJ, Gustafson RG. Genetic Population Structure of Pacific Hake, *Merluccius productus*, in the Pacific Northwest. Environ Biol Fishes. 2004; 69:187–99.

98. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci. The Royal Society; 2004; 359:183–95. https://doi.org/10.1098/rstb.2003.1388 PMID: 15101575

99. Launey S, Hedgecock D. High genetic load in the Pacific oyster *Crassostrea gigas*. Genetics. Oxford University Press; 2001; 159:255–65.

100. Plough LV. Genetic load in marine animals: A review. Curr Zool. 2016; 62:567–79. https://doi.org/10.1093/cz/zow096 PMID: 29491946

101. Plough LV. Environmental stress increases selection against and dominance of deleterious mutations in inbred families of the Pacific oyster *Crassostrea gigas*. Mol Ecol. 2012; 21:3974–87.

102. Hedgecock D, Pudovkin AI. Sweepstakes Reproductive Success in Highly Fecund Marine Fish and Shellfish: A Review and Commentary [Internet]. 2011 [cited 2018 May 7]. Available from: http://www.ingentaconnect.com/content/umrsmas/bullmar/2011/00000087/00000004/art00013

103. Hogan JD, Thiessen RJ, Heath DD. Variability in connectivity indicated by chaotic genetic patchiness within and among populations of a marine fish. Mar Ecol Prog Ser. 2010; 417:263–75.

104. Rey C, Darnaude A, Ferraton F, Guinand B, Bonhomme F, Bierne N, et al. Within-generation polygenic selection shapes fitness-related traits across environments in juvenile sea bream. Genes. Multidisciplinary Digital Publishing Institute; 2020; 11:398. https://doi.org/10.3390/genes11040398 PMID: 32272597

105. Bruckner A. The recent status of sea cucumber fisheries in the continental United States of America. SPC Bêche—Mer Inf Bull. 2005; 22.

106. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, Meester LD. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol. 2013; 22:5983–99. https://doi.org/10.1111/mec.12561 PMID: 24128305

107. Nadeau S, Meirmans PG, Aitken SN, Ritland K, Isabel N. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecol Evol. 2016; 6:8649–64. https://doi.org/10.1002/ece3.2550 PMID: 28035257

108. Arndt A, Smith MJ. Genetic diversity and population structure in two species of sea cucumber: differing patterns according to mode of development. Mol Ecol. 1998; 7:1053–64.

109. Yang H, Hamel J-F, Mercier A. The sea cucumber *Apostichopus japonicus*: history, biology and aquaculture. Academic Press; 2015.

110. Bradbury IR, Laurel B, Snelgrove PV, Bentzen P, Campana SE. Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. Proc R Soc B Biol Sci. The Royal Society London; 2008; 275:1803–9. https://doi.org/10.1098/rspb.2008.0216 PMID: 18445556

111. Kinne O. The effects of temperature and salinity on marine and brackish water animals: 2. Salinity and temperature-salinity combinations. Oceanogr Mar Biol Annu Rev. 1964; 2:281–339.

112. Dong Y, Dong S, Meng X. Effects of thermal and osmotic stress on growth, osmoregulation and Hsp70 in sea cucumber (*Apostichopus japonicus Selenka*). Aquaculture. 2008; 276:179–86.

113. Fitzpatrick SW, Bradburd GS, Kremer CT, Salerno PE, Angeloni LM, Funk WC. Genomic and Fitness Consequences of Genetic Rescue in Wild Populations. Curr Biol. 2020; 30:517–522.e5. https://doi.org/10.1016/j.cub.2019.11.062 PMID: 31902732

114. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am Nat. 2016; 188:379–97. https://doi.org/10.1086/688018 PMID: 27622873

115. Xuereb A D'Aloia CC, Andrello M, Bernatchez L, Fortin M-J. Incorporating putatively neutral and adaptive genomic data into marine conservation planning. Conserv Biol. 2021; 35:909–20. https://doi.org/10.1111/cobi.13609 PMID: 32785955

116. Waples RS, Ford MJ, Nichols K, Kardos M, Myers J, Thompson TQ, et al. Implications of Large-Effect Loci for Conservation: A Review and Case Study with Pacific Salmon. J Hered. Oxford University Press US; 2022; 113:121–44. https://doi.org/10.1093/jhered/esab069 PMID: 35575083

117. Zhang L, Feng Q, Sun L, Ding K, Huo D, Fang Y, et al. Differential gene expression in the intestine of sea cucumber (*Apostichopus japonicus*) under low and high salinity conditions. Comp Biochem Physiol Part D Genomics Proteomics. 2018; 25:34–41.

**118.** Kirby MF, Morris S, Hurst M, Kirby SJ, Neall P, Tylor T, et al. The Use of Cholinesterase Activity in Flounder (*Platichthys flesus*) Muscle Tissue as a Biomarker of Neurotoxic Contamination in UK Estuaries. Mar Pollut Bull. 2000; 40:780–91.

**119.** Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, et al. Understanding and estimating effective population size for practical application in marine species management. Conserv Biol. 2011; 25:438–49. https://doi.org/10.1111/j.1523-1739.2010.01637.x PMID: 21284731

**120.** Waples RS, Gaggiotti O. Invited Reiew: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol. Wiley Online Library; 2006; 15:1419–39.

**121.** Spies I, Spencer PD, Punt AE. Where do we draw the line? A simulation approach for evaluating management of marine fish stocks with isolation-by-distance stock structure. Can J Fish Aquat Sci. NRC Research Press; 2015; 72:968–82.