Decoupling contamination and hybridization to characterize population structure in Puget Sound rockfishes (*Sebastes* spp.)

Liz Boggs

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2026

Committee:

Lorenz Hauser

Krista Nichols

Tim Essington

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

University of Washington

Abstract

Decoupling contamination and hybridization to characterize population structure in Puget Sound rockfishes (*Sebastes* spp.)

Liz Boggs

Chair of the Supervisory Committee:
Professor Lorenz Hauser
School of Aquatic and Fishery Sciences

Genetic population structure analysis can provide crucial information about marine species' connectivity that is key to effective management and recovery efforts. Natural hybridization can bias patterns of population structure, although advances in molecular technologies now allow individual-based and genome-wide analyses that facilitate the detection of such interactions. Nevertheless, even advanced molecular genetic methods are still highly sensitive to errors during the processing and bioinformatic filtering of data, including cross-sample contamination and initial data filtering. Signals of contamination are exceptionally difficult to identify especially in hybridizing species, especially when processed and sequenced simultaneously. Here, we investigated the effect of hybridization on spatial genetic structure in three hybridizing rockfish species – Brown Rockfish (*Sebastes auriculatus*), Copper Rockfish (*S. caurinus*), and Quillback Rockfish (*S. maliger*) – in Puget Sound, Washington. Initial analyses indicated potential cross-sample contamination, so we first had to detect and remove

contaminated individuals. In Chapter 1, we crafted a method to evaluate individuals for signatures of contamination using allelic balance, or the ratio of reads supporting reference and alternate alleles in heterozygous genomic positions. This method is independent of typical hybrid signals, removing only the individuals with extreme deviation in allelic balance metrics. We also found that minor allele frequency filtering has substantial impact on a dataset, demonstrating that retention of low-frequency alleles may result in weak population differentiation. In Chapter 2, we used remaining 'clean' samples to investigate genetic population structure. Hybridization was generally widespread at low levels in Puget Sound and was asymmetric from Quillback Rockfish into Brown Rockfish and Copper Rockfish. We further confirmed Admiralty Inlet as an important bathymetric barrier that isolates populations and retains hybrids in Puget Sound but, in contrast to earlier studies, found no evidence for increased incidence of hybridization in peripheral basins of Puget Sound. However, we detected a genetically differentiated population of Copper Rockfish in Hood Canal. Altogether, our results demonstrated genetic structure and limited dispersal in all species, some but not all of which was due to hybridization. Our results have meaningful implications for future management and conservation of these species, especially concerning the design of species recovery efforts and any potential designation of spatial management units.

TABLE OF CONTENTS

LIST OF FIGURES iii
LIST OF TABLES v
LIST OF SUPPLEMENTAL FIGURES vi
LIST OF SUPPLEMENTAL TABLES vii
Chapter 1. IDENTIFICATION AND REMOVAL OF CROSS-CONTAMINATED SAMPLES BASED ON ALLELIC BALANCE IN A RAD-SEQ DATASET 1
1.1 ABSTRACT 1
1.2 INTRODUCTION 2
1.3 METHODS 6
1.3.1 Sampling 6
1.3.2 DNA Extraction, RAD-Seq Library Preparation, and Sequencing 6
1.3.3 Data Processing and Initial Filtering 7
1.3.4 Analysis 9
1.4 RESULTS 12
1.4.1 Rockfish Interspecific Dataset 12
1.4.2 MAF versus MAC Filtering for Intraspecific Datasets 15
1.4.3 Intraspecific Datasets 16
1.5 DISCUSSION 19
1.6 FIGURES 27
1.7 TABLES 51
1.8 SUPPLEMENTAL FIGURES 59
1.9 REFERENCES 61
Chapter 2. INFLUENCE OF PUGET SOUND BASIN STRUCTURE AND HYBRIDIZATION ON POPULATION GENETICS OF THREE PACIFIC ROCKFISH SPECIES (*Sebastes* spp.) 69
2.1 ABSTRACT 69
2.2 INTRODUCTION 70
2.3 METHODS 74
2.3.1 Sampling 74

2.3.2 DNA Extraction, RAD-Seq Library Preparation, and Sequencing *74*
2.3.3 Data Processing and Initial Filtering *75*
2.3.4 Analysis *77*
2.4 RESULTS 78
2.4.1 Interspecific Dataset *78*
2.4.2 Intraspecific Datasets *81*
2.5 DISCUSSION 84
2.5.1 Hybrid ancestry *85*
2.5.2 Intraspecific structure influenced by hybridization and barriers to dispersal *88*
2.6 FIGURES 93
2.7 TABLES 104
2.8 SUPPLEMENTAL FIGURES 107
2.9 SUPPLEMENTAL TABLES 112
2.10 REFERENCES 113

LIST OF FIGURES

Figure 1.1 Map of rockfish sampling regions in the Salish Sea 27
Figure 1.2 Allelic balance of heterozygous loci in clean versus contaminated samples 28
Figure 1.3 Interspecific Brown Rockfish, Copper Rockfish, and Quillback Rockfish contamination filtering analysis 29
Figure 1.4 Interspecific Brown Rockfish, Copper Rockfish, and Quillback Rockfish PCA at unfiltered, mild, and aggressive filtering levels 30
Figure 1.5 Interspecific Brown Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at unfiltered, mild, and aggressive filtering levels 31
Figure 1.6 Interspecific Copper Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at unfiltered, mild, and aggressive filtering levels 32
Figure 1.7 Interspecific Quillback Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at unfiltered, mild, and aggressive filtering levels 33
Figure 1.8 Minor allele count (MAC) versus minor allele frequency (MAF) filtering results in STRUCTURE analysis 34
Figure 1.9 Pacific Herring (*Clupea pallasii*) contamination filtering analysis 35
Figure 1.10 Brown Rockfish contamination filtering analysis 36
Figure 1.11 Brown Rockfish PCA at unfiltered, mild, and aggressive filtering levels 37
Figure 1.12 Brown Rockfish individual observed heterozygosity per sampling group at unfiltered, mild, and aggressive filtering levels 38
Figure 1.13 Brown Rockfish STRUCTURE analysis at K = 2, grouped by sampling population at unfiltered, mild, and aggressive filtering levels 39
Figure 1.14 Delta K and Mean LnP(K) (±1 SD) for Brown Rockfish STRUCTURE analysis 40
Figure 1.15 Copper Rockfish contamination filtering analysis 41
Figure 1.16 Copper Rockfish PCA at unfiltered, mild, and aggressive filtering levels 42
Figure 1.17 Copper Rockfish individual observed heterozygosity per sampling group at unfiltered, mild, and aggressive filtering levels 43
Figure 1.18 Copper Rockfish STRUCTURE analysis at K = 2, grouped by sampling population at unfiltered, mild, and aggressive filtering levels 44
Figure 1.19 Delta K and Mean LnP(K) (±1 SD) for Copper Rockfish STRUCTURE analysis ...45

Figure 1.20 Quillback Rockfish contamination filtering analysis 46
Figure 1.21 Quillback Rockfish PCA at unfiltered, mild, and aggressive filtering levels.....47
Figure 1.22 Quillback Rockfish individual observed heterozygosity per sampling group at unfiltered, mild, and aggressive filtering levels.....48
Figure 1.23 Quillback Rockfish STRUCTURE analysis at K = 2, grouped by sampling population at unfiltered, mild, and aggressive filtering levels.....49
Figure 1.24 Delta K and Mean LnP(K) (±1 SD) for Quillback Rockfish STRUCTURE analysis.....50
Figure 2.1 Map of rockfish sampling regions in the Salish Sea 93
Figure 2.2 STRUCTURE and PCA results for interspecific rockfish analysis.....94
Figure 2.3 Principal component analysis and STRUCTURE analysis for Brown Rockfish 95
Figure 2.4 Principal component analysis and STRUCTURE analysis for Copper Rockfish 96
Figure 2.5 Copper Rockfish PCA featuring only individuals with <1% hybrid ancestry.....97
Figure 2.6 Copper Rockfish interspecific admixture versus Weir and Cockerham pairwise F_{ST} 98
Figure 2.7 Copper Rockfish interspecific hybrid ancestry versus intraspecific ancestry from Georgia Basin cluster.....99
Figure 2.8 Principal component analysis and STRUCTURE analysis for Quillback Rockfish ..100
Figure 2.9 Brown Rockfish K = 3 STRUCTURE analysis distribution in Puget Sound and West Coast 101
Figure 2.10 Copper Rockfish STRUCTURE analysis distribution in Puget Sound at K = 4.....102
Figure 2.11 Quillback Rockfish K = 2 STRUCTURE analysis distribution in Puget Sound.....103

LIST OF TABLES

Table 1.1 Interspecific Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels....51
Table 1.2 Interspecific rockfish average hybrid ancestry (HA) per species....51
Table 1.3 MAC versus MAF filtering observed heterozygosity results....52
Table 1.4 Brown Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels....53
Table 1.5 Summary statistics for Brown Rockfish at unfiltered, mild, and aggressive contamination filtering levels....54
Table 1.6 Copper Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels....55
Table 1.7 Summary statistics for Copper Rockfish at unfiltered, mild, and aggressive contamination filtering levels....56
Table 1.8 Quillback Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels....57
Table 1.9 Summary statistics for Quillback Rockfish at unfiltered, mild, and aggressive contamination filtering levels....58
Table 2.1 Interspecific rockfish average hybrid ancestry per sampling location and species....104
Table 2.2 Summary statistics for Brown, Copper, and Quillback Rockfish....104
Table 2.3 Brown Rockfish pairwise Weir and Cockerham F_{ST} estimates....105
Table 2.4 Copper Rockfish pairwise Weir and Cockerham F_{ST} estimates....105
Table 2.5 Copper Rockfish pairwise Weir and Cockerham F_{ST} estimates with only pure individuals (<1% hybrid ancestry)....105
Table 2.6 Quillback Rockfish pairwise Weir and Cockerham F_{ST} estimates....106

LIST OF SUPPLEMENTAL FIGURES

Supplemental Figure 1.1 Brown Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at unfiltered, mild, and aggressive filtering levels.....59
Supplemental Figure 1.2 Copper Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at unfiltered, mild, and aggressive filtering levels.....60
Supplemental Figure 2.1 Delta K and Mean LnP(K) (±1 SD) for Brown Rockfish STRUCTURE analysis.....107
Supplemental Figure 2.2 Delta K and Mean LnP(K) (±1 SD) for Copper Rockfish STRUCTURE analysis.....108
Supplemental Figure 2.3 Delta K and Mean LnP(K) (±1 SD) for Quillback Rockfish STRUCTURE analysis 109
Supplemental Figure 2.4 Distribution of samples per species.....110
Supplemental Figure 2.5 Copper Rockfish STRUCTURE analysis distribution in Puget Sound at K = 2, K = 3, K = 4, and K = 5 111

LIST OF SUPPLEMENTAL TABLES

Supplemental Table 2.1 Genotype discordance rates between interspecific dataset replicates ... 112

ACKNOWLEDGEMENTS

You never know how exactly a study will unfold until you are in the midst of it, and with that uncertainty may come fascinating discoveries. I am indebted to everyone who has helped me during this project, with all its twists and turns, whether through hands-on support or words of encouragement from many states away. I'd first like to thank Lorenz Hauser, my advisor, for guiding me through a very interesting and often tricky study – I can think of no better way to become a well-rounded scientist than to venture into unexplored territory. Your consistent support and feedback along every step of the way helped me grow my scientific skills exponentially, and for that I am immensely appreciative.

A big thank you to my committee members Krista Nichols and Tim Essington for your helpful edits and advice, especially when my project took a substantial pivot. To the teams at WDFW, NOAA, and DFO Canada, I am grateful for your help in the nascent stages of my project as I gathered as many rockfish fin clips as I could (and you absolutely delivered), with a special thank you to the WDFW groundfish team for taking me on a sampling trip (the rockfish unfortunately had other ideas that day). I would also like to thank Washington Sea Grant, the School of Aquatic and Fishery Sciences, and the Puget Sound Anglers Fidalgo San Juan chapter for funding this project, as it could not have happened without your support.

During my time at UW, I have been surrounded by so many bright, creative, and determined people in the SAFS department. To my fellow MERlab members, and in particular Anna Simeon, Bryan Briones Ortiz, and Chris Setzke, I thank you for helping me untangle my data and making me confident in my lab skills. Finally, to my friends, family, and husband Anthony – thank you for your endless love and support through this project, and for listening to me talk about "my rockfish" for several years. You are all honorary genetics experts now.

Chapter 1. IDENTIFICATION AND REMOVAL OF CROSS-CONTAMINATED SAMPLES BASED ON ALLELIC BALANCE IN A RAD-SEQ DATASET

1.1 ABSTRACT

Cross-sample contamination can confound population genetics analyses by artificially altering metrics commonly used to investigate population structure and hybridization. Several tools exist to identify contamination in human genomic sequencing data, but frequently require complete reference genomes, known population allele frequencies, targeted genomic regions, and/or high genomic coverage. Researchers working on non-model species do not typically have access to this level of information but still need to detect and remove cross-contamination from their datasets; as these data are often used for policy and management decisions, it is especially imperative to avoid false or inflated signals of structure and hybridization. This study aimed to 1) identify cross-contaminated samples (both between- and within-species) based on heterozygotic allele balance in reduced representation sequencing data, and 2) evaluate the effect of the filtering method on several common population genetics analyses. The method is demonstrated in this study using RAD-seq data from three closely related marine fish species that are known to hybridize – Brown Rockfish, Copper Rockfish, and Quillback Rockfish (*Sebastes* spp.) – as well as a published Pacific Herring (*Clupea pallasii*) dataset with identified contaminated individuals. We did detect contamination in some samples, but the effect on qualitative results was limited: patterns in PCA, STRUCTURE analysis, and heterozygosity remained largely consistent across all levels of contamination filtering. On the other hand, minor allele filtering had strong effects: the same datasets filtered using minor allele frequency (MAF) versus minor allele count (MAC)

showed substantial difference in STRUCTURE and heterozygosity results. These results demonstrate that minor allele filtering parameter choices may affect population structure inference more drastically than low-level cross-contamination. We conclude with recommendations for the detection and removal of contamination in routine RAD-seq applications.

1.2 INTRODUCTION

Cross-sample contamination is the unintentional transfer of sample material (e.g. tissue, DNA) from one sample to another. Identifying and removing cross-sample contamination in genetic sequencing data is an important aspect of producing a reliable dataset for population genetics analysis, as contamination can affect population structure and heterozygosity estimates (Petrou et al., 2019; Jun et al., 2012). The process by which this kind of contamination is flagged and eliminated is, however, currently ill-defined and limited in methods for studies involving wildlife and non-model species. While there are several programs used to identify contamination in sequence data (see comparison in Chen et al., 2024), most have been written for human whole genome sequencing (WGS) datasets and require data not available for most non-model species. For example, one of the most popular methods used in clinical sequencing, VerifyBamID (Jun et al., 2012), requires established population allele frequencies – information that is easily accessible for human studies (e.g. 1000 Genomes Project, Auton et al., 2015), but largely absent in wildlife genetics. Reduced-representation sequencing approaches (RRS), such as restriction site associated DNA sequencing (RAD-seq), provide thousands of loci for a fraction of the cost of WGS; however, approaches designed for WGS data may result in an inaccurate assessment of contamination level when used with RRS data due to the coverage difference in sequence data

when compared to WGS (Andrews et al. 2016), further narrowing the options available for accurate cross-contamination analysis.

This kind of contamination can be introduced in virtually every step of the data generation process. When sampling individuals in a field setting, tissue material may be transferred from one sample into the next. Samples are then handled and transferred multiple times in a laboratory setting; DNA extraction, sample normalization (dilution and/or concentration), and library preparation all present opportunities for accidental transfer of sample material or DNA between plate wells. Contamination in the form of index hopping (reads assigned to the incorrect index) can also occur during sequencing, but is typically low with recent improvements to NGS technologies, reportedly in the range of 0.1-2% of reads for current patterned flow cell sequencing systems (Illumina, Inc., 2017).

Interspecific hybridization can complicate the process of identifying contaminated samples in a multi-species dataset. Widespread contamination in non-hybridizing interspecific datasets is generally easier to identify, as samples will often cluster with species that are different from their own in principal component analysis (PCA), though this could also be due to morphological misidentification or sample mislabeling and is not always indicative of cross-contamination. In non-hybridizing intraspecific (single species) datasets, contamination can be identified from inflated observed heterozygosity (Petrou et al., 2019; Jun et al., 2012). When the dataset comprises species that are known to hybridize with each other, teasing apart contaminated samples from true hybrids can be difficult – inflated heterozygosity, ancestry estimates, and PCA cluster outliers/intermediates (individuals situated between species clusters) are also how potential hybrid individuals are commonly identified in hybridization studies (Payseur and Rieseberg, 2016). Our main three study species – Brown Rockfish (*Sebastes*

auriculatus), Copper Rockfish (*Sebastes caurinus*), and Quillback Rockfish (*Sebastes maliger*) – are known to hybridize with each other in Puget Sound (Wray et al., 2024; Schwenke et al., 2018; Buonaccorsi et al., 2002, 2005; Seeb, 1998). Thus, a contamination detection method that is independent of the effects of hybridization is needed to accurately delineate between hybridization signals and cross-contamination.

Allelic balance refers to the ratio of sequencing reads that support the reference and alternative alleles in a heterozygous site – this ratio is expected to be approximately 50:50. This concept has been utilized specifically for contamination detection, including two published studies (Dallavilla et al., 2020; Fiévet et al., 2019) and three preprints on BioRxiv (Jecha et al., updated 2025; Delafoy et al., posted 2021; McCartney-Melstad et al., posted 2021). It has also been used as the basis for a tool designed to detect systemic genotyping error and false positive variant calls (Muyas et al., 2018). When applied to contamination detection, it can be used to detect deviations from the expected 50:50 ratio at heterozygous sites. Contamination shifts allele balance away from 50:50 when the genotype of the contaminating sample differs from the target sample, for example, when a homozygous sample contaminates a heterozygous sample. Such a shift may also be expected in species with a recently duplicated genome or paralogous loci, but shifts would be concentrated on few loci or genome regions. Thus, detectable and consistent shifts in allelic balance from 50:50 across heterozygous sites could indicate contamination. This method is of particular interest for studies involving hybridizing species, as it could avoid unintentional elimination of true hybrids because of inflated heterozygosity.

Another important consideration in population genetics methodology design is the choice of filtering strategy, as different combinations of filtering thresholds can influence downstream results (O'Leary et al., 2018). Filtering is typically applied both at the locus level (removing loci

based on coverage and/or missingness, both within individuals and across a population) and the individual level (removing samples with high missing data). A common locus-level filter involves the minor allele, which is the less frequent allele at a given genomic position. Studies often use a minor allele frequency (MAF) threshold set at 0.05, which removes variants with a frequency of 5% or below in the dataset. These low-frequency variants could be instances of sequencing error or recent mutations that may affect signals of population differentiation (Linck and Battey, 2019). An alternative minor allele filter strategy is minor allele count (MAC), wherein alleles observed fewer than a minimum number of times across all individuals are excluded (commonly set at three or four). MAC-based filtering has been suggested as a way to retain rare variants while still removing singleton alleles that can bias analyses (Linck and Battey, 2019; Shafter et al., 2017). As a result, selecting appropriate filtering parameters to mitigate both bias and noise could be as crucial as mitigating contamination, but a direct comparison between the effects of standard filtering schemata and contamination filtering has not yet been explored.

The goal of this study was to craft and execute an analysis for identifying and removing samples with cross-sample contamination prior to downstream analysis, and to compare the effect of two common minor allele filtering strategies on population genetics analyses. Our contamination filtering method was applied to one interspecific and three intraspecific *Sebastes* spp. RAD-seq datasets, as well as a Pacific Herring (*Clupea pallasii*) dataset in which approximate contamination levels have been established by a previous study (Petrou et al., 2019) to validate the method.

1.3 METHODS

1.3.1 Sampling

389 fin clip and muscle tissue samples of Brown Rockfish, Copper Rockfish, and Quillback Rockfish were assembled from collections at the Washington Department of Fish and Wildlife (WDFW), the National Oceanic and Atmospheric Administration (NOAA), and Fisheries and Oceans Canada (DFO). Samples were stored either in 95% ethanol tubes and kept refrigerated, or as a dry fin clip in an absorbent paper container inside a plastic bag with moisture-absorbing beads stored at room temperature. Nine individuals sequenced in this study were marked as "possible hybrids" by the original collectors in the field, as the whole specimens appeared to have phenotypic traits (usually coloration) that were not consistent with a single species. Samples were assigned to one of seven regions of the coast or Salish Sea (Figure 1.1) based on their capture coordinates or approximate location (i.e. named sampling location). To estimate genotyping error and check for data comparability, 28 individuals from Wray et al. (2024) (hereafter referred to AW) were re-sequenced, and 70 new individuals (11 Brown Rockfish, 23 Copper Rockfish, and 36 Quillback Rockfish) were sequenced twice; libraries of 26 individuals were prepared in two different libraries from the same DNA extract, and 44 individuals were extracted in two separate extraction plates.

1.3.2 DNA Extraction, RAD-Seq Library Preparation, and Sequencing

DNA was extracted from samples using a Nexttec DNA Isolation Kit (nexttec™ Biotechnologie GmbH, Leverkusen, Germany). Within each of our five 96-well extraction plates, a single random well was used as a negative control to check for contamination. This consisted of a no-template control (NTC) method, where all protocol reagents are still added to the well,

but no DNA is added. DNA was quantified on a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA). Extractions were normalized to 125ng in 10µl and assorted into five libraries; four libraries contained randomized sample groups to avoid batch effect (Leigh et al., 2018), while the fifth library contained only samples of poor quality and our five negative controls to avoid amplification bias during the PCR stage. Library preparation followed an adapted version of existing RAD-seq protocols (Ali et al., 2016) known as BestRAD. Genomic DNA was digested with *Sbf*I, and a P1 adapter containing a forward amplification primer, an Illumina primer, and an individual-specific 6 bp barcode was ligated to the fragments at the restriction site end. Fragments were then sheared randomly via sonication in a Diagenode Bioruptor 300 (Hologic Inc., Marlborough, MA) and size-selected to a length of 300-600 bp. Finally, P2 adapters were then ligated to the reverse-end and libraries were amplified via PCR. Each library was quality checked on a 1% agarose gel post-PCR and a 2100 Expert Bioanalyzer (Agilent Technologies, Santa Clara, CA) when completed. Libraries were then sent to the University of Oregon for equimolar pooling and sequencing on a single NovaSeq 6000 S4 lane (paired end, 150 bp).

1.3.3 Data Processing and Initial Filtering

Raw sequences were quality checked and visualized using FastQC v0.12.0 (Andrews, 2010) and MultiQC (Ewels et al., 2016). Sequences were demultiplexed, assigned genotypes and filtered for quality using the STACKS v2.64 process_radtags program (Catchen et al., 2013; Rochette et al., 2019). PCR duplicates were removed using the STACKS clone_filter program. Sequences were then aligned using Bowtie 2 (Langmead and Salzberg, 2012) to the Honeycomb Rockfish genome (*Sebastes umbrosus*; NCBI BioProject accession number: PRJNA562006), as

it is the closest relative to our three species of *Sebastes* that had a chromosome-level genome assembly available at the time of this study. After alignment, individuals with fewer than 300,000 reads were removed based on our within-dataset distribution of read counts to exclude poorly sequenced samples, as their inclusion may cause higher rates of missing data and genotyping error (O'Leary et al., 2018). BAM files generated in this study were combined into a dataset with BAM files from the AW study (retaining only the samples that were also >300,000 reads; 37 Brown Rockfish, 84 Copper Rockfish, and 86 Quillback Rockfish), with replicate samples as separate files to check for differences in genotyping error between the different sequencing runs. This produced a dataset of 542 samples: 105 Brown Rockfish, 201 Copper Rockfish, and 236 Quillback Rockfish (including all replicates). SNP calling and population genetics statistics were calculated using STACKS *gstacks* and *populations* modules for both interspecific and intraspecific datasets.

For the interspecific data set, individuals were assigned to their morphological species identification in population assignment, and all three species were processed together. Data were filtered in the *populations* module to include loci present in >80% individuals (flag R80), in at least one collection region (P1) and only the first SNP per RAD tag (*write-single-snp)*. We chose the first SNP on each RAD tag, because SNPs with the highest F_{ST} may select SNPs with fixed differences between two of the three species and limited differentiation between the other two species. For the three intraspecific datasets**,** the SNP with the highest minor allele frequency within each species was chosen. We chose this method to keep our dataset as comparable as possible to AW, who used this strategy specifically to select different SNPs for intra- and interspecific analyses (Wray et al., 2024). SNPs that were significantly out of Hardy-Weinberg Equilibrium (HWE) were removed using the following procedure: first, p-values were calculated

for each sample at each locus using the R package pegas v1.4 (Paradis, 2010), then p-values were combined for every locus using Fisher's combination of probabilities. The resulting locus specific p-values were adjusted to q-values for false discovery rate (Benjamini & Hochberg, 1995) and loci with q-values less than 0.05 were removed. For both datasets, individual SNPs were called with a minimum mapping quality of 40 and filtered using a minimum read depth of 10. Across all individuals, SNPs were retained with a minimum average read depth of 15 and less than 20% missing data (O'Leary et al. 2018). To test for differences in minor allele filtering strategy, we separately filtered the intraspecific datasets once using minor allele frequency (MAF) < 0.05 and once using minor allele count (MAC) of 3. Genotype discordance between replicate samples was calculated using the script "genoerrorcalc" (https://github.com/bsjodin/genoerrorcalc). Filtered VCFs were generated using VCFtools (Danecek et al. 2011). To convert data to other formats, PGDSpider v3.0.0.0 (Lischer & Excoffier 2012) was used.

1.3.4 Analysis

After standard quality filtering, we sought to filter our datasets for contamination using an allelic balance-based method (see Figure 1.2 for example). We attempted to run existing filtering tools that most closely matched our needs to evaluate their ability to flag samples in a non-human RRS dataset; one tool was absent from its linked GitHub (dna-contamination-detector; Dallavilla et al., 2020), and the other two failed to run in our computing environment, with limited recourse for technical support in both cases (ART-DeCo, Fiévet et al., 2019; vanquish, Jiang and Motsinger-Reif, 2024). We therefore created a custom R script for contamination filtering that calculated several individual-based metrics from VCF files:

reference and alternative allelic read depths for each locus, total read depth per locus, and number of heterozygous sites. From these data, we calculated the proportion of heterozygous sites where allelic read depths deviated significantly ($p < 0.05$) from 0.5 in a two-tailed binomial test (reference and alternate alleles acting as trials with $p = 0.5$). Because the power of this test depends on the total number of reads, we also calculated the mean absolute deviation across loci from 0.5 (referred to here as "AB deviation"). Individuals were visualized in a scatterplot displaying absolute AB deviation against the proportion of significantly imbalanced sites. For our rockfish datasets, the contamination detection script was applied to the dataset prior to removal of individuals with different genetic and morphological species identifications. Identification discordance is somewhat expected for these species (Wray et al., 2024), so this provides an opportunity to examine whether such discrepancies are due to field misidentification or contamination.

A published dataset of Pacific Herring (*Clupea pallasii*) RAD-seq samples (Petrou et al., 2019) was processed through our contamination analysis in order to compare the allelic balance strategy with a previously used contamination detection method based on individual observed heterozygosity that included intentionally contaminated, potentially contaminated, and "clean" samples. The dataset was downloaded from the NCBI Sequence Read Archive database (NCBI Accession PRJNA508972), aligned to the Atlantic Herring (*Clupea harengus*) genome, and processed through genotype calling using STACKS (Catchen et al., 2013). Data were then filtered using the above contamination pipeline, resulting in 2,774 SNPs. Samples were composed of four intentionally contaminated samples (made from mixing fin tissue clips of four different individuals as a "dirty cocktail" positive control), 211 adults collected during or immediately before active spawning events (and thus potentially contaminated with sperm), and

20 samples of presumably uncontaminated juvenile herring from Cherry Point that were reared at the US Geological Survey (USGS) Marrowstone Marine Field Station, WA. 17 wild adult samples and 20 juvenile samples were exposed to an experimental treatment of a 10-minute bleach bath prior to DNA extraction (samples were split into two pieces; null treatment received no bleach bath). While herring is a different species than our study group, testing our plotting method on samples from a non-hybridizing species with approximate per-sample contamination status garnered from a prior study can act as a control for our rockfish sample sets, wherein we do not know the true level of contamination prior to analysis.

To test differing levels of filtering in our rockfish datasets, we flagged individuals using "mild" (only extreme outliers) and "aggressive" (individuals that only deviated slightly from the main "cloud" of samples) levels of filtering. This was done by first plotting the individuals using the AB deviation and proportion of significantly deviated loci metrics, observing the patterns in the data, and plotting diagonal lines based on the pattern to separate data into two filtering levels. These same slopes and intercepts were used for all four rockfish datasets. Unfiltered, mild, and aggressive filtered rockfish datasets were then analyzed using the same approaches. Individual-level expected and observed heterozygosity estimates were obtained from VCFtools (Danecek et al. 2011). A Principal Component Analysis (PCA) was conducted using the R package adegenet v2.1.11 (Jombart, 2008). Per-group observed and expected heterozygosity, F_{IS}, and pairwise F_{ST} values were obtained using the R package hierfstat v0.5-11 (Goudet, 2005); pairwise F_{ST} values between each dataset's sampling groups were calculated with 1,000 bootstraps. STRUCTURE v2.3.4 (Pritchard et al., 2000) was used to estimate admixture, with five replicates for K = 1-N clusters (N = 5 for interspecific, 7 for Brown Rockfish, and 8 for both Copper Rockfish and Quillback Rockfish) and a burn-in of 10,000 iterations and 100,000 MCMC repetitions. We ran

STRUCTURE using the admixture model and without a priori population information; likely K groups were identified with the ΔK statistic (Evanno et al., 2005) and the mean L(K) using Structure Harvester (Earl and vonHoldt, 2012). The effects of contamination filtering on STRUCTURE analysis were assessed at the most likely K group identified in the unfiltered level. To avoid issues arising from label switching across replicate runs in STRUCTURE, we used CLUMPP (Jakobsson and Rosenberg, 2007) to align replicate files. Results were visualized in R using ggplot2 v3.5.2 (Wickham 2016).

1.4 RESULTS

1.4.1 Rockfish Interspecific Dataset

Per-individual average read counts retained from sequences collected in this study after alignment for Brown Rockfish, Copper Rockfish, and Quillback Rockfish samples were 3.02 million, 4.98 million, and 4.65 million, respectively, which was considerably less than the 11.7 million, 18.0 million, and 18.7 million per-individual average reads obtained by AW. After combining all BAM files from reference-aligned individuals from this study and the AW study into a single dataset, 40 samples with fewer than 300,000 reads and 63 samples with more than 25% missing genotypes were removed. This read count cutoff removed the five negative controls from the dataset, as the maximum number of reads retained in a single negative control sample was 195,000 (with 95,000 reads being the minimum in a negative control sample). Nevertheless, this number of reads in negative controls raised concerns about potential contamination throughout the dataset and its influence on downstream analyses. After quality filtering and removal of replicate samples, 8,685 SNPs were identified in this 391-individual dataset.

We flagged 23 individuals for removal (5.4%) in the mild filter – 22 that were above the regression line, and one for having insufficient data (Figure 1.3). In the aggressive filter, 92 individuals were flagged (23.5%). In our initial STRUCTURE analysis, 24 individuals were found to have their full ancestry corresponding to a species that differed from their field identification; morphological identification for these samples were two Brown Rockfish, three Quillback Rockfish, and 19 Copper Rockfish. 17 of these 24 individuals showed complete Quillback ancestry in STRUCTURE, and one individual morphologically identified as Copper Rockfish had roughly equal ancestry from Quillback Rockfish and Copper Rockfish (48% Quillback Rockfish and 51% Copper Rockfish, also previously identified as such in Wray et al. 2024). In PCA, these individuals also clustered with the species that matched their genetic ancestry rather than their morphological identification, so they were labeled as misidentifications and removed from the dataset prior to downstream analyses. Two of these misidentified individuals (~10%) were also in the initial mild filtering group, and six (~6.5%) were in the initial aggressive filtering group. After removing misidentified individuals, our interspecific dataset consisted of 367 individuals prior to filtering specifically for contamination.

Given the relationship between our filtering metrics (proportion of significantly imbalanced sites and overall deviation from an AB of 0.5), we expected contamination to result in a positive correlation, as contamination would likely cause individuals to deviate considerably in both metrics simultaneously. Most individuals in all four rockfish datasets displayed a negative relationship in the distribution, with a smaller proportion exhibiting deviation in both metrics (Figure 1.3). Individuals with a high proportion of significantly imbalanced sites exhibited relatively low AB deviation, whereas individuals with high AB deviation generally had a lower proportion of imbalanced sites. This pattern appeared to be associated with individuals'

overall read depth – individuals with high read depth rarely displayed high AB deviation, while lower depth individuals were distributed more broadly. Additionally, the proportion of significantly imbalanced loci frequently exceeded the null expected 5% significance under a binomial model, even in individuals with extremely low mean AB deviation. After plotting arbitrary diagonal lines to delineate two filtering levels, our mild dataset was filtered from 367 to 347 individuals, and the aggressive dataset was filtered to 281 individuals.

The interspecific mild filtering dataset included seven individuals (three Brown Rockfish, one Copper Rockfish, and three Quillback Rockfish) that were not flagged in the intraspecific analyses (all individuals in intraspecific filtering were present in the interspecific set). Similarly, interspecific aggressive filtering also contained all individuals in the intraspecific aggressive filtering datasets, with additional individuals unique to the interspecific dataset. The three species clustered distinctly in PCA at all three filtering levels, but individuals removed by filtering tended to be located toward the centroid of the ordination space, indicating less differentiation compared to tightly clustered samples (Figure 1.4). The suspected F1 individual was centered between Quillback Rockfish and Copper Rockfish clusters, along with two other Copper Rockfish individuals pulling away from the Copper Rockfish cluster. The individual closest to the F1 was removed in mild filtering, while the other individual (closest to the Copper Rockfish cluster) was removed in aggressive filtering. Several Brown Rockfish samples trailed into the middle, with the furthest removed in mild or aggressive filtering, but the pattern observed in the Brown Rockfish cluster remained in both levels of filtering even after removal.

Interspecific admixture plots (STRUCTURE analysis, Figures 1.5-1.7) were separated by species to better visualize the effect of the filtering levels. Ancestry percentage that is not from the species' primary ancestry cluster (confirmed against individuals that had <1% ancestry from

other clusters, referred to as “pure” individuals) will hereinafter be referred to as ‘hybrid ancestry.’ Averages for hybrid ancestry between species can be found in Table 1.2. In Brown Rockfish (Figure 1.5), individuals removed in mild filtering had slightly higher average hybrid ancestry (7.57%) than the species average, with one individual containing 19.5% ancestry from Copper and Quillback Rockfish. In Copper Rockfish (Figure 1.6), one individual removed in mild filtering had notably higher Quillback Rockfish ancestry (33.5%) than the species average, while the remaining individuals in this group had 9.55% and 4.11% hybrid ancestry. Quillback Rockfish (Figure 1.7) had the lowest average hybrid ancestry of the three species (0.44%), and mild filtering removed samples with higher-than-average hybrid ancestry (2.20%, with the two highest samples at 6.62% and 6.52% hybrid ancestry). STRUCTURE patterns remained largely the same across all levels of filtering in all three rockfish species. Pairwise F_{ST} (Table 1.1) remained largely stable across filtering levels, though increased slightly with each filtering level.

1.4.2 MAF versus MAC Filtering for Intraspecific Datasets

Prior to running contamination filtering analyses on our intraspecific datasets, we examined the difference between datasets filtered with a minor allele frequency (MAF <0.05) filter or minor allele count (MAC <3) filter. MAC <3 filtering SNP counts were 4,158 for Brown Rockfish, 7,149 for Copper Rockfish, and 6,092 for Quillback Rockfish. For MAF 0.05 filtering, they were 3,817, 7,092, and 3,941 SNPs, respectively; this means that MAF 0.05 filtering reduced the SNP count by 8.20% for Brown Rockfish, 0.80% for Copper Rockfish, and 35.5% for Quillback Rockfish compared to the MAC <3 set. Heterozygosity rates were substantially lower in all three species when filtered with MAC <3 (Table 1.3) and inconsistent with the previous study (Wray et al., 2024) that also contained roughly half of the same individuals used

in this analysis. Additionally, STRUCTURE plots based on MAC filtered loci displayed weak population structure within Quillback Rockfish, with minor differences in ancestry proportions in Copper Rockfish (Figure 1.8). In Quillback Rockfish, using a MAF 0.05 filter resulted in heterozygosity and more distinct STRUCTURE results consistent with the AW study and thus were used as our datasets for the contamination filtering analyses presented in section 1.4.3.

1.4.3 Intraspecific Datasets

For the **Pacific Herring** dataset, we retained 2,774 SNPs in 239 samples (including replicates) after quality filtering using parameters as in Petrou et al. (2019). This dataset displayed a positive correlation between absolute AB deviation and the proportion of loci with a significant deviation in our contamination detection plot (Figure 1.9). All four of the intentionally contaminated samples had both an absolute AB deviation of >0.23 and more than 65% heterozygous sites that were significantly imbalanced. Of the 35 individuals identified as heterozygosity outliers in Petrou et al. (2019), all had both an absolute AB deviation of >0.10 and a proportion of significantly imbalanced sites of >27%. The frequency of heterozygosity outliers increases as deviance in both metrics increases (the 14 most deviant individuals were all flagged as outliers in Petrou et al., 2019), though there are several outliers with low deviance relative to individuals without elevated heterozygosity.

Since our intraspecific rockfish datasets differed in SNP selection from the interspecific dataset, we first compared the overlap in SNPs between them. Intraspecific data of Brown Rockfish shared 16.2%, of Copper Rockfish 18.8%, and of Quillback Rockfish 15.3% SNPs with the interspecific dataset, demonstrating generally low SNP set overlap. All individuals identified at the mild filtering level within each intraspecific dataset were also identified at the interspecific

mild filtering level. For **Brown Rockfish**, we retained 3,817 SNPs in 69 individuals after quality filtering and misidentified individual removal (based on results from the interspecific dataset detailed in section 1.4.1). After contamination filtering, mild and aggressive filtering datasets comprised 64 and 46 individuals, respectively (Figure 1.10). In a PCA (Figure 1.11), three to four groups were visible at the unfiltered level (two small groups in the middle consisted of two and four individuals, respectively). All six West Coast (WC) individuals were completely separated from the remaining individuals at all levels. Patterns remained consistent at all levels of contamination filtering. Approximately half of the individuals in the mild filtering level were higher than average observed heterozygosity (Figure 1.12), while the remaining had average or slightly lower than average heterozygosity. Aggressive filtering removed individuals exclusively at or below average heterozygosity. In STRUCTURE (Figure 1.13), mild filtering removed only individuals with full or nearly full ancestry from one cluster, while aggressive filtering removed several individuals with hybrid ancestry. K = 2 was determined to be the most supported cluster in the unfiltered level, while K = 3 was the most supported cluster in both mild and aggressive levels (Figure 1.14). Pairwise F_{ST} remained stable across filtering levels, with very minor increases in the aggressive filtering level (Table 1.4). F_{IS} and per-region heterozygosity also remained consistent, with marginal increases with increasing filtering level (Table 1.5).

For **Copper Rockfish**, we retained 7,092 SNPs in 123 individuals after quality filtering and misidentified individual removal. After contamination filtering, mild and aggressive filtering datasets comprised 121 and 97 individuals, respectively (Figure 1.15). In PCA, three to four groups were revealed (Figure 1.16) irrespective of filtering, though aggressive filtering removed a large portion of the individuals in one cluster (groups HC and SPS). Since only two individuals were removed in Copper Rockfish mild filtering, individual heterozygosity patterns remained

largely the same, though one mild level individual had substantially lower than average observed heterozygosity (Figure 1.17). Aggressive filtering did not considerably change heterozygosity patterns in five of the seven sampled regions, though HC and SPS saw a decrease in heterozygosity. In STRUCTURE (Figure 1.18), patterns remained consistent across all filtering levels at K = 2, though minor ancestry estimates increased in HC and SPS in aggressive filtering as compared to unfiltered and mild levels (four and three individuals in HC and SPS, respectively). K = 2 was the most supported cluster at the unfiltered level (Figure 1.19), but K = 5 was unambiguously supported in mild filtering; aggressive filtering ΔK values were more ambiguous (K = 2 value 300.9, K = 5 value 308.8). Pairwise F_{ST}, $F_{IS,}$ and per-region heterozygosity remained stable across filtering levels (Table 1.6 and 1.7), with minimal increases as filtering level increases.

For **Quillback Rockfish**, we retained 3,941 SNPs in 161 individuals after quality filtering and misidentified individual removal. After contamination filtering, mild and aggressive filtered datasets comprised 121 and 97 individuals, respectively (Figure 1.20). In PCA, two strong clusters were identified (Figure 1.21) and remained in both levels of filtering. No individuals excluded at either filtering level were obvious outliers in our PCA plots. Individual observed heterozygosity patterns remained largely consistent across all levels of filtering in Quillback Rockfish (Figure 1.22), with neither mild nor aggressive filtering levels removing individuals with exclusively higher or lower than average individual heterozygosity. In STRUCTURE (Figure 1.23), patterns remained consistent and ΔK values unambiguously supported K = 2 across all levels of filtering (Figure 1.24). Pairwise F_{ST} (Table 1.8), $F_{IS,}$ and per-region observed heterozygosity remained stable across filtering levels (Table 1.9).

1.5 DISCUSSION

While cross-sample contamination has been shown to influence population structure and hybridization estimates (Petrou et al., 2019; Jun et al., 2012), our results suggested that low levels of cross-contamination may not necessarily alter patterns in common population genetics analyses. Across both interspecific and intraspecific rockfish datasets, PCA, STRUCTURE, heterozygosity, F_{ST}, and F_{IS} estimates remained stable in contamination filtering, even when aggressive filtering removed a considerable proportion of individuals. In contrast, minor allele filtering strategy had a profound effect on downstream analyses. Intraspecific heterozygosity estimates were considerably lower and inconsistent with previous studies on these species when filtered using minor allele count (MAC) instead of minor allele frequency (MAF) (Table 1.3). When filtered with MAF 0.05, Quillback Rockfish population structure in particular produced stronger STRUCTURE differentiation consistent with PCA clustering (Figure 1.8), as well as with results from Wray et al. (2024), indicating that the inclusion of rarer variants may introduce noise that substantially lowers heterozygosity estimates and obfuscates signals of population structure. These results suggested that filtering strategy may have a larger influence on population estimates than low-level contamination in some RAD-seq datasets. However, outlier individuals in our rockfish datasets showed that contamination could influence ancestry inference by falsely indicating 1) individuals intermediate in PCA and 2) elevated hybrid ancestry in STRUCTURE analysis; this implied that contamination may impact the identification of hybrid individuals, even if dataset-wide ancestry patterns are consistent.

Our results also revealed a depth component to the allelic balance analysis. Individuals with higher average read depth were more likely to have a higher proportion of significantly imbalanced heterozygous sites, as small AB deviations in the 50:50 allelic ratio became more

statistically significant with increasing read depth given the use of a binomial model. This depth pattern was evident in all rockfish filtering analyses, yet these displayed a negative correlation in non-outlier individuals, contrasting with the overall positive correlation in Pacific Herring (Figure 1.9). This could be the much wider range in read depth in rockfish datasets (and also much higher read depth maximum) compared to the Pacific Herring dataset, which may lead to the smaller range of AB deviation observed in the rockfishes. Importantly, RAD-seq data do not produce a perfect 50:50 ratio even in uncontaminated samples, as systematic errors such as sequencing error, amplification bias, or incomplete reference genomes have been shown to contribute to AB deviations (Muyas et al., 2018; McKinney et al., 2016). Substantial deviation in both metrics, however, may be more indicative of contamination than technical variations or biases. The differences between these datasets suggest that contamination signals are influenced by depth and therefore the thresholds for contamination detection may be dataset dependent.

Such signals of allelic imbalance do not necessarily indicate incorrect genotype calls. The STACKS2 pipeline incorporates read depth and allele counts directly into genotype likelihood estimation, and so genotypes with strongly imbalanced allele reads may be less likely to be called as heterozygotes (Rochette et al., 2019; Maruki and Lynch, 2017). The heterozygous loci retained in our datasets therefore excluded sites with extremely uneven allelic balance. Default parameters in STACKS2 dictate that a stack with a minor allele read proportion below 0.05 are called homozygous, between 0.05 and 0.1 are classified as unknown/missing, and above 0.1 are called heterozygous (Catchen et al., 2013). We used these default options for genotype calling in our data, which could explain why individuals with slightly skewed allele balance still produced consistent results in the majority of our analyses. Adjusting these parameters to have stricter minimums and maximums could potentially reduce the retention of contamination-driven

imbalanced sites, though this would likely also increase missing data, which can itself affect downstream analyses (Schmidt et al., 2021; Yi and Latch, 2022). Evaluating the effect of these thresholds and their interplay with contaminated data and quality filtering should be explored in future studies.

Allelic balance may also be affected by paralogous loci (paralogs), which are copies of genome sections that arose from gene or genome duplication events and that mutate independently over time. In taxa with recent genome duplication events, paralogs may still be highly similar and difficult to distinguish; collapsed paralogs can appear as heterozygous genotypes when reads from multiple genomic copies are aligned to the same locus, which can elevate read depth, skew allele ratios, and inflate heterozygosity (McKinney et al., 2016). Paralogous loci are therefore expected to show locus-specific deviations from the 50:50 allelic ratio we expect in diploid loci. Across our datasets, however, we observed inconsistent, individual-specific patterns of allelic imbalance. This indicates that paralogy was unlikely to explain the patterns identified in our contamination filtering analyses, though its influence cannot be fully ruled out.

The overall limited effect of contamination filtering on interspecific population structure analyses may be due to several factors. Only a small proportion (5.4%) of individuals in our interspecific dataset was flagged in mild filtering – which only captured individuals with substantial deviation – suggesting that strongly affected individuals were relatively rare. Furthermore, contamination significant enough to produce allelic imbalances may be insufficient to alter genotype calls at enough loci to impact overall patterns in downstream analyses. Our negative controls retained between 95,000-195,000 reads after alignment, which corresponds to approximately just 2-7% of the average reads retained in our real individuals. If we assume that

similar levels of contamination were present in a portion of our dataset, contamination may not have been significant enough to shift ancestry calculations, which may only impact STRUCTURE-like analyses at levels of 10% or higher (Ward et al., 2026). Additionally, species divergence may reduce the effect that low-level interspecific contamination has on ancestry inference due to the presence of fixed differences between species. It could be simultaneously true that this divergence allows for easier detection of contaminating reads, as contaminants are more likely to differ in genotype and thus skew heterozygous allelic ratios (Ballenghien et al., 2017). Together, this suggests that species divergence may increase sensitivity to contamination detection while also reducing the chance that low-level contamination affects interspecific population structure inference. In the case of hybridization, we showed that individuals with elevated hybrid ancestry identified using PCA and STRUCTURE did not consistently display the allelic imbalances typical for contaminated individuals. Notably, the putative F1 hybrid identified by Wray et al. (2024) was retained across all filtering levels despite its intermediate PCA position and nearly 50/50 ancestry makeup in STRUCTURE. However, two additional individuals in the same group and PCA space as this individual (i.e., intermediate between species) were flagged in mild and aggressive filtering, demonstrating some level of elevated imbalance. These individuals may still have recent hybrid ancestry, considerable levels of contamination, or a combination of both, as our filtering method does not immediately negate the possibility of true hybrid ancestry.

The overlap between individuals flagged in interspecific and intraspecific filtering suggests that our allele balance approach can identify a consistent set of outlier individuals across datasets with very different SNP sets. Despite the low SNP overlap in our dataset types (<20% overlap in all species) and the fact that all three of our study species were aligned to the

same reference genome (*S. umbrosus*, which was not one of our study species), all individuals identified in intraspecific filtering were also flagged in our interspecific dataset, indicating that the contamination signal was robust to the SNP selection strategies and reference genome used here. This indicates that species-specific reference genomes may not be necessary for reliable contamination detection using allele balance methods, at least within closely related taxa. The additional individuals flagged in the interspecific dataset could again reflect increased sensitivity to cross-species contamination – signals of contamination may be stronger in interspecific datasets due to greater genetic divergence as compared to same-species contamination, but more research is needed to support this conclusion.

In all three intraspecific rockfish datasets, patterns remained largely stable across filtering levels throughout our analyses, indicating that low-level cross-sample contamination had limited influence on intraspecific analyses. One possible explanation for this minimal impact is if contamination was primarily inter- rather than intraspecific. During sample DNA extraction and library preparation, samples were randomly checkered across a 96-well plate and minimizing adjacent same-species neighbors (in most cases, wells had no same-species neighbors). Thus, any cross-sample contamination that could have occurred in these steps (i.e., droplet spread when removing films or incubating the plate) is more likely be a different species that neighbors the individual. This potential contamination pathway would also corroborate the seemingly low contamination level, as such droplet spread would likely not heavily impact many individuals at once.

While aggressive filtering also did not change overall patterns observed in PCA and STRUCTURE analysis, it substantially reduced sample size in two Copper Rockfish groups (HC and SPS) that also displayed a higher STRUCTURE minor ancestry proportion compared to their

proportions at mild and unfiltered levels. This suggested that aggressive filtering reducing sizes of some samples may bias ancestry estimates in STRUCTURE, which has been shown to be sensitive to uneven sample sizes (Toyama et al., 2020). Instability in our intraspecific ΔK values further demonstrated this sensitivity; Brown Rockfish and Copper Rockfish ΔK values were ambiguous and inconsistent across filtering levels. Erratic K estimates are a known limitation of the method, which can perform poorly with small or uneven sample size, or when population structure is subtle (Puechmaille, 2016; Evanno et al., 2005). Heterozygosity estimates also increased modestly between filtering levels in all three intraspecific rockfish datasets, which may also indicate bias with decreasing sample size. Heterozygosity estimates can be lower in datasets with more individuals due to the addition of SNPs from individuals with rarer alleles, as most individuals in the dataset will be homozygous for that SNP (Schmidt et al., 2021).

A limitation of our contamination filtering approach is the exclusion of homozygous loci. Cross-sample contamination may substantially affect homozygous loci, as it may generate false heterozygote calls when present at sufficient frequencies (Jun et al., 2012). This may contribute to the inflated heterozygosity estimates seen in other contamination studies, such as Petrou et al. (2019). Restricting our analyses to heterozygous genotype calls may have therefore reduced sensitivity to low-level contamination, particularly when contaminating reads were insufficient in altering the genotype call. However, under the genotype calling thresholds used in STACKS2, low contamination levels (2-7%) in our datasets may have instead resulted in loci classified as missing rather than consistently producing heterozygote calls. Heterozygosity and allelic imbalance were not strongly associated in our rockfish datasets, further suggesting that individuals with notable imbalance did not have widespread contamination-driven genotype

miscalls. A future direction for our allele balance method could incorporate homozygous loci to evaluate its effect.

Comparing minor allele count (MAC <3) to minor allele frequency (MAF 0.05) in filtering had the most profound effect on our analyses as compared to our contamination filtering alone. All rockfish datasets filtered with MAC <3 displayed lower observed heterozygosity and Quillback Rockfish STRUCTURE analysis displayed weaker, less biologically consistent population differentiation than when filtered with MAF 0.05. Quillback Rockfish SNP count was 35.3% higher in the MAC <3 dataset versus the MAF 0.05 dataset, consistent with the stark differences seen in Quillback Rockfish STRUCTURE analysis and observed heterozygosity. However, Brown Rockfish and Copper Rockfish had much smaller increases in SNP count between filtering methods (8.20% and 0.80% respectively), yet both species also experienced substantial differences in heterozygosity between filtering methods. While an effect on heterozygosity is expected when using a MAF 0.05 filter, this indicates that even small proportions of SNPs can have a substantial impact on heterozygosity estimates. Our results contrast with results from Linck and Battey (2019), who found that MAC-based filtering may preserve informative variation that leads to more distinct clustering while still removing singletons. The rare variants retained in our MAC <3 dataset may have added noise to our population structure estimates, leading to a more ambiguous STRUCTURE result. This suggests that the optimal filtering method is dataset-dependent and results should be compared to other analyses like PCA to ensure consistency, as was also suggested by Linck and Battey (2019).

In conclusion, our results indicated that low-level contamination may not substantially affect overall population structure inference but may still influence individual estimates and identification of hybrid individuals in interspecific datasets. We also found that minor allele

filtering method choice can have a considerable impact on heterozygosity estimates and ancestry cluster strength in STRUCTURE analysis. Based on this work, we recommend that studies incorporate and sequence negative controls to identify potential contamination events early, and to carefully consider locus filtering strategy prior to analysis. If contamination is suspected, using multiple methods to examine the dataset and identify outliers (allelic balance, heterozygosity, PCA) can reduce the risk of bias while offering the ability to remove individuals that could influence analyses. Finally, careful consideration of filtering strategy at the locus level is especially critical, as the inclusion of rare variants may substantially impact structure inference.

1.6 FIGURES

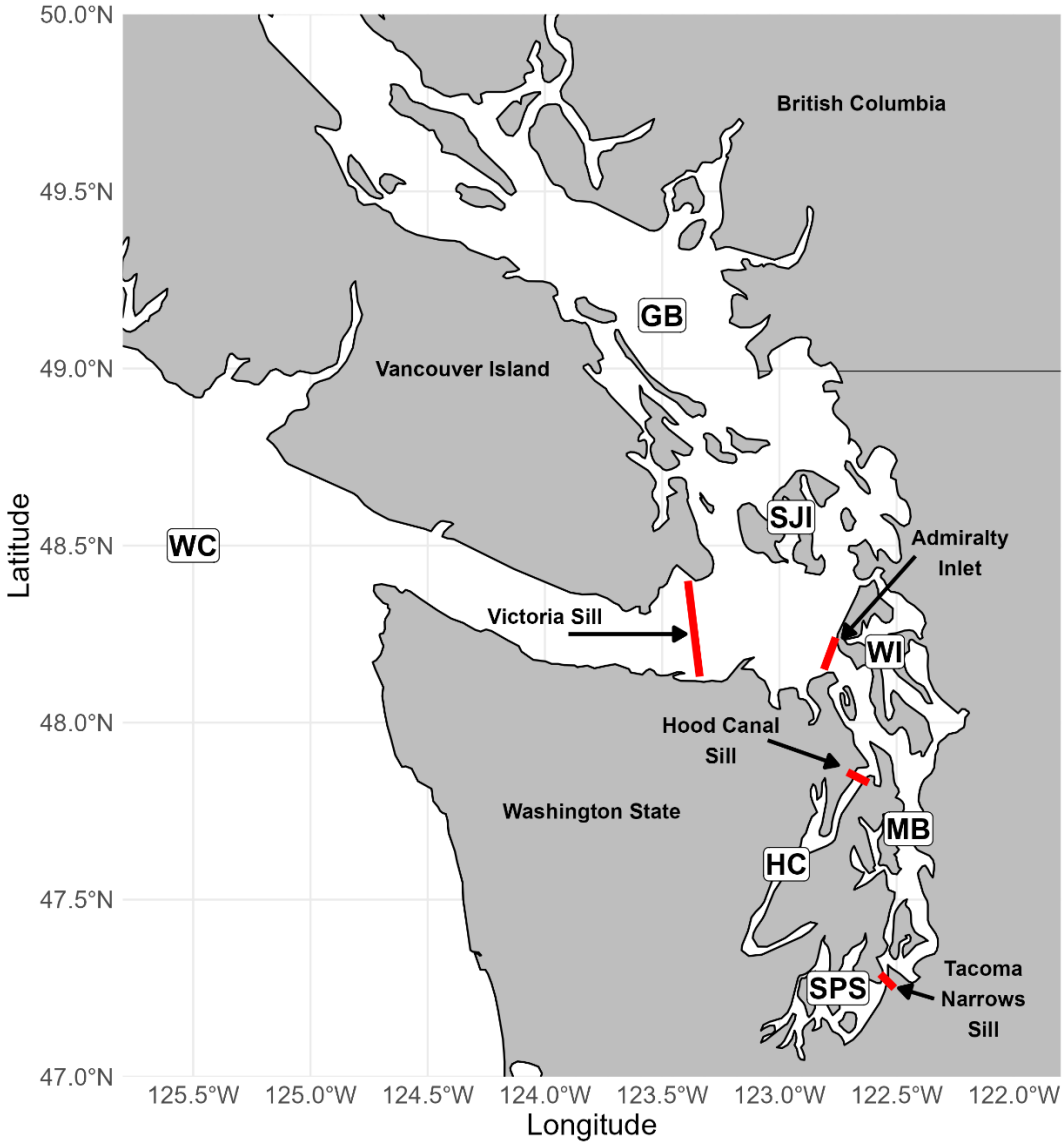


Figure 1.1 Map of rockfish sampling regions in the Salish Sea. Samples were assigned to one of the seven regions according to their sampling coordinates. From the coast, moving inshore: WC = West Coast, GB = Georgia Basin, SJI = San Juan Islands, WI = Whidbey Island, MB = Main Basin, HC = Hood Canal, and SPS = South Puget Sound (south of Tacoma Narrows). Not pictured is Southern California and Mexico, from which 6 Brown Rockfish samples were collected (counted under WC group).

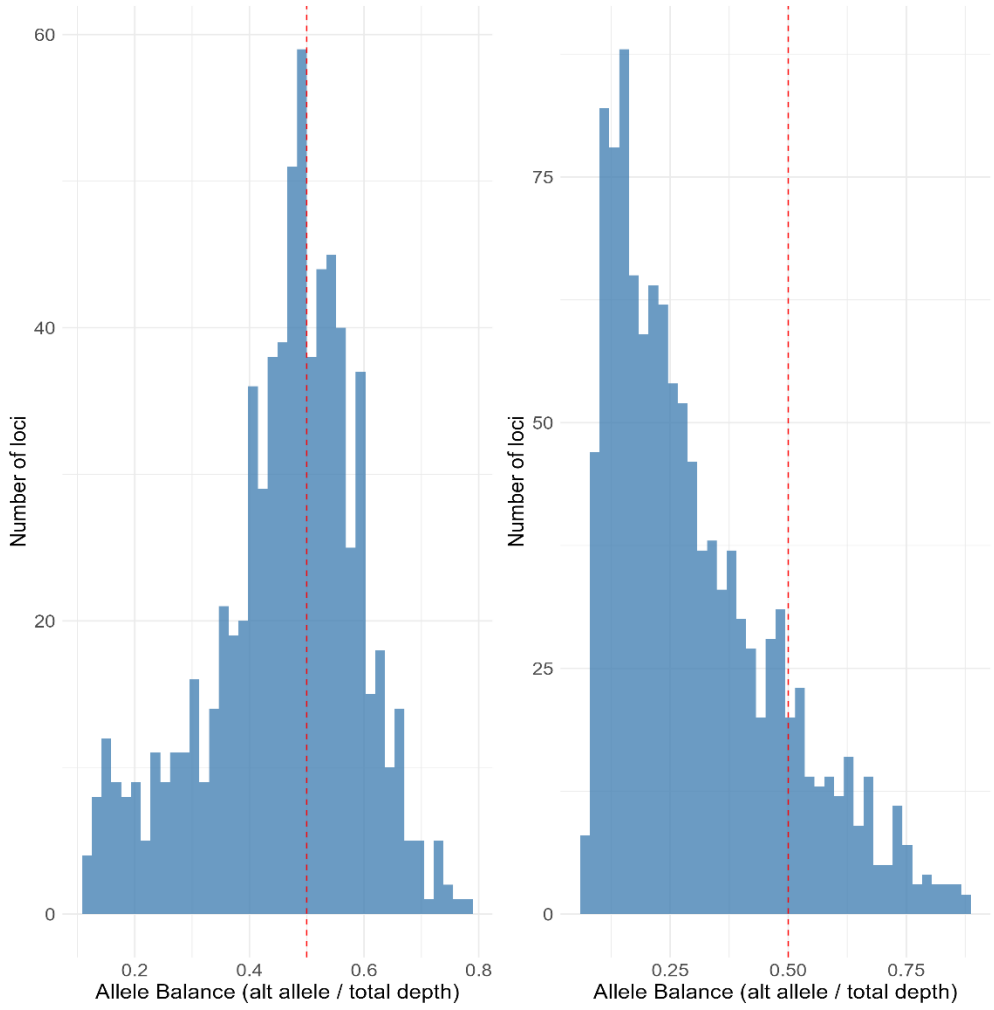


Figure 1.2 Allelic balance of heterozygous loci in clean versus contaminated samples. Plots depict overall allelic balance at heterozygous sites, with the allelic balance ratio per site on the x-axis and the number of loci on the y-axis. The red line marks an allele ratio of 0.5, or 50:50. The left sample is a facility-reared juvenile Pacific Herring (*Clupea pallasii*) sample that was processed through a 10-minute bleach bath prior to DNA extraction. The right sample is a "dirty cocktail" sample, consisting of four Pacific Herring fin clips combined during DNA extraction. Data provided from Petrou et al. (2019).

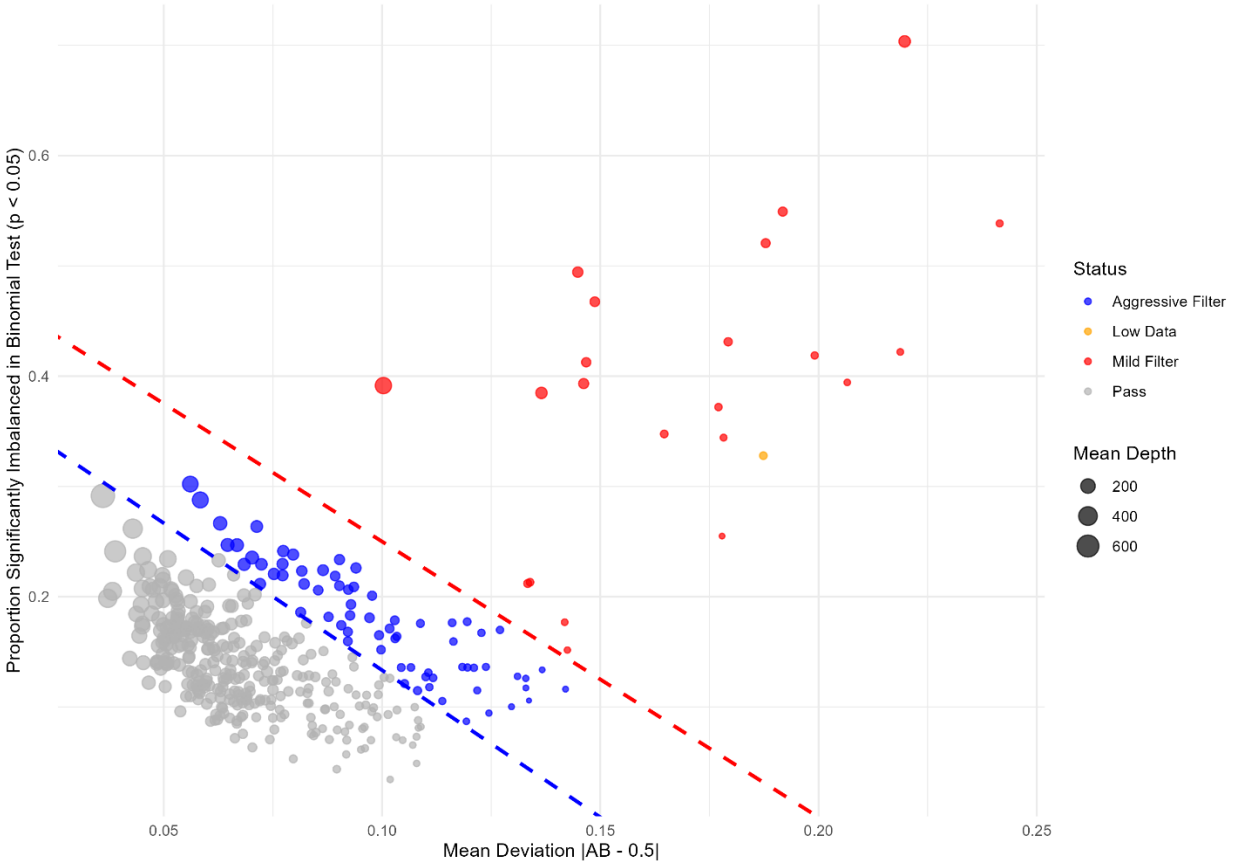


Figure 1.3 Interspecific Brown Rockfish, Copper Rockfish, and Quillback Rockfish contamination filtering analysis. The x-axis metric is the mean absolute deviation from 0.5 of an individual sample's mean allelic balance across all its heterozygous genotype positions. This balance is expected to be at or approximately 0.5 (a 50:50 ratio of reference and alternative reads at each site). The y-axis metric is the sample's proportion of heterozygous positions that had a p-value < 0.05 in a two-tailed binomial test with p = 0.5. Red samples were flagged in mild filtering, and blue samples were flagged in aggressive filtering. Gray samples were not flagged as contaminated in either filtering level. The sample's circle size corresponds to its overall mean depth.

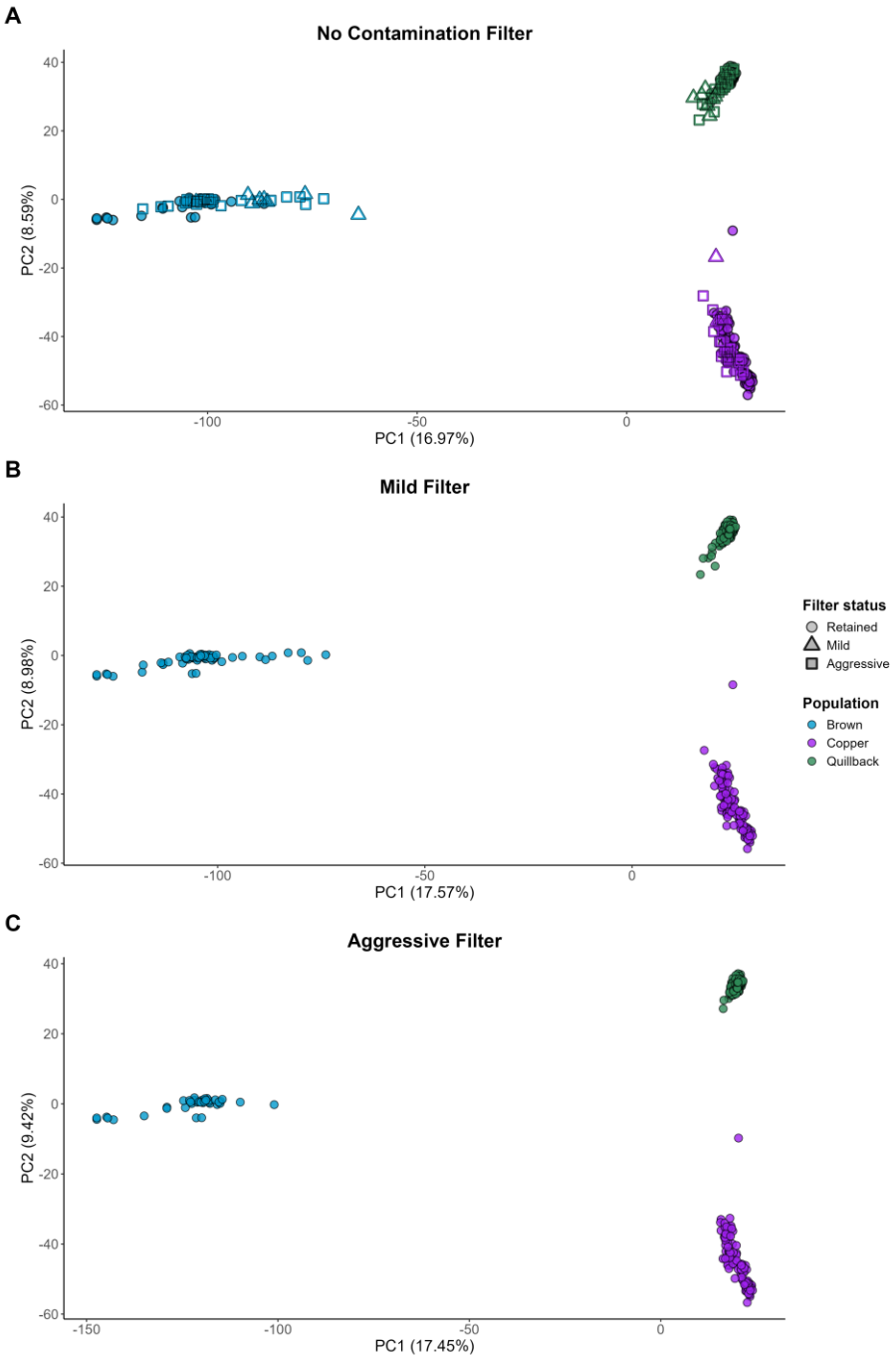


Figure 1.4 Interspecific Brown Rockfish, Copper Rockfish, and Quillback Rockfish PCA at A) unfiltered, B) mild, and C) aggressive filtering levels. Each point is an individual and is colored by its species. Hollow icons denote samples that are filtered out in either mild filtering (triangle) or aggressive filtering (square), and samples that passed contamination filtering are solid circles.

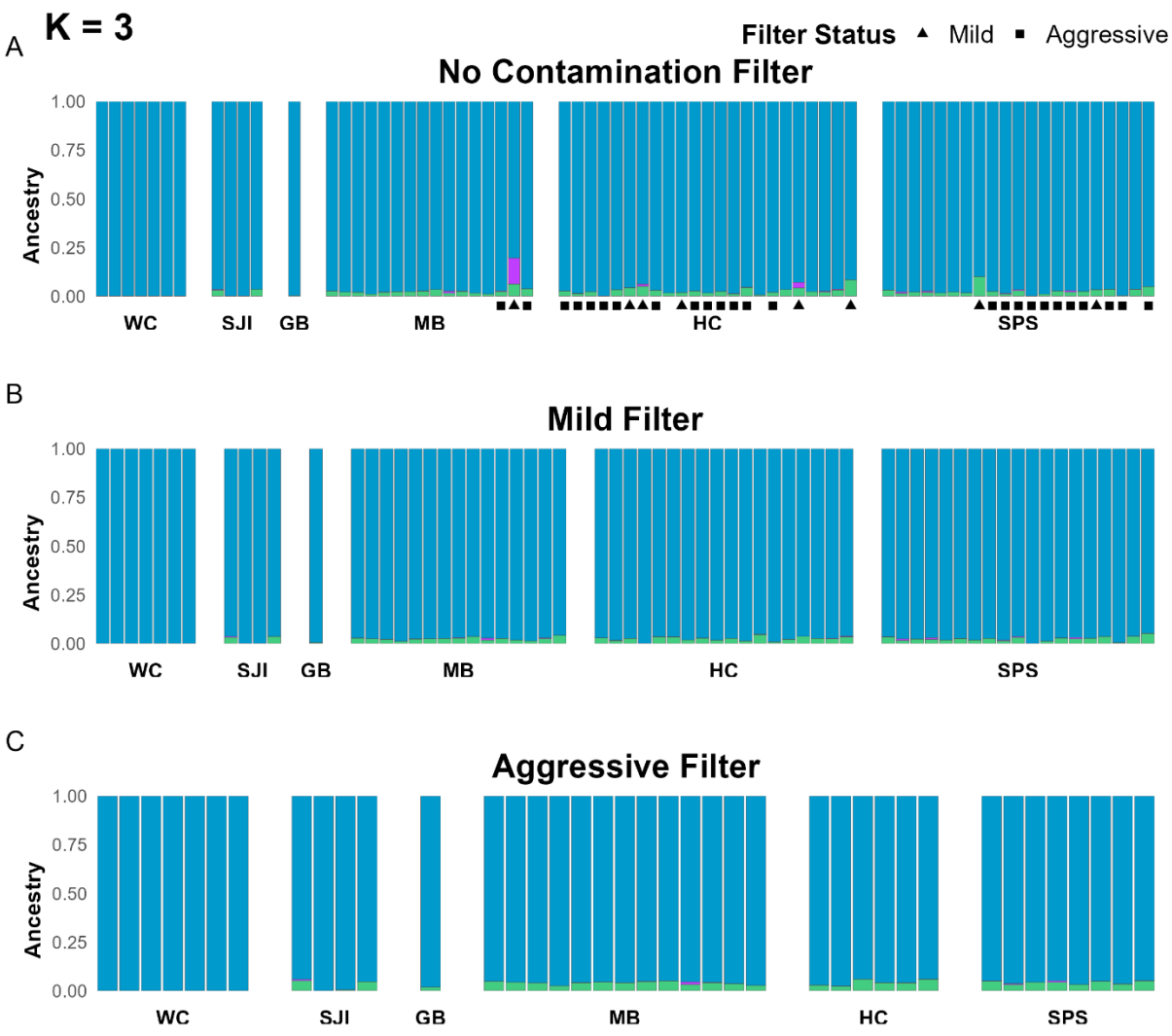


Figure 1.5 Interspecific Brown Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s): blue = Brown Rockfish, purple = Copper Rockfish, green = Quillback Rockfish. Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.

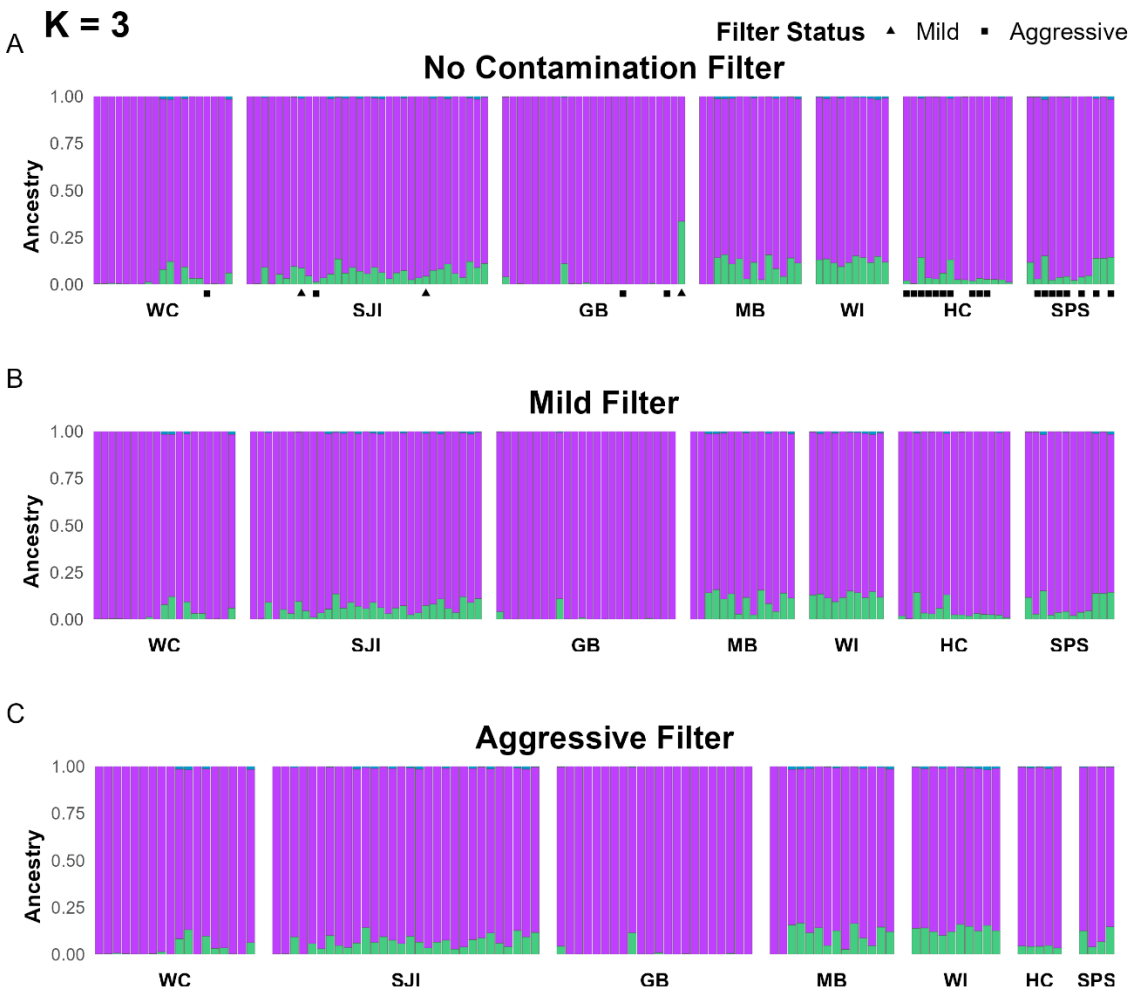


Figure 1.6 Interspecific Copper Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s): blue = Brown Rockfish, purple = Copper Rockfish, green = Quillback Rockfish. Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.

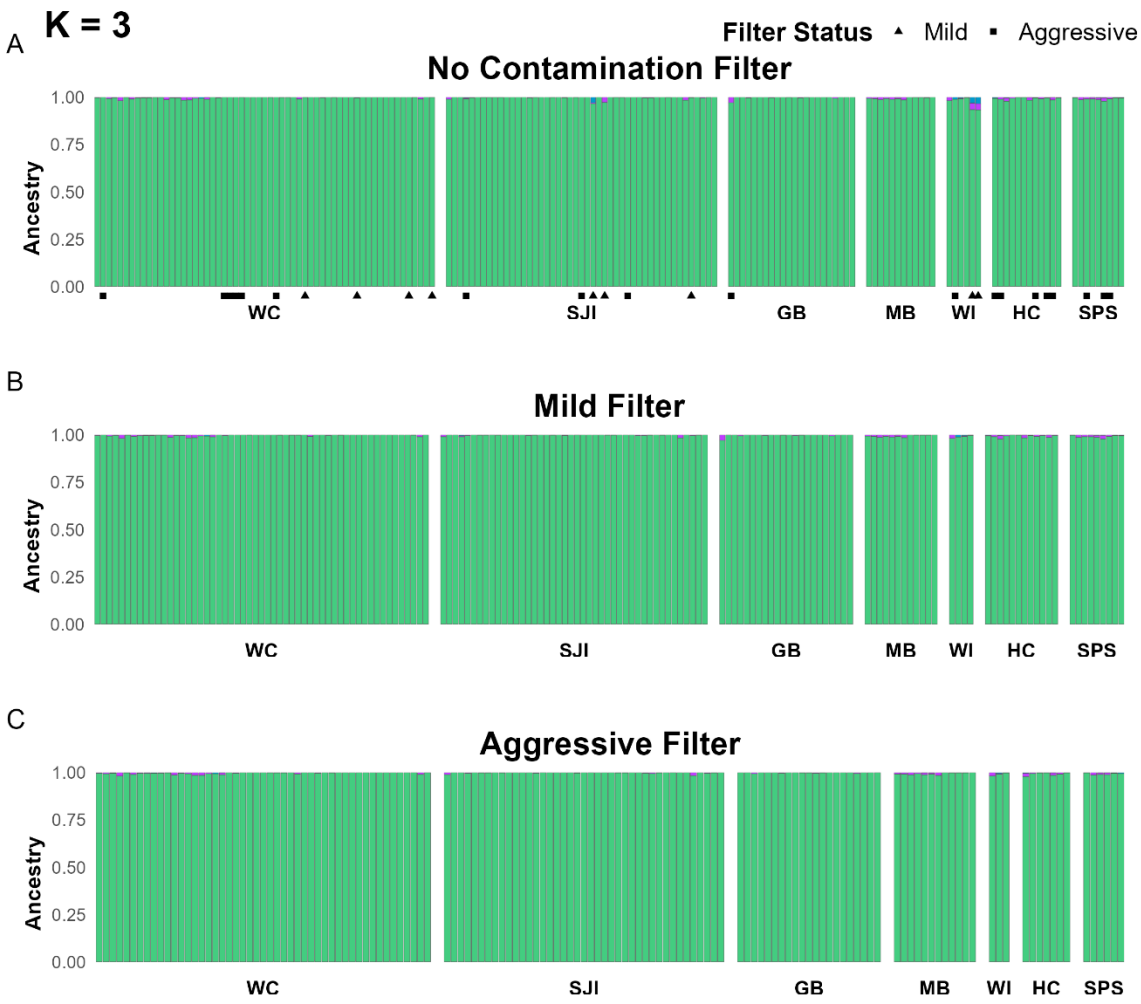


Figure 1.7 Interspecific Quillback Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s): blue = Brown Rockfish, purple = Copper Rockfish, green = Quillback Rockfish. Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.

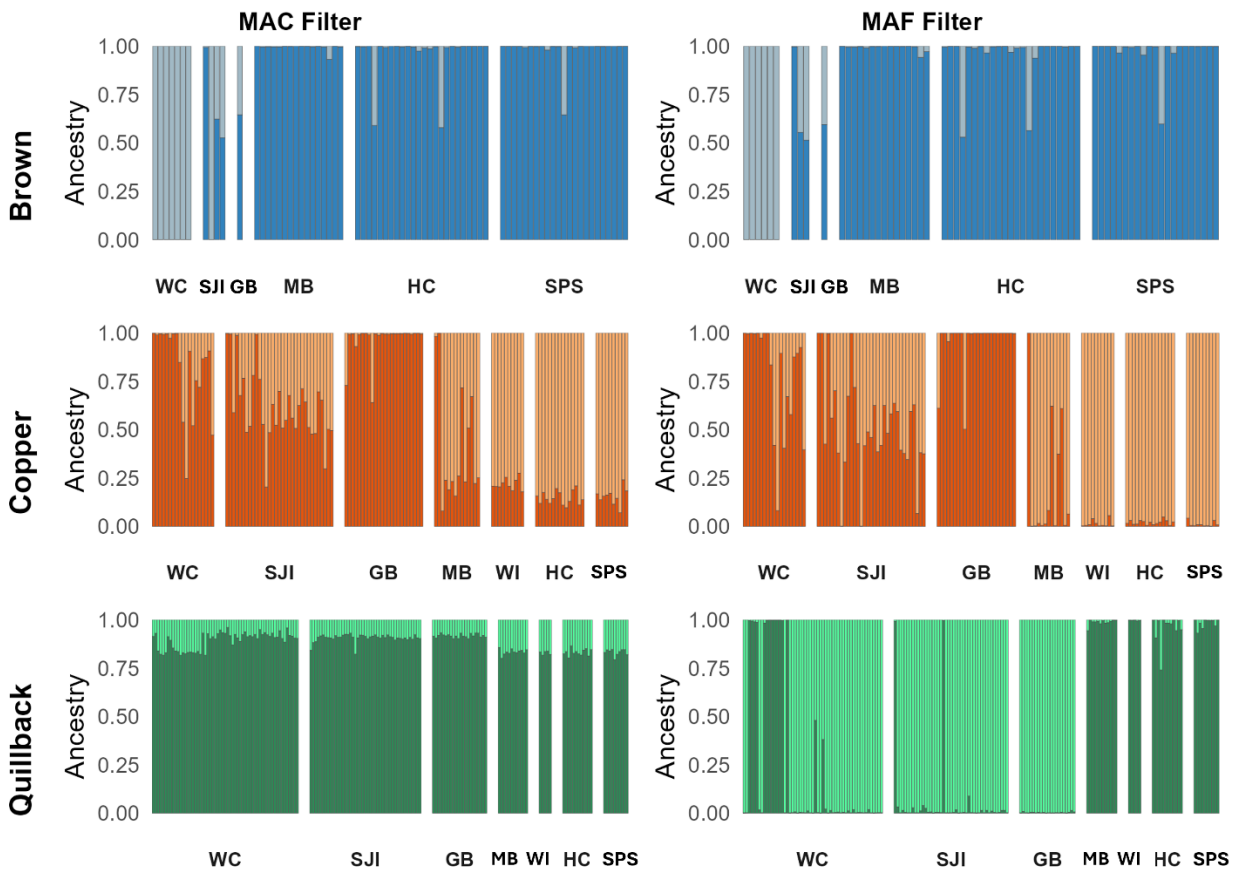


Figure 1.8 Minor allele count (MAC) versus minor allele frequency (MAF) filtering results in STRUCTURE analysis. Column 1 is MAC-filtered results, column 2 is MAF-filtered results. Rows top to bottom are Brown Rockfish, Copper Rockfish, and Quillback Rockfish. Sampling region is ordered along the x-axis, and ancestry proportion is on the y-axis.

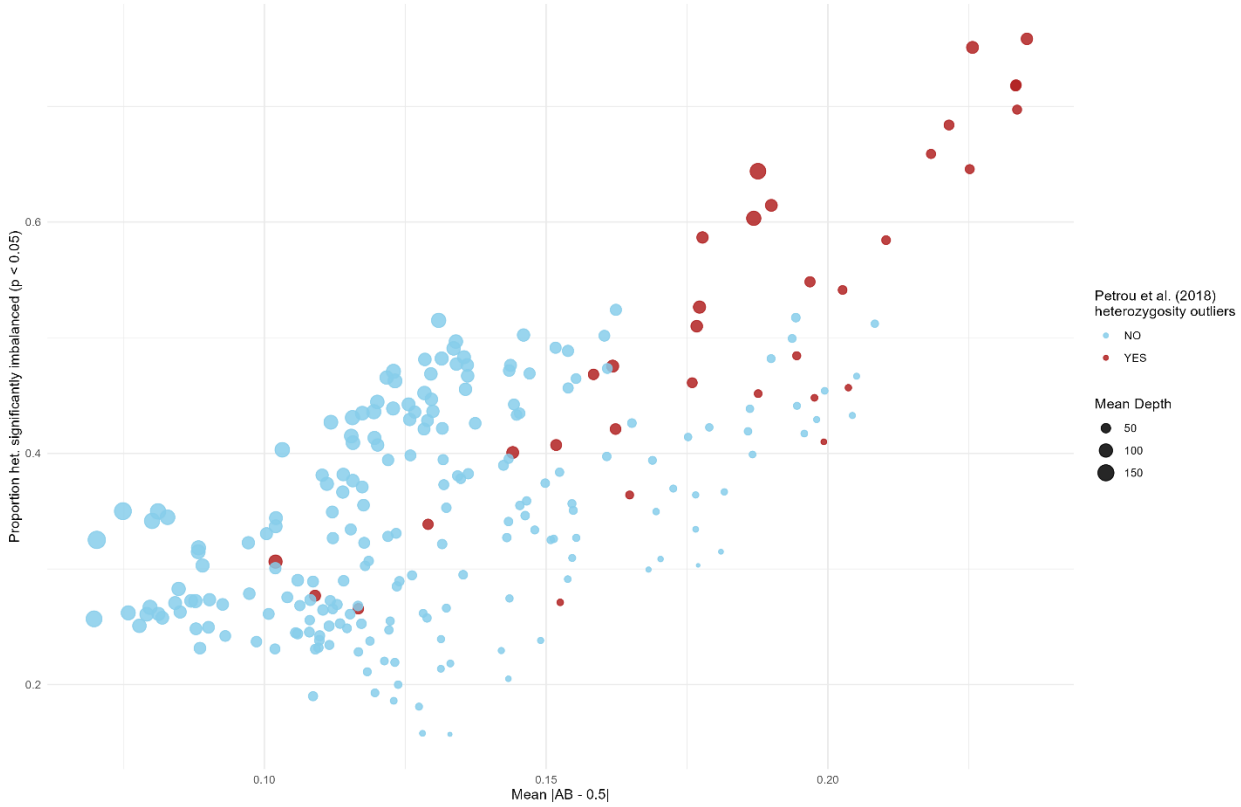


Figure 1.9 Pacific Herring (*Clupea pallasii)* contamination filtering analysis. The x-axis metric is the mean absolute deviation from 0.5 of an individual sample's mean allelic balance across all its heterozygous genotype positions. This balance is expected to be at or approximately 0.5 (a 50:50 ratio of reference and alternative reads at each site). The y-axis metric is the sample's proportion of heterozygous positions that had a p-value < 0.05 in a two-tailed binomial test with p = 0.5. Red samples were identified as contaminated in Petrou et al. (2019) due to excessively elevated heterozygosity (H_O >0.23). Blue samples were not identified as having inflated heterozygosity in the study. The sample's circle size corresponds to its overall mean depth.

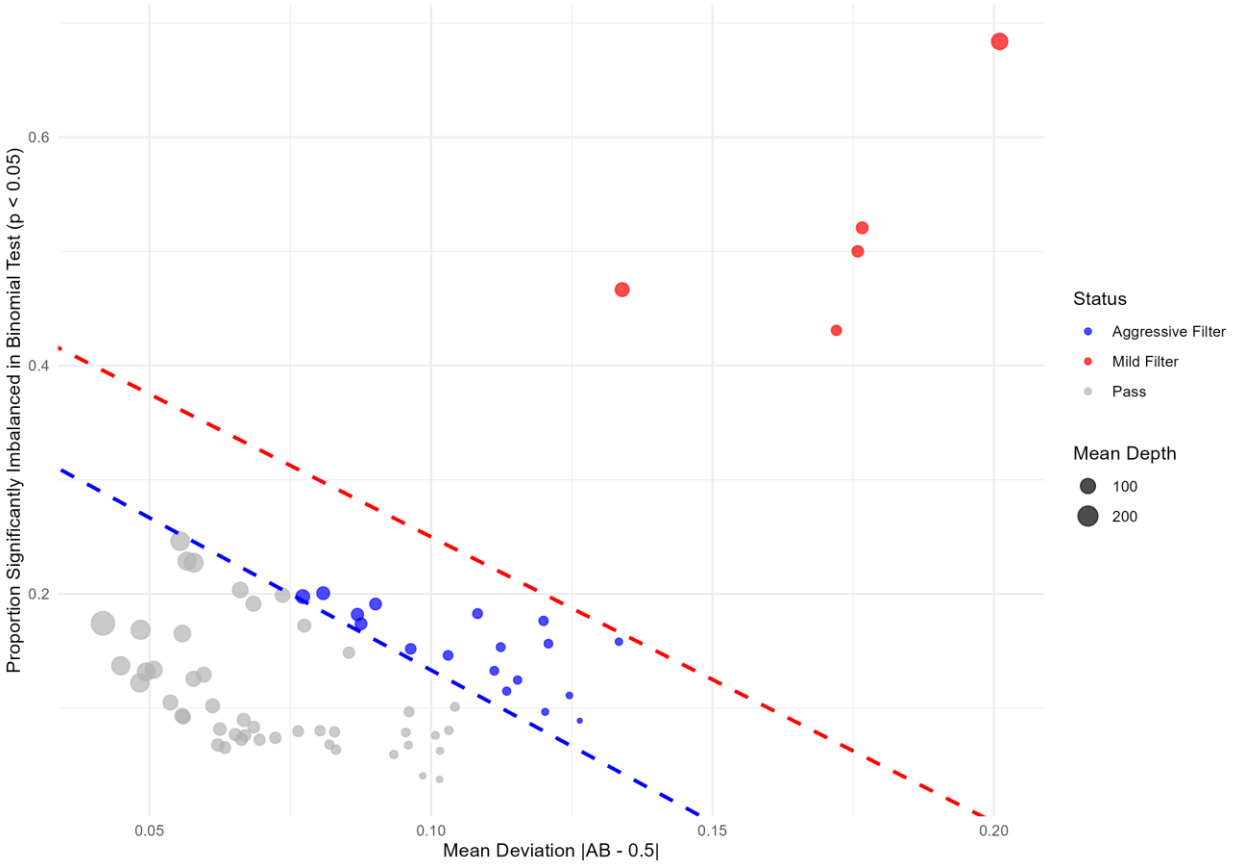


Figure 1.10 Brown Rockfish contamination filtering analysis. The x-axis metric is the mean absolute deviation from 0.5 of an individual sample's mean allelic balance across all its heterozygous genotype positions. This balance is expected to be at or approximately 0.5 (a 50:50 ratio of reference and alternative reads at each site). The y-axis metric is the sample's proportion of heterozygous positions that had a p-value < 0.05 in a two-tailed binomial test with p = 0.5. Red samples were flagged in mild filtering, and blue samples were flagged in aggressive filtering. Gray samples were not flagged as contaminated at either filtering level. The sample's circle size corresponds to its overall mean depth.

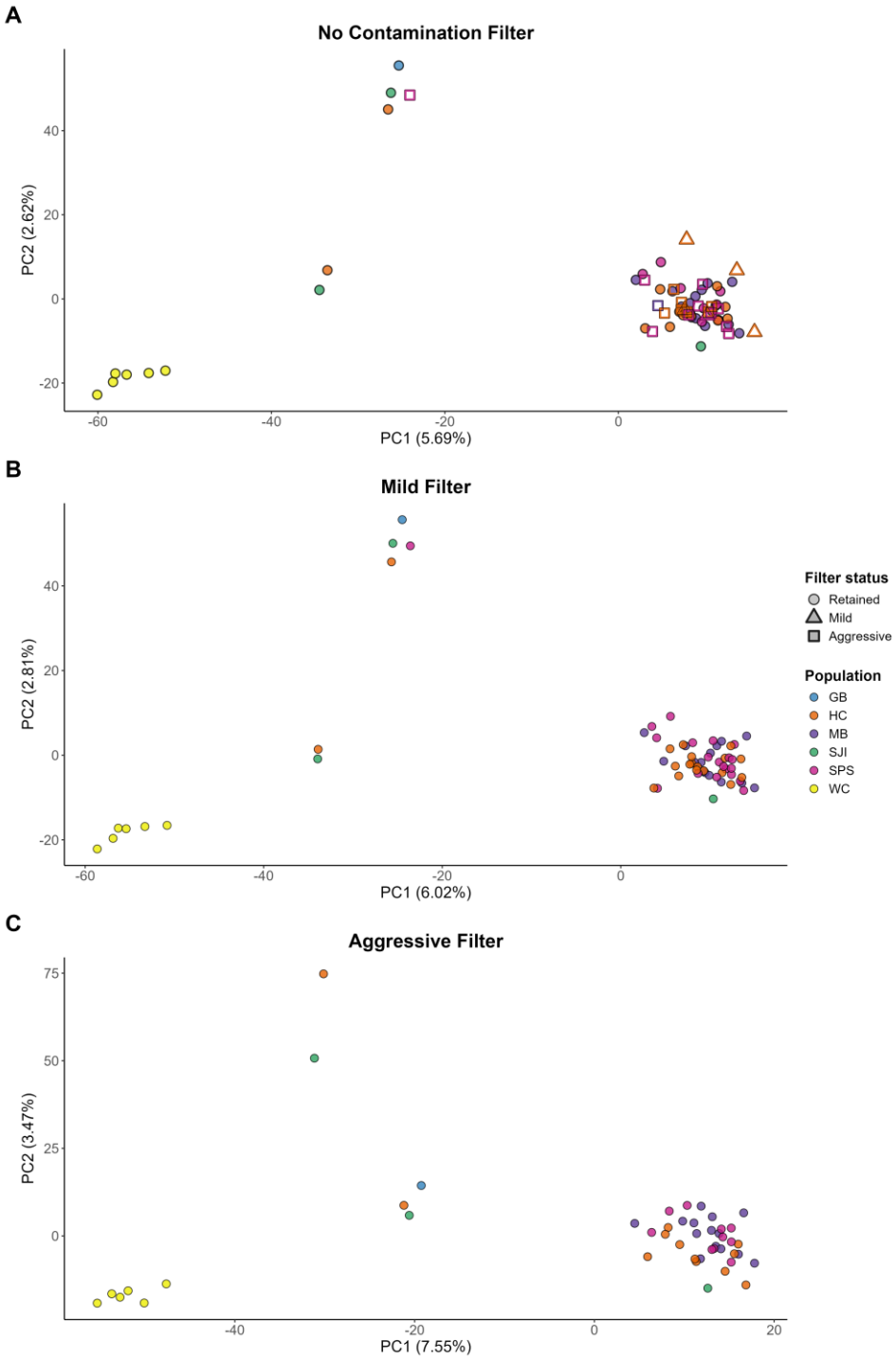


Figure 1.11 Brown Rockfish PCA at A) unfiltered, B) mild, and C) aggressive filtering levels. Each point is an individual and is colored by its sampling location. Hollow icons denote samples that are filtered out in either mild filtering (triangle) or aggressive filtering (square), and samples that passed contamination filtering are solid circles.

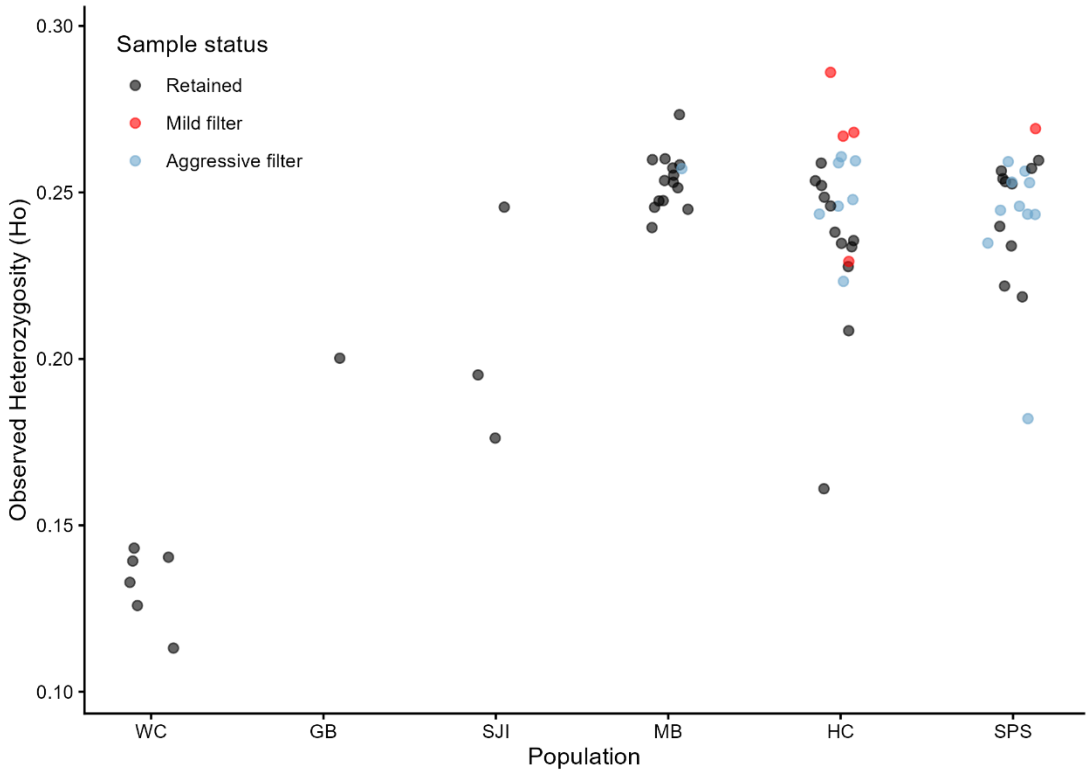


Figure 1.12 Brown Rockfish individual observed heterozygosity per sampling group at unfiltered, mild, and aggressive filtering levels. Each point is an individual and is colored by its filter status: red = removed in mild filtering, blue = removed in aggressive filtering. Points are grouped by sampling location (x-axis). Individual heterozygosity was obtained using VCFtools (Danecek et al., 2011).

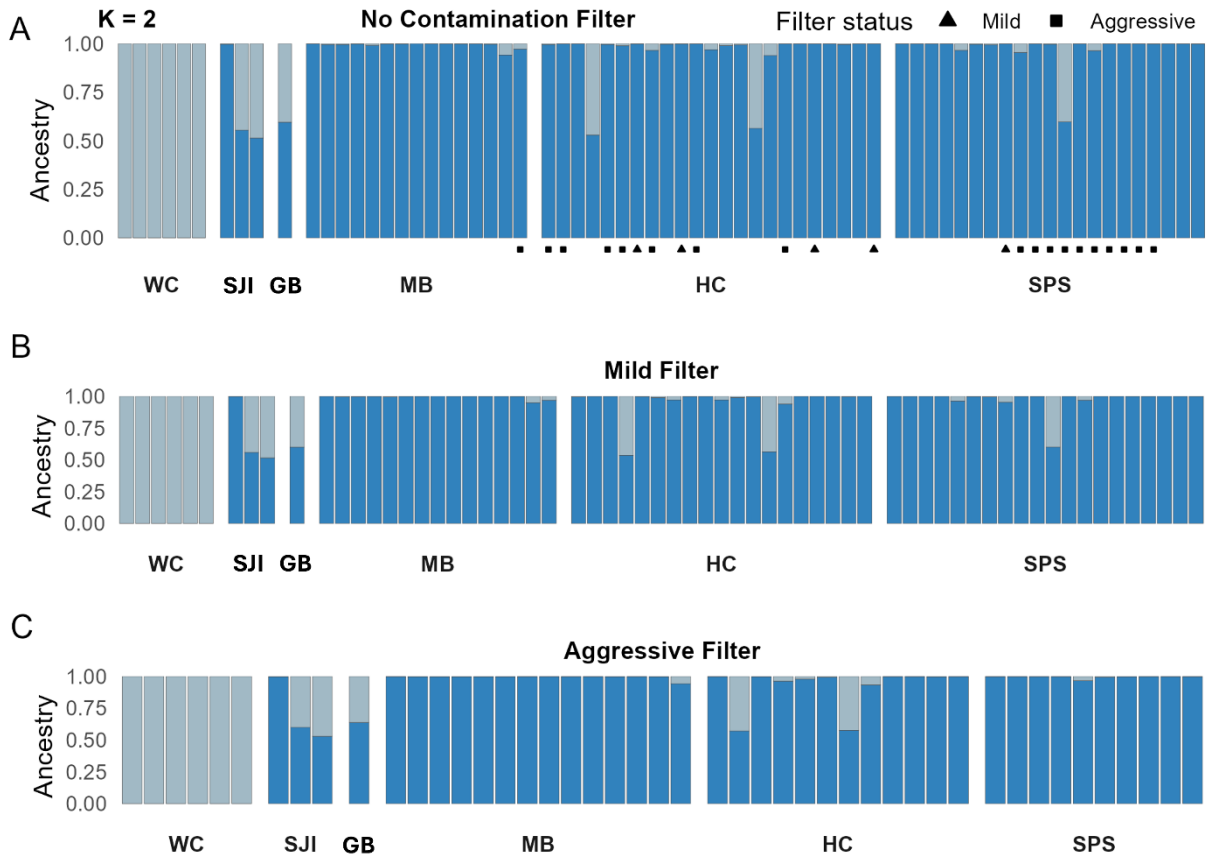


Figure 1.13 Brown Rockfish STRUCTURE analysis at K = 2, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s). Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.

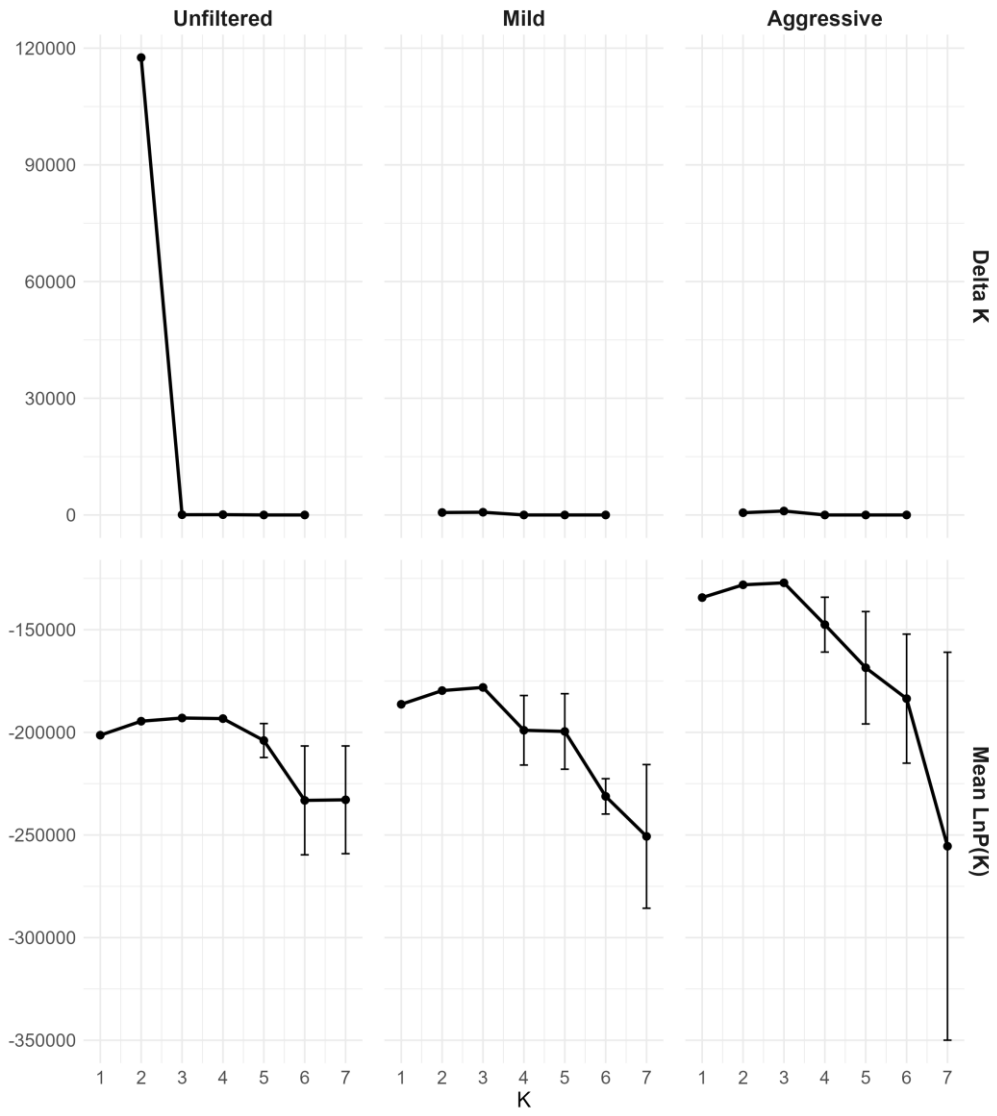


Figure 1.14 Delta K (top) and Mean LnP(K) (bottom, ±1 SD) for Brown Rockfish STRUCTURE analysis. Filtering levels are indicated by column, left to right.

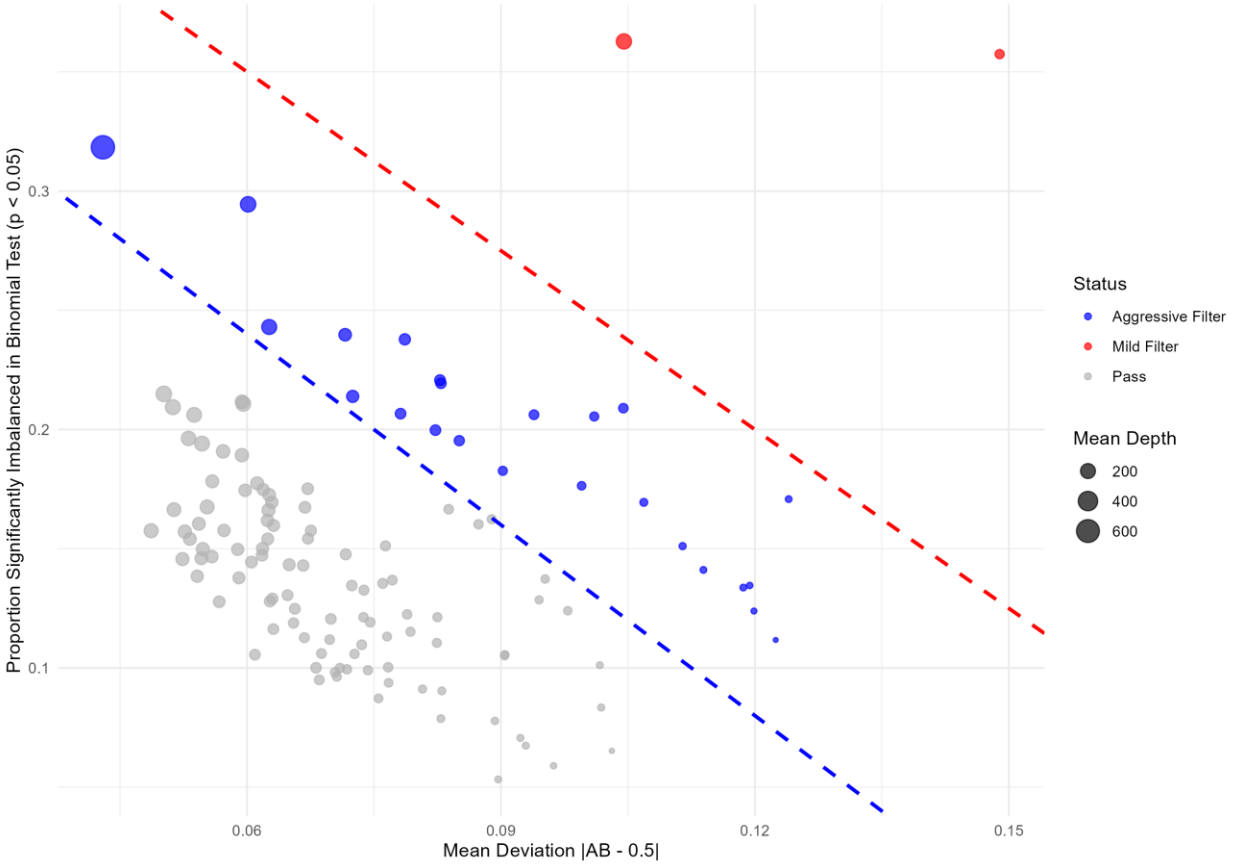


Figure 1.15 Copper Rockfish contamination filtering analysis. The x-axis metric is the mean absolute deviation from 0.5 of an individual sample's mean allelic balance across all its heterozygous genotype positions. This balance is expected to be at or approximately 0.5 (a 50:50 ratio of reference and alternative reads at each site). The y-axis metric is the sample's proportion of heterozygous positions that had a p-value < 0.05 in a two-tailed binomial test with p = 0.5. Red samples were flagged in mild filtering, and blue samples were flagged in aggressive filtering. Gray samples were not flagged as contaminated at either filtering level. The sample's circle size corresponds to its overall mean depth.

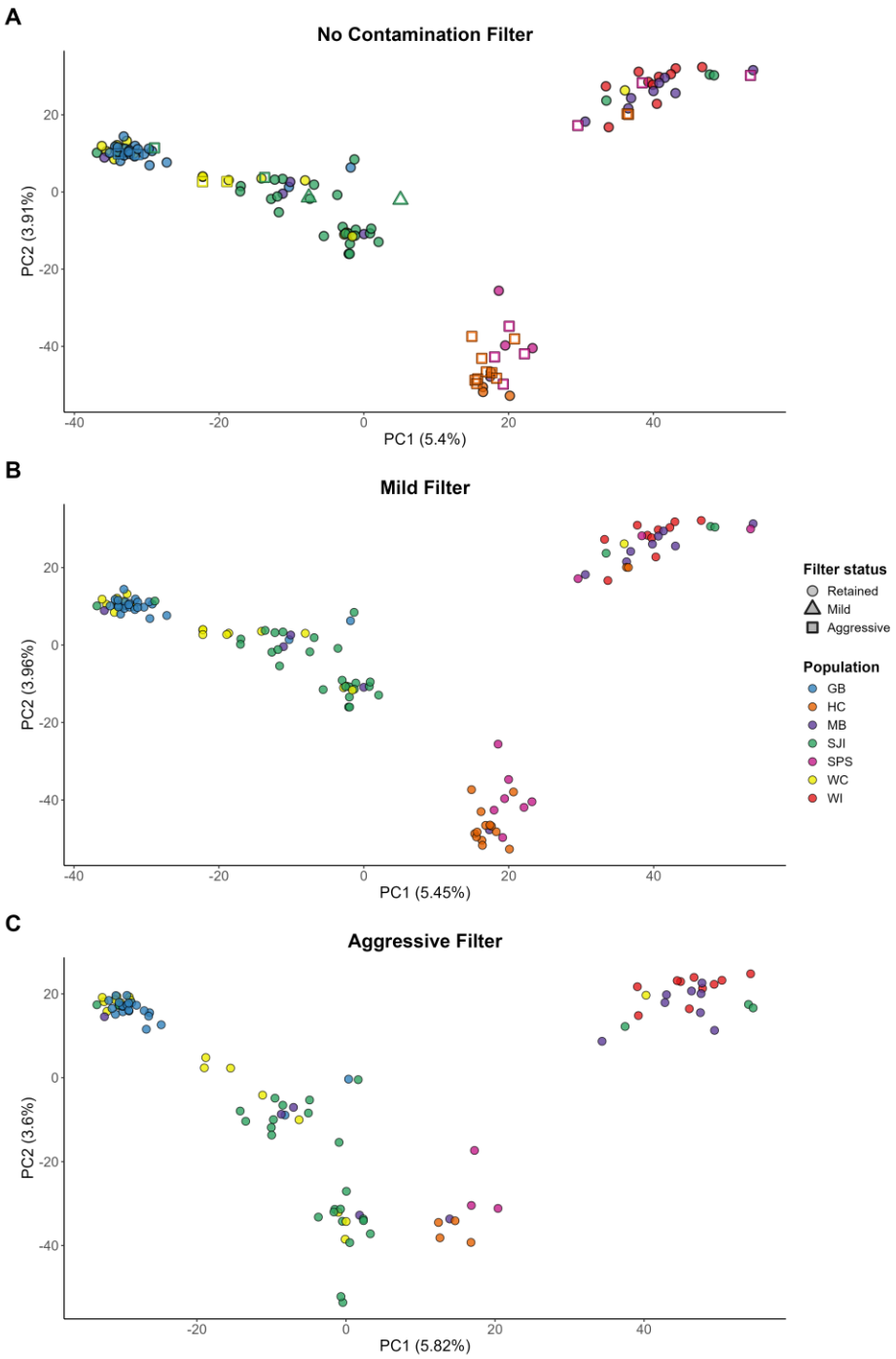


Figure 1.16 Copper Rockfish PCA at A) unfiltered, B) mild, and C) aggressive filtering levels. Each point is an individual and is colored by its sampling location. Hollow icons denote samples that are filtered out in either mild filtering (triangle) or aggressive filtering (square), and samples that passed contamination filtering are solid circles.

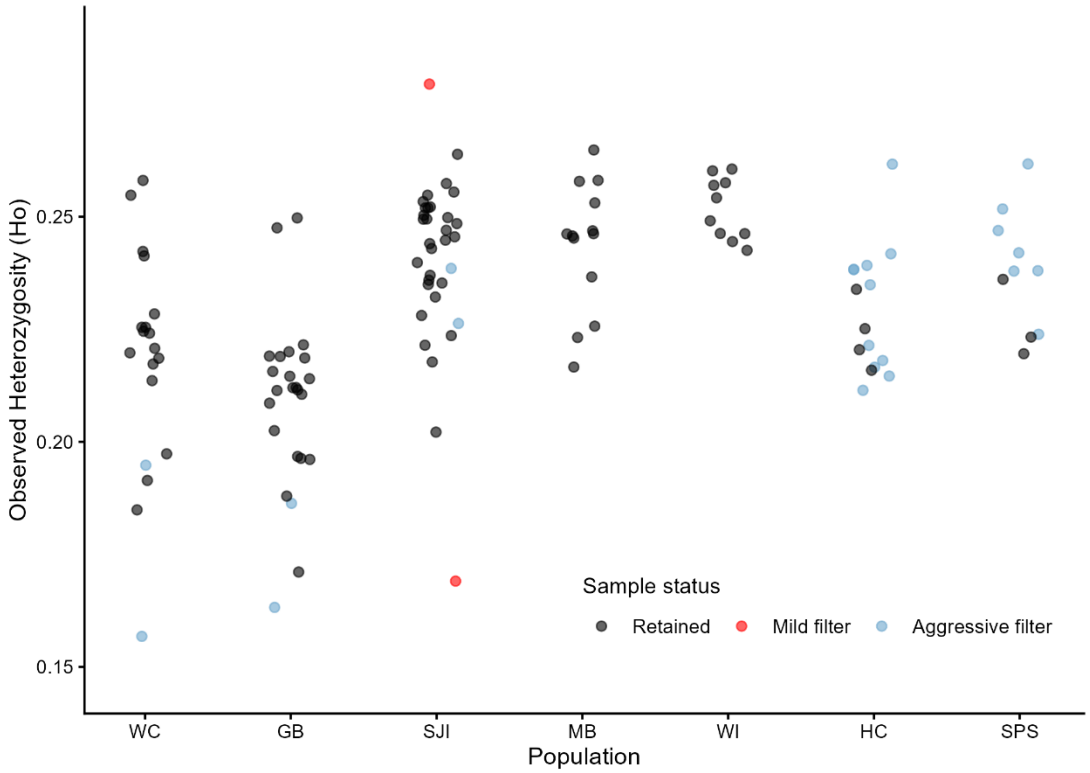


Figure 1.17 Copper Rockfish individual observed heterozygosity per sampling group at unfiltered, mild, and aggressive filtering levels. Each point is an individual and is colored by its filter status: red = removed in mild filtering, blue = removed in aggressive filtering. Points are grouped by sampling location (x-axis). Individual heterozygosity was obtained using VCFtools (Danecek et al., 2011).

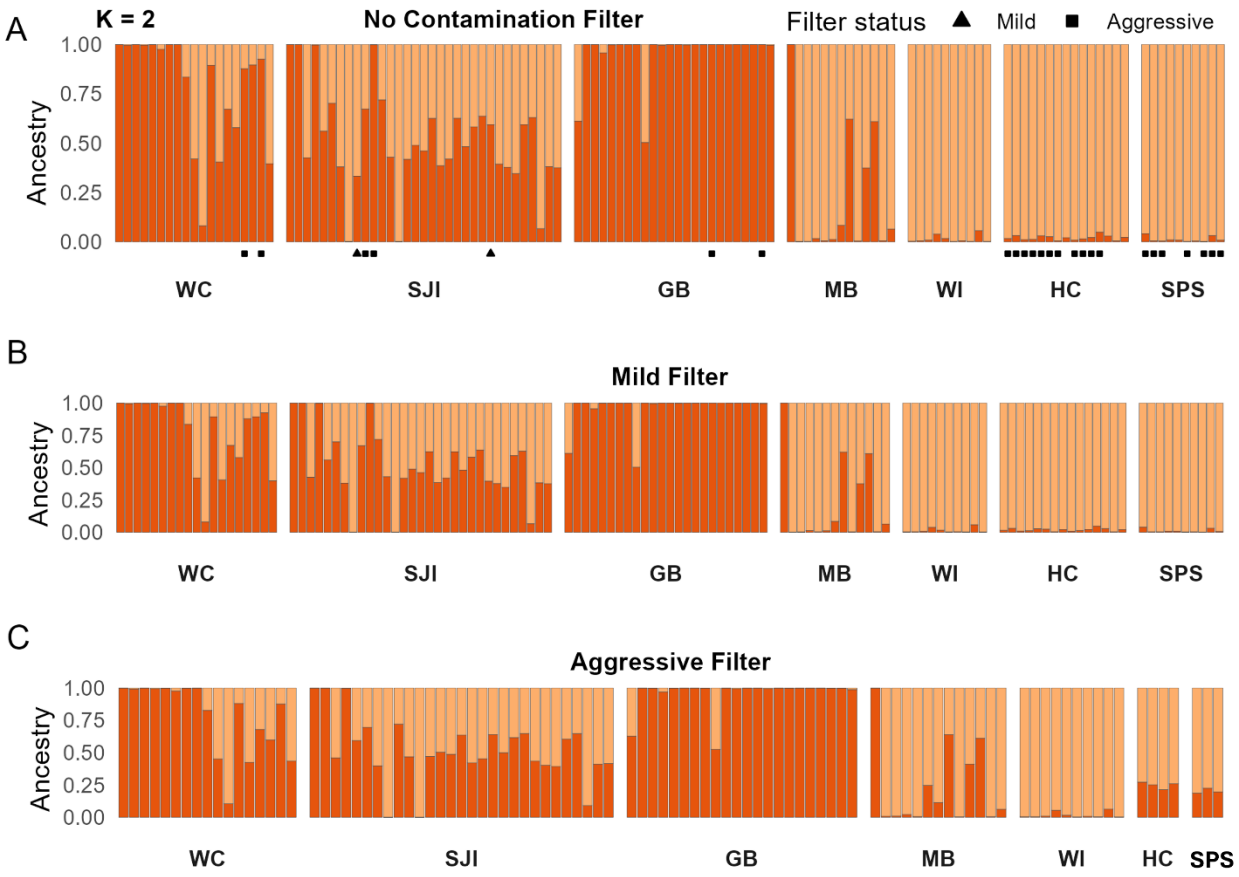


Figure 1.18 Copper Rockfish STRUCTURE analysis at K = 2, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s). Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.

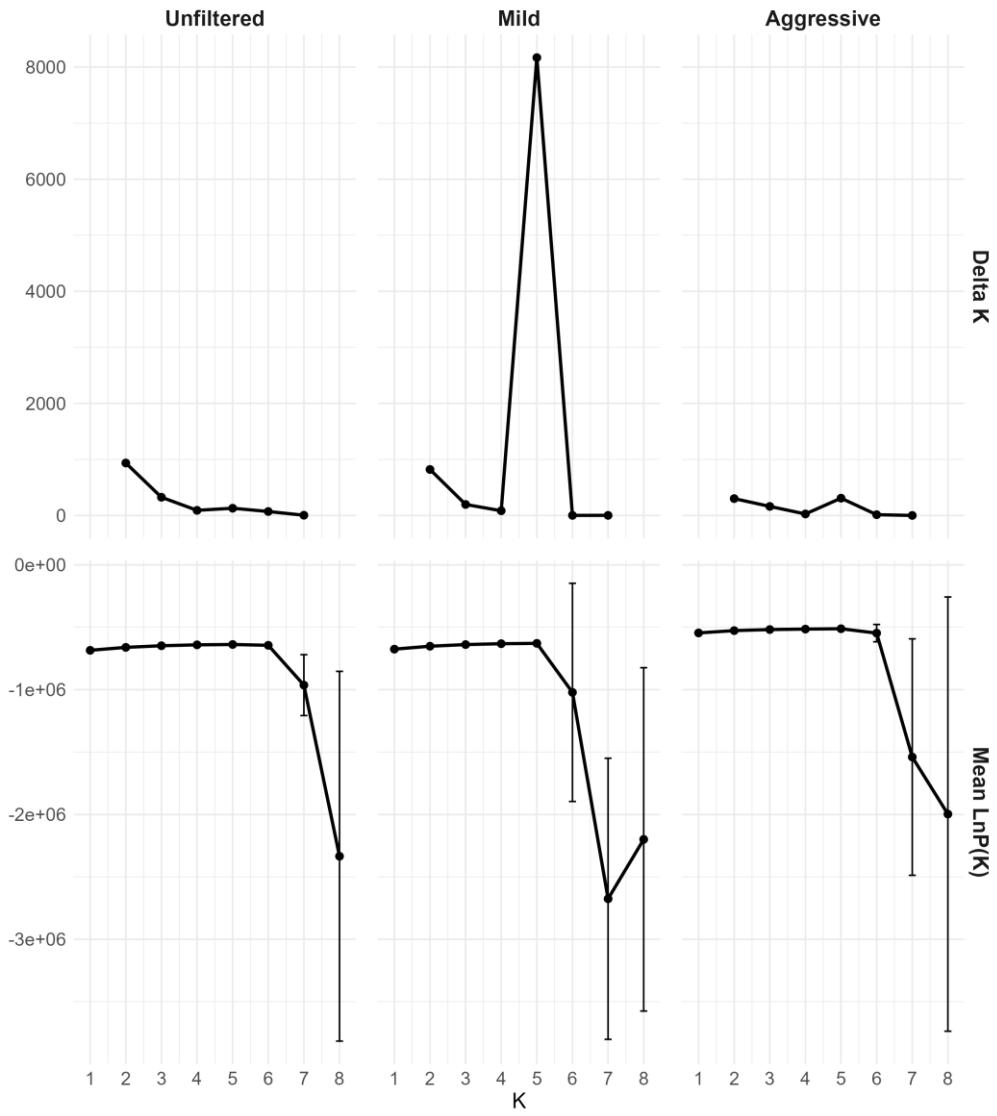


Figure 1.19 Delta K (top) and Mean LnP(K) (bottom, ±1 SD) for Copper Rockfish STRUCTURE analysis. Filtering levels are indicated by column, left to right.

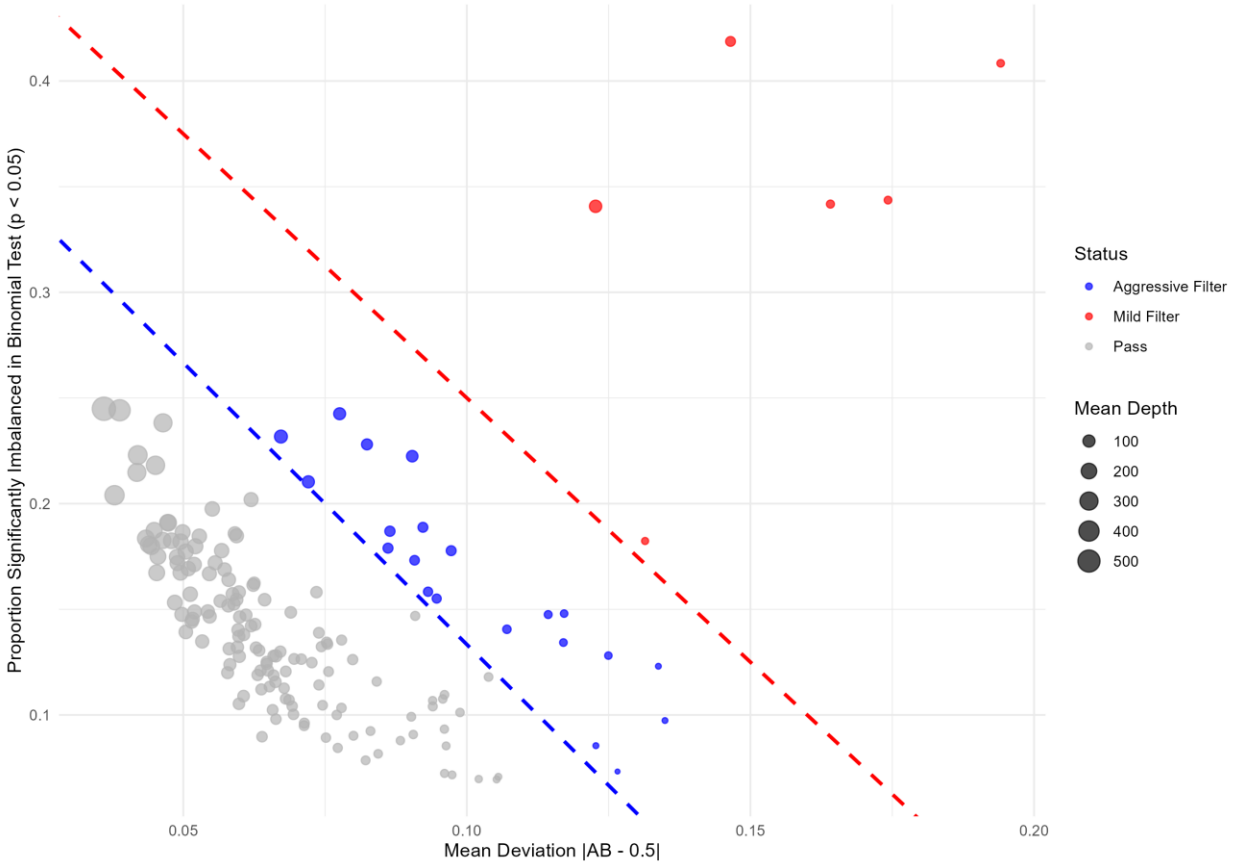


Figure 1.20 Quillback Rockfish contamination filtering analysis. The x-axis metric is the mean absolute deviation from 0.5 of an individual sample's mean allelic balance across all its heterozygous genotype positions. This balance is expected to be at or approximately 0.5 (a 50:50 ratio of reference and alternative reads at each site). The y-axis metric is the sample's proportion of heterozygous positions that had a p-value < 0.05 in a two-tailed binomial test with p = 0.5. Red samples were flagged in mild filtering, and blue samples were flagged in aggressive filtering. Gray samples were not flagged as contaminated at either filtering level. The sample's circle size corresponds to its overall mean depth.

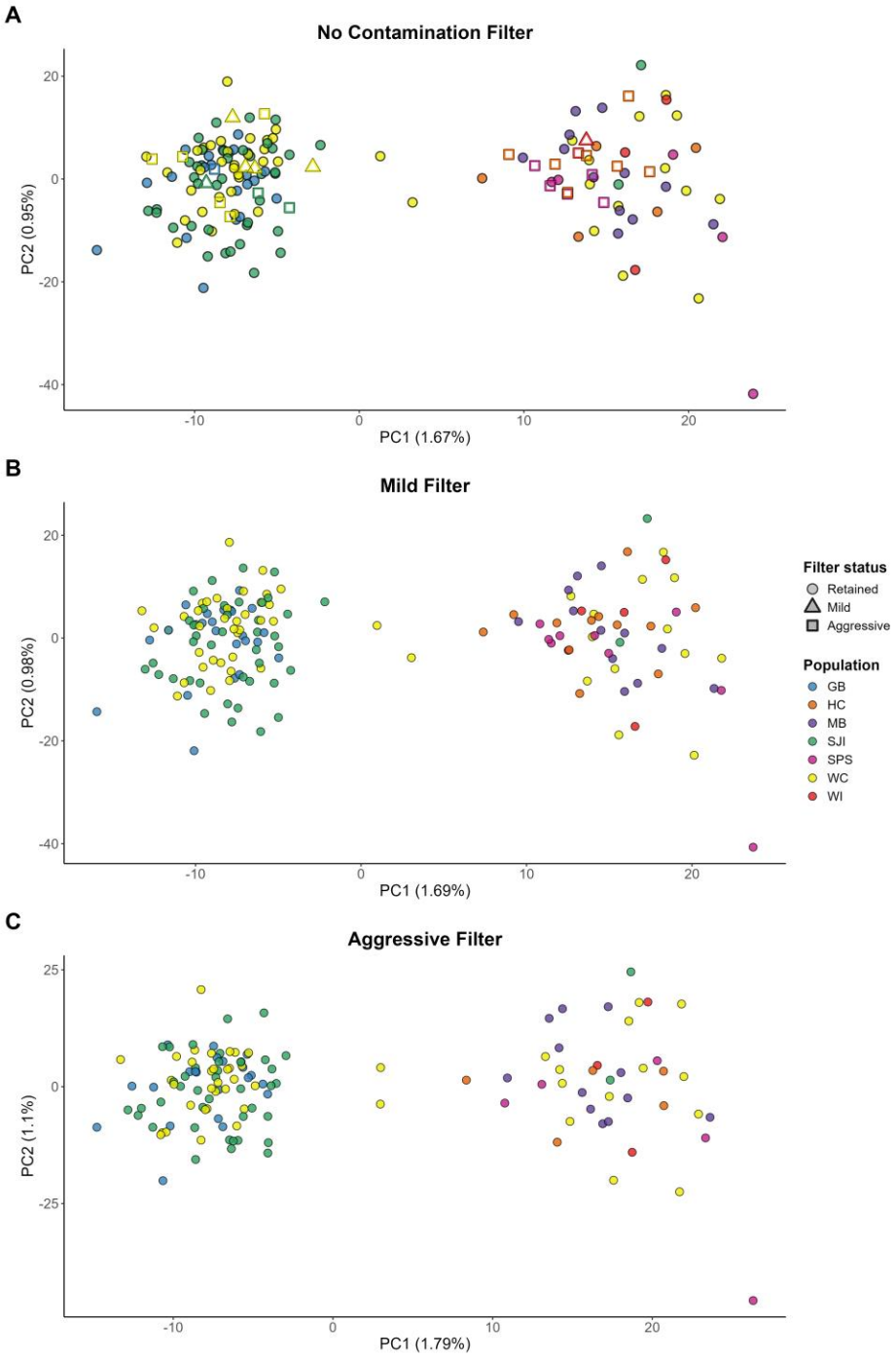


Figure 1.21 Quillback Rockfish PCA at A) unfiltered, B) mild, and C) aggressive filtering levels. Each point is an individual and is colored by its sampling location. Hollow icons denote samples that are filtered out in either mild filtering (triangle) or aggressive filtering (square), and samples that passed contamination filtering are solid circles.

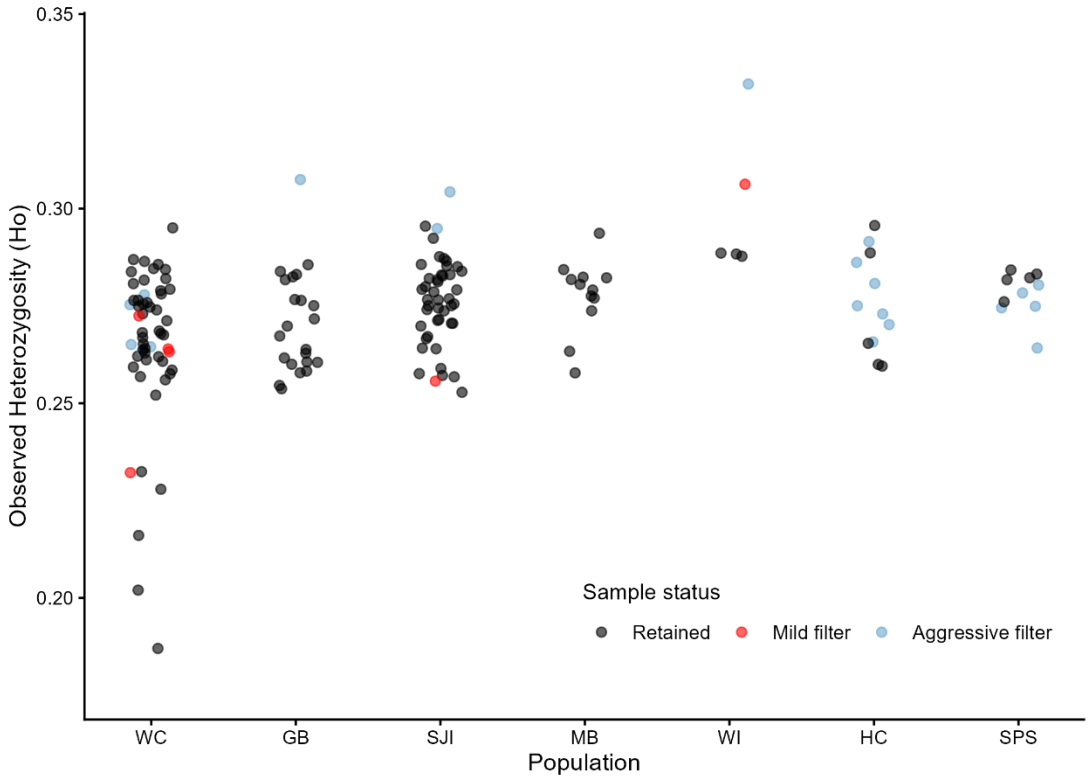


Figure 1.22 Quillback Rockfish individual observed heterozygosity per sampling group at unfiltered, mild, and aggressive filtering levels. Each point is an individual and is colored by its filter status: red = removed in mild filtering, blue = removed in aggressive filtering. Points are grouped by sampling location (x-axis). Individual heterozygosity was obtained using VCFtools (Danecek et al., 2011).

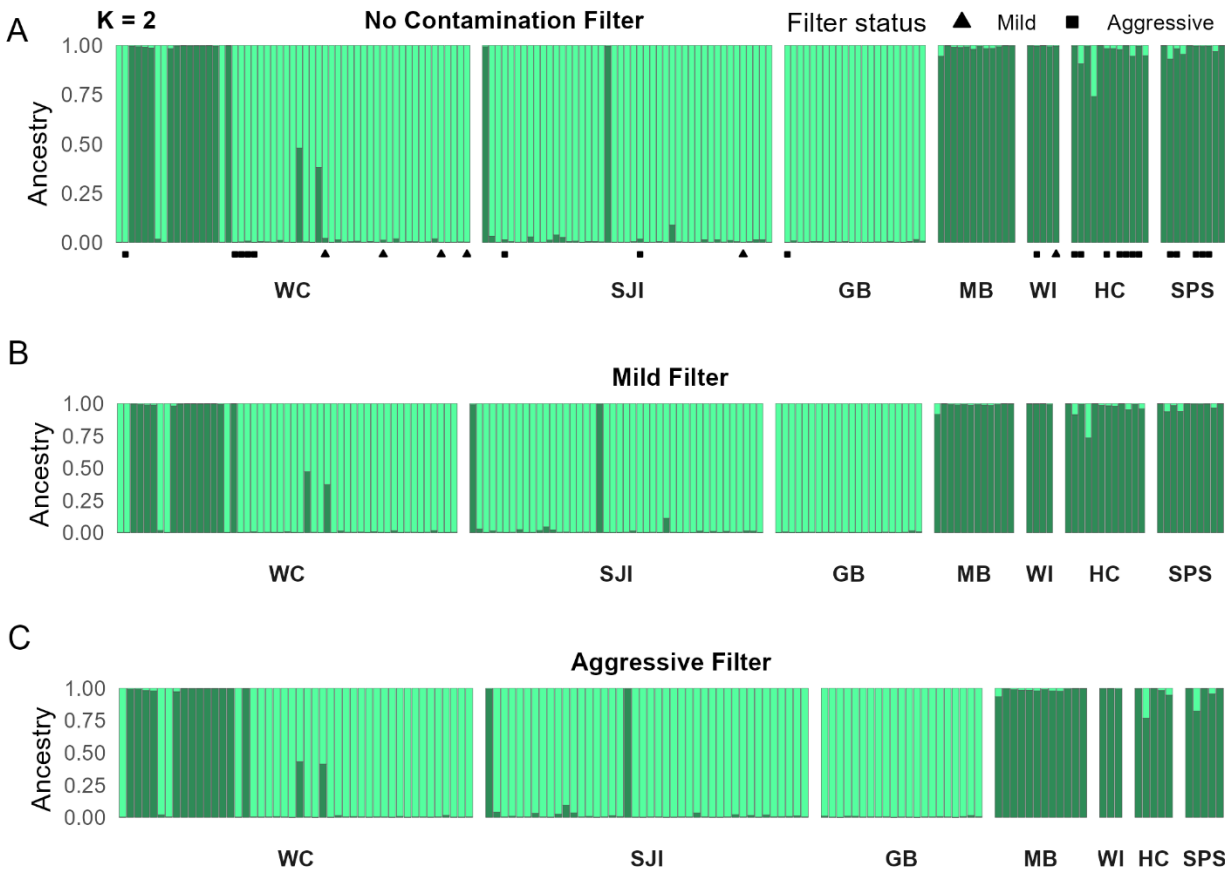


Figure 1.23 Quillback Rockfish STRUCTURE analysis at K = 2, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s). Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.

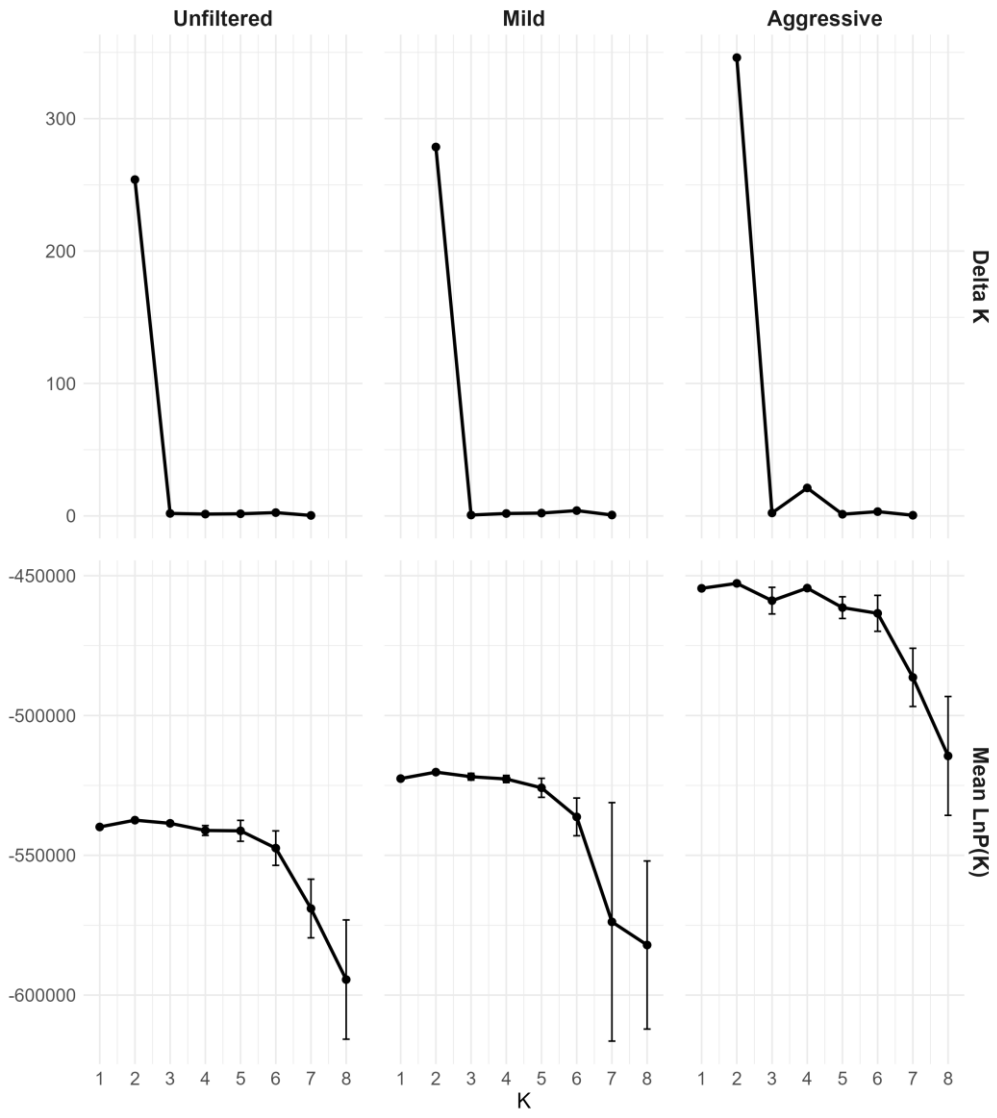


Figure 1.24 Delta K (top) and Mean LnP(K) (bottom, ±1 SD) for Quillback Rockfish STRUCTURE analysis. Filtering levels are indicated by column, left to right.

1.7 TABLES

Table 1.1 Interspecific Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels. Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Interspecific Rockfish Pairwise F_{ST}		
No Contamination Filtering		
	Brown	Copper
Copper	**0.618**	
Quillback	**0.605**	**0.367**
Mild Filtering		
	Brown	Copper
Copper	**0.623**	
Quillback	**0.611**	**0.369**
Aggressive Filtering		
	Brown	Copper
Copper	**0.628**	
Quillback	**0.616**	**0.371**

Table 1.2 Interspecific rockfish average hybrid ancestry (HA) per species. N = individuals per species and dataset and HA = percent hybrid ancestry. Hybrid ancestry is the proportion of STRUCTURE analysis ancestry that does not belong to the species' ancestry cluster, confirmed against samples with <1% ancestry from other clusters.

Interspecific Rockfish Average Hybrid Ancestry (HA)			
	Brown	**Copper**	**Quillback**
N Unfiltered	72	128	167
Unfiltered HA	2.58%	5.75%	0.44%
N Mild	64	125	158
Mild Filtering HA	2.11%	5.50%	0.33%
N Aggressive	39	103	139
Aggressive Filtering HA	3.24%	6.05%	0.29%

Table 1.3 MAC versus MAF filtering observed heterozygosity results. Values are from intraspecific datasets prior to contamination filtering. *N* = individuals per species, MAC = minor allele count, MAF = minor allele frequency, *Ho* = mean observed heterozygosity. Overall *Ho* was calculated using the R package hierfstat v0.5-11 (Goudet, 2005).

	MAC vs MAF Filtering Heterozygosity		
	N	*MAC* H_O	*MAF* H_O
Brown	69	0.163	0.212
Copper	123	0.125	0.232
Quillback	161	0.140	0.278

Table 1.4 Brown Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels. Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Brown Rockfish Pairwise F_{ST}					
No Contamination Filtering					
	SJI	**GB**	**MB**	**WC**	**SPS**
GB	-0.040				
MB	**0.034**	**0.041**			
WC	**0.091**	**0.23**	**0.176**		
SPS	**0.035**	**0.033**	0.001	**0.172**	
HC	**0.024**	**0.026**	0.001	**0.162**	-0.001
Mild Filtering					
	SJI	**GB**	**MB**	**WC**	**SPS**
GB	-0.040				
MB	**0.034**	**0.043**			
WC	**0.091**	**0.23**	**0.179**		
SPS	**0.034**	**0.033**	0.001	**0.173**	
HC	**0.021**	**0.024**	0.002	**0.161**	-0.001
Aggressive Filtering					
	SJI	**GB**	**MB**	**WC**	**SPS**
GB	-0.040				
MB	**0.035**	**0.044**			
WC	**0.091**	**0.23**	**0.181**		
SPS	**0.038**	**0.038**	-0.001	**0.193**	
HC	**0.022**	**0.03**	0.001	**0.17**	-0.001

Table 1.5 Summary statistics for Brown Rockfish at unfiltered, mild, and aggressive contamination filtering levels. *N* = sample size per population, *Ho* = mean observed heterozygosity, *He* = mean expected heterozygosity, F_{IS} = inbreeding coefficient. *Ho*, *He,* and F_{IS} values were calculated using the R package hierfstat v0.5-11 (Goudet, 2005).

	Brown Rockfish Summary Statistics											
	No Contamination Filtering				**Mild Filtering**				**Aggressive Filtering**			
	N	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}
WC	6	0.132	0.139	0.034	6	0.132	0.139	0.034	6	0.132	0.139	0.034
GB	1	0.198	N/A	N/A	1	0.198	N/A	N/A	1	0.198	N/A	N/A
SJI	3	0.204	0.246	0.089	3	0.204	0.246	0.089	4	0.204	0.246	0.089
MB	15	0.251	0.251	-0.003	15	0.251	0.251	-0.003	14	0.251	0.250	-0.004
HC	23	0.241	0.250	0.026	19	0.237	0.248	0.033	12	0.231	0.249	0.051
SPS	21	0.243	0.247	0.009	20	0.241	0.246	0.012	10	0.243	0.246	0.008
Overall	69	0.212	0.225	0.060	64	0.211	0.225	0.062	46	0.210	0.225	0.065

Table 1.6 Copper Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels. Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Copper Rockfish Pairwise F_{ST}						
No Contamination Filtering						
	WC	SJI	GB	WI	MB	SPS
SJI	**0.012**					
GB	**0.011**	**0.033**				
WI	**0.083**	**0.066**	**0.114**			
MB	**0.033**	**0.02**	**0.06**	**0.012**		
SPS	**0.06**	**0.041**	**0.095**	**0.055**	**0.013**	
HC	**0.078**	**0.059**	**0.108**	**0.093**	**0.047**	**0.005**
Mild Filtering						
	WC	SJI	GB	WI	MB	SPS
SJI	**0.011**					
GB	**0.011**	**0.032**				
WI	**0.083**	**0.065**	**0.114**			
MB	**0.033**	**0.02**	**0.06**	**0.012**		
SPS	**0.06**	**0.041**	**0.095**	**0.055**	**0.013**	
HC	**0.078**	**0.059**	**0.108**	**0.093**	**0.047**	**0.005**
Aggressive Filtering						
	WC	SJI	GB	WI	MB	SPS
SJI	**0.011**					
GB	**0.009**	**0.034**				
WI	**0.08**	**0.065**	**0.112**			
MB	**0.03**	**0.019**	**0.059**	**0.012**		
SPS	**0.068**	**0.047**	**0.116**	**0.083**	**0.03**	
HC	**0.09**	**0.068**	**0.129**	**0.113**	**0.06**	**0.012**

Table 1.7 Summary statistics for Copper Rockfish at unfiltered, mild, and aggressive contamination filtering levels. *N* = sample size per population, *Ho* = mean observed heterozygosity, *He* = mean expected heterozygosity, F_{IS} = inbreeding coefficient. *Ho*, *He,* and F_{IS} values were calculated using the R package hierfstat v0.5-11 (Goudet, 2005).

	Copper Rockfish Summary Statistics											
	No Contamination Filtering				**Mild Filtering**				**Aggressive Filtering**			
	N	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}
WC	19	0.218	0.237	0.063	19	0.218	0.237	0.063	17	0.222	0.241	0.059
GB	24	0.209	0.218	0.027	24	0.209	0.218	0.027	22	0.212	0.220	0.023
SJI	32	0.240	0.253	0.039	30	0.241	0.255	0.042	28	0.241	0.255	0.042
MB	13	0.242	0.265	0.064	13	0.242	0.265	0.064	13	0.242	0.265	0.064
WI	10	0.250	0.255	0.013	10	0.250	0.255	0.013	10	0.250	0.255	0.013
HC	15	0.228	0.239	0.038	15	0.228	0.239	0.038	4	0.224	0.227	-0.016
SPS	10	0.236	0.254	0.050	10	0.236	0.254	0.050	3	0.225	0.238	-0.009
Overall	123	0.232	0.246	0.057	121	0.232	0.246	0.057	97	0.231	0.243	0.051

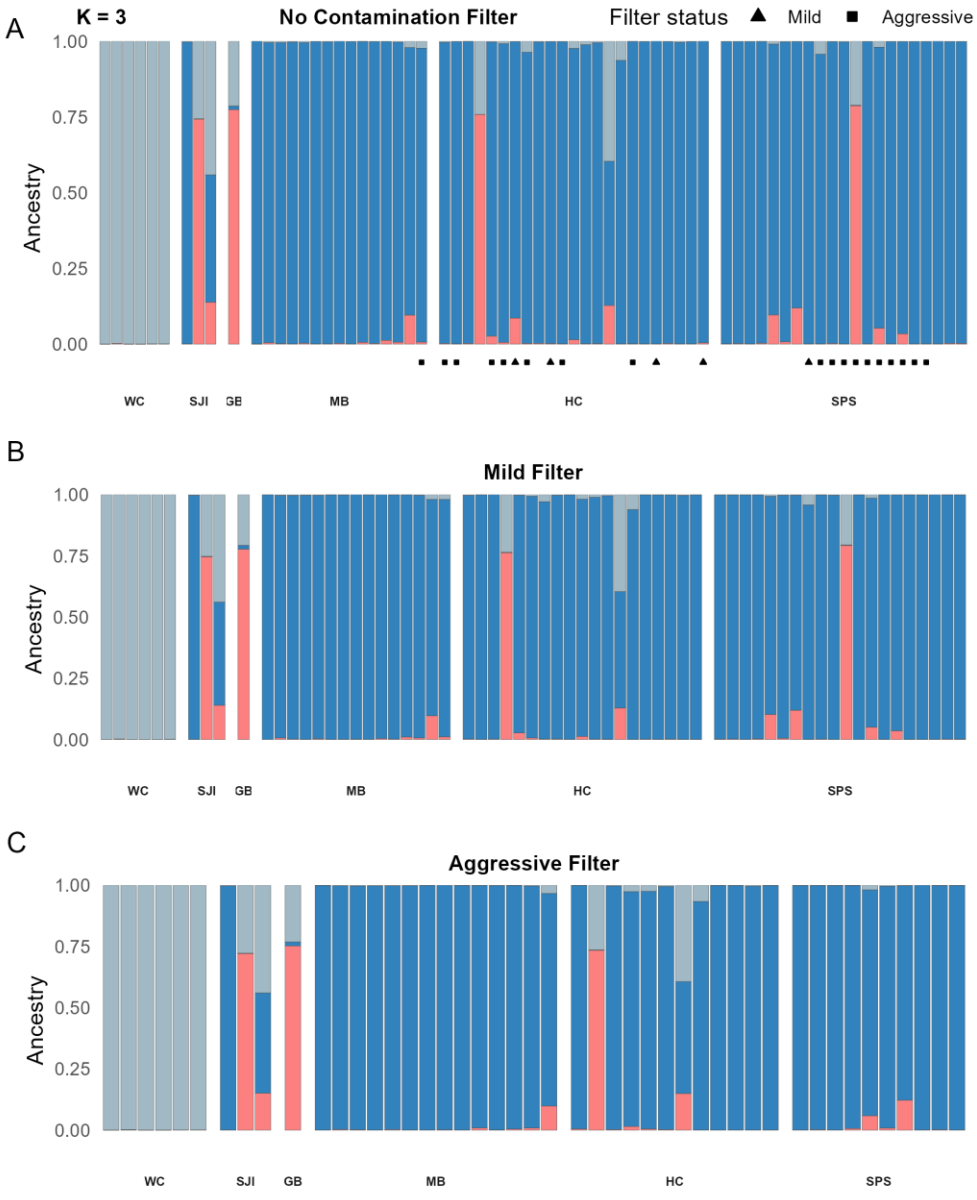
Table 1.8 Quillback Rockfish Pairwise Weir and Cockerham F_{ST} estimates at unfiltered, mild, and aggressive contamination filtering levels. Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Quillback Rockfish Pairwise F_{ST}						
No Contamination Filtering						
	WC	SJI	GB	SPS	WI	MB
SJI	**0.001**					
GB	**0.002**	0.001				
SPS	**0.006**	**0.010**	**0.014**			
WI	**0.006**	**0.010**	**0.015**	0.002		
MB	**0.007**	**0.010**	**0.013**	0.000	-0.004	
HC	**0.006**	**0.011**	**0.015**	0.001	0.002	**0.003**
Mild Filtering						
	WC	SJI	GB	SPS	WI	MB
SJI	**0.001**					
GB	**0.002**	0.001				
SPS	**0.005**	**0.010**	**0.014**			
WI	**0.006**	**0.010**	**0.015**	0.001		
MB	**0.007**	**0.010**	**0.013**	0.000	-0.005	
HC	**0.006**	**0.011**	**0.015**	0.001	0.001	**0.003**
Aggressive Filtering						
	WC	SJI	GB	SPS	WI	MB
SJI	**0.001**					
GB	**0.002**	0.001				
SPS	**0.004**	**0.011**	**0.016**			
WI	**0.008**	**0.013**	**0.019**	**0.008**		
MB	**0.006**	**0.010**	**0.013**	0.000	-0.002	
HC	**0.006**	**0.011**	**0.015**	-0.001	**0.008**	0.005

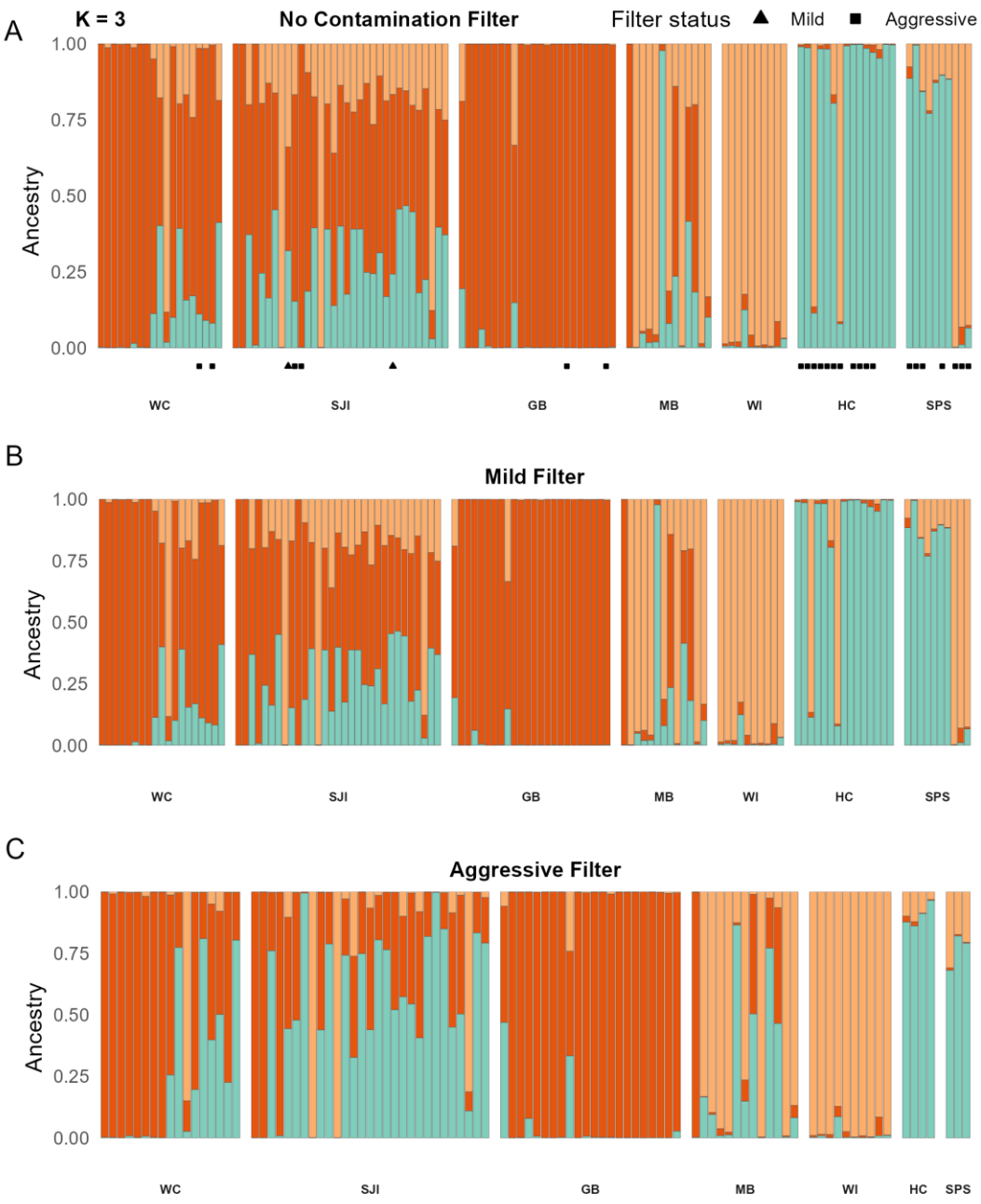
Table 1.9 Summary statistics for Quillback Rockfish at unfiltered, mild, and aggressive contamination filtering levels. *N* = sample size per population, *Ho* = mean observed heterozygosity, *He* = mean expected heterozygosity, F_{IS} = inbreeding coefficient. *Ho*, *He,* and F_{IS} values were calculated using the R package hierfstat v0.5-11 (Goudet, 2005).

	Quillback Rockfish Summary Statistics											
	No Contamination Filtering				**Mild Filtering**				**Aggressive Filtering**			
	N	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}
WC	55	0.266	0.277	0.035	51	0.267	0.278	0.036	46	0.266	0.278	0.037
GB	22	0.271	0.278	0.018	22	0.271	0.278	0.018	21	0.270	0.278	0.022
SJI	45	0.276	0.282	0.018	44	0.277	0.282	0.017	42	0.276	0.282	0.020
MB	12	0.278	0.284	0.012	12	0.278	0.284	0.012	12	0.278	0.284	0.012
WI	5	0.302	0.283	-0.068	4	0.299	0.283	-0.070	3	0.289	0.280	-0.068
HC	12	0.277	0.281	0.011	12	0.277	0.281	0.011	5	0.274	0.281	0.004
SPS	10	0.278	0.283	0.007	10	0.278	0.283	0.007	5	0.282	0.285	-0.008
Overall	161	0.278	0.282	0.011	155	0.278	0.282	0.013	134	0.276	0.282	0.018

1.8 SUPPLEMENTAL FIGURES



Supplemental Figure 1.1 Brown Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s). Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.



Supplemental Figure 1.2 Copper Rockfish STRUCTURE analysis at K = 3, grouped by sampling population at A) unfiltered, B) mild, and C) aggressive filtering levels. Each bar is an individual and is colored by its ancestry cluster(s). Symbols underneath bars in plot A indicate samples that are filtered out in B) mild filtering or C) aggressive filtering; triangle = mild filtering, square = aggressive filtering. Most likely K cluster was determined using ΔK.

1.9 REFERENCES

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. Genetics, 202(2), 389–400. https://doi.org/10.1534/genetics.115.183665

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data (0.12.0).

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

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., … National Eye Institute, N. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLOS ONE*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Ballenghien, M., Faivre, N., & Galtier, N. (2017). Patterns of cross-contamination in a multispecies population genomic project: Detection, quantification, impact, and solutions. *BMC Biology*, *15*(1), 25. https://doi.org/10.1186/s12915-017-0366-6

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Buonaccorsi, V. P., Kimbrell, C. A., Lynn, E. A., & Vetter, R. D. (2002). Population structure of copper rockfish (*Sebastes caurinus*) reflects postglacial colonization and contemporary patterns of larval dispersal. Canadian Journal of Fisheries and Aquatic Sciences, 59(8), 1374–1384. https://doi.org/10.1139/f02-101

Buonaccorsi, V. P., Kimbrell, C. A., Lynn, E. A., & Vetter, R. D. (2005). Limited realized dispersal and introgressive hybridization influence genetic structure and conservation strategies for brown rockfish, *Sebastes auriculatus*. Conservation Genetics, 6(5), 697–713. https://doi.org/10.1007/s10592-005-9029-1

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chen, H., Wang, B., Cai, L., Yang, X., Hu, Y., Zhang, Y., Leng, X., Liu, W., Fan, D., Niu, B., & Zhou, Q. (2024). A comprehensive performance evaluation, comparison, and integration of computational methods for detecting and estimating cross-contamination of human samples in cancer next-generation sequencing analysis. *Journal of Biomedical Informatics*, *152*, 104625. https://doi.org/10.1016/j.jbi.2024.104625

Dallavilla, T., Marceddu, G., Casadei, A., De Antoni, L., & Bertelli, M. (2020). A fast, reliable and easy method to detect within-species DNA contamination. *Acta Bio Medica: Atenei Parmensis*, *91*(Suppl 13), e2020019. https://doi.org/10.23750/abm.v91i13-S.10531

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Delafoy, D., Mercier, J., Larsonneur, E., Wiart, N., Sandron, F., Mejean, T., Meslage, S., Daian, D., Olaso, R., Boland, A., Deleuze, J.-F., & Meyer, V. (2021). *ContaTester: Fast cross contamination estimation and identification for large human sequencing cohorts* (p. 2021.10.01.461647). bioRxiv. https://doi.org/10.1101/2021.10.01.461647

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fiévet, A., Bernard, V., Tenreiro, H., Dehainault, C., Girard, E., Deshaies, V., Hupe, P., Delattre, O., Stern, M.-H., Stoppa-Lyonnet, D., Golmard, L., & Houdayer, C. (2019). ART-DeCo: Easy tool for detection and characterization of cross-contamination of DNA samples in diagnostic next-generation sequencing analysis. *European Journal of Human Genetics*, *27*(5), 792–800. https://doi.org/10.1038/s41431-018-0317-x

Flickinger, M., Jun, G., Abecasis, G. R., Boehnke, M., & Kang, H. M. (2015). Correcting for Sample Contamination in Genotype Calling of DNA Sequence Data. *The American Journal of Human Genetics*, *97*(2), 284–290. https://doi.org/10.1016/j.ajhg.2015.07.002

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics.

Molecular Ecology Notes, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Illumina, Inc. (2017). Effects of Index Misassignment on Multiplexing and Downstream Analysis. https://www.illumina.com/techniques/sequencing/ngs-library-prep/multiplexing/index-hopping.html

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jecha, K., Lavanchy, G., & Schwander, T. (2025). *Decontaminating genomic data for accurate species delineation and hybrid detection in the Lasius ant genus* (p. 2024.11.27.625433). bioRxiv. https://doi.org/10.1101/2024.11.27.625433

Jiang, T., & Motsinger-Reif, and A. A. (2024). Same-species Contamination Detection with Variant Calling Information from Next-generation Sequencing. *Journal of Cancer Science and Clinical Therapeutics*, *8*(1), 59–69. https://www.doi.org/10.26502/jcsct.5079225

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jun, G., Flickinger, M., Hetrick, K. N., Romm, J. M., Doheny, K. F., Abecasis, G. R., Boehnke, M., & Kang, H. M. (2012). Detecting and Estimating Contamination of Human DNA Samples in Sequencing and Array-Based Genotype Data. *The American Journal of Human Genetics*, *91*(5), 839–848. https://doi.org/10.1016/j.ajhg.2012.09.004

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*

Methods, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources*, *18*(4), 778–788. https://doi.org/10.1111/1755-0998.12779

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639 647. https://doi.org/10.1111/1755-0998.12995

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Maruki, T., & Lynch, M. (2017). Genotype Calling from Population-Genomic Sequencing Data. *G3 Genes|Genomes|Genetics*, *7*(5), 1393–1404. https://doi.org/10.1534/g3.117.039008

McCartney-Melstad, E., Bi, K., Han, J., & Foo, C. K. (2021). *VCFcontam: A Machine Learning Approach to Estimate Cross-Sample Contamination from Variant Call Data* (p. 2021.03.12.435007). bioRxiv. https://doi.org/10.1101/2021.03.12.435007

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669. https://doi.org/10.1111/1755-0998.12613

Muyas, F., Bosio, M., Puig, A., Susak, H., Domènech, L., Escaramis, G., Zapata, L., Demidov, G., Estivill, X., Rabionet, R., & Ossowski, S. (2018). Allele balance bias identifies systematic genotyping errors and false disease associations. *Human Mutation*, *40*(1), 115–126. https://doi.org/10.1002/humu.23674

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, *27*(16), 3193–3206. https://doi.org/10.1111/mec.14792

Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Petrou, E. L., Drinan, D. P., Kopperl, R., Lepofsky, D., Yang, D., Moss, M. L., & Hauser, L. (2019). Intraspecific DNA contamination distorts subtle population structure in a marine fish: Decontamination of herring samples before restriction-site associated sequencing and its effects on population genetic statistics. *Molecular Ecology Resources*, *19*(5), 1131–1143. https://doi.org/10.1111/1755-0998.12978

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

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/10.1111/17550998.12512

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Schmidt, T. L., Jasper, M.-E., Weeks, A. R., & Hoffmann, A. A. (2021). Unbiased population

heterozygosity estimates from genome-wide sequence data. *Methods in Ecology and Evolution*, *12*(10), 1888–1898. https://doi.org/10.1111/2041-210X.13659

Schwenke, P. L., Park, L. K., & Hauser, L. (2018). Introgression among three rockfish species (Sebastes spp.) in the Salish Sea, northeast Pacific Ocean. *PLOS ONE*, *13*(3), e0194068. https://doi.org/10.1371/journal.pone.0194068

Seeb, L. (1998). Gene flow and introgression within and among three species of rockfishes, Sebastes auriculatus, S. caurinus, and S. maliger. *Journal of Heredity*, *89*(5), 393–403. https://doi.org/10.1093/jhered/89.5.393

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917. https://doi.org/10.1111/2041-210X.12700

Toyama, K. S., Crochet, P.-A., & Leblois, R. (2020). Sampling schemes and drift can bias admixture proportions inferred by structure. *Molecular Ecology Resources*, *20*(6), 1769–1785. https://doi.org/10.1111/1755-0998.13234

Ward, A. K., Scopel, E. F. C., Shuman, B., Momany, M., & Bensasson, D. (2026). Low-level contamination confounds population genomic analysis. *G3 Genes|Genomes|Genetics*, *16*(4), jkag021. https://doi.org/10.1093/g3journal/jkag021

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (2016). Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wray, A., Petrou, E., Nichols, K. M., Pacunski, R., LeClair, L., Andrews, K. S., Kardos, M., &

Hauser, L. (2024). Contrasting effect of hybridization on genetic differentiation in three rockfish species with similar life history. *Evolutionary Applications*, *17*(7), e13749. https://doi.org/10.1111/eva.13749

Yi, X., & Latch, E. K. (2022). Nonrandom missing data can bias Principal Component Analysis inference of population genetic structure. *Molecular Ecology Resources*, *22*(2), 602–611. https://doi.org/10.1111/1755-0998.13498

Chapter 2. INFLUENCE OF PUGET SOUND BASIN STRUCTURE AND HYBRIDIZATION ON POPULATION GENETICS OF THREE PACIFIC ROCKFISH SPECIES (*Sebastes* spp.)

2.1 ABSTRACT

Hybridization can complicate fisheries management by blurring species boundaries but can also provide important adaptive genetic variation. Brown Rockfish (*Sebastes auriculatus*), Copper Rockfish (*S. caurinus*), and Quillback Rockfish (*S. maliger*) are known to hybridize in Puget Sound. They have not recovered despite a full fishery closure following overfishing, unlike their conspecifics on the coast which are now fully recovered. Population genetic structure is still uncertain given conflicting results regarding the spatial extent of introgression and its influence on population structure. To examine these patterns at finer spatial scales with a larger sample size, we combined previously published restriction site-associated DNA sequencing (RAD-seq) data with 389 newly sequenced individuals sampled from the West Coast and the Salish Sea. Our results confirmed that introgression is generally low-level, widespread, and asymmetric from Quillback Rockfish into Brown and Copper Rockfish, and that intraspecific population structure is species-specific, consistent with a previous RAD-seq study. We found no evidence for previously described higher incidence of hybrids in more isolated basins of Puget Sound. However, we detected another population of Copper Rockfish in Hood Canal, which was not due to increased hybridization. Previous studies included the use of mtDNA, which may be the only trace of hybridization in most individuals. Our findings demonstrate that population structure differs substantially among species and suggest the importance of directional mtDNA introgression in interspecific hybridization.

2.2 INTRODUCTION

Fisheries management and policy are dependent on explicitly defined species and populations, the identification of which are complicated by the presence and extent of hybridization between those species (Allendorf et al., 2001). As genomic research has advanced rapidly and reduced in cost in recent years with the development of high-throughput sequencing, evidence of hybridization in a vast variety of organisms has been uncovered, and outcomes of these hybridization events appear to be more variable than previously thought (Taylor and Larson, 2019). While hybridization may pose threats to wild populations, such as introgression from domestic species (Howard-McCombe et al., 2021) or outbreeding depression, in which offspring from two different populations have reduced fitness (Houde et al., 2011), natural hybridization is an important evolutionary process (Taylor and Larson, 2019) and can contribute positively to genetic diversity. For instance, species that hybridize and repeatedly backcross can retain beneficial alleles (Hedrick, 2018; Grant and Grant, 2010) – known as adaptive introgression – which may be crucial for species to survive in the face of rapidly changing environments (Brauer et al., 2023; Hamilton and Miller, 2015). It is therefore important that the context of hybridization is considered when crafting management tactics for hybridizing species, especially in cases of populations that are declining or threatened.

Pacific rockfishes (*Sebastes* spp.) in Puget Sound, Washington, are a prime example of hybridization complicating efforts for restoring declining populations. Rockfish are a vital component of the Pacific Coast groundfish fishery, comprising approximately 75% of the species in the fishery (PFMC Groundfish FMP, 2025). After rockfish populations sharply declined from overfishing in the 1970s-1980s and crashed in the mid-1990s (Williams et al., 2010), the Puget Sound commercial fishery was closed in 1999. Nevertheless, the populations did not recover, and

in 2010, the recreational fishery was also closed after years of continually declining bag limits (the maximum number of fish an angler can retain in a day). Rockfish are characterized by late maturity, longevity and slow growth (Love et al., 2002; Winemiller, 2005), which can lead to a significant lag in population recovery after a period of overfishing. Since fishery closures, populations along the coast of Washington have recovered, but many species in Puget Sound are still depleted or in a cautionary state (WDFW, 2011). Though the fishery remains closed, rockfishes in Puget Sound were previously managed using distinct population segment (DPS) designations that were determined using data from earlier genetic studies on hybridizing Brown Rockfish, Copper Rockfish, and Quillback Rockfish (Buonaccorsi et al., 2005, 2002; Seeb, 1998). These studies established that these three species of rockfishes exhibit complex and often asymmetric patterns of differentiation and hybridization based on allozyme and microsatellite data (Seeb, 1998; Buonaccorsi et al., 2002, 2005), which was also displayed in a later study that used few nuclear and one mitochondrial loci (Schwenke et al., 2018). Wray et al. (2024) recently provided evidence for species-specific genetic structure of rockfish populations, with some species displaying differentiation between coastal and Puget Sound populations while others did not, such as Black Rockfish and Canary Rockfish (Andrews et al., 2018). Brown Rockfish, Copper Rockfish, and Quillback rockfish all exhibited intraspecific population structure, but spatial patterns differed between species.

Several factors could contribute to the lack of recovery in Puget Sound rockfishes, such as derelict gear, habitat degradation due to nearshore development, pollution, and climate change (Drake et al., 2010). This difference in recovery between coastal and Puget Sound populations may also be due to barriers to connectivity caused by the unique bathymetric and oceanographic features of the region. Puget Sound, an estuarine fjord (Dethier and Schoch, 2004; Burns 1985)

located in the larger marine system known as the Salish Sea, features bathymetric barriers that provide the opportunity for populations to become isolated and, thus, genetically differentiated from those found on the coast – two major barriers being Victoria Sill at the eastern edge of the Strait of Juan de Fuca, and another at Admiralty Inlet between Whidbey Island and Port Townsend. These shallow sills exist between the deep, narrow basins of Puget Sound, leading to changes in water flow (such as gyres) that may limit larval dispersal, particularly in demersal rockfish species (Palsson et al., 2009). Adults of some species are also known to have high site fidelity (Palsson et al., 2009), further reducing the chance for gene flow between populations in these basins. While a lack of connectivity to coastal populations could affect genetic diversity, it may also allow for local adaptation, in which populations become better suited to their local environment than other members of the same species (Blanquart et al., 2013; Forester et al., 2016) – local adaptation could be especially important for species survival in increasingly extreme environmental conditions (Meek et al., 2023).

The bathymetric structure and environmental conditions of Puget Sound are also favorable toward the formation of hybrid zones. Puget Sound features variable temperatures, lower salinity, and less rocky reef habitat than the coast, as well as seasonal periods of anoxia (as cited in Schwenke et al., 2018), which may contribute to overlap in species that are normally separated by depth. Three species of rockfishes found in Puget Sound – Brown Rockfish, Copper Rockfish, and Quillback Rockfish – are known to hybridize based on allozyme and microsatellite work (Drake et al., 2010; Buonaccorsi et al., 2002, 2005; Seeb, 1998). More recent work using mitochondrial and nuclear loci demonstrated asymmetrical introgression (from Quillback Rockfish into Brown Rockfish and Copper Rockfish) that contrasted with the directionality of introgression in an earlier allozyme study (Seeb, 1998). Schwenke et al. (2018) also reported a

high percentage of interspecific hybrids and long-term hybridization, with proportions of hybrids increasing with the number of sills separating capture locations from the open ocean. However, the most recent study on hybridization in these three species – which used restriction site associated DNA sequencing (RAD-seq), a reduced representation sequencing method – reported widespread yet consistent, low levels of admixture from Quillback Rockfish into Brown Rockfish and Copper Rockfish (Wray et al., 2024), again contrasting with previous studies. This study additionally had few to no individuals from basins where Schwenke et al. (2018) reported higher hybridization rates (depending on species and basin), which may have contributed to the lower levels.

To date, the current body of work on this rockfish species complex contains conflicting conclusions, likely due in part to the difference in the markers used, advancements in genomic techniques, and spatial extent of sampling in the Salish Sea. The data presented in Wray et al. (2024) laid a critical foundation for updating management tactics to better reflect each species' respective population structure, but sampling was largely restricted to central Puget Sound, omitting adjacent smaller basins with potentially higher incidence of hybridization (Schwenke et al. 2018). To address these gaps, this study expanded on both sample size and sampling at the geographic basin level. Individuals from Wray et al. (2024) were combined in a dataset with new individuals and divided into the basin of origin in order to examine population structure and hybridization at a finer scale. By including basins that were previously identified as having higher rates of hybrid individuals, as well as substantially larger sample sizes within each basin, we aimed to examine finer-scale population structure in Puget Sound rockfishes and the influence of these species' history of hybridization on these patterns.

2.3 METHODS

2.3.1 Sampling

389 fin clip and muscle tissue samples of Brown Rockfish, Copper Rockfish, and Quillback Rockfish were assembled from collections at the Washington Department of Fish and Wildlife (WDFW), the National Oceanic and Atmospheric Administration (NOAA), and Fisheries and Oceans Canada (DFO). Samples were stored either in 95% ethanol tubes and kept refrigerated, or as a dry fin clip in an absorbent paper container inside a plastic bag with moisture-absorbing beads stored at room temperature. Nine individuals sequenced in this study were marked as "possible hybrids" by the original collectors in the field, as the whole specimens appeared to have phenotypic traits (usually coloration) that were not consistent with a single species. Samples were assigned to one of seven regions of the coast or Salish Sea (Figure 1.1) based on their capture coordinates or approximate location (i.e. named sampling location). To estimate genotyping error and check for data comparability, 28 individuals from Wray et al. (2024) (hereafter referred to AW) were re-sequenced, and 70 new individuals (11 Brown Rockfish, 23 Copper Rockfish, and 36 Quillback Rockfish) were sequenced twice; libraries of 26 individuals were prepared in two different libraries from the same DNA extract, and 44 individuals were extracted in two separate extraction plates. These individuals (re-sequenced individuals from AW and individuals sequenced twice in this study) will hereafter be referred to as "technical replicates."

2.3.2 DNA Extraction, RAD-Seq Library Preparation, and Sequencing

DNA was extracted from samples using a Nexttec DNA Isolation Kit (nexttec™ Biotechnologie GmbH, Leverkusen, Germany). Within each of our five 96-well extraction plates,

a single random well was used as a negative control to check for contamination. This consisted of a no-template control (NTC) method, where all protocol reagents are still added to the well, but no DNA is added. DNA was quantified on a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA). Extractions were normalized to 125ng in 10µl and assorted into five libraries; four libraries contained randomized sample groups to avoid batch effect (Leigh et al., 2018), while the fifth library contained only samples of poor quality and our five negative controls to avoid amplification bias during the PCR stage. Library preparation followed an adapted version of existing RAD-seq protocols (Ali et al., 2016) known as BestRAD. Genomic DNA was digested with SbfI, and a P1 adapter containing a forward amplification primer, an Illumina primer, and an individual-specific 6 bp barcode was ligated to the fragments at the restriction site end. Fragments were then sheared randomly via sonication in a Diagenode Bioruptor 300 (Hologic Inc., Marlborough, MA) and size-selected to a length of 300-600 bp. Finally, P2 adapters were then ligated to the reverse-end and libraries were amplified via PCR. Each library was quality checked on a 1% agarose gel post-PCR and a 2100 Expert Bioanalyzer (Agilent Technologies, Santa Clara, CA) when completed. Libraries were then sent to the University of Oregon for equimolar pooling and sequencing on a single NovaSeq 6000 S4 lane (paired end, 150 bp).

2.3.3 Data Processing and Initial Filtering

Raw sequences were quality checked using FastQC v0.12.0 (Andrews, 2010) and visualized with MultiQC (Ewels et al., 2016). Sequences were demultiplexed, assigned genotypes and filtered for quality using the STACKS v2.64 process_radtags program (Catchen et al., 2013; Rochette et al., 2019). PCR duplicates were removed using the STACKS clone_filter

program. Sequences were then aligned using Bowtie 2 (Langmead and Salzberg, 2012) to the Honeycomb Rockfish genome (*Sebastes umbrosus*; NCBI BioProject accession number: PRJNA562006), as it is the closest relative to our three species of *Sebastes* that had a chromosome-level full genome assembly available at the time of this study. After alignment, individuals with fewer than 300,000 reads were removed based on our within-dataset distribution of read counts to exclude poorly sequenced samples, as their inclusion may cause higher rates of missing data and genotyping error (O'Leary et al., 2018). BAM files generated in this study were combined into a dataset with BAM files from the AW study (retaining only the samples that were also >300,000 reads; 37 Brown Rockfish, 84 Copper Rockfish, and 86 Quillback Rockfish), with replicate samples as separate files to check for differences in genotyping error between the different sequencing runs. This produced a dataset of 542 samples: 105 Brown, 201 Copper, and 236 Quillback Rockfish (including all replicates). SNP calling and population genetics statistics were calculated using STACKS *gstacks* and *populations* modules for both interspecific and intraspecific datasets.

For the interspecific data set, individuals were assigned to their morphological species identification in population assignment, and all three species were processed together. Data were filtered in the *populations* module to include loci present in >80% individuals (flag R80), in at least one population (P1) and only the first SNP per RAD tag (*write-single-snp)*. We chose the first SNP on each RAD tag, because SNPs with the highest F_{ST} may select SNPs with fixed differences between two of the three species and limited differentiation between the other two species. For the three intraspecific datasets**,** the SNP with the highest minor allele frequency within each species was chosen. We chose this method to keep our dataset as comparable as possible to AW, who used this strategy specifically to select different SNPs for intra- and

interspecific analyses (Wray et al., 2024). SNPs that were significantly out of Hardy-Weinberg Equilibrium (HWE) were removed using the following procedure: first, p-values were calculated for each sample at each locus using the R package pegas v1.4 (Paradis, 2010), then p-values were combined for every locus using Fisher's combination of probabilities. The resulting locus specific p-values were adjusted to q-values for false discovery rate (Benjamini & Hochberg, 1995) and loci with q-values less than 0.05 were removed. For both datasets, individual SNPs were called with a minimum mapping quality of 40 and filtered using a minimum read depth of 10. Across all individuals, SNPs were retained with a minimum average read depth of 15 and less than 20% missing data (O'Leary et al., 2018). All three intraspecific datasets were additionally filtered using minor allele frequency (MAF) < 0.05. Genotype discordance between replicate samples was calculated using the script "genoerrorcalc" (https://github.com/bsjodin/genoerrorcalc). Filtered VCFs were generated using VCFtools (Danecek et al., 2011). To convert data to other formats, PGDSpider v3.0.0.0 (Lischer & Excoffier, 2012) was used. All four datasets were then processed through a contamination filtering step as detailed in Chapter 1. In brief, individuals were analyzed for a combination of 1) substantial deviation from a mean allelic balance of 0.5 in heterozygous SNPs (ratio of reference reads to alternate reads) and 2) elevated proportion of heterozygous SNPs with a significant allelic imbalance based on a two-tailed binomial test with p = 0.5. Individuals that deviated strongly in both categories were removed in the "mild" filtering level.

2.3.4 Analysis

Principal component analysis (PCA) was conducted using the R package adegenet v2.1.11 (Jombart, 2008). Per-group observed and expected heterozygosity, F_{IS}, and pairwise F_{ST}

values were obtained using the R package hierfstat v0.5-11 (Goudet, 2005); pairwise F_{ST} values between each dataset's sampling groups were calculated with 1,000 bootstraps. For our analysis comparing hybrid ancestry and F_{ST} in Copper Rockfish, individuals were first divided into hybrid ancestry categories: 'pure' individuals had <1%, low 1-6%, and high hybrid ancestry >6%. Pairwise F_{ST} was calculated between each group and the pure GB cluster, then mean hybrid ancestry percentage of each group was plotted against their F_{ST} values when compared with GB. STRUCTURE v2.3.4 (Pritchard et al., 2000) was used to estimate admixture, with five replicates for K = 1-N clusters (N = 5 for interspecific, 7 for Brown Rockfish, and 8 for both Copper Rockfish and Quillback Rockfish) and a burn-in of 10,000 iterations and 100,000 MCMC repetitions. We ran STRUCTURE using the admixture model and without a priori population information; likely K groups were identified with the ΔK statistic (Evanno et al., 2005) and the mean L(K) using Structure Harvester (Earl and vonHoldt, 2012). To avoid issues arising from label switching across replicate runs in STRUCTURE, we used CLUMPP (Jakobsson and Rosenberg, 2007) to align replicate files. Results were visualized in R using ggplot2 v3.5.2 (Wickham, 2016).

2.4 RESULTS

2.4.1 Interspecific Dataset

Per-individual average read counts retained from sequences collected in this study after reference genome alignment for Brown, Copper, and Quillback Rockfish samples were 3.02 million, 4.98 million, and 4.65 million, respectively. 11.7 million, 18.0 million, and 18.7 million per-individual average read counts were obtained by AW. Average read depth for sequences collected in this study After combining all BAM files from reference-aligned individuals from

this study and the AW study into a single dataset, 40 individuals with fewer than 300,000 reads and 63 individuals with more than 25% missing genotypes were removed. Average read depth for AW individuals was 83.4, and the average for individuals sequenced in this study was 72.5. Overall genotype discordance between replicate samples was an average of 2.77%. Comparisons between individuals first sequenced in Wray et al. (2024) and re-sequenced in this study resulted in a discordance rate of 4.54%, with four individuals >20% discordant (the average read depth across these four original individuals and their respective replicates was 44.6x). Within this study, the discordance rate was 0.92%. After quality filtering and removal of technical replicates (either the replicate first sequenced in the AW study, or the replicate in the second extraction plate or library run for that individual fish was removed), 8,685 SNPs were identified in this 391-individual dataset. In our initial STRUCTURE analysis, 24 individuals had full ancestry corresponding to a species that differed from their field identification; morphological identification for these individuals were two Brown Rockfish, three Quillback Rockfish, and 19 Copper Rockfish. 17 of these 24 individuals showed complete Quillback ancestry in STRUCTURE, and one Copper individual had roughly equal ancestry from Quillback and Copper (48% Quillback and 51% Copper, also previously identified as such in Wray et al., 2024). When plotted using PCA, these individuals also clustered with the species that matched their genetic ancestry rather than morphological identification, so they were labeled as misidentifications and removed from the dataset prior to downstream analyses. After removing these individuals and individuals that were flagged using our contamination filtering strategy described in Chapter 1 section 1.3.3, our final interspecific dataset consisted of 347 individuals (64 Brown Rockfish, 125 Copper Rockfish, and 158 Quillback Rockfish).

Interspecific STRUCTURE analysis revealed distinct ancestry clusters with most sampling groups in each species displaying low but consistent levels of ancestry from other clusters (Figures 2.2A-C). Within sampling groups, individuals are ordered by ID, therefore individuals from Wray et al. (2024) are at the beginning of every group (though some groups do not have individuals from AW). Ancestry percentage that is not from the species' primary ancestry cluster (confirmed against individuals that had <1% ancestry from other clusters, referred to as "pure" individuals) will hereinafter be referred to as 'hybrid ancestry.' Averages across individuals for hybrid ancestry in each species were 2.11% in Brown Rockfish, 5.50% in Copper Rockfish, and 0.33% in Quillback Rockfish (Table 2.1). All Brown Rockfish WC individuals were pure, and the remaining sampling regions all had <2.49% hybrid ancestry. Copper Rockfish had notably higher averages of hybrid ancestry than Brown Rockfish and Quillback Rockfish – WI had an average of 13.3%, which is the highest across all samples in all three species (all WI samples had >10% hybrid ancestry with a maximum of 16.4%). Copper Rockfish WC individuals sampled from Neah Bay contained hybrid ancestry while individuals sampled further east in the Strait of Juan de Fuca were largely pure, despite being further from the coast than Neah Bay. For both Copper Rockfish and Quillback Rockfish, the majority of GB individuals were pure (66.7% and 63.6%, respectively), while HC, SPS, and WI had no pure individuals for either species. Quillback Rockfish had consistently high rates of pure individuals in SJI (52.3%) and WC (47.3%). Of the pure Quillback Rockfish individuals in WC, 69.2% were from the coast of Vancouver Island, further west than the rest of our WC sampling sites. All Quillback Rockfish individuals had lower than 3% hybrid ancestry. In PCA, the three species clustered distinctly, with the suspected F1 individual located between Quillback Rockfish and Copper Rockfish clusters (Figure 2.1D). This F1 was removed during misidentification filtering

(as it contained nearly 50/50 ancestry from Copper Rockfish and Quillback Rockfish in our initial STRUCTURE analysis prior to this filtering), but STRUCTURE results from Wray et al. (2024) display ancestry proportions for this individual.

2.4.2 Intraspecific Datasets

In **Brown Rockfish**, we retained 3,817 SNPs in 64 individuals after filtering for quality, morphological misidentification, and contamination. This dataset shared 16.2% of the SNPs contained in the 8,685-SNP interspecific dataset. In PCA (Figure 2.3A), three groups were visible at the unfiltered level (two small clusters in the middle consisted of two and four individuals each). West Coast (WC) individuals were completely separated from the remaining individuals; all six fish were collected from California and Mexico, therefore genetic distinction from Puget Sound fish may be due to geographic isolation. In a STRUCTURE analysis (Figure 2.3B), three genetic clusters (K = 3) were most supported (Supplemental Figure 2.1). WC contained no ancestry from other clusters, and the main basin (MB) of Puget Sound consisted predominantly of a single cluster. In contrast, SJI, GB, HC and SPS had mixed ancestry; individuals from sites north of Admiralty Inlet (SJI, GB) were especially variable in cluster proportions, though this may be primarily due to the substantially smaller sample sizes for these regions (N = 3 and 1, respectively). Observed heterozygosity was consistently higher in all Puget Sound regions compared to sampling locations north of Admiralty Inlet, with WC showing the lowest value (Table 2.2). In contrast, expected heterozygosity was relatively constant in the Salish Sea, resulting in considerable variation in F_{IS} values. Pairwise F_{ST} values were significant for most sampling location comparisons; WC was highly differentiated from all other sampling

groups, while samples south of Admiralty Inlet showed low and non-significant differentiation (Table 2.3).

In **Copper Rockfish**, we retained 7,092 SNPs in 121 individuals after filtering for quality, morphological misidentification, and contamination. This dataset shared 18.8% of the SNPs contained in the 8,685-SNP interspecific dataset. A PCA revealed four groups (Figure 2.4A); GB individuals clustered tightly together with some individuals from WC and SJI, trailing into a nearby cluster that contained primarily SJI and WC individuals. HC and SPS formed a separate cluster with only three individuals from these groups found in the final cluster, which contained the remaining individuals from primarily WI and MB (with some individuals from other groups, such as SJI). At K = 4, STRUCTURE analysis indicates a differentiated HC and SPS cluster, with mixing between clusters in WC, SJI, and MB; WI and GB displayed little ancestry from other clusters (Figure 2.4B). ΔK strongly supported K = 5, though the fifth cluster is almost exclusively formed by two individuals in SJI and very minor ancestry proportions in others (Supplemental Figure 2.2, 2.5). Observed heterozygosity was lowest in GB and highest in WI (Table 2.2), which corresponded with interspecific average hybrid ancestry for these sampling groups (GB lowest at 0.70%, WI highest at 13.3%; Table 2.1). Pairwise F_{ST} values were significant for all comparisons and revealed three clusters: differentiation was relatively low among groups north of Admiralty Inlet (WC, SJI, GB), between MB and WI, and between HC and SPS. Between these three groups, differentiation was markedly higher and generally consistent with broad clustering in our other analyses.

To assess the effect of hybridization on Copper Rockfish population structure, we repeated the PCA and F_{ST} using datasets with only pure individuals (<1% hybrid ancestry, 40 individuals) as determined by the interspecific STRUCTURE analysis (Fig. 2.2). PCA reduced

the distinguishable clusters to a single main cluster; both HC (of 15 total) and five WC (of 19 total) individuals separated from the main cluster (Figure 2.5). Pairwise F_{ST} values were significant for six of the ten comparisons and >0.1 for all four HC comparisons (Table 2.5), though given that HC was represented by only two individuals in this analysis, these results should be interpreted cautiously. Plotting sampling groups' relative hybrid ancestry (grouped within sampling location as pure, low, or high interspecific admixture) against their pairwise F_{ST} comparisons (see method in 2.3.4) revealed a significant but weak positive relationship (R^2 = 0.328, p-value = 0.01635; Figure 2.6). Our correlation between hybrid ancestry percentage and intraspecific ancestry percentage from the GB cluster displayed a moderate, significant negative correlation when all individuals were included (R^2 = 0.483, p-value <0.001). When HC and SPS were excluded, this became a significant strong correlation (R^2 = 0.898, p-value <0.001; Figure 2.7).

In **Quillback Rockfish**, we retained 3,941 SNPs in 155 individuals after filtering for quality, morphological misidentification, and contamination. This dataset shared 15.3% of the SNPs contained in the 8,685-SNP interspecific dataset. Two strong clusters were seen in PCA: one consisted of exclusively WC, SJI, and GB individuals, and the other consisted of groups south of Admiralty Inlet, plus many individuals from WC and two from SJI (Figure 2.8A). Observed heterozygosity was largely consistent amongst sampling groups, with WI displaying slightly higher heterozygosity in comparison to other groups. STRUCTURE analysis (Figure 2.8B) revealed two clusters that were strongly supported (K = 2; Supplemental Figure 2.3) and displayed minimal mixing. Pairwise F_{ST} values revealed low but significant differentiation between WC and all other groups (lowest against GB and SJI), with higher values in

comparisons between SJI or GB against sampling groups south of Admiralty Inlet (Table 2.6), which showed mostly no significant differentiation amongst each other.

2.5 DISCUSSION

Our analyses provide support that population structure in Brown Rockfish, Copper Rockfish, and Quillback Rockfish in Puget Sound is species-specific and influenced by both hybridization and geographic features. By combining previously analyzed individuals from Wray et al. (2024) and additional individuals from coastal areas and peripheral Puget Sound basins, we were able to further substantiate the directionality of introgression from Quillback Rockfish into Brown Rockfish and Copper Rockfish, as well as the low-level and widespread hybrid ancestry seen in those two species primarily in basins south of Admiralty Inlet. However, our finer scale examination of intraspecific population structure revealed that hybridization alone may not fully explain the structure patterns in Copper Rockfish, as SPS and HC remained differentiated even after allowing for hybridization. In Brown Rockfish, geographic isolation from WC individuals and a potential mixing zone in north Puget Sound appear to drive differentiation, while two distinct populations characterize Quillback Rockfish structure despite absence of noticeable introgression. Altogether, our results demonstrate that asymmetric hybrid ancestry, limited dispersal, and geographic barriers are consistent with the observed fine-scale population structure within these rockfish species.

2.5.1 Hybrid ancestry

Interspecific analyses corroborate previous findings (Wray et al., 2024; Schwenke et al., 2018) that introgression is predominantly asymmetric from Quillback Rockfish into Brown Rockfish and Copper Rockfish, and that this introgression is widespread, ancient, and relatively low in these two species. Copper Rockfish displayed higher levels of introgression than Brown Rockfish, while Quillback Rockfish showed consistently minimal introgression (0.33% average) – these patterns are consistent with Wray et al. (2024) and with prior work identifying uneven introgression rates between these species with this same directionality (Schwenke et al., 2018). These results also confirm Wray et al. (2024) in suggesting that the extreme ranges in the degree of hybridization seen in Schwenke et al. (2018) and Buonaccorsi et al. (2002, 2005) were likely due to their use of mitochondrial DNA (mtDNA) and microsatellites, respectively, as opposed to the thousands of nuclear loci we and Wray et al. (2024) achieved using RAD-seq. Because of their high variability and homoplasy, microsatellites may be unable to differentiate between introgressed individuals and parental species (Nielsen et al., 2014). Elevated hybridization rates in Schwenke et al. (2018) may be due to use of one mitochondrial marker alongside three nuclear introns and one nuclear exon. The discrepancy in mitochondrial and nuclear DNA is known as mitonuclear discordance (Toews and Brelsford, 2012) and has been previously documented in other species of *Sebastes* (Burford and Bernardi, 2008).

Our work supports previous findings that hybrids occur primarily in the basins of Puget Sound and become extremely infrequent north of the San Juan Islands (Wray et al., 2024; Schwenke et al., 2018; Seeb, 1998). Sampling groups in Puget Sound consistently showed higher hybrid ancestry than coastal and north Salish Sea groups, and the majority of our pure individuals originated from WC/GB in Copper Rockfish and WC/SJI in Quillback Rockfish. The

concentration of hybridization in Puget Sound is consistent with mechanisms proposed in Schwenke et al. (2018) and Wray et al. (2024): reduced availability of rocky substrate, periodic hypoxic events, and higher abundance in Quillback Rockfish than Brown Rockfish or Copper Rockfish. Rocky reefs utilized by adults of all three species are less frequent south of Admiralty Inlet (Williams et al., 2010), which can increase overlap among individuals of species that normally occupy different depths (Love et al., 2002). Furthering the potential for species overlap are periodic low oxygen events (hypoxia), which could force Quillback Rockfish (the deepest dwelling of the three species) into shallower depths occupied by Brown Rockfish and Copper Rockfish (Eby & Crowder, 2002). Low oxygen conditions occur sporadically in Puget Sound, most frequently in the Hood Canal basin, and have been documented as far back as pre-1900s (Brandenberger et al., 2011). Thus, environmental conditions and higher abundance in Quillback Rockfish may contribute to directional hybridization in these species.

Brown Rockfish and Quillback Rockfish displayed consistently little variation in hybrid ancestry rates amongst Puget Sound basins as compared to Copper Rockfish, but groups with higher proportions of pure individuals tended to be north of Admiralty Inlet or along the coast, reinforcing the notion that hybridization is primarily concentrated within the more isolated sub-basins of Puget Sound. Schwenke et al. (2018) used underwater sills as a distance metric for investigating hybridization rates with increasing distance from the coast; our results generally align with this proposed hypothesis, though we did not formally evaluate this correlation in our study and, rather, found nuances in frequency and level of hybrid ancestry in our analyses. In Copper Rockfish, ~55% of the SJI group and 21% of the WC group contained >6% hybrid ancestry, while HC and SPS contained just ~13% and ~42% individuals with >6% hybrid ancestry, demonstrating that the frequency of hybrids is not necessarily directly tied to distance

from the coast. Average hybrid ancestry percentage was also not especially elevated in HC for any species and only slightly more elevated in SPS than other basins for Copper Rockfish, both of which have sills at the entrance of their basins (Moore et al., 2008) and would qualify as the most distant from the coast using a "sills as distance" approach. Another notable deviation is the Whidbey Island basin (WI), which showed the highest rate of hybrid ancestry in Copper Rockfish (100% of individuals exceeded 10% hybrid ancestry, with an average of 13.3%) despite lacking a sill at its southern entrance. This suggests that basin-specific factors aside from sills such as water circulation, salinity, depth, and stratification in narrow sub-basins like Whidbey Island basin may play an important role in retaining introgressed ancestry. Hood Canal and Whidbey Island basins experience stronger stratification and salinity variability than other sub-basins of Puget Sound (MacCready et al., 2021; Moore et al., 2008), which may promote the retention of hybrid individuals at higher frequencies than other areas. Importantly, we did not find any new instances of putative F1 hybrid individuals in our PCA and STRUCTURE analyses beyond the single individual previously identified in Wray et al. (2024), even with significantly larger sample sizes and sampling in regions suspected to contain higher rates of hybridization. This strongly supports the conclusions drawn in previous studies that hybridization in these species is not recent (Wray et al., 2024; Schwenke et al., 2018; Buonaccorsi et al., 2002, 2005; Seeb, 1998).

Widespread introgression of Quillback Rockfish into Brown and Copper Rockfish in Puget Sound could also carry adaptive significance. *Sebastes* originated during the late Miocene (~6-10 million years ago) and experienced a rapid radiation event (Hyde and Vetter, 2007), resulting in over 100 closely related taxa. Rapid radiations can benefit from period of isolation and introgression given the novel genetic combinations garnered from hybridization events and

has been documented in East African cichlids (Meier et al., 2017), *Heliconius* butterflies (Dasmahapatra et al., 2012), and Darwin's finches (Lamichhaney et al., 2015). More research involving whole-genome sequencing (to best capture potentially adaptive SNPs; Lowry et al., 2017) and environmental data is needed to investigate adaptive introgression in Puget Sound rockfishes.

2.5.2 Intraspecific structure influenced by hybridization and barriers to dispersal

Intraspecific analyses demonstrated distinct patterns of population structure amongst the three species across our sampling locations. Brown Rockfish WC remained a separate cluster with no ancestry from Salish Sea individuals (Figure 2.9); Brown Rockfish have a reported range from the Gulf of Alaska down the coast to Baja California but are infrequent on the coast of Washington and Oregon (Love et al., 2002), so high differentiation between these WC individuals and the rest of the dataset was expected and previously observed in Wray et al. (2024). One individual in the north Puget Sound region had full WC cluster ancestry in Wray et al. (2024), though this sample was removed during quality filtering in our study (>25% missing data after genotype calling), therefore we were unable to confirm if this individual displayed the same ancestry makeup in our STRUCTURE runs. Though the Puget Sound sampling groups were represented predominantly by a single ancestry cluster, bidirectional mixing in Brown Rockfish was evident in SJI and GB individuals – one SJI individual belonged fully to the Puget Sound cluster while two others showed minor Puget Sound ancestry, and some Puget Sound individuals contained the third ancestry cluster associated with SJI and GB (Figure 2.3). Additionally, the elevated F_{IS} value and lower heterozygosity in SJI were consistent with Wahlund effect as proposed by Wray et al. (2024); alongside an intermediate position between

WC and Puget Sound PCA clusters, this supports the idea that north Puget Sound may function as a mixing zone for Brown Rockfish. Smaller population size in this basin may limit mate choice and encourage mating between hybrid and non-hybrid individuals (Willis et al., 2011). As in Wray et al. (2024), we were unable to examine the effects of hybridization on intraspecific structure due to our small sample size of pure individuals in Puget Sound, even after a substantial increase in sampling of HC and SPS. Future studies also targeting the collection of Brown Rockfish in SJI and GB to increase sample size for structure analysis could help resolve the mixing zone pattern we observed here.

Copper Rockfish analyses support the role of hybridization on intraspecific population structure. The negative relationship between interspecific ancestry percentage and ancestry from the GB cluster (Figure 2.7), which primarily consisted of pure individuals (Table 2.1), indicates that introgression contributes to differentiation in Puget Sound. Notable differentiation between WI and GB emphasizes this relationship, as these were the groups with the highest and lowest average hybrid ancestry, respectively (Table 2.1). However, structure is not fully explained by hybridization alone; in our full-dataset PCA and STRUCTURE analyses, HC and SPS separated distinctly from other clusters (Figure 2.4, 2.10), and the strengthened relationship between GB cluster ancestry and hybrid ancestry when these basins are excluded indicates that other factors are influencing their differentiation.

Genetic differentiation in Hood Canal specifically has been observed in Yelloweye Rockfish (*Sebastes ruberrimus*) and was hypothesized to be influenced by species life history characteristics and water circulation dynamics, which may limit larval dispersal and reduce gene flow between Hood Canal and the main basin of Puget Sound (Andrews et al., 2018). However, a subsequent oceanographic model indicated that larval dispersal patterns may not fully explain the

differentiation in Yelloweye Rockfish (Andrews et al., 2021), suggesting that non-oceanographic factors may contribute to the distinctiveness of the Hood Canal population. Limited dispersal in Copper Rockfish has also previously been proposed as the rationale for differentiation observed in the fjord-like inlets on the west coast of Vancouver Island (Dick et al., 2014), though this study did not evaluate interspecific hybridization and suggested that limited physical movement due to oceanographic and topographical barriers may be drivers of their structure. Copper Rockfish are closely associated with rocky reef habitat as adults and exhibit limited adult movement (Love et al., 2002), as well as limited larval drift due to shallow sills in Puget Sound (Buonaccorsi et al., 2002), making them particularly susceptible to isolation in physically restricted systems. The patterns in Copper Rockfish indicate that while limited dispersal and interspecific hybridization are the dominant drivers of population structure patterns, differentiation at the sub-basin level may reflect additional factors, such as independent glacial refugia, post-glacial colonization history (Buonaccorsi et al., 2002), and the distinct environmental conditions in these basins.

Quillback Rockfish population structure was the most clearly defined of the three species, with two strongly supported genetic clusters; one cluster consisted of WC, SJI, and GB, and the second comprising all Puget Sound sampling groups south of Admiralty Inlet (MB, WI, HC, SPS). The separation of the two populations at Admiralty Inlet is consistent with limited dispersal due to the double sill acting as a barrier to larval dispersal and estuarine exchange (MacCready et al., 2021; Moore et al., 2008) and aligns with the structure identified in Wray et al. (2024). The WC sampling group was an exception to this cluster separation, with individuals approximately evenly distributed between the two clusters (Figure 2.11). Individuals sampled further west along the coast of Vancouver Island tended to cluster with the northern SJI/GB

group, while those further east in the Strait of Juan de Fuca showed ancestry consistent with the populations south of Admiralty Inlet. This was also visible in our PCA and is indicative of the gene flow suggested in Wray et al. (2024), wherein individuals may disperse out of southern Puget Sound and mix in north Puget Sound while inward gene flow from the coast remains limited. This directional but restricted migration out of Admiralty Inlet for both Copper Rockfish and Quillback Rockfish has also been documented in Yelloweye Rockfish (Andrews et al., 2018), suggesting that Admiralty Inlet functions as a dispersal barrier across multiple rockfish species.

The species-specific patterns of intraspecific structure identified here, combined with the influence of interspecific hybridization, should be considered in future management of these species. Currently, commercial and recreational rockfish fisheries in Puget Sound remain closed due to lack of recovery; identifying population structure now allows for the crafting of specific plans that may facilitate improved fisheries management. As discussed in Wray et al. (2024), current management assumes three distinct population segments (DPSs) for these three species: 1) coastal, which encompasses Alaska to California, 2) north Puget Sound, and 3) south Puget Sound, which is south of Admiralty Inlet (Stout et al., 2001). The patterns in our F_{ST} analyses support the foundational hypothesis that population structure in these three species is indeed primarily due to separation between Puget Sound proper and the rest of the Salish Sea out to the coast, but our individual-based analyses using significantly more loci than previous work illuminates more complex structure influenced by hybridization in some species and basins. Our results reaffirm that the established DPSs may be inaccurate for Brown Rockfish, Copper Rockfish, and Quillback Rockfish given the distinct structure each species has displayed, especially with increased sampling in previously data-limited regions. For all three species,

Admiralty Inlet remains the most meaningful boundary, with consistent evidence of limited dispersal and distinct populations on either side. These results are also consistent with Wray et al. (2024) in that there is limited evidence for dispersal from the coast into Puget Sound in these three species, pointing to relative isolation from the coast and reducing the chance of gene flow that would potentially benefit recovery in Puget Sound populations. However, this finding contrasts with other rockfish species, such as Canary (Andrews et al., 2018) and Black Rockfish (Wray et al., 2025), that show little to no population structure across this boundary, and three other rockfish species that show different population distributions (Wray et al., 2025). The species-specific structure we have confirmed in this study, and the sub-basin differentiation observed in Copper Rockfish in particular, suggest that DPS boundaries may require finer-scale delineations to properly manage these hybridizing species in the future. Overall, this study reinforces the need for species-specific management strategies that treat Puget Sound rockfish species as independent units that may require localized recovery attention.

2.6 FIGURES

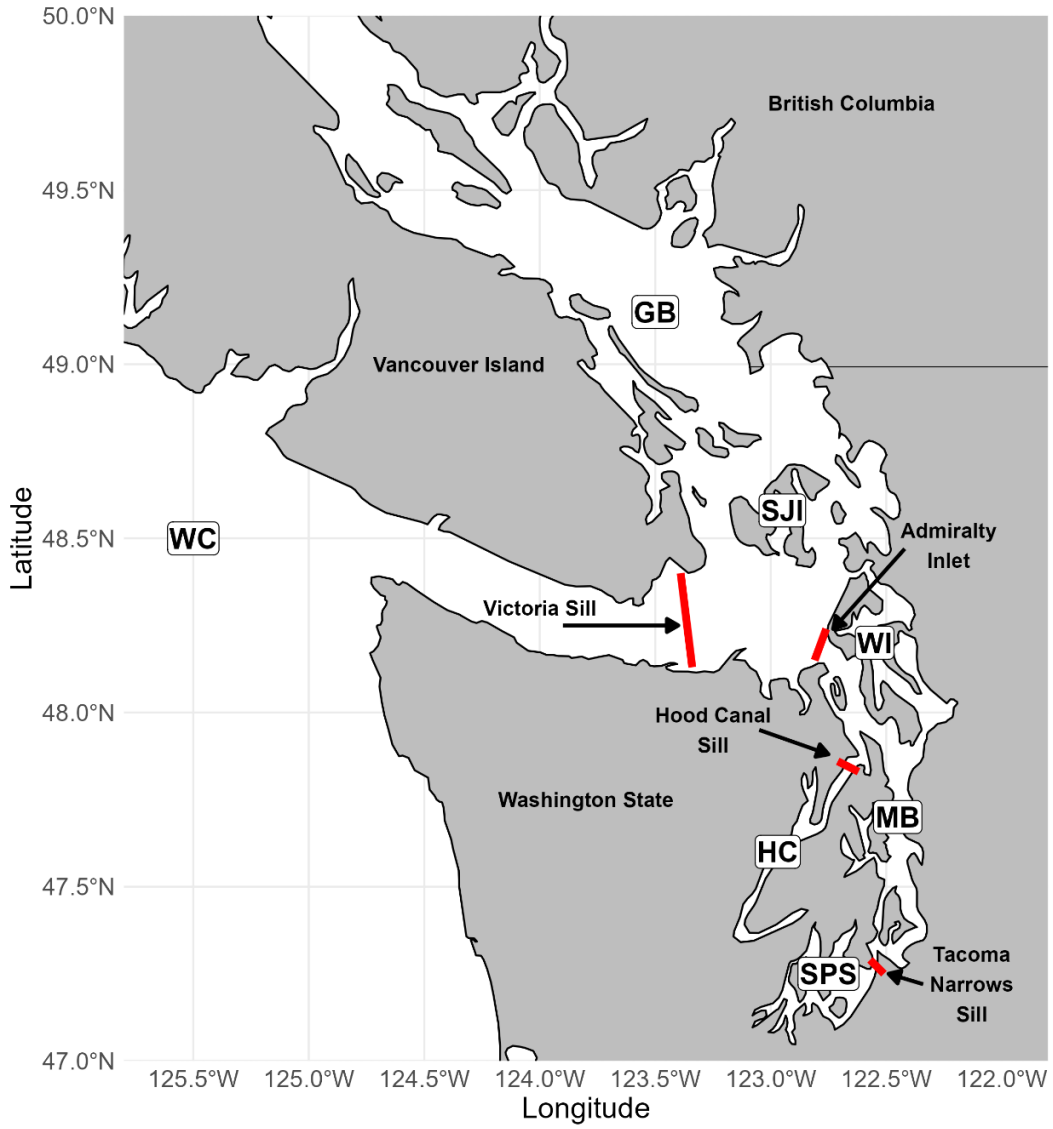


Figure 2.1 Map of rockfish sampling regions in the Salish Sea. Samples were assigned to one of the seven regions according to their sampling coordinates. From the coast, moving inshore: WC = West Coast, GB = Georgia Basin, SJI = San Juan Islands, WI = Whidbey Island, MB = Main Basin, HC = Hood Canal, and SPS = South Puget Sound (south of Tacoma Narrows). Not pictured is Southern California and Mexico, from which 6 Brown Rockfish samples were collected (counted under WC group). See Supplemental for a map of exact sampling locations.

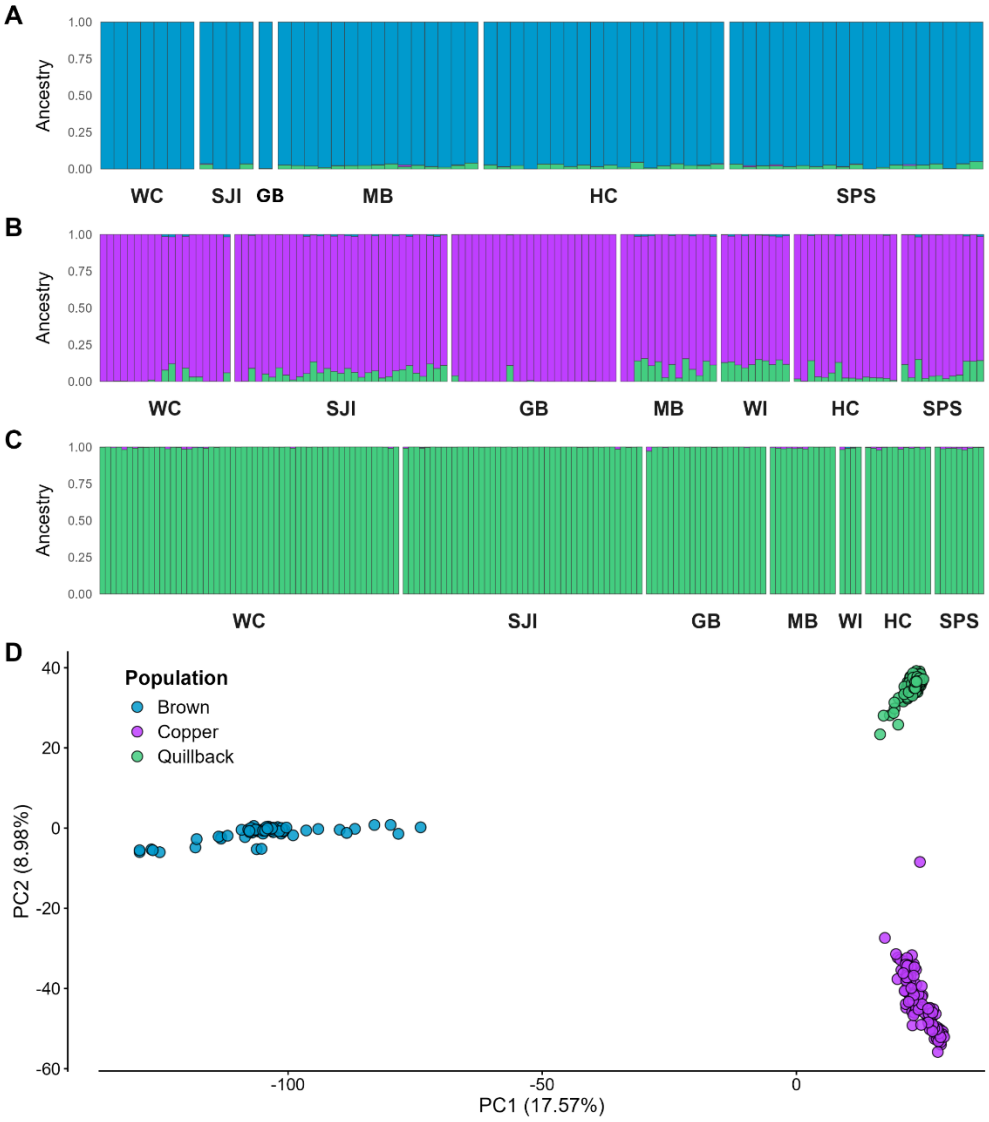


Figure 2.2 STRUCTURE (A-C) and PCA (D) results for interspecific rockfish analysis. A-C: STRUCTURE plots for A) Brown Rockfish (blue ancestry), B) Copper Rockfish (purple ancestry), and C) Quillback Rockfish (green ancestry). Each bar is an individual and is colored by its ancestry cluster(s). Within each group, individuals are ordered by ID; individuals from Wray et al. (2024) are at the beginning, followed by individuals from this study (proportion of individuals from each varies by group). D: Principal component analysis (PCA). Each point is an individual, and color corresponds to in-field species identification. This dataset consisted of 8,685 SNPs.

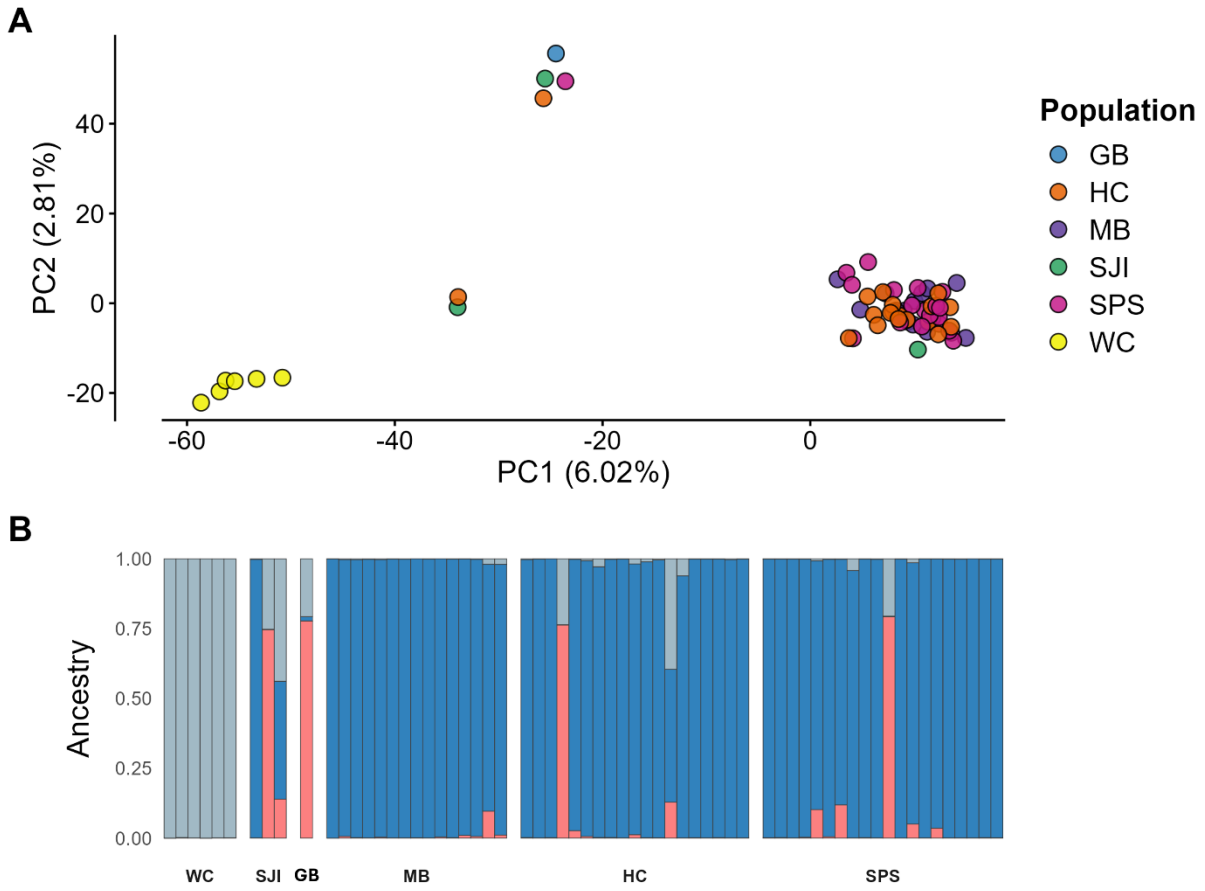


Figure 2.3 Principal component analysis (A) and STRUCTURE (B) analysis for Brown Rockfish. A: PCA; each point is an individual, and color corresponds to sampling location. B: STRUCTURE plot at K = 3 (highest supported using ΔK). Each bar is an individual and is colored by its ancestry cluster(s). Groups are ordered by sampling location, furthest from the coast and moving inshore. WC = West Coast, SJI = San Juan Islands, GB = Georgia Basin, MB = Main Basin, HC = Hood Canal, and SPS = South Puget Sound. Within each group, individuals are ordered by ID; individuals from Wray et al. (2024) are at the beginning, followed by individuals from this study (proportion of individuals from each varies by group). This dataset consisted of 3,817 SNPs and 64 individuals.

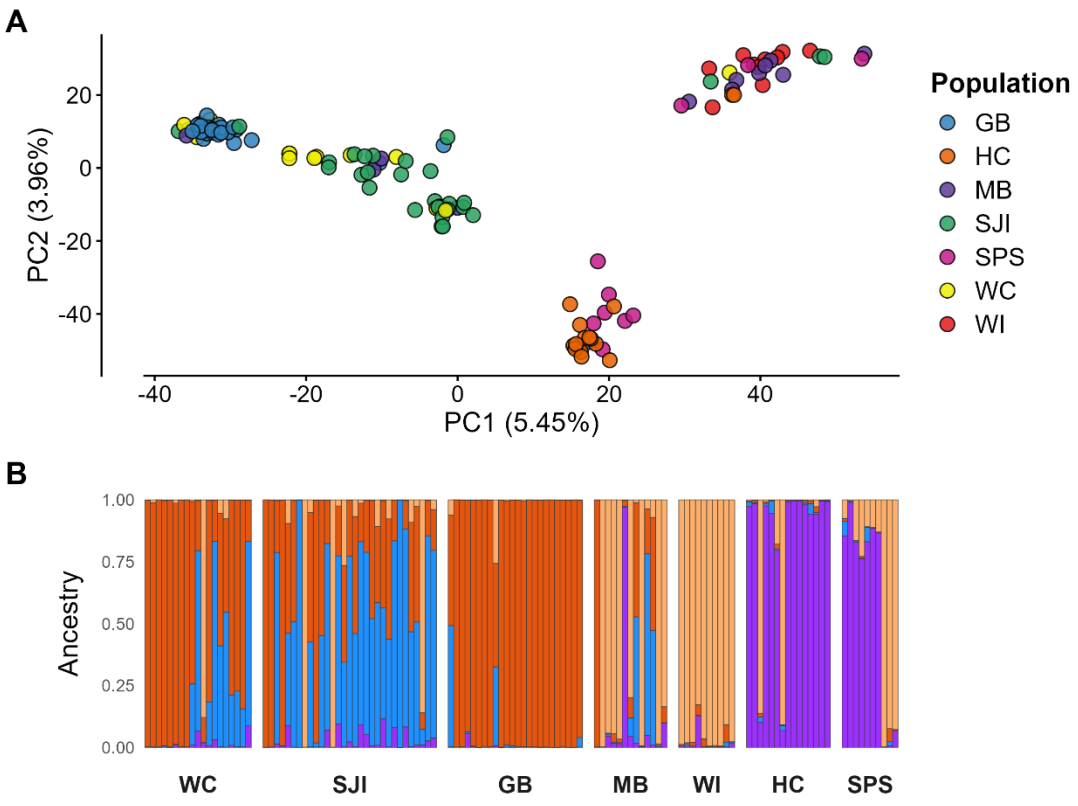


Figure 2.4 Principal component analysis (A) and STRUCTURE (B) analysis for Copper Rockfish. A: PCA; each point is an individual, and color corresponds to sampling location. B: STRUCTURE plot at K = 4 (highest supported using ΔK was K = 5). Each bar is an individual and is colored by its ancestry cluster(s). Groups are ordered by sampling location, furthest from the coast and moving inshore. WC = West Coast, SJI = San Juan Islands, GB = Georgia Basin, MB = Main Basin, WI = Whidbey Island, HC = Hood Canal, and SPS = South Puget Sound. Within each group, individuals are ordered by ID; individuals from Wray et al. (2024) are at the beginning, followed by individuals from this study (proportion of individuals from each varies by group). This dataset consisted of 7,092 SNPs.

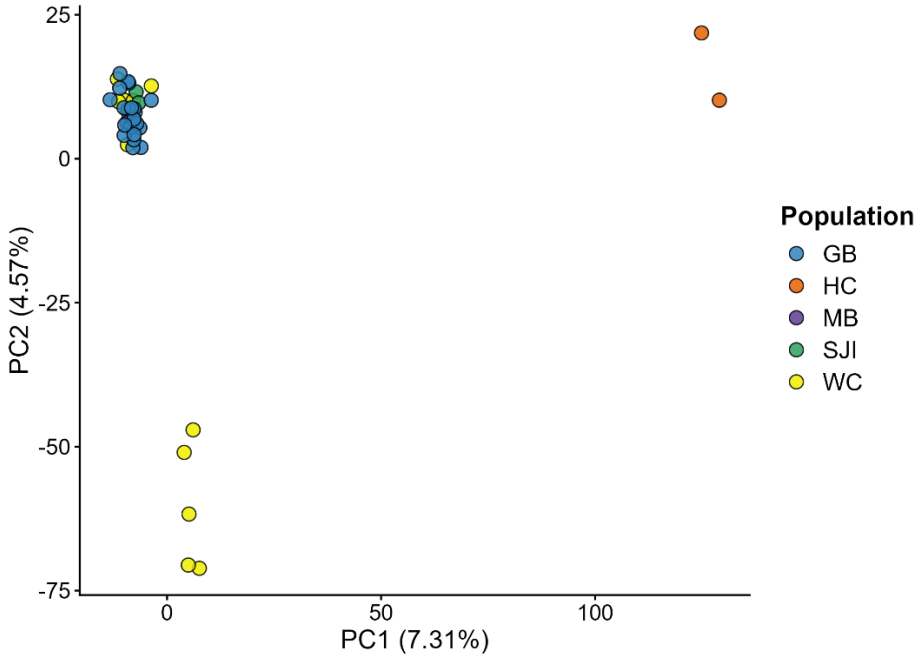


Figure 2.5 Copper Rockfish PCA featuring only individuals with <1% hybrid ancestry.

Each point is an individual, and individuals are colored by sampling location.

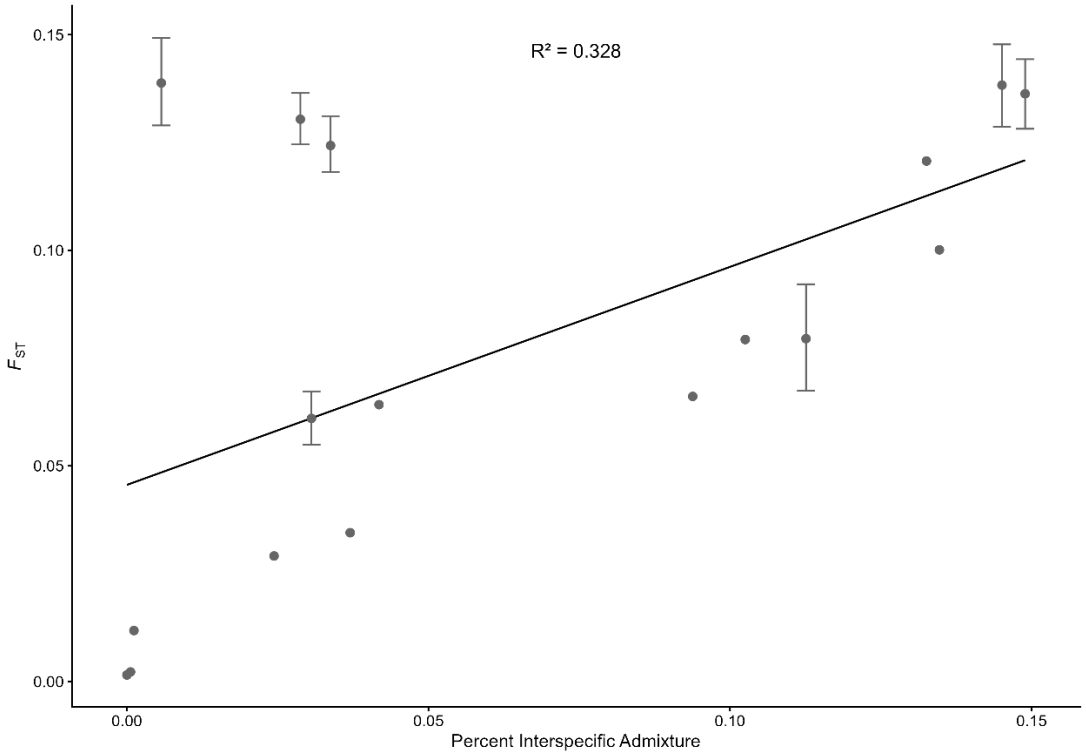


Figure 2.6 Copper Rockfish interspecific admixture (%) versus Weir and Cockerham pairwise F_{ST}. Individuals were grouped by their respective sampling location and then pooled within each group into three categories: pure (<1%), low (1-6%), and high (>6%) interspecific admixture. Because some sampling locations did not have individuals belonging to one or more of the categories (and Brown Rockfish had no WI samples), this resulted in 17 groups. Pairwise F_{ST} values were then calculated between these groups. $R^2 = 0.328$, p-value = 0.01635.

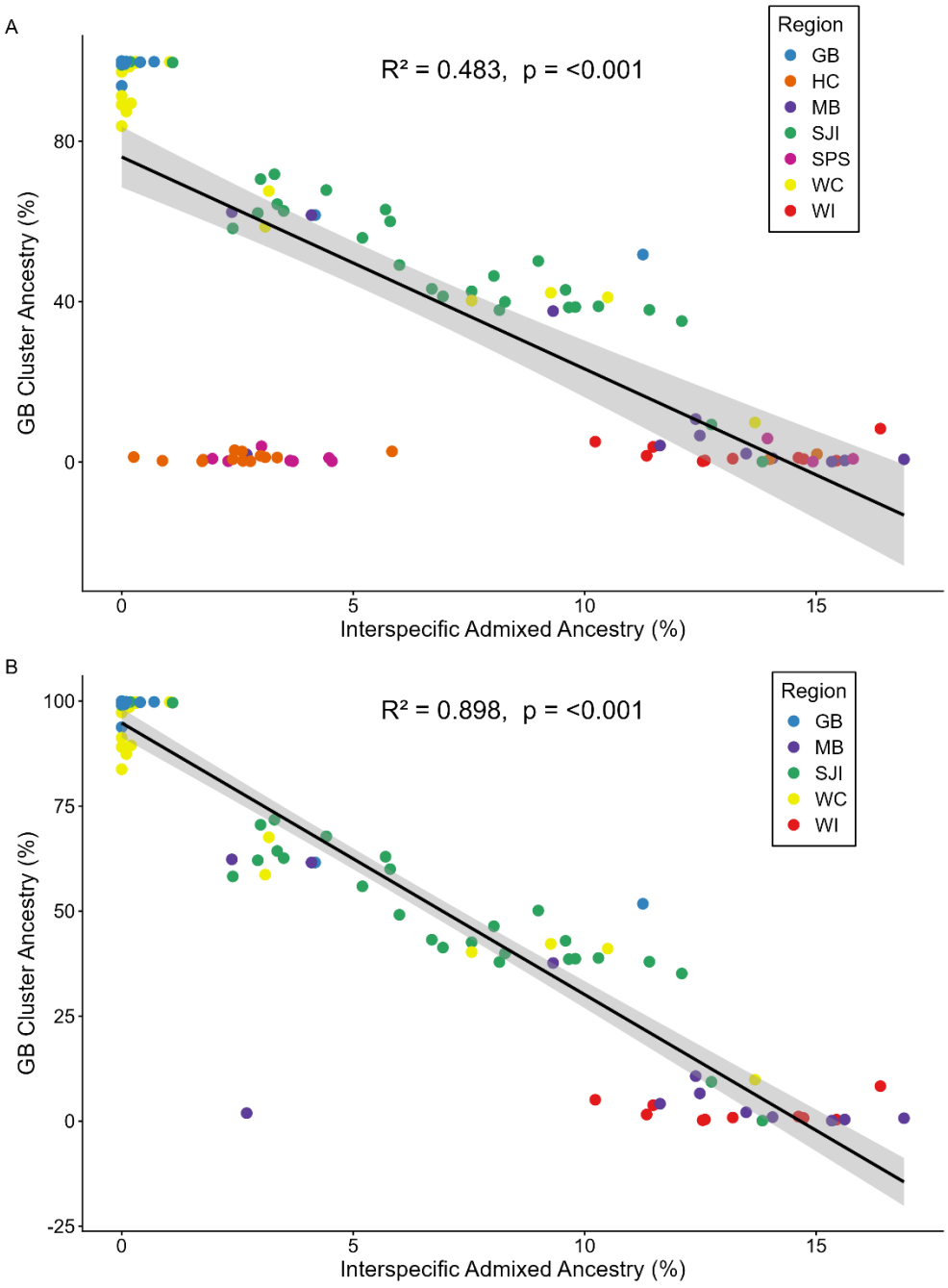


Figure 2.7 Copper Rockfish interspecific hybrid ancestry (%) versus intraspecific ancestry from GB cluster (%). GB = Georgia Basin cluster; each point is an individual, and individuals are colored by sampling location. A) All individuals in the dataset (R^2 = 0.483), and B) HC and SPS removed (R^2 = 0.898). Data on each axis comes from different SNP datasets with only 18.8% overlap.

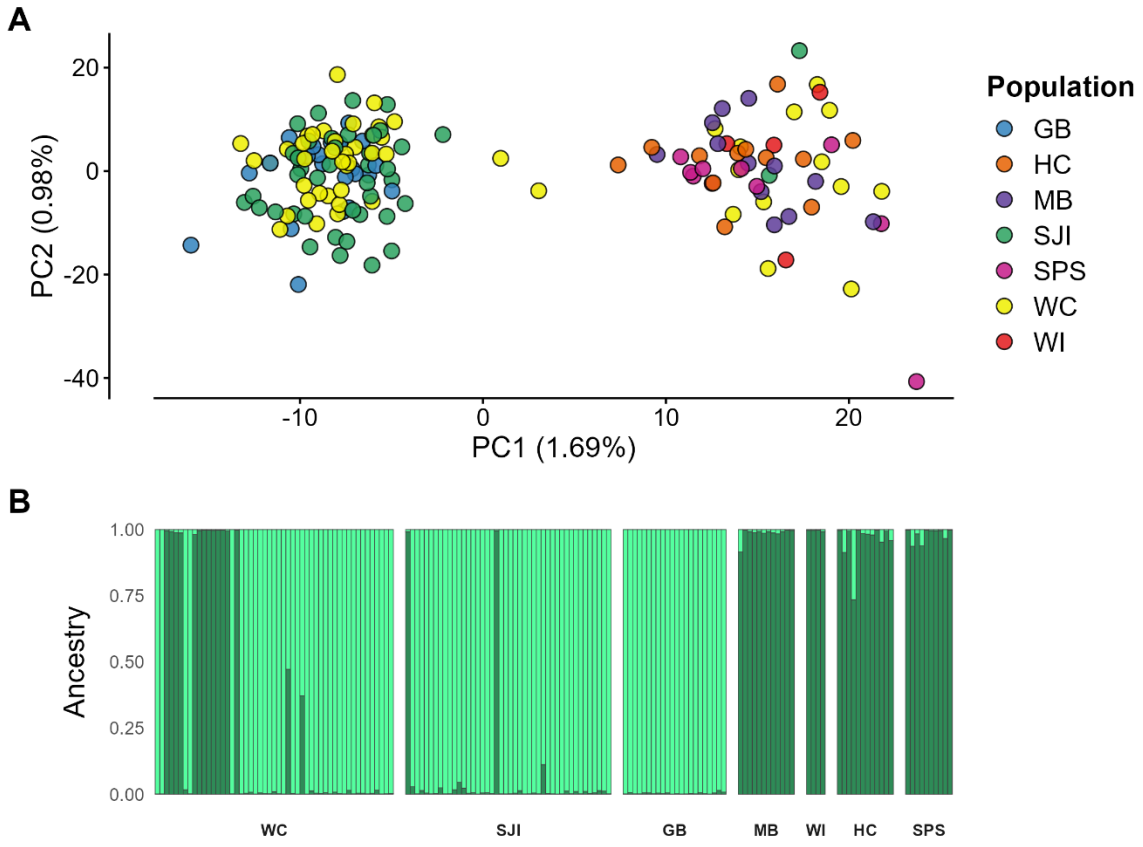


Figure 2.8 Principal component analysis (A) and STRUCTURE (B) analysis for Quillback Rockfish. A: PCA; each point is an individual, and color corresponds to sampling location. B: STRUCTURE plot at K = 2 (highest supported using ΔK). Each bar is an individual and is colored by its ancestry cluster(s). Groups are ordered by sampling region, furthest from the coast and moving inshore. WC = West Coast, SJI = San Juan Islands, GB = Georgia Basin, MB = Main Basin, WI = Whidbey Island, HC = Hood Canal, and SPS = South Puget Sound. Within each group, individuals are ordered by ID; individuals from Wray et al. (2024) are at the beginning, followed by individuals from this study (proportion of individuals from each varies by group). This dataset consisted of 3,941 SNPs.

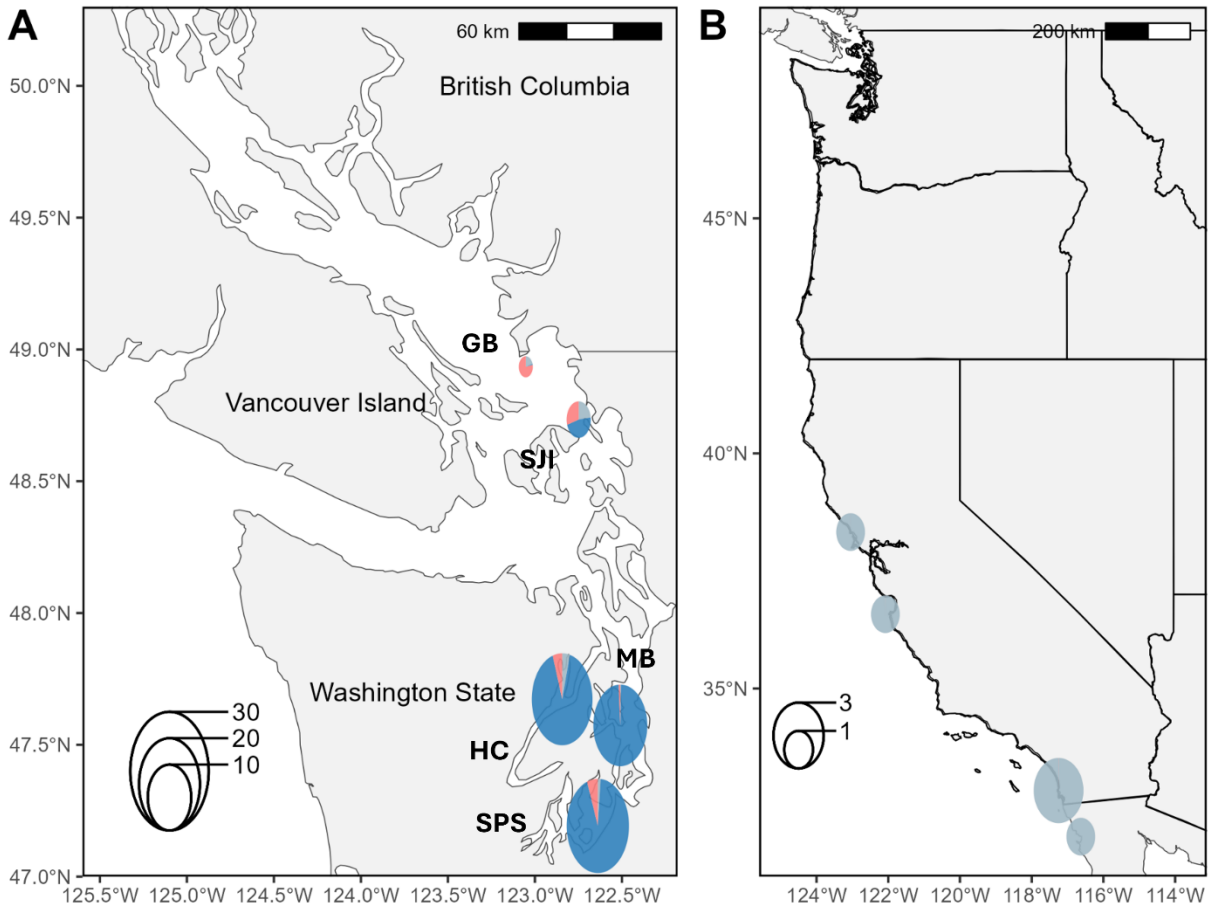


Figure 2.9 Brown Rockfish K = 3 STRUCTURE analysis distribution in A) Puget Sound and B) West Coast (WC). Pie chart colors correspond to STRUCTURE ancestry cluster makeup. Individuals were pooled according to sampling location and size of pie chart indicates sample size.

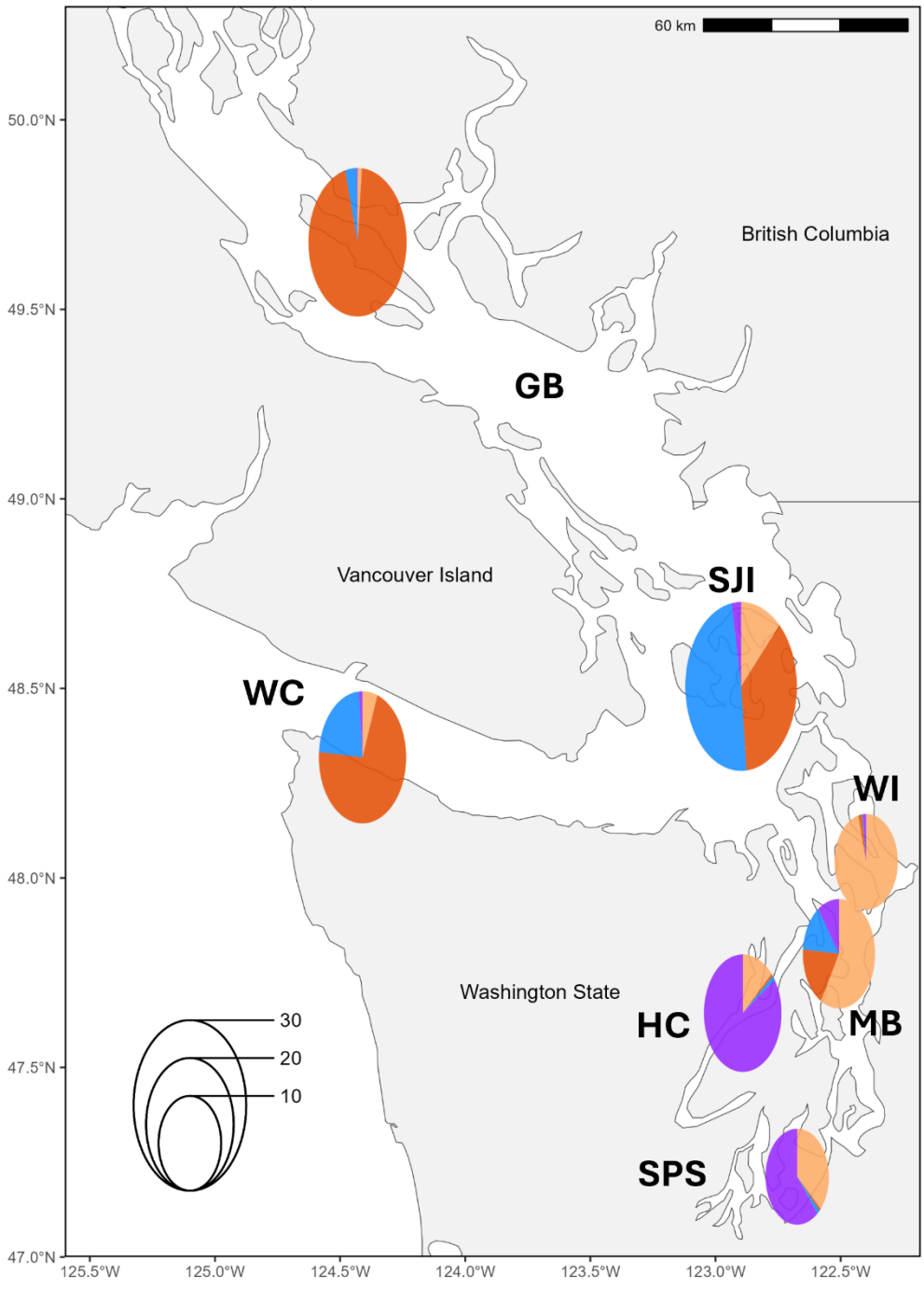


Figure 2.10 Copper Rockfish STRUCTURE analysis distribution in Puget Sound at K = 4. Pie chart colors correspond to STRUCTURE ancestry cluster makeup. Individuals were pooled according to sampling region and size of pie chart indicates sample size.

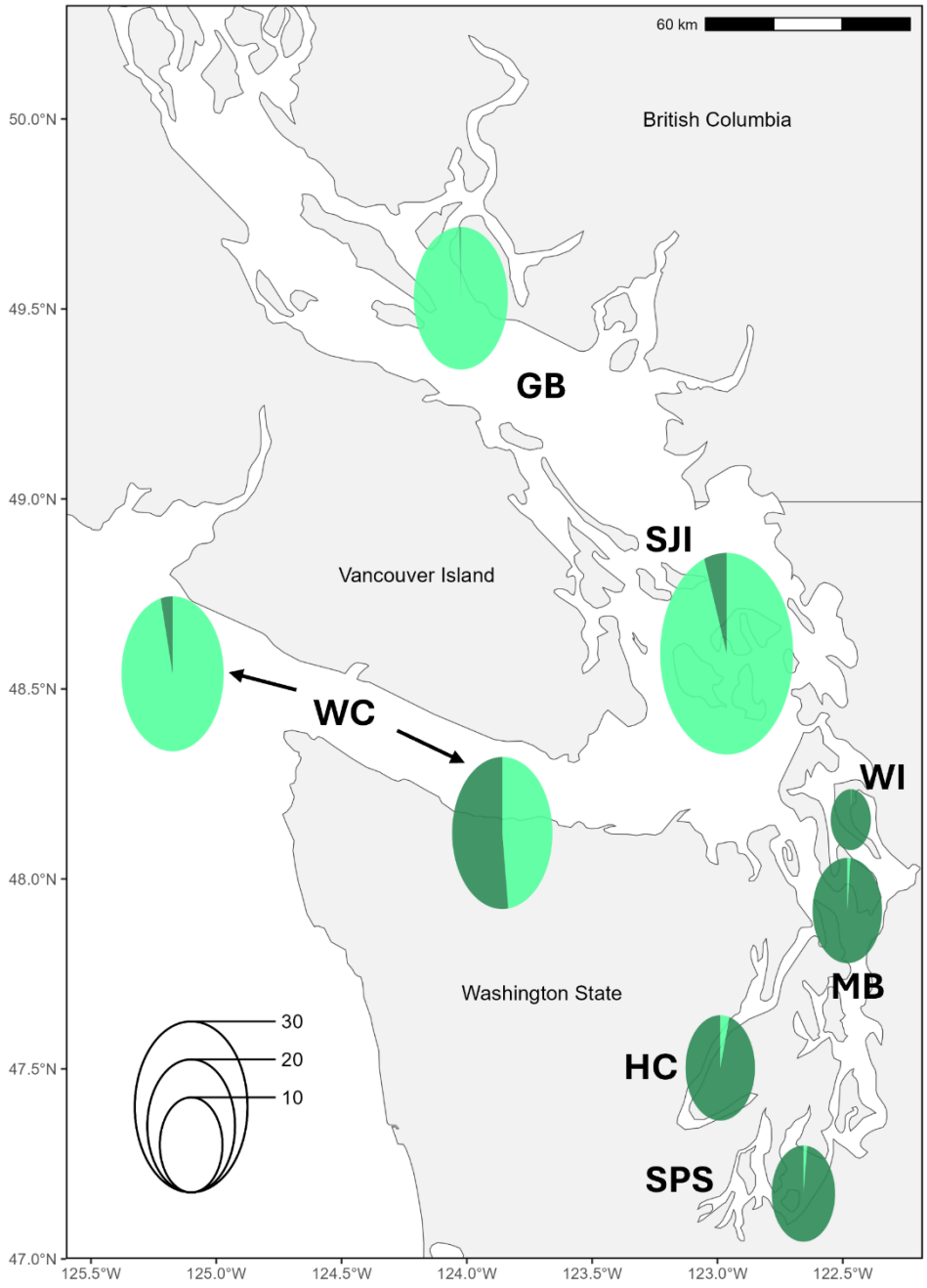


Figure 2.11 Quillback Rockfish K = 2 STRUCTURE analysis distribution in Puget Sound. Pie chart colors correspond to STRUCTURE ancestry cluster makeup. Individuals were pooled according to sampling location and size of pie chart indicates sample size.

2.7 TABLES

Table 2.1 Interspecific rockfish average hybrid ancestry per sampling location and species. N = number of individuals per sampling location for that species, and the average percentage of hybrid ancestry (non-species ancestry from STRUCTURE analysis) to the right of N.

Interspecific Rockfish Average Hybrid Ancestry			
	(N) **Brown**	(N) **Copper**	(N) **Quillback**
WC	(7) **0.00%**	(19) **2.62%**	(55) **0.26%**
SJI	(4) **1.78%**	(31) **6.42%**	(44) **0.13%**
GB	(1) **0.40%**	(24) **0.70%**	(22) **0.20%**
MB	(18) **2.48%**	(14) **9.32%**	(12) **0.58%**
WI	N/A	(10) **13.3%**	(4) **0.95%**
HC	(19) **2.41%**	(15) **4.12%**	(12) **0.70%**
SPS	(20) **2.46%**	(12) **8.08%**	(9) **0.87%**
Overall	(64) **2.11%**	(125) **5.50%**	(158) **0.33%**

Table 2.2 Summary statistics for Brown, Copper, and Quillback Rockfish. *N* = individuals per sampling region, *Ho* = mean observed heterozygosity, *He* = mean expected heterozygosity, F_{IS} = inbreeding coefficient. *Ho*, *He,* and F_{IS} values were calculated using the R package hierfstat v0.5-11 (Goudet, 2005).

	Brown Rockfish				**Copper Rockfish**				**Quillback Rockfish**			
	N	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}	*N*	*Ho*	*He*	F_{IS}
WC	6	0.132	0.139	0.034	19	0.218	0.237	0.063	51	0.267	0.278	0.036
GB	1	0.198	N/A	N/A	24	0.209	0.218	0.027	22	0.271	0.278	0.018
SJI	3	0.204	0.246	0.089	30	0.241	0.255	0.042	44	0.277	0.282	0.017
MB	15	0.251	0.251	-0.003	13	0.242	0.265	0.064	12	0.278	0.284	0.012
WI	N/A	N/A	N/A	N/A	10	0.250	0.255	0.013	4	0.299	0.283	-0.070
HC	19	0.237	0.248	0.033	15	0.228	0.239	0.038	12	0.277	0.281	0.011
SPS	20	0.241	0.246	0.012	10	0.236	0.254	0.050	10	0.278	0.283	0.007
Overall	64	0.211	0.225	0.062	121	0.232	0.246	0.057	155	0.278	0.282	0.013

Table 2.3 Brown Rockfish pairwise Weir and Cockerham F_{ST} estimates. Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Brown Rockfish Pairwise F_{ST}					
	SJI	GB	MB	WC	SPS
GB	-0.040				
MB	**0.034**	**0.043**			
WC	**0.091**	**0.230**	**0.179**		
SPS	**0.034**	**0.033**	0.001	**0.173**	
HC	**0.021**	**0.024**	0.002	**0.161**	-0.001

Table 2.4 Copper Rockfish pairwise Weir and Cockerham F_{ST} estimates. Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Copper Rockfish Pairwise F_{ST}						
	WC	SJI	GB	WI	MB	SPS
SJI	**0.011**					
GB	**0.011**	**0.032**				
WI	**0.083**	**0.065**	**0.114**			
MB	**0.033**	**0.020**	**0.060**	**0.012**		
SPS	**0.060**	**0.041**	**0.095**	**0.055**	**0.013**	
HC	**0.078**	**0.059**	**0.108**	**0.093**	**0.047**	**0.005**

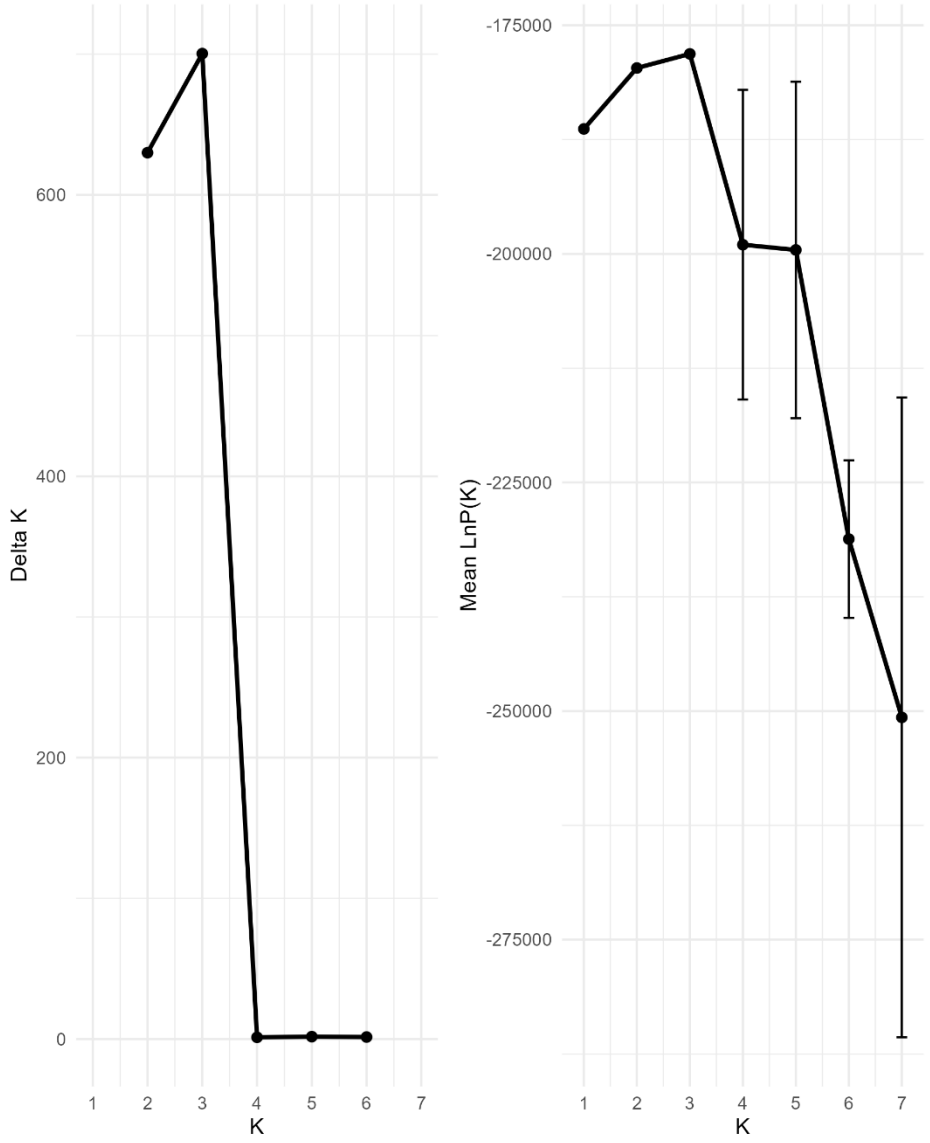
Table 2.5 Copper Rockfish pairwise Weir and Cockerham F_{ST} estimates with only pure individuals (<1% hybrid ancestry). Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Copper Rockfish Pure Individuals (<1% Hybrid Ancestry) Pairwise F_{ST}				
	WC	**SJI**	**MB**	**GB**
SJI	**0.008**			
MB	-0.005	0.012		
GB	**0.012**	0.002	0.002	
HC	**0.116**	**0.132**	**0.117**	**0.139**

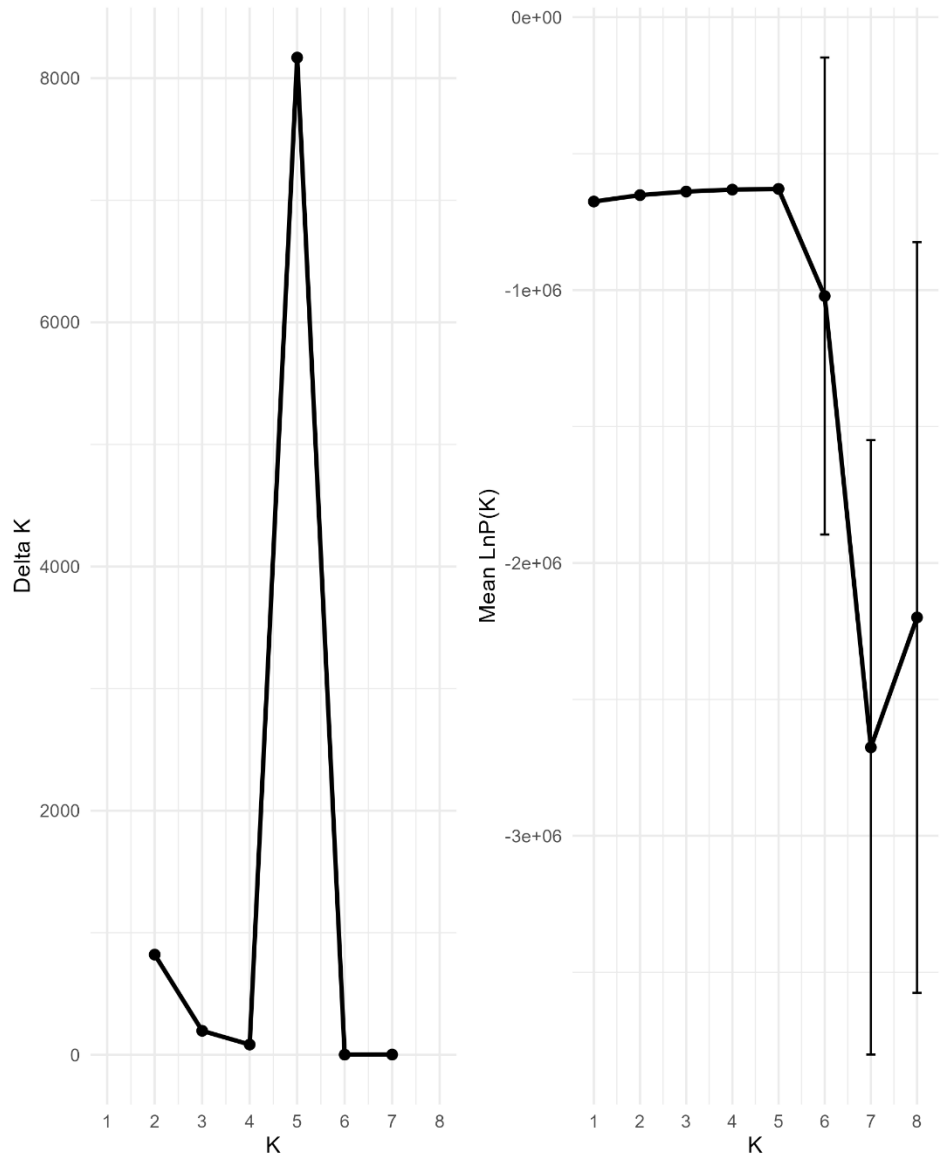
Table 2.6 Quillback Rockfish pairwise Weir and Cockerham F_{ST} estimates. Numbers that are significantly greater than zero are bolded (based on the lower 95% CI limit).

Quillback Rockfish Pairwise F_{ST}						
	WC	SJI	GB	SPS	WI	MB
SJI	**0.001**					
GB	**0.002**	0.001				
SPS	**0.005**	**0.010**	**0.014**			
WI	**0.006**	**0.010**	**0.015**	0.001		
MB	**0.007**	**0.010**	**0.013**	0.000	-0.005	
HC	**0.006**	**0.011**	**0.015**	0.001	0.001	**0.003**

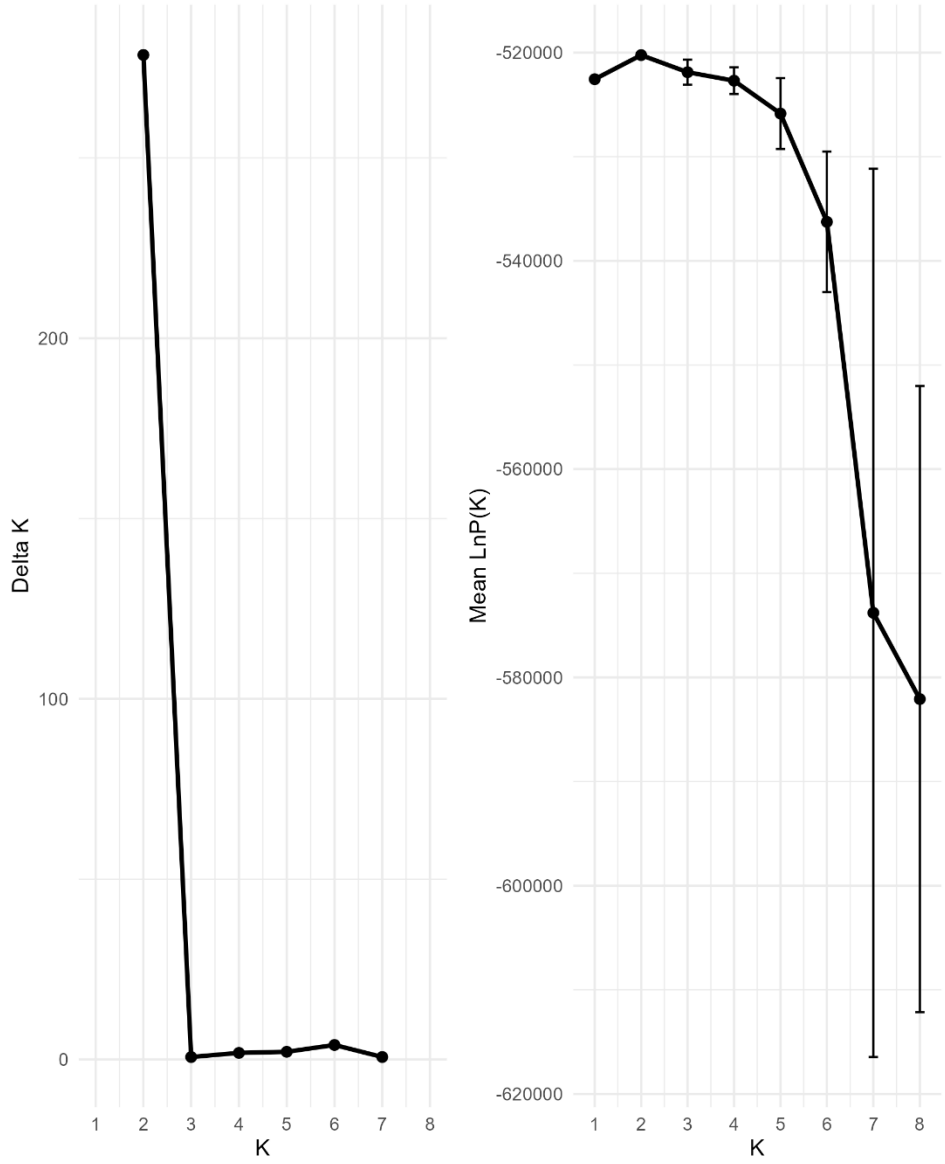
2.8 SUPPLEMENTAL FIGURES



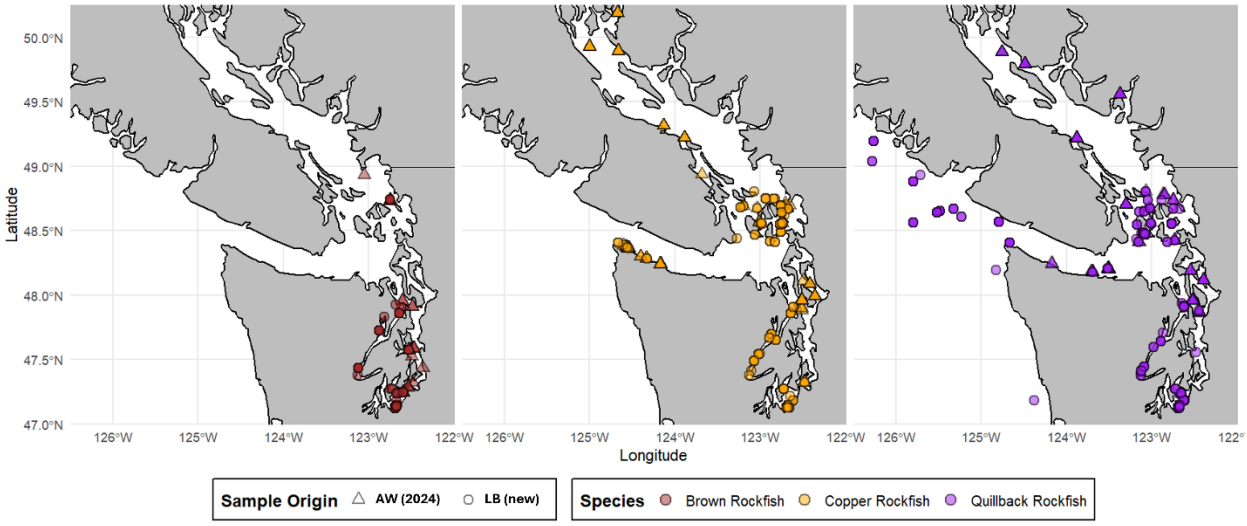
Supplemental Figure 2.1 Delta K (left) and Mean LnP(K) (right, ±1 SD) for Brown Rockfish STRUCTURE analysis.



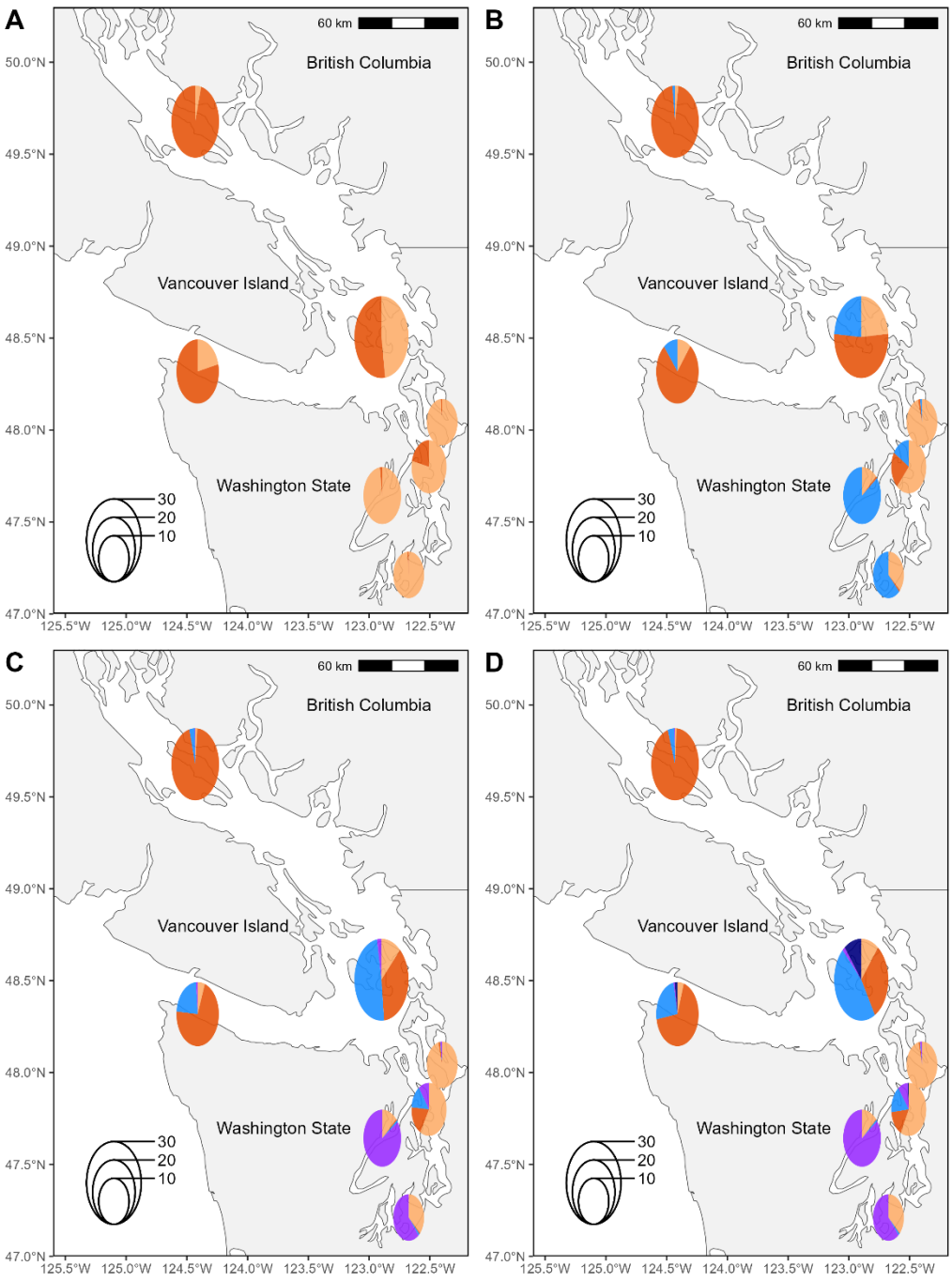
Supplemental Figure 2.2 Delta K (left) and Mean LnP(K) (right, ±1 SD) for Copper Rockfish STRUCTURE analysis.



Supplemental Figure 2.3 Delta K (left) and Mean LnP(K) (right, ±1 SD) for Quillback Rockfish STRUCTURE analysis.



Supplemental Figure 2.4 Distribution of samples per species. Left = Brown Rockfish, middle = Copper Rockfish, right = Quillback Rockfish. Each point is an individual, and points are semi-transparent; individuals that are close in coordinates overlap, therefore opaque points indicate several individuals. Shape of the points indicates if sequenced in Wray et al. (2024) (triangle) or sequenced here, denoted as LB (circle). Six Brown Rockfish samples were sampled from California and Mexico and are not pictured; location can be seen in Figure 2.9.



Supplemental Figure 2.5 Copper Rockfish STRUCTURE analysis distribution in Puget Sound at A) K = 2, B) K = 3, C) K = 4, and D) K = 5. Pie chart colors correspond to STRUCTURE ancestry cluster makeup. Individuals were pooled according to sampling region and size of pie chart indicates sample size. Highest support ΔK value was K = 5.

2.9 SUPPLEMENTAL TABLES

Supplemental Table 2.1 Genotype discordance rates between interspecific dataset replicates. AW replicates comprise 28 individuals from Wray et al., (2024) that were re-sequenced, and technical replicates comprise 70 individuals (11 Brown Rockfish, 23 Copper Rockfish, and 36 Quillback Rockfish) sequenced in this study twice; 26 individuals were prepared in two different libraries from the same DNA extract, and 44 individuals were extracted in two separate extraction plates. Genotype discordance between replicate samples was calculated using the script "genoerrorcalc" (https://github.com/bsjodin/genoerrorcalc).

Genotype Discordance Rates (Interspecific Dataset)			
	Average	Highest	Lowest
Technical	0.92%	12.90%	0.00%
AW vs Re-sequence	4.54%	33.50%	0.11%
Overall	2.77%	33.50%	0.00%

2.10 REFERENCES

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. Genetics, 202(2), 389–400. https://doi.org/10.1534/genetics.115.183665

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. Trends in Ecology & Evolution, 16(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Andrews, K., Bartos, B., Harvey, C. J., Tonnes, D., Bhuthimethee, M., & MacCready, P. (2021). Testing the potential for larval dispersal to explain connectivity and population structure of threatened rockfish species in Puget Sound. *Marine Ecology Progress Series*, *677*, 95 113. https://doi.org/10.3354/meps13858

Andrews, K. S., Nichols, K. M., Elz, A., Tolimieri, N., Harvey, C. J., Pacunski, R., Lowry, D., Yamanaka, K. L., & Tonnes, D. M. (2018). Cooperative research sheds light on population structure and listing status of threatened and endangered rockfish species. *Conservation Genetics*, *19*(4), 865–878. https://doi.org/10.1007/s10592-018-1060-0

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data (0.12.0).

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLOS ONE*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Bargmann, G., Palsson, W., Burley, C., Friedel, D., & Tsou, T. (2011). Final Environmental Impact Statement for the Puget Sound Rockfish Conservation Plan Including Preferred Range of Actions. *Washington Department of Fish & Wildlife*.

https://wdfw.wa.gov/publications/00035.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*, *16*(9), 1195–1205. https://doi.org/10.1111/ele.12150

Brandenberger, J., Louchouarn, P., & Crecelius, E. (2011). Natural and Post-Urbanization Signatures of Hypoxia in Two Basins of Puget Sound: Historical Reconstruction of Redox Sensitive Metals and Organic Matter Inputs. Aquatic Geochemistry, 17, 645–670. https://doi.org/10.1007/s10498-011-9129-0

Brauer, C. J., Sandoval-Castillo, J., Gates, K., Hammer, M. P., Unmack, P. J., Bernatchez, L., & Beheregaray, L. B. (2023). Natural hybridization reduces vulnerability to climate change. *Nature Climate Change*, *13*(3), 282–289. https://doi.org/10.1038/s41558-022-01585-1

Burford, M. O., & Bernardi, G. (2008). Incipient speciation within a subgenus of rockfish (Sebastosomus) provides evidence of recent radiations within an ancient species flock. *Marine Biology*, *154*(4), 701–717. https://doi.org/10.1007/s00227-008-0963-6

Buonaccorsi, V., Kimbrell, C., Lynn, E., & Vetter, R. (2002). Population structure of copper rockfish (Sebastes caurinus) reflects postglacial colonization and contemporary patterns of larval dispersal. *Canadian Journal of Fisheries and Aquatic Sciences*, *59*, 1374–1384. https://doi.org/10.1139/f02-101

Buonaccorsi, V. P., Kimbrell, C. A., Lynn, E. A., & Vetter, R. D. (2005). Limited realized

dispersal and introgressive hybridization influence genetic structure and conservation strategies for brown rockfish, Sebastes auriculatus. *Conservation Genetics*, *6*(5), 697–713. https://doi.org/10.1007/s10592-005-9029-1

Burns, R. The Shape and Form of Puget Sound. (1985). Seattle, WA: Washington Sea Grant Program.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., Zimin, A. V., Hughes, D. S. T., Ferguson, L. C., Martin, S. H., Salazar, C., Lewis, J. J., Adler, S., Ahn, S.-J., Baker, D. A., Baxter, S. W., Chamberlain, N. L., Chauhan, R., Counterman, B. A., … Jiggins, C. D. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94–98. https://doi.org/10.1038/nature11041

Dethier, M. N., & Schoch, G. C. (2005). The consequences of scale: Assessing the distribution of benthic populations in a complex estuarine fjord. *Estuarine, Coastal and Shelf Science*, *62*(1), 253–270. https://doi.org/10.1016/j.ecss.2004.08.021

Dick, S., Shurin, J. B., & Taylor, E. B. (2014). Replicate divergence between and within sounds in a marine fish: The copper rockfish (Sebastes caurinus). *Molecular Ecology*, *23*(3),

575–590. https://doi.org/10.1111/mec.12630

Drake, J. S., Berntson, E. A., Gustafson, R. G., Holmes, E. E. (Elizabeth E., Levin, P. S., Tolimieri, N., Waples, R. S., Sogard, S. M., Williams, G. D. (Gregory D., & Cope, J. M. (2010). Status review of five rockfish species in Puget Sound, Washington: Bocaccio (Sebastes paucispinis), canary rockfish (S. pinniger), yelloweye rockfish (S. ruberrimus), greenstriped rockfish (S. elongatus), and redstripe rockfish (S. proriger). https://repository.library.noaa.gov/view/noaa/3777

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eby, L. A., & Crowder, L. B. (2002). Hypoxia-based habitat compression in the Neuse River Estuary: Context-dependent shifts in behavioral avoidance thresholds. Canadian Journal of Fisheries and Aquatic Sciences, 59(6), 952–965. https://doi.org/10.1139/f02-067

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047 3048. https://doi.org/10.1093/bioinformatics/btw354

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*(1), 104–120. https://doi.org/10.1111/mec.13476

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics.

Molecular Ecology Notes, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Grant, P. R., & Grant, B. R. (2010). Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philosophical Transactions: Biological Sciences*, *365*(1543), 1065–1076.

Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*, *30*(1), 33–41. https://doi.org/10.1111/cobi.12574

Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618. https://doi.org/10.1111/mec.12415

Houde, A. L. S., Fraser, D. J., O'Reilly, P., & Hutchings, J. A. (2011). Relative risks of inbreeding and outbreeding depression in the wild in endangered salmon. *Evolutionary Applications*, *4*(5), 634–647. https://doi.org/10.1111/j.1752-4571.2011.00186.x

Howard-McCombe, J., Ward, D., Kitchener, A. C., Lawson, D., Senn, H. V., & Beaumont, M. (2021). On the use of genome-wide data to model and date the time of anthropogenic hybridisation: An example from the Scottish wildcat. *Molecular Ecology*, *30*(15), 3688–3702. https://doi.org/10.1111/mec.16000

Hyde, J. R., & Vetter, R. D. (2007). The origin, evolution, and diversification of rockfishes of the genus Sebastes (Cuvier). *Molecular Phylogenetics and Evolution*, *44*(2), 790–811. https://doi.org/10.1016/j.ympev.2006.12.026

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population

structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371–375. https://doi.org/10.1038/nature14181

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources*, *18*(4), 778–788. https://doi.org/10.1111/1755-0998.12779

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

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

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

MacCready, P., McCabe, R. M., Siedlecki, S. A., Lorenz, M., Giddings, S. N., Bos, J., Albertson,

S., Banas, N. S., & Garnier, S. (2021). Estuarine Circulation, Mixing, and Residence Times in the Salish Sea. *Journal of Geophysical Research: Oceans*, *126*(2), e2020JC016738. https://doi.org/10.1029/2020JC016738

Meek, M. H., Beever, E. A., Barbosa, S., Fitzpatrick, S. W., Fletcher, N. K., Mittan-Moreau, C.S., Reid, B. N., Campbell-Staton, S. C., Green, N. F., & Hellmann, J. J. (2023). Understanding Local Adaptation to Prepare Populations for Climate Change. *BioScience*, *73*(1), 36–47. https://doi.org/10.1093/biosci/biac101

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), 14363. https://doi.org/10.1038/ncomms14363

Moore, S. K., Mantua, N. J., Newton, J. A., Kawase, M., Warner, M. J., & Kellogg, J. P. (2008). A descriptive analysis of temporal and spatial patterns of variability in Puget Sound oceanographic properties. *Estuarine, Coastal and Shelf Science*, *80*(4), 545–554. https://doi.org/10.1016/j.ecss.2008.09.016

Nielsen, T. G., Kjellerup, S., Smolina, I., Hoarau, G., & Lindeque, P. (2014). Live discrimination of Calanus glacialis and C. finmarchicus females: Can we trust phenological differences? *Marine Biology*, *161*(6), 1299–1306. https://doi.org/10.1007/s00227-014-2419-5

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, *27*(16), 3193–3206. https://doi.org/10.1111/mec.14792

Pacific Fishery Management Council. (2025). Pacific Coast Groundfish Fishery Management Plan, June 2025. https://www.pcouncil.org/documents/2022/08/pacific-coast-groundfish-fishery-management-plan.pdf/

Palsson, W. A., Tsou, T.-S., Bargmann, G. G., Buckley, R. M., West, J. E., Mills, M. L., Cheng, Y. W., & Pacunski, R. E. (2009). The biology and assessment of rockfishes in Puget Sound. *Fish Management Division, Washington Department of Fish and Wildlife.*

Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Schwenke, P. L., Park, L. K., & Hauser, L. (2018). Introgression among three rockfish species (Sebastes spp.) in the Salish Sea, northeast Pacific Ocean. *PLOS ONE*, *13*(3), e0194068. https://doi.org/10.1371/journal.pone.0194068

Seeb, L. (1998). Gene flow and introgression within and among three species of rockfishes, Sebastes auriculatus, S. caurinus, and S. maliger. *Journal of Heredity*, *89*(5), 393–403. https://doi.org/10.1093/jhered/89.5.393

Stout, H.A., B.B. McCain, R.D. Vetter, T.L. Builder, W.H. Lenarz, L.L. Johnson, and R.D. Methot. 2001. Status review of Copper Rockfish, Quillback Rockfish, and Brown Rockfish in Puget Sound, Washington.U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC46, 158 p.

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*(2), 170–177.

https://doi.org/10.1038/s41559-018-0777-y

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*. ISBN 978-3-319-24277-4. https://doi-org/10.1007/978-3-319-24277-4

Williams, G. D., Levin, P. S., & Palsson, W. A. (2010). Rockfish in Puget Sound: An ecological history of exploitation. *Marine Policy*, *34*(5), 1010–1020. https://doi.org/10.1016/j.marpol.2010.02.008

Willis, P. M., Ryan, M. J., & Rosenthal, G. G. (2011). Encounter rates with conspecific males influence female mate choice in a naturally hybridizing fish. *Behavioral Ecology*, *22*(6), 1234–1240. https://doi.org/10.1093/beheco/arr119

Winemiller, K. O. (2005). Life history strategies, population regulation, and implications for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(4), 872–885. https://doi.org/10.1139/f05-040

Wray, A., Petrou, E., Nichols, K. M., Pacunski, R., LeClair, L., Andrews, K. S., Kardos, M., & Hauser, L. (2024). Contrasting effect of hybridization on genetic differentiation in three rockfish species with similar life history. *Evolutionary Applications*, *17*(7), e13749. https://doi.org/10.1111/eva.13749

Wray, A., Petrou, E., Nichols, K. M., Pacunski, R., LeClair, L., Andrews, K. S., Haggarty, D., & Hauser, L. (2025). Divergent Population Structure in Five Common Rockfish Species of Puget Sound, WA Suggests the Need for Species-Specific Management. *Molecular Ecology*, *34*(1), e17590. https://doi.org/10.1111/mec.17590