Oregon Department of Fish and Wildlife

SCIENCE BULLETIN

Number 20XX-XX

[DIVISION or PROGRAM]

[Title] *[No Page Number*]

[*Leave this page blank except statements at the bottom*; *no page number*]

For authors: This document is intended as a default (non-prescriptive) template for all ODFW Science Bulletins (formerly Information Reports and Progress Reports). Report formatting, except the cover pages and ADA statement, may vary; report standards and guidelines should generally follow the style, format, and conventions of the American Fisheries Society, the Wildlife Society, or the organization funding the investigation.

Text within brackets are instructions (if italicized), or denote where to insert text. Delete all such annotations, example text, and this text box upon completing a report.

Before finalizing a report, please review Director's Office Policy 100_02, particularly the requirements for report review and manuscript dissemination. Obtain a report series number (front cover page) from staff designated in the policy.

This report should be cited as [*example*]:

Smith, A. B, C. D. Jones, and E. F. Williams. 2019. Seasonal patterns of Pacific Herring movement along the northern Oregon coast. Science Bulletin 2020-03. Oregon Department of Fish and Wildlife, Salem.

[Title page – suggested format; *note pagination*]

[Title]

[Photograph – optional]

Prepared by

Kathleen G. O'Malley, Oregon State University
David I. Dayan, Oregon State University
Sandra Bohn, Oregon State University
Chris Lorion, Oregon Department of Fish and Wildlife

Oregon Department of Fish and Wildlife
Conservation and Recovery Program
[Local address]
[City, OR 97XXX]

[Month and Year]

[Leave page blank]

CONTENTS

[*Leave page blank*]

INTRODUCTION

Chum salmon (*Oncorhynchus keta*) in Oregon coastal basins are part of the Pacific Coast Evolutionarily Significant Unit (ESU), which includes all populations from the Pacific coasts of Washington and Oregon, as well as populations in the Strait of Juan de Fuca west of the Elwha River (Johnson *et al.* 1997). Although there have been genetic studies of populations in the Pacific Coast and neighboring ESUs (e.g. Johnson *et al.* 1997; Small *et al.* 2011; Johnson *et al.* 2012), historical and contemporary population structure in the Oregon portion of the ESU remains poorly understood. Only three basins on the Oregon coast (Nehalem, Tillamook, and Yaquina) have consistent annual returns of at least several hundred spawners, but smaller numbers of chum are observed regularly or intermittently in at least ten other basins. It is unknown how many of these locations historically supported independent populations, or how they are currently functioning as dependent or independent populations. The Coastal Multi-Species Conservation and Management Plan (ODFW 2014) identified current and historical population structure of coastal chum salmon as a critical uncertainty to address through research. Genetic techniques have the potential to provide insight into chum population structure, but sample quality has hindered previous investigations along the Oregon coast (Johnson *et al.* 2012). In 2019, ODFW initiated chum tissue sample collection in several coastal basins to support a new analysis of genetic relationships among the largest chum populations of the Oregon coast. Understanding relationships among these larger populations is a first step toward understanding population structure among all coastal basins where chum currently occur.

The primary objectives of this research were to:

1) Collect tissue samples from different anatomical locations of chum salmon carcasses to investigate whether certain tissues are more likely to provide higher quality samples for analysis.
2) Analyze a small number of archival chum salmon scale samples to evaluate the potential for investigating historical population structure, using the large number of chum scale samples ODFW has collected through spawning grounds surveys over time.
3) Collect and analyze samples from the three largest coastal chum salmon populations (Nehalem, Tillamook, and Yaquina), and two additional basins (Netarts and Siletz) that often have a substantial number of spawners, to investigate genetic structure of coastal chum salmon populations.
4) Collect and analyze samples from two major Tillamook sub-basins (Kilchis and Miami Rivers) to investigate whether there is significant genetic structure within the basin.

METHODS

Reproducible Research

Detailed logs containing all scripts and data, except raw sequencing data, is available at the github repository at https://github.com/david-dayan/chum_coastal_pilot . A narrative log of

analyses with integrated code, results and commentary is available in the form of R computational notebooks at this repository and the entire repository can be cloned and run locally as a project on Rstudio. Computational notebooks are html files and can be opened with a browser. The repository is also archived at zenodo with a stable identifier DOI: 10.5281/zenodo.5143807

Sampling

Tissue samples were collected from carcasses during spawning surveys in November-December 2013, 2019 and 2020 within the Nehalem, Netarts, Siletz, Yaquina and Coos river basins, as well as along two individual rivers within the Tillamook basin (Miami and Kilchis Rivers) (table 1, figure 1). These spawning surveys are conducted by the Oregon Department of Fish and Wildlife (ODFW) as part of long-term monitoring of chum salmon along Oregon's coastal rivers. Details about spawning survey. Tissues sampled from carcasses include fin clips, operculum punches, scales, muscle, gills and eggs. Twenty archival Yaquina basin scale samples were collected in 2013 and stored dry in paper envelopes until DNA was extracted from ten in 2020. All other tissues were collected in 2019 and 2020 and stored in 95% ethanol. Archival scale samples are excluded from analysis of population genetic structure and genetic diversity. All sampled individuals were of natural-origin.

Table 1: Sampling information including: location, sample sizes before filtering ($n_{initial}$), field calls of the number of males, females and unknown sex per location (n_{male}: n_{female}:$n_{?}$) and sampling dates. Latitude and longitude of sampling locations are the approximate mouth of the river in the respective estuary. Sex determined in the field. Archival scale samples collected in 2013 are not included in this table.

Basin	Sampling Location	$n_{initial}$	n_{male}:n_{female}:$n_{?}$	Sampling Dates (2019, except where noted)	Latitude	Longitude
Nehalem	Foley Creek	50	25:25:0	11/20 and 11/25	45.690	-123.898
Tillamook	Kilchis River	50	25:25:0	11/19 and 11/25	45.499	-123.860
Tillamook	Miami River	50	25:25:0	11/14 – 11/26	45.560	-123.892
Netarts	Whiskey Creek	26	9:17:0	12/11	45.436	-123.954
Siletz	Bear Creek	3	1:1:1	11/27	44.901	-124.021
Yaquina	Mill Creek	64	32:31:1	11/27 – 12/10	44.619	-124.061
Yaquina	Simpson Creek	10	5:5:0	12/10	44.619	-124.061
Coos	?	3	2:1:0	10/15/20 – 11/23/20	43.361	-124.168

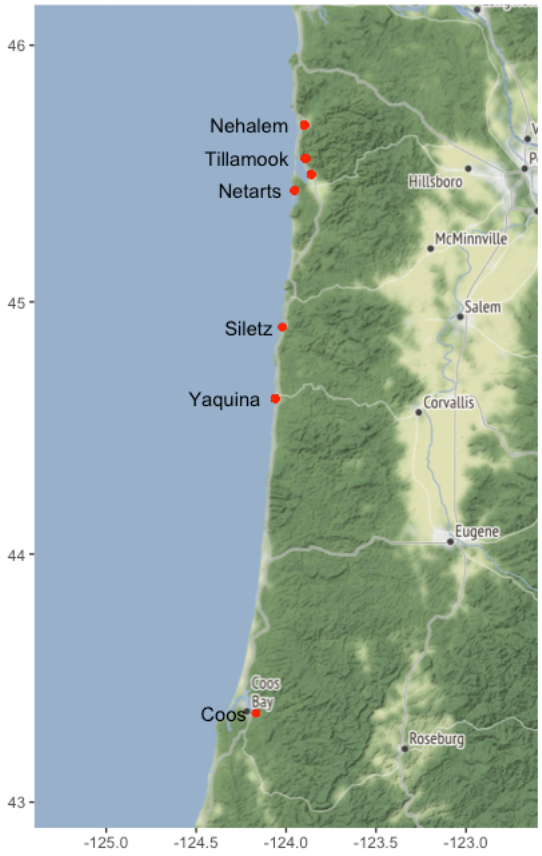


Figure 1. Map of sampling locations with basins labeled (note Tillamook basin is labeled twice, but contains two distinct sampling locations). This map is just a placeholder for the draft.

Genotyping

Genomic DNA was isolated following the protocol of Ivanova et al. (2006), with one exception: scale samples were homogenized using ceramic beads after proteinase K digestion. Samples were genotyped at a panel of previously identified single nucleotide polymorphisms (SNPs) using Genotyping-in-Thousands by sequencing (GT-seq)(Campbell *et al.* 2015). The panel consisted of 350 genetic markers. These markers target SNPs chosen to conduct genetic stock identification and explore population structure among chum salmon populations in Alaska, British Columbia, and Washington (Small 2018). The genotyping protocol followed Campbell *et al.* (2015), except the second polymerase chain reaction (PCR) used Ultra II Q5 master mix (New England Biolabs) to add i5 and i7 adapters. Amplicons were sequenced on an Illumina NextSeq 2000 at University of Oregon in two separate lanes. We included also genotyped negative controls, and approximately 40% of samples as replicates, including samples replicated both within and across lanes.

We used genotyping scripts previously developed by Campbell *et al.* (2015) which are available at https://github.com/GTseq/GTseq-Pipeline/. Genotypes were called from demultiplexed reads with a modified version of GTseq_Genotyper_v3.1pl (available at project github

repository). Individual genotype files were compiled with GTseq_GenoCompile_v3.pl. Genotype quality control was assessed using replicate samples, negative controls.

Briefly, we filtered genotypes on the basis of missingness, sample duplication and the individual fuzziness index (IFI), which estimates the amount of cross-contamination in a given sample and is calculated with the GTseq_Genotyper_v3.1pl script. We also removed sites with poorly calibrated allele correction values or more than three clusters of allele ratios suggestive of a paralogous sequence variant. We took an iterative approach to missingness and IFI filtering and recalculated missingness for all individuals and genotypes between each step (O'Leary *et al.* 2018). We began filtering by removing negative controls and replicate individuals (retained replicate with highest number of on-target reads). Then we removed individuals with more than 30% missing data, then loci with greater than 50% data, and individuals with IFI greater than 10 (i.e. greater than 10% putative background reads). In our second round of filtering, we removed individuals with more than 20% missing data, then removed loci with greater than 20% missing data, and individuals with IFI greater than five. We then examined any marker with greater than 10% missing data, and skewed or high variance in allele ratios among uncalled and heterozygous samples by plotting corrected read counts of alternative alleles. Sites with a strong bias towards one allele among heterozygotes, more than three clusters of allele ratios, or indistinct clusters of allele ratios were removed from the dataset. After genotype quality filtering was complete, we removed monomorphic loci. Further details of our genotype quality filtering is available in the computation notebook titled “Genotyping Notebook” at the archived repository for this manuscript.

We compared efficacy of genotyping success across tissue sample types recently stored in ethanol (2019) and between these samples and archival scale samples (2013). Genotyping efficacy was assessed on the basis of the raw number of reads containing an exact match to a probe and primer sequence from our GT-seq panel (on-target reads), the proportion of on-target reads, and the proportion of samples with less than 20% missing genotypes. Only samples with recorded tissue type were used in the analysis. Significant variation in mean genotyping efficacy among the most common tissue types (fin clips, operculum punches and muscle) and differences in genotyping efficacy between archival scales and ethanol stored tissue was assessed using an ANOVA.

Genetic Diversity and Differentiation

Genetic diversity metrics and differentiation were estimated in R. We estimated per-site observed and expected heterozygosity using *adegenet* (Jombart & Ahmed 2011). Significance testing of differences in heterozygosity among basins used a Monte-Carlo test with 1000 permutations implemented in *adegenet*. Empirical p-values from the Monte-Carlo test were adjusted using FDR. Hardy-Weinberg proportion estimation and significance testing at individual loci was conducted using *pegas* (Paradis 2010). Per-marker and overall F-statistics (F_{ST}, F_{IS}, etc.) were estimated using *hierfstat* and pairwise differentiation among basins was estimated using Weir and Cockerham’s estimator (1984).

Population Genetic Structure

We used both constrained and unconstrained multivariate ordination, as well as a model-based Bayesian clustering approach to interrogate potential population genetic structure among sampling locations. We focused our analysis at different spatial scales. First, we examined structure among six major Oregon coastal river basins (Nehalem, Tillamook, Netarts, Siletz, Yaquina, and Coos), combining multiple sample locations within a basin. Then we examined structure within a single basin (Tillamook), where samples were taken from two separate rivers (Kilchis and Miami). For the spatial analyses (RDA and IBD – detailed below) Kilchis and Miami Rivers we used all available spatial data and each unique sampling location is considered a separate observation.

We used principal component analysis (PCA) to examine orthonormal axes of genetic variation among samples. We conducted two PCAs, one among all samples and a second among only Kilchis River and Miami River samples. Significant axes of genetic variation were determined using the broken stick model and by examining the scree plot of PCA eigenvalues. We also examined spatial patterns genetic variation using two techniques. First, to test for a linear relationship between spatial and genetic distance among the eight sampling locations consistent with a pattern of isolation-by-distance (IBD), we estimated alongshore distances constrained by a 20m isobath using the R package *marmap* and estimated genetic distances using linearized F_{ST} (F_{ST}/(1- F_{ST})). We then conducted a Mantel test between these spatial and genetic distance matrices.

To further examine spatial patterns in the data that may be consistent with IBD, we also conducted redundancy analysis (RDA). RDA is a form of constrained ordination and is conceptually similar to fitting a principal component analysis on the fitted values from a multiple linear regression. The resulting redundant axes are orthogonal and capture the extent to which multiple explanatory variables can be used to explain multiple response variables. In our case, this means that each redundant axis captures an orthogonal component of the relationship between the genetic and spatial variation among individuals. We used the matrix of fully filtered genotypes as response variables and distance-based Moran's eigenvector maps (dbMEMs) determined from sampling locations as explanatory variables in our RDA. dbMEMs are capable of describing spatial variation at multiple scales, including spatial autocorrelation as well as local structures. We used the *moran.randtest* function of *adespatial* to test for dbMEMs with significant Moran's I, then we retained only dbMEMs with positive values of Moran's I as our explanatory variables, because we were only interested in positive spatial autocorrelation that might arise as a consequence of IBD. Significance of the RDA is tested using empirical p-values (permuting the response variables) and dbMEMs retained in the final model are chosen using using forward variable selection. We conducted the RDA using the R package *vegan* [110], using the *rda* and *anova.cca* commands to fit the RDA and to test the global significance of the model and the significance of individual explanatory variables and redundant axes.

In addition to our multivariate approaches to examine population structure we also applied STRUCTURE (Falush *et al.* 2003). Our STRUCTURE runs used a linkage disequilibrium (LD)-pruned dataset; only one marker from any marker pair with r^2 greater than 0.2 was retained. We used an admixture model with correlated allele frequency, no priors, a burn-in of 20,000 iterations, followed by 40,000 iterations and conducted ten replicates for one to six putative ancestral genetic clusters (*k*). Best *k* was chosen by the Evanno method (Evanno *et al.* 2005), and estimated in STRUCTURE HARVESTER (Earl & vonHoldt 2012). Replicate results within each *k* were combined using the clumpak algorithm (Kopelman *et al.* 2015) on the clumpak webserver. If the clumpp algorithm of CLUMPAK detected multimodality across the ten replicates at each *k*, we examined only the most common solution at each *k.*

RESULTS

Genotyping

After demultiplexing reads to individual samples, the raw sequencing dataset consisted of 428,341,672 reads among 383 samples, with 22.5% on-target (containing both primer and probe sequence from the GT-seq amplicon panel). We removed 114 intentional replicates and three negative controls. Then we removed 22 individuals with genotyping success less than 70%, and 18 markers with greater than 50% missingness. In the second round of filtering, we removed an additional eight individuals with genotyping success less than 80%, two markers with greater than 20% missingness and a single individual with IFI greater than five. We removed a single marker with poor clustering of allele ratios consistent with poor primer/probe specificity and four monomorphic markers. The final dataset consisted of 235 individuals and 325 markers, including eight archival scale samples. Mean read depth per marker per individuals in the final dataset was 934 and the median was 448. Sample sizes after filtering and missing data rate per sampling location provided in Table 2.

Table 2: Sample sizes and proportion of uncalled genotypes (missing data rate) per sampling location after filtering. Table includes samples collected in 2019 and 2020 but not the 2013 archival scale samples.

Basin	Sampling Location	n_{final}	n_{male}:n_{female}:$n_{?}$	Missing data rate
Nehalem	Foley Creek	49	25:24:0	0.028
Tillamook	Kilchis River	45	24:21:0	0.060
Tillamook	Miami River	46	22:24:0	0.070
Netarts	Whiskey Creek	13	5:8:0	0.128
Siletz	Bear Creek	3	1:1:1	0.050
Yaquina	Mill Creek	59	31:27:1	0.044
Yaquina	Simpson Creek	9	4:5:0	0.001
Coos	?	3	2:1:0	0.055

Comparison of Tissue Types

We observed no significant differences in genotyping efficacy between major tissue sample types using either total on-target read count or proportion on-target reads as a metric (ANOVA $p > 0.25$). Fin clips performed best, but the mean proportion of on-target reads was only a 13% improvement over operculum punches, and 23% improvement over muscle (Figure 2, Table 3).

Table 3. Mean number and proportion of on-target reads and mean missing data rate across tissue sample types. Fin clips, operculum, punches and muscle collected in 2019 are presented collectively as ethanol stored tissue. The 2020 Coos River samples are not included here.

Tissue Type		n	Mean On-Target Reads	Mean Proportion On-Target	Mean Proportion Missing Data
Ethanol Stored			261140	0.19	0.14
	Fin Clip	234	275931	0.20	0.10
	Operculum Punch	71	248347	0.17	0.22
	Muscle	45	204405	0.16	0.25
Archival Scale		10	142651	0.11	0.25

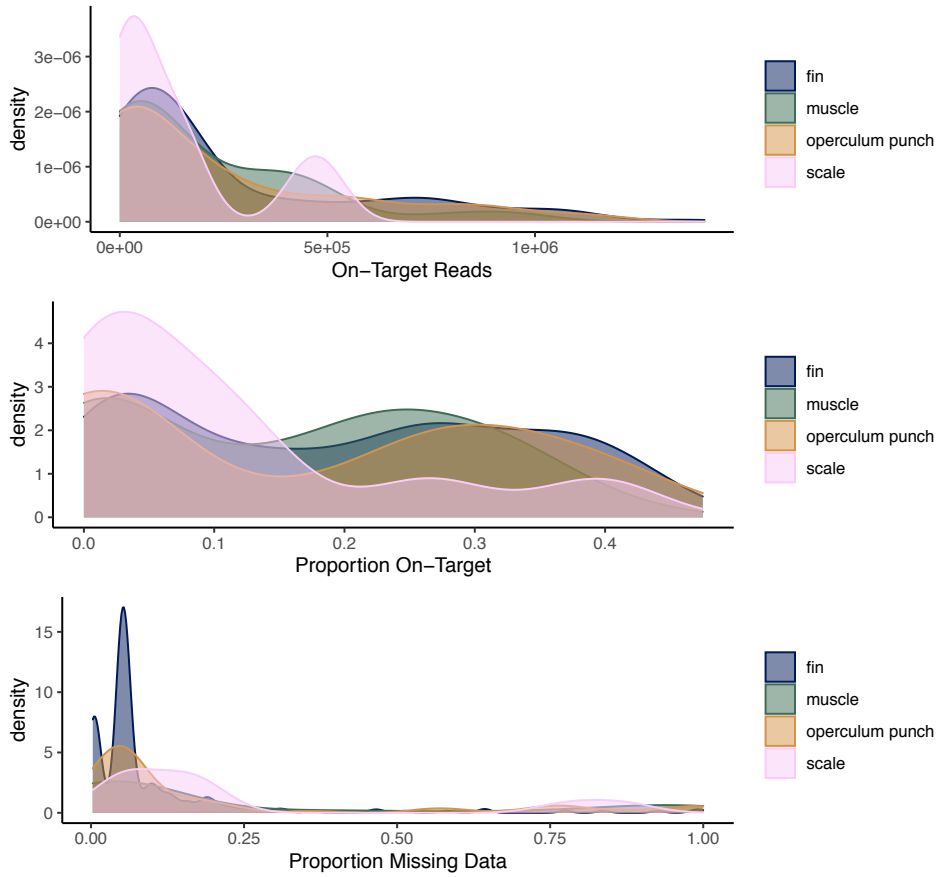


Figure 2: Density plots of genotyping efficacy metrics across tissue sample types.

We also examined differences in genotyping efficacy using DNA extracted from archival scale samples (2013) and DNA extracted from tissues recently stored in ethanol (2019 and 2020).

Archival scale samples have roughly one half the number and proportion of on-target reads and missing data. There was not a significant difference in the number or proportion of on-target reads between ethanol stored and archival scale samples (p = 0.09 and 0.223 respectively, ANOVA). However, there was an interesting pattern of variance in the genotyping success of ethanol stored vs archival scale samples (Figure 2). While both tissue types had similar proportions of samples that completely failed genotyping (greater than 20% missing data, about 20% of samples from each, Figure 2), there was higher variance among scale samples in missing data rate among samples with less than 20% missing data.

Genetic Diversity and Differentiation

Overall genetic diversity as estimated by expected heterozygosity (H_e), was high among markers in the GT-seq panel (figure 3). H_e did not vary significantly between any pair of basins (Figure 3, FDR-corrected Monte-Carlo test p-values > 0.1). While there was generally greater observed than expected heterozygosity within each basin and in the dataset overall, very few markers demonstrated a significant departure from Hardy-Weinberg proportions, including one marker with excess heterozygosity in the Tillamook basin and two markers with excess homozygosity in the Yaquina basin (FDR-corrected Monte-Carlo test p-values < 0.1).

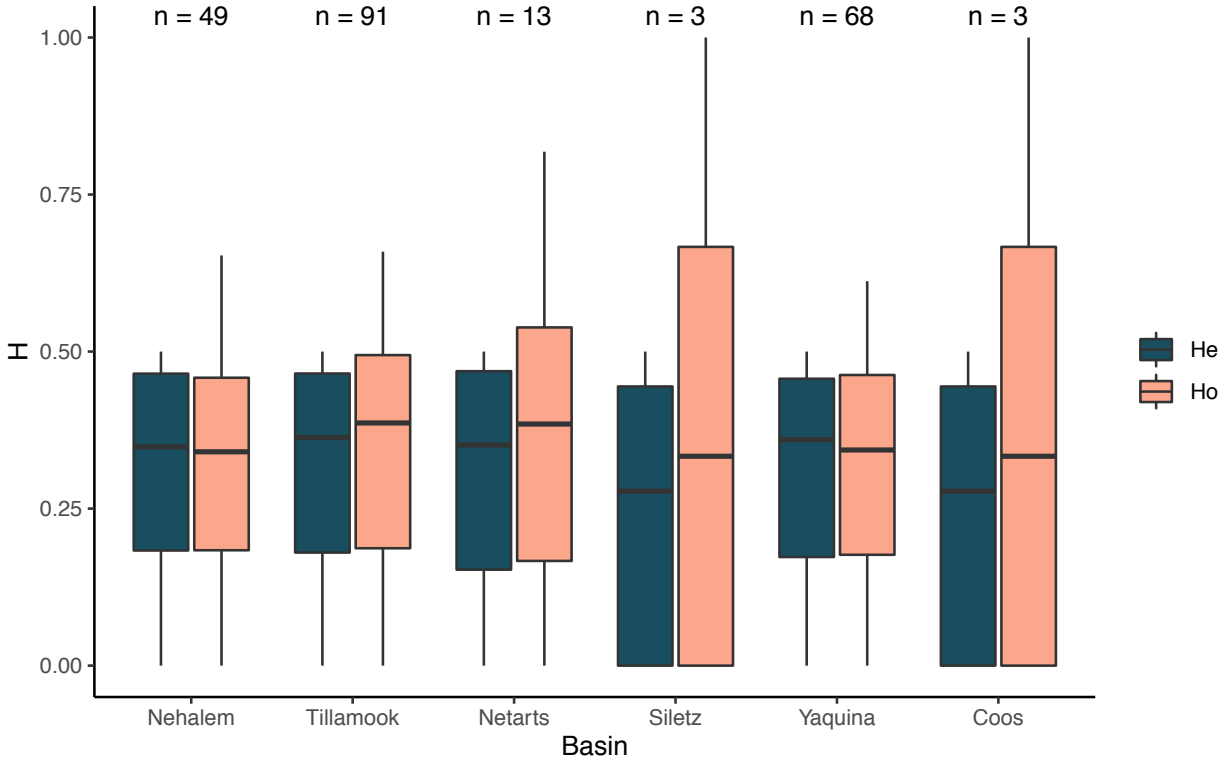


Figure 3. Observed (H_o) *vs.* expected (H_e) heterozygosity across basins. Sample size presented for each basin. Basins are ordered from north to south.

Overall F_{ST} was 0.0203 and F_{IS} was -0.0358, indicating an excess of heterozygosity. Pairwise F_{ST} estimates between basins ranged from 0.003 to 0.042 (Table 4). Within-basin

differentiation was low. F_{ST} between the Kilchis River and Miami River samples was 0.003. This is the same level of differentiation between Nehalem basin samples and the Kilchis and Miami River samples collectively (Tillamook basin).

Table 4. Among basin pairwise differentiation (F_{ST}).

	Nehalem	Tillamook	Netarts	Siletz	Yaquina
Tillamook	0.003				
Netarts	0.015	0.017			
Siletz	0.022	0.029	0.042		
Yaquina	0.014	0.015	0.025	0.017	
Coos	0.027	0.030	0.041	0.033	0.040

Population Genetic Structure

PCA and STRUCTURE -- No principal components of genetic variation were significant using the broken stick model, but visual examination of the scree plot suggested that the first principal component (PC1) may be meaningful (Supplemental Figure 1). PC1 (2.1% of total variance) largely separates Yaquina basin individuals from all other individuals, except the Siletz individuals. Two of the Siletz individuals cluster with the Yaquina individuals along PC1 while the third falls into the second cluster with all other individuals (Figure 4). We did not observe any differences between Kilchis River and Miami River samples in the PCA, nor were we able to identify any structure using a PCA conducted only on Kilchis and Miami Rivers (Supplemental Figure 2).

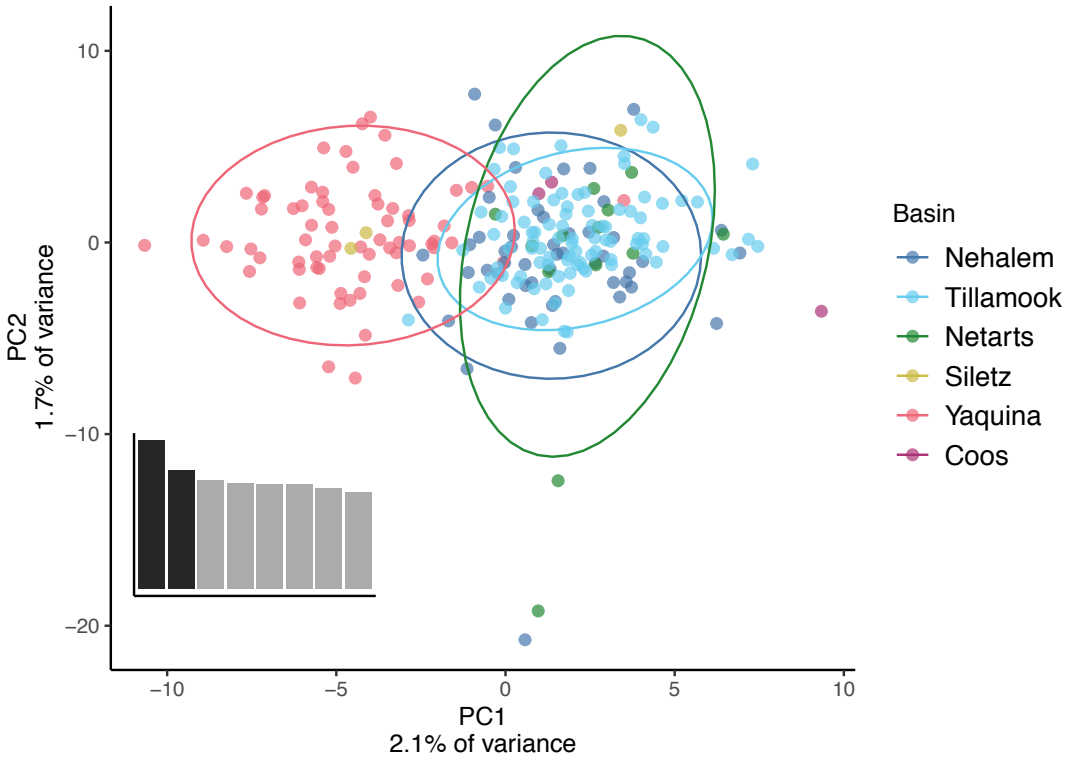


Figure 4: PCA of genetic variation among individuals, colored by basin. Basins are ordered from north to south. Screeplot of first eight eigenvalues presented as bottom left inset. 95% confidence ellipses are drawn for basins with sample size greater than three.

We removed seven markers that demonstrated significant LD with other markers in the data and used this LD-thinned dataset to run STRUCTURE (Figure 5). Best *K* was 2 according to the Evanno (delta *K*) method, however given the level of differentiation and number of markers, we considered all *K*. The most salient pattern we observed in the STRUCTURE results was the presence of an ancestry cluster across all *K* that composed a major proportion of modeled ancestry in most Yaquina and two of three Siletz samples, but only a minor proportion of modeled ancestry in all other basins. However, no individuals derived all of their ancestry from a single cluster and there was high variance in ancestry proportions within all basins, suggestive of limited structure or high admixture.

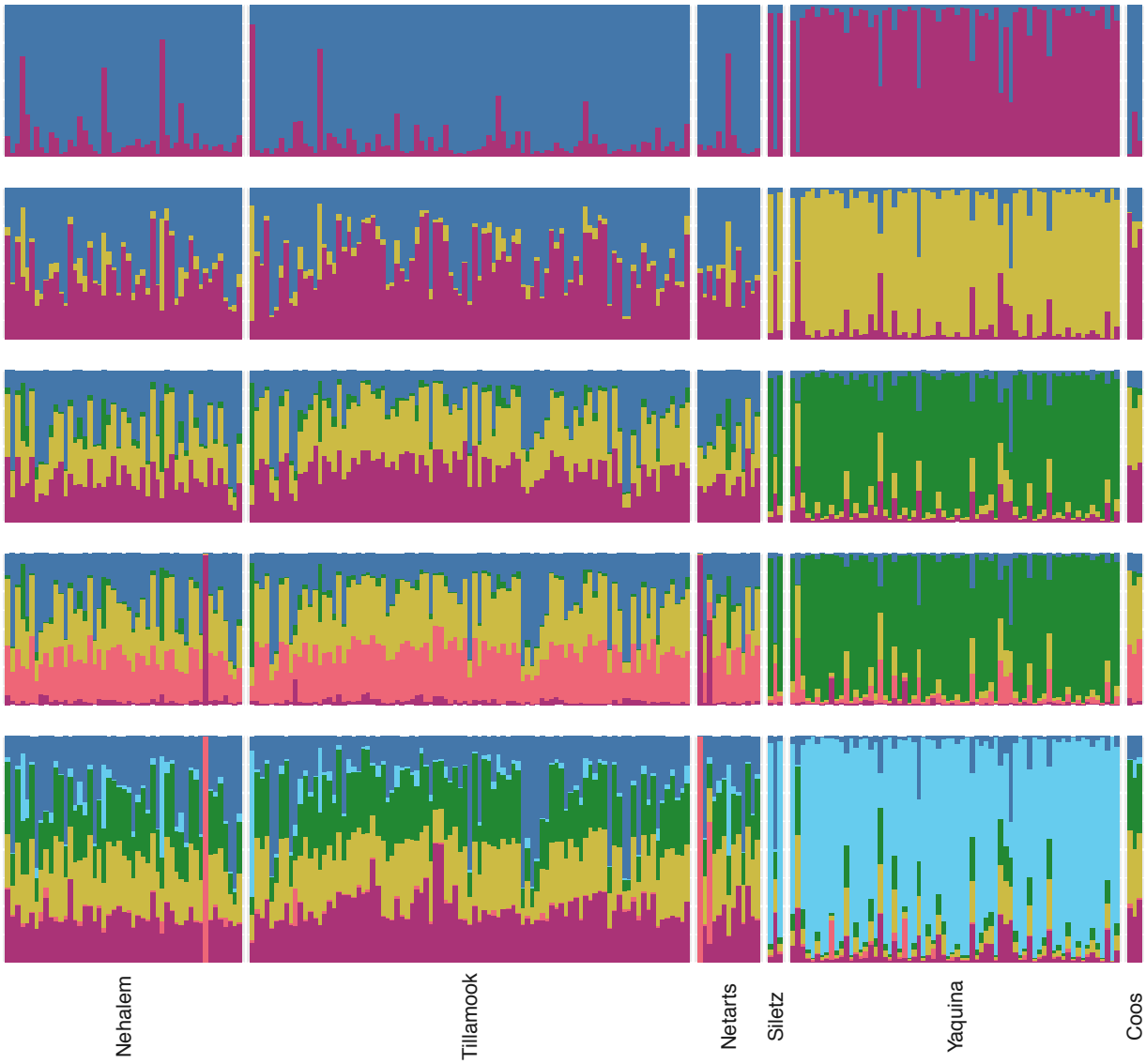


Figure 5. STRUCTURE plots for K = 2 (top) to K = 6 (bottom).

Spatial Analyses -- We found evidence of spatial autocorrelation of genetic variation consistent with a pattern of IBD among samples. The Mantel test between linearized F_{ST} and alongshore distance (km) between sampling sites was significant and highly explanatory (Figure 6, p = 0.04, Mantel r-statistic = 0.55). We identified two dbMEMs that captured spatial autocorrelation (Moran's I > 0) and both dbMEMs were significant after permutation (p = 0.001, 999 permutations). The first dbMEM describes the broadest range pattern of autocorrelation and largely separates Coos River samples from all other samples (Supplemental Figure 3a). The second dbMEM captures finer scale autocorrelation among the remaining samples, (Supplemental Figure 3b). RDA results are summarized in Figure 7. The global model with both dbMEMs was significant (p = 0.001, 999 permutations) and we retained both dbMEMs after variable selection. Both RDA axes were significant (FDR-adjusted p-value = 0.002) and

collectively RDAs constrained 1.76% of variance. Given that the global F_{ST} in the dataset was 0.0203, this means we can explain up to ~87% of among basin genetic variation as spatial autocorrelation among samples. The primary axis of constrained variation (RDA1, 1.0% of total variance) was driven by dbMEM2 which largely captures distance away from Yaquina. The second axis (RDA2, 0.7% of total variance) was driven mostly by dbMEM1 and separates Coos Bay samples from all others.

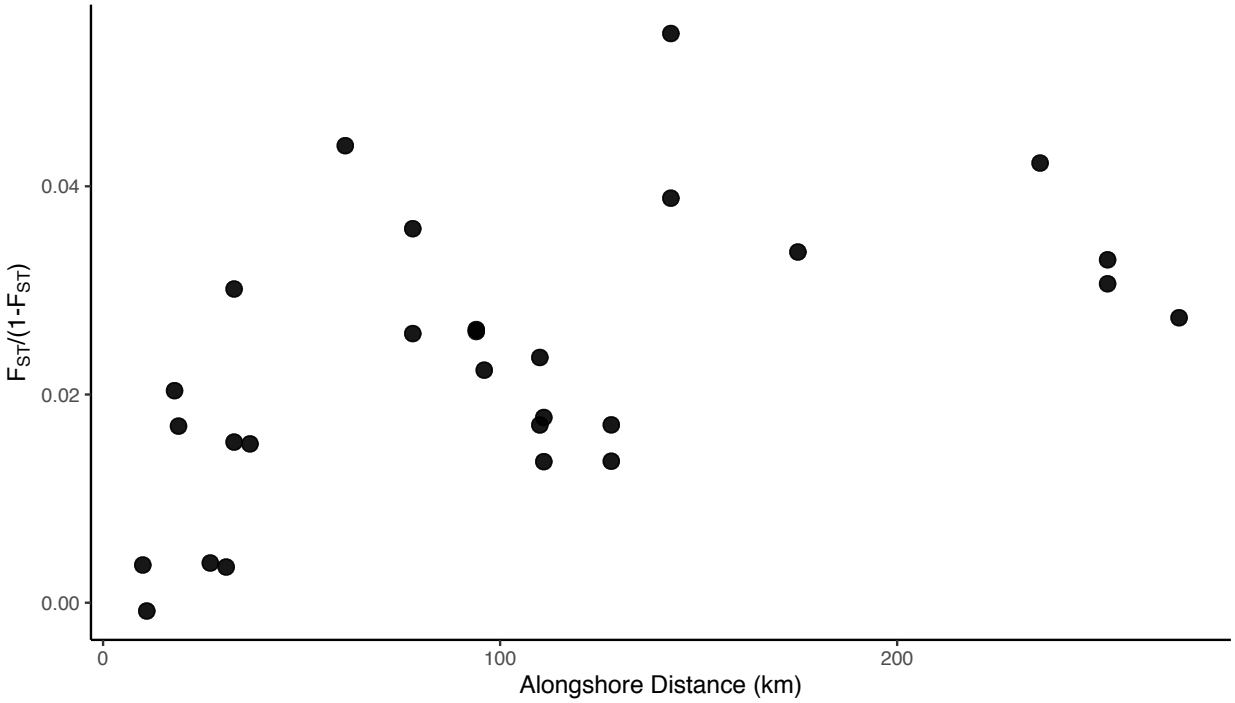


Figure 6. Isolation-by-distance plot. Alongshore distance (km) (constrained by a 20m isobath) between sampling locations, and genetic distance presented as linearized F_{ST} (F_{ST} /(1- F_{ST})).

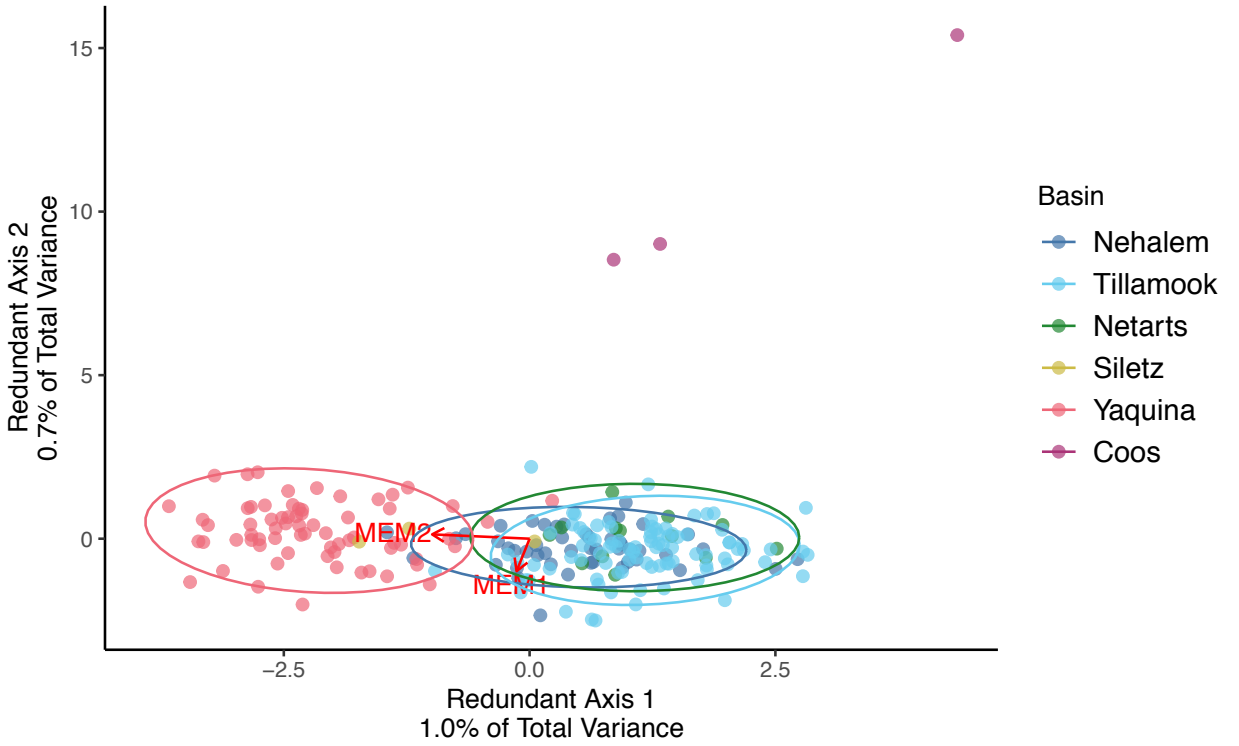


Figure 7. RDA biplot showing individual scores along the first two redundant axes and explanatory variable (dbMEMs). Points are colored by sampling location basin. 95% confidence ellipses are drawn for basins with sample size greater than three. Perpendicular projections of individual points onto red arrows represent explanatory variable values (dbMEMs) for a given individual (type I scaling).

DISCUSSION

The main findings of the study:

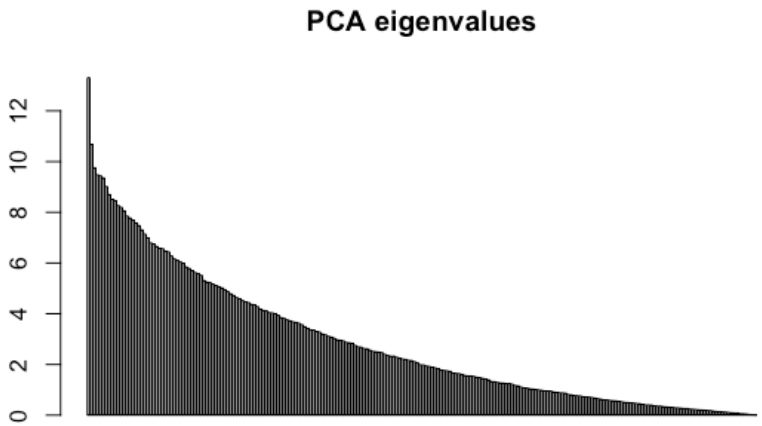
- Objective 1: Tissue Sampling
    - DNA extracted from fin clips, operculum punches and muscle performed similarly. Future chum salmon carcass sampling should prioritize fin clips to streamline sample processing, but if fin tissue is in poor condition, samplers should opt for operculum punches or muscle.
- Objective 2: Archival Scale Samples
    - Scale samples stored dry for eight years provided sufficient DNA quantity and quality to produce GT-seq genotypes
    - However, archival scale samples produced approximately one-half the number and proportion of on-target reads than ethanol stored tissues, suggesting that fewer archival scale samples can be multiplexed per library than estimated from ethanol stored tissues. Therefore, the cost per sample will be higher compared to ethanol-stored tissue samples.
- Objective 3: Genetic Structure of Coastal Chum Salmon Populations
    - A chum salmon GT-seq panel developed to conduct genetic stock identification among chum salmon in Alaska, British Columbia, and Washington is effective for exploring population genetic structure in coastal Oregon chum salmon populations
    - Our results suggest that there is population genetic structure within coastal Oregon chum salmon, but the precise nature of this structure is best described as unresolved since sample sizes were adequate for only three (Nehalem, Tillamook, and Yaquina) of the six basins.
        - We found evidence for low genetic differentiation between Yaquina and the two northern basins (Nehalem and Tillamook)
        - PCA and STRUCTURE results point to two major genetic clusters: one including Yaquina individuals and two of three Siletz individuals, and a second including all other samples
    - An intriguing pattern of isolation-by-distance (IBD) was found to operate at two spatial scales, with different sets of GT-seq markers involved in each, but limited sample sizes at Siletz and Coos rivers limits our confidence in this finding.
        - Long distance: Strong correlation between genetic differentiation and spatial distance between samples points to isolation-by-distance, but this pattern is strongly driven by long spatial distance pairwise comparisons with Coos River, where sample size is small.
        - Short-distance: Isolation-by-distance at shorter scales is largely driven by differences between Yaquina and northern samples (Nehalem, Tillamook and Netarts), which showed limited differentiation from each other. Siletz was intermediate between these two groups, consistent with isolation-by-distance at this spatial scale, but limited sample size for Siletz challenges the strength of this inference.

        - Each redundant axis fitted by the RDA captures an orthogonal component of the relationship between the genetic and spatial variation among individuals. In our results, the first such redundant axis is driven by short-scale spatial distance and mostly clusters Yaquina samples separately from all other samples, with Siletz intermediate. The second axis is driven by long-scale spatial distance and mostly clusters Coos samples from all other samples. These findings suggest that spatial genetic patterns at short and long scales are driven by differentiation at different sets of genes. However, findings at both short and long scales are strongly influenced by data at spatial distances where sample size is small (Siletz and Coos)
    - Future sampling should focus on the same basins to determine if findings are consistent across years but with particular focus on Netarts and Siletz to increase sample sizes. Similarly, additional samples collected south of the Yaquina River (e.g. Coos River) would provide greater insight into the patterns of IBD.
- Objective 4: Genetic Structure within Tillamook Basin
    - There is no evidence of population genetic structure within Tillamook basin based on the 350 markers examined in this study. However, when using a small panel of genetic markers, it should be emphasized that absence of evidence is not evidence of absence. While the overall level differentiation is low, and genome-wide genetic structure is not apparent from the data, ecologically relevant genetic differences may exist at regions of the genome not tagged by the genetic markers analyzed in this study.

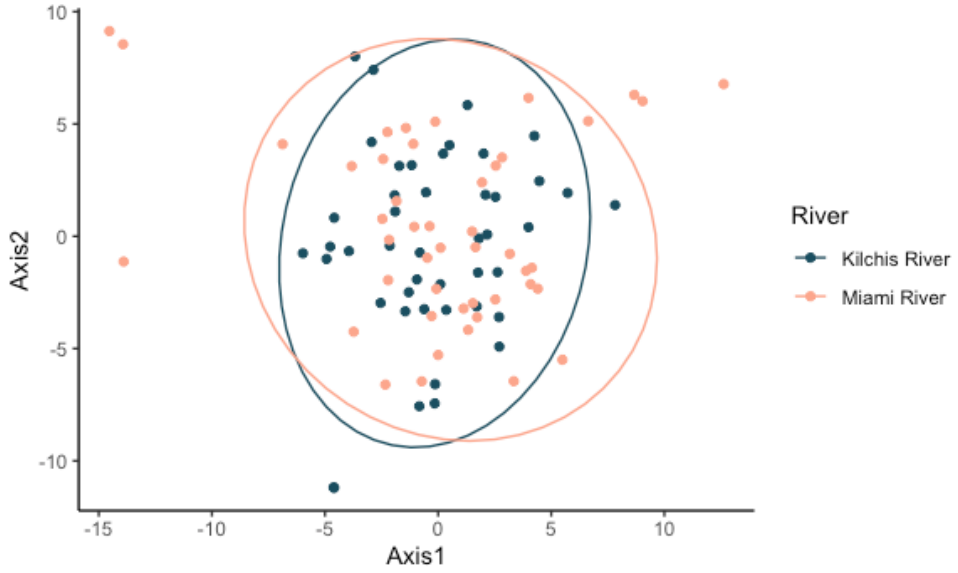
ACKNOWLEDGMENTS

Example text example text.

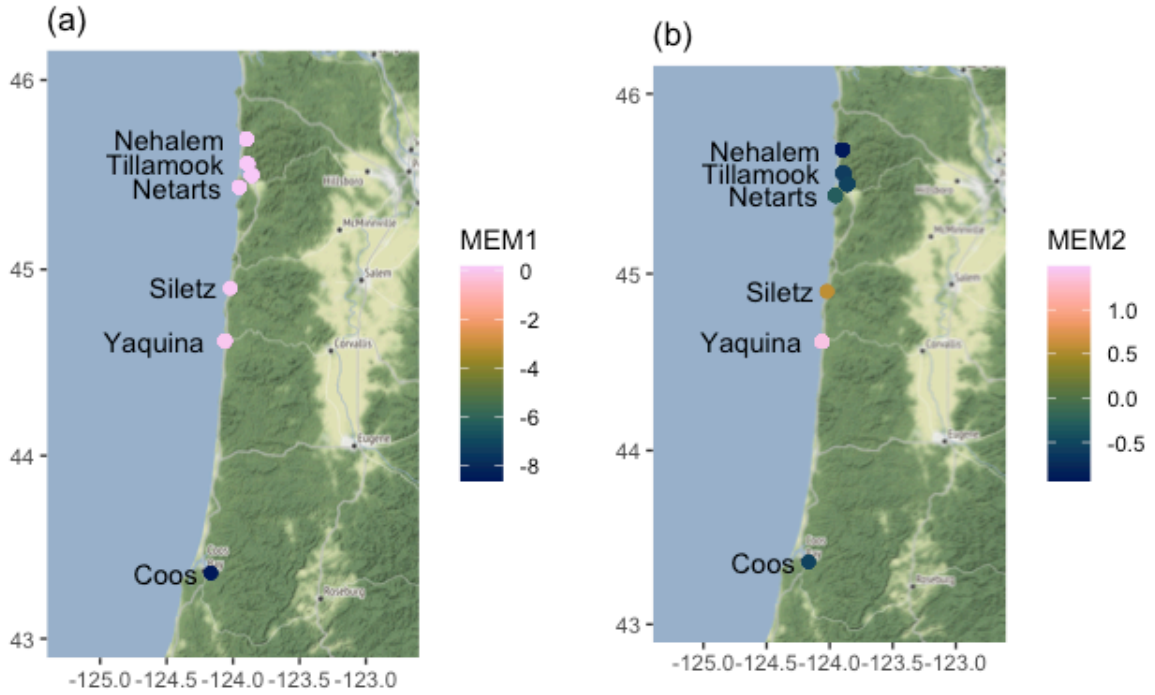
SUPPLEMENTAL FIGURES



Supplemental Figure 1: Screeplot of eigenvalues from PCA of genetic data.



Supplemental Figure 2: First two principal components of PCA using only Kilchis and Miami River samples.



Supplemental Figure 3: Mean value of first two dbMEMs at each sampling location. Text labels correspond to rough position of basin and Tillamook basin contains two sampling locations.

REFERENCES

Campbell NR, Harmon SA, Narum SR (2015) Genotyping-in-thousands by sequencing (gt-seq): A cost effective snp genotyping method based on custom amplicon sequencing. *Mol Ecol Resour* **15**, 855-867.

Earl DA, vonHoldt BM (2012) Structure harvester: A website and program for visualizing structure output and implementing the evanno method. *Conservation Genetics Resources* **4**, 359-361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol Ecol* **14**, 2611-2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* **6**, 998-1002.

Johnson, O.W., W.S. Grant, R.G. Kope, K. Neely, F.W. Waknitz, and R.S. Waples. 1997. Status review of chum salmon from Washington, Oregon, and California. U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-NWFSC No. 32.

Johnson, O.W., A. Elz, J. Hard, and D.S. Stewart. 2012. Why did the chum cross the road? Genetics and life history of chum salmon in the southern portion of their range. North Pacific Anadromous Fish Commission Technical Report No. 8: 135-137.

Jombart T, Ahmed I (2011) Adegenet 1.3-1: New tools for the analysis of genome-wide snp data. *Bioinformatics* **27**, 3070-3071.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across k. *Molecular Ecology Resources*.

ODFW. 2014. Coastal Multi-Species Conservation and Management Plan. Oregon Department of Fish and Wildlife, Salem, OR. Available at: http://www.dfw.state.or.us/fish/CRP/coastal_multispecies.asp

O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS (2018) These aren't the loci you'e looking for: Principles of effective snp filtering for molecular ecologists. *Mol Ecol*.

Paradis E (2010) Pegas: An r package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419-420.

Small, M.P., B. Glaser, T. Hillson, and C. Bowman. 2011. Population genetic structure and recovery of chum salmon in the Lower Columbia River. Washington Department of Fish and Wildlife Molecular Genetics Laboratory, Olympia, WA. Available at: https://pisces.bpa.gov/release/documents/documentviewer.aspx?doc=P122957

Small MW, Kenneth; Pascal, Carita; Seeb, Lisa; Ruff, Casey; Zischke, Jay; Winans, Gary; Jim Seeb (2018) Report to southern fund panel: Chum salmon southern area genetic baseline enhancement part 1 and part 2: Amplicon development, expanded baseline collections, and genotyping.

Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population structure. *Evolution*, 1358-1370.

[Inside of back cover; *no page number*]

[Back cover; *no page number*]

4034 Fairview Industrial Drive SE
Salem, OR 97302