

**Mixed stock and landscape genetics analyses of sockeye salmon in the  
Copper River, Alaska using SNPs**

Michael W. Ackerman

A thesis  
submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

University of Washington  
2010

Program Authorized to Offer Degree:  
School of Aquatic and Fishery Sciences

University of Washington  
Graduate School

This is to certify that I have examined this copy of a master's thesis by

Michael W. Ackerman

and have found that it is complete and satisfactory in all respects,  
and that any and all revisions required by the final  
examining committee have been made.

Committee Members:

---

Lisa Seeb

---

Jim Seeb

---

Thomas Quinn

Date: \_\_\_\_\_

In presenting this thesis in partial fulfillment of the requirements for a master's degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Any other reproduction for any purposes or by any means shall not be allowed without my permission.

Signature\_\_\_\_\_

Date\_\_\_\_\_

# Table of Contents

	<b>Page</b>
List of Figures .....	ii
List of Tables .....	v
Chapter 1: SNPs under diversifying selection provide increased accuracy and precision in mixed stock analyses of sockeye salmon from Copper River, Alaska and nearby coastal areas .....	1
Abstract .....	1
Introduction .....	2
Methods .....	5
Sample collections .....	5
Laboratory analysis .....	5
Pooling collections .....	6
Deviation from Hardy-Weinberg and linkage equilibria .....	6
Diversity and information statistics and detection of non-neutral loci .....	7
Reporting groups for proportional and individual assignment tests .....	8
Improved resolution with non-neutral loci .....	9
Baseline evaluation for mixed stock analysis .....	10
Results .....	12
Sample collections .....	12
Laboratory analysis .....	13
Pooling collections .....	13
Deviation from Hardy-Weinberg and linkage equilibria .....	13
Diversity and information statistics and detection of non-neutral loci .....	14
Reporting groups for proportional and individual assignment tests .....	15
Improved resolution with non-neutral loci .....	15
Baseline evaluation for mixed stock analysis .....	16
Discussion .....	18
Genetic diversity and population structure .....	18
Use of 100% simulations for baseline evaluation .....	19
Linked loci in GSI studies .....	20
Discovery of SNP loci under diversifying selection .....	22
Conclusions .....	23
Chapter 2: Analysis of the spatial genetic structure of sockeye salmon ( <i>Oncorhynchus nerka</i> ) in the Copper River, Alaska .....	37
Abstract .....	37
Introduction .....	37
Methods .....	41
Sampling and genetic data collection .....	41
Statistical analyses .....	41
Results .....	44
Discussion .....	48
References .....	60

## List of Figures

Figure Number	Page
Figure 1.1. Map of sample sites (populations) of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska. Sample site numbers correspond to the population numbers in Table 1.1.....	29
Figure 1.2. Neighbor-joining dendrogram of populations of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska using the full marker panel. Bootstrap values greater than 50 percent and a scale of Cavalli-Sforza and Edwards (1967) distance are shown. Numbers correspond to populations listed in Table 1.1. Gray shading corresponds with proportional assignment reporting groups (Table 1.1). Asterisked populations are outlier populations that were placed in the following reporting groups: 20) CHTNA, 23) DELTA, 38) PWS, 41) ALSEK.....	30
Figure 1.3. $F_{ST}$ values as a function of heterozygosity (ARLEQUIN v3.5; Excoffier and Lischer 2010) for 42 SNP loci (mtDNA SNPs excluded). The dashed line represents the median and the solid lines represent the 95% confidence interval boundaries based on coalescent simulations and using a hierarchical island model. The four loci labeled above the upper 95% confidence interval boundary are candidates for diversifying selection.....	31
Figure 1.4. Distribution of $f_{ORCA}$ scores (Rosenberg et al. 2003) for three different sets of 500 marker panels. For each set of marker panels, 1,000 individuals were simulated, each generated using allele frequencies from one of the 41 baseline populations chosen at random. Each simulated individual was generated using 34 of the 37 neutral loci randomly sampled (with replacement). That individual was then generated using the same 34 neutral loci with an additional three neutral loci (randomly sampled from the same panel of 37 neutral loci) and with an additional three non-neutral loci. The $f_{ORCA}$ score represents the proportion of 1,000 individuals that correctly assign back to the population from which they were generated.....	32

## List of Figures (continued)

Figure Number	Page
Figure 1.5. Mean correct allocation (90% confidence intervals) for mixtures of simulated fish (SPAM v3.7b; updated version of Debevec et al. 2000). An analysis was run for each of the 10 reporting groups in which the simulated mixture was composed entirely of individuals (n=300) generated using allele frequencies from populations within that reporting group, and each population within the reporting group contributing equally to the mixture. Mean correct allocations were estimated from 1,000 simulations. Ninety percent confidence intervals for each reporting group estimate were computed using 1,000 bootstrap resamples and the non-symmetric percentile method.....	33
Figure 1.6. The expected allocation and mean correct assignment allocation and 95% credibility intervals using the neutral and the full marker panels for three known mixtures of 300 individuals each removed from the baseline and assigned back to the baseline (BAYES; Pella and Masuda 2001). Known mixtures are (a) equal number of fish from each of the reporting groups (i.e. 30 fish each) and (b) representative of what might be expected in fisheries in the region (i.e. the Copper River district commercial gillnet fishery) based on total escapement data from the 2005 to 2007 seasons (Fair et al. 2008; Wade et al. 2008).....	34
Figure 1.7. Self-assignment matrices of populations of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska as determined with (a) the neutral marker panel and (b) the full marker panel (GENECLASS2; Piry et al. 2004). The diagonals represent the percentage of self-assigned individuals from a populations and shaded blocks above and below the diagonal indicate percentage of mis-assignments to populations. The black grid lines correspond to the individual assignment reporting groups. Legends depict percentage of assignment in 10% increments.....	35
Figure 1.8. Minor allele frequencies of a) <i>One_GPDH2-187</i> and <i>One_GPDH-201</i> and b) <i>One_MHC2_190</i> and <i>One_MHC2_251</i> for sockeye salmon populations surveyed from the Copper River and adjacent coastal drainages in southcentral Alaska. Population numbers correspond to numbering in Table 1.1 and Figure 2.1.....	36

## List of Figures (continued)

Figure Number	Page
Figure 2.1. Gradient of allelic richness, population groups, and inferred barriers for sockeye salmon from the Copper River, southcentral Alaska. Population symbols indicate the groupings defined by SAMOVA. Barriers indicate partial gene flow barriers transferred from the Delaunay triangulation in BARRIER to approximate locations on the map. Numbers inside the parentheses indicate the robustness of each barrier based on 100 bootstrap samples. The “adaptive barrier” observed above population #17 indicates the additional barrier that was observed when including the candidate SNPs for diversifying selection. The population symbols in parentheses for populations #17 and #19 indicate the groupings that were determined when including the candidate markers for diversifying selection.....	62
Figure 2.2. Gradient of allelic richness, population groups, and inferred barriers for Chinook salmon from the Copper River, southcentral Alaska. Population symbols indicate the groupings defined by SAMOVA. Barriers indicate partial gene flow barriers transferred from the Delaunay triangulation in BARRIER to approximate locations on the map. Numbers inside the parentheses indicate the robustness of each barrier based on 100 bootstrap samples.....	63
Figure 2.3. $F_{ST}$ as a function of heterozygosity (ARLEQUIN v3.5; Excoffier and Lischer 2010) for a) 42 sockeye salmon SNPs and b) 41 Chinook salmon SNPs. The dashed line represents the median and the solid lines represent the 99% confidence interval boundaries based on coalescent simulations and using a finite island model. Loci labeled above the upper 99% confidence interval boundary were determined as candidates for diversifying selection.....	64
Figure 2.4. Pairwise genetic divergence among populations as a function of fluvial distance for a) sockeye salmon and b) Chinook salmon from the Copper River, southcentral Alaska. Chinook salmon data adapted from Seeb et al. (2009b).....	65
Figure 2.5. Allelic richness as a function of upstream migration distance for sockeye and Chinook salmon populations from the Copper River, southcentral Alaska.....	66

## List of Tables

Figure Number	Page
<p>Table 1.1. Reporting group, years sampled, sample size (<i>n</i>), observed and expected heterozygosities, and self-assignment success for sockeye salmon populations surveyed for 45 SNPs from the Copper River and adjacent coastal drainages in southcentral Alaska. Genetic diversity estimates and self-assignment success (GENECLASS; Piry et al. 2004) for each population are given using both the neutral marker panel (37 nuclear neutral DNA loci; three mtDNA loci as composite haplotype for self-assignment success) and the full marker panel (neutral marker panel plus non-neutral loci; <i>One_MHC2_190</i> dropped for diversity estimates; MHC loci as composite phenotype for self-assignment success). Proportional assignment reporting groups are as follows: Slana R. (SLANA), Tanada Lk. (TNADA), Gulkana R. (GLKNA), Tazlina R. (TZLNA), Klutina Lk. tributaries (KLTNA), Klutina and Tonsina Lk. outlets (KLTON), Chitina R. (CHTNA), Lower Copper R., Copper R. Delta, Bering Lk., Eyak Lk. (DELTA), Prince William Sound (PWS), and Alsek R. (ALSEK). Site name and reporting group abbreviations include ‘Cr.’ for creeks, ‘Lk.’ for lakes, and ‘R.’ for rivers.....</p>	25
<p>Table 1.2. Summary of 45 SNP markers (Habicht et al. 2010) from 41 populations of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska. Summary statistics include minor allele frequency range, expected heterozygosity (<math>H_E</math>), observed heterozygosity (<math>H_O</math>), Weir and Cockerham (1984) <math>F_{ST}</math>, informativeness for assignment (<math>I_n</math>), and optimal rate of correct assignment (ORCA). The column ‘CDS’ indicates loci that were determined to candidates for diversifying selection.....</p>	27
<p>Table 2.1. Years sampled, sample size (<i>n</i>), allelic richness (AR), and expected heterozygosities (<math>H_E</math>) for sockeye and Chinook salmon populations from the Copper River, southcentral Alaska. Chinook salmon data adapted from Seeb et al. (2009b).....</p>	55
<p>Table 2.2. Summary of 45 SNPs from 28 sockeye salmon populations and 51 SNPs from 13 Chinook salmon populations from the Copper River, southcentral Alaska. Summary statistics include minor allele frequency range, expected heterozygosity (<math>H_E</math>), and Weir and Cockerham (1984) <math>F_{ST}</math>. A (+) in the comments indicates a candidate for diversifying selection. Chinook salmon SNP data adapted from Seeb et al. (2009b).....</p>	57

## List of Tables (continued)

Figure Number	Page
Table 2.3. AMOVA results for populations when grouped to minimize within-group variation ( $F_{SC}$ ) and to maximize among group variation ( $F_{CT}$ ) using SAMOVA v1.0. Groupings 2a through 10a indicate analyses when candidate markers for diversifying selection were included. The bold rows indicate the groupings where $F_{SC}$ declined markedly and $F_{CT}$ began to plateau and are the groupings depicted in Figures 2.1 and 2.2. The numbers in each group indicate the population ID in Table 1.1.....	60

# **Chapter 1: SNPs under diversifying selection provide increased accuracy and precision in mixed stock analyses of sockeye salmon from Copper River, Alaska and nearby coastal areas**

## **Abstract**

Genetic markers are increasingly used for ascertaining the population of origin of individuals or mixtures of individuals in Pacific salmon mixed stock fisheries. Multi-locus genotype data from single nucleotide polymorphisms (SNPs) are especially useful for admixture analyses. SNPs can be discovered in non-model organisms with relative ease and may be characterized in coding regions of the genome influenced by selection. SNPs influenced by selection may show atypically high levels of differentiation among populations, and thus be particularly valuable for genetic stock identification if neutral loci show no differences between geographically proximate populations. We identified four SNP loci from a panel of 45 as candidates for diversifying selection in sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska. We evaluated the information content of these four loci for use in genetic stock identification and assessed their ability to improve the accuracy and precision of composition estimates. The average measure of informativeness for assignment,  $I_n$ , for neutral loci was 0.019 and the average  $I_n$  for non-neutral loci was 0.064. A simulation-based approach indicated that the addition of non-neutral SNP loci to a neutral marker panel provided significantly higher resolution to assign individuals to population of origin than would have been accomplished by adding an equal number of neutral loci. SNP loci influenced by selection improved the ability to identify straying by individual fish and estimate the composition of Pacific salmon populations in mixed fisheries.

## Introduction

The high fidelity of Pacific salmon (*Oncorhynchus* spp.) to their natal streams allows adaptation to the local spawning and rearing environments, resulting in a range of life history characteristics and genetic variation among discrete populations (Taylor 1991; Dittman and Quinn 1996; Quinn 2005). Marked population differentiation allows researchers to estimate the population of origin of individuals or mixtures of individuals using genetic stock identification (GSI) techniques (Shaklee et al. 1999). Initially, fish are sampled from discrete “reference” populations that might contribute to a mixed fishery of concern and genotyped to establish a genetic baseline. The resolution of GSI given a baseline is then evaluated by assigning individuals (either simulated or real) of known origin back to the reference populations or to reference populations grouped geographically or by life history traits (reporting group). Fish captured from a mixed fishery can then be genotyped and assigned back to reporting group of origin (Pella and Milner 1987).

Two primary assignment methods are used within GSI: 1) each individual is assigned back to a reporting group of origin (individual assignment), and 2) the aggregate proportions of fish from mixed populations are assigned back to reporting groups of origin (proportional assignment; Pella and Masuda 2001; Koljonen et al. 2005). GSI technologies are being used increasingly in management applications to supplement, and in some cases replace, physical tags for estimating the origin of individuals or mixtures of individuals in Pacific salmon mixed fisheries (Seeb et al. 2000; Beacham et al. 2004; Anderson and Garza 2006).

The increased use of and investment in GSI technologies has coincided with a rapid evolution of genetic markers. Initial efforts using GSI in Pacific salmon used allozyme protein polymorphisms to estimate compositions of mixed fisheries (Shaklee and Phelps 1990; Utter and Ryman 1993; Seeb et al. 2000). However, in many cases the limited variability of allozymes did not provide sufficient resolution to address management concerns in some species (Wehrhahn and Powell 1987; Bartley et al. 1992). Further limitations, including lethal sampling, cryopreservation, finite numbers of loci, and low-throughput genotyping, hindered large-scale GSI efforts using allozymes and led researchers to seek alternative genetic markers based on DNA (Seeb et al. 2007). With

declining genotyping costs and the completion of a multi-agency, standardized microsatellite baseline for Chinook salmon (*Oncorhynchus tshawytscha*) across their native range (Seeb et al. 2007), microsatellites began replacing allozymes for many Pacific salmon GSI applications. In addition to microsatellites, researchers began using single nucleotide polymorphism (SNP) markers increasingly to resolve the composition of mixed fisheries (Smith et al. 2005c; Habicht et al. 2010). SNPs represent the most abundant form of variation in the genome of most organisms and can be discovered throughout the genome of non-model organisms with relative ease (Brumfield et al. 2003). SNPs are attractive for large-scale GSI efforts in that their bi-allelic nature allows for highly automated, rapid genotyping (Schlotterer 2004; Van Tassell et al. 2008; Seeb et al. 2009a) with low error rates (Morin et al. 2004; Schlotterer 2004) and allows them to be easily transferred across laboratories. These traits make SNPs ideal for collaboration among agencies involved in Pacific salmon GSI (Narum et al. 2008).

An advantage of SNPs is that they may be characterized in coding regions of the genome potentially influenced by diversifying selection in addition to non-coding regions (Morin et al. 2004; Renaut et al. 2010; Seeb et al. In Press). The utility of genetic markers influenced by diversifying selection for use in GSI has become a topic of interest. By definition, allozymes are located on coding regions of the genome and have been identified as under selection in various model organisms (Hickey and McLean 1980; Rank and Dahlhoff 2002). However, during the era when allozymes were the primary marker used in Pacific salmon GSI, the allozyme loci in use were generally assumed to be neutral (Utter 1991; Utter 1994) with few exceptions (Allendorf and Seeb 2000; Aguilar 2006). Microsatellites are generally located in non-coding regions of the genome, and GSI studies using them have largely ignored the influences of diversifying selection on allele frequencies (Beacham et al. 2004; Langefors 2005; Smith et al. 2007) despite recent evidence of selection at microsatellite loci associated with migration and breeding timing in Chinook salmon (O'Malley et al. 2007) and 'hitch-hiking' selection in Atlantic cod (*Gadus morhua* L.; Nielsen et al. 2006). Recently, SNPs are being discovered in increasing numbers throughout coding regions of the genome in non-model organisms (Parchman et al. 2010), and concurrently, substantial advances are being made in the methods available to detect genetic markers influenced by diversifying selection

(Beaumont and Nichols 1996; Foll and Gaggiotti 2008; Excoffier et al. 2009). Increases in the number of SNPs available and advances in detection of loci under diversifying selection is useful for analyses of Pacific salmon because SNPs influenced by diversifying selection may provide higher resolution for resolving the stock composition of mixed fisheries than neutral markers alone (Habicht et al. 2010).

Little research has been conducted to date to quantify improvements in GSI for a given baseline by incorporating loci influenced by diversifying selection. Assuming populations throughout a region experience similar demographic histories, neutral loci should generally experience similar levels of allele frequency fluctuations among populations due to mutation and drift. Conversely, loci influenced by diversifying selection may show atypically high levels of differentiation among populations (Beaumont and Nichols 1996) and may be particularly valuable for GSI applications where neutral loci show no differences between geographically proximate populations (Habicht et al. 2010). For example, Miller et al. (2001) found that apparent heterogeneity in selection on the major histocompatibility complex (MHC) gene, known to be influenced by selection (Miller and Withler 1997), resulted in strong genetic distinction between nearby populations of sockeye salmon, in contrast to observations at neutral loci. Further, Beacham et al. (2001) noted that two SNPs located on the MHC gene provided comparable or better resolution than eight microsatellite loci in mixture analyses of British Columbia coho salmon fisheries. A similar observation was made later in Fraser River, British Columbia sockeye salmon fisheries (Beacham et al. 2004) where an MHC SNP locus was used in addition to 14 microsatellite loci. However, the improvement in stock composition estimates resulting from the incorporation of variation on the MHC gene was not quantified in each case. With rapid advances in sequencing technologies making SNPs influenced by diversifying selection more readily available, more thorough empirical studies are needed to evaluate the utility of SNPs influenced by diversifying selection in GSI.

In this study, we genotyped 45 SNP loci from 41 sockeye salmon (*Oncorhynchus nerka*) populations (n = 6,224) throughout the Copper River, Alaska and surrounding coastal areas. We used the multi-locus SNP dataset to evaluate the utility of SNPs, including those influenced by diversifying selection, for analyzing the composition of

mixtures of known origin individuals using both proportional and individual assignment. We first identified any SNP loci as  $F_{ST}$  outliers exhibiting higher population divergence than would be expected under neutral processes, a result consistent with diversifying selection. After identifying any SNP loci as candidates for diversifying selection (hereafter referred to as non-neutral SNPs), we examined the information content of individual loci for their utility in population differentiation and GSI to detect if patterns existed between the informativeness of neutral and non-neutral SNPs. We then assessed whether the addition of non-neutral SNP loci added to a neutral marker panel provided significantly higher resolution to assign individuals to population of origin than by adding an equal number of neutral loci. Finally, we quantified any improvements in the accuracy and precision of stock composition estimates using known mixtures by incorporating the SNPs identified as non-neutral. Our results indicated that non-neutral SNP loci will increase the ability to estimate the composition of Pacific salmon populations in mixed fisheries.

## **Methods**

### **Sample collections**

Spawning adult sockeye salmon were collected over multiple years (1991 – 2009) and represented by 62 field collections from throughout the Alsek, Bering, Copper, and Eyak river drainages and Prince William Sound in southcentral Alaska (Table 1.1; Figure 1.1). Tissues collected prior to 2007 were obtained from Alaska Department of Fish and Game (ADF&G) archives and had been collected by multiple agencies. Non-lethal sampling was conducted between the 2007 and 2009 field seasons with the goal of establishing a genetic baseline for use in GSI applications. Axillary processes were collected from adult sockeye salmon, placed in ethanol, and stored at room temperature for subsequent DNA extraction.

### **Laboratory analysis**

Genomic DNA extraction and SNP genotyping using both uniplex and multiplex 5'-nuclease reactions in dynamic arrays followed the methods described in Seeb et al. (2009a). All individuals were genotyped for 45 SNP loci including three mitochondrial

and 42 nuclear loci (Table 1.2). Quality control measures included reanalysis of eight out of every 95 fish (8.4%) for all loci to ensure that genotypes were reproducible, to identify laboratory errors, and to measure rates of inconsistencies during repeated analysis.

### **Pooling collections**

Collections taken between 2005 and 2009 from the same location at similar calendar days across multiple years were pooled as suggested by Waples (1990b). For any location in which multiple collections were made and one of the collections was made previous to the 2005 field season (greater than one full generation), collections were tested for homogeneity at each locus by means of a Pearson's traditional chi-square test using the program CHIFISH (Ryman 2006) and across all loci using pair-wise exact tests for genetic differentiation in GENEPOP v4.0 (Rousset 2008). Markov chain (MC) parameters for the pair-wise exact tests in GENEPOP v4.0 were as follows: dememorization = 10,000, batches = 100, iterations per batch = 5,000. Collections that failed to demonstrate significant departures from homogeneity ( $P > 0.05$ ) were pooled. The pooled and remaining unpooled collections were defined as populations in all subsequent analyses. For all populations, a minimum target sample size of 95 individuals (summed across collection years) was set to achieve acceptable precision of allele frequency estimates (Waples 1990a) and to accommodate the current genotyping platform (Fluidigm Biomark 96.96 Dynamic Array, <http://www.fluidigm.com/products/biomark-main.html>).

### **Deviation from Hardy-Weinberg and linkage equilibria**

Deviations from Hardy-Weinberg expectation (HWE) were evaluated across all loci for each population using exact P-values calculated from the MC method implemented in GENEPOP v4.0. Default parameters were used for the MC algorithm (dememorization = 1,000; batches = 20; iterations per batch = 5,000). Critical values ( $\alpha = 0.05$ ) for HWE tests were adjusted for multiple tests using a step-down sequential Bonferroni correction (Holm 1979).

Tests for linkage disequilibrium between all locus pairs were performed using simulated exact tests in GENEPOP v4.0. A pair of loci were determined to be

significantly out of linkage equilibrium if tests for linkage disequilibrium were significant in more than half of the populations at the  $\alpha = 0.05$  level. If the test was significant between a pair of loci, either the less informative of the locus pair was dropped or the two loci were combined into a composite phenotype for use in baseline evaluation. Composite phenotypes were created by combining genotypes of the two linked loci, ordering them alphabetically, and then assigning a numeric code to the resulting composite phenotype (Habicht et al. 2010). If the genotype for either of the two loci was missing, then the composite phenotype for that fish was excluded from further analysis. To determine whether dropping the less informative locus or combining loci into a composite phenotype resulted in the greatest ability to estimate the population of origin of individuals, leave-one-out tests were performed in GENECLASS2 (Piry et al. 2004) first with each locus dropped separately and then with the locus pair as a composite phenotype. All three mtDNA SNPs were combined into a composite haplotype for baseline evaluation.

### **Diversity and information statistics and detection of non-neutral loci**

The 45 SNP loci were evaluated across the 41 study populations for within-population diversity and information content. Allele frequencies and the observed heterozygosity ( $H_O$ ) and expected heterozygosity ( $H_E$ ) of each locus (excluding mtDNA SNPs) were calculated using GENALEX v6.3 (Peakall and Smouse 2006). The Weir and Cockerham (1984)  $F_{ST}$  statistic for each locus and global  $F_{ST}$  statistics were calculated using FSTAT v2.9.3.2 (Goudet 1995; Goudet 2001). The informativeness for assignment ( $I_n$ ) and optimal rate of correct assignment (ORCA) for each locus were calculated to evaluate the information content of each locus for use in the inference of population of origin of individuals. The  $I_n$  for an allele provides a measure of the potential of assignment for one allele to one population compared to that of an “average” population (a population with allele frequencies calculated using all study individuals). The ORCA gives the probability of correct assignment of an allele using a decision rule with lowest risk and enables the measurement of the contribution of specific alleles to information content. The locus-specific  $I_n$  and ORCA are simply a sum of the allele-specific  $I_n$ s and ORCAs, respectively. The  $I_n$  and ORCA summary statistics are generally highly

correlated to  $F_{ST}$ , but provide an advantage over  $F_{ST}$  in that they are designed specifically for estimating the information content of a locus for the estimation of population of origin of individuals (Rosenberg et al. 2003).

We used ARLEQUIN v3.5 (Excoffier and Lischer 2010) to identify any nuclear SNP loci as candidates for diversifying selection. ARLEQUIN v3.5 uses coalescent simulations to obtain a null distribution of  $F_{ST}$  as a function of expected heterozygosities across loci. We performed 20,000 simulations using 100 demes per group and 10 groups (equal to the number of reporting groups used for proportional assignment) under the assumptions of a hierarchical island model. Loci that fell outside the 95% quantile were determined to be candidate loci for diversifying selection. Excoffier et al. (2009) show that the hierarchical island model produces fewer false positives than the finite island model (Beaumont and Nichols 1996) when study populations exhibit a hierarchical genetic structure. The genetic structure of the populations were defined to be that of the structure defined for reporting groups used for proportional assignment (Table 1.1, see below). Loci not identified as non-neutral were assumed to be neutral. After identifying non-neutral SNP loci, we compared the  $F_{ST}$ ,  $I_n$ , and ORCA of neutral and non-neutral SNP loci to see if patterns existed between the informativeness of the two marker types to discriminate populations ( $F_{ST}$ ) and to aid in the estimation of population membership of individuals ( $I_n$  and ORCA).

### **Reporting groups for proportional and individual assignment tests**

For assignment tests, ten reporting groups were established for proportional assignment and six reporting groups were established for individual assignment. Reporting groups were established based on the geographic structure of the study region (Figure 1.1) and results from an unrooted neighbor-joining (N-J) dendrogram (Figure 1.2). The N-J dendrogram was based on pairwise Cavalli-Sforza and Edwards (1967) genetic distances and bootstrap replicates were calculated between all populations using POPULATIONS (Langella 2001). Consistency of the radial tree topology was estimated using 1,000 bootstrap replicates and the resulting consensus dendrogram was constructed using TREEVIEW (Page 1996). Reporting groups for proportional assignment include: Slana River (SLANA), Tanada River (TNADA), Gulkana River (GLKNA), Tazlina

River (TZLNA), Klutina Lake tributaries (KLTNA), the Klutina Lake and Tonsina Lake outlet populations (KLTON), Chitina River (CHTNA), a group that includes the lower Copper River, Copper River delta, Bering River and Eyak River (DELTA), Prince William Sound (PWS), and the Alsek River (ALSEK).

For individual assignment tests, the TZLNA, KLTNA, and KLTON groups were pooled into a single reporting group and the DELTA, PWS, and ALSEK reporting groups were pooled into a single reporting group. The TZLNA, KLTNA, and KLTON reporting groups all consist of populations that spawn in glacially fed, turbid drainages in the middle Copper River each with a large (14 – 155 km<sup>2</sup>) nursery lake near the headwaters. The DELTA, PWS, and ALSEK reporting groups represent a large “coastal” group. Within reporting groups, population genotypes were maintained separately (i.e. not pooled to recalculate allele frequencies) as recommended by Wood et al. (1987).

### **Improved resolution with non-neutral loci**

The simulation-based approach of Rosenberg (2005) for evaluating the informativeness of marker panels was used to evaluate the resolution added to a neutral marker panel when non-neutral SNPs are added. The goal was to determine whether adding a given number of non-neutral SNPs to a neutral SNP marker panel significantly improved the assignment of individuals to population of origin than by adding an equal number of neutral SNPs. The approach calculates a  $f_{ORCA}$  score (the probability that an individual is assigned to its correct population of origin) for any given set of loci based on population-specific allele-frequencies. After determining which SNP loci were non-neutral, we sampled (with replacement) from the pool of neutral loci to produce 500 marker panels containing exclusively neutral loci. For each marker panel, we simulated 1,000 individuals, each generated using the allele frequencies from one of the baseline populations sampled at random. We then computed which population most likely generated the simulated genotypes of each individual and tabulated how many of the 1,000 individuals correctly assigned back to its correct individual assignment reporting group. The  $f_{ORCA}$  score for each of the 500 sampled marker panels was calculated to obtain a distribution of  $f_{ORCA}$  scores. Finally, we evaluated how the addition of the non-neutral SNP loci to each neutral marker panel improved the  $f_{ORCA}$  for each of the 500

panels versus adding an equal number of neutral loci to the same panel. We report the proportion of individuals that correctly assigned back ( $f_{ORCA}$ ) to the correct reporting group for each marker panel (1. neutral loci panel; 2. neutral loci panel plus additional neutral loci; 3. neutral loci panel plus equal number of non-neutral loci). The simulation procedure of Rosenberg (2005) was programmed and run using R v2.9.2 (R Development Core Team, 2009). Because the process of simulating each individual's genotypes assumes Hardy-Weinberg expectation, the mtDNA composite haplotype was excluded from this analysis. In the case that two neutral loci deviated from linkage equilibrium, the less informative of the two loci was dropped from the analysis. In the case that two non-neutral loci deviated from linkage equilibrium, one of the two loci was sampled randomly for each of the 500 iterations to avoid introducing bias in the case that one of the loci was more informative for individual assignment.

### **Baseline evaluation for mixed stock analysis**

Two marker panels were produced to test the ability to conduct GSI using proportional and individual assignment for populations throughout the study region. The first marker panel, hereafter referred to as the 'neutral' marker panel, excludes non-neutral loci. The second, hereafter referred to as the 'full' marker panel, contains all loci, including those determined to be non-neutral. The three mtDNA SNPs, as a composite haplotype, were included in both the neutral and the full marker panel for baseline evaluation.

We performed 100% simulations as implemented in the program SPAM v3.7b (updated version of Debevec et al. 2000). An analysis was run for each of the 10 reporting groups in which baseline and mixture genotypes were randomly generated from the baseline allele frequencies using parametric bootstrapping with replacement (PB-R) and assuming Hardy-Weinberg expectations. Each mixture ( $n = 300$ ) contained 100% of simulated individuals generated from the reporting group of concern, with each population from the reporting group contributing equally to the mixture. Average estimates of mixture proportions were derived from 1,000 simulations. Individual population estimates were first calculated and then summed into reporting groups using a maximum likelihood model (allocate-sum procedure; Wood et al. 1987). Mean estimates

of mixture proportions and ninety-percent confidence intervals for each reporting group were computed using 1,000 bootstrap iterations and the non-symmetric percentile method. A critical level of 90% mean correct allocation was used to determine if each reporting group was acceptably identifiable (e.g., Seeb et al. 2000).

The strength of the baseline was further tested for proportional assignment using both the neutral and full marker panels by estimating the composition of two known mixtures using the program BAYES (Pella and Masuda 2001). The known mixtures were built by selecting 300 individual fish from the baseline and placing the individuals into a mixture. For each mixture, the baseline was reconstructed with the 300 individuals removed. The first known mixture contained an equal number of fish removed from each reporting group (i.e. 30 fish per reporting group). The second known mixture was constructed to be representative of what might be expected in fisheries in the region (i.e. the Copper River district commercial gillnet fishery). The distribution of this known mixture was constructed using total escapement data from the 2005 to 2007 seasons (Fair et al. 2008; Wade et al. 2008). The composition of each mixture was estimated using a Bayesian model implemented in BAYES. The Bayesian model places a Dirichlet distribution as the “prior” distribution for reporting group proportions, and the parameters for the distribution must be specified. Prior parameters for each reporting group were divided equally among reporting groups, with populations weighted equally within each reporting group. We set the sum of the starting value of reporting group proportions (prior weight) for each Markov Chain Monte Carlo (MCMC) chain to be 1, which is equivalent to adding 1 fish to each mixture (Pella and Masuda 2001). Three independent MCMC chains of 50,000 iterations each with different starting values were run with a burn-in period of 40,000 iterations to remove the influence of the initial starting values. Composition estimates and 95% credibility intervals were calculated from the final 10,000 iterations of the three MCMC chains. The credibility intervals are a direct statement of probability (i.e., a 95% credibility interval has a 95% chance of containing the true answer; Gelman et al. 2000). The adequacy of the burn-in period for each chain was examined using Rafferty and Lewis’ (1996) diagnostic. To ensure that the BAYES output was an acceptable approximation of the stationary posterior distribution and that the composition estimates were valid, the convergence among MCMC chains was

assessed using the Gelman-Rubin shrink factors that are computed for all reporting groups. The shrink factor compares the variation within a chain to the total variation among chains (Gelman and Rubin 1992). A shrink factor less than 1.2 indicates convergence among chains.

To evaluate how including non-neutral loci improved the resolution of the baseline for individual assignment, we conducted leave-one-out tests in GENECLASS2 using both the neutral and the full marker panels. For leave-one-out tests, each individual was removed from the baseline and then assigned to reference population of origin using the Bayesian allele frequency estimation method (Rannala and Mountain 1997). Each individual  $i$  was then given an assignment score for each population  $l$  of  $k$  total populations

$$score_{i,l} = \frac{L_{i,j}}{\sum_{j=1}^k L_{i,j}}$$

where  $L_{i,l}$  is the likelihood that individual  $i$  belongs to population  $l$  (Piry et al. 2004). The population with the highest assignment score for each individual was considered the assigned population. We report the number of individuals that assigned back to the correct population and to the correct reporting group of origin using both the neutral and the full marker panel. When assigning individuals back to reporting group, we simply summed the  $score_{i,l}$  of each individual across all of the populations within each reporting group. We chose to use the GENECLASS2 program because of its ability to incorporate composite phenotypes and composite haplotypes and its ability to perform leave-one-out tests.

## Results

### Sample collections

A total of 6,224 sockeye salmon from 62 field collections conducted throughout the Copper River and adjacent coastal drainages in southcentral Alaska were genotyped for 45 SNPs (Table 1.2). Eleven collections were made between 1991 and 1992. The remaining 51 collections were made between 2005 and 2009.

### **Laboratory analysis**

The overall genotyping failure rate was 1.15% across all individuals for the 45 SNP loci. The discrepancy rate between the original genotyping run and the quality control genotyping run was 0.05%. Assuming an equal error rate in both the original genotyping process and the quality control process, individuals were genotyped with an overall error rate of 0.025%. Due to low discrepancy rates, the original genotype was retained in the case of a discrepancy between the original genotyping run and the quality control genotyping run.

### **Pooling collections**

The 62 collections were distributed among 41 geographic locations. Twenty-three locations were sampled in a single year; the remaining 18 locations were sampled in multiple years. Of the 18 locations sampled in multiple years, eight had at least one collection made prior to 2005 (greater than one generation). Seven of those spanned a period of at least 16 years (1991 to 2007; approximately four generation). For all eight locations, temporal collections failed to demonstrate significant departures from homogeneity ( $\alpha = 0.05$ ), resulting in a total of 41 populations that were used in all subsequent analyses (Table 1.1). We achieved the target sample size of 95 individuals for 35 of the 41 populations (average sample size = 152).

### **Deviation from Hardy-Weinberg and linkage equilibria**

There were no significant deviations from HWE out of 1,514 tests after correction for multiple tests (across loci and populations). Before correcting for multiple tests, 72 of the 1,514 tests were significant (76 would be expected by chance at  $\alpha = 0.05$ ). Of these, the only notable pattern occurred in the Klutina Lake inlet (#14) population, with six loci deviating from HWE. No other population had more than four loci deviating from HWE.

Significant linkage disequilibrium was found between two pairs of nuclear SNP loci in more than half of the populations. Tests for the first pair (*One\_GPDH-201* and *One\_GPDH2-187*) were significant in 38 of the 41 populations. Tests for the second pair (*One\_MHC2\_190* and *One\_MHC2\_251*) were significant in 25 of the 41 populations. Leave-one-out tests indicated that combining *One\_MHC2\_190* and *One\_MHC2\_251* into

a composite phenotype and dropping *One\_GPDH2-187* provided the greatest ability to assign individuals to population of origin. *One\_MHC2\_190* and *One\_MHC2\_251* were combined into a composite phenotype for baseline evaluation and *One\_GPDH2-187* was dropped from all remaining analyses.

### **Diversity and information statistics and detection of non-neutral loci**

All loci were polymorphic ( $\geq 0.05$  minor allele frequency) in at least six of the 41 populations. For heterozygosity across populations by locus, a minimum  $H_E$  of 0.012 was observed for *One\_RH2op-395*, and a maximum  $H_E$  of 0.461 was observed for *One\_Prl2* (Table 1.2). For average heterozygosity across loci by population, a minimum  $H_E$  of 0.155 was observed in the Fish Creek population (#8), and a maximum  $H_E$  of 0.303 was observed in the Miners Lake population (#35; Table 1.1). The average  $F_{ST}$  across all loci was 0.098 with a minimum of 0.017 at *One\_RAG1-103* and a maximum of 0.297 at *One\_MHC2\_251*.

Four of the 42 nuclear SNP loci (*One\_GHII-2165*, *One\_MHC2\_190*, *One\_MHC2\_251*, *One\_Zp3b-49*) fell outside the 95% quantile based on coalescent simulations and were determined to be candidates for diversifying selection (Table 1.2; Figure 1.3). The average  $F_{ST}$  across neutral loci was 0.089, and the average  $F_{ST}$  across non-neutral loci was 0.262. Using the neutral SNP loci (mtDNA SNPs as composite haplotype), the global  $F_{ST}$  for the 41 populations was 0.090 (95% CI: 0.079, 0.101). After adding the non-neutral loci (MHC loci as composite phenotype) to the neutral loci, the global  $F_{ST}$  for all 41 populations increased to 0.113 (95% CI: 0.088, 0.139). According to  $I_n$ , the non-neutral loci were the four highest-ranked loci for the amount of information provided for assignment of alleles to one population compared to that of the “average” population. The average  $I_n$  for neutral loci was 0.019 and the average  $I_n$  for non-neutral loci was 0.064. Similarly, *One\_GHII-2165*, *One\_MHC2\_190*, and *One\_MHC2\_251* were the three highest ranked loci and *One\_Zp3b-49* was the eight-ranked locus according to ORCA. The average ORCA across neutral loci was 0.035 and the average ORCA across non-neutral loci was 0.046 (Table 1.2).

### **Reporting groups for proportional and individual assignment tests**

All reporting groups, with the exception of KLTON, were defined primarily by the geographic structure of the study region. KLTON was determined according to similarities in life-history types of the two populations (#17 Klutina Lake outlet and #18 Tonsina Lake outlet). The relationship of these two populations was depicted in the N-J dendrogram (Figure 1.2). Each of these populations spawns in the mainstem river in and below the outlets of their respective lakes. Besides KLTON, the genetic structure of sockeye salmon populations throughout the Copper River and adjacent coastal drainages corresponded closely with the geographic structure of the region (i.e., populations in the same reporting group were genetically more similar than populations in different reporting groups) with few exceptions (Figure 1.2). For example, the Long Lake (#20) population was highly divergent and did not cluster closely with the other two populations from the CHTNA reporting group (#19 Bear Island and #21 Tebay River). Similarly, the Steamboat Lake (#23) population clustered more closely with populations within PWS than DELTA although Steamboat Lake is located in the Bremner River within the latter reporting group. Populations within the ALSEK reporting group exhibited an interesting pattern of geographic structure. The two populations located in the upper Alsek River (#39 Upper Tatshenshini River and #40 Klukshu River – late weir) were more similar to populations in the middle Copper River (KLTNA and KLTON reporting groups) than to the East Alsek River (#41) population located in the lower Alsek River drainage. The East Alsek River (#41) population appeared to be more similar to populations from the lower Copper River and adjacent coastal drainages. The high levels of divergence among sockeye salmon populations throughout the study region allowed some flexibility in the formation of reporting groups to allow reporting groups to follow more closely the geographic structure of the region.

### **Improved resolution with non-neutral loci**

Four of the 42 nuclear DNA SNPs were determined to be candidates for diversifying selection, leaving 38 neutral nuclear DNA SNPs. However, because the procedure of Rosenberg (2005) assumes independence of loci, *One\_GPDH2-187* was dropped from this analysis, leaving a pool of 37 neutral loci to sample from.

Additionally, one of the two MHC loci were selected randomly (the MHC pair was in linkage disequilibrium) to add to each of the 500 randomly generated marker panels so that three non-neutral loci were added to each panel (*One\_GHII-2165*, *One\_Zp3b-49*, and one of the MHC loci sampled at random). In total, 500 sets of three marker panels were generated. The first consisted of 34 neutral loci sampled at random (with replacement) from the pool of 37 neutral loci. The second consisted of the same 34 neutral loci, plus an additional three neutral loci sampled at random (with replacement) from the same pool of 37 neutral loci. The third consisted of the same 34 neutral loci, plus three non-neutral loci. Each simulated individual was generated using the three marker panels and allele frequency data from one of the 41 reference populations chosen at random.

The addition of three non-neutral SNP loci provided significantly greater (p-value < 0.001) resolution to assign individuals to population of origin than adding three neutral loci to a panel of 34 neutral loci (Figure 1.4). The average  $f_{ORCA}$  score for the 34 neutral loci marker panels was 0.639 (SD = 0.035). The average  $f_{ORCA}$  score for the 34 neutral loci plus an additional three neutral loci was 0.649 (SD = 0.039), a 1.6% increase over 34 neutral loci. The average  $f_{ORCA}$  score for the 34 neutral loci plus an additional three non-neutral loci was 0.750 (SD = 0.027), a 17.4% increase over 34 neutral loci. This translates to approximately a 10% increase in the ability to assign individuals to population of origin by adding three non-neutral loci than by adding three neutral loci based upon allele frequencies from the 41 reference populations.

### **Baseline evaluation for mixed stock analysis**

Two marker panels ('neutral' and 'full') were evaluated for their use in proportional and individual assignment applications throughout the study region in southcentral Alaska. The neutral marker panel contained 38 loci (37 neutral nuclear DNA loci; three mtDNA SNPs as a composite haplotype). The full marker panel contained 41 loci (neutral marker panel plus four non-neutral loci; MHC loci as composite phenotype). Proof tests using SPAM v3.7b indicated that the 10 reporting groups can be adequately assigned (>90% mean correct assignment for 100% simulations; e.g. Seeb et al. 2007) using both the neutral and the full marker panels (Figure 1.5). Using the neutral marker panel, the mean allocation of simulated mixtures to the correct reporting group was

96.4% (average SD = 0.034). With the full marker panel, the mean allocation of simulated mixtures to the correct reporting group was 97.9% (average SD = 0.012). The minimum lower 90% confidence interval of the estimates for a reporting group using the neutral marker panel was 88.1% (CHTNA) and using the full marker panel was 92.5% (ALSEK). On average, the inclusion of the non-neutral loci decreased misallocation of mixtures to incorrect reporting groups by 40.5% (3.6% for the neutral marker panel; 2.1% for the full marker panel).

Stock composition estimates for each of the two known mixtures using the full marker panel were generally closer (21 of 30 estimates) to the expected proportions, and the average standard deviation across reporting groups decreased for both known mixtures with the inclusion of the non-neutral loci (Figure 1.6a-b). Using the neutral marker panel, the average standard deviation for each of the two known mixtures was 3.08% and 2.38%, respectively. Using the full marker panel, the average standard deviation for each of the two known mixtures was 2.45% and 2.26%, respectively. In only one incident (Figure 1.6b) did the expected proportion fall outside of the 95% credibility interval for a reporting group (CHTNA, neutral marker panel).

Leave-one-out tests showed that the number of individuals assigned to the incorrect reporting group was reduced by 33.6% after incorporating the non-neutral SNP loci. Using the neutral marker panel, 2,670 of 6,224 individuals (42.9%) assigned to the correct population, and 4,680 individuals (75.2%) assigned to the correct reporting group. With the full marker panel, 3,100 individuals (49.8%) assigned to the correct population, and 5,198 individuals (83.5%) assigned to the correct reporting group. The full marker panel resulted in an increase in the correct assignment of individuals of 430 (6.9%) to population and 518 (8.3%) to reporting group. Results of individual assignment tests were displayed as a matrix of population assignment percentage (Figure 1.7a-b) with boxes delineating reporting groups and self-assignment to reference populations lying on the diagonal. Mis-allocation using the full marker panel created a pattern of blocks that corresponded more closely to the reporting groups than when using the neutral marker panel. Examining patterns of self-assignment, populations from the upper- and middle-Copper River drainage consistently had higher self-assignment success than populations from the lower Copper River and adjacent coastal drainages.

## **Discussion**

### **Genetic diversity and population structure**

In general, sockeye salmon populations from the Copper River and adjacent coastal drainages exhibited a pattern of genetic isolation-by-distance. That is, along a continuum, populations were genetically more similar to nearby populations than to more distant populations. Further, populations nearer to the coast (i.e. with a shorter upstream migration distance) generally exhibited greater within-population genetic diversity and lesser among-population diversity than populations spawning further inland. However, interesting exceptions to this general trend demonstrated the influences of historical demographic processes.

The two populations located the upper Alsek River drainage (#39 Upper Tatshenshini River and #40 Klukshu River – late weir) were more similar to populations from the Klutina and Tonsina river drainages than to the East Alsek River population (#41) that spawns in the lower Alsek River (Figure 1.2). When evaluating the baseline for individual and proportional assignment, exchange between the two upper Alsek River populations and the KTLNA and KLTON reporting groups represented one of the largest sources of misallocation during GSI for the given baseline. Similarities between the two groups may be explained by historical founding processes. Upper portions of the Copper and Alsek Rivers may have once been part of a northern glacial refuge (Beringia) for Pacific salmon in the region (Lindsey and McPhail 1986; Smith et al. 2001). During the Pleistocene Epoch, large portions of northwestern North America were repeatedly buried in ice. During the height of the most recent of the glaciations, known as the McConnell/McCauley Glaciation throughout the Yukon Territory and Alaska, ice-free areas persisted north of the glaciated regions near the present day upper Yukon, Copper, and Alsek River drainages (Smith et al. 2001). Although it is unclear whether Pacific salmon occupied the region during the McConnell/McCauley Glaciations, Smith et al. (2001) provided evidence that some salmon populations may have persisted in the region in small numbers. The genetic similarity between the two upper Alsek River populations and populations from the Klutina and Tonsina river drainages provides supporting evidence that sockeye salmon populations in the upper Alsek River and the upper- to

middle-Copper River may have been founded by individuals from the northern glacial refuge, as opposed to populations that persisted in ice-free areas along the glaciated coast (Wood et al. 1994; Carney et al. 1997).

Within the Klutina and Tonsina river drainages, another interesting exception to the general isolation-by-distance trend occurred. The two populations that spawn in and below the outlets of their respective lakes, the Klutina Lake outlet (#17) and Tonsina Lake outlet (#18) populations, are both likely of the ‘sea/river ecotype’ (versus ‘lake ecotype’) as defined by Wood et al. (2008). The lake ecotype is the typical anadromous form of sockeye salmon that spends about half its life rearing in freshwater lakes before migrating to sea (Burgner 1991) whereas the sea/river ecotype is also anadromous, but rears in freshwater for a much shorter duration and typically inhabits side channels and backwater habitats during the freshwater stage (Gilbert 1913). The Klutina Lake outlet population is genetically more similar to the Tonsina Lake outlet population (Figure 1.2) than it is to the five populations (#12-16) that spawn in tributaries above Klutina Lake, each of which are likely of the lake ecotype. The ‘recurrent evolution’ hypothesis put forth by Wood (1995) claims that the sea/river ecotype is the ancestral ‘straying’ form and typically exhibits greater within-population diversity and smaller among-population differences. Both of the ‘sea/river’ ecotype populations (Klutina and Tonsina Lake outlet populations) exhibit greater heterozygosity (Table 1.1) and allelic richness (data not shown) and weaker population structure than the Klutina Lake populations, supportive of the recurrent evolution hypothesis. These findings suggest that: 1) the Klutina Lake outlet and Tonsina Lake outlet populations diverged more recently than either of the populations diverged from nearby lake ecotype populations, 2) greater contemporary straying occurs between the two sea/river ecotype populations than occurs between the Klutina Lake sea/river and Klutina Lake lake ecotypes, or 3) a combination of the two.

### **Use of 100% simulations for baseline evaluation**

We used the PB-R method as implemented in SPAM v3.7b for simulating mixtures when evaluating the resolution of the baseline for proportional assignment. The PB-R method uses allele frequencies from reference population samples to simulate mixtures of individuals and to create a new “re-sampled” baseline (Anderson et al. 2008).

Anderson et al. (2008) show that this method often overestimates the expected accuracy of proportional assignment because it does not account for sampling error in baseline allele frequencies. Allele frequencies from baseline samples will exhibit higher divergence between the two samples, on average, than the true allele frequency divergence between those populations. Anderson et al. (2008) show that if this sampling error is not accounted for, the observed pairwise  $F_{ST}$  between two samples taken from a pair of populations will be inflated by an approximate magnitude of  $\frac{1}{2n}$  (Wright 1978; Chakraborty and Leimar 1987) where  $n$  is the number of individuals sampled. This inflation term will be most pronounced in cases of small sample sizes and low levels of population divergence. Because of the high levels of population divergence seen in the study populations and the large sample sizes, inflation of  $F_{ST}$  due to sampling error should be fairly negligible. If we assume a sample size of 95 individuals per sample (achieved for 35 of the 41 baseline populations) and a parametric pairwise  $F_{ST}$  of 0.1 (0.118 was the average pairwise  $F_{ST}$  observed), the estimated  $F_{ST}$  that would be used to simulate individuals in the mixture would be 0.103, an inflation of only 3% when compared with the true pairwise  $F_{ST}$ . Anderson et al. (2008) further acknowledge that the PB-R method's bias is much less pronounced when assignment of mixtures is made to aggregate reporting groups as was done here as opposed to individual populations. We feel that because of the large sample sizes, high levels of genetic differentiation among populations, and assignment of simulated individuals to reporting region, the use of the PB-R method as implemented in SPAM v3.7b is warranted.

### **Linked loci in GSI studies**

Across the study region, two pairs of SNP loci were significant for linkage disequilibrium in more than half of the populations, the first pair being *One\_GPDH-201* and *One\_GPDH2-187*, and the second pair being *One\_MHC2\_190* and *One\_MHC2\_251*. How to handle pairs of linked loci for use in GSI is an issue that is often met with an arbitrary solution. Habicht et al. (2010) outlined the linkage scenarios that can be encountered. If linkage disequilibrium between a pair of markers only appears in a minority of the populations, then including both loci in GSI applications should not inflate the ability to assign individuals or groups of individuals to an inappropriate level.

Conversely, if linkage disequilibrium appears in the majority of the populations, then dropping one locus from the analysis should not result in an inappropriate loss of power for assignment. However, if the disequilibrium appears in the majority of the populations and the linkage phase between the two loci varies among populations, then dropping a locus might result in an inappropriate loss of power for assignment. The goal is to preserve the information due to divergent phasing among populations while avoiding an overstatement of power due to pseudo-replication (Hurlbert 1984).

For the study region, we determine that dropping *One\_GPDH2-187* and combining *One\_MHC2\_190* and *One\_MHC2\_251* into a composite phenotype provided the greatest ability to assign individuals to population of origin (data not shown). *One\_MHC2\_190* and *One\_MHC2\_251* show fairly sporadic patterns of linkage phase throughout the study populations (Figure 1.8a), and combining the locus pair into a composite phenotype provided a marked improvement in the ability to assign individuals to population and to reporting group of origin. By combining the MHC loci into a composite phenotype, we preserved the information held in the variable pattern of linkage phase observed among the study populations. The results observed for *One\_GPDH-201* and *One\_GPDH2-187* were not as clear. Dropping *One\_GPDH2-187* provided the greatest ability to assign individuals to population of origin. However, when we evaluated the assignment of individuals to reporting group of origin, we observed that combining *One\_GPDH-201* and *One\_GPDH2-187* into a composite phenotype slightly improved the ability to assign individuals to reporting group of origin, but decreased the ability to assign individuals to population of origin. It appears that there is some information contained in the linkage phase of the GPDH markers throughout the study populations (Figure 1.8b), but the linkage phase appears to be more stable than that observed at the MHC loci. While our decision for how to treat each of the locus pairs exhibiting linkage disequilibrium was determined with an arbitrary solution (leave-one-out method), results demonstrate that examining patterns of linkage disequilibrium between locus pairs among reference populations may provide a good first step in evaluating the information content of linked loci for GSI.

### **Discovery of SNP loci under diversifying selection**

Four of the 42 nuclear DNA SNP loci were determined to be candidates for diversifying selection. *One\_GHII-2165* is located on the growth hormone gene responsible for the production of growth hormones essential to somatic growth in vertebrates (Yowe and Epping 1995). Variation on the growth hormone gene may be influenced by selection in a range of teleost fishes (Clements et al. 2004). *One\_MHC2\_190* and *One\_MHC2\_251* are located on genes responsible for the production of major histocompatibility complex (MHC) class II molecules associated with the immune system response in vertebrates (Bernatchez and Landry 2003). Several studies support the adaptive nature of MHC markers associated with pathogen mediated selection in salmon (Miller et al. 2001; Dionne et al. 2009; Evans and Neff 2009). *One\_Zp3b-49* is located within the zona pellucida gene family associated with sperm recognition in the chorion surrounding the oocyte in fishes (Hyllner et al. 2001; Sun et al. 2010).

Although only 9.5% of the nuclear DNA SNPs in this study were determined to be candidates for diversifying selection, next-generation sequencing technologies provide promise for identifying non-neutral SNPs at increasing rates in the near future. Advances in next-generation sequencing technologies have allowed the discovery of large numbers of SNPs at increasing rates in species where little genomic information has previously been available (Renaut et al. 2010). Further, these technologies are being applied increasingly to transcribed regions of the genome (Barbazuk et al. 2007; Sanchez et al. 2009; Seeb et al. In Press), allowing researchers to target the discovery of SNPs influenced by selection. With increasing numbers of non-neutral SNPs likely becoming available for Pacific salmon in the near future, it is essential to understand and recognize how these markers can be used in GSI. Non-neutral SNPs frequently have higher levels of allele-frequency divergence than markers influenced exclusively by stochastic and demographic processes. Rosenberg et al. (2003) showed that  $F_{ST}$  is highly correlated with measures of marker information content ( $I_n$  and ORCA). Thus, non-neutral SNPs will be highly informative for inference of population and reporting group of origin for fisheries mixtures. Incorporating non-neutral SNPs into marker panels to be used in GSI may allow researchers to answer a broader spectrum of research and management questions

with a fewer number of SNPs and reduced genotyping costs. As more non-neutral SNPs become available in the near future, more empirical studies will be needed to evaluate the utility of non-neutral SNPs to benefit GSI applications for Pacific salmon.

A potential disadvantage of non-neutral SNPs is that they potentially have more unstable allele frequencies than neutral markers. This may require more frequent sampling and genotyping of reference populations, increasing field and laboratory time and costs. In this study, we did not observe any issues with temporal changes in allele frequencies of the four non-neutral SNPs in the seven populations where temporal collections spanned a period of at least 16 years. Although it is unclear as to how many generations it would take on average for non-neutral SNP loci to demonstrate temporal allele frequency differences as opposed to neutral markers, the span is likely long enough that alternative concerns (i.e. demographic processes) would concurrently initiate re-sampling of reference populations.

## **Conclusions**

We demonstrate here that non-neutral SNP markers are highly effective in Pacific salmon GSI. The addition of the SNP loci identified as candidates for diversifying selection to a neutral marker panel provided significantly greater resolution to assign individuals to population of origin than adding an equal number of neutral loci. Further, incorporating non-neutral SNPs provided increased accuracy and precision of composition estimates in proportional assignment tests. To reduce laboratory time and genotyping costs, it is often desirable to reduce the number of markers needed to infer the origin of individuals or mixtures of individuals (Paschou et al. 2007). Incorporating non-neutral SNPs into Pacific salmon GSI may allow researchers to answer a broad spectrum of research and management questions with a reduced number of markers. Our results indicated that SNPs influenced by diversifying selection will increase the ability to identify straying by individual fish and estimate the composition of Pacific salmon populations in mixed fisheries.

## **Acknowledgements**

This project was a collaborative effort between the University of Washington, School of Aquatic and Fishery Sciences (UW SAFS) and the Alaska Department of Fish and Game (ADFG), Gene Conservation Laboratory. Many thanks to Chris Habicht, Tom Quinn, Daniel Schindler, Jim Seeb, and Lisa Seeb for insightful discussions regarding this chapter. Thanks to Eric Grau, Jim Jasper, and Bill Templin for their quantitative expertise and assistance with R-programming. Reviews from Lisa Creelman, Daniel Gomez-Uchida, Molly McGlaulin and Matt Smith greatly improved earlier versions of this manuscript. We are grateful for laboratory contributions from staff at UW SAFS and ADFG. Thanks to the US Fish and Wildlife Service staff in Copper Center, AK for assistance with field collections from Tanada and Long Lakes. This report was partially funded by the Alaska Sustainable Salmon Fund under Study #45869 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, administered by the Alaska Department of Fish and Game. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, the U.S. Department of Commerce, or the Alaska Department of Fish and Game. Additional project funding was provided by the Gordon and Betty Moore Foundation and the State of Alaska. Funding for MWA was provided by the H. Mason Keeler Endowment for Excellence (UW).

Table 1.1. Reporting group, years sampled, sample size ( $n$ ), observed and expected heterozygosities, and self-assignment success for sockeye salmon populations surveyed for 45 SNPs from the Copper River and adjacent coastal drainages in southcentral Alaska. Genetic diversity estimates and self-assignment success (GENECLASS; Piry et al. 2004) for each population are given using both the neutral marker panel (37 nuclear neutral DNA loci; three mtDNA loci as composite haplotype for self-assignment success) and the full marker panel (neutral marker panel plus non-neutral loci; *One\_MHC2\_190* dropped for diversity estimates; MHC loci as composite phenotype for self-assignment success). Proportional assignment reporting groups are as follows: Slana R. (SLANA), Tanada Lk. (TNADA), Gulkana R. (GLKNA), Tazlina R. (TZLNA), Klutina Lk. tributaries (KLTNA), Klutina and Tonsina Lk. outlets (KLTON), Chitina R. (CHTNA), Lower Copper R., Copper R. Delta, Bering Lk., Eyak Lk. (DELTA), Prince William Sound (PWS), and Alesk R. (ALSEK). Site name and reporting group abbreviations include ‘Cr.’ for creeks, ‘Lk.’ for lakes, and ‘R.’ for rivers.

Map #	Population	Reporting Group	Years Sampled	$n$	Observed Heterozygosity		Expected Heterozygosity		Self Assignment Success (%)	
					Full	Neutral	Full	Neutral	Full	Neutral
1	Bad Crossing #2	SLANA	2009	114	0.205	0.199	0.213	0.208	0.561	0.404
2	Bone Cr.	SLANA	2005, 2008	56	0.225	0.221	0.229	0.226	0.464	0.232
3	Mentasta Lk.	SLANA	2008	197	0.234	0.236	0.233	0.235	0.721	0.614
4	Tanada Lk. - outlet	TNADA	2009	95	0.240	0.235	0.251	0.245	0.779	0.642
5	Tanada Lk. - east shore	TNADA	2009	95	0.248	0.237	0.254	0.245	0.495	0.526
6	Keg Cr.	GLKNA	2009	66	0.231	0.223	0.231	0.221	0.742	0.742
7	Swede Lk.	GLKNA	2008	200	0.208	0.195	0.207	0.195	0.885	0.915
8	Fish Cr.	GLKNA	2008	169	0.151	0.147	0.155	0.150	0.953	0.947
9	EF Gulkana R. - early	GLKNA	2008, 2009	152	0.244	0.238	0.242	0.235	0.526	0.461
10	EF Gulkana R. - late	GLKNA	2008, 2009	190	0.244	0.230	0.241	0.229	0.605	0.568
11	Mendeltna Cr.	TZLNA	2008, 2009	189	0.214	0.204	0.215	0.204	0.852	0.847
12	1884 Lk.	KLTNA	2008	82	0.254	0.248	0.253	0.246	0.280	0.232
13	Bear Hole	KLTNA	2008	143	0.254	0.254	0.253	0.252	0.552	0.420
14	Klutina Lk. - inlet area	KLTNA	2008, 2009	100	0.268	0.256	0.268	0.260	0.310	0.280
15	St. Anne Cr.	KLTNA	2005, 2008	299	0.256	0.253	0.258	0.255	0.485	0.344
16	Mahlo R.	KLTNA	2008	192	0.253	0.249	0.250	0.245	0.485	0.240

Table 1.1. Continued.

Map #	Population	Reporting Group	Years Sampled	<i>n</i>	Observed Heterozygosity		Expected Heterozygosity		Self Assignment Success (%)	
					Full	Neutral	Full	Neutral	Full	Neutral
17	Klutina Lk. - outlet	KLTON	2008, 2009	159	0.276	0.263	0.273	0.263	0.799	0.491
18	Tonsina Lk. - outlet	KLTON	2009	154	0.272	0.262	0.279	0.265	0.247	0.175
19	Bear Island	CHTNA	2009	30	0.283	0.264	0.278	0.265	0.567	0.267
20	Long Lk.	CHTNA	2005	189	0.235	0.226	0.231	0.221	0.958	0.947
21	Tebay R.	CHTNA	2008	197	0.283	0.268	0.280	0.265	0.376	0.274
22	Salmon Cr. (Bremner R.)	DELTA	2008	99	0.279	0.270	0.275	0.267	0.253	0.162
23	Steamboat Lk.	DELTA	2008	177	0.275	0.260	0.275	0.262	0.435	0.356
24	Clear Cr. at 40-mile	DELTA	2007, 2008	188	0.276	0.269	0.270	0.263	0.154	0.074
25	McKinley Lk.	DELTA	2007	94	0.256	0.246	0.249	0.240	0.489	0.447
26	Salmon Cr. (McKinley Lk.)	DELTA	1991, 2007	184	0.265	0.257	0.263	0.255	0.261	0.196
27	Tokun Lk.	DELTA	2008, 2009	190	0.250	0.243	0.245	0.238	0.779	0.679
28	Martin Lk.	DELTA	1991, 2007, 2008	286	0.268	0.260	0.261	0.253	0.427	0.416
29	Martin R. slough	DELTA	2008	95	0.286	0.278	0.281	0.270	0.242	0.242
30	Eyak Lk. - Hatchery Cr.	DELTA	1991	38	0.265	0.254	0.254	0.242	0.263	0.079
31	Eyak Lk. - middle arm	DELTA	1991, 2007, 2008	184	0.279	0.270	0.280	0.271	0.168	0.136
32	Eyak Lk. - south beaches	DELTA	1991, 2007, 2008	189	0.283	0.273	0.281	0.271	0.079	0.106
33	Bering Lk.	DELTA	1991, 2009	190	0.264	0.254	0.258	0.251	0.279	0.184
34	Kushtaka Lk.	DELTA	2007, 2008	190	0.251	0.244	0.262	0.251	0.295	0.353
35	Miners Lk.	PWS	1991, 2009	191	0.295	0.286	0.303	0.290	0.471	0.335
36	Coghill Lk.	PWS	1991, 1992	288	0.296	0.285	0.298	0.289	0.253	0.177
37	Main Bay	PWS	1991	96	0.294	0.284	0.288	0.276	0.208	0.156
38	Eshamy Lk.	PWS	1991, 2008	191	0.216	0.217	0.224	0.224	0.969	0.948
39	Upper Tatshenshini R.	ALSEK	2003	95	0.260	0.250	0.261	0.250	0.305	0.274
40	Klukshu R. - late weir	ALSEK	2006	95	0.242	0.244	0.239	0.239	0.800	0.611
41	East Alsek R.	ALSEK	2000	96	0.287	0.283	0.285	0.278	0.625	0.625

Table 1.2. Summary of 45 SNP markers (Habicht et al. 2010) from 41 populations of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska. Summary statistics include minor allele frequency range, expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_O$ ), Weir and Cockerham (1984)  $F_{ST}$ , informativeness for assignment ( $I_n$ ), and optimal rate of correct assignment (ORCA). The column ‘CDS’ indicates loci that were determined to candidates for diversifying selection.

Marker	Minor Allele Frequency Range	$H_E$	$H_O$	$F_{ST}$	$I_n$	ORCA	CDS
<i>One_ACBP-79</i>	0.085 - 0.643	0.446	0.456	0.077	0.016	0.038	
<i>One_ALDOB-135</i>	0.030 - 0.606	0.324	0.320	0.101	0.018	0.038	
<i>One_COI</i> <sup>1</sup>	0.000 - 0.484	-	-	0.117	0.030	0.036	
<i>One_ctgf-301</i>	0.000 - 0.114	0.041	0.041	0.035	0.007	0.027	
<i>One_Cytb_17</i> <sup>1</sup>	0.000 - 0.558	-	-	0.200	0.037	0.038	
<i>One_Cytb_26</i> <sup>1</sup>	0.000 - 0.705	-	-	0.146	0.037	0.042	
<i>One_E2-65</i>	0.005 - 0.716	0.356	0.359	0.078	0.020	0.042	
<i>One_GHII-2165</i>	0.006 - 0.900	0.362	0.362	0.281	0.064	0.046	+
<i>One_GPDH-201</i> <sup>2</sup>	0.092 - 0.539	0.419	0.418	0.061	0.014	0.035	
<i>One_GPDH2-187</i>	0.039 - 0.611	0.345	0.349	0.111	0.025	0.038	
<i>One_GPH-414</i>	0.008 - 0.562	0.275	0.278	0.096	0.023	0.038	
<i>One_hsc71-220</i>	0.024 - 0.616	0.378	0.381	0.086	0.022	0.039	
<i>One_HGFA-49</i>	0.000 - 0.321	0.220	0.216	0.065	0.017	0.032	
<i>One_HpaI-71</i>	0.143 - 0.801	0.446	0.449	0.087	0.020	0.040	
<i>One_HpaI-99</i>	0.000 - 0.379	0.225	0.217	0.068	0.019	0.034	
<i>One_IL8r-362</i>	0.000 - 0.227	0.040	0.041	0.077	0.011	0.030	
<i>One_KPNA-422</i>	0.029 - 0.391	0.271	0.266	0.042	0.009	0.033	
<i>One_LEI-87</i>	0.003 - 0.489	0.313	0.313	0.080	0.020	0.036	
<i>One_MARCKS-241</i>	0.000 - 0.096	0.043	0.043	0.040	0.008	0.027	
<i>One_MHC2_190</i>	0.000 - 0.946	0.310	0.303	0.268	0.069	0.047	+
<i>One_MHC2_251</i> <sup>3</sup>	0.016 - 0.991	0.343	0.340	0.297	0.080	0.048	+
<i>One_Ots213-181</i>	0.135 - 0.775	0.410	0.405	0.072	0.017	0.040	
<i>One_p53-534</i>	0.000 - 0.272	0.083	0.085	0.072	0.014	0.031	
<i>One_ins-107</i>	0.198 - 0.892	0.447	0.450	0.113	0.024	0.041	
<i>One_Prl2</i>	0.036 - 0.907	0.461	0.469	0.081	0.019	0.046	
<i>One_RAG1-103</i>	0.000 - 0.066	0.033	0.034	0.017	0.005	0.026	
<i>One_RAG3-93</i>	0.000 - 0.220	0.166	0.166	0.058	0.017	0.030	
<i>One_RFC2-102</i>	0.039 - 0.568	0.330	0.333	0.084	0.020	0.037	
<i>One_RFC2-285</i>	0.000 - 0.297	0.133	0.135	0.070	0.014	0.032	
<i>One_RH2op-395</i>	0.000 - 0.088	0.012	0.012	0.047	0.005	0.027	
<i>One_serpin-75</i>	0.000 - 0.158	0.064	0.064	0.034	0.009	0.028	
<i>One_STC-410</i>	0.000 - 0.790	0.274	0.276	0.156	0.034	0.044	

Table 1.2. Continued.

Marker	Minor Allele Frequency Range	H <sub>E</sub>	H <sub>O</sub>	F <sub>ST</sub>	I <sub>n</sub>	ORCA	CDS
<i>One_STR07</i>	0.005 - 0.845	0.353	0.358	0.154	0.036	0.045	
<i>One_Tf_ex11-750</i>	0.077 - 0.769	0.424	0.420	0.107	0.029	0.041	
<i>One_Tf_in3-182</i>	0.000 - 0.140	0.049	0.049	0.046	0.009	0.028	
<i>One_U301_92</i>	0.018 - 0.638	0.274	0.272	0.116	0.029	0.040	
<i>One_U401-224</i>	0.090 - 0.758	0.453	0.464	0.086	0.019	0.041	
<i>One_U404-229</i>	0.006 - 0.229	0.130	0.132	0.042	0.009	0.030	
<i>One_U502-167</i>	0.000 - 0.113	0.041	0.041	0.039	0.008	0.027	
<i>One_U503-170</i>	0.000 - 0.326	0.131	0.127	0.071	0.018	0.032	
<i>One_U504-141</i>	0.060 - 0.559	0.350	0.341	0.101	0.023	0.037	
<i>One_U508-533</i>	0.000 - 0.342	0.152	0.153	0.085	0.020	0.033	
<i>One_VIM-569</i>	0.024 - 0.300	0.226	0.222	0.040	0.009	0.031	
<i>One_ZNF-61</i>	0.024 - 0.703	0.334	0.339	0.104	0.021	0.041	
<i>One_Zp3b-49</i>	0.000 - 0.709	0.361	0.365	0.200	0.044	0.042	+

<sup>1</sup> mtDNA loci; <sup>2</sup> Linkage between the GPDH loci was found to be significant and this locus was dropped during baseline evaluation; <sup>3</sup> Linkage between the MHC loci was found to be significant and these loci were phenotyped for baseline evaluation

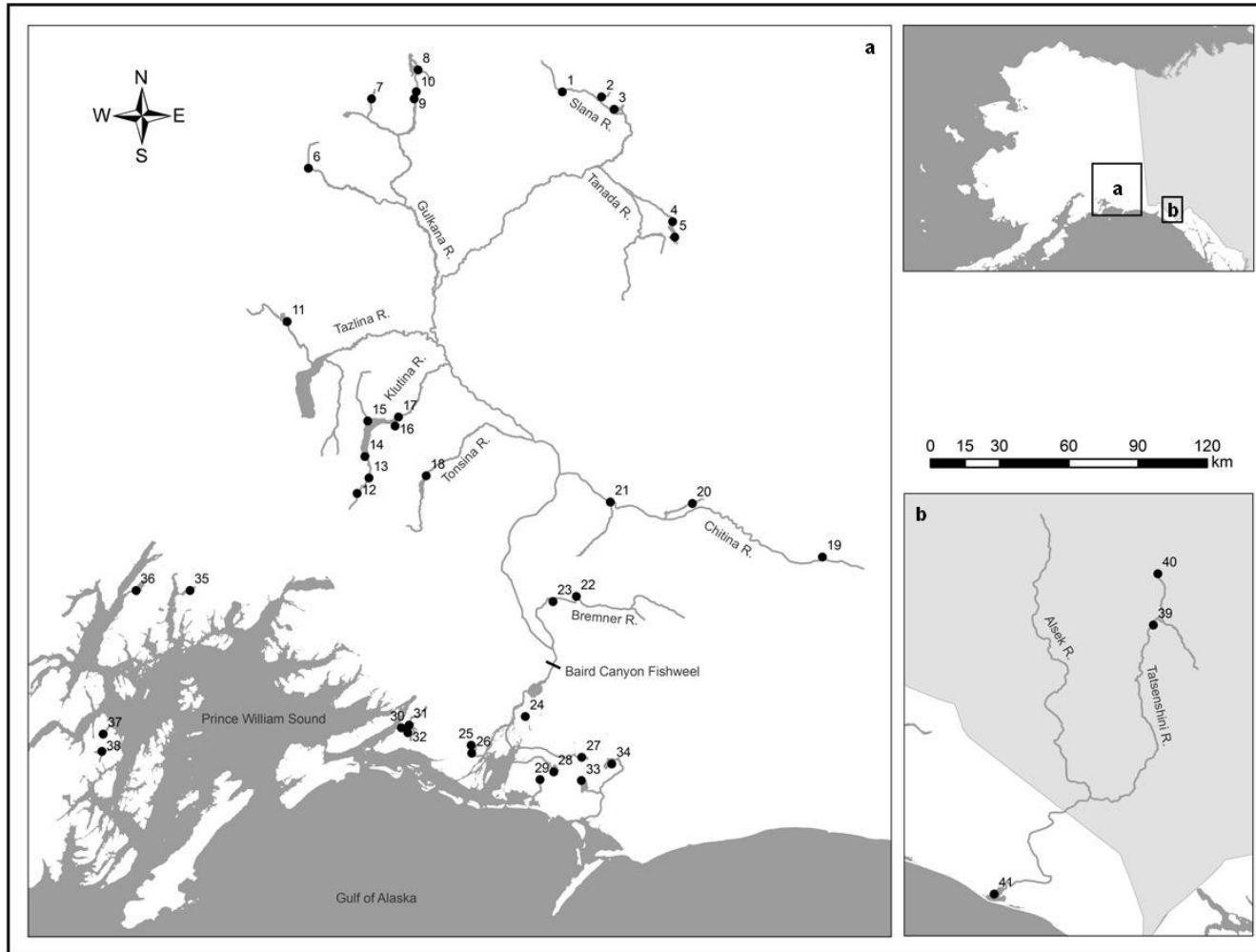


Figure 1.1. Map of sample sites (populations) of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska. Sample site numbers correspond to the population numbers in Table 1.1.

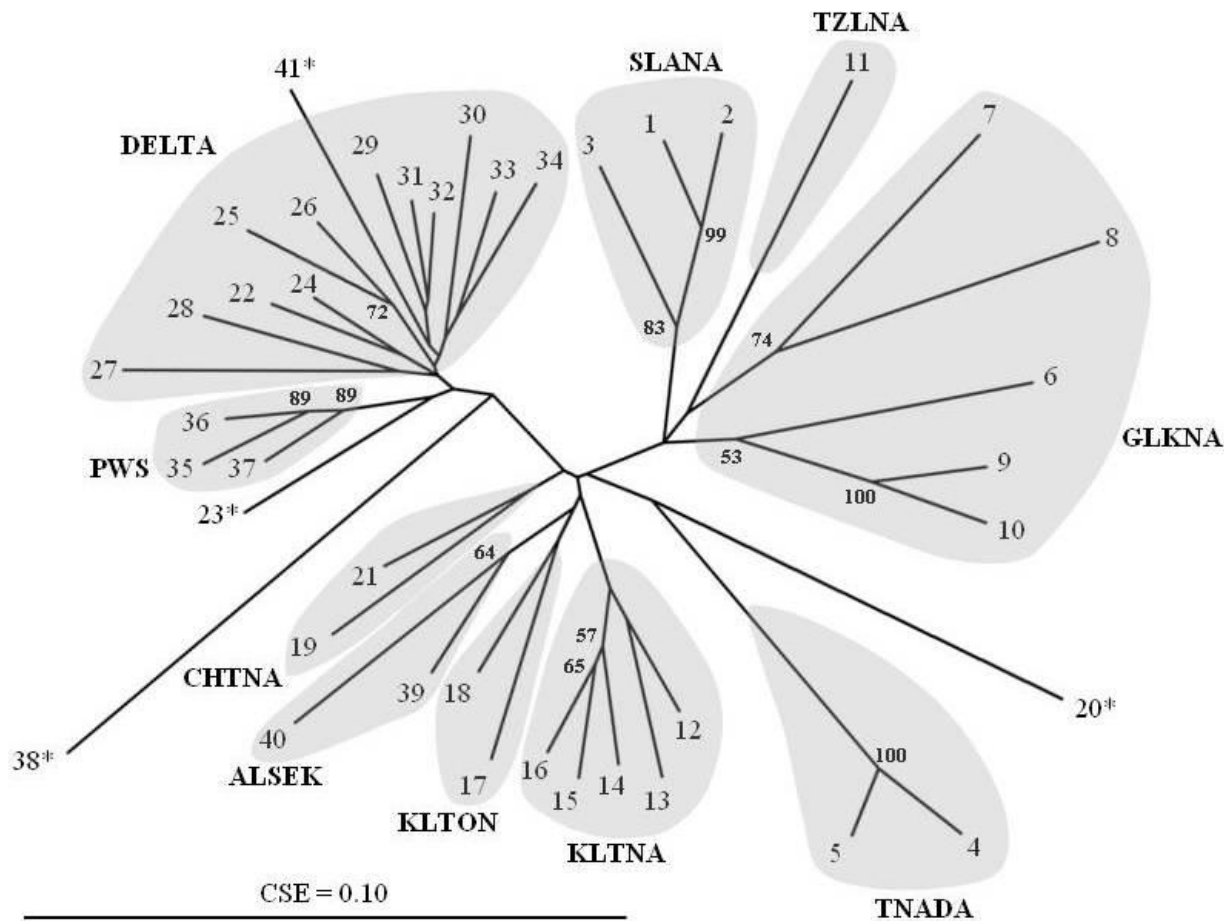


Figure 1.2. Neighbor-joining dendrogram of populations of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska using the full marker panel. Bootstrap values greater than 50 percent and a scale of Cavalli-Sforza and Edwards (1967) distance are shown. Numbers correspond to populations listed in Table 1.1. Gray shading corresponds with proportional assignment reporting groups (Table 1.1). Asterisked populations are outlier populations that were placed in the following reporting groups: 20) CHTNA, 23) DELTA, 38) PWS, 41) ALSEK.

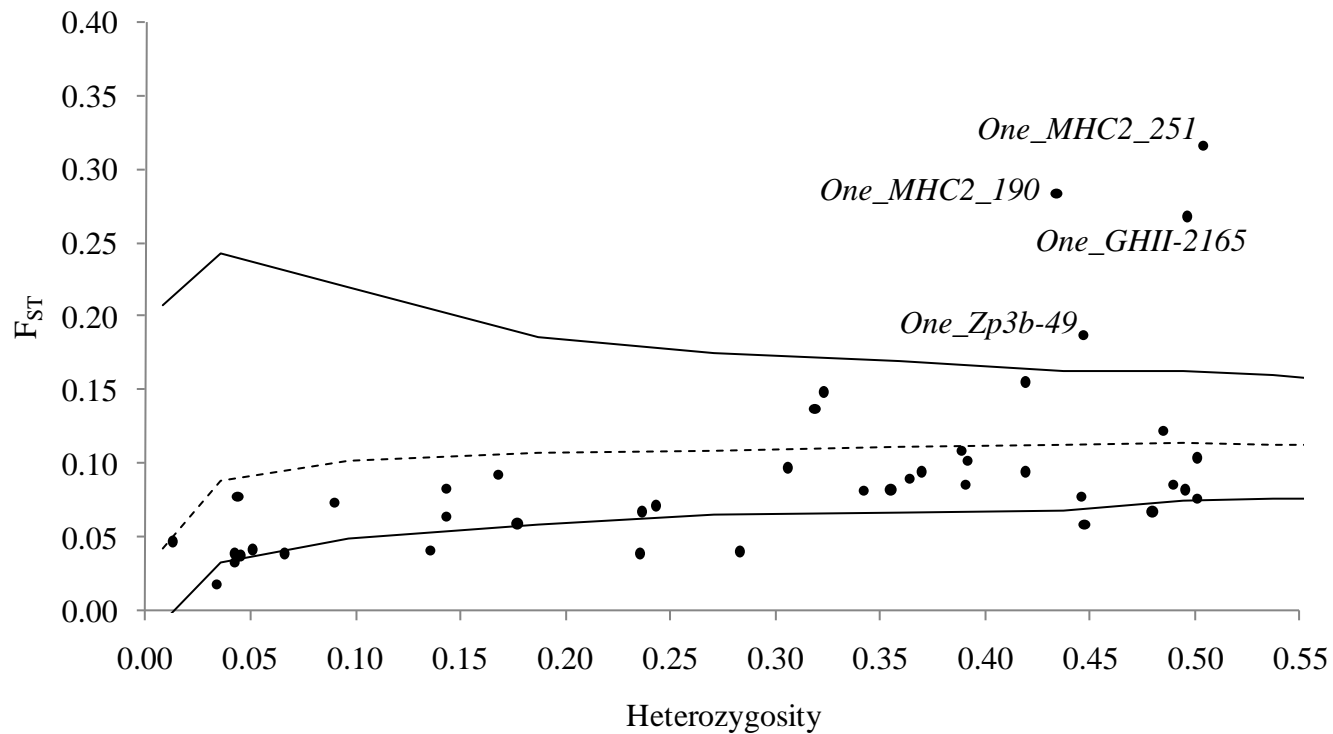


Figure 1.3.  $F_{ST}$  values as a function of heterozygosity (ARLEQUIN v3.5; Excoffier and Lischer 2010) for 42 SNP loci (mtDNA SNPs excluded). The dashed line represents the median and the solid lines represent the 95% confidence interval boundaries based on coalescent simulations and using a hierarchical island model. The four loci labeled above the upper 95% confidence interval boundary are candidates for diversifying selection.

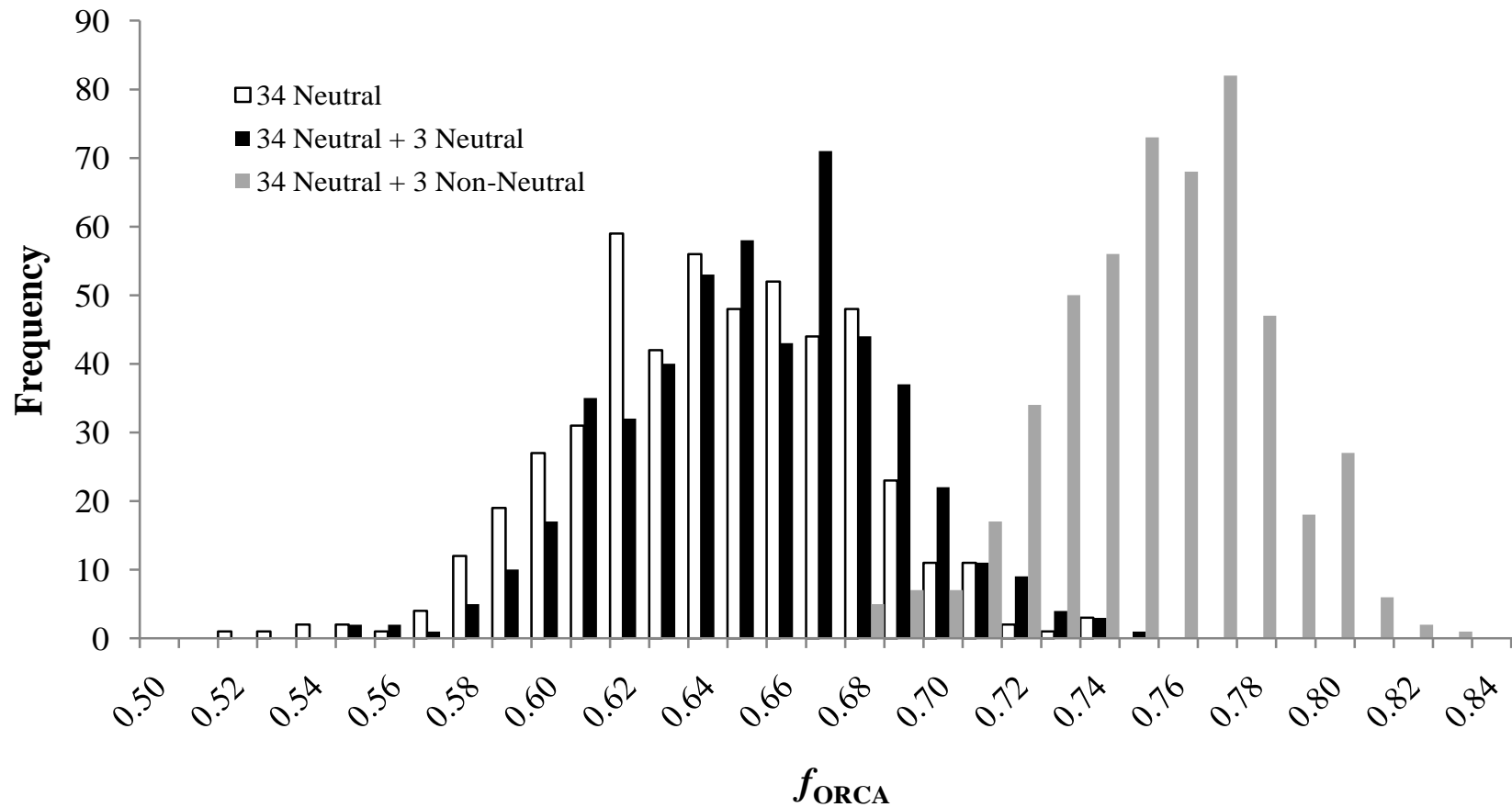


Figure 1.4. Distribution of  $f_{ORCA}$  scores (Rosenberg et al. 2003) for three different sets of 500 marker panels. For each set of marker panels, 1,000 individuals were simulated, each generated using allele frequencies from one of the 41 baseline populations chosen at random. Each simulated individual was generated using 34 of the 37 neutral loci randomly sampled (with replacement). That individual was then generated using the same 34 neutral loci with an additional three neutral loci (randomly sampled from the same panel of 37 neutral loci) and with an additional three non-neutral loci. The  $f_{ORCA}$  score represents the proportion of 1,000 individuals that correctly assign back to the population from which they were generated.

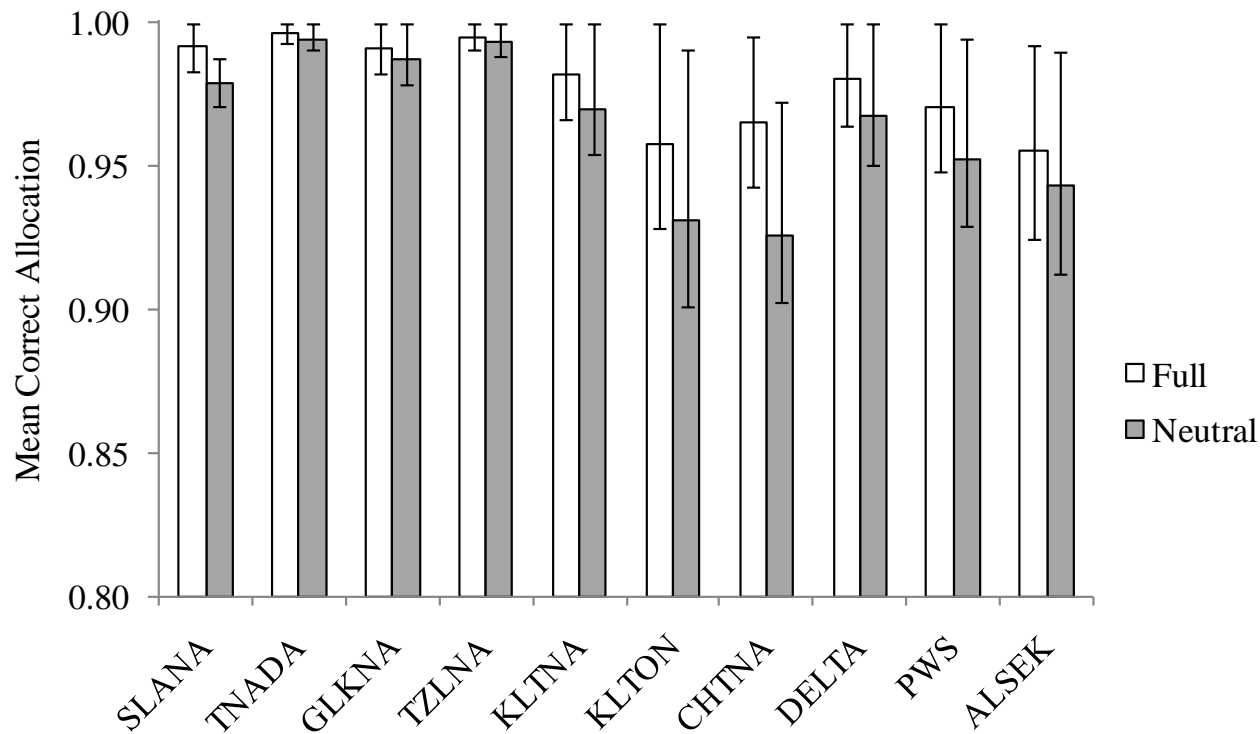


Figure 1.5. Mean correct allocation (90% confidence intervals) for mixtures of simulated fish (SPAM v3.7b; updated version of Debevec et al. 2000). An analysis was run for each of the 10 reporting groups in which the simulated mixture was composed entirely of individuals (n=300) generated using allele frequencies from populations within that reporting group, and each population within the reporting group contributing equally to the mixture. Mean correct allocations were estimated from 1,000 simulations. Ninety percent confidence intervals for each reporting group estimate were computed using 1,000 bootstrap resamples and the non-symmetric percentile method.

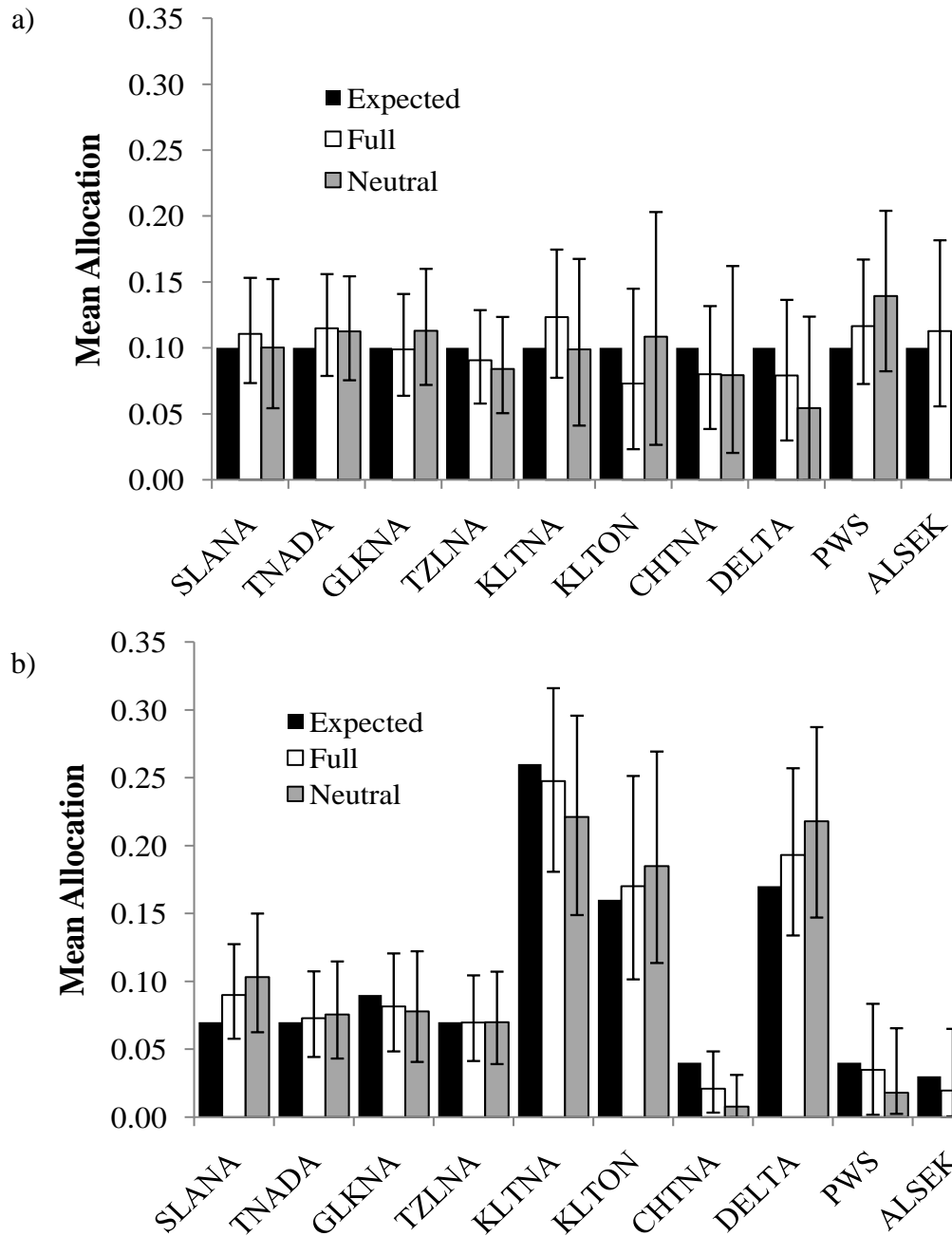


Figure 1.6. The expected allocation and mean correct assignment allocation and 95% credibility intervals using the neutral and the full marker panels for three known mixtures of 300 individuals each removed from the baseline and assigned back to the baseline (BAYES; Pella and Masuda 2001). Known mixtures are (a) equal number of fish from each of the reporting groups (i.e. 30 fish each) and (b) representative of what might be expected in fisheries in the region (i.e. the Copper River district commercial gillnet fishery) based on total escapement data from the 2005 to 2007 seasons (Fair et al. 2008; Wade et al. 2008).

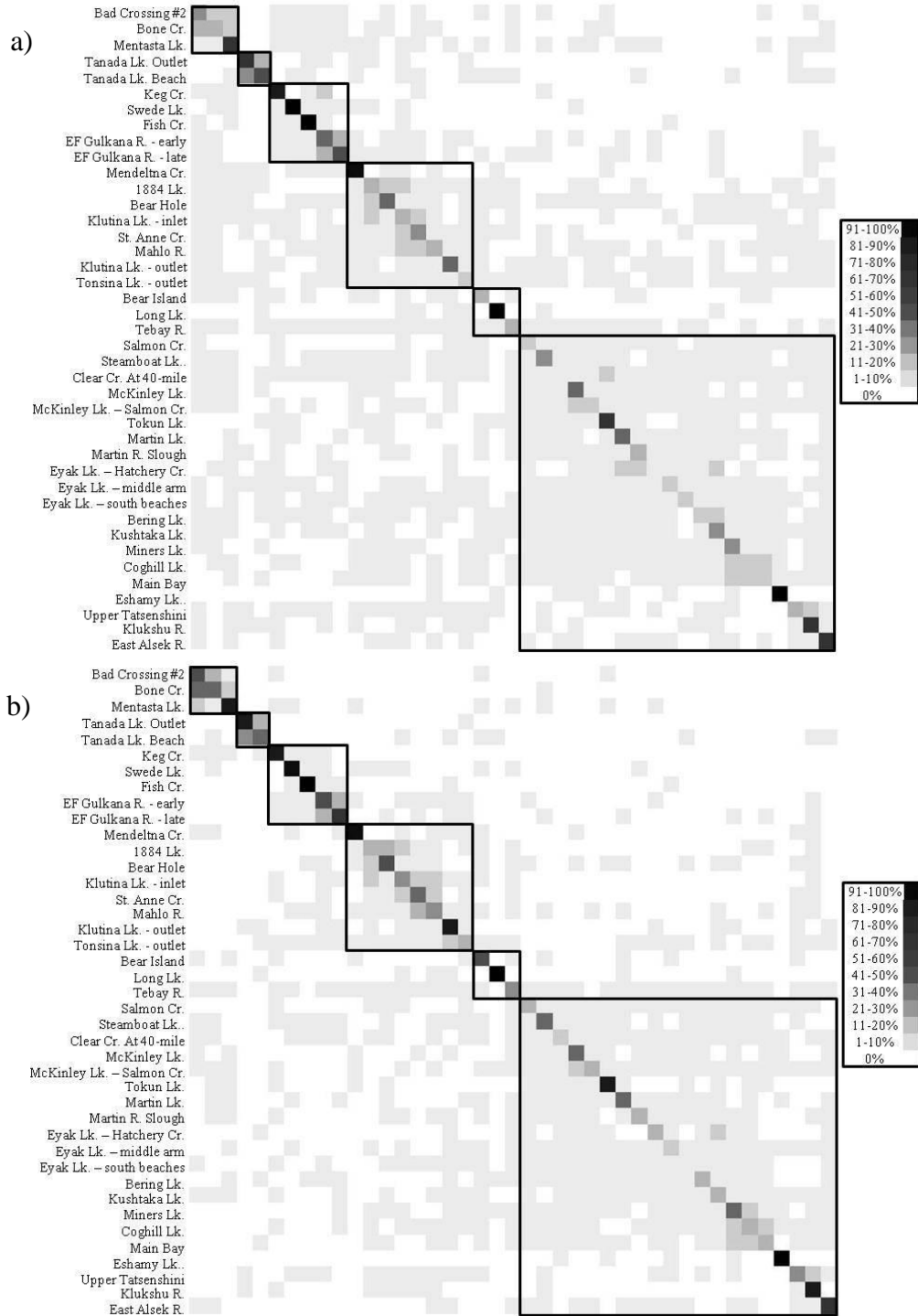


Figure 1.7. Self-assignment matrices of populations of sockeye salmon from the Copper River and adjacent coastal drainages in southcentral Alaska as determined with (a) the neutral marker panel and (b) the full marker panel (GENECLASS2; Piry et al. 2004). The diagonals represent the percentage of self-assigned individuals from a populations and shaded blocks above and below the diagonal indicate percentage of mis-assignments to populations. The black grid lines correspond to the individual assignment reporting groups. Legends depict percentage of assignment in 10% increments.

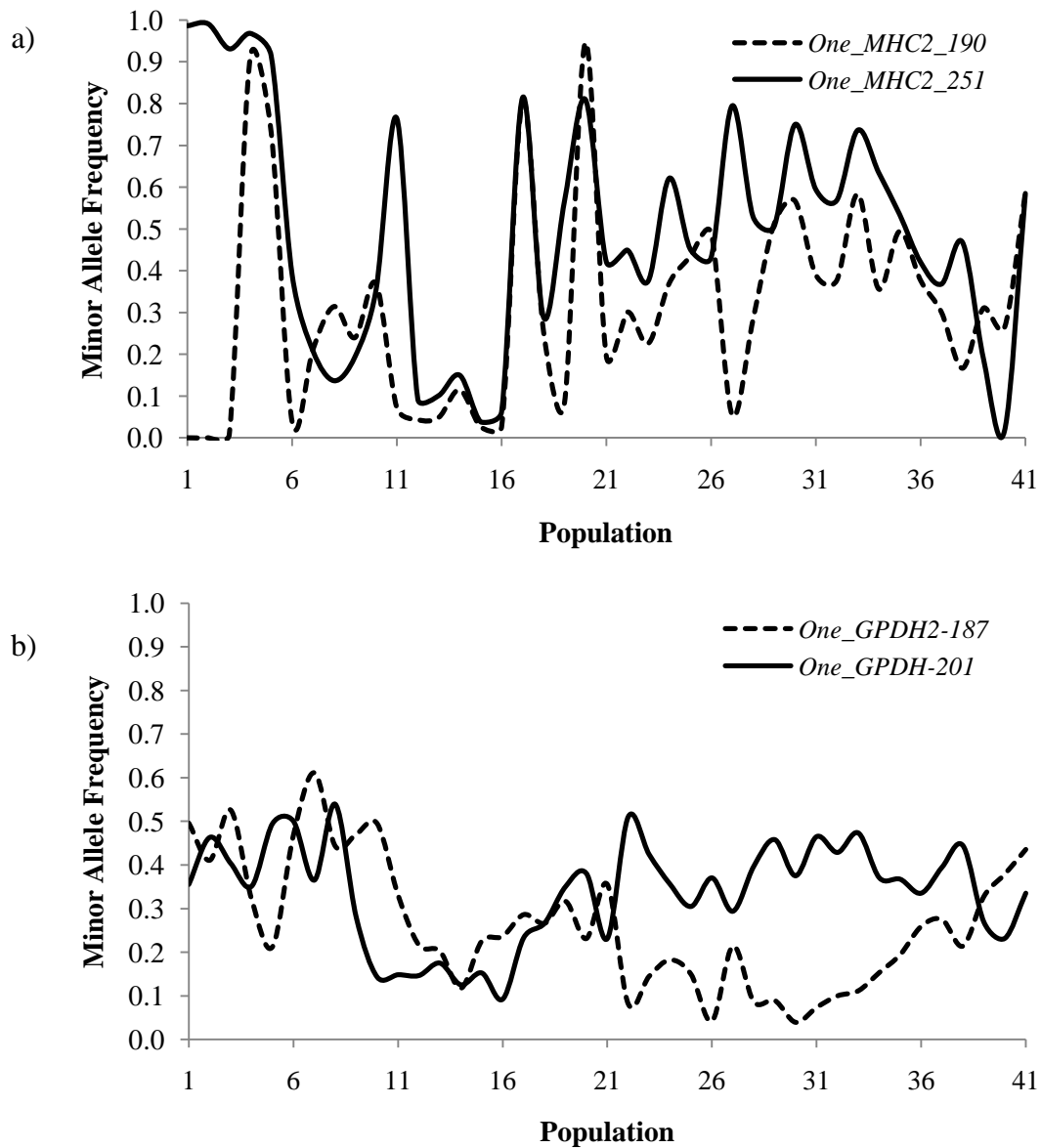


Figure 1.8. Minor allele frequencies of a) *One\_GPDH2-187* and *One\_GPDH-201* and b) *One\_MHC2\_190* and *One\_MHC2\_251* for sockeye salmon populations surveyed from the Copper River and adjacent coastal drainages in southcentral Alaska. Population numbers correspond to numbering in Table 1.1 and Figure 2.1.

## **Chapter 2: Analysis of the spatial genetic structure of sockeye salmon (*Oncorhynchus nerka*) in the Copper River, Alaska**

### **Abstract**

Identifying the spatial distribution of genetic variation across the landscape is an essential step in informing species conservation. Comparison of closely related and geographically overlapping species can be particularly useful in cases where landscape may similarly influence their genetic structure. Congruent patterns among species may highlight the importance landscape plays in determining spatial genetic structure whereas contrasting patterns may emphasize the importance of species-specific adaptation to local environments. In this study, we examined the interacting roles of demography and adaptation in determining the spatial distribution of genetic variation in two closely related and geographically overlapping species in a pristine environment. Using single nucleotide polymorphism (SNP) loci exhibiting both neutral and presumably adaptive variation, we evaluated the spatial genetic structure of sockeye salmon in the Copper River, southcentral Alaska and compare it to existing data available for Chinook salmon from the same drainage. Overall, both species exhibited significant patterns of isolation by distance; the spatial distribution of populations largely determined the distribution of genetic variation across the landscape. Further, both species exhibited largely congruent patterns of within- and among-population genetic diversity, highlighting the role that landscape heterogeneity and historical processes play in determining spatial genetic structure. When SNPs exhibiting adaptive variation were incorporated into landscape genetics analyses, adaptive differences among geographically proximate sockeye salmon populations were observed. Results were evaluated in the context of conservation efforts with an emphasis on reproductive isolation, historical processes, and local adaptation.

## Introduction

Identifying the distribution of genetic variation across the landscape is an essential step in informing species conservation. Landscape genetics methods facilitate our understanding of the role landscape heterogeneity plays in shaping the genetic structure of natural populations (Manel et al. 2003; Storfer et al. 2007). Comparison of closely related and geographically overlapping species can be particularly useful in cases where the landscape may similarly (or differentially) influence the genetic structure of multiple species (Marten et al. 2006). Congruent patterns among species may highlight the importance that landscape (including historical processes) plays in determining spatial genetic structure through demographic processes such as gene flow and genetic drift (Waples 1987; Petren et al. 2005; Gagnon and Angers 2006). Conversely, contrasting patterns may emphasize the importance of species-specific adaptation to local environments (Short and Caterino 2009). In either case, an understanding of the interacting roles that gene flow, genetic drift, and local adaptation play in shaping genetic structure can greatly improve our knowledge of a species and improve conservation efforts (Schwartz et al. 2009).

Traditional landscape genetics analyses have focused on divergence at loci assumed to be neutral to infer demographic history and population structure (Manel et al. 2003). However, overemphasis on the role of reproductive isolation and under-emphasis on adaptive differences among populations has received scrutiny in conservation applications (Hard 1995; Pearman 2001; Schwartz et al. 2009). In this study, we examined the interacting roles of demography and adaptation in determining the spatial distribution of genetic variation in two closely related and geographically overlapping species in a pristine environment. Using single nucleotide polymorphism (SNP) loci exhibiting both neutral and presumably adaptive variation, we evaluated the spatial genetic structure of sockeye salmon (*Oncorhynchus nerka*) in the Copper River, southcentral Alaska and compare it to existing data available for Chinook salmon (*O. tshawytscha*) from the same drainage.

Pacific salmon are distributed throughout most river systems along the west coast of North America from central California north to Alaskan and Canadian rivers flowing into the Arctic Ocean (Behnke and Tomelleri 2002). Five species of Pacific salmon are

anadromous and semelparous. The two species in this study, sockeye salmon (*Oncorhynchus nerka*) and Chinook salmon (*O. tshawytscha*), support economically and culturally valuable fisheries in the Copper River, southcentral Alaska. The high fidelity of Pacific salmon to their natal spawning and rearing environments results in genetic variation among discrete populations (Taylor 1991; Dittman and Quinn 1996; Quinn 2005). Juvenile sockeye salmon generally spend a year or more rearing in freshwater before emigrating to the marine environment; although a 'sea-type' race occurs in which juveniles migrate to the marine environment in their first year (Quinn 2005; Wood et al. 2008). Adults sockeye salmon usually mature and return to spawn in their natal freshwater environment at four to five years of age. Chinook salmon exhibit a similar life-history as sockeye salmon with one notable exception; juvenile sockeye salmon generally rear in lakes whereas juvenile Chinook salmon generally rear in fluvial environments (Behnke and Tomelleri 2002; Quinn 2005).

The Copper River, with a drainage area of approximately 64,000 km<sup>2</sup>, provides an ideal heterogeneous landscape to evaluate interactions between genetic and spatial variation in Pacific salmon. From its origins in the Wrangell, Chugach, and Alaska mountain ranges, the Copper River flows approximately 450 km in a generally southward direction before emptying into the Gulf of Alaska. During its course, the river primarily drains two different environments. The upper Copper River and its main tributaries, the Slana, Tanada, Chistochina and Gulkana rivers, drain large portions of tundra and are generally characterized by lower gradient. The middle and lower Copper River and its main tributaries, the Tazlina, Klutina, Tonsina, and Chitina rivers, drain a more forested, mountainous landscape and are largely fed by glacial meltwater and are characterized by higher gradient, glacially turbid reaches. Based on differences between the upper and middle/lower Copper River environments, we hypothesize that sockeye salmon will exhibit an inland-coastal transition in the distribution of genetic variation, and that abrupt differences in allele frequencies (i.e., barriers) may occur between the two environments.

In this study we used a combination of population genetics and landscape genetics methods to assess the spatial distribution of genetic diversity for sockeye salmon in the Copper River. Our analyses and discussion focus primarily on sockeye salmon and evaluating the roles of spatial variation and local adaptation in shaping their hierarchical

genetic structure in the Copper River drainage. The sockeye salmon data was drawn from a genetic baseline previously developed to document population structure and perform mixture analyses in the study region (Chapter 1). In addition, a comparable multi-locus genotype dataset was available for Chinook salmon in the Copper River (Seeb et al. 2009b). We used a subset of the Chinook salmon dataset from Seeb et al. (2009b) to evaluate how landscape (including historical processes) may similarly or differentially shape the genetic structure of the two species. Results from each species were used to evaluate the interacting roles of landscape heterogeneity and local adaptation in determining genetic variation.

On a large scale, we predict that genetic isolation by distance will be significant for sockeye salmon due to: 1) exchange of genetic material between populations will be more likely to occur among geographically proximate populations than among more distant populations and 2) the presence of genetic barriers. Second, we predict that an inland-to-coastal transition of within-population diversity (allelic richness) will occur that will coincide with the transition in environments between the upper and middle/lower Copper River. Third, we predict that sockeye and Chinook salmon will exhibit largely congruent patterns of hierarchical genetic structure, highlighting the importance that landscape heterogeneity and historical geologic processes plays in shaping the genetic structure in multiple species. And fourth, incorporating SNPs exhibiting adaptive variation into the hierarchical structure analyses, we predict that important adaptive differences between geographically proximate populations will be observed.

Spatial and genetic data were analyzed to test these four predictions. Data indicated that the spatial distribution of sockeye salmon populations in the Copper River drainage primarily determined their spatial genetic structure. Patterns of genetic diversity indicated that sockeye salmon populations throughout the upper Copper River exhibited less within-population diversity and greater population structure than populations in the lower Copper River, likely resulting from greater reproductive isolation and historical founder effects. Further, SNPs exhibiting adaptive variation highlighted interesting adaptive differences among geographically proximate sockeye salmon populations. Chinook salmon populations exhibited largely congruent patterns of hierarchical genetic structure demonstrating that landscape heterogeneity and historical processes similarly

influence genetic variation in both species in the Copper River. Results were evaluated in the context of conservation with an emphasis on demographic isolation, historical processes, and local adaptation.

## **Methods**

### **Sampling and genetic data collection**

The sockeye salmon genetic data were drawn from chapter 1, but includes only populations that spawn within the Copper River drainage (excluding Alsek, Bering, and Eyak rivers and Prince William Sound). The Chinook salmon data were drawn from Seeb et al. (2009b) with one population removed. To increase baseline coverage for mixture analyses, Seeb et al. (2009b) included individuals from the Tonsina River that were captured and sampled at fishwheels approximately 70 km above the mouth of the Copper River. The individuals were radiotagged at the fishwheel and were later determined to return to the mainstem Tonsina River using radiotelemetry (Savereide 2005). This population was removed as we were not performing mixture analyses. The remaining data were all from adults captured near the spawning grounds. In total, data were available for adult sockeye and Chinook salmon from a total of 58 field collections made from spawning populations during 2004-2009 (Table 2.1). During field collection, a GPS waypoint or a physical description of each location was taken to create a geographic information system (GIS) data shapefile of the populations in ArcGIS 9.3 (ESRI, Redlands, California).

### **Statistical analyses**

Allele frequencies were calculated across all loci for sockeye salmon populations in FSTAT v2.9.3.2 (Goudet 1995; Goudet 2001). Tests for deviation from HWE were evaluated across all loci (mtDNA SNPs excluded) for each population using exact P-values calculated from the Markov Chain (MC) method implemented in GENEPOP v4.0 (Rousset 2008). Default parameters were used for the MC algorithm (dememorization = 1,000; batches = 20; iterations per batch = 5,000). Critical values ( $\alpha = 0.05$ ) were adjusted for multiple tests using a step-down sequential Bonferroni correction (Holm 1979). Tests for linkage disequilibrium between all locus pairs were performed using

simulated exact tests in GENEPOP v4.0. A pair of loci were determined to be significantly out of linkage equilibrium if tests for linkage disequilibrium were significant in more than half of the populations at the  $\alpha = 0.05$  level.

Loci were evaluated for within-population diversity and for among population divergence across all populations. The expected heterozygosity ( $H_E$ ) of each nuclear locus was estimated using GENALEX 6.3 (Peakall and Smouse 2006). The Weir and Cockerham (1984)  $F_{ST}$  for each locus was estimated using FSTAT v2.9.3.2. Based on within-population and among-population diversity statistics, we differentiated between those SNPs exhibiting neutral variation and SNPs exhibiting adaptive variation. Candidate SNP loci for directional selection (hereafter referred to as adaptive SNPs) were identified using a genome scan approach as implemented in ARLEQUIN v3.5 (Excoffier and Lischer 2010). ARLEQUIN v3.5 uses coalescent simulations to obtain a null distribution of  $F_{ST}$  as a function of expected heterozygosities across loci. We performed 20,000 simulations using 100 demes per group under the assumptions of a finite island model. We were only interested in high  $F_{ST}$  outliers; low  $F_{ST}$  outliers suggestive of balancing selection (Beaumont and Balding 2004) were ignored. Any SNPs that fell above the 99% quantile were determined to be adaptive SNPs. Seeb et al. (2009b) identified one SNP (*Ots\_MHC-2*) exhibiting adaptive variation at the  $\alpha = 0.05$  level and using the infinite-alleles model as implemented in FDIST2 (M.A. Beaumont, University of Reading, UK; <http://www.rubic.rdg.ac.uk/~mab/software.html>). We performed a similar but newly described analysis at the  $\alpha = 0.01$  level using the finite island model in ARLEQUIN v3.5. All remaining SNPs were assumed to be neutral.

Fluvial distance and  $F_{ST}/(1 - F_{ST})$  are expected to follow a linear model when gene flow is geographically limited (Rousset 1997; Guillot et al. 2009). Seeb et al. (2009) showed that this relationship holds true for Chinook salmon in the Copper River drainage. We wanted to determine whether this relationship is also significant for sockeye salmon in the Copper River using a Mantel test implemented in GENALEX 6.3. A significant isolation-by-distance relationship would indicate that the spatial distribution of populations largely determines the hierarchical population structure in sockeye salmon. One hundred permutations were performed to test the statistical significance of the pairwise matrix correlations (Sokal and Rohlf 1995). As the goal was to examine the

dispersal ability or patterns of straying, adaptive SNPs were eliminated from the isolation-by-distance analysis.

Decreases in allelic richness may be indicative of demographic isolation among populations or population bottlenecks (including founder effects). To examine for potential gradients of within-population diversity across the Copper River landscape for each species, we spatially interpolated the observed allelic richness of populations across the drainage using the ordinary kriging (Isaaks and Srivastava 1989) function in the geostatistical analyst toolbox in ArcGIS 9.3. Further, allelic richness was plotted as a function of upstream migration distance for each species. Upstream migration distance was measured as the fluvial distance from the mouth of the Copper River to each population's spawning habitat and was measured using ArcGIS 9.3. The allelic richness for each population was calculated in FSTAT v2.9.3.2. Each population's allelic richness was calculated by averaging the allelic richness across loci within each population. Mitochondrial and adaptive SNPs were eliminated from allelic richness analyses.

SAMOVA v1.0 (Spatial Analysis of Molecular Variance; Dupanloup et al. 2002) was used to evaluate patterns of spatial genetic structure for each species. SAMOVA performs a series of AMOVA analyses to define population groups that are geographically and genetically similar and are maximally differentiated from each other using a simulated annealing process (Kirkpatrick et al. 1983). The goal is to minimize the amount of variation that occurs within groups ( $F_{SC}$ ) and maximize the amount of variation that occurs among groups ( $F_{CT}$ ). The number of groups to be defined ( $k$ ) is a user-defined variable; we conducted analyses of  $k$  equals 2 through 10 for each species and examined patterns of  $F_{SC}$  and  $F_{CT}$  for each  $k$  to determine the appropriate number of population groupings. The analysis was initially conducted for each species using exclusively neutral SNPs. Following, the analysis was re-run incorporating adaptive SNPs to determine if adaptive variation among populations changed population groupings suggesting signals of adaptive differences among geographically proximate populations. Because latitude and longitude data do not reflect the true spatial relationships in a fluvial system, we calculated surrogate  $x$  and  $y$  coordinates using a pairwise matrix of waterway distances to project the position of the populations on a

multidimensional scale (MDS) using the program NTSYS (Rohlf 1994). The populations were spatially referenced using the first and second MDS values.

The program BARRIER v2.2 (Manni et al. 2004) was used to identify any barriers to gene flow on the landscape and to complement the SAMOVA analysis. BARRIER uses Monmonier's maximum difference algorithm (Monmonier 1973) to detect areas of abrupt change in population divergence along edges of a Delaunay triangulation (Delaunay 1934) connecting population pairs. Simulations suggest the Monmonier's approach is better than the simulated annealing approach (SAMOVA) for identifying barriers, especially in cases where population divergence follows an isolation-by-distance model (Dupanloup et al. 2002; Manni et al. 2004). The first and second MDS values used in the SAMOVA analysis were also used to create the Delaunay triangulation used in BARRIER. To account for patterns of isolation-by-distance in each species, we used a matrix of residuals (based on the observed isolation-by-distance relationship) rather than the observed  $F_{ST}$ s as suggested by Manni et al. (2004). By using the residuals, the influence of isolation-by-distance is removed during the identification of barriers. To determine the robustness of identified barriers, 100 matrices of residuals were generated by bootstrapping across loci. Bootstrap sampling across loci and calculation of pairwise observed, expected, and residual  $F_{ST}$ s were performed in R v2.9.2 (R Development Core Team, 2009). R-scripts were written and provided by J. Olsen and J. Bromaghin (USFWS, Anchorage, Alaska). Similar to the SAMOVA analysis, the BARRIER analysis was first conducted for each species using only neutral SNPs to identify any 'demographic barriers'. The analysis was then re-run incorporating adaptive SNPs to determine if any additional 'adaptive barriers' were identified, suggestive of abrupt differences in allele frequencies at adaptive SNPs among populations. We identify any barriers that were consistently identified with greater than 50% bootstrap support.

## **Results**

Multi-locus genotype data were available for 28 sockeye salmon ( $n = 4,100$ ) populations from throughout the Copper River drainage. For sockeye salmon, sample sizes ranged from 30 to 299 individuals and averaged 146. Additionally, data were available for 13 Chinook salmon ( $n = 1,184$ ) populations throughout the Copper River

drainage (Table 2.1, Figure 2.2). No data were available for Chinook salmon from the lower Copper River (below the confluence with the Chitina River) as there is little evidence of the species in that region (Savereide 2005).

All SNPs identified as polymorphic in the original baselines (Chapter 1; Seeb et al. 2009b) were used in this study and include 45 sockeye salmon SNPs and 41 Chinook salmon SNPs. For sockeye salmon, there were no significant deviations from HWE out of 1,006 tests (28 populations at 42 loci; mtDNA SNPs excluded) after correction for multiple tests. Before correction for multiple tests, 41 of the 1,006 tests were significant (50 would be expected by chance). Of the 41 significant tests (before correction), deviations were not consistently found either among specific locus pairs or within specific populations. As observed by Seeb et al. (2009b), all Chinook salmon SNP loci conformed to HWE after correction for multiple tests.

Significant linkage disequilibrium was detected between two sockeye salmon locus pairs and three Chinook salmon locus pairs. For sockeye salmon, significant linkage disequilibrium was detected between *One\_GPDH-201* and *One\_GPDH2-187* in more than half of the populations. Consequently, *One\_GPDH-201* was dropped from all subsequent analyses as it was the less informative of the locus pair. Further, significant linkage disequilibrium was detected between *One\_MHC2\_190* and *One\_MHC2\_251* in more than half of the populations. Both MHC loci were retained for further analyses as each were later determined to exhibit adaptive variation (see below). For both the SAMOVA and BARRIER analyses, *One\_MHC2\_190* and *One\_MHC2\_251* were incorporated as a composite phenotype. The composite phenotype was created by combining genotypes of the two linked loci, ordering them alphabetically, and then assigning a numeric code to the resulting composite phenotype (Habicht et al. 2010). In Chinook salmon, significant linkage disequilibrium was found between *Ots\_FGF6A* and *Ots\_FGF6B*, *Ots\_HSP90B-100* and *Ots\_HSP90B-385* as well as *Ots\_MHC1* and *Ots\_MHC2* (Seeb et al. 2009b). *Ots\_FGF6B*, *Ots\_HSP90B-385*, and *Ots\_MHC1* were dropped from subsequent analyses as they were the less informative of the respective locus pairs (Seeb et al. 2009b). Additionally, significant linkage disequilibrium was found between *Ots\_MHC-2* and *Ots\_LWSop-638*, but both loci were retained for further

analyses as the structural relationship between the two loci is uncertain (Seeb et al. 2009b).

Based on the  $F_{ST}$  outlier analysis, three sockeye salmon SNPs and one Chinook salmon SNP fell above the 99% quantile and were determined to be adaptive SNPs. Across sockeye salmon loci, a minimum  $H_E$  was observed for *One\_RH2op-395* (0.007) and a maximum was observed for *One\_Prl2* (0.451). A minimum  $F_{ST}$  was observed for *One\_RAG1-103* (0.021) and a maximum was observed for *One\_MHC2\_251* (0.380). Based on coalescent simulations performed in ARLEQUIN v3.5, three sockeye salmon SNPs (*One\_GHII-2165*, *One\_MHC2\_190*, *One\_MHC2\_251*) exhibited greater among-population diversity that expected under neutral processes and were determined to be adaptive SNPs. Results from Chinook salmon corroborated the findings of Seeb et al. (2009b) using the infinite-alleles model in FDIST2. One Chinook salmon SNP locus (*Ots\_MHC-2*) fell above the 99% quantile and was determined to be adaptive (Table 2.2; Figure 2.3.).

Both sockeye salmon ( $r^2 = 0.241$ ; p-value < 0.001) and Chinook salmon ( $r^2 = 0.362$ ; p-value < 0.001; Seeb et al. 2009b) exhibited significant isolation by distance relationships in the Copper River (Figure 2.4). That is, according to the neutral variation observed among populations, gene flow was more likely to occur between geographically proximate populations than between more distant populations. For sockeye salmon, the linear relationship was  $F_{ST}/(1-F_{ST}) = (0.00009 \times \text{km}) + 0.0206$  where km is the number of fluvial kilometers between a pair of populations. For Chinook salmon, the linear relationship was  $F_{ST}/(1-F_{ST}) = (0.00008 \times \text{km}) + 0.01$ . Pairwise residual matrices used in the BARRIER analyses were calculated based on these observed relationships.

Sockeye and Chinook salmon populations that spawn nearer to the coast appeared to have higher allelic richness than populations that spawn further inland. Among sockeye salmon populations, the minimum allelic richness (1.61) was observed in the Fish Creek (8) population within the Gulkana River drainage and the maximum (1.92) was observed in the Clear Creek at 40-mile (24) population located in the lower Copper River. Allelic richness tended to decrease as upstream migration distance increased for populations in both species, although the relationships did not appear to be linear. In particular, a marked decrease in allelic richness was observed in populations > 375 km

upriver (Figure 2.5). Similarly, Seeb et al. (2009b) showed that significant differences in allelic richness occurred among three geographic regions in Chinook salmon (upper Copper River < Gulkana River < lower Copper River) and that regions nearer to the coast exhibited greater allelic richness. Spatial patterns of allelic richness interpolated across the Copper River landscape (Figures 2.1 and 2.2) further supported these findings.

Using exclusively SNPs exhibiting neutral variation, the SAMOVA results suggested that nine sockeye salmon and six Chinook salmon population groups were appropriate when attempting to maximize the among-group variation ( $F_{CT}$ ) and minimize the within-group variation ( $F_{SC}$ ; Figures 2.1 and 2.2, respectively). Although the  $F_{CT}$  continued to increase slightly and the  $F_{SC}$  continued to decrease slightly as the number of groups was increased for each species, the  $F_{CT}$  largely plateaued and the  $F_{SC}$  markedly decreased at these respective groupings (Table 2.3). Findings from the SAMOVA analysis reflected what would be expected based on the within-population diversity (allelic richness) analyses. Two large sockeye salmon population groupings dominated the middle/lower Copper River region in which greater within-population diversity was observed among populations. The two groupings include; 1) seven populations that all spawn below the confluence of the Chitina and Copper Rivers and 2) nine populations throughout the Klutina, Tonsina, and Chitina river drainages (middle Copper River). Among the sockeye salmon populations that spawn > 375 km upstream of the mouth of the Copper River that exhibit decreased allelic richness, SAMOVA identified six population groups, none of which were larger than three populations. The final population group for sockeye salmon was a single population, Long Lake (20), that spawns in the Chitina River drainage.

After incorporating the non-neutral SNPs into the SAMOVA analysis, results suggested that nine sockeye salmon and six Chinook salmon population groups were still appropriate. For Chinook salmon, the population groups remained the same after *Ots\_MHC-2* was included with the neutral SNPs. However, two sockeye salmon populations changed groups, suggesting adaptive differences between geographically proximate populations. First, the Klutina Lake outlet (17) population that previously grouped with the large group encompassing the Klutina River, Tonsina River, and two populations from the Chitina River transferred to a group with the Keg Creek (6), East

Fork Gulkana River – early broodstock (9), and East Fork Gulkana River – late broodstock (10) populations after incorporating the non-neutral SNPs. Similarly, the Bear Island (19) population that belonged to the same large group as the Klutina Lake outlet population transferred to a group with the Mendeltna Creek (11) population in the Tazlina River.

The BARRIER results for both species largely corroborate the findings from SAMOVA. For sockeye salmon, five barriers consistently occurred with greater than 50% bootstrap support, and all barriers occurred between SAMOVA groupings (Figure 2.1). Of the five barriers, four occurred in the upper Copper River above 375 km. The remaining barrier occurred below the Long Lake population that formed its own distinct population group in the middle Copper River according to the SAMOVA analysis. Further, similar to observed in the SAMOVA analysis, an additional ‘adaptive barrier’ was observed above the population that spawns in the mainstem Klutina River below Klutina Lake (17) after incorporating the adaptive SNPs, separating itself from the five other populations that spawn in tributaries and backwater habitats of those tributaries above Klutina Lake. For Chinook salmon, four barriers consistently occurred with greater than 50% bootstrap support, and each of the barriers occurred between SAMOVA groupings (Figure 2.2) one of which occurred near the 375 km area separating the upper and the middle/lower Copper River. After incorporating the non-neutral SNPs into the Chinook salmon barrier analysis, no additional ‘adaptive barriers’ were identified.

## **Discussion**

Under an isolation by distance model, the genetic differences among populations are inversely related to the exchange of genetic information, which depends largely on the geographic proximity between populations. Decreases in the exchange of genetic information owing to isolation by distance can result from two factors: 1) individuals are more likely to exchange genetic material with geographically proximate populations than with more distant populations, and 2) barriers limiting gene flow in particular locations (Dupanloup et al. 2002; Guillot et al. 2009). In Pacific salmon, approximately 95-99% of individuals home to their natal spawning habitat as adults; the remainder are termed ‘strays’ and may reproduce with non-natal populations (Quinn 2005). The location that

strays return to spawn can be largely determined by the proximity of their natal river (Quinn 1993; Unwin and Quinn 1993), meaning strays are more likely to reproduce with proximate populations than with more distant populations. Further, the presence of in-stream velocity or ecological barriers has been shown to hinder the exchange of genetic material between proximate populations in Pacific salmon (Ramstad et al. 2004; Lin et al. 2008). For sockeye and Chinook salmon in the Copper River, the observed isolation by distance pattern appeared to result from a combination of spatial patterns of straying and the presence of barriers to gene flow. Along a continuum, populations tended to be more similar to proximate populations than they were to more distant populations, suggesting that the exchange of genetic information was more common in geographically proximate populations. Also, in each species abrupt changes in allele frequencies were also observed in geographically proximate populations, suggestive of barriers. In either regard, it appears that the spatial distribution of populations in the Copper River largely determines the distribution of hierarchical genetic variation in both species.

Three barriers in particular identified in the sockeye salmon BARRIER analysis show areas of abrupt changes in allele frequencies in proximate populations. For example, the BARRIER analysis detected a barrier (100% bootstrap support) downstream of the sockeye salmon population in Fish Creek (8) in the Gulkana River (Figure 2.1). The barrier was corroborated by the SAMOVA analysis; the Fish Creek population formed its own distinct population group (Table 2.3). The Fish Creek population occurs in close proximity (~ 13 km) to two downstream populations, the East Fork Gulkana River – early (9) population and East Fork Gulkana River – late (10) population that were genetically similar to each other. However, the Fish Creek population appeared to be genetically distinct suggesting that very little gene flow occurs between the two groups despite their close proximity. Similarly, a barrier (94% bootstrap support) was identified below the Long Lake (20) sockeye salmon population. The Long Lake population formed its own distinct population group (Table 2.3), despite the fact that the population spawns between two other populations in the Chitina River, each of which were identified as part of the large middle Copper River population group by SAMOVA. The Long Lake population is unique in that individuals reach the spawning grounds in August and September, similar to many other populations throughout the middle Copper River, but

have the longest known spawning duration (August through April) of any sockeye salmon population in North America, spawning through the winter and into spring during large returns ([www.nps.gov/wrst/parkmgmt/long-lake-fish-weir.htm](http://www.nps.gov/wrst/parkmgmt/long-lake-fish-weir.htm)). In Pacific salmon, differences in spawning timing between populations can be a major contributing factor in determining genetic population structure (O'Malley et al. 2007; Creelman et al. In Review). The genetic distinction of the Long Lake population is likely a result of its temporal separation from surrounding populations. This is in contrast to the general spatial pattern observed throughout the Copper River and highlights the role that differences in spawning timing can play in shaping genetic structure on a local scale. Finally, the Chinook salmon BARRIER analysis detected a barrier (97% bootstrap support) downstream of the Mendeltna Creek population (9) that occurs in close proximity (~ 12 km) to the Kaina Creek population (10). The Kaina Creek population was in a group that included two populations from the adjacent Klutina and Tonsina rivers according to the SAMOVA analysis. The above mentioned barriers highlight the contribution of in-stream barriers, including 'temporal barriers', to the observed isolation by distance pattern in each species.

Both sockeye and Chinook salmon exhibited a similar inland-to-coastal gradient of within-population diversity; populations in the upper Copper River had less within-population diversity than populations in the middle and lower Copper River. In particular, populations >375 km upriver had a marked decrease in allelic richness (Figure 2.5) and generally exhibited lower expected heterozygosity (Table 2.1). The transition in allelic richness observed in both species roughly coincided with the confluence of the Tazlina and Copper rivers where the Copper River drainage transitions from a primarily tundra environment in the upper drainage to a more forested, mountainous environment in the middle/lower Copper River. An example of the influence of this transition area are the sockeye (9) and Chinook (11) salmon populations that spawn in Mendeltna Creek within the Tazlina River drainage; each of these populations formed its own discrete population group in the SAMOVA analyses. Interestingly, at lower values of k, the sockeye salmon population in Mendeltna Creek grouped with populations from the middle/lower Copper River whereas the Chinook salmon population grouped with populations in the upper Copper River. Results from the among-population analyses conducted in SAMOVA and

BARRIER support the observed transition area, particularly in sockeye salmon. Sockeye salmon populations above the transition area exhibited greater hierarchical population structure and four of the five barriers identified using neutral SNPs were located in the upper Copper River. Populations of both species exhibited decreased within-population diversity above the transition area.

The lesser within-population diversity of salmon populations in Mendeltna Creek and the upper Copper River may be attributed to two factors: 1) populations throughout the upper Copper River tend to be smaller and more geographically isolated, promoting less gene flow among populations and greater influence from genetic drift, and 2) founder effects. Above the transition area, populations tend to be geographically isolated relative to populations below (Wade et al. 2008) promoting decreased gene flow and increased genetic drift. However, historic processes contributing to founder effects also likely contribute to the decreased within-population diversity. In chapter 1, it was postulated that sockeye salmon populations in the upper Copper River may have been founded by individuals that persisted in a northern glacial refuge (Lindsey and McPhail 1986). During the height of the most recent glaciations, known as the McConnell/McCauley Glaciation throughout the Yukon Territory and Alaska, ice-free areas persisted north of the ice in regions near the present day upper Alsek, Copper, Mackenzie and Yukon river drainages (Smith et al. 2001). Although it is unclear whether sockeye and Chinook salmon occupied the region during the glaciations, Smith et al. (2001) provided evidence that some coho salmon (*Oncorhynchus kisutch*) populations may have persisted in the region in small numbers. If sockeye and Chinook salmon individuals were able to persist in the same region, populations in the upper Copper River may have been founded by individuals from the northern glacial refuge as the ice receded. Similarly, as ice receded north, populations in the upper Copper River may have been founded by strays from the middle/lower Copper River or coastal populations. In either regard, the observed decrease in within-population diversity in salmon populations in the upper Copper River is likely a result of both contemporary (gene flow and genetic drift) and historical (founder effects) processes.

STREAMTREE (Kalinowski et al. 2008) analyses conducted by Seeb et al. (2009b) for Chinook salmon in the Copper River supported the BARRIER results

reported in this study. Seeb et al. (2009b) used STREAMTREE to map Cavalli-Sforza and Edwards (CSE; 1967) chord distances onto sections of the Copper River drainage connecting populations within the drainage. The two stream reaches identified in the STREAMTREE analysis as having the greatest genetic discontinuities were identified by BARRIER (barriers 1 and 4) with greater than 50% bootstrap support. Further, the remaining two barriers identified by BARRIER (barriers 2 and 3) were also located in areas identified by STREAMTREE as having increased chord distances relative to surrounding stream reaches. Finally, STREAMTREE noted increased chord distances in the transition area above the confluence of the Tazlina and Copper rivers; reaches of the Copper River above the Tazlina and the lower Gulkana River each showed genetic discontinuities. Both STREAMTREE and BARRIER proved to be useful for identifying barriers to gene flow in the Copper River drainage.

Incorporating SNPs exhibiting adaptive variation into the SAMOVA and BARRIER analyses for sockeye salmon identified interesting adaptive differences in the Klutina River drainage. When using exclusively neutral SNPs, the sockeye salmon population that spawns in the mainstem Klutina River below the outlet of Klutina Lake (17) grouped with populations throughout the middle Copper River which included five populations that spawn in tributaries above Klutina Lake. However, when the SNPs exhibiting adaptive variation were incorporated into the analyses, the mainstem Klutina River population changed groups and joined three populations from the Gulkana River drainage (Figure 2.1). Additionally, BARRIER detected an additional ‘adaptive barrier’ above this population. The adaptive differences observed between the population below Klutina Lake and the five populations that spawn in tributary habitats above Klutina Lake are primarily due to allele frequency differences observed at *One\_MHC2\_190* and *One\_MHC2\_251* (Chapter 1). Interestingly, the Klutina Lake outlet population is of the sea/river ecotype as defined by Wood et al. (2008) whereas the five populations spawning above Klutina Lake are of the lake ecotype. The lake ecotype is the typical form of sockeye salmon that spends about half its life in a nursery lake before emigrating to the marine environment to mature, whereas the sea/river ecotype is a rarer form that rears in the freshwater environment for a much shorter and more variable period of time. Large differences in MHC allele frequencies between geographically proximate populations of

the two ecotypes have been observed elsewhere throughout the range of sockeye salmon (Creelman et al. In Review; McGlaufflin et al. In Review). The two sockeye salmon MHC SNPs are found in one exon (*One\_MHC2\_190*) and one intron (*One\_MHC2\_251*) within a Major Histocompatibility Complex (MHC) class II gene (Miller and Withler 1996; Miller et al. 2001). Several studies support the adaptive nature of MHC polymorphisms related to pathogen-mediated selection (Miller et al. 2001; Dionne et al. 2009; Gomez-Uchida et al. In Review). Differences in the spawning and rearing habitats (i.e. temperature regimes, flow velocity) between the two ecotypes may provide a mechanism for the adaptive differences observed among populations occurring in the Klutina River drainage.

The description of conservation units has previously been criticized for the overemphasis of reproductive isolation based on information from loci assumed to be neutral and the under-emphasis of adaptive variation among populations (Allendorf and Luikart 2007). The new field of landscape genomics aims to include large numbers of loci exhibiting both neutral and adaptive variation to not only make inferences regarding gene flow and drift, but also to tease apart information regarding evolutionary forces and local adaptation influencing natural populations (Schwartz et al. 2009). Although we do not suggest that the methods presented here constitute a landscape genomics analysis, we do emphasize that the adaptive differences observed among the Klutina Lake sockeye salmon populations highlights the role of local adaptation in determining genetic variation. In the past, landscape geneticists have largely viewed loci under selection as nuisances that should be avoided (Schwartz et al. 2009), but we show that incorporating adaptive loci into landscape genetics analyses can provide valuable insight into adaptive processes. Variation at adaptive loci should be recognized when planning conservation efforts.

We identified largely congruent patterns of spatial genetic structure between sockeye and Chinook salmon in the Copper River, highlighting the role that landscape heterogeneity plays in determining spatial patterns of genetic variation among the closely related species. On a large scale, both species exhibited significant patterns of isolation by distance demonstrating that the spatial distribution of populations was the primary determinant of the genetic structure in each species. Further, inland populations of both

species exhibited decreased within-population diversity resulting from a combination of reproductive isolation and historical processes including founder effects. On a smaller scale, temporal differences observed in the Chitina River sockeye salmon populations and adaptive differences observed in the Klutina River sockeye salmon populations highlighted the role that differences in spawning timing and local adaptation can play in determining genetic structure. Our results demonstrate that including genetic divergence at adaptive loci together with divergence at neutral loci can enhance and help determine priorities when identifying units of conservation.

### **Acknowledgements**

This project was a collaborative effort between the University of Washington, School of Aquatic and Fishery Sciences (UW SAFS) and the Alaska Department of Fish and Game (ADFG), Gene Conservation Laboratory. We are grateful for laboratory contributions from staff at UW SAFS and ADFG. This report was partially funded by the Alaska Sustainable Salmon Fund under Study #45869 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, administered by the Alaska Department of Fish and Game. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, the U.S. Department of Commerce, or the Alaska Department of Fish and Game. Additional project funding was provided by the Gordon and Betty Moore Foundation and the State of Alaska. Funding for MWA was provided by the H. Mason Keeler Endowment for Excellence (UW).

Table 2.1. Years sampled, sample size ( $n$ ), allelic richness (AR), and expected heterozygosities ( $H_E$ ) for sockeye and Chinook salmon populations from the Copper River, southcentral Alaska. Chinook salmon data adapted from Seeb et al. (2009b).

ID	Population	Years Sampled	$n$	AR	$H_E$
<i>Sockeye Salmon Populations</i>					
1	Bad Crossing #2	2009	114	1.70	0.213
2	Bone Cr.	2005, 2008	56	1.75	0.228
3	Mentasta Lk.	2008	197	1.72	0.236
4	Tanada Lk. - outlet	2009	95	1.72	0.249
5	Tanada Lk. - east shore	2009	95	1.75	0.246
6	Keg Cr.	2009	66	1.67	0.226
7	Swede Lk.	2008	200	1.64	0.203
8	Fish Ck.	2008	169	1.61	0.156
9	EF Gulkana R. - early broodstock	2008, 2009	152	1.70	0.243
10	EF Gulkana R. - late broodstock	2008, 2009	190	1.72	0.241
11	Mendeltna Cr.	2008, 2009	189	1.67	0.215
12	1884 Lk.	2008	82	1.88	0.255
13	Bear Hole	2008	143	1.84	0.259
14	Klutina Lk. inlet	2008, 2009	100	1.86	0.266
15	St. Anne Cr.	2005, 2008	299	1.87	0.263
16	Mahlo R.	2008	192	1.88	0.256
17	Klutina Lk. - outlet	2008, 2009	159	1.90	0.270
18	Tonsina Lk. - outlet	2009	154	1.91	0.271
19	Bear Island	2009	30	1.88	0.271
20	Long Lk.	2005	189	1.74	0.226
21	Tebay R.	2008	197	1.91	0.274
22	Salmon Cr. (Bremner R.)	2008	99	1.89	0.256
23	Steamboat Lk.	2008	177	1.89	0.258
24	Clear Ck. at 40-mile	2007, 2008	188	1.92	0.258
25	McKinley Lk.	2007	94	1.85	0.233
26	Salmon Cr. (McKinley Lk.)	2007	94	1.88	0.238
27	Tokun Lk.	2008, 2009	190	1.83	0.234
28	Martin Lk.	2007, 2008	190	1.86	0.243
<i>Chinook Salmon Populations</i>					
1	Bone Cr.	2004, 2005	78	1.69	0.227
2	Otter Cr.	2005	126	1.77	0.241
3	Indian Cr.	2004, 2005	49	1.70	0.231
4	EF Chistochina R.	2004	132	1.74	0.227

Table 2.1. Continued.

ID	Population	Years Sampled	<i>n</i>	AR	H <sub>E</sub>
5	Sinona Creek	2004, 2005	159	1.65	0.212
6	MF Gulkana R.	2004	76	1.76	0.241
7	PF Gulkana R.	2004	88	1.80	0.239
8	Gulkana R.	2004	46	1.82	0.256
9	Mendeltna Cr.	2004	144	1.83	0.270
10	Kaina Cr.	2004	75	1.83	0.256
11	Manker Cr.	2004, 2005	62	1.89	0.274
12	Little Tonsina R.	2004, 2006	81	1.88	0.258
13	Tebay R.	2004, 2005, 2006	68	1.83	0.230

Table 2.2. Summary of 45 SNPs from 28 sockeye salmon populations and 41 SNPs from 13 Chinook salmon populations from the Copper River, southcentral Alaska. Summary statistics include minor allele frequency range, expected heterozygosity ( $H_E$ ), and Weir and Cockerham (1984)  $F_{ST}$ . A (+) in the comments indicates a candidate for diversifying selection. Chinook salmon SNP data adapted from Seeb et al. (2009b).

Marker	Minor Allele Frequency Range	$H_E$	$F_{ST}$	Comments	Reference <sup>f</sup>
<i>Sockeye salmon single nucleotide polymorphisms</i>					
<i>One_ACBP-79</i>	0.085 - 0.643	0.440	0.093		2
<i>One_ALDOB-135</i>	0.030 - 0.606	0.334	0.118		2
<i>One_COI<sup>b</sup></i>	0.000 - 0.308	-	0.115	mtDNA	2
<i>One_ctgf-301</i>	0.000 - 0.114	0.029	0.040		2
<i>One_Cytb_17</i>	0.000 - 0.543	-	0.229	mtDNA	2
<i>One_Cytb_26<sup>b</sup></i>	0.000 - 0.543	-	0.126	mtDNA	2
<i>One_E2-65</i>	0.027 - 0.716	0.362	0.088		3
<i>One_GHII-2165</i>	0.006 - 0.860	0.352	0.253	+	2
<i>One_GPDH-201<sup>c</sup></i>	0.092 - 0.539	0.400	0.074	LD	3
<i>One_GPDH2-187</i>	0.042 - 0.611	0.368	0.117	LD	3
<i>One_GPH-414</i>	0.008 - 0.562	0.248	0.092		2
<i>One_hsc71-220</i>	0.111 - 0.616	0.421	0.066		2
<i>One_HGFA-49</i>	0.000 - 0.321	0.188	0.060		3
<i>One_HpaI-71</i>	0.143 - 0.801	0.434	0.094		2
<i>One_HpaI-99</i>	0.000 - 0.379	0.212	0.094		2
<i>One_IL8r-362</i>	0.000 - 0.103	0.023	0.042		1
<i>One_KPNA-422</i>	0.056 - 0.391	0.263	0.049		2
<i>One_LEI-87</i>	0.003 - 0.347	0.268	0.069		2
<i>One_MARCKS-241</i>	0.000 - 0.096	0.052	0.040		1
<i>One_MHC2_190</i>	0.000 - 0.946	0.245	0.365	+	2
<i>One_MHC2_251</i>	0.009 - 0.962	0.307	0.380	+	2
<i>One_Ots213-181</i>	0.135 - 0.775	0.401	0.093		2
<i>One_p53-534</i>	0.000 - 0.272	0.085	0.086		2
<i>One_ins-107</i>	0.108 - 0.803	0.437	0.143		3
<i>One_Prl2</i>	0.036 - 0.907	0.451	0.105		2
<i>One_RAG1-103</i>	0.000 - 0.063	0.028	0.021		2
<i>One_RAG3-93</i>	0.000 - 0.194	0.129	0.059		2

Table 2.2. Continued.

Marker	Minor Allele Frequency Range	H <sub>E</sub>	F <sub>ST</sub>	Comments	Reference <sup>f,g</sup>
<i>One_RFC2-102</i>	0.039 - 0.568	0.332	0.103		3
<i>One_RFC2-285</i>	0.000 - 0.297	0.154	0.078		3
<i>One_RH2op-395</i>	0.000 - 0.088	0.007	0.070		2
<i>One_serpin-75</i>	0.000 - 0.158	0.076	0.034		3
<i>One_STC-410</i>	0.000 - 0.414	0.225	0.119		2
<i>One_STR07</i>	0.037 - 0.661	0.338	0.133		2
<i>One_Tf_ex11-750</i>	0.077 - 0.769	0.422	0.130		2
<i>One_Tf_in3-182</i>	0.000 - 0.103	0.031	0.040		2
<i>One_U301_92</i>	0.018 - 0.638	0.288	0.123		2
<i>One_U401-224</i>	0.090 - 0.758	0.437	0.089		1
<i>One_U404-229</i>	0.006 - 0.229	0.123	0.055		1
<i>One_U502-167</i>	0.000 - 0.113	0.034	0.045		1
<i>One_U503-170</i>	0.000 - 0.326	0.118	0.099		1
<i>One_U504-141</i>	0.069 - 0.559	0.366	0.113		1
<i>One_U508-533</i>	0.000 - 0.260	0.136	0.068		1
<i>One_VIM-569</i>	0.024 - 0.298	0.228	0.047		2
<i>One_ZNF-61</i>	0.024 - 0.703	0.344	0.129		1
<i>One_Zp3b-49</i>	0.047 - 0.709	0.381	0.209		3
<i>Chinook salmon single nucleotide polymorphisms</i>					
<i>Ots_AsnRS-60</i>	0.260 - 0.654	0.464	0.069		3
<i>Ots_C3N3</i>	0.000 - 0.076	-	0.063	mtDNA	4
<i>Ots_E2-275</i>	0.003 - 0.230	0.200	0.043		3
<i>Ots_ETIF1A</i>	0.104 - 0.408	0.345	0.048		1
<i>Ots_FARSLA-220</i>	0.022 - 0.256	0.206	0.046		2
<i>Ots_FGF6A</i>	0.000 - 0.439	0.291	0.143	LD	1
<i>Ots_FGF6B<sup>d</sup></i>	0.000 - 0.243	0.134	0.069	LD	-
<i>Ots_GH2</i>	0.288 - 0.559	0.454	0.027		4
<i>Ots_GnRH-271</i>	0.000 - 0.065	0.038	0.027		3
<i>Ots_GPDH-338</i>	0.000 - 0.033	0.012	0.012		3
<i>Ots_GPH318</i>	0.028 - 0.423	0.249	0.114		2
<i>Ots_GST-207</i>	0.000 - 0.073	0.030	0.036		2
<i>Ots_GTH2B_550</i>	0.276 - 0.702	0.459	0.082		1
<i>Ots_hnRNPL-533</i>	0.000 - 0.298	0.204	0.062		2
<i>Ots_HSP90B-100</i>	0.000 - 0.073	0.051	0.022		2
<i>Ots_HSP90B-385<sup>e</sup></i>	0.000 - 0.023	0.011	0.008		2
<i>Ots_IGF-I.1-76</i>	0.235 - 0.626	0.468	0.060		3

Table 2.2 Continued.

Marker	Minor Allele Frequency Range	H <sub>E</sub>	F <sub>ST</sub>	Comments	Reference <sup>f,g</sup>
<i>Ots_il-Iracp-166</i>	0.138 - 0.604	0.437	0.112		3
<i>Ots_ins-115</i>	0.003 - 0.167	0.108	0.049		3
<i>Ots_LWSop-638</i>	0.000 - 0.171	0.119	0.061		3
<i>Ots_MetA</i>	0.000 - 0.102	0.052	0.040		-
<i>Ots_MHC1</i>	0.148 - 0.757	0.386	0.120	LD	5
<i>Ots_MHC2</i>	0.000 - 0.758	0.210	0.431	LD, +	5
<i>Ots_NOD1</i>	0.176 - 0.381	0.422	0.016		1
<i>Ots_Ots2</i>	0.000 - 0.025	0.009	0.020		4
<i>Ots_P450</i>	0.304 - 0.538	0.473	0.019		4
<i>Ots_P53</i>	0.255 - 0.661	0.460	0.054		4
<i>Ots_PGK54</i>	0.042 - 0.346	0.194	0.056		1
<i>Ots_Prl2</i>	0.153 - 0.500	0.414	0.065		4
<i>Ots_RAG3</i>	0.000 - 0.170	0.119	0.063		1
<i>Ots_S7-1</i>	0.235 - 0.712	0.470	0.049		1
<i>Ots_SClkF2R2-135</i>	0.269 - 0.557	0.469	0.020		3
<i>Ots_SERPC1-209</i>	0.007 - 0.131	0.122	0.015		2
<i>Ots_SL</i>	0.258 - 0.616	0.474	0.040		4
<i>Ots_SWS1op-182</i>	0.313 - 0.486	0.477	0.012		3
<i>Ots_Tnsf</i>	0.067 - 0.254	0.285	0.017		4
<i>Ots_u202-161</i>	0.033 - 0.276	0.219	0.057		3
<i>Ots_U212-158</i>	0.000 - 0.121	0.061	0.074		3
<i>Ots_u4-92</i>	0.013 - 0.141	0.100	0.019		3
<i>Ots_u6-75</i>	0.007 - 0.205	0.115	0.040		3
<i>Ots_Zp3b-215</i>	0.000 - 0.066	0.024	0.038		3

<sup>a</sup>Dropped from analyses because monomorphic (< 0.02 minor allele frequency in all populations);

<sup>b</sup>Mitochondrial DNA SNPs dropped from analyses in which combining mtDNA SNPs into a composite haplotype was inappropriate; <sup>c</sup>Linkage disequilibrium between the GPDH loci was found to be significant and this locus was dropped during analyses; <sup>d</sup>Linkage disequilibrium between the FGF6 loci was found to be significant and this locus was dropped during analyses; <sup>e</sup>Linkage disequilibrium between the HSP90B loci was found to be significant and this locus was dropped during analyses; <sup>f</sup>Sockeye reference: (1) Habicht et al. 2010, (2) Elfstrom et al. 2006, (3) Smith et al. 2005a; <sup>g</sup>Chinook reference: (1) Narum et al. 2008, (2) Smith et al. 2007, (3) Smith et al. 2005a, (4) Smith et al. 2005b

Table 2.3. AMOVA results for populations when grouped to minimize within-group variation ( $F_{SC}$ ) and to maximize among group variation ( $F_{CT}$ ) using SAMOVA v1.0. Analyses for groupings 2a through 10a included candidate markers for diversifying selection. The bold rows indicate the groupings where  $F_{SC}$  declined markedly and  $F_{CT}$  began to plateau and are the groupings depicted in Figures 2.1 and 2.2. The numbers in each group indicate the population ID in Table 1.1.

Species	Groups	Group Composition	$F_{ST}$	$F_{SC}$	$F_{CT}$
sockeye	2	[1-3,6-28] [4-5]	0.175	0.097	0.086
	3	[1-3,6-19,21-28] [4-5] [20]	0.168	0.090	0.086
	4	[1-3,6,8-19,21-28] [4-5] [7] [20]	0.158	0.084	0.082
	5	[1-3,6,9-19,21-28] [4-5] [7] [8] [20]	0.152	0.079	0.080
	6	[1-3,6,9-10,12-19,21-28] [4-5] [7] [8] [11] [20]	0.146	0.074	0.078
	7	[1-3] [4-5] [6,9-19,21] [7] [8] [20] [22-28]	0.125	0.049	0.080
	8	[1-3] [4-5] [6,9-10] [7] [8] [11-19,21] [20] [22-28]	0.120	0.038	0.085
	<b>9</b>	<b>[1-3] [4-5] [6,9-10] [7] [8] [11] [12-19,21] [20] [22-28]</b>	<b>0.118</b>	<b>0.031</b>	<b>0.090</b>
	10	[1-3] [4-5] [6,9-10] [7] [8] [11] [12-16,18,21] [17,19] [20] [22-28]	0.116	0.028	0.090
	2a	[1-19,21-28] [20]	0.200	0.119	0.092
	3a	SAME AS NEUTRAL	0.198	0.109	0.100
	4a	[1-3,6-18,21-28] [4-5] [19] [20]	0.192	0.110	0.092
	5a	[1-3] [4-5] [6-19,21] [20] [22-28]	0.156	0.072	0.090
	6a	[1-3] [4-5] [6-7,9-19,21] [8] [20] [22-28]	0.153	0.067	0.093
	7a	SAME AS NEUTRAL	0.150	0.059	0.097
	8a	[1-3] [4-5] [6,9-10,12-19,21] [7] [8] [11] [20] [22-28]	0.148	0.051	0.101
	<b>9a</b>	<b>[1-3] [4-5] [6,9-10,17] [7] [8] [11,19] [12-16,18,21] [20] [22-28]</b>	<b>0.140</b>	<b>0.039</b>	<b>0.106</b>
10a	[1-3] [4-5] [6,9-10] [7] [8] [11,19] [12-16,18,21] [17] [20] [22-28]	0.140	0.035	0.109	

Table 2.3. Continued.

Species	Groups	Group Composition	F <sub>ST</sub>	F <sub>SC</sub>	F <sub>CT</sub>
Chinook	2	[1-5] [6-13]	0.075	0.031	0.045
	3	[1-5] [6-12] [13]	0.074	0.025	0.050
	4	[1-5] [6-8] [9-12] [13]	0.068	0.020	0.049
	5	[1-5] [6-8] [9] [10-12] [13]	0.066	0.017	0.050
	<b>6</b>	<b>[1-4] [5] [6-8] [9] [10-12] [13]</b>	<b>0.062</b>	<b>0.010</b>	<b>0.052</b>
	7	[1] [2-4] [5] [6-8] [9] [10-12] [13]	0.060	0.008	0.053
	8	[1] [2] [3-4] [5] [6-8] [9] [10-12] [13]	0.058	0.003	0.056
	9	[1] [2] [3-4] [5] [6-8] [9] [10,12] [11] [13]	0.057	0.002	0.056
	10	[1] [2] [3-4] [5] [6,8] [7] [9] [10,12] [11] [13]	0.057	0.001	0.056
	2a	[1-9] [10-13]	0.100	0.044	0.059
	3a	[1-5] [6-9] [10-13]	0.085	0.024	0.063
	4a	[1-5] [6-9] [10-12] [13]	0.084	0.020	0.065
	5a	SAME AS NEUTRAL	0.081	0.017	0.066
	<b>6a</b>	SAME AS NEUTRAL	<b>0.076</b>	<b>0.010</b>	<b>0.066</b>
	7a	SAME AS NEUTRAL	0.073	0.008	0.066
	8a	SAME AS NEUTRAL	0.071	0.003	0.068
	9a	SAME AS NEUTRAL	0.070	0.003	0.068
	10a	[1] [2] [3-4] [5] [6-8] [9] [10] [11] [12] [13]	0.070	0.002	0.068

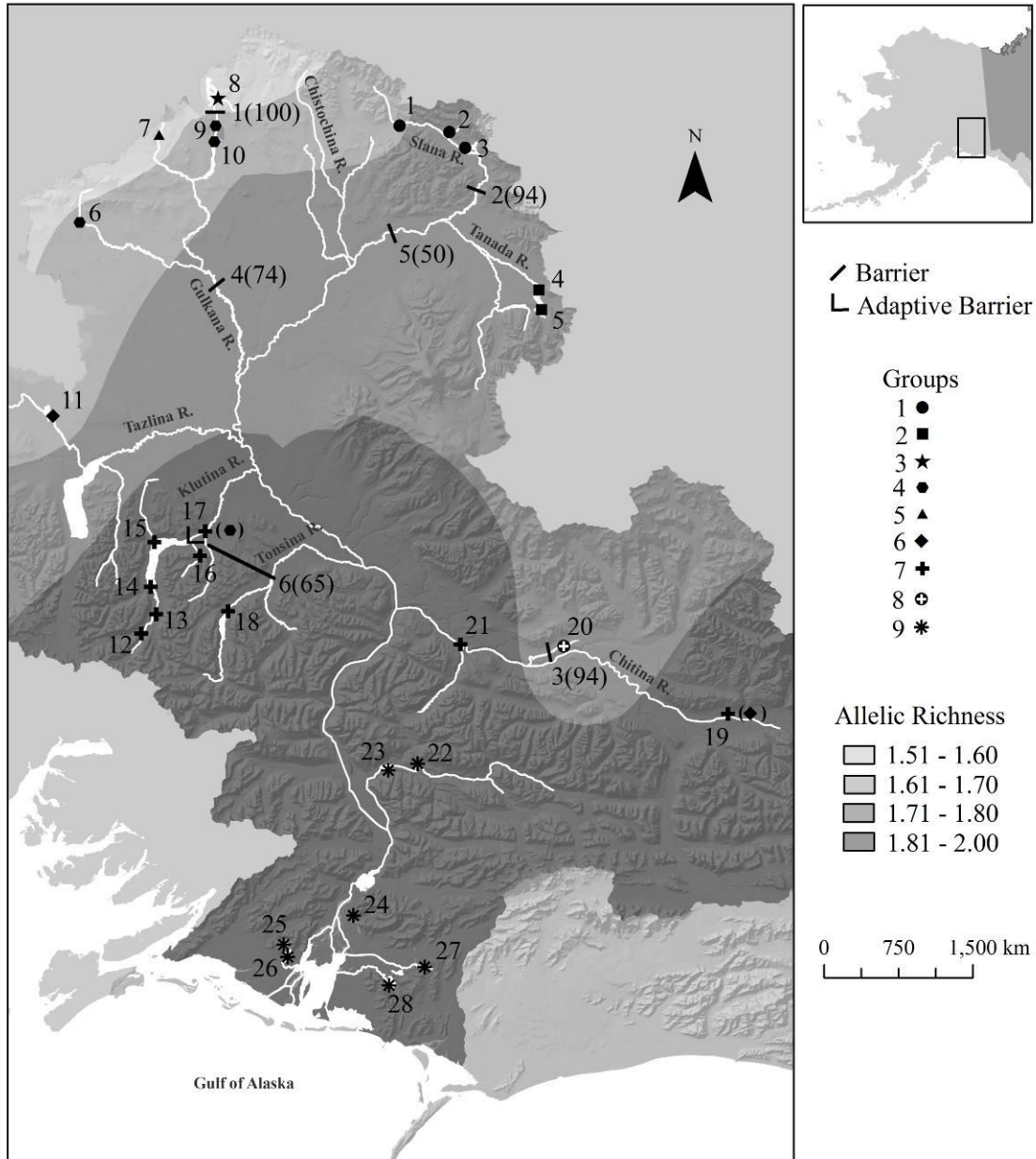


Figure 2.1. Gradient of allelic richness, population groups, and inferred barriers for sockeye salmon from the Copper River, southcentral Alaska. Population symbols indicate the groupings defined by SAMOVA. Barriers indicate abrupt changes in allele frequencies transferred from the Delaunay triangulation in BARRIER to approximate locations on the map. Numbers inside the parentheses indicate the robustness of each barrier based on 100 bootstrap samples. The “Adaptive Barrier” observed above population #17 indicates the additional barrier that was observed when including the candidate SNPs for diversifying selection. The population symbols in parentheses for populations #17 and #19 indicate the groupings that were determined when including the candidate markers for diversifying selection.

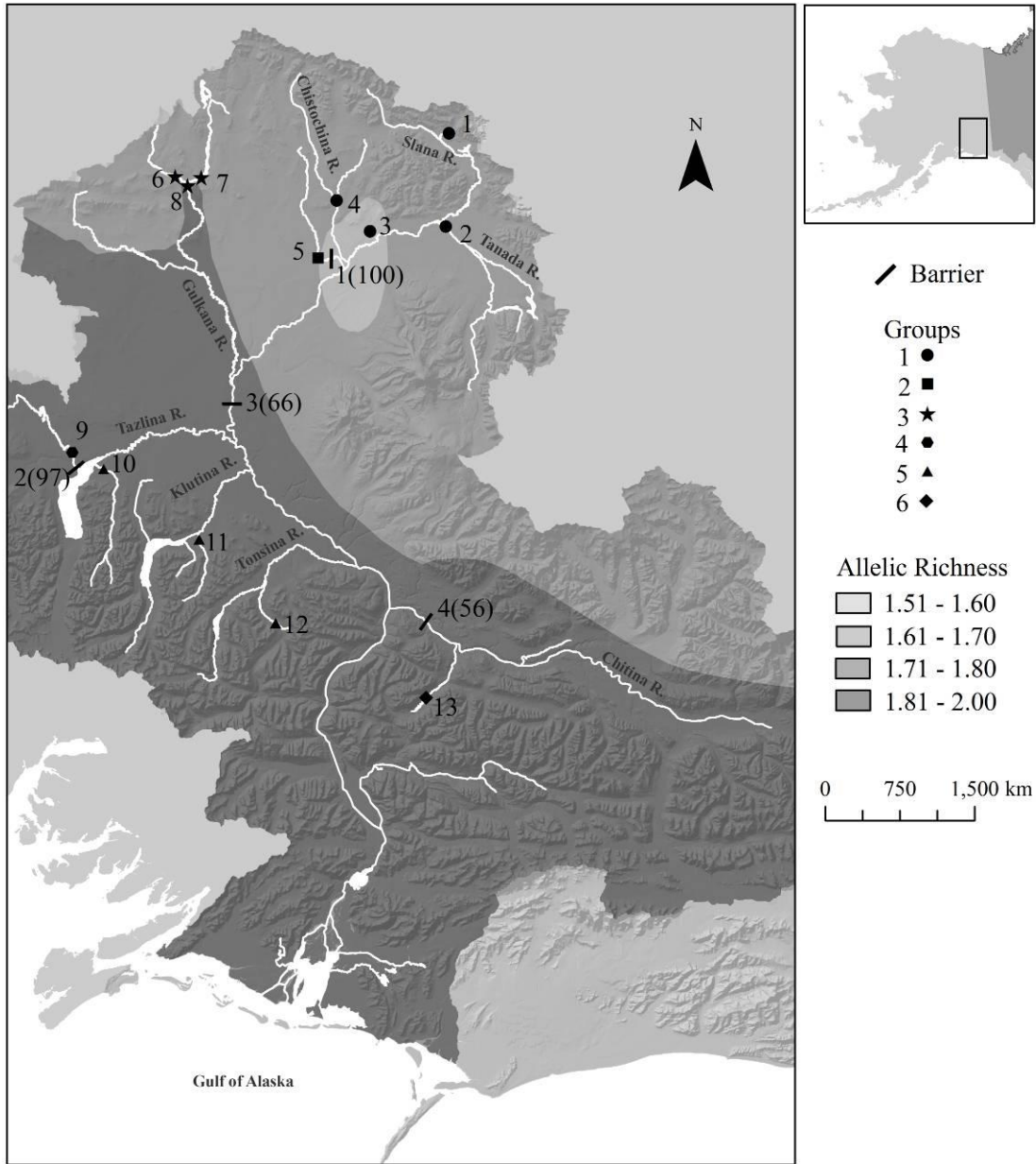


Figure 2.2. Gradient of allelic richness, population groups, and inferred barriers for Chinook salmon from the Copper River, southcentral Alaska. Population symbols indicate the groupings defined by SAMOVA. Barriers indicate abrupt changes in allele frequencies transferred from the Delaunay triangulation in BARRIER to approximate locations on the map. Numbers inside the parentheses indicate the robustness of each barrier based on 100 bootstrap samples.

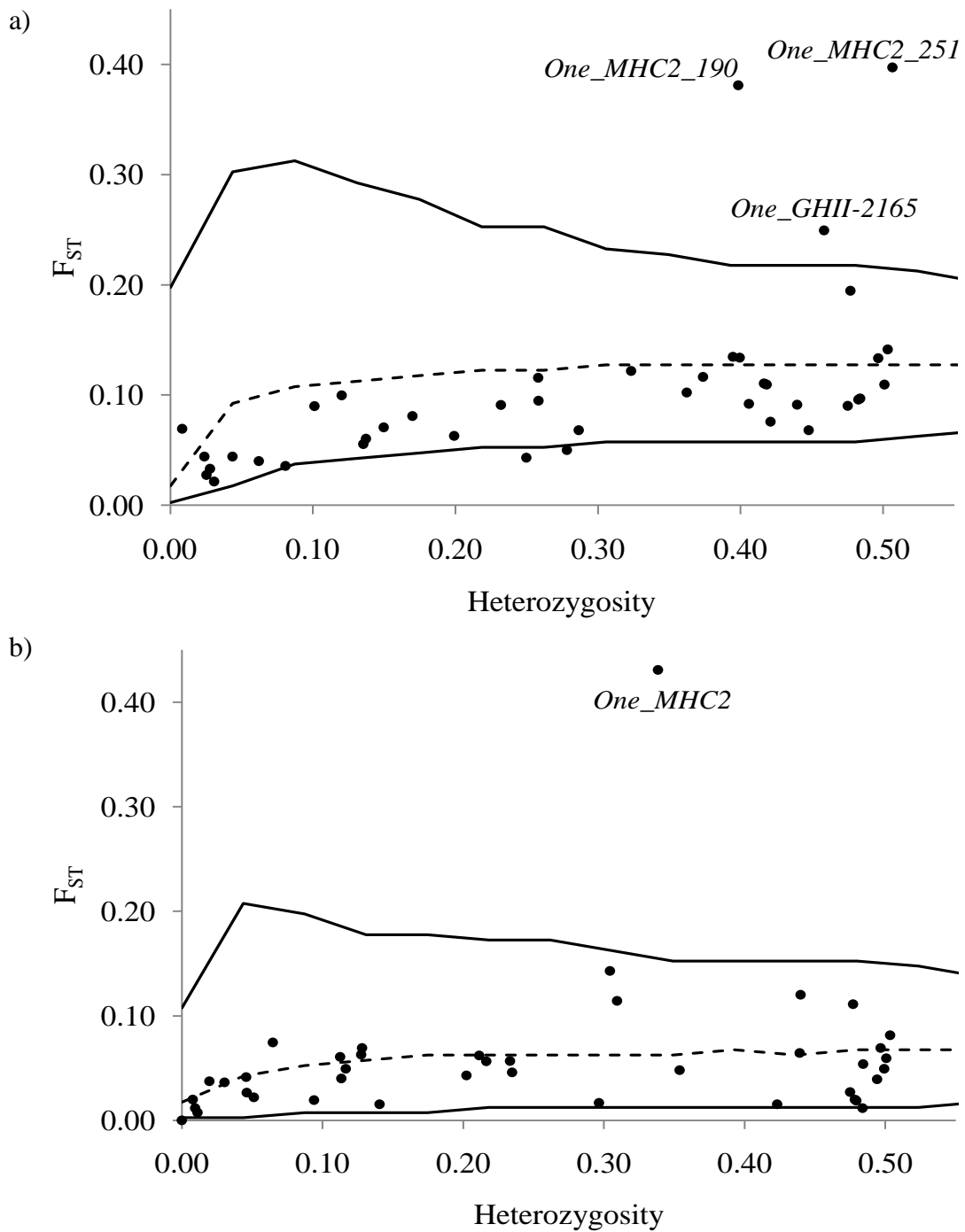


Figure 2.3.  $F_{ST}$  as a function of heterozygosity (ARLEQUIN v3.5; Excoffier and Lischer 2010) for a) 42 sockeye salmon SNPs and b) 41 Chinook salmon SNPs. The dashed line represents the median and the solid lines represent the 99% confidence interval boundaries based on coalescent simulations and using a finite island model. Loci labeled above the upper 99% confidence interval boundary were determined as candidates for diversifying selection.

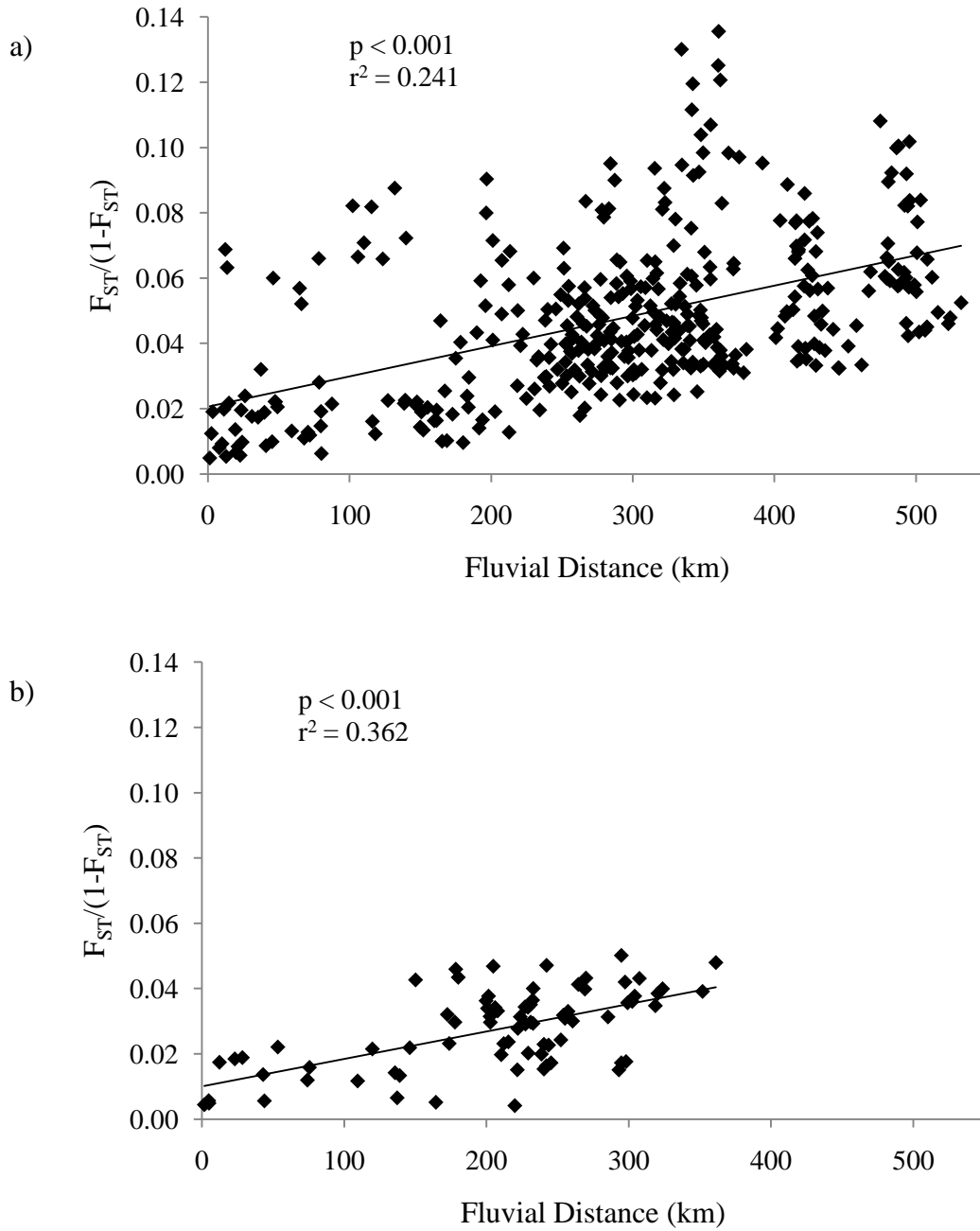


Figure 2.4. Pairwise genetic divergence among populations as a function of fluvial distance for a) sockeye salmon and b) Chinook salmon from the Copper River, southcentral Alaska. Chinook salmon data adapted from Seeb et al. (2009b).

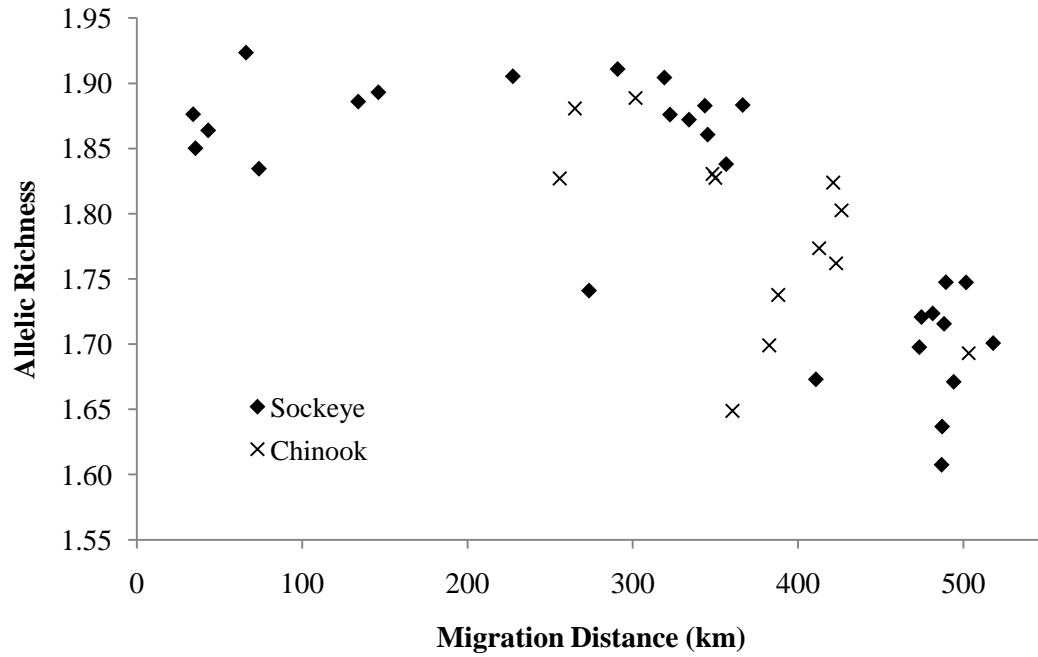


Figure 2.5. Allelic richness as a function of upstream migration distance for sockeye and Chinook salmon populations from the Copper River, southcentral Alaska.

## References

- Aguilar, A. 2006. Analysis of F-ST outliers at allozyme loci in Pacific salmon: implications for natural selection. *Environmental Biology of Fishes* 76(2-4):329-339.
- Allendorf, F. W., and G. Luikart. 2007. *Conservation and the Genetics of Populations*. Blackwell, Oxford.
- Allendorf, F. W., and L. W. Seeb. 2000. Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution* 54(2):640-651.
- Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172(4):2567-2582.
- Anderson, E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences* 65(7):1475-1486.
- Barbazuk, W. B., S. J. Emrich, H. D. Chen, L. Li, and P. S. Schnable. 2007. SNP discovery via 454 transcriptome sequencing. *Plant Journal* 51(5):910-918.
- Bartley, D. M., B. Bentley, P. G. Olin, and G. A. E. Gall. 1992. Population genetic-structure of coho salmon (*Oncorhynchus kisutch*) in California. *California Fish and Game* 78(3):88-104.
- Beacham, T. D., J. R. Candy, K. J. Supernault, T. Ming, B. Deagle, A. Schulze, D. Tuck, K. H. Kaukinen, J. R. Irvine, K. M. Miller, and R. E. Withler. 2001. Evaluation and application of microsatellite and major histocompatibility complex variation for stock identification of coho salmon in British Columbia. *Transactions of the American Fisheries Society* 130(6):1116-1149.
- Beacham, T. D., M. Lapointe, J. R. Candy, B. McIntosh, C. MacConnachie, A. Tabata, K. Kaukinen, L. T. Deng, K. M. Miller, and R. E. Withler. 2004. Stock identification of Fraser River sockeye salmon using microsatellites and major histocompatibility complex variation. *Transactions of the American Fisheries Society* 133(5):1117-1137.
- Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13(4):969-980.
- Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* 263(1377):1619-1626.

- Behnke, R. J., and J. R. Tomelleri. 2002. Trout and salmon of North America. Free Press, New York.
- Bernatchez, L., and C. Landry. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology* 16(3):363-377.
- Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18(5):249-256.
- Burgner, R. L. 1991. Life history of sockeye salmon (*Oncorhynchus nerka*). Pages 3-117 in C. Groot, and L. Margolis, editors. Pacific salmon life histories. University of British Columbia Press, Vancouver.
- Carney, B. L., A. K. Gray, and A. J. Gharrett. 1997. Mitochondrial DNA restriction site variation within and among five populations of Alaskan coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences* 54(4):940-949.
- Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis - models and estimation procedures. *Evolution* 21(3):550-&.
- Chakraborty, R., and O. Leimar. 1987. Genetic variation within a subdivided population. Pages 89-102 in N. Ryman, and F. M. Utter, editors. Population genetics and fishery management. University of Washington Press, Seattle, WA.
- Clements, M. D., H. L. Bart, and D. L. Hurley. 2004. Isolation and characterization of two distinct growth hormone cDNAs from the tetraploid smallmouth buffalofish (*Ictiobus bubalus*). *General and Comparative Endocrinology* 136(3):411-418.
- Creelman, E., L. Hauser, L. W. Seeb, R. Simmons, and W. D. Templin. In Review. Temporal and geographic genetic divergence: Characterizing sockeye salmon populations in the Chignik watershed, Alaska, using single nucleotide polymorphisms. *Transactions of the American Fisheries Society*.
- Debevec, E. M., R. B. Gates, M. Masuda, J. Pella, J. Reynolds, and L. W. Seeb. 2000. SPAM (Version 3.2): Statistics program for analyzing mixtures. *Journal of Heredity* 91(6):509-510.
- Delaunay, B. 1934. Sur la sphere vide. *Bulletin of the Academy of Sciences of the USSR* 7:793-800.

- Dionne, M., K. M. Miller, J. J. Dodson, and L. Bernatchez. 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society B-Biological Sciences* 364(1523):1555-1565.
- Dittman, A. H., and T. P. Quinn. 1996. Homing in Pacific salmon: Mechanisms and ecological basis. *Journal of Experimental Biology* 199(1):83-91.
- Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11(12):2571-2581.
- Elfstrom, C. M., C. T. Smith, and J. E. Seeb. 2006. Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon. *Molecular Ecology Notes* 6(4):1255-1259.
- Evans, M. L., and B. D. Neff. 2009. Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* 18(22):4716-4729.
- Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103(4):285-298.
- Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10(3):564-567.
- Fair, L. F., S. D. Moffitt, M. J. Evenson, and J. Erickson. 2008. Escapement goal review of Copper and Bering rivers, and Prince William Sound Pacific salmon stocks, 2008. Alaska Department of Fish and Game, Fishery Manuscript No. 08-02, Anchorage.
- Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180(2):977-993.
- Gagnon, M. C., and B. Angers. 2006. The determinant role of temporary proglacial drainages on the genetic structure of fishes. *Molecular Ecology* 15(4):1051-1065.
- Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457-511.
- Gelman, A. B., J. S. Carlin, H. S. Stern, and D. B. Rubin. 2000. Bayesian data analysis. Chapman and Hall, Boca Raton, Florida.

- Gilbert, C. H. 1913. Age at maturity of the Pacific coast salmon of the genus *Oncorhynchus*. Report of the British Columbia Commissioner of Fisheries 1912:57-70.
- Gomez-Uchida, D., M. J. Smith, C. Habicht, J. E. Seeb, T. P. Quinn, and L. W. Seeb. In Review. Single nucleotide polymorphisms mirror geographic and ecological complexity among Alaskan sockeye salmon populations.
- Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86(6):485-486.
- Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at <http://www.unil.ch/izea/software/fstat.html>.
- Guillot, G., R. Leblois, A. Coulon, and A. C. Frantz. 2009. Statistical methods in spatial genetics. *Molecular Ecology* 18(23):4734-4756.
- Habicht, C., L. W. Seeb, K. W. Myers, E. Farley, and J. E. Seeb. 2010. Summer-fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single nucleotide polymorphisms (SNPs). *Transactions of the American Fisheries Society* 139:1171-1191.
- Hard, J. J. 1995. A quantitative genetic perspective on the conservation of intraspecific diversity. Pages 304-326 in J. L. Nielsen, editor. *Evolution and the aquatic ecosystem: defining unique units in population conservation*. American Fisheries Society Symposium 17, Bethesda, MD.
- Hickey, D. A., and M. D. McLean. 1980. Selection for ethanol tolerance and ADH allozymes in natural populations of *Drosophila melanogaster*. *Genetical Research* 36(1):11-15.
- Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6(2):65-70.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54(2):187-211.
- Hyllner, S. J., L. Westerlund, P. E. Olsson, and A. Schopen. 2001. Cloning of rainbow trout egg envelope proteins: Members of a unique group of structural proteins. *Biology of Reproduction* 64(3):805-811.
- Isaaks, E. H., and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

- Kalinowski, S. T., M. H. Meeuwig, S. R. Narum, and M. L. Taper. 2008. Stream trees: a statistical method for mapping genetic differences between populations of freshwater organisms to the sections of streams that connect them. *Canadian Journal of Fisheries and Aquatic Sciences* 65(12):2752-2760.
- Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science* 220(4598):671-680.
- Koljonen, M. L., J. J. Pella, and M. Masuda. 2005. Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo salar*) catches from DNA microsatellite data. *Canadian Journal of Fisheries and Aquatic Sciences* 62(9):2143-2158.
- Langefors, A. H. 2005. Adaptive and neutral genetic variation and colonization history of Atlantic salmon, *Salmo salar*. *Environmental Biology of Fishes* 74(3-4):297-308.
- Langella, O. 2001. POPULATIONS (Populations genetic software): individuals or populations distances based on allelic frequencies, phylogenetic trees, file conversions. Available at [http://www.bioinformatics.org/project/?group\\_id=84](http://www.bioinformatics.org/project/?group_id=84). .
- Lin, J., T. P. Quinn, R. Hilborn, and L. Hauser. 2008. Fine-scale differentiation between sockeye salmon ecotypes and the effect of phenotype on straying. *Heredity* 101(4):341-350.
- Lindsey, C. C., and J. D. McPhail. 1986. Zoogeography of fishes of the Yukon and Mackenzie basins. Pages 639-674 in C. H. Hocutt, and E. O. Wiley, editors. *The zoogeography of North American freshwater fishes*. Wiley, New York.
- Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* 18(4):189-197.
- Manni, F., E. Guerard, and E. Heyer. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology* 76(2):173-190.
- Marten, A., M. Brandle, and R. Brandl. 2006. Habitat type predicts genetic population differentiation in freshwater invertebrates. *Molecular Ecology* 15(9):2643-2651.
- McGlaufflin, M. T., D. E. Schindler, L. W. Seeb, C. T. Smith, C. Habicht, and J. E. Seeb. In Review. Spawning habitat and geography influence population structure and juvenile migration timing of sockeye salmon in the Wood River lakes, Alaska. *Transactions of the American Fisheries Society*.

- Miller, K. M., K. H. Kaukinen, T. D. Beacham, and R. E. Withler. 2001. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* 111(1-3):237-257.
- Miller, K. M., and R. E. Withler. 1996. Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics* 43(6):337-351.
- Miller, K. M., and R. E. Withler. 1997. MHC diversity in Pacific salmon: Population structure and trans-species allelism. *Hereditas* 127(1-2):83-95.
- Monmonier, M. S. 1973. Maximum-difference barriers: An alternative numerical regionalization method. *Geographical Analysis* 5(3):245-261.
- Morin, P. A., G. Luikart, R. K. Wayne, and S. N. P. W. Grp. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19(4):208-216.
- Narum, S. R., M. Banks, T. D. Beacham, M. R. Bellinger, M. R. Campbell, J. Dekoning, A. Elz, C. M. Guthrie, C. Kozfkay, K. M. Miller, P. Moran, R. Phillips, L. W. Seeb, C. T. Smith, K. Warheit, S. F. Young, and J. C. Garza. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* 17(15):3464-3477.
- Nielsen, E. E., M. M. Hansen, and D. Meldrup. 2006. Evidence of microsatellite hitchhiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms. *Molecular Ecology* 15(11):3219-3229.
- O'Malley, K. G., M. D. Camara, and M. A. Banks. 2007. Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* 16(23):4930-4941.
- Page, R. D. M. 1996. TreeView: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12(4):357-358.
- Parchman, T. L., K. S. Geist, J. A. Grahn, C. W. Benkman, and C. A. Buerkle. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *Bmc Genomics* 11:180.
- Paschou, P., E. Ziv, E. G. Burchard, S. Choudhry, W. Rodriguez-Cintron, M. W. Mahoney, and P. Drineas. 2007. PCA-correlated SNPs for structure identification in worldwide human populations. *Plos Genetics* 3(9):1672-1686.

- Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6(1):288-295.
- Pearman, P. B. 2001. Conservation value of independently evolving units: Sacred cow or testable hypothesis? *Conservation Biology* 15(3):780-783.
- Pella, J., and M. Masuda. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. *Fishery Bulletin* 99:151-167. BAYES can be obtained at <ftp://ftp.afsc.noaa.gov/sida/mixture-analysis/bayes/>.
- Pella, J. J., and G. B. Milner. 1987. Use of genetic marks in stock composition analysis. Pages 274-276 in N. Ryman, and F. Utter, editors. *Population genetics and fisheries management*. University of Washington Press, Seattle, WA.
- Petren, K., P. R. Grant, B. R. Grant, and L. F. Keller. 2005. Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. *Molecular Ecology* 14(10):2943-2957.
- Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95(6):536-539.
- Quinn, T. P. 1993. A review of homing and straying of wild and hatchery-produced salmon. *Fisheries Research* 18(1-2):29-44.
- Quinn, T. P. 2005. *The behavior and ecology of Pacific salmon and trout*. American Fisheries Society, Bethesda, MD; University of Washington Press, Seattle, WA.
- Rafferty, A. E., and S. M. Lewis. 1996. Implementing MCMC. Pages 115-130 in W. R. Gilks, S. Richardson, and Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Chapman and Hall, Inc., London.
- Ramstad, K. M., C. A. Woody, G. K. Sage, and F. W. Allendorf. 2004. Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. *Molecular Ecology* 13(2):277-290.
- Rank, N. E., and E. P. Dahlhoff. 2002. Allele frequency shifts in response to climate change and physiological consequences of allozyme variation in a montane insect. *Evolution* 56(11):2278-2289.
- Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* 94(17):9197-9201.

- Renaut, S., A. W. Nolte, and L. Bernatchez. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology* 19 (Suppl. 1):115-131.
- Rohlf, F. J. 1994. NTSYS-pc. Numerical taxonomy system for cluster and ordination analyses. Version 2.2. Exeter Software, Setauket, New York.
- Rosenberg, N. A. 2005. Algorithms for selecting informative marker panels for population assignment. *Journal of Computational Biology* 12(9):1183-1201.
- Rosenberg, N. A., L. M. Li, R. Ward, and J. K. Pritchard. 2003. Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics* 73(6):1402-1422.
- Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145(4):1219-1228.
- Rousset, F. 2008. GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8(1):103-106.
- Ryman, N. 2006. CHIFISH: a computer program testing for genetic heterogeneity at multiple loci using chi-square and Fisher's exact test. *Molecular Ecology Notes* 6(1):285-287.
- Sanchez, C. C., T. P. L. Smith, R. T. Wiedmann, R. L. Vallejo, M. Salem, J. B. Yao, and C. E. Rexroad. 2009. Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *Bmc Genomics* 10.
- Savereide, J. W. 2005. Inriver abundance, spawning distribution, and run timing of Copper River Chinook salmon, 2002-2004 final report for study 02-015. U.S. Fish and Wildlife Service, Office of Subsistence Management, Fishery Information Service Division, Anchorage.
- Schlotterer, C. 2004. The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics* 5(1):63-69.
- Schwartz, M. K., G. Luikart, K. S. McKelvey, and S. A. Cushman. 2009. Landscape Genomics: A Brief Perspective. S. A. Cushman, and F. Huettmann, editors. *Spatial Complexity, Informatics, and Wildlife Conservation*. Springer.
- Seeb, J., C. Pascal, L. W. Seeb, E. Grau, W. D. Templin, T. Harkins, and S. Roberts. In Press. Next generation transcriptome sequencing and high-resolution melt

analyses enhance SNP discovery in chum salmon, *Oncorhynchus keta*. *Molecular Ecology Resources*:000-000.

- Seeb, J. E., C. E. Pascal, R. Ramakrishnan, and L. W. Seeb. 2009a. SNP genotyping by the 5'-nuclease reaction: advances in high throughput genotyping with non-model organisms. Pages 277-292 in A. Komar, editor. *Methods in Molecular Biology, Single Nucleotide Polymorphisms*, 2nd edition. Humana Press.
- Seeb, L. W., A. Antonovich, A. A. Banks, T. D. Beacham, A. R. Bellinger, S. M. Blankenship, A. R. Campbell, N. A. Decovich, J. C. Garza, C. M. Guthrie, T. A. Lundrigan, P. Moran, S. R. Narum, J. J. Stephenson, K. J. Supernault, D. J. Teel, W. D. Templin, J. K. Wenburg, S. E. Young, and C. T. Smith. 2007. Development of a standardized DNA database for Chinook salmon. *Fisheries* 32(11):540-552.
- Seeb, L. W., N. A. DeCovich, A. W. Barclay, C. T. Smith, and W. D. Templin. 2009b. Timing and origin of Chinook salmon stocks in the Copper River and adjacent coastal fisheries using DNA markers. Annual report for study 04-207. UFWWS Office of Subsistence Management, Fisheries Resource Monitoring Program. Alaska Department of Fish and Game, Fishery Data Series No. 09-58, Anchorage.
- Seeb, L. W., C. Habicht, W. D. Templin, K. E. Tarbox, R. Z. Davis, L. K. Brannian, and J. E. Seeb. 2000. Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the *Exxon Valdez* oil spill. *Transactions of the American Fisheries Society* 129(6):1223-1249.
- Shaklee, J. B., T. D. Beacham, L. Seeb, and B. A. White. 1999. Managing fisheries using genetic data: case studies from four species of Pacific salmon. *Fisheries Research* 43(1-3):45-78.
- Shaklee, J. B., and S. R. Phelps. 1990. Operation of a large-scale, multiagency genetic stock identification program. *American Fisheries Society Symposium* 7:817-830.
- Short, A. E. Z., and M. S. Caterino. 2009. On the validity of habitat as a predictor of genetic structure in aquatic systems: a comparative study using California water beetles. *Molecular Ecology* 18(3):403-414.
- Smith, C. T., A. Antonovich, W. D. Templin, C. M. Elfstrom, S. R. Narum, and L. W. Seeb. 2007. Impacts of marker class bias relative to locus-specific variability on population inferences in Chinook salmon: A comparison of single-nucleotide polymorphisms with short tandem repeats and allozymes. *Transactions of the American Fisheries Society* 136(6):1674-1687.

- Smith, C. T., C. M. Elfstrom, L. W. Seeb, and J. E. Seeb. 2005a. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology* 14(13):4193-4203.
- Smith, C. T., R. J. Nelson, C. C. Wood, and B. F. Koop. 2001. Glacial biogeography of North American coho salmon (*Oncorhynchus kisutch*). *Molecular Ecology* 10(12):2775-2785.
- Smith, C. T., J. E. Seeb, P. Schwenke, and L. W. Seeb. 2005b. Use of the 5'-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. *Transactions of the American Fisheries Society* 134(1):207-217.
- Smith, C. T., W. D. Templin, J. E. Seeb, and L. W. Seeb. 2005c. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of US and Canadian Chinook salmon caught in Yukon River fisheries. *North American Journal of Fisheries Management* 25(3):944-953.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry*. Freeman, San Francisco, California.
- Storfer, A., M. A. Murphy, J. S. Evans, C. S. Goldberg, S. Robinson, S. F. Spear, R. Dezzani, E. Delmelle, L. Vierling, and L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. *Heredity* 98(3):128-142.
- Sun, Y., H. Yu, Q. Zhang, J. Qi, Q. Zhong, Y. Chen, and C. Li. 2010. Molecular characterization and expression pattern of two zona pellucida genes in half-smooth tongue sole (*Cynoglossus semilaevis*). *Comparative Biochemistry and Physiology, Part B* 155:316-321.
- Taylor, E. B. 1991. A review of local adaptation in salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture* 98(1-3):185-207.
- Unwin, M. J., and T. P. Quinn. 1993. Homing and straying patterns of Chinook salmon (*Oncorhynchus tshawytscha*) from a New Zealand hatchery: spatial distribution of strays and effects of release date. *Canadian Journal of Fisheries and Aquatic Sciences* 50(6):1168-1175.
- Utter, F., and N. Ryman. 1993. Genetic-markers and mixed stock fisheries. *Fisheries* 18(8):11-21.
- Utter, F. M. 1991. Biochemical genetics and fishery management - an historical perspective. *Journal of Fish Biology* 39:1-20.
- Utter, F. M. 1994. Perspectives of molecular genetics and fisheries into the 21st century. *Reviews in Fish Biology and Fisheries* 4(3):374-378.

- Van Tassell, C. P., T. P. L. Smith, L. K. Matukumalli, J. F. Taylor, R. D. Schnabel, C. T. Lawley, C. D. Haudenschild, S. S. Moore, W. C. Warren, and T. S. Sonstegard. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* 5(3):247-252.
- Wade, G. D., J. J. Smith, K. M. van den Broek, and J. W. Savereide. 2008. Spawning distribution and run timing of Copper River sockeye salmon, 2007 final report for study 05-501. U.S. Fish and Wildlife Service, Office of Subsistence Management, Fisheries Resource Monitoring Program, Anchorage.
- Waples, R. S. 1987. A multispecies approach to the analysis of gene flow in marine shore fishes. *Evolution* 41(2):385-400.
- Waples, R. S. 1990a. Conservation genetics of Pacific salmon. III. Estimating effective population size. *Journal of Heredity* 81(4):277-289.
- Waples, R. S. 1990b. Temporal changes of allele frequency in Pacific salmon - implications for mixed-stock fishery analysis. *Canadian Journal of Fisheries and Aquatic Sciences* 47(5):968-976.
- Wehrhahn, C. F., and R. Powell. 1987. Electrophoretic variation, regional differences, and gene flow in the coho salmon (*Oncorhynchus kisutch*) of southern British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* 44(4):822-831.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358-1370.
- Wood, C. C. 1995. Life history variation and population structure in sockeye salmon. *American Fisheries Society Symposium* 17:195-216.
- Wood, C. C., J. W. Bickham, R. J. Nelson, C. J. Foote, and J. C. Patton. 2008. Recurrent evolution of life history ecotypes in sockeye salmon: implications for conservation and future evolution. *Evolutionary Applications* 1(2):207-221.
- Wood, C. C., S. McKinnell, T. J. Mulligan, and D. A. Fournier. 1987. Stock identification with the maximum-likelihood mixture model - sensitivity analysis and application to complex problems. *Canadian Journal of Fisheries and Aquatic Sciences* 44(4):866-881.
- Wood, C. C., B. E. Riddell, D. T. Rutherford, and R. E. Withler. 1994. Biochemical genetic survey of sockeye salmon (*Oncorhynchus nerka*) in Canada. *Canadian Journal of Fisheries and Aquatic Sciences* 51:114-131.

- Wright, S. 1978. Evolution and the genetics of populations, volume IV. Variability within and among natural populations. University of Chicago Press, Chicago.
- Yowe, D. L., and R. J. Epping. 1995. Cloning of the barramundi growth hormone encoding gene - a comparative analysis of higher and lower vertebrate GH genes. *Gene* 162(2):255-259.