Can biophysical models of pelagic larval dispersal explain the observed population structure; case studies from the Gulf of Alaska

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Bioinformatics and Computational Biology
in the
College of Graduate Studies
University of Idaho
by
Jacek Maselko

Approved by:
Major Professor: Erkan Buzbas, Ph.D. and Paul Hohenlohe, Ph.D.
Committee Members: Terry Soule, Ph.D., Ron Heintz, Ph.D.
Department Administrator: Paul Hohenlohe, Ph.D.

May 2022

Abstract

Numerous marine fish species have a characteristic pelagic larval dispersal stage. Understanding how this life history strategy affects the observed population structure of the adult groups and the adaptive potential of the species as a whole is therefore of paramount importance. In this study, I initially apply RAD-seq genomic analysis to examine the young of the year aggregates of Pacific ocean perch (Sebastes alutus) collected in 2014 and 2015 in the eastern Gulf of Alaska. I discover that these samples, even from the same haul, contain distinct genetic population mixtures indicating pelagic life stage sympatry. I also discover differences in selection strength between the two years, indicating that the maintenance of a portfolio of adaptive alleles may provide resilience of populations to natural environmental variability, where each adult cohort's genetic composition is influenced by the environmental conditions experienced during their first year at sea.

The apparent disconnect between pelagic stage sympatry and adult stage allopatry motivated the development of a stochastic spatio-temporal genetic model to understand the effect of biophysical dispersal on the population structure. Here, I develop the spatio-temporal genetic model utilizing a dispersal matrix and an allele frequency matrix which is then tracked over a number of generations. I then validate the genetic model with a suite of eight synthetic dispersal matrices and examine the inference via isolation by distance regression, STRUCTURE admixture analysis, and principal component analysis. This lead to unique insight into how each of these commonly used inference methods differs in their ability to differentiate among the synthetic candidate models. I then propose a log likelihood model selection framework based on the beta distribution as a viable alternate to determine which of the candidate dispersal models best explain the observed population structure based on pairwise F_{ST} values.

Finally, I demonstrate the application of the newly developed spatio-temporal genetic model to calculate the expected population structure for three fish species in the Gulf of Alaska, namely, Pacific ocean perch (*Sebastes alutus*), arrowtooth flounder (*Atheresthes stomias*), and Pacific cod (*Gadus macrocephalus*). I use a biophysical dispersal matrix based on the Regional Ocean Modeling System (ROMS) and species specific ontogenic life stage behavior combined in a previously developed mode, the Dispersal Model for Early Life

History Stages (DisMELS) to calculate the expected genetic differentiation. I then describe the expected population structure for these three species and apply the PCA, STRUCTURE admixture, and IBD regression inference. This is followed by the comparison of Pacific ocean perch and Pacific cod biophysical model based expected population structure and the observed genetic datasets which reconciles previously contradictory studies. I also demonstrate the application of this spatio-temporal genetic model to determine optimal sampling strategy in the log likelihood model selection framework. The results presented here also suggest that the biophysical based dispersal may be the primary driver behind the observed population structure in the marine species with life history strategies characterized by pelagic larval dispersal.

Acknowledgments

Dr. Erkan Buzbas, who humored my many false starts and rabbit holes. Erkan's challenging and thought provoking discussions were instrumental in the successful completion of this project.

Dr. Ron Heintz, who challenged me to take on this project, with a simple question: "What are all those young of the year Pacific ocean perch doing so far offshore", and supported me throughout this pursuit. To whom I want to apologize for not finishing this sooner, prior to his retirement from NOAA.

Dr. Paul Hohenlohe, for his suggestions and ideas and guidance in RAD-seq analysis, as well as his stalwart support and encouragement throughout.

Dr. Terry Soule, for his guidance on evolutionary algorithms giving me a unique insights into the evolution of fish behavior.

Dr. Kim Andrews for her guidance during the RAD-seq sequencing and analysis without which, my skillset would notably still be lacking any pipetting or unix batch programming.

Dedication

I would like to dedicate this work to my daughters, Izabela and Zosia Maselko, Their unwavering support and at times leading family intervention to make sure I did not quit were paramount to my finishing. To my daughters, dad quitting, was just not an option they were willing to tolerate.

Table of Contents

Abstract ii

Acknowledgments iii

Dedication iv

List of Tables viii

List of Figures ix

Statement of Contribution xiv

Chapter 1: Long-lived Marine Species may be Resilient to Environmental Variability through a Temporal Portfolio Effect 1

- Abstract 1
- Introduction 1
- Materials and Methods 5
  - Sample Collection and Processing 5
  - Molecular Analysis 5
  - Sequencing and Data Processing 6
- Results 8
  - Bioinformatics and Population Grouping 8
  - Genotype-Environment Association 9
  - Gene Ontology Enrichment 10
- Discussion 10
  - Sympatry and Population Structure 10
  - Genome Environment Association 12
  - Gene Ontology Enrichment 14
  - Fluctuating Selection and Maintenance of Adaptive Diversity 15
  - Conclusions 16
- Tables and Figures 18
- Supplemental Materials 26

References 31
Chapter 2: Beyond Isolation by Distance; Population Structure and Dispersal 38
Abstract 38
Introduction 38
Methods 40
Theoretical Population Demographic Framework 40
G_{ST} Population Differentiation 43
Population Structure and Larval Dispersal Models 44
Parameter Input Values for Larval Dispersal Models 45
Observed $\boldsymbol{G''_{ST}}$ and Null Dispersal Models 47
Dispersal Model Validation 47
Results 48
Time to Stationarity 48
Population Structure Through Regression 48
Population Structure Through PCA 49
Population Structure Through STRUCTURE Analysis 50
Cohort Specific Selection 51
Beta Distribution Parameters Describe Population Structure 51
Discussion 52
Algorithm 1 57
Tables and Figures 58
References 66
Chapter 3: Population Structure Through Biophysical Larval Dispersal Models 75
Abstract 75
Introduction 75
Methods 78
Results 81

Expected Population Structure..... 81
Model Validation..... 82
Effect of Sampling Design..... 83
Discussion 84
Tables and Figures..... 90
References 106
Appendix 1..... 114

List of Tables

Table 1-1 Pairwise F_{ST} values (F_{ST} below diagonal and p-value above the diagonal) between sNMF-derived populations and sampling year. There is little genetic differentiation as indicated by F_{ST} within a population between years (bolded F_{ST} and corresponding p-values)..... 24

Table 1-2 Number of private alleles in the given year specific to each sNMF-derived population. Private alleles is the number of alleles that were detected in only one year. For example, 127 is the number of alleles detected in population A in 2014, but not in 2015..... 24

Table 1-3 Number of private alleles in common between the sNMF-derived populations that were detected in 2014 samples but not in 2015. For example, 30 is the number of alleles detected in 2014 in both populations A and B, that were not detected in 2015 in either A or B population ... 24

Table 1-4 Results of LFMM analysis and the number of putative loci under divergent selection in 2014 and 2015. Note that sampling date and latitude (both are confounded as sampling was generally in the northward direction) were consistently associated with selection pressure in both years. However, other environmental factors only showed signatures of selection in 2015. 25

Table 2-1 Estimated time to stationarity under the various dispersal models. 58

Table 2-2 Estimated mean, variance and the calculated beta distribution parameters α, β of the pairwise $G"_{ST}$ estimates under various simulated population structure models for the single age demographic. The values in parentheses contain the standard deviation of the simulated values. 58

Table 2-3. Estimated mean, variance and the calculated beta distribution parameters α, β of the pairwise $G"_{ST}$ estimates under various simulated population structure models for the age structure (1:10) demographic. The values in parentheses contain the standard deviation of the simulated values. 59

Table 3-1 Model parameters for the three species. Life history parameters were minimum and maximum spawning ages. The Years to Stationarity was calculated by running the simulation until the slope of the mean and variance of all pairwise population F_{ST} values did not change. For Pacific ocean perch simulation, the stationarity is not achieved even after 30,000 years, but I did not carry the simulation further. 90

Table 3-2 Results of sample design evaluation for a subset of 5 zones to determine which subset best explains the overall population structure. 90

List of Figures

Figure 1-1 Locations of the 2014 (yellow) and 2015 (orange) collection of the young-of-the-year Pacific ocean perch. 18

Figure 1-2 sNMF ancestry analysis revealed 4 ancestral populations represented by fish collected in both 2014 and 2015: (a) cross-entropy plot for the number of populations in sample; (b) sNMF population ancestry barplot; and (c) PCA analysis with the colors corresponding to the sNMF-derived majority ancestry populations in panel b. Note that both years were included concurrently in the analysis. For clarity year designation is omitted as each population cluster contains both years interspersed throughout. 20

Figure 1-3 Spatial distribution and the proportional representation of the putative populations in the samples in 2014 (a) and 2015 (b) collections. 21

Figure 1-4 Relationships of body length and weight in the 2014 (black and 2015 (red) young-of-the-year Pacific ocean perch. In 2015, smaller length fish had significantly ($p<0.0001$) smaller mass than in 2014, indicating environmental factors in 2015 may have negatively influenced their condition. 22

Figure 1-5 The selective sieve. The five colored plates represent various hypothetical environmental forces (such as temperature, chlorophyll density, etc), that are highly variable among years. This represent the different selection pressures encountered by the Pacific Ocean Perch during initial pelagic life stage. Each years' cohort therefore contains the alleles that were selected for during their first year. Populations of long-lived adults representing multiple cohorts maintain genetic diversity as a result of this temporal variation in selection. 23

Figure 2-1 . Theoretical population dispersal models. The arrows indicate the direction of the dispersal. The top panel shows the theoretical specification of the dispersal models with the $\delta_{i\rightarrow j}$ parameters indicating the magnitude of the pairwise dispersal between the *ith* and *jth* population. The bottom panels are the directed graph diagram of the dispersal connectivity matrices and the heat map showing the magnitude of dispersal connectivity between the source and sink populations. 60

Figure 2-2 Time to stationarity plots for the mean and variance of pairwise G''_{ST} of the dispersal models. Stationarity was determined by the outcome of a slope test with no significant slope detected. a) left panel are the results with no age structure; b) right panel are populations with 1 to 10 aged cohorts interbreeding. The results of the slope test are in table 2-1. 61

Figure 2-3 IBD regression plots of the linearized $G''_{ST}/(1-G''_{ST})$ on unit distance. The mean dispersal (m), R^2 and the slope p-value for each model are displayed below each corresponding graph. In

panel a) the models consist of a single cohort, while in panel b) the populations are age structured with cohorts 1:10 interbreeding. 61
Figure 2-4 PCA results sampling 100 individuals from each population with the discrete populations colored separately. Panel a) indicates the results of a single cohort age structure; b) indicates ages 1:10 with 10 individuals sampled from each cohort equating to 100 total per population. The model parameters were: N_e=1,000, loci=1,000, N_em=5, n=100. 62
Figure 2-5 STRUCTURE results with best number of putative populations selected by STRUCTURE indicated by separate colors for each of the dispersal models with no age structure and 100 samples per population. a) The PCA plot with individual samples colored corresponding to the STRUCTURE population assignments; b) The STRUCTURE based admixture results. The model parameters were: N_e=1,000, loci=1,000, N_em=5, ages=1:1, samples=100. 62
Figure 2-6 PCA and STRUCTURE results for dispersal models with age structures (cohorts 1:10) with sampling all age classes (n=10) for 100 samples per populations. 63
Figure 2-7 PCA and STRUCTURE results for dispersal models with age structure (cohorts 1:10), but sampling a single age (1) with 100 samples per population. 63
Figure 2-8 The confounding effect of strong cohort specific selection (s=0.5), demonstrate the lack of population discrimination power. Panel a) shows the PCA plot with colors denoting the populations, panel b) shows the PCA plot with the colors denoting the ten cohorts. Panel c) indicates the PCA plots with STRUCTURE derived putative populations denoted in panel d. 64
Figure 2-9 STRUCTURE admixture results under strong selection (s=0.5) and sampling a single cohort. Note that only *total isolation* dispersal model results in a differentiation of the 10 populations, with all other models only identifying a single population. 64
Figure 2-10 The Beta distribution parameter space of the pairwise G''_{ST} under various dispersion models using 100 simulations and 1,000 loci. Panel a) is the parameter space of the non-age structured population, while panel b) is the parameter space for the age structured (1:10) populations. Insets show the mean G''_{ST} and standard deviation for each of the dispersal model and the respective age structure. Note that both x and y axis are on a log scale. 65
Figure 3-1 Map of the Gulf of Alaska depicting the modeled area. a) Conceptual model of the larval dispersal where each of the $\delta_{i\rightarrow k}$ parameters describe the probability of larvae dispersing from ith population and settling and recruiting to the kth population. The $\delta_{i\rightarrow k}$ parameters are obtained from the individual based DisMELS (*GitHub - WStockhausen/DisMELS: A Java-Based Framework for Developing/Running Individual-Based Models(IBMs) for Marine Species with Pelagic Early Life Stages.*, n.d.) model runs that incorporate ocean circulation and life history parameters specific to the species. b) Map depicting the 13 modeled spatial zones in the Gulf of

Alaska of larval release and settlement used in the DisMELS model to calculate the $\delta_{i\rightarrow k}$ parameters. c) shows an example of DisMELS model graphical output where the left panel are the pathways during early larval dispersal stages and right hand panel are the subsequent stages. Yellow indicating pathways of "successful" larvae and magenta being the "unsuccessful" runs. 91

Figure 3-2 Graphical representation of the D dispersal probability matrices obtained from the DisMELS model for the three species, Pacific ocean perch (POP), Arrow tooth flounder (ATF), and Pacific cod. The top row heat maps depict the strength of dispersal with darker colors denoting higher probability. The bottom row are the corresponding directed graphs that show a different representation of the connectivity among the DisMELS population zones. 92

Figure 3-3 The expected IBD regression plots for the three species using the pairwise linearized $G''_{ST}/(1-G''_{ST})$ on the cumulative distance between zones (km). Based on the biophysical dispersal model, there is an expected significant IBD relationship for POP and Pacific cod, however, ATF shows no significant genetic differentiation with distance. 93

Figure 3-4 PCA plot of the genetic simulation results and sampling 100 individuals from each population-zone across all age classes. The PCA plot colors correspond to the population-zone location colors on the map to the right. 94

Figure 3-5 STRUCTURE derived putative populations for Pacific ocean perch (POP) based on sampled 100 individuals per DisMELS population zone. The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. PCA representation shows 3 distinct clusters with a possible fourth one. The STRUCTURE plot show a single (green), distinct population group (in the farthest southeastern zone), and immediately followed by admixture from another population (orange) in the adjacent zone to the northwest. The third population group (blue) appears off of cross sound. All three putative groups are then admixed in relatively equal proportions resulting in a concurrent detection of distinct groups from Yakutat to east of Kodiak island, at which point only two admixed populations are evident to the west. The main break in genotype distribution appears around Kenai Peninsula. 95

Figure 3-6 STRUCTURE derived putative populations for Arrow tooth flounder (ATF) based on sampled 100 individuals per DisMELS population zone. The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. The PCA plot shows 4 distinct population clusters with obscured groups in the middle of the plot, possibly indicating a non-linear variation in the genotype groups. The

STRUCTURE admixture analysis indicates 6 distinct groups with increasing admixture from the eastern to western Gulf of Alaska. The zones in the eastern part of Gulf of Alaska are composed of distinct, putative populations, with little admixture among them. However from outside of Prince William Sound, to the west, there is only a single, dominant population identified with little admixture from the eastern populations. The map depicts where the corresponding, colored putative populations are spatially distributed. 96

Figure 3-7 STRUCTURE derived putative populations for Pacific cod based on sampled 100 individuals per DisMELS population zone. The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. PCA plot suggest 3 genotype clusters, but they are not concordant with the STRUCTURE derived clusters. STRUCTURE analysis shows a single population throughout the eastern Gulf of Alaska which then progressive admixture originating from zone 13 (Cook Inlet). The farthest west zones indicate the presence of some admixture from other putative populations. The map depicts where the corresponding, colored putative populations are spatially distributed. 97

Figure 3-8 Comparing the observed and modeled population structure for Pacific ocean perch. The observed data set was adapted from (Palof et al., 2011) where the sampled locations were matched to the DisMELS population zones (1, 2, 4, 6, 9, 10, 12). The matching subset of the linearized pairwise observed F_{ST} values for both the observed and modeled populations was then regressed on the coast distance. Both observed and modeled populations had positive and significant scale of genetic differentiation with distance ($R^2 = 0.332, p = 0.001$ and $R^2 = 0.157, p = 0.037$ respectively). The observed and simulated pairwise F_{ST} values had a correlation of 0.184 (Mantel test $p = 0.129$). The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. Both STRUCTURE and PCA plots indicate the presence of three populations clusters with clusters A and B (orange and blue) dominating eastern Gulf of Alaska, while clusters B and C (blue and green) dominating western part with the dominant break occurring outside of Prince William Sound. 98

Figure 3-9 Comparing the observed and modeled population structure for Pacific cod. The observed data set of linearized G''_{ST} was adapted from (Drinan et al., 2018) where the sampled locations were matched to the DisMELS population zones (1, 6, 8, 11, 12). The matching subset of the linearized pairwise observed F_{ST} values for both the observed and modeled populations was then regressed on the coast distance. Both observed and modeled populations had positive and

significant scale of genetic differentiation with distance ($R^2 = 0.713, p = 0.002$ and $R^2 = 0.673, p = 0.004$ respectively). There was significant positive (Mantel test $p = 0.03$) correlation ($r = 0.72$) between the observed and simulated pairwise G''_{ST} values. The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. The STRUCTURE plot show a single distinct population group in the southeastern and central part of the State, a second group in the Kodiak area. Note that the PCA plot appears to show 3 clusters, STRUCTURE analysis was only able to identify 2 distinct groups with admixture among them, which did not appear to match the PCA clusters. 100

Figure 3-10 Results of the effect of sampling location to determine "best" and "worst" subset of five locations to sample under the POP model. The full model included all 12 population zones with the best subset (top row) had the highest log-likelihood (228.9) when populations 1, 4, 5, 6, 7 are sampled and lowest log likelihood (-2,921.5) when populations 4, 8, 9, 10, 12 are sampled. Note that the worst fit model results in a highly significant apparent IBD relationship, but only a single STRUCTURE derived putative population. 102

Figure 3-11 Results of effect of sampling locations to determine "best" and "worst" subset of five locations to sample under the ATF model. The full model included all 12 population zones with the best subset (top row) had the highest log-likelihood (104.75) when populations 1, 9, 10, 11, 12 are sampled and lowest log likelihood (-27,796.64) when populations 8, 9, 10, 11, 12 are sampled. The best subset inference would result in identifying two distinct populations groups with little admixture, however, a panmictic population results from the worst sampled zones. 103

Figure 3-12 Results of the effect of sampling location to determine "best" and "worst" subset of five locations to sample under the Pacific cod model. The full model included all 13 population zones with the best subset (top row) had the highest log-likelihood (223.4) when populations 7, 8, 9, 10, 13 are sampled and lowest log likelihood (-1,331.1) when populations 1, 2, 3, 5, 6 are sampled. The best model would result in a non-significant IBD relationship, but strong PCA and STRUCTURE clustering. In contrast, the worst model results in a significant IBD relationship, but a single cluster detected with PCA, and some STRUCTURE detected admixture from another population which is not evident in the PCA plot. 104

Figure 3-13 Distribution of putative populations based on STRUCTURE analysis of adult POP collections from 2017 and 2019 (Timm, L). Populations A-D correspond to the populations identified in larval aggregates described in chapter 1 of this dissertation. 105

Statement of Contribution

Chapter 1 of this dissertation was published in Ecology and Evolution:

Maselko, J., Andrews, K. R., & Hohenlohe, P. A. (2020). Long-lived marine species may be resilient to environmental variability through a temporal portfolio effect. Ecology and evolution, 10(13), 6435-6448.

Jacek Maselko: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (equal); Resources (equal); Software (equal); Writing-original draft (lead); Writing-review & editing (equal).

Kimberly R. Andrews: Conceptualization (supporting); Data curation (supporting); Formal analysis (supporting); Methodology (supporting); Project administration (equal); Writing-review & editing (equal).

Paul A. Hohenlohe: Conceptualization (supporting); Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (equal); Project administration (lead); Writing-review & editing (equal).

Chapters 2 and 3 were solely the work of Jacek Maselko

Chapter 1: Long-lived Marine Species may be Resilient to Environmental Variability through a Temporal Portfolio Effect

"Long-lived marine species may be resilient to environmental variability through a temporal portfolio effect." Ecol Evol. 2020; 10: 6435– 6448. https://doi.org/10.1002/ece3.6378

Abstract

Maintenance of a portfolio of adaptive alleles may provide resilience of populations to natural environmental variability. We used Pacific ocean perch (POP; *Sebastes alutus*) to test for the maintenance of adaptive variation across overlapping generations. POP are a long-lived species characterized by widespread larval dispersal in their first year and a longevity of over 100 years. In order to understand how early marine dispersal affects POP survival and population structure, we used Restriction Site Associated DNA sequencing (RADseq) to obtain 11,146 single-nucleotide polymorphisms (SNPs) from 401 young-of-the-year (YOY) POP collected during surveys conducted in 2014 (19 stations) and 2015 (4 stations) in the eastern Gulf of Alaska. Population clustering analysis showed that the POP samples represented four distinct ancestral populations mixed throughout the sampling area. Based on prior work on larval dispersal of POP, these larvae are most likely from distinct parturition locations that are mixing during their pelagic dispersal life stage. Latent factor mixed models revealed that POP larvae face significant selection during their first year at sea, which were specific to the year of their birth. Thus each adult cohort's genetic composition is heavily influenced by the environmental conditions experienced during their first year at sea. Long-lived species relying on broadcast spawning strategies may therefore be uniquely resilient to environmental variability by maintaining a portfolio of cohort-specific adaptive genotypes, and age truncation due to overfishing of older cohorts may have detrimental effect on the population viability.

Introduction

Understanding the resilience of biological marine resources to changing oceanographic conditions is central to ecosystem-based fisheries management and the implementation of adaptive sustainable harvest strategies (Link 2002; Levin and Möllmann 2015). The ability of populations to respond to disturbances in their habitat is in part determined by the genetic diversity present in the population (Parker et al. 2000; Hoffman

and Sgro 2011). This genetic diversity, exhibited by a portfolio of available gene variants, allows for a quick response if selectively advantageous variants are already present in the population (Sunday et al. 2011; Pacifici et al. 2017). Understanding the response of marine populations to environmental perturbations will allow us to readily assess the resilience or vulnerability of these populations and species.

Measuring differential survival between subpopulations reveals how environmental conditions can influence the overall productivity of exploited populations. For example, Schindler et al. (2010) demonstrated that environmental conditions favored the production of discrete salmon populations residing in Bristol Bay, Alaska. Maintenance of a portfolio of locally adapted genotypes ensured adult returns to the region, but returns to different streams were maximized under different environmental conditions. This maintenance of a portfolio of adaptive alleles may be a key aspect of resilience of populations to natural environmental variability.

The idea of a temporal portfolio effect, in which adaptive variation is maintained by overlapping generations in a temporally variable environment, has been studied in general (Ellner and Hairston 1994). For example, many freshwater zooplankton taxa have relatively short-lived adults that may be subject to strong selection, but eggs can remain viable for decades in sediment, resulting in persistent egg banks that are relatively buffered from environmental variation (Brendonck and DeMeester 2003). We hypothesize that a similar effect may occur in marine fish species with highly dispersive larvae and long-lived, relatively sedentary adults. The genetic composition of each recruitment cohort may reflect relatively strong selection during the larval stage, while the adult population would maintain genetic variation reflecting multiple cohorts. Here we test this hypothesis using genomic methods for detecting population structure and adaptive loci.

Genomic data allow us to scan for individual and population-level differences across the whole genome, and genomics is becoming integral in answering a wide array of previously unresolved questions in conservation biology with numerous applications in fisheries (Wenne et al., 2007; Barrio et al. 2016; Jasonowicz et al., 2016; Valenzuela-Quiñonez, 2016; Kumar and Kocour, 2017). It is now possible to estimate, with a high level of precision and certainty, the demographic structure of fish populations at small spatio-temporal scales, and to identify local adaptation from genomic data (Wang and Höök, 2009;

Barrio et al., 2016; Catchen et al. 2017; McKinney et al. 2017). RADseq approaches have been extensively used to describe various biological and ecological phenomena, such as phylogeography, population differentiation and structure, population and individual admixture (composition of lineages), genetic diversity, and outlier loci detection, among others (Alexander et al. 2009; Andrews et al. 2016; Narum et al. 2013).

Our model species, the Pacific ocean perch (POP; *Sebastes alutus*), is a long-lived species with its oldest individuals being over 100 years old (Conrath and Knoth 2013). POP are the most abundant and economically important rockfish species in the Gulf of Alaska (Conrath and Knoth 2013) with landings in excess of 55,000 tons in 2017 (NOAA 2019). The fishery is managed using an age-structured model where the vital population rates are derived from the abundances of different ages in the catch (Megrey 1988; Hulson et al. 2017), but the relationship between the abundance of spawning fish and their offspring cohort is highly variable and unpredictable. This extreme annual fluctuation in success and failures of various year classes has been noted as a characteristic of this and many other commercially exploited species (Westrheim 1958; Carlson and Haight 1976).

POP in the Gulf of Alaska live on the upper slope of the continental shelf. They spawn from September through November, with parturition occurring in April through May the following year, when larvae rise from demersal spawning habitats on the continental shelf break (150 – 400m depth) to surface waters. They then become part of the ichthyoplankton and within a few weeks metamorphose to a young-of-the-year form (YOY). They are carried in the surface waters by currents and settle out of the water column in nearshore rocky habitat by the end of their first year (Carlson and Haight 1976; Major and Shippen 1970). During their shoreward movement, larvae grow rapidly and allocate significant amounts of energy to creating lipid tissue. This lipid tissue is apparently lost during settlement (Moss et al. 2016), suggesting energy acquisition and growth are important determinants of settlement success (Hoey and McCormick 2004). They remain in the nearshore habitat for the next few years until they join the discrete adult schools residing on the continental shelf and slope (Love et al. 2002). They reach sexual maturity at eight to ten years of age and repeatedly spawn until their hundredth year or longer (Hulson et al. 2017). These adult schools are genetically differentiated and the degree of their differentiation (F_{ST}) is correlated to the geographic distance between them (Palof et al. 2011).

Larval dispersal pathways in POP may be highly variable from year-to-year since they mostly depend on ocean currents in a given year (Mundy et al. 2010). An important prediction of ocean current and dispersal models (Stockhausen 2009; Stockhausen and Hermann 2007) is that in each year, the larvae at a given pelagic location are comprised of mixtures of individuals from different spawning locations indicating a high degree of mixing among them. However, population genetic studies of young-of-the-year and adults indicate there is limited mixing among subpopulations (Palof et al., 2011; Kamin et al., 2014). The Kamin et al. (2014) follow up study examined the YOY POP catches corresponding to locations near the adults caught by Palof et al. (2011). Their work showed that the collections of YOY POP were most related to the linearly closest adult populations. Either widely dispersed juveniles are able to return to their natal areas, or survival is maximized among locally retained larva, possibly due to local adaptation.

Here, we test whether POP larvae exhibit signatures of selection that could allow for the maintenance of a portfolio of adaptive variation in the multi-cohort adult population. We examined YOY POP collected from the eastern Gulf of Alaska during two years (2014 and 2015), when the oceanographic conditions were drastically different, with 2014 being an average temperature year, and 2015 being anomalously warm which is expected to have a negative impact on the fish (Cavole et al. 2016; Gentemann et al. 2017; Jones et al. 2018). We evaluated the potential for differences in selection strength for YOY POP across years by testing whether the fish differed in physiological conditions in 2014 and 2015, measured as a body condition index based on weight-length relationships, and total lipid content. We then used genotype environment association (GEA) tests with RADseq genomic data to test for differences in selection acting on the genome to favor different phenotypes between the two years. Finally, we identified candidate biological pathways on which selection was acting in the two different YOY cohorts. We predicted that the strength of selection would be higher in 2015 than 2014 due to the unusually high 2015 sea temperatures, resulting in poor body condition and a greater number of SNPs associated with environmental variables and physiological condition in the 2015 dataset. These findings may explain the difference in recruitment for the 2014 and 2015 cohorts as estimated in the 2017 stock assessment (Hulson et al., 2017).

Materials and Methods

Sample Collection and Processing

Young-of-the-year (YOY) POP were collected during NOAA oceanographic surveys in the summer of 2014 and 2015 (Figure 1-1). POP larvae and YOY were distinguished from congenerics (Kendall et al., 2007) using diagnostic SNPs (Garvin et al., 2011) prior to inclusion in this study. The resulting sample size of identified YOY POP was 399 fish in 2014 and 108 in 2015.

Fish length, weight, and lipid content for each identified POP fish were measured at Auke Bay Laboratories in Juneau, AK. Lipid content was extracted using the Folch method (Folch et al. 1957) and quantified using the colorimetric sulpho-phospho-vanillin (SPV) method (Chabrol and Charonnet, 1937). Condition index was calculated as the residual value from a log(weight)~log(length) linear regression. This index accounts for the different lengths due to age of the YOY fish where a positive residual indicates better body condition than expected (Froese, 2006). Finally, for DNA analysis, a small tissue plug was extracted through an incision in the abdominal wall that included the heart tissue and stored in 95% ethanol.

Molecular Analysis

DNA was extracted from the muscle plug from the 515 individual fish into 96-well plates with the QIAGEN DNeasy Blood and Tissue Kits as described by the manufacturer (QIAGEN, Inc.). In brief, small pieces of tissue (~20 mg) were excised from each muscle plug. The tissue pieces were digested in a proteinase solution for at least 3 hours at 55ºC. Protease digestions were performed in 96 well plates. After digestion, the samples were purified with either QIAxtractor or Corbett X-tractor robot producing eluted DNA which was stored at -20 ºC.

RADseq library preparation was done for all 515 samples plus eight samples that were replicates according to Ali et al. (2016) and refined by Andrews et al. (2018) using the Sbf1 restriction enzyme, which cuts at an eight-base recognition site. Custom eight-base biotinylated barcodes were ligated to the cut site allowing multiplexing of groups of 96 samples. The multiplexed samples were then sheared to 400 bp using Covaris M220 sonicator. This was followed by a Streptavidin bead assay to exclude sheared fragments that did not include the biotinylated barcodes. Illumina's NEBNext ultra DNA library prep kit

was then used to add Illumina adapters with indexes unique to each of the multiplexed groups of 96 samples to allow further pooling and Illumina sequencing compatibility. 150 bp paired end sequencing was done on two lanes at the Berkeley Genomics Center Laboratory (https://qb3.berkeley.edu/gsl/) using Illumina HiSeq 4000.

Sequencing and Data Processing

We followed the bioinformatic pipeline described in Andrews at al. (2018; Figure 1-1), with slight modification for STACKS version 2.0 (Catchen et al. 2013). Briefly, a custom PERL script was used to flip the raw reads so that each 140 bp read was aligned starting with the barcode, and the Sbf1 cutsite sequence. STACKS 2.0 (Catchen et al. 2013) program *process_radtags* was used to demultiplex the raw reads followed by program *clone_filter* to remove PCR duplicates. BOWTIE2 version 2.3.4.3 (Langmead and Salzberg; 2012) was used to align the sequences to *Sebastes nigrocinctus* reference genomes downloaded from the ncbi database (https://www.ncbi.nlm.nih.gov/genome/14568). The *S. nigrocinctus* aligned reads were then processed using the refmap.pl pipeline in STACKS 2.0. Filtering of the final set of SNPs was done using POPULATIONS module in STACKS 2.0 with the minimum percent of individuals genotyped at a locus in a population set at 10% and the minimum global minor allele frequency of SNPs set at 0.1. Subsequent analysis was conducted using R statistical software (R Core Team 2016) using data in *genepop* format exported from POPULATIONS module.

CLUSTER analysis was conducted using package *adegenet* (Jombart et Devillard; 2010) and *poppr* (Kamvar et al., 2014, 2015) using all samples, including the seven remaining replicate pairs (one replicate was not recovered during sequencing) to select the optimal set of filters for removing individuals and loci based on the level of missing data. These filter settings were varied until the CLUSTER plot showed the paired replicates to be most closely related. This resulted in removal of loci which were absent in at least 15% of individuals and genotypes having more than 20% of total identified loci missing. For subsequent analyses, only one from each pair of replicate samples with the most loci was retained. We used the R package s*equoia* (Huisman, 2017) to identify related individuals, up to half – siblings; this program is specifically designed to use large SNP datasets and does not require a parent to be present in the sample. This was done for each of the two cohorts in

order to make sure no related individuals were included in the Genome-Environment Association (GEA) tests.

We estimated the number of ancestral populations represented in the sample using the *LEA* R package (Frichot and François 2015). The analysis employed population clustering analysis with sparse non-negative matrix factorization optimization (sNMF) (Frichot et al. 2014) to estimate number of ancestral populations represented in the sample. The number of populations was determined from the cross entropy criteria and Cattell's rule (Cattell, 1966) from the sNMF output. We favored the sNMF routine because it is robust to departures from Hardy-Weinberg equilibrium as compared to Bayesian and Maximum Likelihood approaches (Frichot et al. 2014). We also compared the sNMF results to STRUCTURE 2.3.4 (Pritchard, Stephens & Donnelly 2000) derived population clustering.

We examined whether selection pressure is consistent from year-to-year by testing for a difference in the number of private alleles, or homozygous loci in each year. The number of private alleles that were only found in 2014 but not in 2015 was quantified specifically to each sNMF derived population and across all SNPs. To compute whether the number of private alleles was significantly different between years, we needed to account for the difference in sample sizes between the years. We wrote a custom bootstrap routine in R to create a null distribution of the expected number of lost alleles in the smaller sample size by selecting without replacement from the larger year's sample, the reduced sample size. The significance (p=0.05) was then based on where the observed number of private alleles lies in the null distribution.

Latent factors mixed model (LFMM) algorithm in R package *LEA* (Frichot, 2014) was conducted to identify loci influenced by selection. For subsequent analysis we imputed any remaining missing data (3.5% in 2014 and 4.1% in 2015). The missing genotypes were imputed using the random forest algorithm in the R packages *randomForestSRC* and *radiator* (Gosselin, 2018). We used the R package *hierfstat* (Goudet, 2005) to estimate pairwise F_{ST} according to Nei (1987). Significance of F_{ST} was calculated through 1,000 permutations of population indices. PCA analysis was conducted using the dudi.pca routine in *ade4* R package (Dray 2007). Environmental variables included in the GEA included sample date and latitude, sea water temperature, and chlorophyll concentration. Phenotypic metrics were also included in the GEA, including percent lipid content and condition index. This analysis

was done for 2014 and 2015 data separately with four latent factors to account for population structure while testing for genome-environment association. This was followed by nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) search of nucleotide sequences and their corresponding protein coding gene regions where selection may be occurring. Loci annotation and BLAST searches of the associated 140 bp sequences were accepted when below the nucleotide and protein e-value threshold of $1x10^{-10}$. BLAST e-value score is the probability that the similarity is due to chance.

The gene ontology (GO) enrichment analysis was used to determine whether the groups of genes associated with each of the environmental variables was enriched for certain biological processes. This analysis was done by querying geneontology.org database using zebra fish (*Danio rerio*) as a reference organism, and the alpha level was set at p=0.05 with no multiple test correction applied. Subsequently, the biocyc.org and informatics.jax.org were queried to determine general biological functions of the gene aggregates.

Results

Bioinformatics and Population Grouping

The total number of raw Illumina sequencing reads for the six plates was 2,983 million, or on average 497 million per plate. The proportion of reads with a correct barcode and restriction enzyme cut site varied from 69% to 83% per plate with an average of 76%. Alignment to the *S. nigrocinctus* reference genome resulted in 79% overall alignment rate, with the percentage of aligned reads per sample ranging from 56 to 77% (mean = 71%). Filtering of individuals with high percentages of missing genotypes ($\geq$ 15%) and SNPs with low genotyping rates ($\leq$ 20%) resulted in the final sample size of 398 individual fish (321 in 2014 and 77 in 2015) and 11,146 SNPs.

The ancestry analysis revealed the presence of 4 discrete spawning populations. sNMF ancestry analysis in LEA revealed 4 populations based on cross-entropy criteria (Figure 1-2a, 1-2b). PCA analysis supported the K=4 sNMF derived putative population clusters (Figure 1-2c). STRUCTURE (Pritchard et al. 2000) analysis also supported K=4 populations, but with greater admixture of population 2 and 3 than was estimated via sNMF algorithm. All four of these populations were represented in both the 2014 and 2015 collections (Figure 1-3). Pairwise F_{ST} values for genetic differentiation among putative population-year combinations revealed consistent differentiation between populations in each

year (Table 1-1). Additionally, this difference was conserved across years, meaning little differentiation as measured by F_{ST} was observed within a population, between years. These findings support the results of the ancestry analysis and provide evidence that the 2014 and 2015 collections are composed of similar mixtures of discrete spawning populations.

Relatedness analysis showed no related individuals (up to half siblings) in the collections. This indicates that the discrete sNMF derived populations are not simply groups of closely related individuals. Furthermore, the results of this analysis ensured that no related individuals are included in the subsequent genotype-environment association models, which is thought to result in higher false positive rates due to lack of independence among the samples (Newman et al. 2001; Voight and Pritchard 2005).

Fewer private alleles were detected in 2015 than in 2014 and this pattern was significant when adjusting for the smaller sample size in 2015 (Table 1-2). This analysis was done separately for each sNMF-derived population, and we detected private alleles in common among all four populations (Table 1-3) indicating the same suite of alleles was not detected in 2015. This suggests that 2015 selection was stronger as compared to 2014, leading to loss of deleterious alleles in the 2015 cohort, which is consistent with the more abnormal oceanic conditions observed in 2015 than in 2014 (Cavole et al. 2016; Gentemann et al. 2017; Jones et al. 2018).

Genotype-Environment Association

The results of LFMM analysis linking environmental and phenotypic variables to SNP variants indicated similar patterns of association with latitude and collection date in both years (Table 1-4). Of the 76 SNPs associated with these variables in 2014 and 305 in 2015, ten were shared between years (Supplementar materials, Table 1-1). The loci common in both years were significantly associated with latitude and collection date only. However, because of the sampling being conducted in a generally south to north direction, sampling date and latitude are collinear. This may possibly indicate a temporal gradient of selection where less fit individuals, those with deleterious alleles die off during their first months of life. Therefore the fish collected at later dates may be a subset of the fitter individuals as compared to earlier collections. Or there may be a true latitudinal gradient, or a combination of both factors contributing in various proportions to a selection gradient. Chlorophyll concentration and seawater temperature did not appear to influence loci in 2014, but were

associated with 100 loci in 2015. It is important to note the Gulf of Alaska experienced unusually warm temperatures in 2015 (Cavole et al. 2016), marked by large sea bird die-offs (Jones et al. 2018).

In 2015, the fish experienced poorer growing condition as compared to 2014 (Cavole et al. 2016). This was evident in their weight for a given length when examining the condition index graphs (Figure 1-4). Linear regression analysis indicated a significantly ($p<0.05$) lower intercept and steeper slope in 2015 suggesting that smaller fish had poorer condition in 2015, but larger fish appeared to be unaffected. Whether the smaller fish died off and only larger fish survived is uncertain, although there appears to be a genetic basis of selection where a number of loci were identified as being associated with fish body condition (% lipid and condition index). This was not observed for the fish collected in 2014.

Gene Ontology Enrichment

BLAST search resulted in only six loci being associated with known genes in 2014 and 24 in 2015 (Supplementary materials, Table 1-2). The broad-scale biological processes associated with the gene ontology (GO) enrichment are listed in (Supplementary materials, Table 1-3), while detailed information and fine-scale biological processes associated with the gene subsets may be found in the supplementary materials. The majority of genes were associated with developmental processes: 4 out 5 in 2014 and 108 out of 168 in 2015 (Supplementary materials, Table 1-3). Intracellular processes were associated with all environmental gradients in both years (see supplementary materials). Various developmental processes were mostly associated with collection date, chlorophyll-a concentration, latitude, water temperature, as well as tissue lipid percentage. Growth associated processes were mostly associated with chlorophyll-a concentration, collection date, collection latitude, water temperature and tissue lipid percent, but not condition index. Metabolic related processes were associated with condition index, latitude, and temperature. Fatty acid and lipid metabolism processes were only associated with a temperature gradient.

Discussion

Sympatry and Population Structure

One of the surprising findings of this study was the strong genetic clustering where the individuals of respective clusters were dispersed among the sampling locations, as well as conserved between the two years (Figure 1-3). This is consistent with the predictions from

the DisMELS model (Stockhausen 2009), but in contrast with the findings of Kamin et al. (2014), where the collections were mapped to the closest adult groups and no genetic clustering was detected. However, their study only used twelve microsatellite markers and therefore may have lacked statistical power to detect the finer-scale genetic clustering as the RADseq approach we employed here. This inference is supported by the low F_{ST} values detected here (F_{ST} ranging from 0.008 to 0.032 between clusters), because detection of low F_{ST} values can require markers with high power. Also, the Kamin et al. (2014) study treated each haul collection as a sampling unit and conducted tests on the allele frequencies among the hauls, transects, locations, and years, but did not examine genetic clustering based on individual admixture analysis. However, the presence of genetic structure in our study is consistent with Palof et al. (2011), who detected isolation-by-distance population structure in the adults. The complete mixing among the genetically distinct groups of YOYs would be expected to result in a lack of population structure within just a few generations if the mixed fish maintained their grouping through settlement, recruitment and spawning. Our observations are consistent with both the DisMELS (Stockhausen 2009) and Palof et al. (2011) results indicating dispersal is not the primary mechanism by which POP population structure is maintained.

Our study suggests that distinct POP populations that are sympatric during the larval and YOY stage are likely geographically segregated and genetically differentiated during spawning. The presence of genetic clusters in spite of larval stage sympatry may indicate that once the fish settle out in the nearshore rearing habitat, they may be able to home-in to their natal locations over the following few years. If homing to their natal locations begins after fish settle out of the water column into their nearshore rearing habitat, then the mixtures of genotypes would be evident among larvae as they advected towards shore by cross-shelf currents.

The homing behavior in adult *Sebastes* spp. has been well documented (i.e. Carlson and Haight 1972; Matthews 1990; Carlson at al. 1995). It is unknown, however, when this behavior begins. Schools of age 1+ fish are spatially segregated (Carlson and Haight 1976), although it is unknown if those individuals are from a single or multiple source populations. It may be that these single cohort schools are composed of individuals from multiple sourced populations and like salmon, leave the school when natal location is nearby.

Homing behavior would result in genetic isolation and population structure consistent with our observations. Westrheim (1975) noted that POP schools were separated by bathymetry and would not cross deep trenches once in demersal stage. Withler et al. (2001) also described POP populations that were genetically distinct, yet lived within close proximity of each other, even when sampled in different seasons. Therefore if larvae from discrete nearby parturition locations, separated by bathymetric features such as canyons and ridges, were jointly entrapped in the oceanic currents, these clusters would resemble our observations. If homing to their natal locations begins after fish settle out of the water column into their nearshore rearing habitat, then the mixtures of genotypes would be evident among larvae as they advected towards shore by cross-shelf currents.

Another explanation for the fate of these YOY fish is that they are entrained in the coastal current and mesoscale eddies and fail to find suitable rearing habitat prior to winter settlement and are therefore destined to die, and our sampled fish were already the "swimming dead". The selection that we observed would then be the sign of various phenotypes dying at different rates, while the unobserved fish, the ones that did not get advected away from natal grounds and mixed with other similar-fated YOYs, are the only ones that successfully reach suitable nearby rearing habitat. Westrheim (1958), and Carlson and Haight (1976) noted the extreme successes and failures among POP year classes, which perhaps may be indicative of advection rate away from the natal grounds or high larval mortality, assuming consistent spawning population.

Genome Environment Association

Fish employing broadcast spawning strategies characterized by larval and juvenile pelagic drift in ocean currents are subject to large interannual variability in oceanic conditions (Stockhausen et al., 2018). Stockhausen et al. (2018) refers to this as "running the gauntlet", as it is during this critical life stage that these fish are most vulnerable, experiencing the highest rates of mortality. This vulnerability is not only due to the vagaries of physical transport, but also due to their physiological condition where they must meet energetic demands of acquiring sufficient lipid reserves in order to move to inshore nursery areas.

During years favorable ocean conditions with ample food availability, such as 2014 for POP, mortality may be low and selection weak, allowing most phenotypes to survive

through the pelagic phase and into nearshore settlement. However, during years of unfavorable ocean conditions, such as the unusual warming, low primary productivity, and low food availability in 2015 for POP, mortality may be high. If this increase in mortality is especially high for certain phenotypes, the selection may be strong, with only the most favorable phenotypes surviving to settlement.

Our results show consistent selective forces along the sampling date/latitude gradient in both 2014 and 2015 for POP with 10 of the 381 putative selective loci being in common in both years (Table 1-4). The LFMM analysis was done independently for each of the years and finding the same putative selected loci in both years is surprising. And although the LFMM method purportedly accounts for demographic factors such as population mixtures, the date/latitude gradient association could be due adult spawning populations being differentiated at these loci. Based on timing and location of spawning, their progeny may follow the spatio-temporal pattern identified by GEA. This is further supported by the distribution of the sNMF identified genetic clusters in relation to their distribution as seen in Figure 1-3. Alternatively, this may indicate that the spawning adult populations contain a high proportion of alleles at those loci that in 2014 and 2015 years were deleterious to the YOY progeny encountering the environmental conditions during their pelagic developmental stage. Since POP are very long-lived and may even spawn into their 100th year (Conrath and Knoth, 2013; Hulson et al., 2017; Heppell et al., 2010), some of the alleles in the parental population are expected to have been selectively advantageous during their respective first year at sea; therefore, the alleles that were advantageous when the parents were YOY may be deleterious in some oceanic conditions encountered by their progeny decades later. It is then expected that patterns of selection as displayed by the subsets of selected alleles would be cohort-specific.

Interannual differences in the strength of selection pressure was evident when comparing the 2014 and 2015 YOY POP fish. Due to the larger sample size in 2014 (321) than in 2015 (77), we would expect more putative selected loci in 2014 just due to the increase in statistical power, but that was not the case. In 2014, the oceanic conditions were typical (Cavole et al., 2016), with large YOY abundances in the ocean, and no putative selected loci were identified aside from those associated with collection date/latitude. However, in 2015, the oceanic conditions were abnormal with warmer sea surface

temperatures (Gentemann et al., 2017) and were marked by large seabird die-off (Jones et al. 2018). This likely resulted in stronger selective pressure on YOY in 2015 and this is supported by the greater number of putative selected alleles. Therefore by the time the 2014 and 2015 cohorts settled out in the nearshore, we expect that most individuals have gene variants that were most favorable and selected for by the conditions encountered in that year.

The difference in the change in condition index indicates different growth conditions between the two years. In 2015, the smaller fish had less mass than in 2014, but the larger fish had equivalent mass in both years (Figure 1-4). This indicates that in 2015, a much warmer year than in 2014, the smaller fish were unable to gain weight as compared to the same sized fish in 2014. If the temperatures were still within optima for POP YOY growth, then smaller size suggests smaller-sized prey items were either unavailable or of insufficient nutritional value to support the higher growth rates predicted by the higher temperature in 2015. However the larger fish in both years were equally successful at gaining mass. This suggests that the environment in 2015 imposed a larger variance in fitness and therefore much stronger selection pressure, and this is consistent with the greater number of putative selected loci in 2015 than in 2014. This is further supported by the recruitment estimates in the 2017 stock assessment with 2014 cohort being 87.5 million and 38.2 million in 2015 (Hulson et al., 2017).

Gene Ontology Enrichment

The GO enrichment analysis yielded particularly interesting and intuitive results. The selective processes identified here act during the developmental and high growth larval life stage, and 87% of the general biological processes associated with the LFMM identified putative selected genes directly corresponded to development and growth. The remaining 7% and 5% were associated with intracellular processes and metabolism respectively. Furthermore, in 2014, we did not identify any biological processes associated with growth or metabolism, indicating that the early life conditions were favorable across the habitat surveyed, with little selection acting on those gene variants. The numerous processes identified in 2015, however, may be indicative of unfavorable oceanic conditions, leading to a significant loss of phenotypes with the deleterious gene variants. Because 2015 was an unusually warm year (Gentemann et al., 2017), it is not surprising that these warmer temperatures would directly affect metabolic processes. This is underscored by our finding of

associations between fatty acid, lipid metabolism, and temperature for 2015, but not for 2014.

Fluctuating Selection and Maintenance of Adaptive Diversity

Our results suggest the presence of a temporal portfolio effect, where a multi-age population with overlapping generations maintains a portfolio of genotypes (Ellner and Hairston 1994). The interannual variation in oceanic conditions and its effects on the selection of POP during their first year at sea prior to settlement may be thought of as a "Selective Sieve" (Figure 1-5), where each year presents different sets of selection pressures during the early developmental life stage. The selection pressures, in the form of various environmental drivers such as ocean temperatures, productivity (chlorophyll-a) and their timing vary from year to year resulting in some phenotypes being detrimental in one year, but advantageous in another when encountering highly diverse pelagic habitats. The selective sieve is therefore specific to the year of the POP pelagic life stage, and therefore unique to each cohort which then contains alleles favored by the conditions of their first year. In these long-lived species with lifespans of over 100 years, in any one year the larval cohort at parturition may be the result of breeding across dozens of spawning aged cohorts (~ 8 to 100 years old or more). The parental cohorts contain many alleles that are representative of the selection due to oceanic conditions during their first year at sea. At parturition, the POP larvae contain all of these alleles, however, from parturition to settlement, some of the alleles prove to be deleterious as the oceanic conditions do not favor them, and only a subset of the larvae containing the advantageous alleles survive until settlement. This is an example of fluctuating selection (Bell 2010; Kawecki 2000; Lande 2007), where the direction of selection is constantly changing between generations. The species' life history of long reproductive period relative to the time-scale of fluctuating selection maintains genetic diversity that is adaptive across a range of environmental variation.

Population viability in fish employing broadcast spawning strategies is especially vulnerable to changing oceanographic conditions. Ocean currents may advect YOY far offshore where they will fail to reach their shelf-slope nursery areas. Using ROMS-based models, Stockhausen et al. (2018) showed that up to 70% of the YOY failed to reach suitable nursery habitats prior to wintertime and were not expected to survive. The ones that are not advected out of reach of nursery habitat must still acquire sufficient lipid reserves in order to

settle out and overwinter. Interannual differences in ocean temperatures, prey and predator abundances and composition will also affect whether the YOY will survive to reach their nursery habitats with sufficient lipid reserves to overwinter and eventually recruit to the population. Maintaining a high diversity in phenotypes through cohort-specific selection may be thought of as a form of diversification bet hedging response to a fluctuating natural selection as described by Simons (2009).

These results underscore the importance of maintaining many cohorts in order to maximize the population resilience to environmental variability. POP are vulnerable to age truncation where older fish are more likely to be fished since they are exposed to fishing longer (Berkeley 2004). The importance of maintaining older age classes in marine fishes has long been recognized as being a factor in their recruitment (Longhurst, 2002; Hixon et al., 2013). Hanselman et al. (2005) noted that age truncation has occurred in POP due to unrestricted fishing in the past which led to disproportional absence of 40+ year old fish. However, the mechanism of adaptation through maintenance of age-specific advantageous alleles would be compromised if whole cohorts are inadvertently fished by depriving populations of the advantageous alleles specific to that cohort.

The uniqueness of the demonstrated cohort-specific selection signatures may allow for reconstruction of past oceanographic conditions based on the alleles present in a given cohort. The 2015 cohort will therefore represent the alleles favored (or conversely lost) during especially warm oceanic conditions as experienced during 2015. It may be possible that by examining allele frequencies in an adult cohort, of for example 50-year-old fish, the selection pressures encountered during their YOY stage may be revealed. Furthermore, ageing of adults based on cohort-specific allelic signatures may also be possible by maintaining cohort-specific selected allelic signatures. This may prove especially useful since otolith ageing of POP adults is fairly error prone especially for older fish (>20 years old) (Beamish, 1979; Stanley, 1986).

Conclusions

We found evidence for different selective pressures for POP YOY across two different years that had very different environmental conditions. These results provide evidence that long-lived marine species such as POP may be resilient to natural environmental variability by maintaining a portfolio of adaptive alleles resulting from

selection encountered by each cohort during their most vulnerable life stage from parturition to settlement. However, this resilience may be limited to the environmental conditions that prevailed in the last few centuries. The “selective sieve” framework may provide valuable insights into other species employing similar life history strategies. Hoffman and Sgro (2011) note that species facing strong but fluctuating selection pressures, such as YOY POP during the pelagic life stage, will have a difficult time adapting. Here, we demonstrated an exception where due to the way POP are able to maintain these selected alleles may allow them to be especially adaptable under fluctuating environmental conditions. POP have proven to be an ideal model species for investigating portfolio effects. By examining relative strengths of selection among discrete populations and adult cohorts, it allows us to jointly examine spatial and temporal portfolio effects. In the future, we plan to sample across adult populations to link genetic variation to larval cohorts and adult habitat/geographic population structure.

Tables and Figures

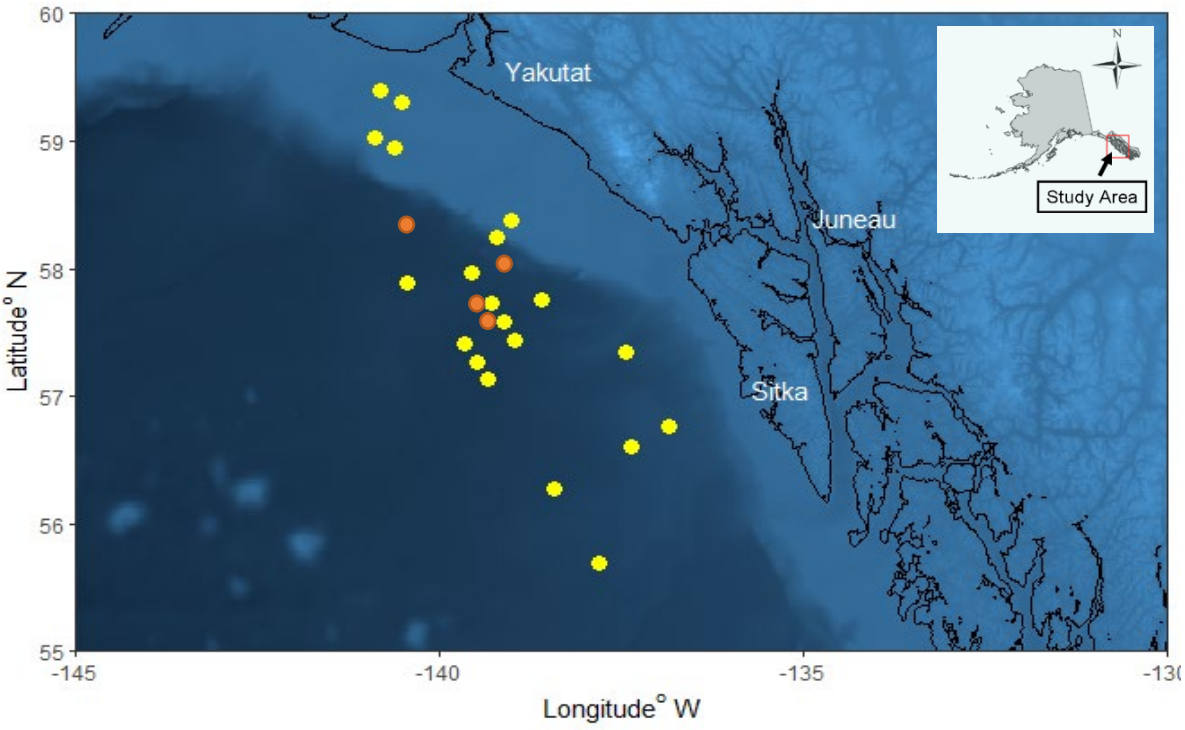
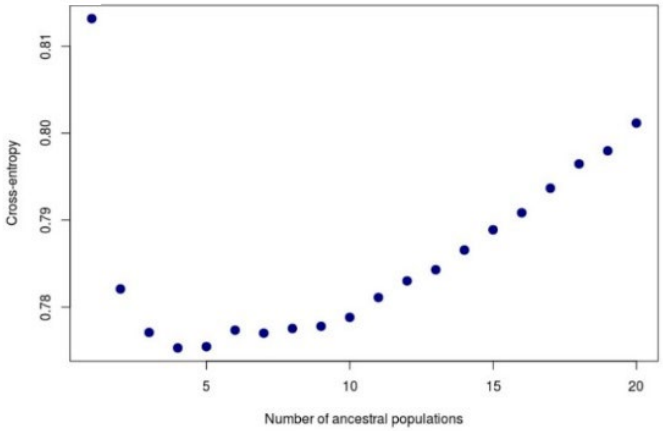
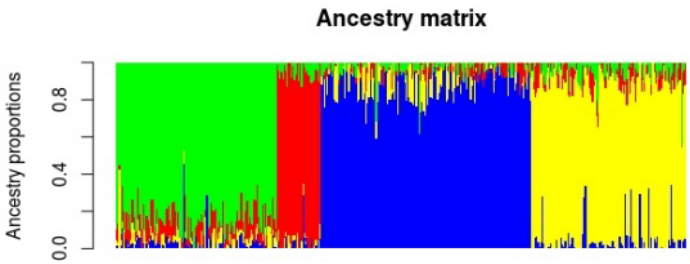


Figure 1-1 Locations of the 2014 (yellow) and 2015 (orange) collection of the young-of-the-year Pacific ocean perch.

(a)


Cross-entropy
0.78
0.79
0.80
0.81
5
10
15
20
Number of ancestral populations

(b)


Ancestry matrix
Ancestry proportions
0.0
0.4
0.8

(c)

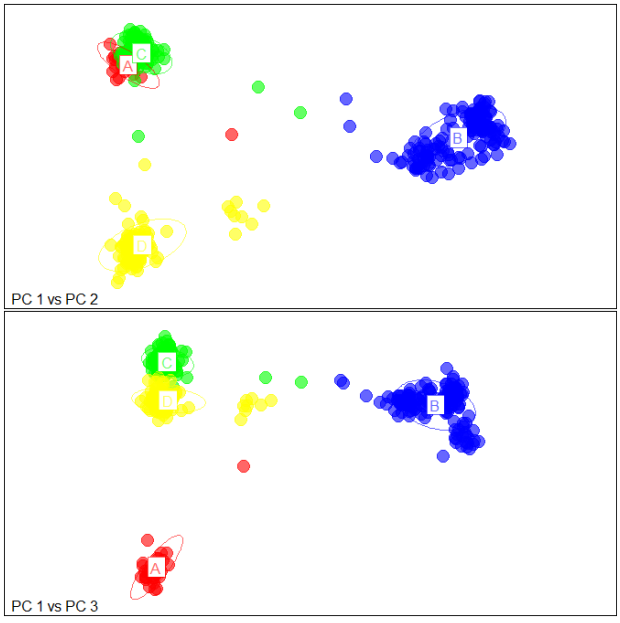

A
C
B
D
PC 1 vs PC 2
C
D
B
A
PC 1 vs PC 3

Figure 1-2 sNMF ancestry analysis revealed 4 ancestral populations represented by fish collected in both 2014 and 2015: (a) cross-entropy plot for the number of populations in sample; (b) sNMF population ancestry barplot; and (c) PCA analysis with the colors corresponding to the sNMF-derived majority ancestry populations in panel b. Note that both years were included concurrently in the analysis. For clarity year designation is omitted as each population cluster contains both years interspersed throughout.

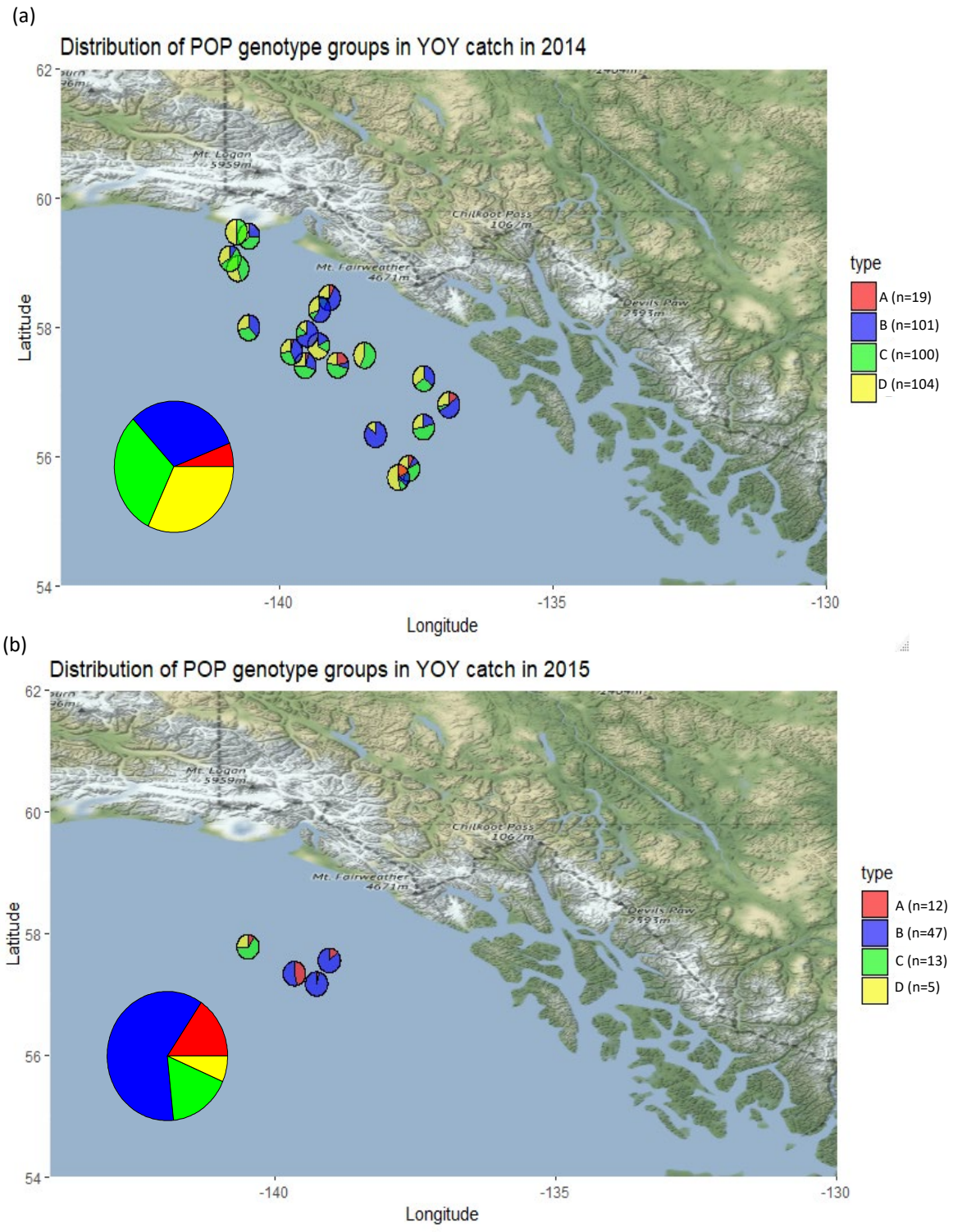


Figure 1-3 Spatial distribution and the proportional representation of the putative populations in the samples in 2014 (a) and 2015 (b) collections.

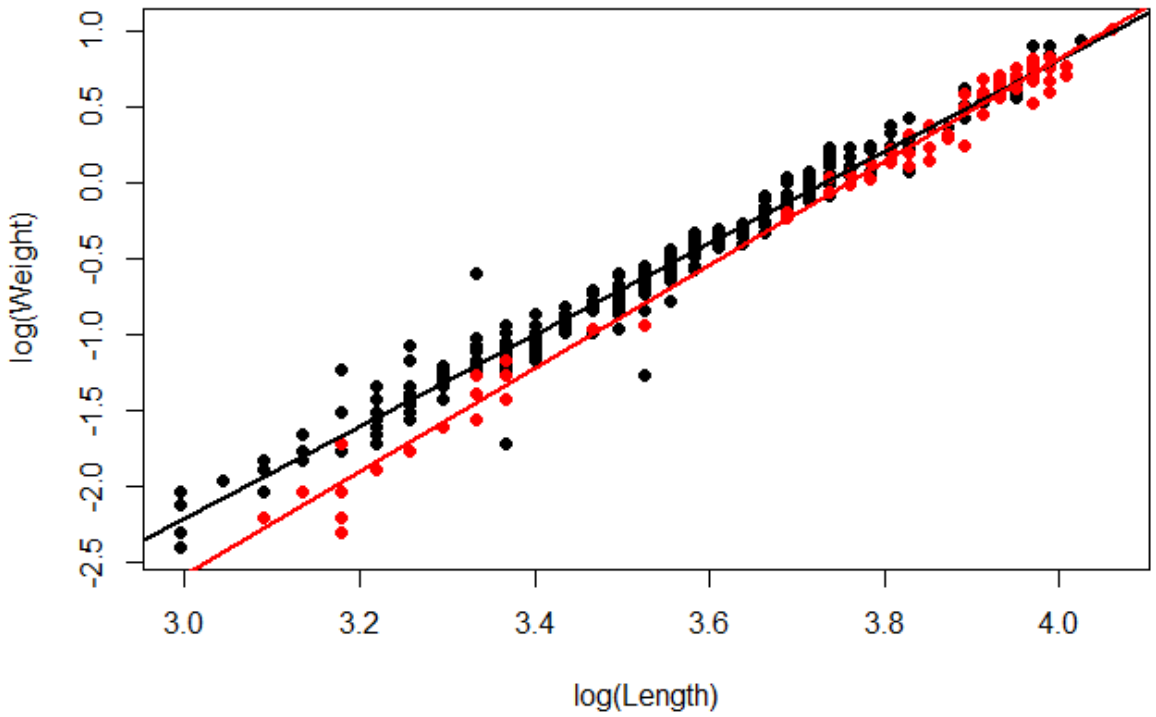


Figure 1-4 Relationships of body length and weight in the 2014 (black and 2015 (red) young-of-the-year Pacific ocean perch. In 2015, smaller length fish had significantly ($p<0.0001$) smaller mass than in 2014, indicating environmental factors in 2015 may have negatively influenced their condition.

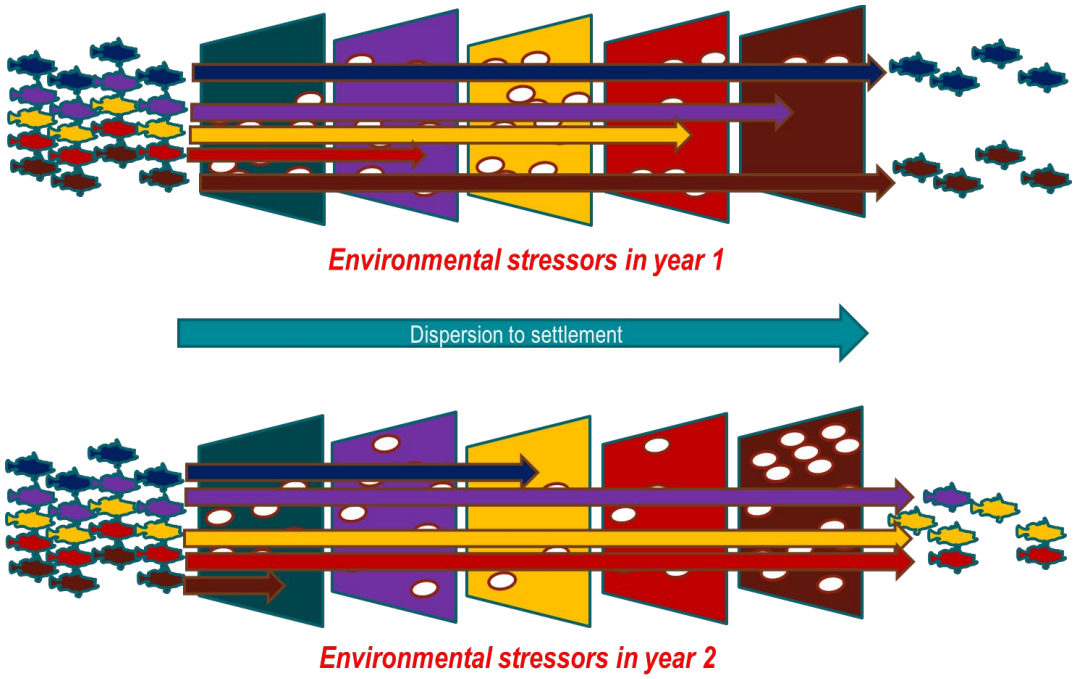


Figure 1-5 The selective sieve. The five colored plates represent various hypothetical environmental forces (such as temperature, chlorophyll density, etc), that are highly variable among years. This represent the different selection pressures encountered by the Pacific Ocean Perch during initial pelagic life stage. Each years' cohort therefore contains the alleles that were selected for during their first year. Populations of long-lived adults representing multiple cohorts maintain genetic diversity as a result of this temporal variation in selection.

Table 1-1 Pairwise F_{ST} values (F_{ST} below diagonal and p-value above the diagonal) between sNMF-derived populations and sampling year. There is little genetic differentiation as indicated by F_{ST} within a population between years (bolded F_{ST} and corresponding p-values).

	2014-A	2014-B	2014-C	2014-D	2015-A	2015-B	2015-C	2015-D
2014-A	*	0.001	0.001	0.001	**0.021**	0.001	0.001	0.001
2014-B	0.023	*	0.001	0.001	0.001	**0.329**	0.001	0.001
2014-C	0.030	0.014	*	0.001	0.001	0.001	**0.484**	0.001
2014-D	0.023	0.009	0.012	*	0.001	0.011	0.001	**0.303**
2015-A	**0.001**	0.026	0.032	0.026	*	0.001	0.001	0.001
2015-B	0.023	**0.000**	0.013	0.007	0.026	*	0.001	0.017
2015-C	0.030	0.014	**0.000**	0.011	0.032	0.013	*	0.001
2015-D	0.023	0.009	0.011	**0.000**	0.026	0.008	0.012	*

Table 1-2 Number of private alleles in the given year specific to each sNMF-derived population. Private alleles is the number of alleles that were detected in only one year. For example, 127 is the number of alleles detected in population A in 2014, but not in 2015.

	Sample Size		Private Alleles	
Population	2014	2015	2014	2015
A	100	47	127	1[1]
B	103	5	2,101	0[1]
C	19	12	604	149[1]
D	99	13	401	4

[1]Indicates that there are significantly fewer alleles lost in 2015 than expected by chance.

Table 1-3 Number of private alleles in common between the sNMF-derived populations that were detected in 2014 samples but not in 2015. For example, 30 is the number of alleles detected in 2014 in both populations A and B, that were not detected in 2015 in either A or B population

Population	A	B	C
B	30		
C	4	181	
D	5	180	74

Table 1-4 Results of LFMM analysis and the number of putative loci under divergent selection in 2014 and 2015. Note that sampling date and latitude (both are confounded as sampling was generally in the northward direction) were consistently associated with selection pressure in both years. However, other environmental factors only showed signatures of selection in 2015.

	Selected Loci	
Gradient	**2014**	**2015**
Latitude	14	56
Sampling date	62	101
Temperature	0	16
Chlorophyll	0	100
% Lipid	0	27
Condition Index	0	5

Supplemental Materials

Supplemental Table 1-1 Putative selected loci and their associated environmental gradient that were in common in both 2014 and 2015 years.

			Environmental Gradient		
Locus	**Year**	**Date**	**Latitude**	**Chlorophyll**	**% Lipid**
10101	2014	X			
	2015		X		
2088	2014	X			
	2015	X	X	X	X
24006	2014		X		
	2015	X		X	
25342	2014	X			
	2015	X		X	
2579	2014	X			
	2015		X	X	
28846	2014	X	X		
	2015	X	X	X	X
29782	2014	X	X		
	2015	X	X	X	
343	2014	X			
	2015	X		X	
8626	2014	X			
	2015	X	X	X	
922	2014	X	X		
	2015	X		X	

Supplemental Table 1-2 List of BLAST matches for the selected loci from LFMM in 2014 and 2015.

2014	Date	Lat
biogenesis of lysosome-related organelles complex 1 subunit 4-like (BLOC-1)	X	
lysine-specific demethylase 5B-B-like, transcript variant X4 (KDM5B)	X	
neurochondrin (NCDN)	X	
GREB1-like protein	X	
kelch-like protein 36 (KLHL36)	X	
WD repeat domain 3 (WDR3)	X	X

2015	Date	Lat	Temp	Lipid	Chl	Condition Index
collagen type IV alpha 2 chain (col4a2)	X					
DnaJ heat shock protein family (Hsp40) member A2		X	X	X	X	
keratin, type II cytoskeletal 8-like (KRT8)	X				X	
MDS1 and EVI1 complex locus (mecom)	X				X	
phosphoribosyltransferase domain-containing protein 1 (PRTFDC1)		X				
STE20-like serine/threonine-protein kinase-like (SLK)	X	X			X	
teneurin transmembrane protein 1 (tenm1)	X	X			X	
alanyl-tRNA synthetase (aars)						X
aldehyde reductase (AKR1A1) gene	X	X		X	X	
alsin-like (ALS2)	X			X	X	
enkurin, TRPC channel interacting protein (enkur)		X				
erb-b2 receptor tyrosine kinase 4 (erbb4)	X					
myotubularin-related protein 13-like (sbf2)		X				
NGFI-A binding protein 1 (nab1)	X	X		X	X	X
nucleoprotein (TPR)	X	X	X	X	X	X
voltage-dependent L-type calcium channel subunit alpha-1D (CACNA1D)		X				

putative helicase mov-10-B.2 (MOV-10)		X				
retromer complex component B (vps26b)	X	X			X	
solute carrier family 25 member 46 (slc25a46)	X					
T-box 20 (tbx20)	X	X	X	X	X	
teneurin-1-like (TENM1)	X	X			X	
trypsin domain containing 1 (tysnd1)			X			
ubiquitin specific peptidase 32 (usp32), transcript variant X7	X				X	
zinc finger CCCH-type containing 3 (zc3h3)	X	X		X	X	

Supplemental Table 1-3 Summary of gene ontology and enrichment analysis showing broad biological processes associated with selected genes from LFMM analysis. The numbers in the table indicate how many distinct biological functions were found in each category and do not indicate significance or importance of the associated putatively selected alleles. "Total processes" indicates the total number of distinct biological processes identified, whereas "Discrete processes" indicates the number of broad biological process groups.

Broad Biological Process	**2014**	**2015**						
	Date	**Chl**	**Date**	**Condition**	**Latitude**	**Lipid %**	**Temp**	**Total**
intracellular processes	1	39	38	22	72	29	12	213
Development								
cell development		6	5	2	2	1	1	17
cell adhesion		1	1		1			3
extracellular matrix organization			1					1
determination of symmetry						2	2	4
embryo development		8	6	1		3	3	21
developmental growth		3	6		1	6	6	22
anatomical structure/organ development	1	2	2					5
epithelium development		2	2			2	2	8
cardiac development		13	11		13	16	16	69
vasculogenesis		2	2		1	4	4	13
liver development		1	1		1	1	1	5
renal system development	1	20	19			1	1	42
digestive system development		4	4		4	4	4	20
gland development	2							2
brain development	2							2
neuron generation			4					4
immune system development						1	1	2
pigment biosynthesis					2			2
Growth								

non-developmental growth						1	1	2
muscle formation		5	5		5	7	7	29
growth regulation		1	1		1	1	1	5
organ growth		2	2		2	2	2	10
cell growth		3	3			1	1	8
blood production						1	1	2
Metabolism								
metabolic processes				13	15		8	36
fatty acid metabolism							3	3
lipid metabolism							2	2
Total processes	7	112	113	38	120	83	77	550
Discrete processes	5	16	18	4	13	18	21	

References

Alexander, D.H., Novembre, J. & K. Lange. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19:1655–1664.

Ali, O.A., O'Rourke S.M., Amish, S.J., Meek, M.H., Gordon, L., Carson, J. and Miller, M. R. (2016). RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. Genetics 202(2): 389-400.

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

Andrews, K.R., Adams J, Cassirer F, Plowright R, Gardner C, Dwire M, Hohenlohe PA, & Waits L. (2018). A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RADseq data. Molecular Ecology Resources 18: 1263-1281.

Barrio, A.M., Lamichhaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H., Dainat, J., Ekman, D., Höppner, M., Jern, P. & Martin, M., (2016). The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. eLife 2016;5:e12081.

Beamish, R. J. (1979). New information on the longevity of Pacific ocean perch (*Sebastes alutus*). Journal of the Fisheries Board of Canada, 36(11), 1395-1400.

Bell, G. (2010). Fluctuating selection: the perpetual renewal of adaptation in variable environments. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1537), 87-97.

Berkeley, S. A., Hixon, M. A., Larson, R. J., & Love, M. S. (2004). Fisheries sustainability via protection of age structure and spatial distribution of fish populations. Fisheries, 29(8), 23-32.

Brendonck L, De Meester L. 2003. Egg banks in freshwater zooplankton: evolutionary and ecological archives in the sediment. Hydrobiologia, 491: 65-84.

Carlson, H. R., & Haight, R. E. (1972). Evidence for a home site and homing of adult yellowtail rockfish, *Sebastes flavidus*. Journal of the Fisheries Board of Canada, 29(7), 1011-1014.

Carlson, H. R., & Haight, R. E. (1976). Juvenile life of Pacific ocean perch, *Sebastes alutus*, in coastal fiords of southeastern Alaska: Their environment, growth, food habits, and schooling behavior. Transactions of the American Fisheries Society, 105(2): 191-201.

Carlson, H. R., Haight, R. E., & Helle, J. H. (1995). Initial behavior of displaced yellowtail rockfish *Sebastes flavidus* in Lynn Canal, Southeast Alaska. Alaska Fisheries Research Bulletin, 2(1): 76-80.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.

Catchen, J.M., Hohenlohe, P.A., Bassham, S, Amores, A, Cresko, WA. (2013). Stacks: an Analysis tool set for population genomics. Molecular Ecology 22: 3124-3140.

Catchen, J., Hohenlohe, P.A., Bernatchez, L., Funk, W.C., Andrews, K.R., & Allendorf, F. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Molecular Ecology Resources 17: 362-365

Cavole, L. M., Demko, A. M., Diner, R. E., Giddings, A., Koester, I., Pagniello, C. M., Paulsen, M-L., Ramirez-Valdez, A., Schwenck, S. M., Yen, N. K., Zill, M. E., & Franks, P. J. S. (2016). Biological impacts of the 2013–2015 warm-water anomaly in the Northeast Pacific: winners, losers, and the future. Oceanography, 29(2), 273-285.

Chabrol, E., Charonnet R (1937) Une novelle reaction pour l'etude des lipides. Presse Med 45:1713

Conrath, C. L., & Knoth, B. (2013). Reproductive biology of Pacific ocean perch in the Gulf of Alaska. Marine and Coastal Fisheries, 5(1), 21-27.

Cure, K., Thomas, L., Hobbs, J. P. A., Fairclough, D. V., & Kennington, W. J. (2017). Genomic signatures of local adaptation reveal source-sink dynamics in a high gene flow fish species. Scientific Reports, 7(1), 8618.

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. Briefings in Functional Genomics, 9(5-6), 416-423.

Dray S, Dufour A (2007). The ade4 package: Implementing the duality diagram for ecologists. Journal of Statistical Software, 22(4), 1–20. doi: 10.18637/jss.v022.i04.

Ellner S, Hairston NG. 1994. Role of overlapping generations in maintaining genetic variation in a fluctuating environment. The American Naturalist, 143: 403-417.

Folch, Jordi, M. Lees, & G. H. Sloane Stanley. (1957). A simple method for the isolation and purification of total lipides from animal tissues. Journal of Biological Chemistry, 226(1): 497-509.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. Genetics, 196(4), 973-983.

Frichot E, Francois O. (2015). LEA: an R package for landscape and ecological association studies. Methods in Ecology and Evolution 6: 925– 929.

Froese, R. (2006). Cube law, condition factor and weight–length relationships: History, meta-analysis and recommendations. Journal of Applied Ichthyology 22(4): 241-253.

Garvin, M.R., Marcotte, R.W., Palof, K.J., Riley, R.J., Kamin, L.M. & Gharrett, A.J. (2011) Diagnostic single nucleotide polymorphisms (SNPs) identify Pacific ocean perch and delineate blackspotted and rougheye rockfish. Trans. Am. Fish. Soc. 140:984–988.

Gentemann, C. L., Fewings, M. R., & García-Reyes, M. (2017). Satellite sea surface temperatures along the West Coast of the United States during the 2014–2016 northeast Pacific marine heat wave. Geophysical Research Letters, 44(1), 312-319.

Gosselin, T. (2018) radiator: RADseq data exploration, manipulation and visualization using R. R package version 0.0.11. Retrieved from https://github.com/thierrygosselin/radiator.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes, 5(1), 184-186.

Groot, C., & Margolis, L. (Eds.). (1991). Pacific salmon life histories. UBC Press. 608 p.

Hanselman, D., Heifetz, J., Fujioka, J. T., & Ianelli, J. N. (2005). Gulf of Alaska Pacific ocean perch. Stock assessment and fishery evaluation report for the groundfish fisheries of the Gulf of Alaska, 525-578. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. North Pacific Fishery Management Council, 605W. 4th Ave., Suite 306, Anchorage, AK 99501-2252.

Heppell, S. S., Heppell, S. A., Spencer, P. D., Smith, W. D., & Arnold, L. (2010). Assessment of female reproductive effort and maternal effects in Pacific ocean perch *Sebastes alutus*: Do big old females matter? North Pacific Research Board Final Report, NPRB Project 629. Oregon Sea Grant College Program.

Hixon, M. A., Johnson, D. W., & Sogard, S. M. (2013). BOFFFFs: on the importance of conserving old-growth age structure in fishery populations. ICES Journal of Marine Science, 71(8), 2171-2185.

Hoey, A. S., & McCormick, M. I. (2004). Selective predation for low body condition at the larval-juvenile transition of a coral reef fish. Oecologia 139(1):23-29.

Hoffmann, A. A., & Sgro, C. M. (2011). Climate change and evolutionary adaptation. Nature, 470(7335), 479.

Huisman, J. (2017). Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Molecular Ecology Resources, 17(5): 1009-1024. R package version 1.0.2, Retrieved from https://github.com/JiscaH/sequoia.

Hulson, P.J., Hanselman, D.H., Lunsford, C.R. and Fissel, B. (2017). Assessment of the Pacific Ocean perch stock in the Gulf of Alaska. P. 913-992. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska. North Pacific Fishery Management Council, 605W. 4th Ave., Suite 306, Anchorage, AK 99501-2252.

Jasonowicz, A.J., Goetz, F.W., Goetz, G.W. & Nichols, K.M. (2016). Love the one you're with: Genomic evidence of panmixia in the sablefish (*Anoplopoma fimbria*). Canadian Journal of Fisheries and Aquatic Sciences, 999:1-11.

Jombart T., Devillard S., & Balloux F. (2010) Discriminant analysis of principal components: a New method for the analysis of genetically structured populations. BMC Genetics 11:94

Jones, T., Parrish, J. K., Peterson, W. T., Bjorkstedt, E. P., Bond, N. A., Ballance, L. T., Bowes, V. J., Hipfner, M., Burgess, H. K., Dolliver, J. E., Lindquist, K., Lindsey, J., Nevins, H. M., Robertson, R. R. Roletto, J., Wilson, L., Joyce, T., & J. Harvey. (2018). Massive mortality of a planktivorous seabird in response to a marine heatwave. Geophysical Research Letters, 45(7), 3193-3202.

Kamin, L. M., Palof, K. J., Heifetz, J., & Gharrett, A. J. (2014). Interannual and spatial variation in the population genetic composition of young-of-the-year Pacific ocean perch (*Sebastes alutus*) in the Gulf of Alaska. Fisheries Oceanography, 23(1), 1-17.

Kamvar, Z. N., Tabima, J.F., Grünwald, N.J. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. https://doi.org/10.7717/peerj.281.

Kamvar Z. N., Brooks JC & Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. 6:208.

Kawecki, T. J. (2000). The evolution of genetic canalization under fluctuating selection. Evolution, 54(1), 1-12.

Kendall, A.W., Kondzela, C., Li, Z., Clausen, D. & Gharrett, A.J. (2007) Genetic and morphological identification of pelagic juvenile rockfish collected from the Gulf of Alaska. U.S. Department of Commerce, NOAA Professional Paper NMFS 9, 26 p.

Kumar, G., & Kocour, M. (2017). Applications of next-generation sequencing in fisheries research: a Review. Fisheries Research, 186, 11-22.

Lande, R. (2007). Expected relative fitness and the adaptive topography of fluctuating selection. Evolution, 61(8), 1835-1846.

Langmead B, & Salzberg S. (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357-359.

Levin, P.S, & C. Möllmann C. (2015) Marine ecosystem regime shifts: challenges and opportunities for ecosystem-based management. Philosophical Transactions of the Royal Society B: Biological Sciences. 370(1659):20130275.

Link, J.S. (2002) What does ecosystem-based fisheries management mean. Fisheries, 27(4):18-21.

Longhurst, A. (2002). Murphy's Law revisited: Longevity as a factor in recruitment to fish populations. Fisheries Research, 56(2), 125-131.

Love, M. S., Yoklavich, M., & Thorsteinson, L. K. (2002). The Rockfishes of the Northeast Pacific. University of California Press. 416 p.

Major, R. L., and H. H. Shippen. (1970). Synopsis of biological data on Pacific ocean perch, Sebastodes alutus. FAO Fisheries Synopsis No. 79, NOAA Circular 347, 38 p.

Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: Matching biological questions with appropriate techniques. Trends in Ecology & Evolution, 20(3), 136-142.

Marshall, T. C., Slate, J. B. K. E., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology, 7(5), 639-655.

Matthews, K. R. (1990). A telemetric study of the home ranges and homing routes of copper and quillback rockfishes on shallow rocky reefs. Canadian Journal of Zoology, 68(11), 2243-2250.

McKinney, G. J., Larson, W. A., Seeb, L. W. & Seeb, J. E. (2017), RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). Mol Ecol Resour, 17: 356–361. doi:10.1111/1755-0998.12649.

Megrey, B. A. (1988). Review and comparison of age-structured stock assessment models. NWAFC Processed Rep., 88-21, 116 p. Northwest and Alaska Fish. Cent., Natl. Mar. Fish. Serv., NOAA, 7600 Sand Point Way NE, Seattle, WA 98115-0070.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology, 22(11), 2841-2847.

Nei, M. (1987). Molecular Evolutionary Genetics. Columbia University Press. 512 p.

NOAA (2019). Commercial Fishery Statistics – Annual Commercial Landings by Group: https://www.fisheries.noaa.gov/national/sustainable-fisheries/commercial-fisheries-landings

Newman, D. L., Abney, M., McPeek, M. S., Ober, C., & Cox, N. J. (2001). The importance of genealogy in determining genetic associations with complex traits. The American Journal of Human Genetics, 69(5), 1146-1148.

Pacifici, M., Foden, W. B., Visconti, P., Watson, J. E., Butchart, S. H., Kovacs, K. M., Scheffers, B. R., Hole, D. G., Martin, T. G., Akçakaya, H. R., Corlett, R. T., Huntley, B., Bickford, D., Carr, J. A., Hoffmann, A. A., Midgley, G. F., Pearce-Kelly, P., Pearson, R. G., Williams, S. E., Willis, S. G. Young, B., & Rondinini, C. (2015). Assessing species vulnerability to climate change. Nature Climate Change, 5(3), 215.

Palof, K.J., Heifetz, J., & Gharrett, A.J. (2011). Geographic structure in Alaskan Pacific ocean perch (*Sebastes alutus*) indicates limited life-time dispersal. Mar. Biol. 158:779-792.

Parker, S.J., Berkeley, S.A., Golden, J.T. et al (2000). Management of Pacific rockfish. Fisheries 25:22–29

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Moss, J. H., Shotwell, S. K., Heintz, R. A., Atkinson, S., Debenham, C., Fournier, W., Golden, N., Heifetz, J., Mueter, F. J., Pirtle, J. L., Reid, J. A., Slater, L., Sreenivasan, A., Will, A., Zaleski, M., & Zimmermann, M. (2016). Surviving the Gauntlet: A comparative study of the pelagic, demersal, and spatial linkages that determine groundfish recruitment and diversity in the Gulf of Alaska ecosystem (NPRB GOA Project G81 Upper Trophic Level Final Report). North Pacific Research Board, Anchorage, AK.

Rooper, C. N., Boldt, J. L., Batten, S., & Gburski, C. (2012). Growth and production of Pacific ocean perch (*Sebastes alutus*) in nursery habitats of the Gulf of Alaska. Fisheries Oceanography, 21(6), 415-429.

Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., & Planes, S. (2009). Estimating connectivity in marine populations: an Empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. Molecular Ecology, 18(8), 1765-1776.

Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. Nature, 465(7298), 609.

Sigler, M. F., M. P. Eagleton, T. E. Helser, J. V. Olson, J. L. Pirtle, C. N. Rooper, S. C. Simpson, & R. P. Stone. (2017). Alaska Essential Fish Habitat Research Plan: A Research Plan for the National Marine Fisheries Service's Alaska Fisheries Science Center and Alaska Regional Office. AFSC Processed Rep. 2015-05, 22 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Simons, A. M. (2009). Fluctuating natural selection accounts for the evolution of diversification bet hedging. Proceedings of the Royal Society B: Biological Sciences, 276(1664), 1987-1992.

Stanley, R. D. (1986). A comparison of age estimates derived from the surface and cross-section methods of otolith reading for Pacific ocean perch (*Sebastes alutus*), p. 187-196. *In* Proceedings of the International Rockfish Symposium. Alaska Sea Grant College Program, AK-SG-87-02.

Stockhausen, W. T., & Hermann, A. J. (2007). Modeling larval dispersion of rockfish: a Tool for marine reserve design. Biology, assessment, and management of North Pacific rockfishes, 251-273.Alaska Sea Grant College Program, AK-SG-07-01.

Stockhausen, B. (2009). DisMELS: A Dispersal Model for Early Life History Stages. http://www.afsc.noaa.gov/Quarterly/jfm2009/divrptsREFM5.htm. Last accessed July 25, 2016.

Stockhausen, W. T., Coyle, K. O., Hermann, A. J., Doyle, M., Gibson, G. A., Hinckley, S., Ladd, C., & Parada, C. (2018). Running the gauntlet: Connectivity between natal and nursery areas for Pacific ocean perch (*Sebastes alutus*) in the Gulf of Alaska, as inferred from a biophysical individual-based model. Deep-sea Research Part II: Topical Studies in Oceanography, 165:74-88.

Sunday, J. M., Crim, R. N., Harley, C. D., & Hart, M. W. (2011). Quantifying rates of evolutionary adaptation in response to ocean acidification. PLOS ONE, 6(8), e22881.

Valenzuela-Quiñonez, F., (2016). How fisheries management can benefit from genomics? Briefings in Functional Genomics, 15(5): 352-357. p.elw006.

Voight, B. F., & Pritchard, J. K. (2005). Confounding from cryptic relatedness in case-control association studies. PLOS Genetics. https://doi.org/10.1371/journal.pgen.0010032.

Wang, H.-Y., & Höök, T. O. (2009). Eco-genetic model to explore fishing-induced ecological and evolutionary effects on growth and maturation schedules. Evolutionary Applications, 2(3), 438–455.

Wenne, R., Boudry, P., Hemmer-Hansen, J., Lubieniecki, K.P., Was, A., & Kause, A. (2007). What role for genomics in fisheries management and aquaculture? Aquat. Living Resour., 20 3 241-255.

Westrheim, S. J. (1958). On the biology of the Pacific ocean perch, *Sebastodes alutu*s (Gilbert) (Doctoral dissertation, University of Washington).

Westrheim, S. J. (1975). Reproduction, maturation, and identification of larvae of some *Sebastes* (Scorpaenidae) species in the northeast Pacific Ocean. Journal of the Fisheries Board of Canada, 32(12), 2399-2411.

Withler, R., Beacham, T., Schulze, A., Richards, L., & Miller, K. (2001). Co-existing populations of Pacific ocean perch, *Sebastes alutus*, in Queen Charlotte Sound, British Columbia. Marine Biology, 139(1), 1-12.

Chapter 2: Beyond Isolation by Distance; Population Structure and Dispersal

Abstract

In order to test whether biophysical dispersal models may describe the population structure of marine fish species, I develop a spatio-temporal stochastic genetic model and validate it with a suite of synthetic dispersal models. I then apply PCA, STRUCTURE admixture analysis and linearized F_{ST} on distance regression and show that no single method is adequate to describe the underlying population structure model. The effect of multiple age cohorts, and cohort specific selection is shown to have a negative impact on the population structure inference leading to lack of differentiation as well as phantom populations. Lastly, I show how the beta distribution parameters may be used to differentiate candidate population dispersal models in a likelihood model selection framework.

Introduction

Inferring population structure is one of the primary goals of population genetics. Population structure indicates any deviation from the expectation of random mating resulting in population heterogeneity. It is now commonplace that a first step after sequencing genotypes is to examine the principal component (PCA) plots of sample genotypes for any patterns that may result from non-random mating. PCA partitions the total variation of the genomic sequence samples in orthogonal space and by plotting the first two components is able to visually show underlying patterns in the data in this dimensionally reduced space. Oftentimes, this allows the identification of genomic clusters, or individual groups with sufficient genetic differentiation. As well, STRUCTURE analysis (Pritchard, 2010; Pritchard et al., 2000), infers population structure by assigning individuals to distinct populations as well as identify admixed individuals by grouping individuals into groups satisfying the Hardy-Weinber equilibrium. Finally, under limited dispersal among geographically divided populations the population structure oftentimes can be described as isolation by distance where the populations separated by geographic distance are proportionally genetically diverged, and this is measured through a linearized F_{ST} regression.

Isolation by distance, first described by Wright (Sewall Wright, 1943), abbreviated IBD is a commonly used measure to assess population structure, with inference usually

limited to its presence/absence based on significance of the genetic differentiation (F_{ST}) regression on spatial distance (i.e. Euclidean geographic distance) (Bradburd et al., 2013; I. R. Bradbury & Bentzen, 2007; Duforet-Frebourg & Slatkin, 2016; Jenkins et al., 2010; Rousset, 1997; Slatkin & Barton, 1989). However, when the genetic differentiation is more complex, or nonlinear, this regression may not result in the apparent significant linear relationships, possibly resulting in a failure to detect population structure.

The nonlinear dispersal, possibly independent of geographic distance has been explored through isolation by resistance (McRae, 2006), isolation by environment (Sexton et al., 2014; Wang & Bradburd, 2014) and is of primary interest in the field of landscape and seascape genetics. Isolation by resistance, unlike geographic distance, aims to estimate the landscape resistance pathways among populations (McRae, 2006), while isolation by environment estimates connectivity based on the similarity of environments (Wang & Bradburd, 2014). These methods exploit the relationship between the physical connectivity matrix and the genetic population differentiation (i.e. pairwise F_{ST} matrix).

The most often used approach to estimate the relationship between the connectivity matrix and the genetic distance matrix is the Mantel's test (Mantel, 1967; Sokal, 1979) or partial Mantel's test (Smouse et al., 1986). And although popular in landscape genetics methodology (Balkenhol et al., 2009), numerous simulation studies have shown that this test tends to suffer from a high type 1 error rate and autocorrelation (Castellano & Balletto, 2002; Guillot & ßois Rousset, 2012; Harmon & Glor, 2010; Raufaste & Rousset, 2001; Rousset, 2002). However, its use continues to persist among researchers (Stéphanie Manel & Holderegger, 2013; Storfer et al., 2010). Other methods used to compare the environmental distance or connectivity to genetic distance are BIO-ENV (K R Clarke & Ainsworth, 1993), RELATE (K. Robert Clarke et al., 2008), multiple regression (Bradburd et al., 2013; Legendre & Casgrain, 1994), and others. See (Balkenhol et al., 2009) for a review and simulation study comparing these methods in the landscape genetics framework.

Here, I estimate the association between the environmental connectivity and genetic connectivity through the expected genetic population differentiation resulting from a set of modeled environmental connectivity matrices. Although not a novel approach (Slatkin, 1980, 1993), this model differs in that I do not consider migration with the associated exit of genotypes, but offspring dispersal, where individuals are never leaving their natal population,

only their genetic propagules, i.e. larvae. In addition, I model a life history characteristic in numerous marine species with multiple spawning cohorts, and test its effect on population differentiation.

Methods

Theoretical Population Demographic Framework

Consider the demographic life history model that is characterized by larval stage dispersal, but no subsequent migration, so that the allele frequencies in each of the populations only change due to immigration from dispersal and no emigration. The demographic life history may then be modeled as follows:

1. Spawning aggregate at each discrete population may consist of multiple cohorts.
2. The larvae from each population then disperse according to some dispersal parameter matrix.
3. Selection occurs during the larval stage, and is specific to the year of spawning, and hence cohort-specific.
4. The new, age 0 cohort at each population is then the mixture of offspring surviving the pelagic stage to settlement at the destination population.
5. Adult stage mortality is negligible.

The demographic process is modeled as a deterministic or a stochastic process described by two matrices $\boldsymbol{D}$, and $\mathbf{X}_{\boldsymbol{r}}^{(\boldsymbol{t})}$. The $\boldsymbol{D}$ matrix is a K*K matrix representing dispersal probabilities amongst all *K* populations. The dispersal $\boldsymbol{D}$ matrix remains fixed throughout the simulation. The $\mathbf{X}_{\boldsymbol{r}}^{(\boldsymbol{t})}$matrix is a *K*L* matrix of *K* populations and *L* allele frequencies at time t, where t denotes the spawning adults and t+1 are the larval offspring during dispersal in a given year. There are *multiple* $\mathbf{X}_{\boldsymbol{r}}^{(\boldsymbol{t})}$matrices, one for each of the *R* cohorts and only the youngest cohort is updated at each time step, while the other cohorts are incremented by each time step until they are dropped when they "age out" of the model.

Consider each population has a dispersal parameter vector δ_k which is the probability of dispersal into each of the *K* populations, including itself so that:

$$\sum_{k=1}^{K} \delta_{i \to k} = 1$$

And the full dispersal matrix is then:

$$\boldsymbol{D} = \begin{bmatrix} \delta_{1\to1} & \cdots & \delta_{1\to K} \\ \vdots & \ddots & \vdots \\ \delta_{K\to1} & \cdots & \delta_{K\to K} \end{bmatrix}$$

Therefore the allele contribution to the next cohort (*t+1*) in the *kth* target population is the sum of the larval allele contributions from all populations (*i=1, 2, 3,.., K*), so that the resulting allele frequency $x_k^{(t+1)}$ may be written as:

$$f\left(x_k^{(t+1)} | x_1^{(t)}, x_2^{(t)}, \cdots x_K^{(t)}, \delta_{1\to k}, \delta_{2\to k}, \cdots, \delta_{K\to k}\right) = \sum_{i=1}^{K} x_i^{(t+1)} \frac{\delta_{i\to k}}{\sum_{i=1}^{K} \delta_{i\to k}} \quad (1)$$

The function above describes the probability of the *x* allele being introduced to the *kth* population from all populations, in a given year and is assumed to remain unchanged in that cohort until they age out of the model.

When multiple cohorts form a spawning aggregate, the frequency of the *x* allele in the spawning aggregate population i at time *t*, is therefore the joint contribution of allele frequencies of the mature cohorts in the spawning group *R*, where *R* represents the spawning age cohorts:

$$x_i^{(t)} = \frac{1}{R} \sum_{r=1}^{R} x_{i,r}^{(t)} \quad (2)$$

$x_i^{(t)}$ is then the allele frequency in the spawning group from the *ith* population. Mutation is introduced through mutation parameter μ, to the allele frequencies in the spawning aggregate:

$$x_i^{(t+1)} = x_i^{(t)}(1-\mu) + \left(1 - x_i^{(t)}\right)\mu$$

The selection parameter $s^{(t+1)}$ is assumed to be specific to the age 0 cohort, mimicking the strong selection experienced by each cohort during their pelagic larval duration stage regardless of source population since they are expected to experience similar pelagic ocean conditions. This models the life history bottleneck that is characteristic of these pelagic larval dispersal marine species. The selection parameter is year specific, and applied equally to all loci. In a given year, it is drawn from a uniform distribution [-s, s]. The allele frequency may then be written as: $x_i^{(t+1)s}$, so that:

$$x_i^{(t+1)s} = \frac{\left(1 + s^{(t+1)}\right) x_i^{(t+1)}}{(1 + s^{(t+1)}) x_i^{(t+1)} + (1 - s^{(t+1)})\left(1 - x_i^{(t+1)}\right)} \quad (3)$$

The above formulation with *s* bounded by [-1, 1] allows for very strong selection of the $x_i^{(t+1)}$ allele at time *t+1* when $\left[0 \ll s^{(t+1)} \leq 1\right]$ and conversely a strong selection of the alternate $\left(1 - x_i^{(t+1)}\right)$ allele $\left[-1 \leq s^{(t+1)} \ll 0\right]$. For example when $s^{(t+1)} = 1$, the alternate allele is assumed to be lethal resulting in $x_i^{(t+1)s} = 1$, but when $s^{(t+1)} = -1$, the "main" allele is lethal and $x_i^{(t+1)s} = 0$. When $s^{(t+1)} = 0$, there is no selection at that loci and selection parameter does not change the allele frequencies. Note that the selection is assumed to act only on the larval aggregates in a given year, and not other cohorts (i.e. spawning aggregates). Therefore, the allele frequencies in each cohort, regardless of age remain a function of the selection parameter during their first year of life, or larval stage. The larval allele frequency matrix of all the larvae leaving a given population is then denoted as: $\chi^{(t+1)}$. Stochasticity occurs at two stages in the model, first, when the parents at each population are chosen to represent the spawning cohort aggregate such that:

$x_{i,r}^{(t+1)} = \frac{a_{i,r}^{(t+1)}}{2N}$, and $a_{i,r}^{(t+1)} \sim Binom(2N, x_{i,r}^{(t)})$ where $x_{i,r}^{(t)}$ is the allele frequency in equation 2, or the allele frequencies in the in the i^{th} population spawning aggregate.

Secondly, a random realization of the fraction of the larval aggregates from each of the K populations arriving at the given destination population (*k*) with probability, $\delta_{i \to k}$:

$$\tau_i^{(t+1)} \sim multinomial(N, \delta_{i \to (1:K)})$$

Where *N* is the larval cohort size, $\tau_i^{(t+1)}$ is a *K* length row vector of the number of individuals from *ith* to all *K* populations, $\delta_{i \to (1:K)}$ is the row probability vector of dispersal, and $\tau_i^{(t+1)}$ is a vector of the number of larvae where $\sum_{k=1}^{K} \tau_{i \to k}^{(t+1)} = N$. This results in a *K*K* matrix $\mathbf{T}^{(t+1)}$ representing the number of larvae dispersing among all the populations. Where the $\boldsymbol{D}$ matrix is row stochastic, the $\mathbf{T}$ matrix is column stochastic, and therefore has to be column normalized so that proportional allocation from all *K* populations to *kth* population sums to 1, and written as $\mathbf{T}_{\boldsymbol{u}}^{(t+1)}$.

The *K*L* allele frequency matrix of the age 0 cohort at each of the *K* populations is then the matrix product of the transpose of $\mathbf{T}_{\boldsymbol{u}}^{(t+1)}$ and $\boldsymbol{\chi}^{(t+1)}$:

$$\mathbf{X}_{\boldsymbol{c=0}}^{(t+1)} = \left(\mathbf{T}_{\boldsymbol{u}}^{(t+1)}\right)^{\boldsymbol{T}} \boldsymbol{\chi}^{(t+1)} \tag{4}$$

$$f\left(x_k^{(t+1)}\right) = \sum_{i=1}^{K} x_{i\to k}^{(t+1)s} \frac{\tau_{i\to k}^{(t+1)}}{\sum_{i=1}^{K} \tau_{i\to k}^{(t+1)}} \tag{5}$$

Function $f\left(x_k^{(t+1)}\right)$ is the transition probability of allele frequencies to the next cohort at *kth* population location and can be modeled in a deterministic or stochastic process of species with life history strategy described by early life stage dispersal. See algorithm 1 for full model implementation.

G_{ST} Population Differentiation

The fixation index, commonly referred to as F_{ST} is used to describe the genetic differentiation, especially due to population structure (S. Wright, 1951). Here I use the G_{ST} formulation based on gene diversity between two population which is an F_{ST} analog more readily adapted for multiple biallelic loci which is becoming the most common type of molecular data (Jakobsson et al., 2013b; Meirmans & Hedrick, 2011; Nei, 1973).

The average G_{ST} between two populations (i, k) averaged across L loci may then be calculated as follows:

$$\bar{H}_{S_{i,k}} = 1 - \frac{1}{2L}\sum_{l=1}^{L}(x_{il}^2 + x_{kl}^2)$$

$$\bar{H}_{T_{i,k}} = \frac{1}{L}\sum_{l=1}^{L}\left(1 - \left(\frac{x_{il} + x_{kl}}{2}\right)^2\right)$$

$$G_{ST_{i,k}} = \frac{\bar{H}_{T_{i,k}} - \bar{H}_{S_{i,k}}}{\bar{H}_{T_{i,k}}} \tag{6}$$

Here I use a corrected version of the standardized Hedrick's $G"_{ST}$ (Hedrick, 2006), since G_{ST} has been shown to produce biased results especially in pairwise population comparisons as would be the case when applying this method to sampled populations from larger, unknown metapopulations (Meirmans & Hedrick, 2011):

$$G"_{ST_{i,k}} = \frac{2\left(\bar{H}_{T_{i,k}} - \bar{H}_{S_{i,k}}\right)}{\left(2\bar{H}_{T_{i,k}} - \bar{H}_{S_{i,k}}\right)\left(1 - \bar{H}_{S_{i,k}}\right)} \tag{7}$$

The choice of the initial allele frequency generating function may influence the scale of the pairwise $G"_{ST}$. For example, the choice of Beta(1, 1) or uniform on (0, 1), resulted in maximal $G"_{ST} \approx 0.33$, while Beta(0.5, 0.5), maximal $G"_{ST} \approx 0.46$. $G"_{ST}$ has been originally proposed (Hedrick, 2006; Meirmans & Hedrick, 2011) as an alternate to F_{ST} (S. Wright,

1951) and G_{ST} (Nei, 1973) to measure population genetic differentiation because they noted that their maximum value was dependent on within population heterozygosity, H_S. However, (Alcala & Rosenberg, 2019; Jakobsson et al., 2013a) subsequently show that the $G"_{ST}$ likewise is influenced by the frequency of the most frequent allele and derive the relationship between the number of subpopulations, the major allele frequency and the maximum $G"_{ST}$. Therefore when calculating $G"_{ST}$, and converting the allele frequencies to major allele frequencies ($0.5 < x_i < 1$), the maximal value of $G"_{ST}$ would be smaller than when using minor allele frequencies ($0 < x_i < 0.5$). In this simulations, I do not convert the allele frequencies and use the raw frequencies from the simulation with initial frequencies generated from Jeffrey's prior (Beta(0.5, 0.5).

Population Structure and Larval Dispersal Models

The larval dispersal model (Figure 2-1) allows for various population structure models by changing the dispersal from *ith* to *kth* population, or $\delta_{i\to k}$ parameters. For example, ***random dispersal*** occurs when $\forall\delta_{i\to k} \propto \frac{1}{K}$; ***stepping stone***, $\delta_{i\to i+1} \neq 0, \forall\delta_{i\to k} = 0$. Therefore by varying the dispersal parameters, different population structure due to larval dispersal may be modeled (Figure 2-1 lower panel). Note for ***random dispersal***, for each generation, the dispersal parameters ($\delta_{i\to k}$) were selected at random from a uniform distribution (0, 1), which resulted in the overall average $\delta_{i\to k} \propto \frac{1}{K}$. Because the dispersal into a given population may be greater than dispersal out of a population, the total proportion of larvae entering a population may exceed 1.0, and the combined contribution of larvae into a given population includes donors from multiple populations which are then rescaled by the population size to calculate the allele frequencies in the new cohort as described above.

The simplest and probably the most well studied migration model is the ***isolation by distance***, or *IBD* first proposed by Sewall Wright (1946, 1943). Here, I employ a ***symmetrical IBD***, where the larvae disperse amongst populations with the dispersal being proportional to distance between them, in all directions.

The ***stepping stone*** migration model is another well studied variation on the ***IBD*** model where migrants only move between adjacent populations (Kimura & Weiss, 1964). The gene flow may be linear or two dimensional as well as unidirectional or bidirectional with the rate of migration being equal between the adjacent populations. Here, I use the

unidirectional case of dispersal between source population and the adjacent "downstream" population, i.e. $\delta_{i\to i+1} > 0, \delta_{i\to k} \cong 0, \forall(k \gg i)$.

The ***circular stepping stone*** model is a variation on the ***stepping stone*** gene flow in that the dispersal from the last population enter the first population. In nature, we may encounter this type of dispersal in shoreline habitat organisms residing around a lake or an island.

In the ***isolation barrier*** dispersal model, there exists a gene flow barrier with two independently evolving groups of populations with no interaction between the two sets. However, the populations in each group also undergo an ***isolation by distance*** dispersal as described above.

Source island refugia is a case where a single population is isolated from inflow of migrants, but is able to contribute migrants to other populations. This may occur in riverine systems where a single population is isolated by a barrier (i.e. waterfalls, or fast moving water sections) and once swept are unable to return. In contrast, ***sink island refugia*** may be thought of as the opposite in that a single population receives dispersed larvae into, but no larvae leave (aka the hotel California effect). Additionally, in both models, the non-refugia populations are totally isolated among themselves with no gene flow except into the refugia population (***sink island refugia)*** or out of from a single ***source island refugia*** population**.** Lastly, ***random dispersal*** *and* ***total isolation*** may be considered as the null models to be tested against. In ***random dispersal*** the probability of dispersal among populations is chosen at random during each time step, including remaining in the source population. In contrast, ***total isolation*** is characterized by no dispersal, where all progeny remains at the source population.

Parameter Input Values for Larval Dispersal Models

The choice of population size (*N*) is of particular importance since both affect the scale of differentiation and eventual decay under different dispersal scenarios. In order to create a candidate dispersal model generating $G"_{ST}$ values on the same scale as observed, the choice of population size (*N*) and migration rate (*m*) is of particular importance. Slatkin, (1993) describes the relationship of gene flow (*m*) and population size (*N*) as a function of differentiation between two island model populations at equilibrium $N_e m \approx \frac{1}{4}(\frac{1}{G_{ST}} - 1)$, where *Nm* is the number of migrants at each generation step. Here I used this relationship to

describe the expected $G"_{ST}$ distributions for populations under various two dimensional dispersal models. Rearranging the above equation, $G"_{ST} \approx \frac{1}{1+4N_e m}$, which is then used to inform the choice of the number of migrants per generation ($N_e m$) given a desired mean $G"_{ST}$ which is based on the observed population pairwise $G"_{ST}$ to be tested. Waples (Waples & Gaggiotti, 2006) in their simulation shown that migration rate of 5 migrants per generation (Nm=5) resulted in consistent population assignment inference when using STRUCTURE (Pritchard et al., 2000) to test for population differentiation. In this simulation therefore the migration rate per generation entering any population (m) was set at 0.005 and an effective population size (N_e) of 1,000 individuals each.

The other critical parameter is the number of time steps required for the simulation. Depending on the dispersal model, the transitions among all populations require different number of generations until mean pairwise $G"_{ST}$ and variance become stable. This weak stationarity of the dispersal model assures that a dispersal equilibrium among all pairwise populations has been sufficiently reached. I use the one sample slope test (Warton et al., 2012) of the mean and variance of pairwise $G"_{ST}$ over the previous 100 generations, and once the slope is no longer significantly different from 0, I determine that asymptotic behavior has been achieved. This ensured that no further simulations would result in change in parameter estimates.

All other parameters (i.e. spawning age, number of cohorts, number of populations) can be set to match the life history of the species. Here I initially test the dispersal models with no age structuring, where all parents are removed after spawning resulting in discrete, non-overlapping generations. I then introduce age structuring, with each population composed of 10 cohorts aged 1:10 which randomly contribute to the spawning aggregate and next generation of offspring. The number of time steps and generations under these models are exchangeable and equal to 1 since the offspring at each time step enter the breeding cohort at the next time step. Note that time to stationarity is recalculated for each change in age structuring. See algorithm 1 for model implementation.

All coding, simulation and analysis was carried out in *R* (R Core Team, 2021). STRUCTURE analysis was done using an improved and faster algorithm in *LEA* (Frichot & Francois, 2015), and graph aesthetics employed the packages *ggplot2* (Wickham, 2016), *gridExtra* (Auguie, 2017), *lemon* (Edwards, 2020), *RColorBrewer* (Neuwirth, 2014)*, qgraph*

(Epskamp et al., 2012), *scales* (Wickham & Seidel, 2020), *ggplotify* (Yu, 2021), and *cowplot* (Wilke, 2020).

Observed G''_{ST} and Null Dispersal Models

Given a vector of observed pairwise G''_{ST} values between populations, each of the candidate dispersal models with their corresponding α and β parameters of the Beta distribution may be evaluated how well each explains the observed data. Here, I use log likelihood approach as the basis of model fit which is proportional to the probability that the observed data was generated by the candidate model (i.e. $p(D|dispersal\ model = k)$). Consider a vector X of observed pairwise G''_{ST} values for n populations, $X = \{G_{ST_{1,2}}, G_{ST_{1,3}}, \cdots G_{ST_{n-1,n}}\}$ and K candidate dispersal models: $M_k \sim Beta(\alpha_k, \beta_k)$. The likelihood is:

$$\mathcal{L}_{\mathrm{k}} = \prod_{X} \frac{x^{\alpha_k - 1}(1-x)^{\beta_k - 1}}{B(\alpha_k, \beta_k)} \tag{8}$$

$$\log(\mathcal{L}_{\mathrm{k}}) = -\mathrm{n}\log(B(\alpha_k, \beta_k)) + \left(\sum_{i=1}^{n} (\alpha_k - 1) log x_i + (\beta_k - 1)\log(1 - x_i)\right) \tag{9}$$

Because the α, β parameters may take on very large values due to small G''_{ST}, and necessitates the use of Stirling's formula, the asymptotic approximation of the beta function for large α and β parameters:

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \frac{(\alpha - 1)!\,(\beta - 1)!}{(\alpha + \beta - 1)!} \approx \sqrt{2\pi}\,\frac{\alpha^{\alpha - \frac{1}{2}}\beta^{\beta - \frac{1}{2}}}{(\alpha + \beta)^{\alpha + \beta - \frac{1}{2}}} \tag{10}$$

$$\begin{aligned} \log\big(\mathrm{B}(\alpha, \beta)\big) \approx\ & log\sqrt{2\pi} + \left(\alpha - \frac{1}{2}\right) log\alpha + \left(\beta - \frac{1}{2}\right) log\beta \\ & - \left(\alpha + \beta - \frac{1}{2}\right) log(\alpha + \beta) \end{aligned} \tag{11}$$

Dispersal Model Validation

Candidate larval dispersal models were generated for various dispersal scenarios (Figures 2-1, 2-2). The Beta distribution parameters (α_k, β_k) were estimated (Appendix 1) from the mean and variance of the pairwise G''_{ST} values for each dispersal model once the simulation reached stationarity. The Beta distribution parameters were then estimated for each of the stationary dispersal models. Subsequently, an independent set of pairwise G''_{ST}

values was calculated for dispersal model. These were then treated as "observed" pairwise G''_{ST} values to test the model fit using the log likelihood approach described above.

Results

Time to Stationarity

The allele frequencies were initially generated at random for each population from Beta (0.5,0.5), with all age cohorts within a population having equal frequencies. The number of generations until stationarity was highly variable depending on the demographic model as well as population age structure and ranged from 100 to 1,900 for single cohort age structure, and 200 to 10,500 for 10 cohort age structured populations (Figure 2-2, Table 2-1). *Total isolation* required the fewest generations to reach stationarity with 100 and 200 for one and 10 cohort models respectively. This is expected, since only drift would result in a change in allele frequencies from the initial conditions. Similarly, *random dispersal* took 100 and 200 generations respectively to reach stationarity indicating speed with which populations mixed under this model with both age structures. *Stepping stone* model took the longest to reach stationarity with 1,900 and 10,500 generations for single and ten cohorts respectively. This indicated that the propagation of alleles via larval dispersal took the most generations to asymptote as the two farthest populations required the most steps between them. This is evident when examining the connectivity of each of the models displayed in figure 2-1.

Notable is the non-monotonic behavior of the average variance of pairwise G''_{ST} for both *stepping stone* and *circular stepping stone* models, which is unlike that of the other tested dispersal models. The behavior is characterized by initially rapid increase in the variance followed by a long decay as the variance reaches stationarity. These two dispersal models also are characterized by the longest number of generations until stationarity.

Population Structure Through Regression

The IBD regression plots (Figure 2-3) show how the population dispersal models as described by the relationship of the pairwise $\frac{G''_{ST}}{1-G''_{ST}}$ and the distance among the populations. I used unit distance, where the adjacent populations are one unit apart, and the first and 10th population are 10 units apart. *Random*, *total isolation*, *island refugia sink*, *island refugia source*, are indistinguishable from each other with an insignificant slope relationship. *Symmetrical IBD*, *stepping stone*, *circular stepping stone*, and *isolation barrier* models show

a positive relationship of genetic differentiation with distance. The *circular stepping stone* was the most distinct, with a semi-circular plot pattern which may possibly be used as a diagnostic for these type of dispersals. However, the *isolation barrier* model although statistically significant ($p<0.05$), is not indicative of a genetic differentiation with distance since it only captures the pairwise differentiation within each of the two population groups which are under IBD, and the differentiation between the two groups.

This demonstrates that the IBD regression plots may adequately discriminate population structure models with a distinct linear increase in genetic differentiation with distance. However, other population structure models are poorly differentiated, and may even result in false positive assertion of isolation by distance.

Population Structure Through PCA

PCA plots of 100 sampled individuals from each populations showed distinct population clustering for each of the dispersal models, except for *random dispersal*, as expected (Figure 2-4). However, without a priori population color coding, some of the groupings are less evident. For example, under *symmetrical IBD* and *stepping stone* models, the two farthest apart populations would be unidentifiable through PCA plots alone. As well, the population groupings in the *isolation barrier* model, with the left population aggregate undifferentiated, while the right side is composed of three populations. There is no difference in the model between the populations in the left and right group of the *isolation barrier* model. Likewise, two populations appear to overlap in the *total isolation* model, but this overlap is wholly arbitrary and not indicative of the model specification.

Age structure increased dispersion pattern in the PCA plots (Figure 2-4b), further increasing the inability to differentiate among distinct populations. However, except for the *random dispersal* model where the proportion of variance explained by each of the PCA axis is unchanged, all other models show a decrease. The *isolation barrier* model is now only composed of two identifiable population groups, as compared to four under a single cohort model, while the *circular stepping stone* model is sufficiently dispersed to misidentify any distinct groups. *Stepping stone* and *symmetrical IBD* are now also more diffused, leading to an apparent gradient in differentiation, with no distinct population groupings.

In contrast, *total isolation* and *island refugia sink* dispersal result in a more distinct population groupings with all ten populations readily identifiable under the 10 cohort model. However, only 8 population groups are identifiable in the no age structured models.

Population Structure Through STRUCTURE Analysis

STRUCTURE analysis for single cohort age structure shows good agreement between the modeled populations and the putative populations (Figure 2-5). Color designation of the STRUCTURE derived putative populations on PCA ordination plots (Figure 2-5a) shows the corresponding distinct groups or ambiguity in differentiating the populations based on PCA and STRUCTURE admixture (Figure 2-5b). As expected, *random dispersal* model results in a single putative population, while *total isolation* is correctly specifying all ten populations. *Island refugia source* and *island refugia sink* dispersal models also were correctly specified using STRUCTURE, but with greater levels of admixture with the *Island refuge sink* model indicating greatest admixture in the sink population (Figure 2-5b). The *symmetrical IBD* and *circular stepping stone* models show large admixture in each of the putative populations, but still correctly identify the original modeled populations. However both *stepping stone* and *isolation barrier* models fail to discriminate among two adjacent populations, resulting in a misspecification of the number of original populations.

The addition of age structuring to the demographic process resulted in a greater dispersion in the PCA plots and poorer population discrimination of population admixture for some dispersal models when sampling all populations (Figure 2-6). The 10 cohort model inference of *random dispersal*, *total isolation*, *island refugia sink*, and *island refugia source* remained unchanged when compared to a single cohort models. *Circular steeping stone* and *symmetrical IBD* associated with only three out of the original ten population groups. All of the sampled individuals in the *circular stepping stone* were identified as highly admixed with no clear separation among them. The *symmetrical IBD* model had clear separation of the first and last three populations and strong admixture in the middle four populations (Figures 2-5, 2-6). *Stepping stone* model was identified as a single population, with some admixture in the individuals from the last modeled population. The samples from the *isolation barrier* were estimated as two separate populations with no admixture among them, but no population structure within each of the groups. There was no difference in the PCA patterns and

STRUCTURE based inference when sampling only a single cohort of out of the ten modeled (Figure 2-7).

Cohort Specific Selection

Cohort specific selection, mimicking the annual variability in selection experienced by numerous marine species during their pelagic larval stage resulted in PCA groupings by age under strong (s=0.5) selection (Figure 2-8), with the populations only identifiable for the *total isolation* model. Under total isolation, the first PCA axis, explaining 1.28% of variation separated the ten age cohorts, whereas the second axis, explaining 0.62% of variation separated the 10 populations. STRUCTURE analysis identified all ten populations correctly, but with an admixture of an 11^{th} "population" due to the strong cohort specific selection effect. All other dispersal models resulted in two or three putative "populations", but with the groups identified being admixtures of age cohorts. Sampling a single cohort from each population did not improve the population identification except for *total isolation* which resulted in correct identification of the ten populations without the admixture of 11^{th} (Figure 2-9). All other dispersal models resulted in a single cluster as identified by PCA and STRUCTURE. Decreasing selection strengths increased the resolution of differentiation among the populations and this was specific to the dispersal model, however, sampling a single cohort versus a mixture of all ten did not improve population separation.

Beta Distribution Parameters Describe Population Structure

The estimated parameters of the beta distribution for each of the dispersal models (Figure 2-1) were unique to each of the models (Figure 2-10, Tables 2-2, 2-3). For the models with no age structure, the mean α ranged from 1.97 under the *stepping stone* model to 657 for *total isolation*, while mean β ranged from 9.43 for the *isolation barrier* and 235,636 for the *random dispersal* model. As expected, the mean genetic differentiation, $G"_{ST}$ was smallest for *random dispersal* (2.32e-5) and largest for *total isolation* (0.472). *Random dispersal* model also had the smallest variance (1.16e-10), while *isolation barrier* had the largest variance (0.0125).

Similarly for the age structured models, the largest mean α for *total isolation* (674), while smallest for *stepping stone* model (0.68), while mean β ranged from 13.15 for the *isolation barrier* and 2.15e8 for the *random dispersal* model. Genetic differentiation, $G"_{ST}$ was smallest for random dispersal (2.23e-7) and largest for *total isolation* (0.444), while the

variance was again smallest for *random dispersal* (1.16e-15) and largest for *symmetrical IBD* (0.00145).

The Beta distribution parameter space is non-overlapping for all the candidate models, while the mean G''_{ST} has considerable overlap (Figure 2-10) indicating that using the beta distribution maximum likelihood would differentiate among candidate dispersal models. Cross validation of log likelihood of candidate dispersal models verified this with 100% correct assignments.

Discussion

Population structure, or a difference in allele frequencies between subpopulations, is of paramount importance in conservation and management biology. If a population is deemed to be panmictic, having no apparent genetic subdivision or genetic structure, and large effective population size, N_e, then it is less susceptible to genetic drift and able to maintain a large number of potentially adaptive alleles and have greater genetic diversity. In contrast, a population of equal size, but divided into subpopulations with limited gene flow among them, will be more susceptible to genetic drift, and have a smaller genetic diversity and hence more susceptible to perturbations of their adapted niche. This leads to important implications in conservation and management biology.

Conservation biologists will focus on the structured population because of their potential lack of resilience, whereas management biologists will be concerned with local extirpation due to overexploitation. However, determining what is a population and how to measure its structure is not yet completely resolved (Waples & Gaggiotti, 2006; Whitlock, 1999). In this study as well, even when controlling population parameters, discerning specific dispersal system underlying the population structure is still elusive.

The population structure under the various dispersal models is difficult to differentiate when displayed in the ubiquitously applied linearized G''_{ST} on distance regression (Figure 2-3) where a small change in dispersal model may be indistinguishable. This was especially apparent when comparing *stepping stone* vs. *symmetrical IBD*, or *source island refugia* vs. *isolation barrier*. The differences between the *stepping stone* vs. *symmetrical IBD* regression plots are subtle, yet the underlying dispersal mechanism is vastly different. The *stepping stone* model is simply a unidirectional dispersal to the adjacent population only, whereas the *symmetrical IBD* is bidirectional with dispersal strength

proportional to the distance from source population. The extirpation of a single population in the *stepping stone* model would eliminate gene flow between the two groups alienated population groups, whereas gene flow would only be marginally slowed under the *symmetrical IBD* model. As well, the *source island refugia* is characterized by a single population dispersing to all other populations equally, but not receiving any gene flow unto itself. In the *isolation barrier* dispersal model, however, there exist two population groups each under IBD, but with no mixing between them. The IBD regression test for significance would lead to opposite conclusions, however, with *isolation barrier* having significant slope, but not *island refugia source*. This demonstrates the limit of interpretability of the IBD regression framework as it has been recognized (Meirmans, 2012).

Principal components analysis (PCA) is a part of most applied population genetics studies and readily applied to SNP data where thousands of biallelic loci are reduced to a two dimensional plane. The interpretation of the resulting patterns, however, has been controversial and the subject of numerous studies (Elhaik, 2021; Gauch et al., 2019; McVean, 2009; Novembre & Stephens, 2008; Reich et al., 2008). As well, in this study, the dispersal models are not readily distinguishable, and some of the inherent PCA biases readily observable. For example, in figure 2-4a, the PCA projection of the *isolation barrier* model where five populations are isolated from each, but with equal dispersal among them, would lead to an erroneous conclusion that the populations in the left most cluster are less differentiated than the more distinct groups in the right hand cluster, but correctly display the two separated population clusters. Likewise *total isolation*, *island refugia sink* and *island refugia source* are indistinguishable in the PCA ordination plot, and would most likely be classified as isolated populations. These results show that PCA plots do not adequately distinguish among various population structure models, while generating compelling plots. In practice, especially when sampling mixed population aggregates of age structured species, the representative sample size of individuals from each population is oftentimes unknown. This exacerbates population unidentifiability problem and may lead to erroneous conclusions of no population structure.

STRUCTURE admixture analysis (Pritchard et al., 2000) is arguably the most applied tool to infer the number of populations represented in a mixture of individual genotypes and has been readily adapted since its introduction(Gilbert, 2016). Unlike PCA, STRUCTURE is

a model based approach that minimizes the Hardy-Weinberg and uses linkage equilibrium within arbitrary clusters thereby determining the number of clusters K, and assign samples to each cluster (Pearse & Crandall, 2004). PCA and STRUCTURE have become the most used unsupervised clustering methods to determine distinct population groups represented in a sample of unknown mixtures of genotypes (Gilbert, 2016; Lawson et al., 2018; Rosenberg et al., 2005). However, some researchers began to question how accurate are STRUCTURE based population structure inferences (Puechmaille, 2016).

Here I find that STRUCTURE performed well in discriminating the number of modeled populations while also providing unique insight into the dispersal dynamics. For example *island refugia sink* models the dispersal from neighboring populations into a single population, with no dispersal out, and this can be seen in the STRUCTURE plot with the high admixture in that population only (Figure 2-5b). However STRUCTURE failed to identify gene flow out of a single population in the i*sland refugia source* model and consequently this dispersal model was undistinguishable from *total isolation*. High degree of admixture was correctly identified in the models with high dispersal among all the populations

STRUCTURE had a tendency to underestimate the number of population clusters which was contrary to (Frantz et al., 2009) who found that it lead to overestimation of population structure with more putative populations than present in the sample. The only case where STRUCTURE identified these "phantom" populations was in the models with strong cohort specific selection (Figure 2-8d). It was surprising, however, that although PCA plots showed distinct genetic cluster groups corresponding to groups of individual age classes, these clusters were not identified as distinct groups in STRUCTURE. As well, in the *total isolation* model, the groups of cohorts and populations are clearly visible in the PCA plot (Figure 2-8a, b), but STRUCTURE largely ignores the effect of cohort specific selection, and correctly identifies all population clusters, while indicating the presence of an additional admixture group. This shows that STRUCTURE is robust in identifying population structure when sampling isolated populations with strong cohort specific selection.

Cohort specific selection may be common in r selected species with pre-adult stages experiencing highly variable environment leading to highly variable juvenile mortality, as is common in many marine fishes. I modeled this effect here, rather crudely, by applying a

randomly generated selection parameter (s), where $s_i \sim uniform(-s, s)$ on all alleles in the i^{th} generation. The resulting PCA and STRUCTURE patterns then represent the effects of analysis on selected loci only. When multiple cohorts are sampled, the selection effect overwhelms the dispersal effect with distinct PCA clusters, but no STRUCTURE clustering (Figure 2-8). When analyzing a single cohort, the cohort cluster effect disappears and single population is apparent (Figure 2-9). As well, comparing results using neutral loci only (Figure 2-6) and selected loci (Figure 2-8) may yield different, but informative results. As well, when strong cohort selection is present, analyzing multiple cohorts jointly, and separately may elucidate complex population dynamics and demography.

The distribution of pairwise G''_{ST} values was unique to each of the candidate models which motivates the idea of being able to test model fit of observed data and candidate dispersal models. The choice of a beta distribution seemed natural since like G''_{ST} shares the same support of values bounded by 0 and 1, but a normal distribution may be a suitable approximation as well. Here, cross validation results for the dispersal models resulted in 100% accurate assignment for both beta and normal distributions likelihood testing. However, I noted considerable multi-modality in the distribution of the observed values, and mixture distributions may be more suitable when discerning among models with more subtle differences than I test here.

The pattern in the beta distribution α, β parameter space (Figure 2-10) shows that the dispersal models may be an adequate statistic to summarize the dispersal models and testing framework for observed data. The shift in the distribution (Figure 2-10a, b) shows the effect of the age structure on the parameter estimates which shows how age structure affects population structure inference as assessed with PCA, STRUCTURE, and G''_{ST} statistics. The age structure must therefore be taken into account when designing a hypothesis testing framework or to infer the dispersal matrices that may have generated the observed data.

Incorporating Approximate Bayesian Computation (ABC) approaches (Beaumont et al., 2002), may allow the estimation of the dispersal matrices that are likely to have generated the observed G''_{ST} dataset. The $\hat{\alpha}$, and $\hat{\beta}$ parameters for the observed G''_{ST} dataset and simulated various population connectivity/dispersal matrices. The location of the $\hat{\alpha}$, and $\hat{\beta}$ parameters in the candidate models parameter space (i.e. Figure 2-10) may inform the initial candidate dispersal model for ABC simulation. The expected G''^{*}_{ST} dataset calculated from

the simulated dispersal matrices could then be generated with α^*, and β^* test parameters. The Euclidean distance from the $\hat{\alpha}$, and $\hat{\beta}$ parameters and the α^*, and β^* test parameters would then be minimized by adjusting the dispersal matrix parameters. Alternatively, the log likelihood may then be used as the test statistic to assess model fit as described here and assess the best model fit from a suite of algorithm generated candidate models.

In nature, species undergo much more complex demographic life histories than simulation possibilities and not all methods may be adequately informative. Further complicating the issue of discerning population structure through dispersal is that populations may not be in dispersal drift equilibrium because any change in demographic history will take many generations before it can be measured through F_{ST} analogues (Varvio et al., 1986; Whitlock & McCauley, 1999). Additionally, populations likely start from a panmictic state and then undergo drift, mutation, migration, and selection which leads to the observed genetic population structuring. In this simulation, I start with wholly differentiated populations, then apply drift, mutation, and dispersal which allowed the simulation to reach an equilibria at a much faster rate than if the simulation started with equal allele frequencies. And yet, it still took thousands of generations to reach equilibria (Figure 2-2), especially in the age structured populations. This underscores that these simulations are an idealized, simplistic models, but useful nonetheless to test various methodologies that are commonly employed for population structure inference in natural populations. As well, in this study, I demonstrate the necessity of not relying on a single method when determining population structure.

Algorithm 1

1:	Initialize population size, *N*; number of populations *K*; number of simulated generations *T*; set spawning cohort ages: $R \in (R_{min}: R_{max})$; set *K*L* allele frequencies matrices **X** specific to each population and equal among cohorts, $x_{kl} \sim beta(0.5,0.5)$;set generation *t*=0; select dispersal matrix, **D**.
2:	**for** each year step t in 1 to *T* **do**
3:	Create allele frequencies in the spawning aggregates $x_{k,\rho}$ for each of the *K* populations:
4:	If using selection, select $s^{(t)} \sim Uniform[-s, s]$
5:	**for** each population *k* in *K* **do**
6:	**for** each spawning age group $r \in (R_{min}: R_{max})$ **do**
7:	For each of *L* loci, draw number of alleles in the *r* spawning cohort group in population *k*: $a_{k,r,l} \sim Binom(2N, x_{k,r,l})$
8:	**end for**
9:	Calculate matrix *K*L* matrix $\boldsymbol{X_l}$, or the *L* allele frequencies of larvae released at each of the *K* spawning locations: $x_{k,l} = \frac{1}{2NR}\sum_{r=1}^{R} a_{k,r,l}$
10:	If including mutation rate, μ: $x_{k,l} = x_{k,l}(1-\mu) + (1 - x_{k,l})\mu$
11:	If including selection, *s*: $x_k^{(t)s} = \frac{(1+s^{(t)})x_k^{(t)}}{(1+s^{(t)})x_k^{(t)} + (1-s^{(t)})\left(1-x_k^{(t)}\right)}$
	end for
12:	Create **T** *K*K* matrix by selecting the number of larvae (cohort 0) from each *K* source populations entering each of the *K* destination populations based on the row probabilities in matrix $\boldsymbol{D}$, or $\mathbf{D}_k$: $\boldsymbol{\tau_k} \sim multinom(N, \delta_k)$
13:	Create allele frequencies in the age 0 cohort for each of the *L* loci and *K* populations, or *K*L* matrix $\boldsymbol{X_{c=0}}$ by matrix multiplication of transposed *K*K* matrix of the fraction of the dispersing larvae, $\mathbf{T_u}$ and the *K*L* larval allele frequency matrix $\boldsymbol{\chi^{(t)}}$ $\mathbf{X}_{\boldsymbol{c=0}}^{(\boldsymbol{t+1})} = \mathbf{T}_{\boldsymbol{u}}^{(\boldsymbol{t+1})^{\boldsymbol{T}}} \boldsymbol{\chi^{(t)}}$
14:	For all populations, drop oldest cohort age $c = C_{max}$ and increment the ages of cohorts $c \in (0: C_{max-1})$ by 1; reset $x_{k,\rho} = 0$.
15:	**end for**

Tables and Figures

Table 2-1 Estimated time to stationarity under the various dispersal models.

Dispersal model	single cohort	cohorts 1:10
Random	200	200
Island refugia sink	300	800
Total Isolation	100	200
Symmetrical IBD	900	4,900
Circular Stepping stone	1,500	8,800
Stepping stone	1,900	10,500
Island refugia source	1,400	8,000
Isolation Barrier	1,300	7,800

Table 2-2 Estimated mean, variance and the calculated beta distribution parameters (α, β) of the pairwise G''_{ST} estimates under various simulated population structure models for the single age demographic. The values in parentheses contain the standard deviation of the simulated values.

Population Structure	μ (sd)	ν (sd)	α (sd)	β (sd)
Circular Stepping stone	0.0937 (0.0024)	9.87e-04 (6.97e-5)	8.17 (0.28)	79.14 (4.24)
Island refugia sink	0.4397 (0.0032)	8.05e-03 (9.35e-5)	13.33 (0.08)	16.98 (0.27)
Island refugia source	0.2801 (0.0054)	4.45e-04 (6.35e-5)	131.98 (18.2)	339.51 (48.23)
Isolation Barrier	0.1892 (0.0047)	1.25e-02 (8.74e-4)	2.2 (0.08)	9.43 (0.59)
Random	2.32e-05 (3.30e-06)	1.16e-10 (5.34e-1)	5.34 (1.54)	235,636.36 (82,692.4)
Stepping stone	0.0775 (0.0031)	2.78e-03 (2.82e-4)	1.97 (0.09)	23.48 (1.72)
Symmetrical IBD	0.1267 (0.0029)	3.89e-03 (3.05e-4)	3.58 (0.18)	24.69 (1.67)
Total Isolation	0.4723 (0.0034)	1.94e-04 (4.69e-5)	657.66 (159.23)	735.09 (179.8)

Table 2-3. Estimated mean, variance and the calculated beta distribution parameters (α, β) of the pairwise G''_{ST} estimates under various simulated population structure models for the age structure (1:10) demographic. The values in parentheses contain the standard deviation of the simulated values.

Population Structure	μ (sd)	ν (sd)	α (sd)	β (sd)
Circular stepping stone	0.0087 (0.0003)	2.17e-5 (1.56e-6)	3.54 (0.03)	403.36 (16.18)
Island refugia sink	0.4036 (0.0037)	7.25e-3 (1.01e-4)	13.31 (0.06)	19.67 (0.34)
Island refugia source	0.1145 (0.0017)	2.53e-5 (5.83e-6)	496.05 (121.2)	3,834.22 (930.39)
Isolation barrier	0.0837 (0.003)	5.12e-3 (3.97e-4)	1.2 (0.01)	13.15 (0.66)
Random	2.23e-7 (1.14e-8)	1.16e-15 (3.68e-16)	47.79 (15.01)	2.15e+08 (6.63e+8)
Stepping stone	0.0152 (0.0005)	3.34e-4 (2.17e-5)	0.68 (0.01)	44.18 (1.66)
Symmetrical IBD	0.0477 (0.0016)	1.45e-3 (1.21e-4)	1.49 (0.04)	29.74 (1.69)
Total Isolation	0.444 (0.0037)	1.75e-4 (4.12e-5)	674.14 (151.67)	843.97 (189.56)

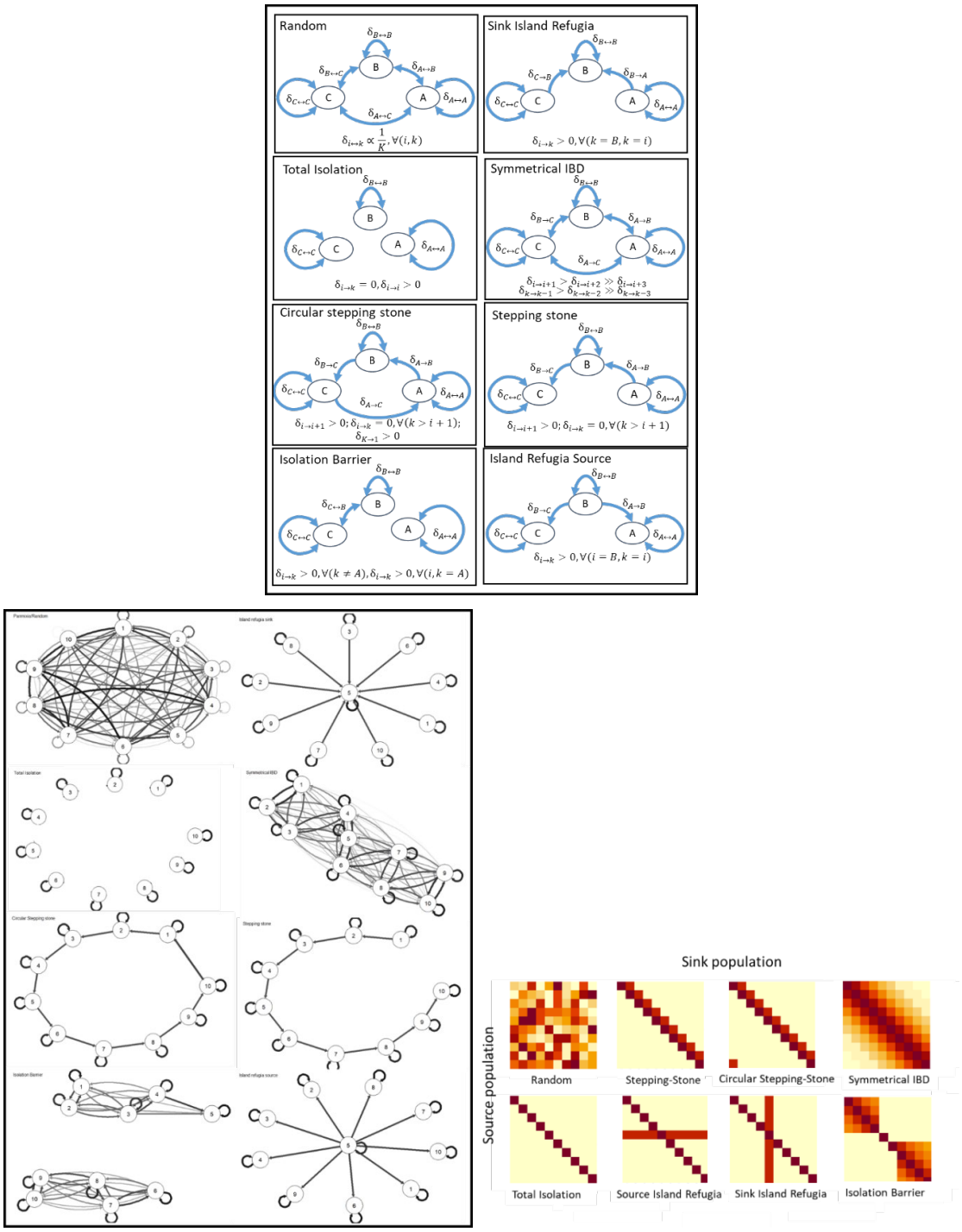


Figure 2-1 . Theoretical population dispersal models. The arrows indicate the direction of the dispersal. The top panel shows the theoretical specification of the dispersal models with the $\delta_{i\to j}$ parameters indicating the magnitude of the pairwise dispersal between the *ith* and *jth* population. The bottom panels are the directed graph diagram of the dispersal connectivity matrices and the heat map showing the magnitude of dispersal connectivity between the source and sink populations.

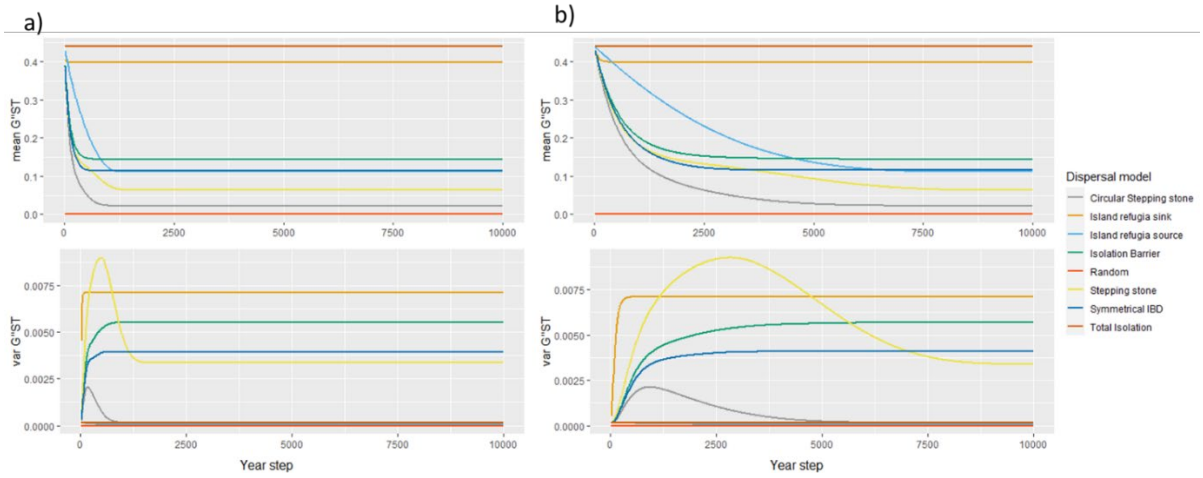


Figure 2-2 Time to stationarity plots for the mean and variance of pairwise G''_{ST} of the dispersal models. Stationarity was determined by the outcome of a slope test with no significant slope detected. a) left panel are the results with no age structure; b) right panel are populations with 1 to 10 aged cohorts interbreeding. The results of the slope test are in table 2-1.

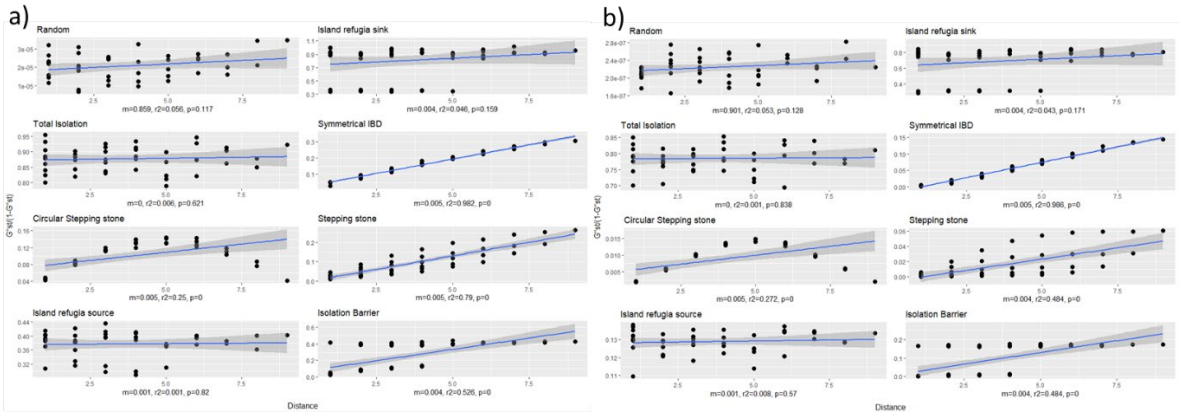


Figure 2-3 IBD regression plots of the linearized $G''_{ST}/(1-G''_{ST})$ on unit distance. The mean dispersal (m), R^2 and the slope p-value for each model are displayed below each corresponding graph. In panel a) the models consist of a single cohort, while in panel b) the populations are age structured with cohorts 1:10 interbreeding.

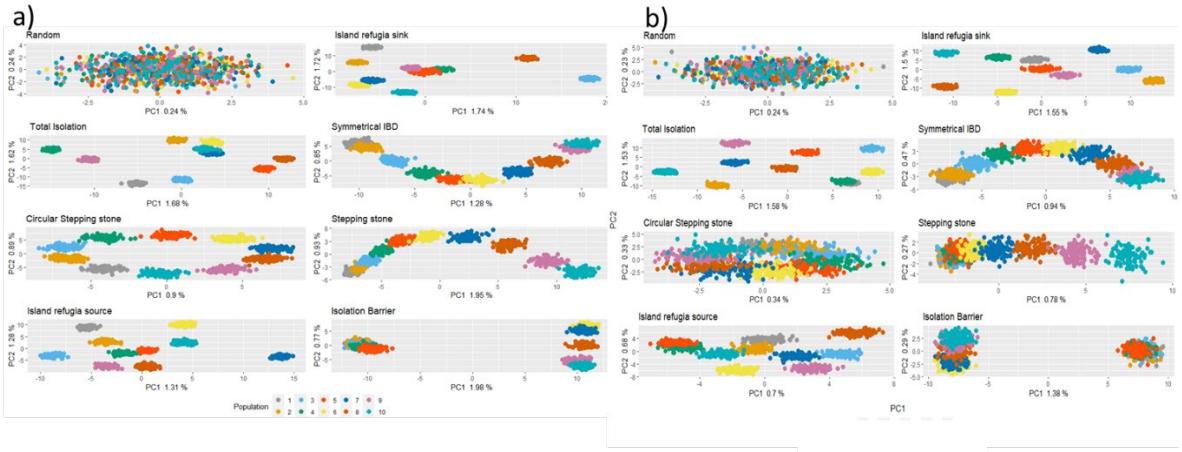


Figure 2-4 PCA results sampling 100 individuals from each population with the discrete populations colored separately. Panel a) indicates the results of a single cohort age structure; b) indicates ages 1:10 with 10 individuals sampled from each cohort equating to 100 total per population. The model parameters were: N_e=1,000, loci=1,000, N_em=5, n=100.

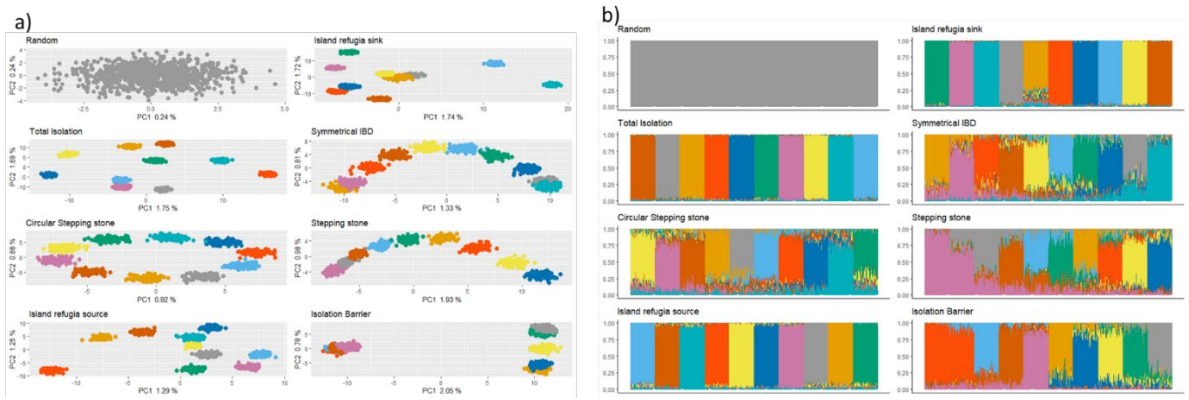


Figure 2-5 STRUCTURE results with best number of putative populations selected by STRUCTURE indicated by separate colors for each of the dispersal models with no age structure and 100 samples per population. a) The PCA plot with individual samples colored corresponding to the STRUCTURE population assignments; b) The STRUCTURE based admixture results. The model parameters were: N_e=1,000, loci=1,000, N_e m=5, ages=1:1, samples=100.

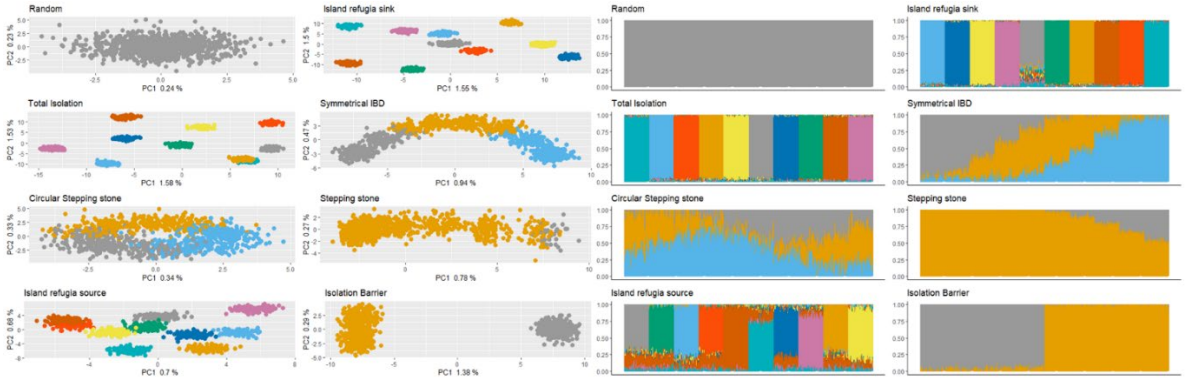


Figure 2-6 PCA and STRUCTURE results for dispersal models with age structures (cohorts 1:10) with sampling all age classes (n=10) for 100 samples per populations.

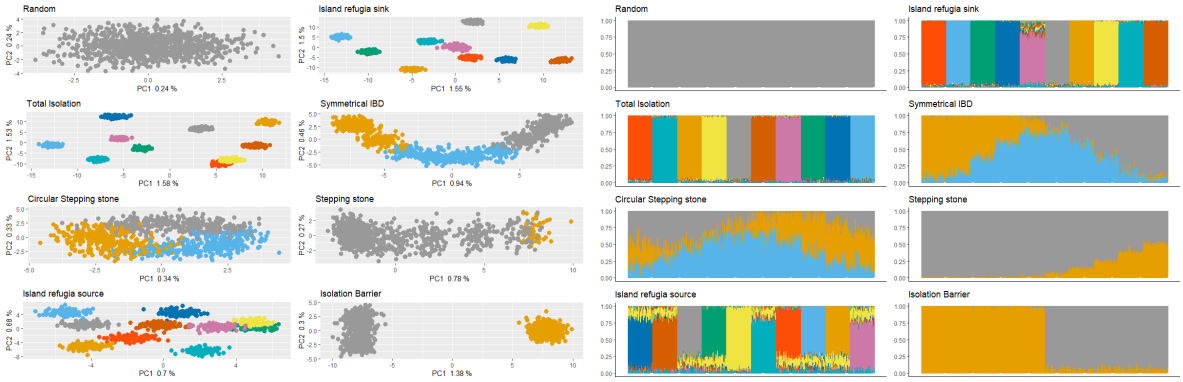


Figure 2-7 PCA and STRUCTURE results for dispersal models with age structure (cohorts 1:10), but sampling a single age (1) with 100 samples per population.

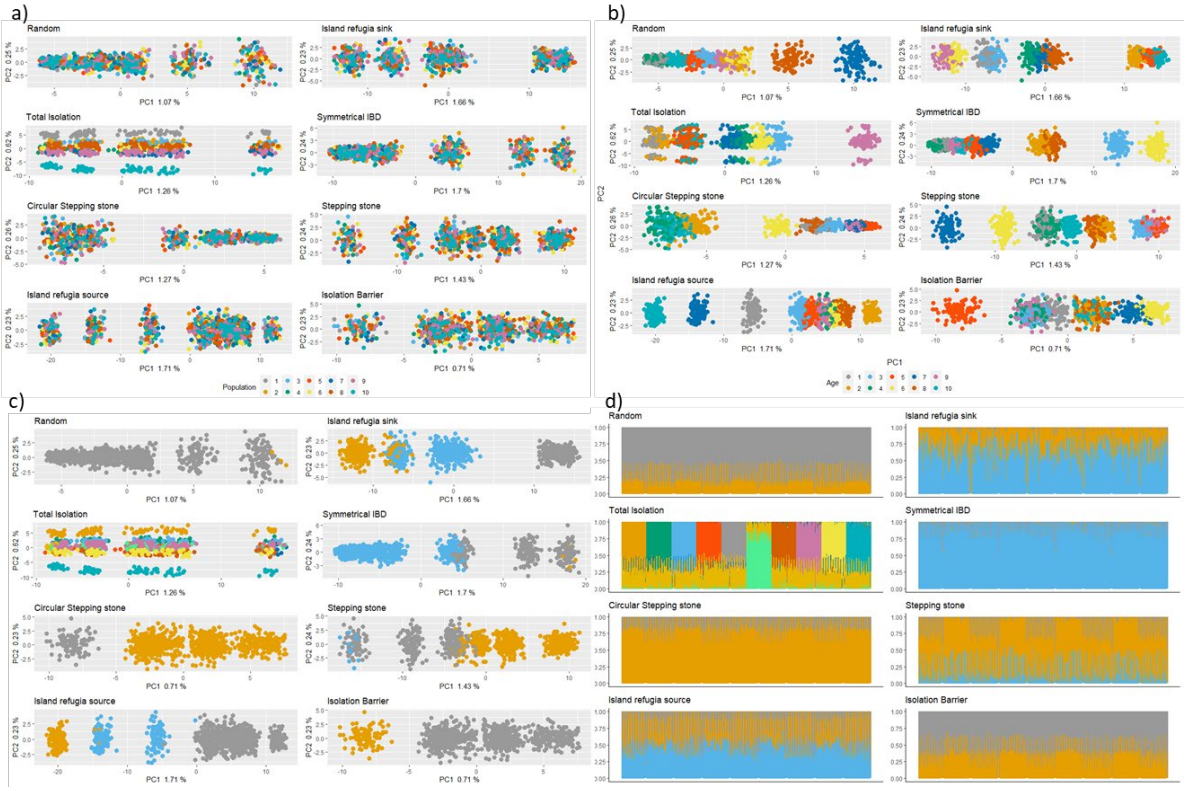


Figure 2-8 The confounding effect of strong cohort specific selection (s=0.5), demonstrate the lack of population discrimination power. Panel a) shows the PCA plot with colors denoting the populations, panel b) shows the PCA plot with the colors denoting the ten cohorts. Panel c) indicates the PCA plots with STRUCTURE derived putative populations denoted in panel d.

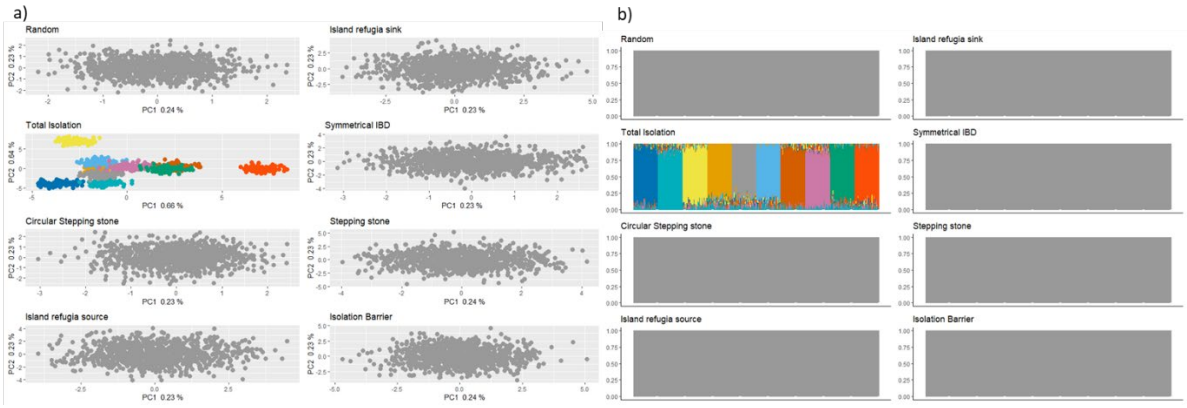


Figure 2-9 STRUCTURE admixture results under strong selection (s=0.5) and sampling a single cohort. Note that only *total isolation* dispersal model results in a differentiation of the 10 populations, with all other models only identifying a single population.

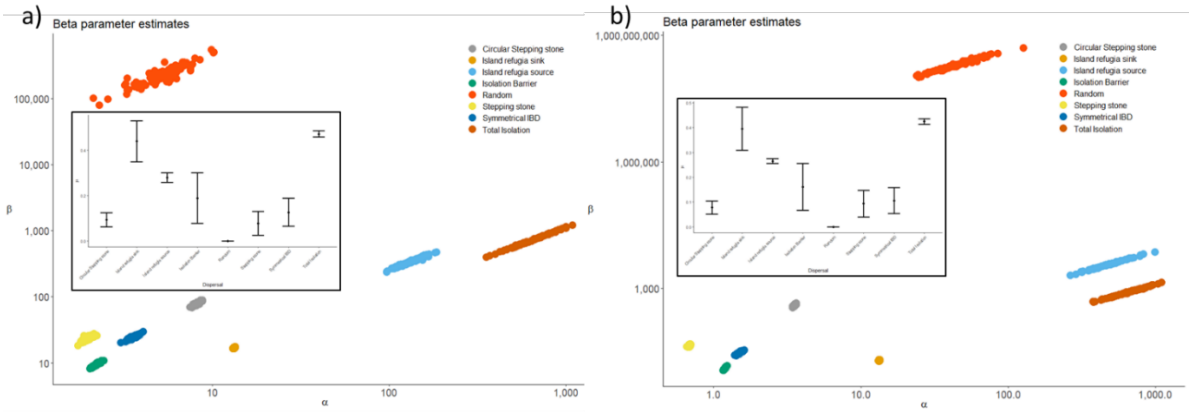


Figure 2-10 The Beta distribution parameter space of the pairwise G''_{ST} under various dispersion models using 100 simulations and 1,000 loci. Panel a) is the parameter space of the non-age structured population, while panel b) is the parameter space for the age structured (1:10) populations. Insets show the mean G''_{ST} and standard deviation for each of the dispersal model and the respective age structure. Note that both x and y axis are on a log scale.

References

Alcala, N., & Rosenberg, N. A. (2019). , Jost's D, and FST are similarly constrained by allele frequencies: A mathematical, simulation, and empirical study. *Molecular Ecology*, *28*(7), 1624–1636. https://doi.org/10.1111/MEC.15000

Arrowtooth Flounder | NOAA Fisheries. (n.d.). Retrieved March 5, 2022, from https://www.fisheries.noaa.gov/species/arrowtooth-flounder

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://cran.r-project.org/package=gridExtra

Balkenhol, N., Waits, L. P., Dezzani, R. J., Balkenhol, N., & Waits, L. P. (2009). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, *32*. https://doi.org/10.1111/j.1600-0587.2009.05807.x

Baltazar-Soares, M., Hinrichsen, H. H., & Eizaguirre, C. (2018). Integrating population genomics and biophysical models towards evolutionary-based fisheries management. *ICES Journal of Marine Science*, *75*(4), 1245–1257. https://doi.org/10.1093/ICESJMS/FSX244

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035. https://doi.org/10.1111/j.1937-2817.2010.tb01236.x

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., Rochette, R., & Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). *Molecular Ecology*, *25*(20), 5073–5092. https://doi.org/10.1111/mec.13811

Blackiston, D. J., Casey, S. S., & Weiss, E. R. (2008). Retention of Memory through Metamorphosis: Can a Moth Remember What It Learned As a Caterpillar? *PLoS ONE*, *3*(3), 1736. https://doi.org/10.1371/journal.pone.0001736

Blood, D. M., Matarese, A. C., & Busby, M. S. (2007). Spawning, egg development, and early life history dynamics of arrowtooth flounder (Atheresthes stomias) in the Gulf of Alaska. *NOAA Professional Paper NMFS 7*, *NOAA Profe*(May), 28.

Bonanomi, S., Overgaard Therkildsen, N., Retzel, A., Berg Hedeholm, R., Pedersen, M. W., Meldrup, D., Pampoulie, C., Hemmer-Hansen, J., Grønkjaer, P., & Nielsen, E. E. (2016). Historical DNA documents long-distance natal homing in marine fish. *Molecular Ecology*, *25*(12), 2727–2734. https://doi.org/10.1111/mec.13580

Bouwens, K. A., Smith, R. L., Paul, A. J., & Rugen, W. (1999). Length at and Timing of Hatching and Settlement for Arrowtooth Flounders in the Gulf of Alaska. *Alaska Fishery Research Bulletin*, *6*(1). http://www.state.ak.us/adfg/geninfo/pubs/afrb/afrbhome.htm

Bradburd, G. S., & Ralph, P. L. (2019). *Spatial Population Genetics: It's About Time*.

https://doi.org/10.1146/annurev-ecolsys-110316

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, *67*(11), 3258–3273. https://doi.org/10.1111/evo.12193

Bradbury, I., & Laurel, B. (2007). Defining "natal homing" in marine fish populations: comment on Svedäng et al. (2007). *Marine Ecology Progress Series*, *349*, 307–308. https://doi.org/10.3354/meps07281

Bradbury, I. R., & Bentzen, P. (2007). Non-linear genetic isolation by distance: implications for dispersal estimation in anadromousand marine fish populations. *Marine Ecology Progress Series*, *340*, 245–257. https://doi.org/10.3354/MEPS340245

Castellano, S., & Balletto, E. (2002). IS THE PARTIAL MANTEL TEST INADEQUATE? In *Evolution* (Vol. 56, Issue 9).

Clarke, K. Robert, Somerfield, P. J., & Gorley, R. N. (2008). Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *Journal of Experimental Marine Biology and Ecology*, *366*(1–2), 56–69. https://doi.org/10.1016/j.jembe.2008.07.009

Clarke, K R, & Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. In *MARINE ECOLOGY PROGRESS SERIES Mar. Ecol. Prog. Ser* (Vol. 92).

Conrath, C. L., & Knoth, B. (2013). Reproductive Biology of Pacific Ocean Perch in the Gulf of Alaska. *Https://Doi.Org/10.1080/19425120.2012.751941*, *5*(5), 21–27. https://doi.org/10.1080/19425120.2012.751941

Cunningham, K. M., Canino, M. F., Spies, I. B., & Hauser, L. (2009). Genetic isolation by distance and localized fjord population structure in Pacific cod (Gadus macrocephalus): Limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(1), 153–166. https://doi.org/10.1139/F08-199

Debenham, C., Moss, J., & Heintz, R. (2019). *Ecology of age-0 arrowtooth flounder (Atheresthes stomias) inhabiting the Gulf of Alaska.*

Diopere, E., Vandamme, S. G., Hablützel, P. I., Cariani, A., Van Houdt, J., Rijnsdorp, A., Tinti, F., Volckaert, F. A. M., & Maes, G. E. (2018). Seascape genetics of a flatfish reveals local selection under high levels of gene flow. *ICES Journal of Marine Science*, *75*(2), 675–689. https://doi.org/10.1093/icesjms/fsx160

Drinan, D. P., Gruenthal, K. M., Canino, M. F., Lowry, D., Fisher, M. C., & Hauser, L. (2018). Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (Gadus macrocephalus). *Evolutionary Applications*, *11*(8), 1448–1464. https://doi.org/10.1111/eva.12639

Duforet-Frebourg, N., & Slatkin, M. (2016). Isolation-by-distance-and-time in a stepping-stone model. *Theoretical Population Biology*, *108*, 24–35.

https://doi.org/10.1016/j.tpb.2015.11.003

Dunn, J. R., & Matarese, A. C. (1987). A review of the early life history of Northeast Pacific gadoid fishes. *Fisheries Research*, *5*(2–3), 163–184. https://doi.org/10.1016/0165-7836(87)90038-5

Edwards, S. M. (2020). *lemon: Freshing Up your "ggplot2" Plots*. https://cran.r-project.org/package=lemon

Elhaik, E. (2021). Why most Principal Component Analyses (PCA) in population genetic studies are wrong. *BioRxiv*, 2021.04.11.439381. https://doi.org/10.1101/2021.04.11.439381

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). {qgraph}: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*(4), 1–18.

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, *46*, 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Frichot, E., & Francois, O. (2015). {LEA}: an {R} package for {L}andscape and {E}cological {A}ssociation studies. *Methods in Ecology and Evolution*. http://membres-timc.imag.fr/Olivier.Francois/lea.html

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/J.TREE.2012.05.012

Gauch, H. G., Qian, S. I., Piepho, H.-P., Zhou, L., & Chen, R. (2019). *Consequences of PCA graphs, SNP codings, and PCA variants for elucidating population structure*. https://doi.org/10.1371/journal.pone.0218306

Gilbert, K. J. (2016). *Identifying the number of population clusters with STRUCTURE: problems and solutions*. https://doi.org/10.1111/1755-0998.12521

GitHub - wStockhausen/DisMELS: A Java-based framework for developing/running individual-based models(IBMs) for marine species with pelagic early life stages. (n.d.). Retrieved July 2, 2020, from https://github.com/wStockhausen/DisMELS

Guillot, G., & ßois Rousset, F. (2012). *Dismantling the Mantel tests*. https://doi.org/10.1111/2041-210x.12018

Harmon, L. J., & Glor, R. E. (2010). POOR STATISTICAL PERFORMANCE OF THE MANTEL TEST IN PHYLOGENETIC COMPARATIVE ANALYSES. *Evolution*, *64*(7), 2173–2178. https://doi.org/10.1111/j.1558-5646.2010.00973.x

Hedrick, P. W. (2006). Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 67–93. https://doi.org/10.1146/annurev.ecolsys.37.091305.110132

Hinckley, S., Stockhausen, W. T., Coyle, K. O., Laurel, B. J., Gibson, G. A., Parada, C., Hermann, A. J., Doyle, M. J., Hurst, T. P., Punt, A. E., & Ladd, C. (2019). Connectivity between spawning and nursery areas for Pacific cod (Gadus macrocephalus) in the Gulf of Alaska. *Deep Sea Research Part II: Topical Studies in Oceanography*, *165*, 113–126. https://doi.org/10.1016/j.dsr2.2019.05.007

Jakobsson, M., Edge, M. D., & Rosenberg, N. A. (2013a). *The Relationship Between F ST and the Frequency of the Most Frequent Allele*. https://doi.org/10.1534/genetics.112.144758

Jakobsson, M., Edge, M. D., & Rosenberg, N. A. (2013b). The relationship between FST and the frequency of the most frequent allele. *Genetics*, *193*(2), 515–528. https://doi.org/10.1534/genetics.112.144758

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., Knight, S., Knox, J., Long, T., Mannino, M., McGuire, M., Riffle, A., Segelsky, S., Shappell, L., Sterner, A., Strickler, T., & Tursi, R. (2010). A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, *33*(2), no-no. https://doi.org/10.1111/j.1600-0587.2010.06285.x

Kamin, L. M., Palof, K. J., Heifetz, J., & Gharrett, A. J. (2014). Interannual and spatial variation in the population genetic composition of young-of-the-year Pacific ocean perch (Sebastes alutus) in the Gulf of Alaska. *Fisheries Oceanography*, *23*(1), 1–17. https://doi.org/10.1111/fog.12038

Kenchington, E. L. (2003). The effects of fishing on species and genetic diversity. In *Responsible fisheries in the marine ecosystem* (pp. 235–253). CABI. https://doi.org/10.1079/9780851996332.0235

Kimura, M., & Weiss, G. H. (1964). THE STEPPING STONE MODEL OF POPULATION STRUCTURE AND THE DECREASE OF GENETIC CORRELATION WITH DISTANCE. *Genetics*, *49*(4), 561–576. https://doi.org/10.1093/genetics/49.4.561

Knutsen, H., Catarino, D., Rogers, L., Sodeland, M., Mattingsdal, M., Jahnke, M., Hutchings, J. A., Mellerud, I., Espeland, S. H., Johanneson, K., Roth, O., Hansen, M. M., Jentoft, S., André, C., & Jorde, P. E. (2022). Combining population genomics with demographic analyses highlights habitat patchiness and larval dispersal as determinants of connectivity in coastal fish species. *Molecular Ecology*. https://doi.org/10.1111/mec.16415

Laurel, B. J., Hurst, T. P., Copeman, L. A., & Davis, M. W. (2008). The role of temperature on the growth and survival of early and late hatching Pacific cod larvae (Gadus macrocephalus). *Journal of Plankton Research*, *30*(9), 1051–1060. https://doi.org/10.1093/plankt/fbn057

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1), 3258. https://doi.org/10.1038/s41467-018-05257-7

Legendre, P., & Casgrain, Ph. (1994). MODELING BRAIN EVOLUTION FROM BEHAVIOR: A PERMUTATIONAL REGRESSION APPROACH. In *Evolution* (Vol. 48, Issue 5).

Leis, J. M., Siebeck, U., & Dixson, D. L. (2011). How Nemo Finds Home: The Neuroecology of Dispersal and of Population Connectivity in Larvae of Marine Fishes. *Integrative and Comparative Biology*, *51*(5), 826–843. https://doi.org/10.1093/icb/icr004

Levin, L. A. (2006). Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, *46*(3), 282–297. https://doi.org/10.1093/icb/icj024

Manel, Stephanie, Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, *20*(3), 136–142. https://doi.org/10.1016/j.tree.2004.12.004

Manel, Stéphanie, & Holderegger, R. (2013). Ten years of landscape genetics. In *Trends in Ecology and Evolution* (Vol. 28, Issue 10, pp. 614–621). Elsevier Current Trends. https://doi.org/10.1016/j.tree.2013.05.012

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, *27*(2 Part 1).

Maselko, J., Andrews, K. R., & Hohenlohe, P. A. (2020). Long-lived marine species may be resilient to environmental variability through a temporal portfolio effect. *Ecology and Evolution*. https://doi.org/10.1002/ece3.6378

Maselko, Jacek, Andrews, K. R., & Hohenlohe, P. A. (2020). Long-lived marine species may be resilient to environmental variability through a temporal portfolio effect. *Ecology and Evolution*, *10*(13), 6435–6448. https://doi.org/10.1002/ece3.6378

McRae, B. H. (2006). ISOLATION BY RESISTANCE. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, *5*(10), 1000686. https://doi.org/10.1371/journal.pgen.1000686

Meirmans, P. G. (2012). The trouble with isolation by distance. In *Molecular Ecology* (Vol. 21, Issue 12, pp. 2839–2846). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. In *Molecular Ecology Resources* (Vol. 11, Issue 1, pp. 5–18). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12), 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. https://cran.r-

project.org/package=RColorBrewer

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*(5), 646–649. https://doi.org/10.1038/ng.139

Pacific Cod | NOAA Fisheries. (n.d.). Retrieved March 5, 2022, from https://www.fisheries.noaa.gov/species/pacific-cod

Pacific Ocean Perch | NOAA Fisheries. (n.d.). Retrieved March 5, 2022, from https://www.fisheries.noaa.gov/species/pacific-ocean-perch

Palof, K. J., Heifetz, J., & Gharrett, A. J. (2011). Geographic structure in Alaskan Pacific ocean perch (Sebastes alutus) indicates limited lifetime dispersal. *Marine Biology*, *158*(4), 779–792. https://doi.org/10.1007/s00227-010-1606-2

Pante, E., & Simon-Bouhet, B. (2013). *marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R*. https://doi.org/10.1371/journal.pone.0073051

Pearse, D. E., & Crandall, K. A. (2004). Beyond FST: Analysis of population genetic data for conservation. *Conservation Genetics*, *5*(5), 585–602. https://doi.org/10.1007/s10592-003-1863-4

Pritchard, J. K. (2010). *Software For Inferring Population Structure*. http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.4/html/structure.html

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). *Inference of Population Structure Using Multilocus Genotype Data*. http://www.stats.ox.ac.uk/pritch/home.html.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627. https://doi.org/10.1111/1755-0998.12512

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Raufaste, N., & Rousset, F. O. (2001). ARE PARTIAL MANTEL TESTS ADEQUATE? In *BRIEF COMMUNICATIONS Evolution* (Vol. 55, Issue 8).

Reich, D., Price, A. L., & Patterson, N. (2008). Principal component analysis of genetic data. *Nature Genetics 2008 40:5*, *40*(5), 491–492. https://doi.org/10.1038/ng0508-491

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005). Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure. *PLoS Genetics*, *1*(6), e70. https://doi.org/10.1371/journal.pgen.0010070

Rousset. (2002). PARTIAL MANTEL TESTS: REPLY TO CASTELLANO AND

BALLETTO. *Https://Doi.Org/10.1554/0014-3820(2002)056[1874:PMTRTC]2.0.CO;2*, *56*(9), 1874–1875. https://doi.org/10.1554/0014

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228. https://doi.org/10.1093/genetics/145.4.1219

Schiavina, M., Marino, I. A. M., Zane, L., & Melià, P. (2014). Matching oceanography and genetics at the basin scale. Seascape connectivity of the Mediterranean shore crab in the Adriatic Sea. *Molecular Ecology*, *23*(22), 5496–5507. https://doi.org/10.1111/mec.12956

Schroeder, I. D., Santora, J. A., Moore, A. M., Edwards, C. A., Fiechter, J., Hazen, E. L., Bograd, S. J., Field, J. C., & Wells, B. K. (2014). Application of a data-assimilative regional ocean modeling system for assessing California Current System ocean conditions, krill, and juvenile rockfish interannual variability. *Geophysical Research Letters*, *41*(16), 5942–5950. https://doi.org/10.1002/2014GL061045

Schunter, C., Carreras-Carbonell, J., MacPherson, E., TintorÉ, J., Vidal-Vijande, E., Pascual, A., Guidetti, P., & Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. https://doi.org/10.1111/j.1365-294X.2011.05355.x

Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., Von Der Heyden, S., & Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, *554*(July), 1–19. https://doi.org/10.3354/meps11792

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*(1), 1–15. https://doi.org/10.1111/evo.12258

Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, *9*(4), 347–404. https://doi.org/10.1016/j.ocemod.2004.08.002

Slatkin, M. (1980). THE DISTRIBUTION OF MUTANT ALLELES IN A SUBDIVIDED POPULATION. *Genetics*, *95*(2).

Slatkin, M. (1993). Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*. https://doi.org/10.2307/2410134

Slatkin, M., & Barton, N. H. (1989). *A Comparison of Three Indirect Methods for Estimating Average Levels of Gene Flow* (Vol. 43, Issue 7).

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology*, *35*(4), 627. https://doi.org/10.2307/2413122

Sokal, R. R. (1979). Testing Statistical Significance of Geographic Variation Patterns.

Systematic Zoology, *28*(2), 227. https://doi.org/10.2307/2412528

Spies, I., & Punt, A. E. (2015). The utility of genetics in marine fisheries management: A simulation study based on Pacific cod off Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(9), 1415–1432. https://doi.org/10.1139/CJFAS-2014-0050

Stephenson, R., Kenchington, E., Branch, S., & Canada, O. (2000). International Council for the Exploration of the Sea CM 2000/Mini:07 Defining the Role of ICES in Supporting Biodiversity Conservation. *ICES Journal of Marine Science*.

Stewart, J. R., & Lister, A. M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, *16*(11), 608–613. https://doi.org/10.1016/S0169-5347(01)02338-2

Stockhausen, W. (2007). *Modeling Larval Dispersion of Rockfish: A Tool for Marine Reserve Design?* https://doi.org/10.4027/bamnpr.2007.15

Stockhausen, W. T., Coyle, K. O., Hermann, A. J., Blood, D., Doyle, M. J., Gibson, G. A., Hinckley, S., Ladd, C., & Parada, C. (2019). Running the gauntlet: Connectivity between spawning and nursery areas for arrowtooth flounder (Atheresthes stomias) in the Gulf of Alaska, as inferred from a biophysical individual-based model. *Deep Sea Research Part II: Topical Studies in Oceanography*, *165*, 127–139. https://doi.org/10.1016/j.dsr2.2018.05.017

Stockhausen, W. T., Coyle, K. O., Hermann, A. J., Doyle, M., Gibson, G. A., Hinckley, S., Ladd, C., & Parada, C. (2019). Running the gauntlet: Connectivity between natal and nursery areas for Pacific ocean perch (Sebastes alutus) in the Gulf of Alaska, as inferred from a biophysical individual-based model. *Deep Sea Research Part II: Topical Studies in Oceanography*, *165*, 74–88. https://doi.org/10.1016/j.dsr2.2018.05.016

Storfer, A., Murphy, Melanie A. Spear, Stephen F. Holderegger, R., & Waits, L. P. (2010). Landscape genetics: where are we now? *Molecular Ecology*, *19*(17), 3496–3514. https://doi.org/10.1111/j.1365-294X.2010.04691.x

Sunday, J. M., Calosi, P., Dupont, S., Munday, P. L., Stillman, J. H., & Reusch, T. B. H. (2014). Evolution in an acidifying ocean. *Trends in Ecology & Evolution*, *29*(2). https://doi.org/10.1016/j.tree.2013.11.001

Svedäng, H., Righton, D., & Jonsson, P. (2007). Migratory behaviour of Atlantic cod Gadus morhua: natal homing is the prime stock-separating mechanism. *Marine Ecology Progress Series*, *345*, 1–12. https://doi.org/10.3354/MEPS07140

Thomson, J. A. (1962). On the Fecundity of Pacific Cod (Gadus macrocephalus Tilesius) from Hecate Strait, British Columbia. *Journal of the Fisheries Research Board of Canada*, *19*(3), 497–500. https://doi.org/10.1139/f62-028

Varvio, S. L., Chakraborty, R., & Nei, M. (1986). Genetic variation in subdivided populations and conservation genetics. *Heredity*, *57*(2), 189–198. https://doi.org/10.1038/hdy.1986.109

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Waples, R. S., & Gaggiotti, O. (2006). INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419–1439. https://doi.org/10.1111/J.1365-294X.2006.02890.X

Warton, D. I., Duursma, R. A., Falster, D. S., & Taskinen, S. (2012). SMATR 3-an R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution*, *3*, 257–259. https://doi.org/10.1111/j.2041-210X.2011.00153.x

Werner, F. E., Page, F. H., Lynch, D. R., Loder, J. W., Lough, R. G., Perry, R. I., Greenberg, D. A., & Sinclair, M. M. (1993). Influences of mean advection and simple behavior on the distribution of cod and haddock early life stages on Georges Bank. *Fisheries Oceanography*, *2*(2), 43–64. https://doi.org/10.1111/j.1365-2419.1993.tb00120.x

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration: FST≠1/(4Nm+1). *Heredity*, *82*(2), 117–125. https://doi.org/10.1038/sj.hdy.6884960

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

Wickham, H., & Seidel, D. (2020). *scales: Scale Functions for Visualization*. https://cran.r-project.org/package=scales

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."* https://cran.r-project.org/package=cowplot

Withler, R. E., Beacham, T. D., Schulze, A. D., Richards, L. J., & Miller, K. M. (2001). Co-existing populations of Pacific ocean perch, Sebastes alutus, in Queen Charlotte Sound, British Columbia. *Marine Biology*, *139*(1), 1–12. https://doi.org/10.1007/S002270100560

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*(4), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

WRIGHT, S. (1946). Isolation by distance under diverse systems of mating. *Genetics*, *31*(January), 39–59.

Wright, Sewall. (1943). ISOLATION BY DISTANCE*. *Genetics*, *28*, 114–138.

Yu, G. (2021). *ggplotify: Convert Plot to "grob" or "ggplot" Object*. https://cran.r-project.org/package=ggplotify

Chapter 3: Population Structure Through Biophysical Larval Dispersal Models

Abstract

Combining the biophysical dispersal models with a stochastic genetic population component allows for unique insights into the effect of larval dispersal on population structure. Here I demonstrate the application of the spatio-temporal genetic model developed and validated in chapter 2, to calculate the expected population structure for three marine fish species, Pacific ocean perch (*Sebastes alutus*), arrowtooth flounder (*Atheresthes stomias*), and Pacific cod (*Gadus macrocephalus)* in the Gulf of Alaska. I then use the resulting genetic differentiation among the biophysical model zones to demonstrate the utility of this methodology by comparing inference on population structure as evident from PCA, STRUCTURE and linearized F_{ST} on geographic distance regression. This is followed by the comparison of Pacific ocean perch (*Sebastes alutus*) and Pacific cod (*Gadus macrocephalus)* expected population structure and compared to the observed datasets which reconciles previously contradictory studies. I also demonstrate the application of this model to determine optimal sampling strategy. The results presented here also suggest that the biophysical based dispersal may be the primary driver behind the observed population structure in the marine species with life history strategies characterized by pelagic larval dispersal. I find a surprising concordance between the observed and modeled population structure based solely on the Gulf of Alaska biophysical larval dispersal model.

Introduction

Understanding population structure in managed fish stocks is of paramount importance for sustainable management. Managing a panmictic population as multiple stocks is inefficient, while managing multiple stocks as a single stock may lead to overexploitation and even extirpation (Kenchington, 2003). However, accounting for stock structure when managing fisheries may allow for higher exploitation rates (Spies & Punt, 2015). Historically, the predominant approach has been to manage fisheries despite oftentimes inadequate knowledge of the target species complex population structure. This lead to calls by some researchers for integration of the stock structure into management in order to

maintain sustainable fisheries and biodiversity (Kenchington, 2003; Stephenson et al., 2000). Today, this is even more imperative during the recent accelerated environmental change in the ocean and need to understand species and population resilience (Sunday et al., 2014).

Genetic approaches have been integral to elucidate stock structure since population structure was the subject of population genetic models since their inception (S. Wright, 1951; S. Wright, 1943). However, traditional approaches of using genetics to delineate stock structures may not adequately capture the species population structure characterized by complex gene flow such as pelagic stage dispersal in fish populations. The most often used method to examine whether an isolation by distance (IBD) population structure exists is through $F_{ST}/(1 - F_{ST})$ on distance regression (I. R. Bradbury & Bentzen, 2007; Rousset, 1997). Bradbury & Bentzen (2007) through their simulations of one-dimensional stepping stone dispersal simulations and meta-analysis of eighteen anadromous species, found that in fact the IBD patterns are non-linear, with slope decreasing with increasing dispersal distance. The lack of robustness in $F_{ST}/(1 - F_{ST})$ regression to reliably estimate population differentiation with distance as well as advances in increased computational power has led to inference based on a landscape/seascape genetics approaches (Balkenhol et al., 2009; Jenkins et al., 2010). These approaches have since become the standard methods estimate the relationships of population genetic differentiation, dispersal, and environmental and physical effects (Stephanie Manel et al., 2005; Stéphanie Manel & Holderegger, 2013).

Larval dispersal is difficult to study through traditional methods. Not only is larval dispersal most often found in r-selected species with a very small proportion of large brood surviving, but the dispersal may also involve multiple life stages such as eggs, larvae, and juveniles making any tagging attempts futile. This has resulted in an increasing integration of biophysical particle tracking models to estimate larval dispersal (Bradburd & Ralph, 2019; Levin, 2006). Werner et al. (1993) were one of the first to demonstrate the utility of biophysical larval dispersal models to explain larval densities for Cod and Haddock. Since then, there have been a number of successful follow up studies (Benestan et al., 2016; Diopere et al., 2018; Knutsen et al., 2022; Schiavina et al., 2014; Schunter et al., 2011), with the ultimate goal of integrating population genomics and biophysical models towards evolutionary-based fisheries management (Baltazar-Soares et al., 2018)

Here I present an application of combining biophysical larval dispersal model in a spatio-temporal genetic framework developed in chapter 2, to calculate the expected population genetic differentiation. The Dispersal Model for Early Life Stages (DisMELS) was developed at the Alaska Fisheries Science Center (NOAA/NMFS) (https://github.com/wStockhausen/DisMELS). It is an Individual Based Model (IBM) that simulates the pelagic dispersal life stage of marine fishes. It combines the Regional Ocean Modeling System (ROMS)(Schroeder et al., 2014) output of physical oceanographic processes that are time frame specific with species specific life history biological parameters to simulate the dispersal of eggs and/or larvae from spawning/parturition locations, through pelagic juvenile stages, and eventual settlement in nearshore nursery areas (W. T. Stockhausen, Coyle, Hermann, Blood, et al., 2019). I focus on three commercially exploited species in the Gulf of Alaska for which the DisMELS results are available, the Pacific ocean perch (POP)(W. T. Stockhausen, Coyle, Hermann, Doyle, et al., 2019), Pacific cod (Hinckley et al., 2019), and arrowtooth flounder (ATF)(W. T. Stockhausen, Coyle, Hermann, Blood, et al., 2019), each with very different early life histories, and calculate the expected population structure based on the estimated larval dispersal from DisMELS model results.

Currently, these IBMs are limited by the absence of tools to empirically test their predictions (Hinrichsen et al., 2011; Coyle et al., 2013). In the case of DisMELS predictions, the larvae at a given location are aggregates of larval particles originating from various zones. The IBM (Stockhausen and Hermann, 2007; Stockhausen, 2009) describing dispersal of Pacific Ocean Perch (Sebastes alutus) larvae suggest a high degree of mixing among populations. The recent discovery of pelagic mixtures of multiple genotype clusters in POP in the eastern Gulf of Alaska (J. Maselko et al., 2020) supports these DisMELS model predictions. However, previous genetic studies reveal a strong pattern of isolation by distance (Kamin et al., 2014; Palof et al., 2011), which is thought to be inconsistent with the model results (W. T. Stockhausen, Coyle, Hermann, Blood, et al., 2019). Resolving this discrepancy is critical for understanding how juvenile fish recruit to adult populations. The primary goal of this study was therefore to determine whether DisMELS based dispersal can explain the observed population structure in the Gulf of Alaska for POP and Pacific cod, with the

secondary goal being to evaluate how the choice of sampling locations may influence population structure inference.

The spatio-temporal genetic model employed here allows for a comparison of the results to the observed genetic differentiation as measured by pairwise F_{ST} based on microsatellite data for POP (Palof et al., 2011), and RAD-seq SNP data for Pacific cod (Drinan et al., 2018). The model results are genetic differentiation between the simulated zones, and a number of simulated zones are concordant with the sampling locations used in the above studies for POP and Pacific cod which allows for direct comparison. I also show the effect of the choice of sampling locations on the population structure inference and propose a method to evaluate the optimal sampling strategy that reflects the overall population structure given a subsample of locations. Given that there is no available genetic data on arrowtooth flounder, the method proposed here, may well inform future sampling strategy in order to best elucidate their population structure.

Methods

Dispersal matrices were obtained from DisMELS model runs (W. Stockhausen, 2007). DisMELS is an individual based model (IBM) that combines the regional oceanic modeling system (ROMS)(Shchepetkin & McWilliams, 2005) outputs which provide three-dimensional model of oceanographic currents, temperature and salinity. The ROMS output is then overlaid with species specific early life stage biological processes that influence the larval survival and transport. These processes include water temperature, diel vertical migration and temperature-dependent survival. The results of the species specific DisMELS model runs are then trajectories of "individual" particles from spawning locations, pelagic duration, and settlement locations. Individuals that then successfully reach the settlement locations are based on the duration of larval particle at suitable settlement nursery habitats (W. Stockhausen, 2007).

The Gulf of Alaska DisMELS comprised of thirteen, ~150km wide along shore zones including spawning and nursery areas. The locations within the zones were then modeled as suitable habitats of egg deposition (arrowtooth flounder and Pacific cod) or parturition (Pacific ocean perch) areas as well as species specific suitable settlement areas (Figure 3-1b). The DisMELS output (Figure 3-1c) was generated for a 16 year period (1996-2011) for

Pacific ocean perch and arrowtooth flounder (W. T. Stockhausen, Coyle, Hermann, Blood, et al., 2019; W. T. Stockhausen, Coyle, Hermann, Doyle, et al., 2019), and 19 year period (1995-2013) for Pacific cod (Hinckley et al., 2019). All the model runs for each species were then averaged to obtain the mean pairwise dispersal parameters $\delta_{i\to k}$, between ith spawning and kth settlement location (Figure 3-1a). The resulting graphical representation of the strength of the connectivity among the zones for each of the species is depicted in Figure 3-2 with the top row heatmaps showing relative strength of connectivity and a directed graph diagram on the bottom row indicating direction of the connectivity only.

The three species (POP, ATF, and Pacific cod) all share early life history characterized by pelagic stage dispersal, but having very different ontogeny. Pacific ocean perch females are internally fertilized in the winter and then give live birth to anywhere from 10,000 to 300,000 larvae per female in the spring (Conrath & Knoth, 2013). The POP IBM then simulates the life history from this parturition stage where the particles are released from 300-600m isobaths and their pathways to nearshore settlement stage juveniles. Successful "particles" were those that as settlement stage juveniles, were entrapped for a minimum of 30 days near suitable nursery areas between September and November (Figure 3-2; W. T. Stockhausen, Coyle, Hermann, Blood, et al., 2019).

In contrast, arrowtooth flounder spawn in the water column between December and February at depths >400m (Blood et al., 2007). Each female releases 250,000-2,400,000 eggs in prolonged spawning events (Bouwens et al., 1999; Debenham et al., 2019). The arrowtooth flounder DisMELS IBM then simulated the episodic egg deposition from December to April at depths of 300 to 700m to nearshore juvenile settlement from August to October (Figure 1; W. T. Stockhausen, Coyle, Hermann, Blood, et al., 2019).
Pacific cod spawn from April to May on the sea bottom at around 200m depth, where the eggs are then attached to the substrate (Dunn & Matarese, 1987; Hinckley et al., 2019). Each female lays approximately 1-3 million eggs (Thomson, 1962). They hatch approximately 2 weeks later and move up in the water column (Laurel et al., 2008). They reach inshore nursery areas by July (Figure 4; Hinckley et al., 2019).

The mean DisMELS generated dispersal matrix $D_{species}$ reflected a very high dispersal rate (>0.65) (Table 3-1) which would result in a panmictic population and therefore

the off diagonal elements needed to be rescaled by a factor $\rho_{species}$. In order to maximize the signal of the dispersal matrix, the expected migration rate of 0.005 was chosen based on the results of the spatio-temporal genetic model verification conducted in chapter 2. The retention factor was calculated as the ratio of the DisMELS dispersal based on the average column mean less diagonal elements to expected migration rate of 0.005. The off diagonal elements were then divided by $\rho_{species}$ (Table 3-1). The adjusted $D_{species}$ were then row normalized in order to reflect the probabilities of dispersal from ith to kth zone.

The specific model parameters used were based on the verified model behavior (see chapter 2), with 1,000 loci, mutation rate of 1e-6, population size of 1,000, and migration as described above of Nm=5, and stochasticity=TRUE. The age structure was chosen to reflect the life history of each species:, Pacific ocean perch spawning ages of 10-98 years (*Pacific Ocean Perch | NOAA Fisheries*), arrowtooth flounder, 7-27 years (*Arrowtooth Flounder | NOAA Fisheries*, n.d.), and Pacific cod, 5-20 years (*Pacific Cod | NOAA Fisheries*).

The spatio-temporal model was then ran for each species for a maximum of 30,000 year steps and stationarity, calculated as no change in the mean $G_{ST}^{"}$ and variance $G_{ST}^{"}$ for the previous 100 year steps. The linearized $\frac{G_{ST}^{"}}{1-G_{ST}^{"}}$ was then plotted against distance (km) between the zones, where distance was calculated as the midpoint distance between adjacent zones, and the distance between non-adjacent zones as the sum of the distances between the intermediate zones. Subsequent analysis was based on sampling 100 genotypes from each zone where the number of alleles per locus (0=aa, 1=Aa, 2=AA) were drawn from a binomial distribution with probability equal to the allele frequency.

PCA analysis was conducted using the R (R Core Team, 2021) pcomp() function and plotted using ggplot2() (Wickham, 2016). I used LEA (Frichot & Francois, 2015) snmf() function to calculate the STRUCTURE like admixture coefficients. Mapping of putative populations in the DisMELS designated zones throughout the Gulf of Alaska was done using the marmap (Pante & Simon-Bouhet, 2013) package.

The Pacific ocean perch observed pairwise F_{ST} data was based on microsatellite analysis from (Palof et al., 2011). The Pacific cod observed pairwise F_{ST} data was based on RAD-seq analysis (Drinan et al., 2018). Both of the observed datasets were culled to retain

only the observations from locations corresponding to the DisMELS designated zones. Pacific ocean perch observed data corresponded to the DisMELS population zones (1, 2, 4, 6, 9, 10, 12), and the Pacific cod data corresponded to zones 1, 6, 8, 11, and 12. This allowed for a direct comparison of the observed and model predicted population structure inference using PCA, STRUCTURE, and regression analysis. STRUCTURE analysis was performed separately on the subset of the zones corresponding to the observed datasets.

Sampling strategy evaluation was done by selecting the subset of five zones that when sampled would result in the population structure inference that was the best and worst model fit as compared to sampling all zones. The model fit was calculated using the log likelihood framework. First I calculated the parameters of the beta distribution from the pairwise $G_{ST}^{"}$ based on sampling all zones. I then calculated the log likelihood for all the combinations of sampling of five zones only. The best and worst fit models were determined by the subset resulting in the highest and lowest log likelihoods respectively. STRUCTURE analysis was performed on the subsets identified as the best and worst model fit.

Results

The number of year steps required to achieve weak stationarity determined as the point when the mean $G"_{ST}$ for the previous 100 year steps has not changed, was the maximum simulation runs of 67,300 for POP, 9,500 for ATF, and 19,500 for Pacific cod (Table 3-1). This corresponds to approximately 4,500 generation for POP, assuming a generation time of 15 years, 1,000 generations for ATF and 2,000 for Pacific cod assuming generation time of 10 years for both species.

Expected Population Structure

The results of the spatio-temporal genetic model show a significant IBD relationship of genetic differentiation with distance between zones for POP ($R^2 = 0.092, p = 0.024$), and Pacific cod ($R^2 = 0.787, p < 0.001$), but not ATF ($R^2 = 0.002, p = 0.776$) (Figure 3-3). However, principal component plots show the presence of distinct clusters for the three species (Figure 3-4). Pacific ocean perch model appears to show four groups with zones 1, 2, and 3 being separate, distinct groups and a single group in the zones 4-12. Likewise, arrowtooth flounder model shows distinct groups for zones 1, 2, 3, and single group of zones 4-12. In contrast, the Pacific cod model, shows a more of an isolation by distance pattern (see

chapter 2), with gradual transition of genetic differentiation from zones 1 to 12, with zone 13 being more distinct.

STRUCTURE analysis based on 100 genotypes sampled from each zone showed very different admixture patterns among the three species modeled. For Pacific ocean perch, STRUCTURE analysis identified 3 putative population groups (Figure 3-5). There were three distinct groups associated with zones 1-3 that then persisted throughout the Gulf of Alaska. Zone 1 had little admixture, with zone 2 admixture of a second genetic group followed the third population group (blue) off of cross sound. All three putative groups are then admixed in relatively equal proportions resulting in a concurrent detection of distinct groups from Yakutat and further west. The model also predicted that most of the zones consisted of the three sympatric genotypes (zone 4, 5, 6, 7, 8, 10, 12) originating from zone 1, 2, and 3. Likewise two distinct genotypes present at zones 9, and 11 consisting of genotypes found in zone 2 and 3.

Arrowtooth flounder resulted in 6 putative populations with five unique with little admixtures in zones 1-5, and sixth admixed population spread out over western Gulf of Alaska zones 6-12 (Figure 3-6). Unlike Pacific ocean perch, there were no zones where apparent sympatry of genotypes was present. In contrast to STRUCTURE which identified 6 population groups, PCA distinguished only 4 clusters, with zones 1-3 being most distinct and zones 4-12 grouped in a single cluster.

STRUCTURE analysis of Pacific cod model showed a much greater admixture in the majority of the zones, except zone 1-3 being the only zones with little admixture (Figure 3-7). These genotypes were then propagated throughout the Gulf of Alaska and are found in all the zones, except zone 13 in Cook Inlet. Two additional genotypes are identified in zone 11, 12, and 13 with zones 11, and 12 containing all 3 genotypes, while zone 13, the only zone that did not contain the prevalent genotype from eastern Gulf of Alaska. PCA also indicated the presence of three clusters, but they were not concordant with the STRUCTURE derived population groups.

Model Validation

There were eight sample locations from (Palof et al., 2011) Pacific ocean perch microsatellite data set that corresponded to the DisMELS zones, namely zones 1, 2, 4, 6, 9,

10, 11, and 12 (Figure 3-8). The linearized F_{ST} regression on zone distance, showed a significant positive slope for the Palof ($R^2 = 0.332, \mathrm{p} = 0.001$) and DisMELS simulated ($R^2 = 0.157, \mathrm{p} = 0.037$) dataset. The observed and simulated pairwise F_{ST} values had a correlation of 0.184 (Mantel test $p = 0.129$). The PCA plot indicates the presence of three population clusters, while the STRUCTURE analysis indicates the presence of all three groups corresponding to the genotypes in the full dataset. These genotypes are also predicted to be found throughout the sampling area with zone 1 containing only two, while the rest of the zones containing all three sympatric genotypes.

There were five sites for the observed Pacific cod RAD seq data from (Drinan et al., 2018) study that corresponded to the DisMELS zones 1, 6, 8, 11, and 12 (Figure 3-9). Both observed and DisMELS derived genotypes showed a significant positive slope when regressing the linearized F_{ST} on zone distance, indicative of isolation by distance with $R^2 = 0.713$ ($p = 0.002$), and $R^2 = 0.646$ ($p = 0.005$) for the observed and modeled data respectively. Unlike the full DisMELS simulated data which represented three putative genotypes, the subset only contained two with sympatry in zones 11 and 12 in the Aleutian chain. The observed (Drinan et al., 2018) and DisMELS derived pairwise F_{ST} matrices were significantly correlated with $r = 0.72$ (Mantel test $p = 0.03$).

Effect of Sampling Design

The effect of the choice of the subset of five zones to sample from all thirteen was dependent on which species model was used (Table 3-2). For Pacific ocean perch, the five zones which best ($\log(\mathcal{L}) = 228.9$) described the full dataset were 1, 4, 5, 6, and 7 (Figure 3-10, top) with a log likelihood of 228.9. This resulted in a significant linear IBD relationship of genetic differentiation with distance ($R^2 = 0.482; p = 0.031$), and two distinct STRUCTURE derived populations. The sampled zones are also predicted to contain the two genotypes in equal proportions in zones 4, 5, 6, 7, and a single genotype in zone 1. In contrast, the worst model fit ($\log(\mathcal{L}) = -2{,}921.5$) was when sampling zones 4, 8, 9, 10, and 12 (Figure 3-10, bottom). This resulted in an apparent panmictic population inference when examining PCA and STRUCTURE plots, with the single genotype distributed throughout the sampling area. However, the resulting regression plot showed a significant slope with ($R^2 = 0.962; p < 0.001$) but with pairwise F_{ST} values two orders of magnitude smaller than in the

best model. Note that zone 4 is included in both the best and worst fit models. In the presence of the zones in the best model fit, both PCA and STRUCTURE identifies two sympatric population clusters, while in the worst model fit, only a single cluster is identified.

The five zones that were most representative ($\log(\mathcal{L}) = 104.75$) of the overall population structure of arrowtooth flounder were 1, 9, 10, 11, and 12 (Figure 3-11, top). These five zones represented only two putative populations out of total of 6 identified with the full model as indicated by the STRUCTURE and PCA plots. However, unlike when sampling all zones, the best model subset displayed a significant genetic differentiation with distance ($R^2 = 0.95; p < 0.001$). In contrast, the worst model fit ($\log(\mathcal{L}) = -27{,}796.6$) was when sampling zones 8, 9, 10, 11, and 12 (Figure 3-11, bottom). This resulted in no apparent genetic differentiation with zone distance ($R^2 = 0.044; p = 0.562$), and an apparent panmictic population as indicated by both PCA and STRUCTURE analysis.

For Pacific cod, sampling zones 7, 8, 9, 10, and 13 would result in inference most like ($\log(\mathcal{L}) = 223.4$) the full data set (Figure 3-12, top). They showed significant differentiation with distance ($R^2 = 0.048; p = 0.544$), and two distinct STRUCTURE derived populations in zones 7 and 13, and admixture of the two in the other zones. PCA plot, however, indicated a possible presence of a third population cluster, which was not indicated in the STRUCTURE analysis. Sampling sites 1, 2, 3, 5, and 6 resulted in the worst model fit ($\log(\mathcal{L}) = -1{,}331.1$), but with a significant association between genetic differentiation and zone distance($R^2 = 0.709; p = 0.002$) (Figure 3-11, bottom). However, STRUCTURE analysis indicated the presence of a single dominant population genotype in zones 1-3 with slight admixture of a second genotype in zones 5 and 6, but no presence of two genotypes in any of the zones. PCA plot as well, suggested a presence of a single putative population cluster.

Discussion

Combining the spatio-temporal genetic model with the IBM larval dispersal allowed for novel insights into the long term effects of dispersal on population structure. The results presented here demonstrate that the DisMELS based larval dispersal may accurately describe the observed linear isolation by distance pattern in the Gulf of Alaska Pacific cod and to a lesser degree Pacific ocean perch populations. Additionally, the seemingly inconsistent

findings of minimal dispersal leading to genetic population structure (Kamin et al., 2014; Palof et al., 2011) and DisMELS predicted long distance larval advection are in fact consistent, especially in light of some recent findings where sympatric populations were found in young of the year pelagic aggregates as well as adult populations using RAD-seq and WGS approaches respectively (Maselko et al., 2020; Timm, L in prep).

The Pacific cod population structure in the Gulf of Alaska has been consistently described as having an isolation by distance genetic differentiation with limited gene flow among adjacent populations (Cunningham et al., 2009; Drinan et al., 2018). This is also supported by the simulation results presented here, with strong concordance between simulated and observed genetic differentiation patterns (Figure 3-9). In fact, when controlling for sampling location, both observed and modeled populations had positive and significant IBD genetic differentiation with distance ($R^2 = 0.713, \mathrm{p} = 0.002$ and $R^2 = 0.646, \mathrm{p} = 0.005$ respectively). The two datasets were also positively and significantly correlated ($r = 0.72$, Mantel test $p = 0.03$) suggesting that the model accurately captured the observed population structure when sampling corresponding sites. This is unsurprising since the DisMELS generated dispersal matrix (Figure 3-2) shows a decay in dispersal rates with increasing distance to nonadjacent zones. This is indicative of the classic isolation by distance model (Sewall Wright, 1943) that has been extensively applied ever since (Jenkins et al., 2010). Therefore it appears that the DisMELS model captured the Pacific cod larval dispersal accurately, suggesting that the majority of their dispersal may in fact occur during the pelagic larval and juvenile stage at relatively small distances.

DisMELS predicted larval dispersal for Pacific ocean perch and arrowtooth flounder was considerably more complex (Figure 3-2), yielding little insight of the expected population structure or difference between the two species based solely on the dispersal matrices. However, once the spatio-temporal genetic model was applied, the differences in the expected population structures became apparent with arrowtooth flounder showing a strong pattern of distinct, isolated populations in the eastern Gulf of Alaska and a single putative population west of Kayak Island with some admixture from the eastern populations (Figure 3-6). This pattern was not evident when examining the IBD regression plots for ATF (Figure 3-3, middle), which showed no significant IBD relationship, demonstrating the

complex population structures resulting from larval dispersal. POP did have a significant IBD relationship (Figure 3-3, top), however, the expected population structure for POP is dominated by little admixture in the southeastern Gulf of Alaska, but high admixture in the western and the Aleutians (Figure 3-5). This suggests that both arrowtooth flounder and to a lesser degree, Pacific ocean perch, may be vulnerable to local extirpation in the eastern Gulf of Alaska, but less so in the western part. However, this pattern could be due to DisMELS model truncation, since no influx of larvae from south of Dixon entrance is modeled. Therefore the populations in zones 1-4 may have under estimated admixture from populations outside the area.

Recent study by Timm, L (in prep) of whole genome sequencing data of Pacific ocean perch adult collections from 2017 and 2019 shows a distribution of five distinct genotype groups throughout the Gulf of Alaska (Figure 3-13). Note that the genotype groups A, B, C, and D correspond to the POP larval aggregates (Jacek Maselko et al., 2020). It also shows that many of the genotype groups are co-occurring in close proximity, which is predicted by the modeled results here. This supports the conclusion presented here that the observed population structure are greatly influenced by the oceanographic currents during the pelagic larval dispersal and the IBM may explain most of the observed variation and distribution of the genotypes including sympatry.

The model predicted high admixture in the Pacific ocean perch populations are indicated in the apparent sympatry of different population genotypes (Figure 3-5) with population B originating near Cross Sound being the dominant genotype in all populations to the west. As well, POP larval collections were composed of four distinct genetic groups (Maselko et al., 2020) indicating sympatry during both the pelagic larval and adult stages (Figure 3-13). However, this seemed inconsistent with the previous genetic studies (Kamin et al., 2014; Palof et al., 2011; Withler et al., 2001) which indicated an isolation by distance pattern. In the Palof et al. (2011) study, there were distinct adult populations with limited gene flow. Likewise, in the (Kamin et al., 2014), juvenile POP, were found to be in close geographic proximity to their genetically similar adult groups. As well, (Withler et al., 2001) described distinct adult POP populations at small spatial scales in British Columbia suggesting little dispersal among them. However, the genetic model results here, suggest

long distance dispersal of >1,000 km, yet the resulting genetic differentiation inference are similar to the observed patterns (Figure 3-8).

Comparing the POP IBD regressions shows a similar pattern between the observed POP microsatellite data (Palof et al., 2011), and the modeled SNP data presented here when controlling for sampling locations (Figure 3-8). Both the observed and modeled data sets show significant positive slope ($R^2 = 0.332, \mathrm{p} = 0.001$ and $R^2 = 0.157, \mathrm{p} = 0.037$ respectively), but they were not significantly correlated ($r = 0.184$, Mantel test $p = 0.129$). Unlike, Pacific cod, where both data sets were based on SNPs, the lack of concordance may also be influenced by the difference in genetic markers used since Palof et al. (2011) used microsatellites, and the spatio-temporal genetic model used here simulates SNPs. Nonetheless, the surprising agreement between the modeled population structure and the predicted mixtures of genotypes in the observed larval aggregates (Maselko et al., 2020) and adult collections (Timm, L, in prep) (Figure 3-13) suggests that, similarly to the concordance found in Pacific cod, the utility of the DisMELS model results to predict expected population structure due to IBM derived dispersal.

Some of the discrepancy for the Pacific ocean perch model is that the model results are of genetic differentiation when dispersal-drift-mutation equilibrium is reached. However, for POP, the weak stationarity was not achieved until 67,300 year steps which is possibly unrealistic considering the Bering land bridge, and ice age were influencing the Gulf of Alaska in the Pleistocene as recently as 10,000 years ago (Stewart & Lister, 2001). Therefore, the lack of dispersal-drift-mutation equilibrium could be a shortcoming of this model, since it is highly unlikely that these populations are in the drift-migration (dispersal) equilibrium (Whitlock & McCauley, 1999). However, it may be possible to determine the number of year steps parameter by comparing the resulting population structure at a given time to the observed population structure using the model selection criteria used here. This may elucidate additional insight into the species life history and population structure.

The retention parameter, which is used to adjust the off-diagonal elements of the dispersal matrix so that the average total dispersal into each zone was 0.005, allowed for the maximum resolution of STRUCTURE and PCA results as showed in chapter 2 and (Waples & Gaggiotti, 2006). This shows that there may be considerable larval retention that is not

captured by the DisMELS model which in raw form predicts panmictic population due to the resulting magnitude of dispersal (>0.8). It could also mean that since the DisMELS does not take survival into account, that there is mortality proportional with distance from natal populations. Maselko et al. (2020) did find that there is a significant association of allele loss with collection date and latitude but not in the change in population group composition. This indicates that all populations were equally affected. Alternatively, and perhaps more contentiously, apparent retention could also be due to homing behavior after juvenile settlement in the nursery habitat, distant from their natal populations as they may be. The homing ability of marine fish is commonplace, although not fully understood especially when large portion of dispersal occurs during larval stages. Homing has been well documented in salmon, eel, and even recently Atlantic cod (Bonanomi et al., 2016; Svedäng et al., 2007). The natal homing ability of marine reef fish, where the larvae use soundscape, sun, magnetic field, and olphactory acuity to find settlement habitats has received the most attention (Leis et al., 2011). And although the barrier to imprinting due to metamorphosis from larvae to adults is still contentious in fish (I. Bradbury & Laurel, 2007), it has been well documented in arthropods (Blackiston et al., 2008). The presence of large larval aggregates composed of distinct population groups as observed by Maselko et al. (2020) with no discernable difference in the larval condition among the groups suggests that adult homing behavior in POP and not larval stage mortality is the most likely explanation.

As expected, the choice of sampling locations has large effect on the population structure inference, and underscores the importance of sampling over a broad geographic scales. Note that both POP and ATF population structure is defined by a single outgroup in the southeastern Gulf of Alaska and the remaining genotypes from zones in the western Gulf of Alaska. The poorest fit for both ATF and POP was when sampling only the western Gulf of Alaska resulting in an apparent panmictic population inference. This underscores the importance of incorporating these types of dispersal models when designing a sampling strategy, especially when trying to designate management units for exploitation, where the incorrect population structure inference may lead to overexploitation or even extirpation.

This model could also be used in a management strategy evaluation by identifying populations most at risk for loss of diversity due to overexploitation, and conversely the most

resilient ones. Whilst the DisMELS model results can be used to estimate the relative dispersal and immigration of larvae among the populations, combining the spatio-temporal genetic model may illuminate the long term loss of genetic diversity in the metapopulations due to local extirpation. For example, using the genetic population delineations (i.e. STRUCTURE, PCA) to describe management units may result in a preservation of genetic diversity, which may be especially important during accelerated environmental change (Sunday et al., 2014).

Finally, the spatio-temporal genetic model employed here relies on the evolution of neutral markers as a function of dispersal, thereby ignoring any local adaptation (Selkoe et al., 2016). However, local adaptation may play a role in larval retention or homing (Funk et al., 2012), but is not accounted for here. Incorporating a spatially explicit selection parameters, simulating local adaptation may allow for a further refinement of the model, as well as inform zones of at risk populations.

Tables and Figures

Table 3-1 Model parameters for the three species. Life history parameters were minimum and maximum spawning ages. The Years to Stationarity was calculated by running the simulation until the slope of the mean and variance of all pairwise population F_{ST} values did not change. For Pacific ocean perch simulation, the stationarity is not achieved even after 30,000 years, but I did not carry the simulation further.

Larval Dispersal Model	Minimum Age	Maximum Age	Years to Stationarity	Raw m	Retention Factor ρ	Final m
Pacific ocean perch	10	98	67,300	0.845	169.0	0.005
Arrow tooth flounder	5	20	9,500	0.857	171.4	0.005
Pacific cod	6	27	19,500	0.623	124.6	0.005

Table 3-2 Results of sample design evaluation for a subset of 5 zones to determine which subset best explains the overall population structure.

POP	α	β	$Log(\mathcal{L})$	Sampled Zones
Full Set	0.41	24.49		
Best Fit	0.60	25.05	228.86	1, 4, 5, 6, 7
Worst Fit	1.05	3,082.85	-2,921.51	4, 8, 9, 10, 12
ATF	α	β	$Log(\mathcal{L})$	
Full Set	0.93	7.26		
Best Fit	0.45	4.07	104.75	1, 9, 10, 11, 12
Worst Fit	1.80	3,392.99	-27,796.64	8, 9, 10, 11, 12
Pacific cod	α	β	$Log(\mathcal{L})$	
Full Set	0.58	21.41		
Best Fit	0.65	21.77	223.44	7, 8, 9, 10, 13
Worst Fit	1.57	930.99	-1,331.07	1, 2, 3, 5, 6

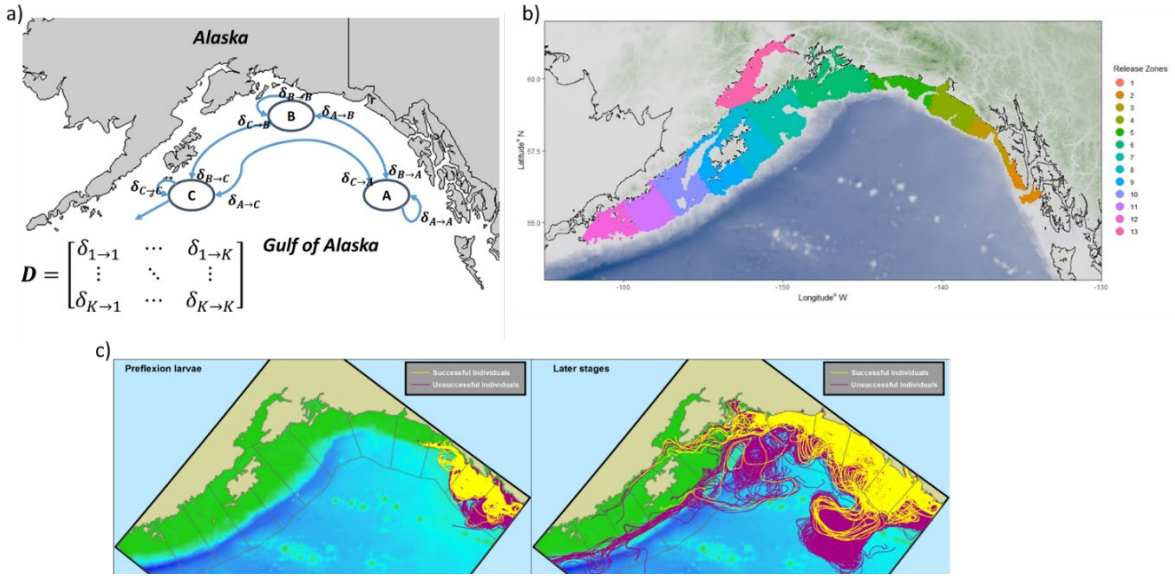


Figure 3-1 Map of the Gulf of Alaska depicting the modeled area. a) Conceptual model of the larval dispersal where each of the $\delta_{i\to k}$ parameters describe the probability of larvae dispersing from ith population and settling and recruiting to the kth population. The $\delta_{i\to k}$ parameters are obtained from the individual based DisMELS (*GitHub - WStockhausen/DisMELS: A Java-Based Framework for Developing/Running Individual-Based Models(IBMs) for Marine Species with Pelagic Early Life Stages.*, n.d.) model runs that incorporate ocean circulation and life history parameters specific to the species. b) Map depicting the 13 modeled spatial zones in the Gulf of Alaska of larval release and settlement used in the DisMELS model to calculate the $\delta_{i\to k}$ parameters. c) shows an example of DisMELS model graphical output where the left panel are the pathways during early larval dispersal stages and right hand panel are the subsequent stages. Yellow indicating pathways of "successful" larvae and magenta being the "unsuccessful" runs.

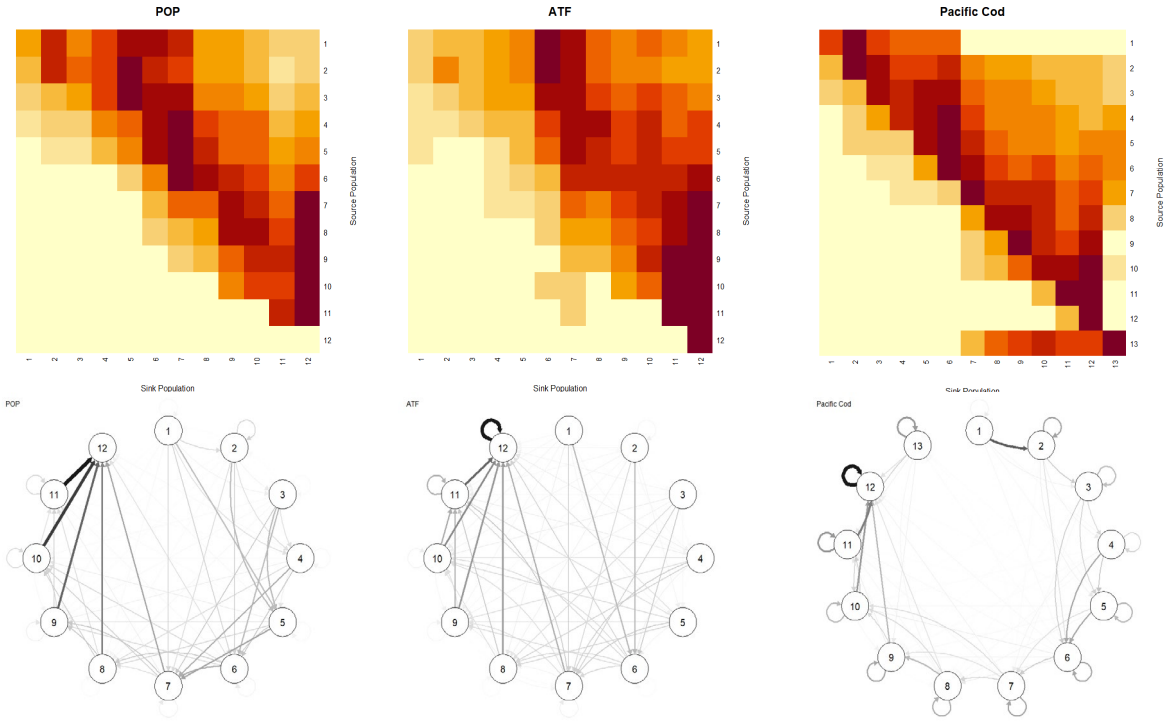


Figure 3-2 Graphical representation of the D dispersal probability matrices obtained from the DisMELS model for the three species, Pacific ocean perch (POP), Arrow tooth flounder (ATF), and Pacific cod. The top row heat maps depict the strength of dispersal with darker colors denoting higher probability. The bottom row are the corresponding directed graphs that show a different representation of the connectivity among the DisMELS population zones.

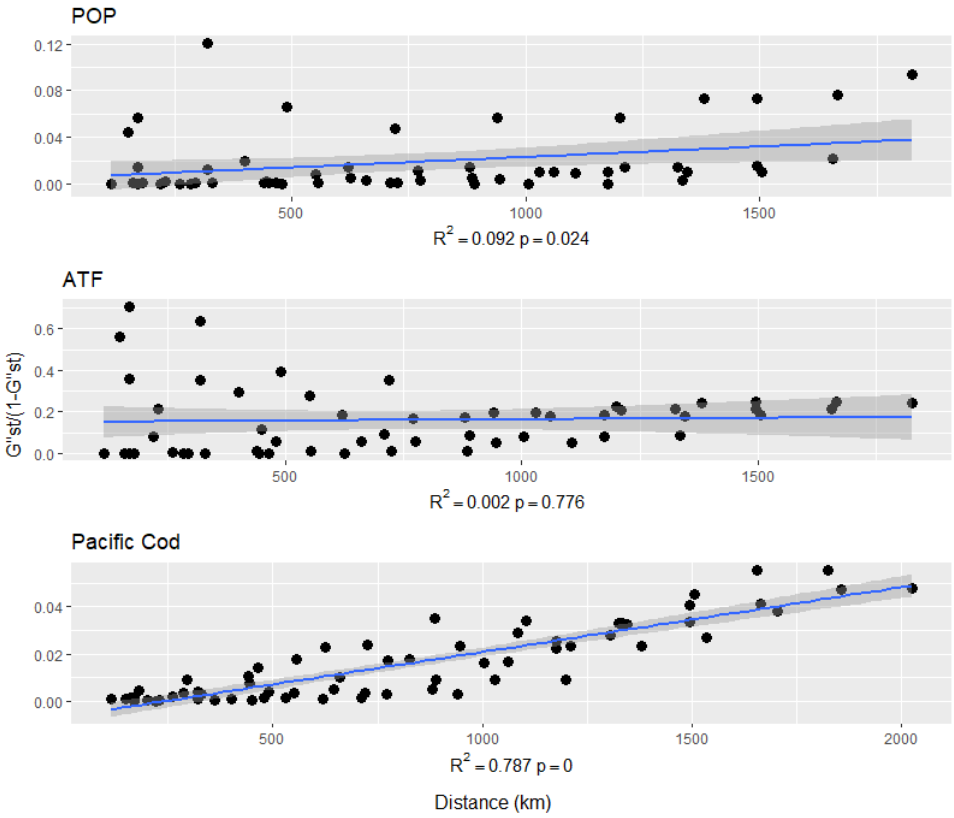


Figure 3-3 The expected IBD regression plots for the three species using the pairwise linearized $G"_{ST}/(1 - G"_{ST})$ on the cumulative distance between zones (km). Based on the biophysical dispersal model, there is an expected significant IBD relationship for POP and Pacific cod, however, ATF shows no significant genetic differentiation with distance.

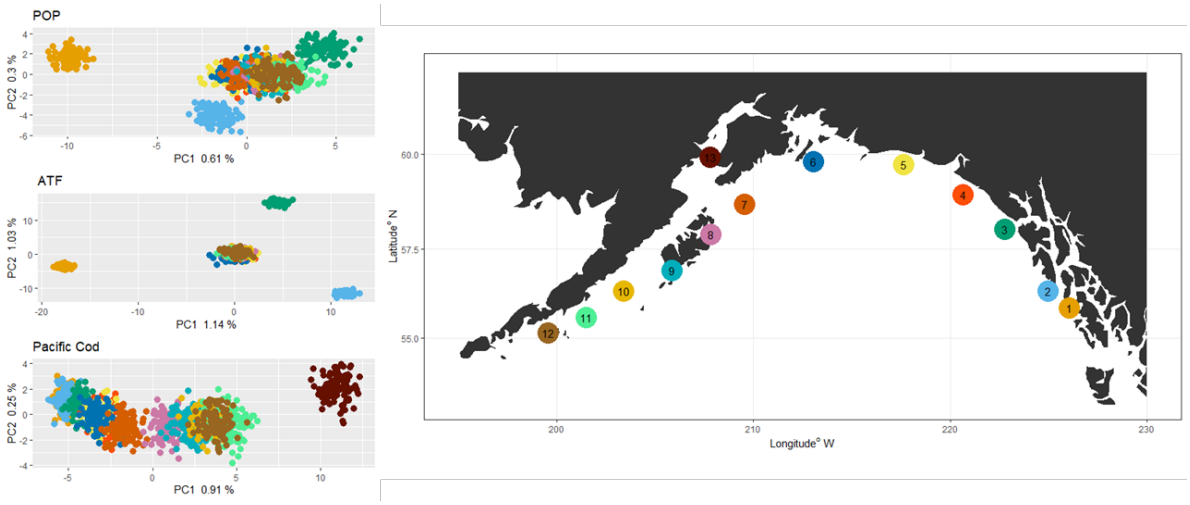


Figure 3-4 PCA plot of the genetic simulation results and sampling 100 individuals from each population-zone across all age classes. The PCA plot colors correspond to the population-zone location colors on the map to the right.

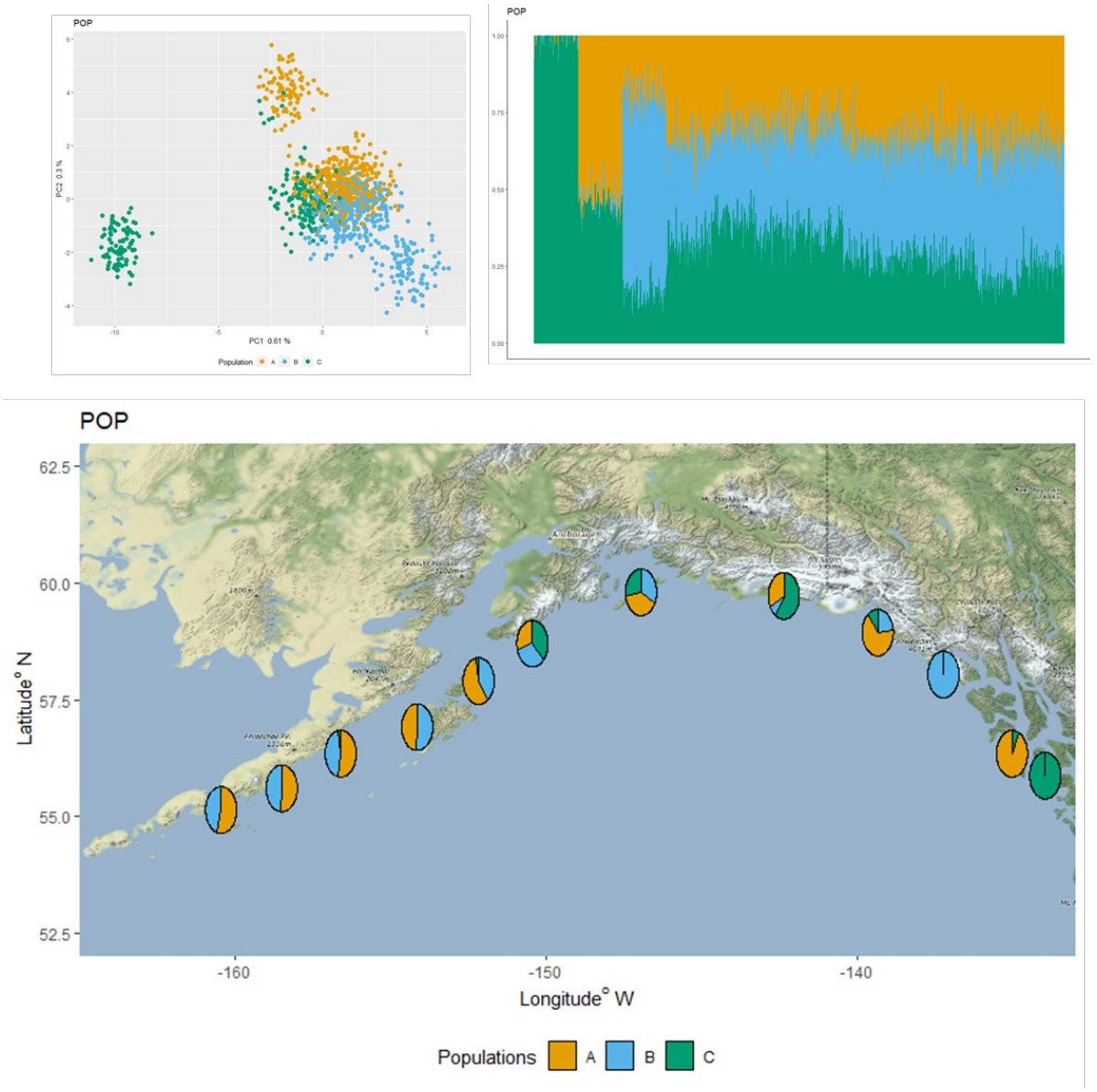


Figure 3-5 STRUCTURE derived putative populations for Pacific ocean perch (POP) based on sampled 100 individuals per DisMELS population zone. The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. PCA representation shows 3 distinct clusters with a possible fourth one. The STRUCTURE plot show a single (green), distinct population group (in the farthest southeastern zone), and immediately followed by admixture from another population (orange) in the adjacent zone to the northwest. The third population group (blue) appears off of cross sound. All three putative groups are then admixed in relatively equal proportions resulting in a concurrent detection of distinct groups from Yakutat to east of Kodiak island, at which point only two admixed populations are evident to the west. The main break in genotype distribution appears around Kenai Peninsula.

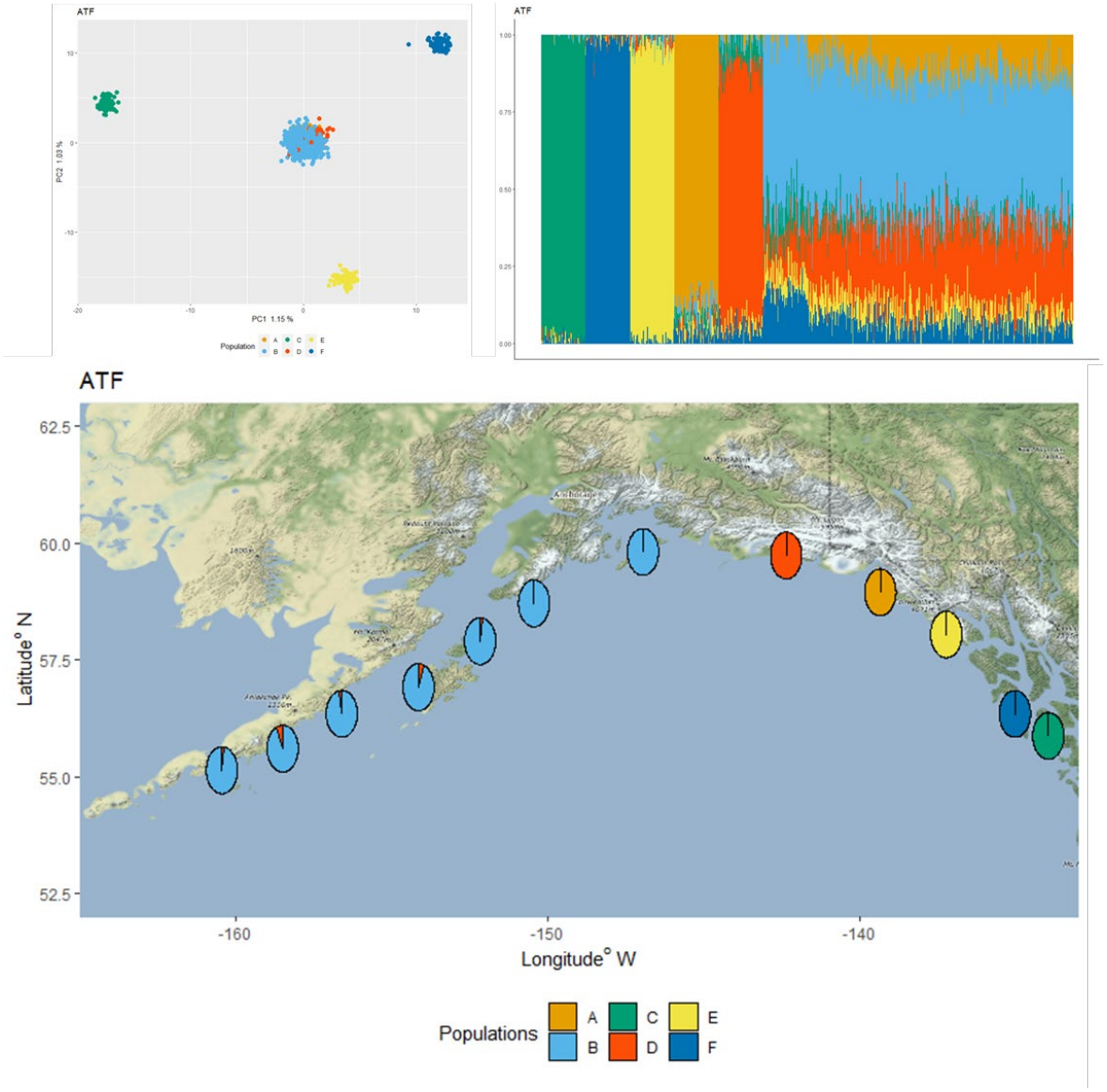


Figure 3-6 STRUCTURE derived putative populations for Arrow tooth flounder (ATF) based on sampled 100 individuals per DisMELS population zone. The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. The PCA plot shows 4 distinct population clusters with obscured groups in the middle of the plot, possibly indicating a non-linear variation in the genotype groups. The STRUCTURE admixture analysis indicates 6 distinct groups with increasing admixture from the eastern to western Gulf of Alaska. The zones in the eastern part of Gulf of Alaska are composed of distinct, putative populations, with little admixture among them. However from outside of Prince William Sound, to the west, there is only a single, dominant population identified with little admixture from the eastern populations. The map depicts where the corresponding, colored putative populations are spatially distributed.

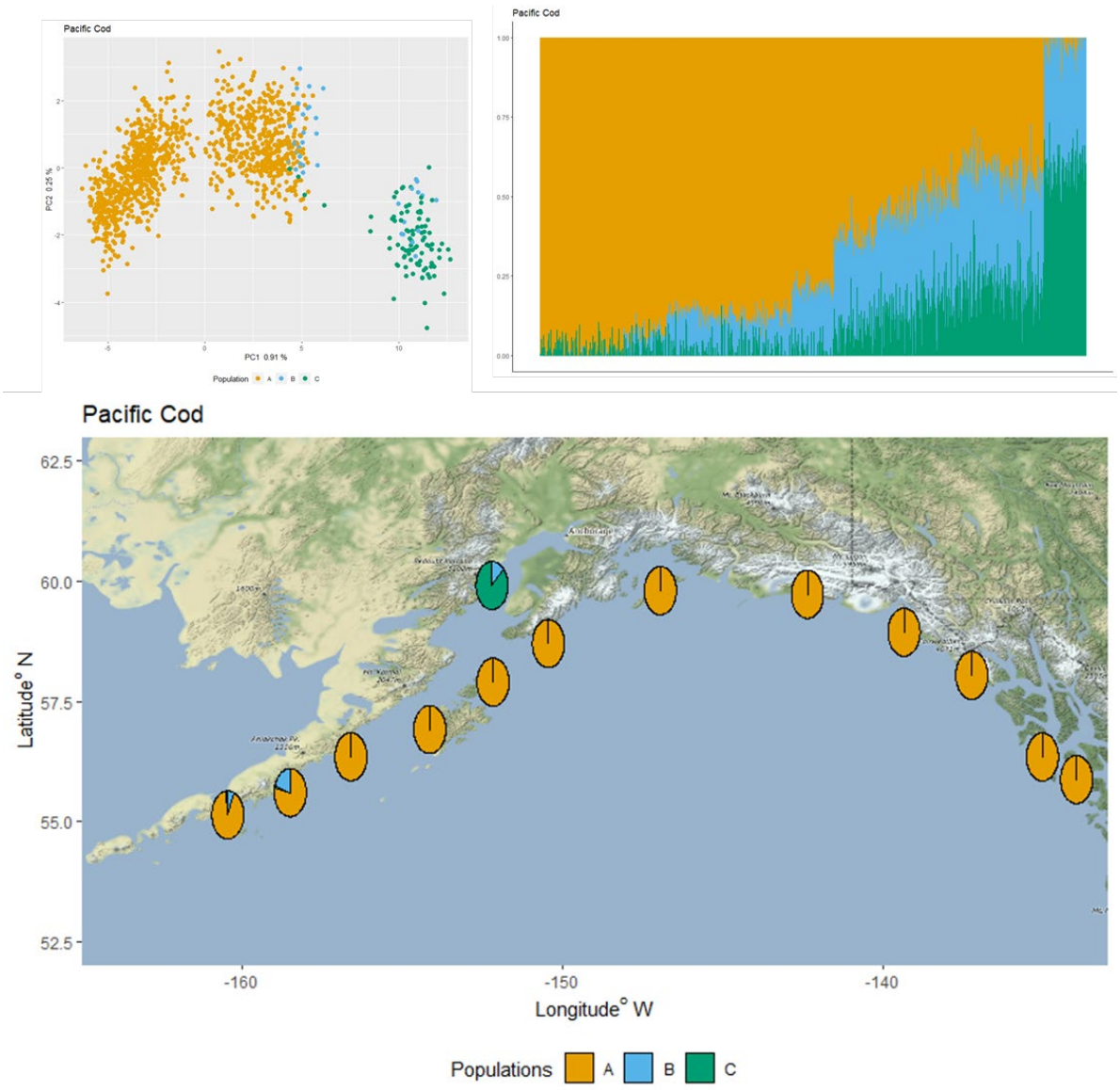


Figure 3-7 STRUCTURE derived putative populations for Pacific cod based on sampled 100 individuals per DisMELS population zone. The map depicts where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. PCA plot suggest 3 genotype clusters, but they are not concordant with the STRUCTURE derived clusters. STRUCTURE analysis shows a single population throughout the eastern Gulf of Alaska which then progressive admixture originating from zone 13 (Cook Inlet). The farthest west zones indicate the presence of some admixture from other putative populations. The map depicts where the corresponding, colored putative populations are spatially distributed.

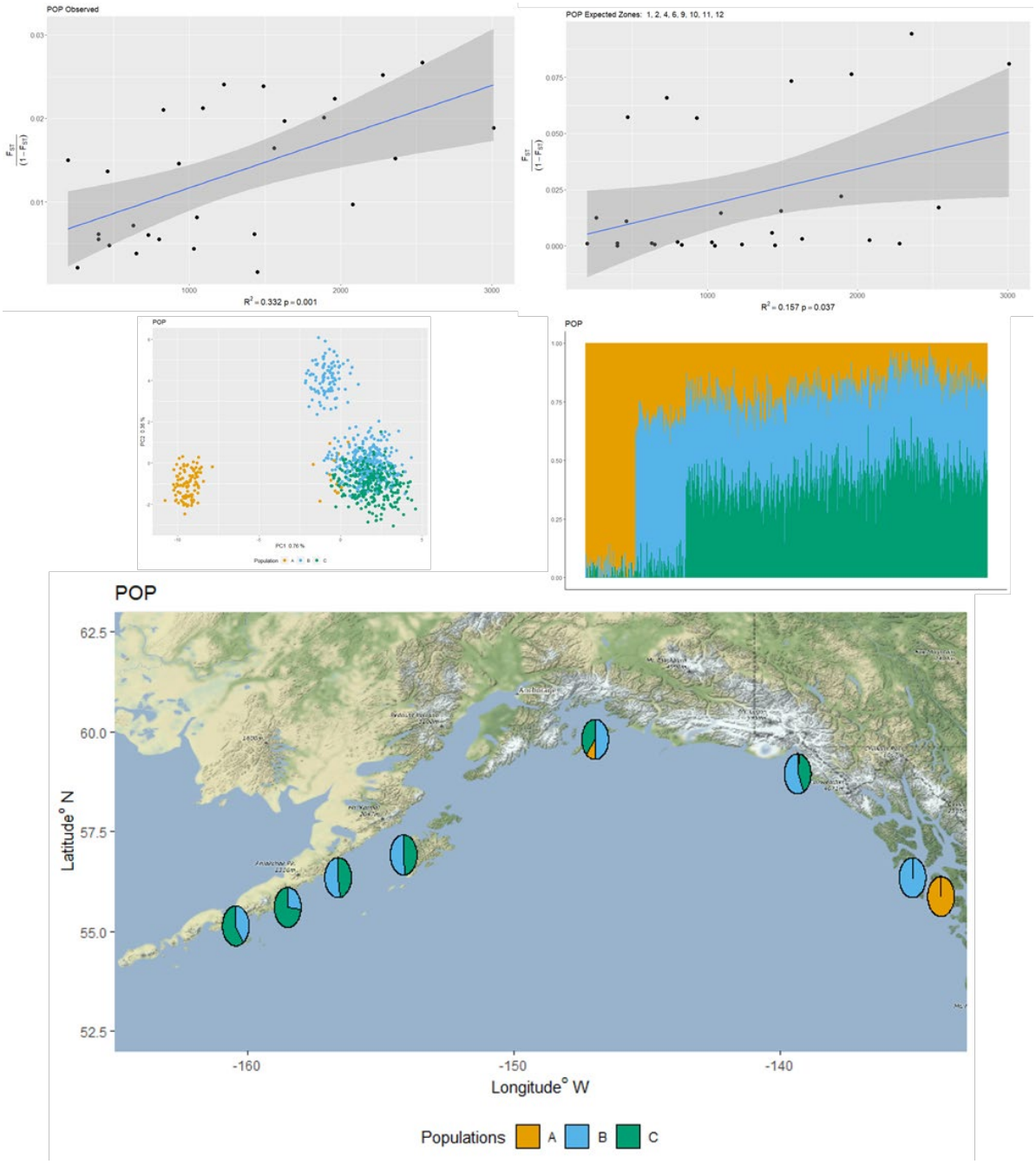


Figure 3-8 Comparing the observed and modeled population structure for Pacific ocean perch. The observed data set was adapted from (Palof et al., 2011) where the sampled locations were matched to the DisMELS population zones (1, 2, 4, 6, 9, 10, 12). The matching subset of the linearized pairwise observed F_{ST} values for both the observed and modeled populations was then regressed on the coast distance. Both observed and modeled populations had positive and significant scale of genetic differentiation with distance ($R^2 = 0.332, \mathrm{p} = 0.001$ and $R^2 = 0.157, \mathrm{p} = 0.037$ respectively). The observed and simulated pairwise F_{ST} values had a correlation of 0.184 (Mantel test $p = 0.129$). The map depicts

where the corresponding, colored putative populations are proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. Both STRUCTURE and PCA plots indicate the presence of three populations clusters with clusters A and B (orange and blue) dominating eastern Gulf of Alaska, while clusters B and C (blue and green) dominating western part with the dominant break occurring outside of Prince William Sound.

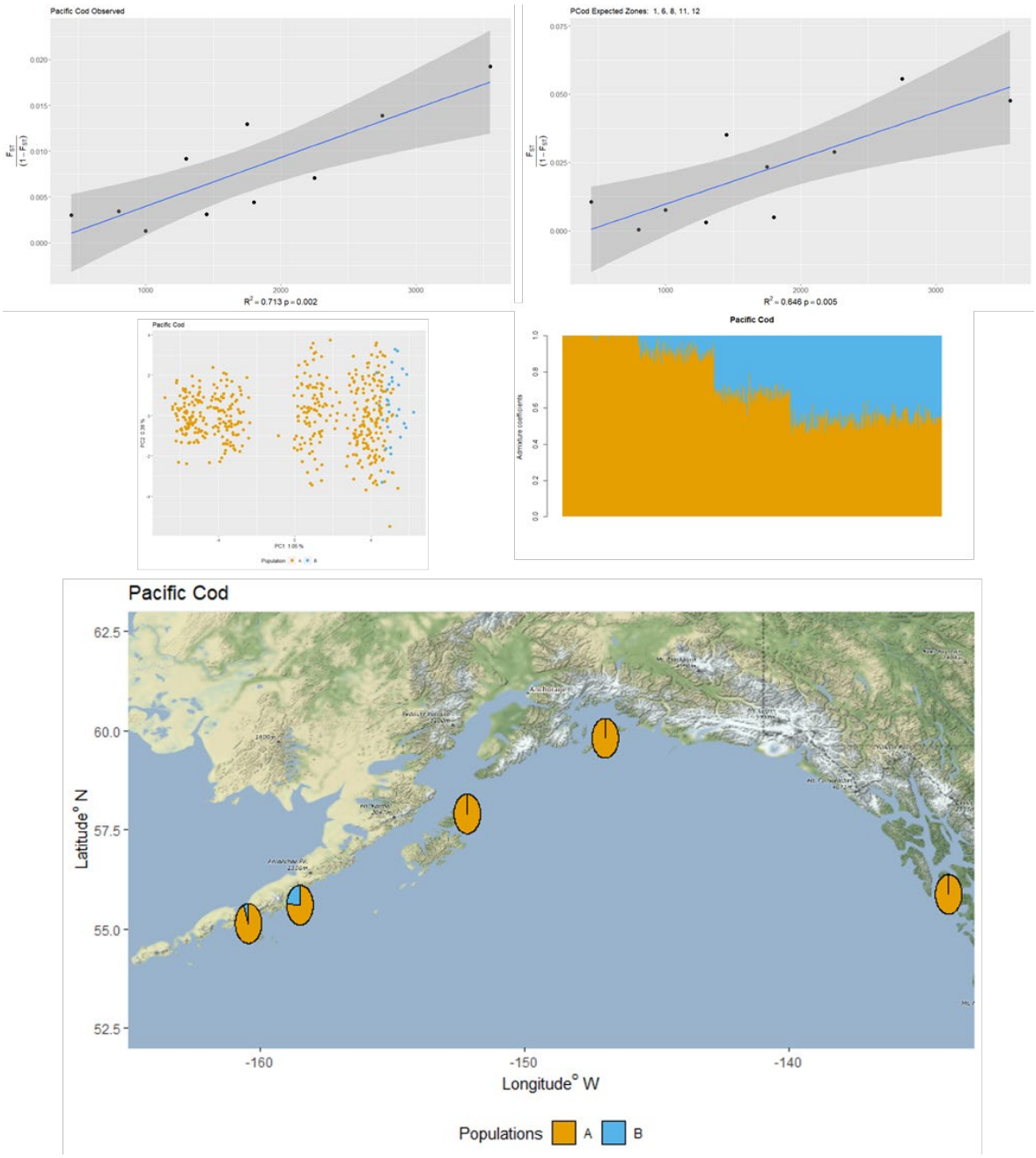


Figure 3-9 Comparing the observed and modeled population structure for Pacific cod. The observed data set of linearized $G"_{ST}$ was adapted from (Drinan et al., 2018) where the sampled locations were matched to the DisMELS population zones (1, 6, 8, 11, 12). The matching subset of the linearized pairwise observed F_{ST} values for both the observed and modeled populations was then regressed on the coast distance. Both observed and modeled populations had positive and significant scale of genetic differentiation with distance ($R^2 = 0.713, \mathrm{p} = 0.002$ and $R^2 = 0.673, \mathrm{p} = 0.004$ respectively). There was significant positive (Mantel test $p = 0.03$) correlation ($r = 0.72$) between the observed and simulated pairwise $G"_{ST}$ values. The map depicts where the corresponding, colored putative populations are

proportionally found in the sampling area and is based on the 100 genotypes population assignments from STRUCTURE analysis at each sampling location. The STRUCTURE plot show a single distinct population group in the southeastern and central part of the State, a second group in the Kodiak area. Note that the PCA plot appears to show 3 clusters, STRUCTURE analysis was only able to identify 2 distinct groups with admixture among them, which did not appear to match the PCA clusters.

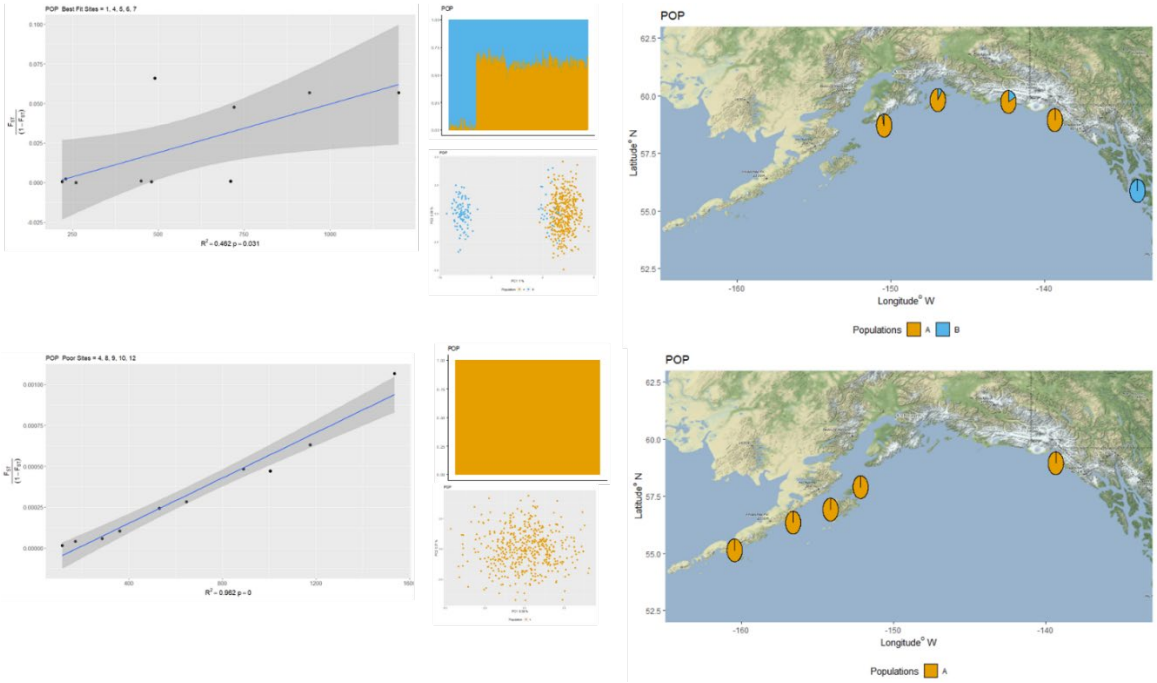


Figure 3-10 Results of the effect of sampling location to determine "best" and "worst" subset of five locations to sample under the POP model. The full model included all 12 population zones with the best subset (top row) had the highest log-likelihood (228.9) when populations 1, 4, 5, 6, 7 are sampled and lowest log likelihood (-2,921.5) when populations 4, 8, 9, 10, 12 are sampled. Note that the worst fit model results in a highly significant apparent IBD relationship, but only a single STRUCTURE derived putative population.

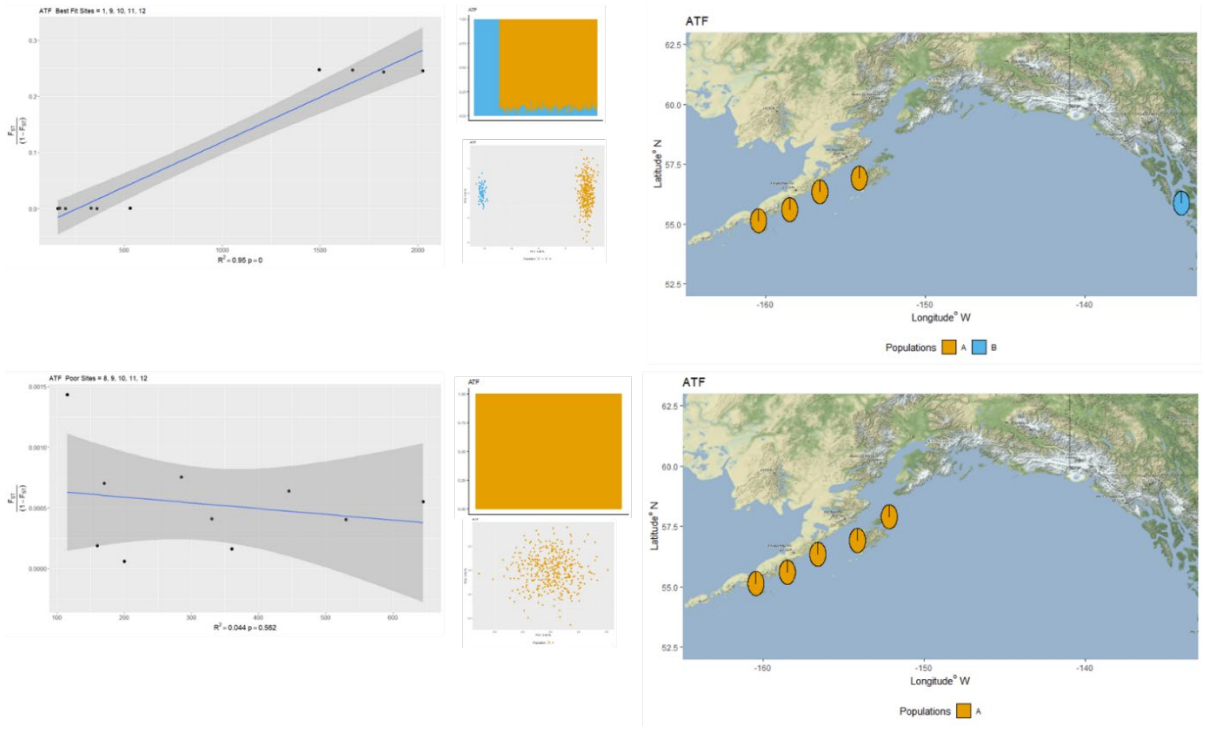


Figure 3-11 Results of effect of sampling locations to determine "best" and "worst" subset of five locations to sample under the ATF model. The full model included all 12 population zones with the best subset (top row) had the highest log-likelihood (104.75) when populations 1, 9, 10, 11, 12 are sampled and lowest log likelihood (-27,796.64) when populations 8, 9, 10, 11, 12 are sampled. The best subset inference would result in identifying two distinct populations groups with little admixture, however, a panmictic population results from the worst sampled zones.

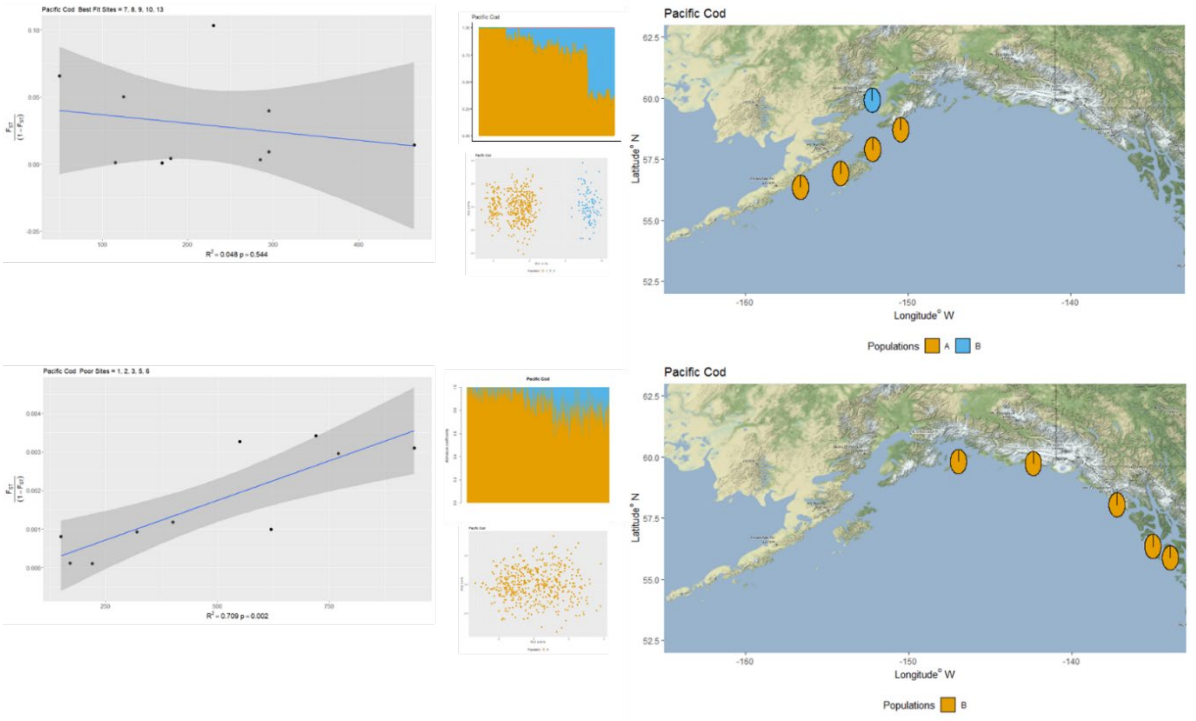


Figure 3-12 Results of the effect of sampling location to determine "best" and "worst" subset of five locations to sample under the Pacific cod model. The full model included all 13 population zones with the best subset (top row) had the highest log-likelihood (223.4) when populations 7, 8, 9, 10, 13 are sampled and lowest log likelihood (-1,331.1) when populations 1, 2, 3, 5, 6 are sampled. The best model would result in a non-significant IBD relationship, but strong PCA and STRUCTURE clustering. In contrast, the worst model results in a significant IBD relationship, but a single cluster detected with PCA, and some STRUCTURE detected admixture from another population which is not evident in the PCA plot.

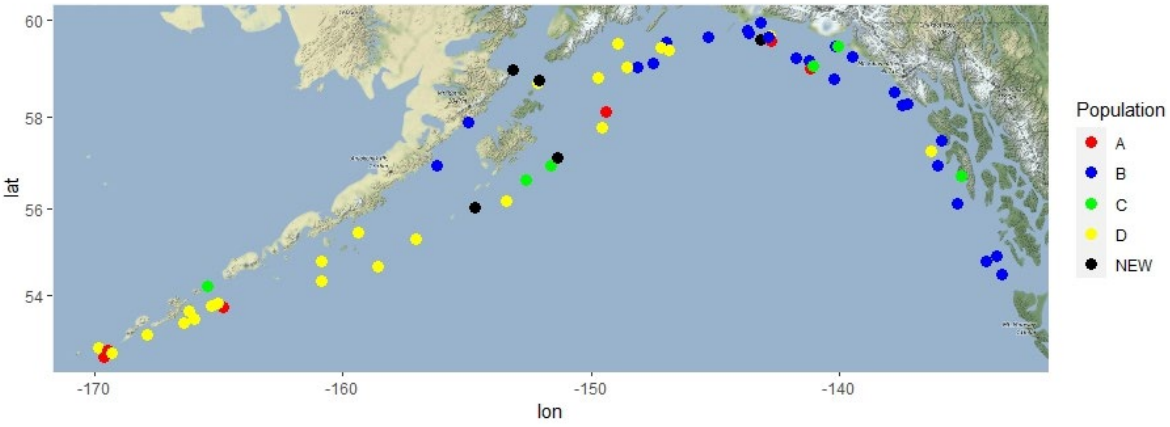


Figure 3-13 Distribution of putative populations based on STRUCTURE analysis of adult POP collections from 2017 and 2019 (Timm, L). Populations A-D correspond to the populations identified in larval aggregates described in chapter 1 of this dissertation.

References

Alcala, N., & Rosenberg, N. A. (2019). , Jost's D, and FST are similarly constrained by allele frequencies: A mathematical, simulation, and empirical study. *Molecular Ecology*, *28*(7), 1624–1636. https://doi.org/10.1111/MEC.15000

Arrowtooth Flounder | NOAA Fisheries. (n.d.). Retrieved March 5, 2022, from https://www.fisheries.noaa.gov/species/arrowtooth-flounder

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://cran.r-project.org/package=gridExtra

Balkenhol, N., Waits, L. P., Dezzani, R. J., Balkenhol, N., & Waits, L. P. (2009). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, *32*. https://doi.org/10.1111/j.1600-0587.2009.05807.x

Baltazar-Soares, M., Hinrichsen, H. H., & Eizaguirre, C. (2018). Integrating population genomics and biophysical models towards evolutionary-based fisheries management. *ICES Journal of Marine Science*, *75*(4), 1245–1257. https://doi.org/10.1093/ICESJMS/FSX244

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035. https://doi.org/10.1111/j.1937-2817.2010.tb01236.x

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., Rochette, R., & Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). *Molecular Ecology*, *25*(20), 5073–5092. https://doi.org/10.1111/mec.13811

Blackiston, D. J., Casey, S. S., & Weiss, E. R. (2008). Retention of Memory through Metamorphosis: Can a Moth Remember What It Learned As a Caterpillar? *PLoS ONE*, *3*(3), 1736. https://doi.org/10.1371/journal.pone.0001736

Blood, D. M., Matarese, A. C., & Busby, M. S. (2007). Spawning, egg development, and early life history dynamics of arrowtooth flounder (Atheresthes stomias) in the Gulf of Alaska. *NOAA Professional Paper NMFS 7*, *NOAA Profe*(May), 28.

Bonanomi, S., Overgaard Therkildsen, N., Retzel, A., Berg Hedeholm, R., Pedersen, M. W., Meldrup, D., Pampoulie, C., Hemmer-Hansen, J., Grønkjaer, P., & Nielsen, E. E. (2016). Historical DNA documents long-distance natal homing in marine fish. *Molecular Ecology*, *25*(12), 2727–2734. https://doi.org/10.1111/mec.13580

Bouwens, K. A., Smith, R. L., Paul, A. J., & Rugen, W. (1999). Length at and Timing of Hatching and Settlement for Arrowtooth Flounders in the Gulf of Alaska. *Alaska Fishery Research Bulletin*, *6*(1). http://www.state.ak.us/adfg/geninfo/pubs/afrb/afrbhome.htm

Bradburd, G. S., & Ralph, P. L. (2019). *Spatial Population Genetics: It's About Time*.

https://doi.org/10.1146/annurev-ecolsys-110316

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, *67*(11), 3258–3273. https://doi.org/10.1111/evo.12193

Bradbury, I., & Laurel, B. (2007). Defining "natal homing" in marine fish populations: comment on Svedäng et al. (2007). *Marine Ecology Progress Series*, *349*, 307–308. https://doi.org/10.3354/meps07281

Bradbury, I. R., & Bentzen, P. (2007). Non-linear genetic isolation by distance: implications for dispersal estimation in anadromousand marine fish populations. *Marine Ecology Progress Series*, *340*, 245–257. https://doi.org/10.3354/MEPS340245

Castellano, S., & Balletto, E. (2002). IS THE PARTIAL MANTEL TEST INADEQUATE? In *Evolution* (Vol. 56, Issue 9).

Clarke, K. Robert, Somerfield, P. J., & Gorley, R. N. (2008). Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *Journal of Experimental Marine Biology and Ecology*, *366*(1–2), 56–69. https://doi.org/10.1016/j.jembe.2008.07.009

Clarke, K R, & Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. In *MARINE ECOLOGY PROGRESS SERIES Mar. Ecol. Prog. Ser* (Vol. 92).

Conrath, C. L., & Knoth, B. (2013). Reproductive Biology of Pacific Ocean Perch in the Gulf of Alaska. *Https://Doi.Org/10.1080/19425120.2012.751941*, *5*(5), 21–27. https://doi.org/10.1080/19425120.2012.751941

Cunningham, K. M., Canino, M. F., Spies, I. B., & Hauser, L. (2009). Genetic isolation by distance and localized fjord population structure in Pacific cod (Gadus macrocephalus): Limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(1), 153–166. https://doi.org/10.1139/F08-199

Debenham, C., Moss, J., & Heintz, R. (2019). *Ecology of age-0 arrowtooth flounder (Atheresthes stomias) inhabiting the Gulf of Alaska.*

Diopere, E., Vandamme, S. G., Hablützel, P. I., Cariani, A., Van Houdt, J., Rijnsdorp, A., Tinti, F., Volckaert, F. A. M., & Maes, G. E. (2018). Seascape genetics of a flatfish reveals local selection under high levels of gene flow. *ICES Journal of Marine Science*, *75*(2), 675–689. https://doi.org/10.1093/icesjms/fsx160

Drinan, D. P., Gruenthal, K. M., Canino, M. F., Lowry, D., Fisher, M. C., & Hauser, L. (2018). Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (Gadus macrocephalus). *Evolutionary Applications*, *11*(8), 1448–1464. https://doi.org/10.1111/eva.12639

Duforet-Frebourg, N., & Slatkin, M. (2016). Isolation-by-distance-and-time in a stepping-stone model. *Theoretical Population Biology*, *108*, 24–35.

https://doi.org/10.1016/j.tpb.2015.11.003

Dunn, J. R., & Matarese, A. C. (1987). A review of the early life history of Northeast Pacific gadoid fishes. *Fisheries Research*, *5*(2–3), 163–184. https://doi.org/10.1016/0165-7836(87)90038-5

Edwards, S. M. (2020). *lemon: Freshing Up your "ggplot2" Plots*. https://cran.r-project.org/package=lemon

Elhaik, E. (2021). Why most Principal Component Analyses (PCA) in population genetic studies are wrong. *BioRxiv*, 2021.04.11.439381. https://doi.org/10.1101/2021.04.11.439381

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). {qgraph}: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*(4), 1–18.

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, *46*, 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Frichot, E., & Francois, O. (2015). {LEA}: an {R} package for {L}andscape and {E}cological {A}ssociation studies. *Methods in Ecology and Evolution*. http://membres-timc.imag.fr/Olivier.Francois/lea.html

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/J.TREE.2012.05.012

Gauch, H. G., Qian, S. I., Piepho, H.-P., Zhou, L., & Chen, R. (2019). *Consequences of PCA graphs, SNP codings, and PCA variants for elucidating population structure*. https://doi.org/10.1371/journal.pone.0218306

Gilbert, K. J. (2016). *Identifying the number of population clusters with STRUCTURE: problems and solutions*. https://doi.org/10.1111/1755-0998.12521

GitHub - wStockhausen/DisMELS: A Java-based framework for developing/running individual-based models(IBMs) for marine species with pelagic early life stages. (n.d.). Retrieved July 2, 2020, from https://github.com/wStockhausen/DisMELS

Guillot, G., & ßois Rousset, F. (2012). *Dismantling the Mantel tests*. https://doi.org/10.1111/2041-210x.12018

Harmon, L. J., & Glor, R. E. (2010). POOR STATISTICAL PERFORMANCE OF THE MANTEL TEST IN PHYLOGENETIC COMPARATIVE ANALYSES. *Evolution*, *64*(7), 2173–2178. https://doi.org/10.1111/j.1558-5646.2010.00973.x

Hedrick, P. W. (2006). Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 67–93. https://doi.org/10.1146/annurev.ecolsys.37.091305.110132

Hinckley, S., Stockhausen, W. T., Coyle, K. O., Laurel, B. J., Gibson, G. A., Parada, C., Hermann, A. J., Doyle, M. J., Hurst, T. P., Punt, A. E., & Ladd, C. (2019). Connectivity between spawning and nursery areas for Pacific cod (Gadus macrocephalus) in the Gulf of Alaska. *Deep Sea Research Part II: Topical Studies in Oceanography*, *165*, 113–126. https://doi.org/10.1016/j.dsr2.2019.05.007

Jakobsson, M., Edge, M. D., & Rosenberg, N. A. (2013a). *The Relationship Between F ST and the Frequency of the Most Frequent Allele*. https://doi.org/10.1534/genetics.112.144758

Jakobsson, M., Edge, M. D., & Rosenberg, N. A. (2013b). The relationship between FST and the frequency of the most frequent allele. *Genetics*, *193*(2), 515–528. https://doi.org/10.1534/genetics.112.144758

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., Knight, S., Knox, J., Long, T., Mannino, M., McGuire, M., Riffle, A., Segelsky, S., Shappell, L., Sterner, A., Strickler, T., & Tursi, R. (2010). A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, *33*(2), no-no. https://doi.org/10.1111/j.1600-0587.2010.06285.x

Kamin, L. M., Palof, K. J., Heifetz, J., & Gharrett, A. J. (2014). Interannual and spatial variation in the population genetic composition of young-of-the-year Pacific ocean perch (Sebastes alutus) in the Gulf of Alaska. *Fisheries Oceanography*, *23*(1), 1–17. https://doi.org/10.1111/fog.12038

Kenchington, E. L. (2003). The effects of fishing on species and genetic diversity. In *Responsible fisheries in the marine ecosystem* (pp. 235–253). CABI. https://doi.org/10.1079/9780851996332.0235

Kimura, M., & Weiss, G. H. (1964). THE STEPPING STONE MODEL OF POPULATION STRUCTURE AND THE DECREASE OF GENETIC CORRELATION WITH DISTANCE. *Genetics*, *49*(4), 561–576. https://doi.org/10.1093/genetics/49.4.561

Knutsen, H., Catarino, D., Rogers, L., Sodeland, M., Mattingsdal, M., Jahnke, M., Hutchings, J. A., Mellerud, I., Espeland, S. H., Johanneson, K., Roth, O., Hansen, M. M., Jentoft, S., André, C., & Jorde, P. E. (2022). Combining population genomics with demographic analyses highlights habitat patchiness and larval dispersal as determinants of connectivity in coastal fish species. *Molecular Ecology*. https://doi.org/10.1111/mec.16415

Laurel, B. J., Hurst, T. P., Copeman, L. A., & Davis, M. W. (2008). The role of temperature on the growth and survival of early and late hatching Pacific cod larvae (Gadus macrocephalus). *Journal of Plankton Research*, *30*(9), 1051–1060. https://doi.org/10.1093/plankt/fbn057

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1), 3258. https://doi.org/10.1038/s41467-018-05257-7

Legendre, P., & Casgrain, Ph. (1994). Modeling Brain Evolution From Behavior: A Permutational Regression Approach. In *Evolution* (Vol. 48, Issue 5).

Leis, J. M., Siebeck, U., & Dixson, D. L. (2011). How Nemo Finds Home: The Neuroecology of Dispersal and of Population Connectivity in Larvae of Marine Fishes. *Integrative and Comparative Biology*, *51*(5), 826–843. https://doi.org/10.1093/icb/icr004

Levin, L. A. (2006). Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, *46*(3), 282–297. https://doi.org/10.1093/icb/icj024

Manel, Stephanie, Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, *20*(3), 136–142. https://doi.org/10.1016/j.tree.2004.12.004

Manel, Stéphanie, & Holderegger, R. (2013). Ten years of landscape genetics. In *Trends in Ecology and Evolution* (Vol. 28, Issue 10, pp. 614–621). Elsevier Current Trends. https://doi.org/10.1016/j.tree.2013.05.012

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, *27*(2 Part 1).

Maselko, J., Andrews, K. R., & Hohenlohe, P. A. (2020). Long-lived marine species may be resilient to environmental variability through a temporal portfolio effect. *Ecology and Evolution*. https://doi.org/10.1002/ece3.6378

Maselko, Jacek, Andrews, K. R., & Hohenlohe, P. A. (2020). Long-lived marine species may be resilient to environmental variability through a temporal portfolio effect. *Ecology and Evolution*, *10*(13), 6435–6448. https://doi.org/10.1002/ece3.6378

McRae, B. H. (2006). ISOLATION BY RESISTANCE. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, *5*(10), 1000686. https://doi.org/10.1371/journal.pgen.1000686

Meirmans, P. G. (2012). The trouble with isolation by distance. In *Molecular Ecology* (Vol. 21, Issue 12, pp. 2839–2846). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. In *Molecular Ecology Resources* (Vol. 11, Issue 1, pp. 5–18). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12), 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. https://cran.r-project.org/package=RColorBrewer

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*(5), 646–649. https://doi.org/10.1038/ng.139

Pacific Cod | NOAA Fisheries. (n.d.). Retrieved March 5, 2022, from https://www.fisheries.noaa.gov/species/pacific-cod

Pacific Ocean Perch | NOAA Fisheries. (n.d.). Retrieved March 5, 2022, from https://www.fisheries.noaa.gov/species/pacific-ocean-perch

Palof, K. J., Heifetz, J., & Gharrett, A. J. (2011). Geographic structure in Alaskan Pacific ocean perch (Sebastes alutus) indicates limited lifetime dispersal. *Marine Biology*, *158*(4), 779–792. https://doi.org/10.1007/s00227-010-1606-2

Pante, E., & Simon-Bouhet, B. (2013). *marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R*. https://doi.org/10.1371/journal.pone.0073051

Pearse, D. E., & Crandall, K. A. (2004). Beyond FST: Analysis of population genetic data for conservation. *Conservation Genetics*, *5*(5), 585–602. https://doi.org/10.1007/s10592-003-1863-4

Pritchard, J. K. (2010). *Software For Inferring Population Structure*. http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.4/html/structure.html

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). *Inference of Population Structure Using Multilocus Genotype Data*. http://www.stats.ox.ac.uk/pritch/home.html.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627. https://doi.org/10.1111/1755-0998.12512

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Raufaste, N., & Rousset, F. O. (2001). ARE PARTIAL MANTEL TESTS ADEQUATE? In *BRIEF COMMUNICATIONS Evolution* (Vol. 55, Issue 8).

Reich, D., Price, A. L., & Patterson, N. (2008). Principal component analysis of genetic data. *Nature Genetics 2008 40:5*, *40*(5), 491–492. https://doi.org/10.1038/ng0508-491

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005). Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure. *PLoS Genetics*, *1*(6), e70. https://doi.org/10.1371/journal.pgen.0010070

Rousset. (2002). PARTIAL MANTEL TESTS: REPLY TO CASTELLANO AND BALLETTO. *Https://Doi.Org/10.1554/0014-3820(2002)056[1874:PMTRTC]2.0.CO;2*, *56*(9), 1874–1875. https://doi.org/10.1554/0014

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228. https://doi.org/10.1093/genetics/145.4.1219

Schiavina, M., Marino, I. A. M., Zane, L., & Melià, P. (2014). Matching oceanography and genetics at the basin scale. Seascape connectivity of the Mediterranean shore crab in the Adriatic Sea. *Molecular Ecology*, *23*(22), 5496–5507. https://doi.org/10.1111/mec.12956

Schroeder, I. D., Santora, J. A., Moore, A. M., Edwards, C. A., Fiechter, J., Hazen, E. L., Bograd, S. J., Field, J. C., & Wells, B. K. (2014). Application of a data-assimilative regional ocean modeling system for assessing California Current System ocean conditions, krill, and juvenile rockfish interannual variability. *Geophysical Research Letters*, *41*(16), 5942–5950. https://doi.org/10.1002/2014GL061045

Schunter, C., Carreras-Carbonell, J., MacPherson, E., TintorÉ, J., Vidal-Vijande, E., Pascual, A., Guidetti, P., & Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. https://doi.org/10.1111/j.1365-294X.2011.05355.x

Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., Von Der Heyden, S., & Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, *554*(July), 1–19. https://doi.org/10.3354/meps11792

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*(1), 1–15. https://doi.org/10.1111/evo.12258

Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, *9*(4), 347–404. https://doi.org/10.1016/j.ocemod.2004.08.002

Slatkin, M. (1980). THE DISTRIBUTION OF MUTANT ALLELES IN A SUBDIVIDED POPULATION. *Genetics*, *95*(2).

Slatkin, M. (1993). Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*. https://doi.org/10.2307/2410134

Slatkin, M., & Barton, N. H. (1989). *A Comparison of Three Indirect Methods for Estimating Average Levels of Gene Flow* (Vol. 43, Issue 7).

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology*, *35*(4), 627. https://doi.org/10.2307/2413122

Sokal, R. R. (1979). Testing Statistical Significance of Geographic Variation Patterns. *Systematic Zoology*, *28*(2), 227. https://doi.org/10.2307/2412528

Spies, I., & Punt, A. E. (2015). The utility of genetics in marine fisheries management: A

simulation study based on Pacific cod off Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(9), 1415–1432. https://doi.org/10.1139/CJFAS-2014-0050

Stephenson, R., Kenchington, E., Branch, S., & Canada, O. (2000). International Council for the Exploration of the Sea CM 2000/Mini:07 Defining the Role of ICES in Supporting Biodiversity Conservation. *ICES Journal of Marine Science*.

Stewart, J. R., & Lister, A. M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, *16*(11), 608–613. https://doi.org/10.1016/S0169-5347(01)02338-2

Stockhausen, W. (2007). *Modeling Larval Dispersion of Rockfish: A Tool for Marine Reserve Design?* https://doi.org/10.4027/bamnpr.2007.15

Stockhausen, W. T., Coyle, K. O., Hermann, A. J., Blood, D., Doyle, M. J., Gibson, G. A., Hinckley, S., Ladd, C., & Parada, C. (2019). Running the gauntlet: Connectivity between spawning and nursery areas for arrowtooth flounder (Atheresthes stomias) in the Gulf of Alaska, as inferred from a biophysical individual-based model. *Deep Sea Research Part II: Topical Studies in Oceanography*, *165*, 127–139. https://doi.org/10.1016/j.dsr2.2018.05.017

Stockhausen, W. T., Coyle, K. O., Hermann, A. J., Doyle, M., Gibson, G. A., Hinckley, S., Ladd, C., & Parada, C. (2019). Running the gauntlet: Connectivity between natal and nursery areas for Pacific ocean perch (Sebastes alutus) in the Gulf of Alaska, as inferred from a biophysical individual-based model. *Deep Sea Research Part II: Topical Studies in Oceanography*, *165*, 74–88. https://doi.org/10.1016/j.dsr2.2018.05.016

Storfer, A., Murphy, Melanie A. Spear, Stephen F. Holderegger, R., & Waits, L. P. (2010). Landscape genetics: where are we now? *Molecular Ecology*, *19*(17), 3496–3514. https://doi.org/10.1111/j.1365-294X.2010.04691.x

Sunday, J. M., Calosi, P., Dupont, S., Munday, P. L., Stillman, J. H., & Reusch, T. B. H. (2014). Evolution in an acidifying ocean. *Trends in Ecology & Evolution*, *29*(2). https://doi.org/10.1016/j.tree.2013.11.001

Svedäng, H., Righton, D., & Jonsson, P. (2007). Migratory behaviour of Atlantic cod Gadus morhua: natal homing is the prime stock-separating mechanism. *Marine Ecology Progress Series*, *345*, 1–12. https://doi.org/10.3354/MEPS07140

Thomson, J. A. (1962). On the Fecundity of Pacific Cod (Gadus macrocephalus Tilesius) from Hecate Strait, British Columbia. *Journal of the Fisheries Research Board of Canada*, *19*(3), 497–500. https://doi.org/10.1139/f62-028

Varvio, S. L., Chakraborty, R., & Nei, M. (1986). Genetic variation in subdivided populations and conservation genetics. *Heredity*, *57*(2), 189–198. https://doi.org/10.1038/hdy.1986.109

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Waples, R. S., & Gaggiotti, O. (2006). INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419–1439. https://doi.org/10.1111/J.1365-294X.2006.02890.X

Warton, D. I., Duursma, R. A., Falster, D. S., & Taskinen, S. (2012). SMATR 3-an R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution*, *3*, 257–259. https://doi.org/10.1111/j.2041-210X.2011.00153.x

Werner, F. E., Page, F. H., Lynch, D. R., Loder, J. W., Lough, R. G., Perry, R. I., Greenberg, D. A., & Sinclair, M. M. (1993). Influences of mean advection and simple behavior on the distribution of cod and haddock early life stages on Georges Bank. *Fisheries Oceanography*, *2*(2), 43–64. https://doi.org/10.1111/j.1365-2419.1993.tb00120.x

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration: FST≠1/(4Nm+1). *Heredity*, *82*(2), 117–125. https://doi.org/10.1038/sj.hdy.6884960

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

Wickham, H., & Seidel, D. (2020). *scales: Scale Functions for Visualization*. https://cran.r-project.org/package=scales

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."* https://cran.r-project.org/package=cowplot

Withler, R. E., Beacham, T. D., Schulze, A. D., Richards, L. J., & Miller, K. M. (2001). Co-existing populations of Pacific ocean perch, Sebastes alutus, in Queen Charlotte Sound, British Columbia. *Marine Biology*, *139*(1), 1–12. https://doi.org/10.1007/S002270100560

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*(4), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

WRIGHT, S. (1946). Isolation by distance under diverse systems of mating. *Genetics*, *31*(January), 39–59.

Wright, Sewall. (1943). ISOLATION BY DISTANCE*. *Genetics*, *28*, 114–138.

Yu, G. (2021). *ggplotify: Convert Plot to "grob" or "ggplot" Object*. https://cran.r-project.org/package=ggplotify

Appendix 1.

Deriving parameter estimation of the Beta(α,β) distributed G_{ST} values:

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

$$\alpha = \mu\alpha + \mu\beta$$

$$\mu\beta = \alpha - \mu\alpha$$

$$\beta = \frac{\alpha(1 - \mu)}{\mu} = \alpha\left(\frac{1}{\mu} - 1\right)$$

Substituting for β:

$$\sigma^2 = \frac{\alpha\alpha\left(\frac{1}{\mu} - 1\right)}{\left(\alpha + \alpha\left(\frac{1}{\mu} - 1\right)\right)^2\left(\alpha + \alpha\left(\frac{1}{\mu} - 1\right) + 1\right)}$$

$$\sigma^2 = \frac{\alpha^2\left(\frac{1}{\mu} - 1\right)}{\left(\alpha\left(\frac{1}{\mu}\right)\right)^2\left(\alpha\left(\frac{1}{\mu}\right) + 1\right)}$$

$$\sigma^2 = \frac{\alpha^2\left(\frac{1}{\mu} - 1\right)}{\alpha^2\left(\frac{1}{\mu}\right)^2\left(\alpha\left(\frac{1}{\mu}\right) + 1\right)}$$

$$\sigma^2 = \frac{\left(\frac{1}{\mu} - 1\right)}{\left(\frac{1}{\mu}\right)^2\left(\alpha\left(\frac{1}{\mu}\right) + 1\right)}$$

$$\sigma^2 = \frac{\left(\frac{1}{\mu} - 1\right)\mu^2}{\frac{\alpha}{\mu} + 1}$$

$$\sigma^2\left(\frac{\alpha}{\mu} + 1\right) = \left(\frac{1}{\mu} - 1\right)\mu^2$$

$$\frac{\alpha}{\mu} + 1 = \frac{\left(\frac{1}{\mu} - 1\right)\mu^2}{\sigma^2}$$

$$\alpha = \frac{\left(\frac{1}{\mu} - 1\right)\mu^3}{\sigma^2} - \mu$$

$$\alpha = \frac{(1-\mu)\mu^2}{\sigma^2} - \mu$$

Substituting for α:

$$\beta = \left(\frac{(1-\mu)\mu^2}{\sigma^2} - \mu\right)\left(\frac{1}{\mu} - 1\right)$$

$$\beta = \mu\left(\frac{(1-\mu)\mu}{\sigma^2} - 1\right)\left(\frac{1}{\mu} - 1\right)$$

$$\beta = \left(\frac{(1-\mu)\mu}{\sigma^2} - 1\right)(1-\mu)$$

$$\beta = \frac{(1-\mu)^2\mu}{\sigma^2} - 1 + \mu$$