Fishery Manuscript Series No. 12-06

Genetic Baseline for Upper Cook Inlet Sockeye Salmon: 96 SNPs and 10,000 Fish

by

Andrew W. Barclay

and

Christopher Habicht

December 2012

Alaska Department of Fish and Game **Divisions of Sport Fish and Commercial Fisheries**

Symbols and Abbreviations

The following symbols and abbreviations, and others approved for the Système International d'Unités (SI), are used without definition in the following reports by the Divisions of Sport Fish and of Commercial Fisheries: Fishery Manuscripts, Fishery Data Series Reports, Fishery Management Reports, and Special Publications. All others, including deviations from definitions listed below, are noted in the text at first mention, as well as in the titles or footnotes of tables, and in figure or figure captions.

Weights and measures (metric)

centimeter	cm
deciliter	dL
gram	g
hectare	ha
kilogram	kg
kilometer	km
liter	L
meter	m
milliliter	mL
millimeter	mm

Weights and measures (English)

cubic feet per second	ft^3/s
foot	ft
gallon	gal
inch	in
mile	mi
nautical mile	nmi
ounce	oz
pound	lb
quart	qt
yard	yd

Time and temperature

day	d
degrees Celsius	°C
degrees Fahrenheit	°F
degrees kelvin	K
hour	h
minute	min
second	s

Physics and chemistry

all atomic symbols	
alternating current	AC
ampere	A
calorie	cal
direct current	DC
hertz	Hz
horsepower	hp
hydrogen ion activity (negative log of)	pH
parts per million	ppm
parts per thousand	ppt, ‰
volts	V
watts	W

General

Alaska Administrative Code	AAC
all commonly accepted abbreviations	e.g., Mr., Mrs., AM, PM, etc.
all commonly accepted professional titles	e.g., Dr., Ph.D., R.N., etc.
at	@
compass directions:	
east	E
north	N
south	S
west	W
copyright	©
corporate suffixes:	
Company	Co.
Corporation	Corp.
Incorporated	Inc.
Limited	Ltd.
District of Columbia	D.C.
et alii (and others)	et al.
et cetera (and so forth)	etc.
exempli gratia (for example)	e.g.
Federal Information Code	FIC
id est (that is)	i.e.
latitude or longitude	lat. or long.
monetary symbols (U.S.)	$, ¢
months (tables and figures): first three letters	Jan,...,Dec
registered trademark	®
trademark	™
United States (adjective)	U.S.
United States of America (noun)	USA
U.S.C.	United States Code
U.S. state	use two-letter abbreviations (e.g., AK, WA)

Mathematics, statistics

all standard mathematical signs, symbols and abbreviations	
alternate hypothesis	H_A
base of natural logarithm	*e*
catch per unit effort	CPUE
coefficient of variation	CV
common test statistics	(F, t, χ^2, etc.)
confidence interval	CI
correlation coefficient (multiple)	R
correlation coefficient (simple)	r
covariance	cov
degree (angular)	°
degrees of freedom	df
expected value	*E*
greater than	>
greater than or equal to	≥
harvest per unit effort	HPUE
less than	<
less than or equal to	≤
logarithm (natural)	ln
logarithm (base 10)	log
logarithm (specify base)	$\log_2$, etc.
minute (angular)	'
not significant	NS
null hypothesis	H_O
percent	%
probability	P
probability of a type I error (rejection of the null hypothesis when true)	α
probability of a type II error (acceptance of the null hypothesis when false)	β
second (angular)	"
standard deviation	SD
standard error	SE
variance	
population	Var
sample	var

FISHERY MANUSCRIPT SERIES NO. 12-06

GENETIC BASELINE FOR UPPER COOK INLET SOCKEYE SALMON: 96 SNPS AND 10,000 FISH

by
Andrew W. Barclay and Christopher Habicht
Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory, Anchorage

Alaska Department of Fish and Game
Division of Sport Fish, Research and Technical Services
333 Raspberry Road, Anchorage, Alaska, 99518-1565

December 2012

Laboratory analysis for this project was funded by the Western Alaska Salmon Stock Identification Project (WASSIP; http://www.adfg.alaska.gov/index.cfm?adfg=wassip.main) and statistical analyses were funded by State of Alaska.

The Fishery Manuscript series was established in 1987 by the Division of Sport Fish for the publication of technically-oriented results of several years' work undertaken on a project to address common objectives, provide an overview of work undertaken through multiple projects to address specific research or management goal(s), or new and/or highly technical methods, and became a joint divisional series in 2004 with the Division of Commercial Fisheries. Fishery Manuscripts are intended for fishery and other technical professionals. Fishery Manuscripts are available through the Alaska State Library and on the Internet http:/www.adfg.alaska.gov/sf/publications/ This publication has undergone editorial and peer review.

Andrew W. Barclay,
Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory, 333 Raspberry Road, Anchorage, AK 99518, USA

Christopher Habicht,
Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory, 333 Raspberry Road, Anchorage, AK 99518, USA

This document should be cited as:
Barclay, A. W. and C. Habicht. 2012. Genetic baseline for Upper Cook Inlet sockeye salmon: 96 SNPs and 10,000 fish. Alaska Department of Fish and Game, Fishery Manuscript Series No. 12-06, Anchorage.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This report updates the previously reported genetic baseline used in mixed stock analysis (MSA) of sockeye salmon *Oncorhynchus nerka* originating from all major spawning systems in Upper Cook Inlet, Alaska. The baseline was augmented with new samples that provide better representation of existing populations and add 10 new populations. All individuals in the baseline and test mixtures were genotyped for 39 previously reported single nucleotide polymorphic (SNP) markers and 57 new SNP markers representing 92 loci. These markers reveal population structure similar to that observed in previous analyses. Eight reporting groups important to fisheries management were tested and met the criteria to be used in MSA. This baseline can be used for MSA of sockeye salmon harvested in Upper Cook Inlet fisheries and for inriver mixtures. Finally, this baseline can be combined with other baselines throughout the Pacific Northwest for MSA of mixtures that might contain sockeye salmon from Cook Inlet.

Key words: Upper Cook Inlet, Alaska, sockeye salmon, *Oncorhynchus nerka*, mixed stock analysis, MSA, single nucleotide polymorphism, SNP

INTRODUCTION

BACKGROUND

Since the early 1990s the department's Gene Conservation Laboratory has actively developed and refined mixed stock analysis (MSA) applications to provide improved stock composition information for use in the management of commercial fisheries. These efforts have focused on chum *Oncorhynchus keta*, Chinook *O. tshawytscha*, and sockeye *O. nerka* salmon (e.g., Seeb et al. 2004; Templin et al. 2005; Habicht et al. 2007a; Habicht et al. 2007b; Dann et al. 2009). The department now conducts MSA projects throughout the state and maintains extensive tissue archives from spawning populations for all 3 species.

One of the earliest MSA projects was initiated by the department in 1992 for sockeye salmon in Upper Cook Inlet (UCI) following the *Exxon Valdez* oil spill using allozyme (protein) markers (Seeb et al. 1997). Building on the earlier genetic studies of Grant et al. (1980) and Wilmot and Burger (1985), the project was designed to detect the contribution of Kenai River sockeye salmon to the commercial harvest. Over the course of the project, the department sampled approximately 8,300 sockeye salmon from 54 spawning populations between 1992 and 1997 and provided a detailed analysis of population structure of sockeye salmon in UCI using allozyme analyses (Seeb et al. 2000). The data revealed a substantial amount of genetic diversity among populations of UCI with the diversity distributed both within and among major drainages. The data supported a model of population structure generally organized around the lakes in which juvenile sockeye salmon rear (nursery lakes).

These allozyme data, paired with the MSA statistical methods available at that time, were able to differentiate among populations spawning in the major sockeye salmon-producing regions: Yentna/Susitna, Kenai, and Kasilof rivers, and groups of minor river drainages including those in West Cook Inlet, Northeast Cook Inlet, and Knik Arm. Single-region mixtures of simulated fish (based on population-specific allele frequencies) subjected to MSA allocated on average 91% to the correct region. However, when samples were taken from fish captured at fish wheels within the Kenai, Kasilof, Susitna, and Yentna rivers, allocations to the reporting group of origin averaged 85%. In addition, stock composition estimates from fish sampled in drift and set gillnet fisheries showed higher day-to-day variability than was expected by the fishery managers. This combination of results did not provide the managers with the confidence necessary to use these data for management decisions regarding UCI sockeye salmon.

Concurrent with these fishery monitoring activities, the department actively focused on research to improve the techniques of MSA, including: 1) development and evaluation of genetic markers for improved resolution of stock identification, 2) development of statistical techniques for more accurate and precise estimation of stock composition, and 3) development of the infrastructure to support high-throughput and low-error genotyping.

In 2007, the department released the first set of results of an initiative begun in July, 2005 to apply improved MSA techniques to estimate UCI sockeye salmon stock composition in commercial harvests for selected periods from 2005 through 2007 (Habicht et al. 2007b). In this initiative, single nucleotide polymorphism (SNP) markers replaced allozymes. This first set of results, based on 45 SNP markers, used the same reporting groups identified with the allozyme data.

After these results were released, the Gene Conservation Laboratory upgraded the baseline with additional samples, modified reporting groups, and incorporated more state-of-the-art statistical methods for MSA. The baseline was augmented with additional samples and 1 linked marker was excluded from the analysis (44 SNPs used). Reporting groups were modified to better incorporate stock composition information into stock-specific production models for evaluating escapement goals, estimating exploitation rates, and forecasting future runs. Bayesian methods were used exclusively in these mixture analyses.

In 2010, the department released 2 sets of results using this upgraded baseline. The first set of results included an updated baseline analysis and stock composition estimates based on genetic markers from sockeye salmon collected in selected periods of the Central and Northern district commercial fisheries and from the offshore test fishery between 2005 and 2008 (Barclay et al. 2010a). Later that year, a second set of stock composition estimates was released for the 2009 commercial and offshore test fisheries using the same baseline and analysis methods (Barclay et al. 2010b).

In 2012, a new coastwide baseline was published for the Western Alaska Salmon Stock Identification Program (WASSIP; Dann et al. 2012). This baseline doubled the number of markers screened for sockeye salmon populations from Cape Suckling to Kotzebue Sound. This baseline also incorporated new baseline samples (from additional sampling years and populations) and implemented improved methods to detect and handle linked loci. Since the last baseline upgrade, additional test mixtures were also used to evaluate baseline performance for MSA in UCI. Here we reanalyze the baseline collections for MSA in UCI, taking advantage of these new data and methods, and provide a new baseline for future fishery and inriver mixture analyses.[1]

[1] Product names used in this publication are included for completeness but do not constitute product endorsement. The Alaska Department of Fish and Game does not endorse or recommend any specific company or their products.

DEFINITIONS

To reduce confusion associated with the methods, results, and interpretation of this study, basic definitions of commonly used genetic terms are offered here.

Allele. Alternative form of a given gene or DNA sequence.

Allozyme. Variant form of a protein enzyme encoded at a given locus. Allozymes are usually distinguished by protein electrophoresis and histochemical staining techniques.

Bootstrapping. A method of resampling data with replacement to assess the variation of parameters of interest.

Credibility Interval. In Bayesian statistics, a credibility interval is a posterior probability interval. Credibility intervals are a direct statement of probability: i.e., a 90% credibility interval has a 90% chance of containing the true answer. This is different than the confidence intervals used in frequentist statistics.

F_{ST}. Fixation index is an estimate of the reduction in heterozygosity due to random genetic drift among populations; the proportion of the variation at a locus attributable to divergence among populations.

Linkage Disequilibrium. A state that exists in a population when alleles at different loci are not distributed independently in the population's gamete pool, often because the loci are physically linked.

Genetic Marker. A known DNA sequence that can be identified by a simple assay.

Genotype. The set of alleles for 1 or more loci for a fish.

Hardy-Weinberg Expectations (H-W). The genotype frequencies that would be expected from given allele frequencies assuming: random mating, no mutation (the alleles don't change), no migration or emigration (no exchange of alleles between populations), infinitely large population size, and no selective pressure for or against any traits.

Heterozygosity. The proportion of individuals in a population that have 2 allele forms (are heterozygous) at a particular marker; a measure of variability.

Locus (plural, loci). A fixed position or region on a chromosome.

Linked Markers. Markers showing gametic disequilibrium.

Mixed stock analysis (MSA). Method using allele frequencies from populations and genotypes from mixture samples to estimate stock compositions of mixtures.

Polymerase chain reaction (PCR). A method to amplify a single copy or few copies of a locus across several orders of magnitude, generating millions of copies of DNA.

Population. A locally interbreeding group that has little interbreeding with other spawning aggregations other than the natural background stray rate, is uniquely adapted to a spawning habitat, and has inherently unique attributes (Ricker 1958) that result in different productivity rates (Pearcy 1992; NRC 1996). This population definition is analogous to the spawning aggregations described by Baker et al. (1996) and the demes by NRC (1996).

Reporting Group. A group of populations in a genetic baseline to which portions of a mixture are allocated during MSA; constructed based on a combination of management needs and

genetic distinction. See definition for Salmon Stock for breakdown of reporting groups (stocks) in Upper Cook Inlet.

Single nucleotide polymorphism (SNP). A DNA sequence variation occurring when a single nucleotide (A, T, C, or G) differs among individuals or within an individual between paired chromosomes.

Salmon Stock. A locally interbreeding group of salmon (population) that is distinguished by a distinct combination of genetic, phenotypic, life history, and habitat characteristics or an aggregation of 2 or more interbreeding groups (populations) which occur within the same geographic area and is managed as a unit (from 5 AAC 39.222(f)). For purposes of this study, *stocks* in Upper Cook Inlet were delineated based on the major population or aggregation of populations for which the department estimates escapement or for a population or aggregation of populations which occur in a geographic area for which the department does not estimate escapement. Upper Cook Inlet stocks are defined as: 1) the largest producer on the west side (Crescent River; *Crescent*), 2) the remaining West Cook Inlet producers (*West*), 3) the lakes with weirs in the Susitna/Yentna rivers (Judd/Chelatna/Larson lakes) and the Mama and Papa Bear Lakes and Talkeetna Sloughs population (*JCL*), 4) the remaining producers in the Susitna/Yentna rivers (*SusYen*), 5) the only major creek with a weir in the Knik/Turnagain/Northeast Cook Inlet area (Fish Creek; *Fish*), 6) the remaining Knik/Turnagain/Northeast Cook Inlet producers (*KTNE*), 7) the composite of all populations within the Kenai River (*Kenai*), and 8) the composite of all populations within the Kasilof River (*Kasilof*).

DEVELOPMENT OF GENETIC MARKERS

A suite of genetic markers have been used over the years for MSA applications in Pacific salmon (reviewed in Habicht et al. 2007b). Single nucleotide polymorphism (SNP) applications in MSA studies of Pacific salmon have become increasingly common (Smith et al. 2005b; Smith et al. 2007; Narum et al. 2008; Habicht et al. 2007b; Dann et al. 2009; Barclay et al. 2010a; Barclay et al. 2010b; Habicht et al. 2010). The department developed assays for SNP markers for sockeye salmon (Smith et al. 2005a; Elfstrom et al. 2006; Habicht et al. 2010), and these markers are now used by U.S. laboratories for projects on sockeye salmon by the Pacific Salmon Commission in the Northern Boundary region. This same method has been used by the department with sockeye salmon in Bristol Bay both inseason to estimate relative stock contributions passing through the Port Moller test fishing area, and postseason to estimate the commercial-catch stock contributions in fisheries for brood tables used to establish escapement goals (Dann et al. 2009). This same set of SNP assays was used in UCI to analyze a subset of the samples reported in Habicht et al. (2007b) and Barclay et al. (2010a, 2010b). A new suite of SNP markers was used in this study, which includes a subset of these SNP assays. This new suite of 96 SNP markers was screened for a sockeye salmon baseline from populations from Cape Suckling to Kotzebue Sound under WASSIP (Dann et al. 2012).

STATISTICAL DEVELOPMENTS

Different statistical methods have been developed over the years for MSA applications in Pacific salmon (reviewed in Habicht et al. 2007b). Conditional maximum likelihood methods (Pella and Milner 1987) have been used to directly estimate the stock composition of sockeye salmon mixtures in UCI, or to provide a prior for Bayesian analysis (Pella and Masuda 2001; Koljonen

et al. 2005; reviewed in Habicht et al. 2007b). Use of conditional likelihood methods to produce priors for Bayesian analysis has been abandoned because results might provide more optimistic measures of accuracy than is warranted due to the double use of mixture information. In the most recent analyses, sockeye salmon mixture analyses rely solely on Bayesian methods (Dann et al. 2009; Barclay et al. 2010a; Barclay et al. 2010b). Priors for Bayesian analyses in these more recent analyses have been based on best available information including professional judgment and information from previous mixtures.

INFRASTRUCTURE IMPROVEMENTS

Genotyping technologies for SNPs have been developing at a rapid rate and are now faster than those for any other marker class (Ranade et al. 2001; Melton 2003; Wang et al. 2009; Hauser et al. 2011). SNP genotypes can be assayed by a variety of methods, typically with exceedingly low error rates (Habicht et al. 2007b; Dann et al. 2009), and these assays are readily transferred and repeatable across instruments and laboratories. Recently, the department installed highly automated technology to further reduce genotyping error rates and increase throughput.

The movement to high-throughput analyses has also required the department to develop a laboratory database and implement quality control measures to ensure data integrity and measure genotyping error rates. Both of these components were used and are reviewed in this study.

METHODS

TISSUE SAMPLING

Tissue samples suitable for genetic analyses (from here on referred to as *genetic samples*) were collected and subsequently either frozen (heart, muscle, liver and eye; samples collected prior to 2006) or preserved in ethanol (axillary fin). Frozen tissues were sampled into individual vials and ethanol-preserved samples were placed collectively into 125–500 ml containers with 1 or more containers for each collection site for each year.

Baseline

Baseline genetic samples were collected from spawning populations of sockeye salmon by the department using gillnets and beach seines (Table 1; Figure 1). Most collections were made in the 1990s and have been reported in Seeb et al. (2000), Habicht et al. (2007b) and Barclay et al. (2010a). These populations represent most of the known genetic diversity of the species in this region, influenced by both geographic (location) and temporal (early- and late-spawning) forces. Collections selected for inclusion in the current study represent all the populations previously identified in Barclay et al. (2010a) with additional collections made between 2009 and 2011 from unrepresented or underrepresented populations. Target sample size for baseline populations was 95 individuals across all years to achieve acceptable precision for the allele frequency estimates (Allendorf and Phelps 1981; Waples 1990a).

Test Mixtures

Test mixture genetic samples were collected from sockeye salmon captured within drainages during their homeward or seaward migrations. Homeward-migrating fish were sampled from adults captured in fish wheels operating on the Crescent, Yentna, Susitna, Kenai, and Kasilof rivers, at a weir on Fish Creek, and from fish seined at 2 lake-spawning locations on Crescent

Lake (Table 2). Seaward-migrating fish were sampled from smolt captured in an inclined-plane trap operating on the Kenai River.

Fish wheel samples were generally collected in proportion to the fish wheel catch throughout the run. These fish wheels were all located below the spawning sites in each river (Table 2; Figure 1) and are thought to capture only fish destined to spawn within the rivers where the fish wheels operate. The seine samples collected at Crescent Lake were originally sampled as a baseline collection from 2 different spawning locations but were used as a test mixture, and not included in this version of the baseline, because they likely represented multiple populations. Because the Crescent River fish wheel sample was <100 fish we combined it with the sample from the 2 lake-spawning locations in Crescent Lake. The inclined-plane trap samples taken from the Kenai River were generally collected in proportion to the catch over a 1-week period.

LABORATORY ANALYSIS

Assaying Genotypes

We extracted genomic DNA from tissue samples using a DNeasy® 96 Tissue Kit by QIAGEN® (Valencia, CA). We screened 96 SNP markers (Table 3) using Fluidigm® 96.96 Dynamic Arrays (http://www.fluidigm.com). The Fluidigm® 96.96 Dynamic Array contains a matrix of integrated channels and valves housed in an input frame. On one side of the frame are 96 inlets to accept the sample DNA from individual fish and on the other are 96 inlets to accept the assays for a unique SNP marker. Once in the wells, the components are pressurized into the chip using the IFC Controller HX (Fluidigm). The 96 samples and 96 assays are then systematically combined into 9,216 parallel reactions. Each reaction is a mixture of 4µl of assay mix (1×DA Assay Loading Buffer [Fluidigm], 10×TaqMan® SNP Genotyping Assay [Applied Biosystems], and 2.5×ROX [Invitrogen)]) and 5µl of sample mix (1×TaqMan® Universal Buffer [Applied Biosystems], 0.05×AmpliTaq® Gold DNA Polymerase [Applied Biosystems], 1×GT Sample Loading Reagent [Fluidigm], and 60–400ng/µl DNA) combined in a 7.2nL chamber. To increase loading efficiency and reduce liquid handling, 96-well plates of assay mix and reaction mix (working stocks) were made to load up to 40 and 100 Dynamic Arrays, respectively. Thermal cycling was performed on an Eppendorf IFC Thermal Cycler as follows: 70°C for 30 min for Hot-Mix step, initial denaturation of 10 min at 96°C followed by 40 cycles of 96° for 15 s and 60° for 1 min. The Dynamic Arrays were read on a Fluidigm® EP1™ System after amplification and scored using Fluidigm® SNP Genotyping Analysis software.

Assays that failed to amplify on the Fluidigm system were reanalyzed on the Applied Biosystems platform. Each reaction on this platform was performed in 384-well reaction plates in a 5µL volume consisting of 5–40ng/µl of template DNA, 1×TaqMan® Universal PCR Master Mix (Applied Biosystems), and 1×TaqMan® SNP Genotyping Assay (Applied Biosystems). Thermal cycling was performed on a Dual 384-Well GeneAmp® PCR System 9700 (Applied Biosystems) as follows: an initial denaturation of 10 min at 95°C followed by 50 cycles of 92°C for 1 s and annealing/extension temperature for 1 min. The plates were scanned on an Applied Biosystems Prism 7900HT Sequence Detection System after amplification and scored using Applied Biosystems' Sequence Detection Software (SDS) version 2.2. Genotypes produced on both platforms were imported and archived in the Gene Conservation Laboratory Oracle database, LOKI.

Laboratory Failure Rates and Quality Control

The overall failure rate was calculated by dividing the number of failed single-locus genotypes by the number of assayed single-locus genotypes. An individual genotype was considered a failure when a fish at a single locus was not given an allele call during the scoring process.

Quality control (QC) measures were instituted to identify laboratory errors and to determine the reproducibility of genotypes. For collections not previously analyzed for SNPs (not reported in Barclay et al. 2010a), the process involved the reanalysis of 8 out of every 96 fish (1 row per 96-well plate) for all markers by staff not involved with the original analysis. For collections that were analyzed previously for a subset of the SNPs (reported in Barclay et al. 2010a), 2 QC steps were implemented. The first step was used to ensure that the correct DNA extractions were screened. Newly collected genotypes of all individuals were compared with genotypes in the database for the 39 SNP markers that overlapped the two analyses. The second step was used to ensure that the correct assays were used for the 57 new SNP markers. For each SNP assay working stock plate used to load the Fluidigm® 96.96 Dynamic Arrays in the original analysis, 1 extraction plate was chosen for reanalysis for all markers by staff not involved with the original analysis.

Laboratory errors found during the QC process were corrected and data were corrected in the database. Inconsistencies not attributable to laboratory error were recorded, but original genotyping scores were maintained in the database.

Assuming that the inconsistencies among analyses (original vs. QC genotyping) were due equally to errors in original genotyping and errors during the QC genotyping and that these analyses are unbiased, error rates in the original genotyping can be estimated as ½ the rate of inconsistencies. Because baseline collections were genotyped on several projects and were subjected to different quality control measures, we report QC results for 2 recent baseline projects.

STATISTICAL ANALYSIS

Data Retrieval and Quality Control

We retrieved genotypes from LOKI and imported them into *R* (R Development Core Team 2010). All subsequent analyses were performed in *R* unless otherwise noted.

Prior to statistical analysis, we performed 3 analyses to confirm the quality of the data used. First we identified SNP markers that were invariant in all individuals. We excluded these markers from further statistical analyses. Second, we identified individuals that were missing substantial genotypic data. We used the 80% rule (missing data at 20% or more of loci; Dann et al. 2009) to identify individuals missing substantial genotypic data. We removed these individuals from further analyses because we suspect these samples have poor quality DNA. The inclusion of individuals with poor quality DNA might introduce genotyping errors into the baseline and reduce the accuracies of mixed stock analyses.

The final QC analysis identified individuals with duplicate genotypes and removed them from further analyses. Duplicate genotypes can occur as a result of sampling or extracting the same individual twice, and were defined as pairs of individuals sharing the same alleles in $\geq$ 95% loci screened. The individual with the most missing genotypic data from each duplicate pair was

removed from further analyses. If both individuals had the same amount of genotypic data the first individual was removed from further analysis.

Baseline Development

Hardy-Weinberg Equilibrium

For each locus within each baseline collection, we tested for conformance to Hardy-Weinberg expectations (HWE) using Monte Carlo simulation with 10,000 iterations in the *adegenet* package (Jombart 2008). We combined probabilities for each collection across loci and for each locus across collections using Fisher's exact test (Sokal and Rohlf, 1995) and removed collections and loci that violated HWE after correcting for multiple tests with Bonferroni's method ($\alpha = 0.05$ / # of collections) from subsequent analyses.

Pooling Collections into Populations

When appropriate, we pooled some collections to obtain better estimates of allele frequencies following a step-wise protocol. First we pooled collections from the same geographic location, sampled at similar calendar dates but in different years, as suggested by Waples (1990b). We then tested for differences in allele frequencies between pairs of geographically proximate collections that were collected at similar calendar dates and might represent the same population. We defined collections as being *geographically proximate* if they were within 12 km (water distance) of each other and within the same river or lake. We used Fisher's exact test (Sokal and Rohlf 1995) of allele frequency homogeneity and based our decisions on a summary across loci using Fisher's method. When these tests indicated no significant difference between collections ($P > 0.01$), we pooled them. When all individual collections within a pooled collection were geographically proximate to other collections we followed the same protocol until we found significant differences between the pairs of collections being tested. After this pooling protocol, we considered these final collections as populations. Finally, we tested populations for conformance to HWE following the same protocol described above to ensure that our pooling was appropriate, and that tests for linkage disequilibrium would not result in falsely positive results due to departure from HWE.

Linkage Disequilibrium

We tested for linkage disequilibrium between each pair of nuclear markers in each population to ensure that subsequent analyses would be based on independent markers. We used the program *Genepop* version 4.0.11 (Rousset 2008) with 100 batches of 5,000 iterations for these tests and summarized the frequency of significant linkage disequilibrium between pairs of SNPs ($P < 0.05$). We considered pairs to be linked if they exhibited linkage in more than half of all populations. We also examined the correlation coefficient r between the first allele in each linked pair of SNPs in each population to visualize the pattern of linkage across the geographic range of the baseline. We used the *BRugs* package to estimate the error around these correlation coefficient estimates (Thomas et al. 2006) and visualized these results with barplots using the *gplots* package (Warnes 2010).

When SNP pairs were found to be linked, we either removed 1 of the linked SNPs or combined the genotypes of the pair into a composite, haploid marker in further analyses if the pattern of linkage provided information useful for mixed stock analysis. We used f_{ORCA} as our measure of information. f_{ORCA} assesses the rate of correct allocation of simulated individuals to defined reporting groups based upon the markers in question (Rosenberg 2005). Because combinations

of alleles from 2 or more markers can exist in more forms than single markers (9 possible haplotypes vs. 2 alleles for SNPs), composite markers generally have higher f_{ORCA} values than the single markers that form them. Simple comparisons of these values would always suggest combining linked pairs into composite markers. However, there is a cost associated with composite markers as estimates of 8 haplotype frequencies are less precise than estimates of 1 allele frequency for a given sample size and 2 assays are required in the laboratory rather than a single assay.

To account for these costs, and to ensure that we combined only SNP pairs that provided significantly more information than the single SNPs in question, we compared the difference between f_{ORCA} values of the composite marker and the single SNP with the greater f_{ORCA} value in the pair [$\Delta = f_{ORCA\text{-}pair} - \max(f_{ORCA\text{-}SNP1}, f_{ORCA\text{-}SNP2})$]. This difference ($\Delta$) was our test statistic. Since we did not know the distribution of Δ for all pairs of markers, we conducted a sampled randomization test (Sokal and Rohlf 2005). We calculated Δ for each pair of nuclear markers (4,278 pairs) to empirically define the test statistic distribution, and set the 90th quantile of the distribution as a critical value (Δ_{90}). We then combined linked SNPs into composite, haploid loci if Δ was greater than this critical value and dropped the SNP with the lower f_{ORCA} value if Δ was less than the critical value. SNPs that did not exhibit linkage disequilibrium with any other SNP and SNPs that were combined were defined as loci for the remaining analyses.

Analysis of genetic structure

Analysis of variance

We examined the temporal variation of allele frequencies with a hierarchical, 3-level Analysis of Variance (ANOVA). We treated the temporal samples as subpopulations based on the method described in Weir (1996). This method allowed the quantification of the sources of total allelic variation and permitted the calculation of the between-subpopulation component of variance and the assessment of its magnitude relative to the between-population component of variance. This analysis was conducted using the software package *GDA* (Lewis and Zaykin 2001).

Visualization of genetic distances

To visualize genetic distances among populations, F_{ST} estimates were calculated from allele frequencies between all pairs of populations from the final set of independent markers with the package *hierfstat* (Goudet 2011). We constructed 1,000 bootstrapped Neighbor-Joining trees by resampling loci with replacement to assess the stability of tree nodes. We plotted the consensus tree with the *ape* package (Paradis et al. 2004).

Baseline Evaluation for MSA

Reporting groups and reporting group nomenclature

Reporting groups used in this report are the same as those used in Barclay et al. (2010a, 2010b) except that the Mama and Papa Bear lakes and Talkeetna Sloughs (MPBTS) population was moved to the original Judd/Chelatna/Larson lakes (*JCL*) reporting group. Populations were assigned to 2 different types of reporting groups: 1) groups of populations with escapement goals, and 2) regional groups of populations where escapements are not estimated. Populations were assigned into the following 8 reporting groups (stocks): 1) the largest producer of sockeye salmon on the west side (Crescent River; *Crescent*), 2) the remaining West Cook Inlet producers (*West*), 3) the lakes monitored by weirs in the Susitna/Yentna rivers (Judd/Chelatna/Larson lakes

with the addition of the MPBTS population; *JCL*), 4) the remaining producers in the Susitna/Yentna rivers (*SusYen*), 5) the only major creek monitored with a weir in the Knik/Turnagain/Northeast Cook Inlet area (Fish Creek; *Fish*), 6) the remaining Knik/Turnagain/Northeast Cook Inlet producers (*KTNE*), 7) the composite of all populations within the Kenai River (*Kenai*), and 8) the composite of all populations within the Kasilof River (*Kasilof*). Hereafter, when the terms *Crescent*, *West*, *JCL*, *SusYen*, *Fish*, *KTNE*, *Kenai*, and *Kasilof* are used as nouns, they refer to reporting groups (stocks: see definitions).

Testing Baseline for MSA

We assessed the potential of the baseline to identify these reporting groups for MSA applications with proof tests and test mixtures. For proof tests, up to 200 individuals were sampled without replacement from each reporting group and analyzed as a mixture against the reduced baseline. These tests provided an indication of the power of the baseline for MSA under the assumption that all the populations from a reporting group are represented in the baseline.

For test mixtures, we analyzed the test mixture genetic samples (see the Methods section *Tissue Sampling, Test Mixtures*; Table 2, Figure 1). These were the most challenging tests because fish may have originated from populations not represented in the baseline. They may also be the most realistic tests of baseline performance because the proportion of fish from each population in each mixture was more likely to be in proportion to the relative run strength of each population within the drainage for that year.

Stock compositions for the proof tests and the test mixtures were estimated with the program *BAYES* (Pella and Masuda 2001). The Bayesian model implemented by *BAYES* places a Dirichlet distribution as the prior distribution for the stock proportions, and the parameters for this distribution must be specified. We defined prior parameters for each reporting group to be equal (i.e., a *flat* prior) with the prior for each reporting group subsequently divided equally to populations within that reporting group. We set the sum of all prior parameters to 1 (prior weight), which is equivalent to adding 1 fish to each mixture (Pella and Masuda 2001). Populations were maintained separately within these reporting groups as recommended by Wood et al. (1987). Reporting group estimates were calculated by summing population estimates. We ran 5 independent Markov Chain Monte Carlo (MCMC) chains of 40,000 iterations with different starting values and discarded the first 20,000 iterations to remove the influence of the initial start values. We combined the second half of each chain to form the posterior distribution and tabulated mean estimates and 90% credibility intervals from a total of 100,000 iterations. We also assessed the among-chain convergence of these estimates using the Gelman-Rubin shrink factor, which compares the variation within a chain to the total variation among chains (Gelman and Rubin, 1992). If a shrink factor for any stock group estimate was greater than 1.2 we reanalyzed the mixture with 80,000-iteration chains following the same protocol. We repeated this procedure for each reporting group mixture. A critical level of 90% correct allocation was used to determine if the reporting group was acceptably identifiable (Seeb et al. 2000). We visualized these results with barplots using the *gplots* package.

RESULTS

TISSUE SAMPLING

Baseline

Spawning populations of sockeye salmon were collected from throughout UCI (Table 1; Figure 1). About half of the collections (60) were made during the 1990s. Collection efforts resumed in 2006 and 62 collections were made between 2006 and 2011. The 122 collections were taken at 86 locations throughout UCI drainages; individuals from 25 of these locations were collected in multiple years.

Test Mixtures

A total of 11 collections from fish captured at 5 fish wheels, 1 weir, 1 smolt trap, and 2 lake-spawning locations were made from 1992 to 2010 (Table 2).

LABORATORY ANALYSIS

Assaying Genotypes

A total of 10,001 baseline and 2,952 test mixture genetic samples were selected for analysis and assayed for 96 SNP markers (Tables 1 and 2).

Laboratory Failure Rates and Quality Control

For all 122 collections in the baseline, the overall failure rate for genotypes at the 96 SNP markers was 1.4%. The QC analyses for not-previously analyzed collections were comprised of 1,964 individuals (~20% of the current baseline); while the QC analysis for previously analyzed collections were comprised of 6,489 individuals (~87% of the current baseline. For the collections not previously analyzed for SNPs, the QC process detected discrepancy rates of 0.336%; for collections previously analyzed for a subset of the SNPs, the QC process detected discrepancy rates of 0.215%. The combined discrepancy rate between these 2 processes was 0.222%; therefore an estimate of the overall error rate was 0.111%.

STATISTICAL ANALYSIS

Data Retrieval and Quality Control

All SNP markers were variable in the UCI baseline collections. Based upon the 80% rule, 0.84% of individuals were removed for baseline collections, and 0.26% of individuals were removed for test mixture collections. Based upon the 95%-of-loci criterion for detecting duplicate individuals, 0.17% of individuals were removed from baseline collections as duplicate individuals; no duplicate individuals were detected in 106 of the 122 baseline collections (87%), or in the test mixture collections.

Baseline Development

Hardy-Weinberg Equilibrium

Over all nuclear markers and collections, 608 of 11,346 tests performed did not conform to HWE ($\alpha = 0.05$) without adjusting for multiple tests. These were spread over 92 markers and no

markers were out of HWE in more than 13 of the 122 collections. No collections were out of HWE at more than 15 of the 93 markers. After adjusting for multiple tests, all collections conformed to HWE.

Pooling collections into populations

A total of 69 populations were identified after pooling collections taken from the same geographic location over multiple years and geographically proximate collections (pooled collections and collections taken at different sites are referred to as *populations*; Table 1). The only collections taken from the same place over multiple years that were not pooled were from the early and late runs to Goat Creek. Over all nuclear markers and populations, 386 of 6,417 tests did not conform to HWE ($\alpha = 0.05$) without adjusting for multiple tests. These were spread over 89 markers and no markers were out of HWE in more than 11 of the 69 populations. No population was out of HWE at more than 12 of the 93 markers. After adjusting for multiple tests, all populations conformed to HWE.

Linkage disequilibrium

In the tests for linkage disequilibrium, 2 pairs of SNPs showed significant linkage ($P < 0.05$) in more than 50% of populations. These pairs were *One_MHC2-190* and *One_MHC2-251* (81% of populations) and *One_GPDH* and *One_GPDH2* (68% of populations; Table 3; Figure 2). Correlation coefficients *r* between the first alphabetical allele in each linked pair of SNPs varied across reporting groups and ranged from –0.530 to 0.121 for the *One_MHC2* pairs, and –0.318 to –0.096 for the *One_GPDH* pairs (Figure 3). The 90% critical value of the f_{ORCA} difference distribution Δ_{90} was 0.41, which was greater than Δ for the *One_GPDH*/*One_GPDH2* pair (Δ = 0.01) and less than Δ for the *One_MHC2-190*/*One_MHC2-251* pair (Δ = 0.08; Figure 4). The linked *One_MHC2* SNPs were combined to form a composite, haploid locus (*One_MHC2_190-251*) and *One_GPDH2* was dropped from further analysis because it had a lower f_{ORCA} value than *One_GPDH*.

Analysis of genetic structure

Analysis of variance

Twenty-five populations were included in the analysis of temporal variation of allele frequencies. Allele frequencies for all populations appeared to be temporally stable. Within populations, 14 pairs of collections were 10–17 years apart, and 11 were 1–3 years apart (Table 1). The 3-level ANOVA indicated that the ratio of variation among temporal collections to the variation among populations was 1.6%.

Visualization of genetic distances

Genetic relationships among baseline populations are shown in the Neighbor-Joining tree (Figure 5). We observed similar relationships among populations as reported in Barclay et al. (2010a), except in KTNE where 3 new populations were added. The Eagle River population clustered with populations from other spawning streams which drain into Knik Arm (Eska, Bodenburg, and Jim creeks) in 92% of bootstrap trees and the Carmen Lake population clustered with these populations and a nearby spawning stream which drains into Turnagain Arm (Williwaw Creek) below a node with little support. The Chickaloon River population was highly distinct and clustered with Crescent populations below a node with little support.

Baseline Evaluation for MSA

Testing Baseline for MSA

In the analysis of the proof test mixtures, all reporting groups were highly identifiable at ≥ 99% correct assignment (Table 4).

In the analysis of the escapement mixtures, all but the Yentna River fish wheel mixtures assigned at > 98% to the correct reporting group(s) (Table 5). In the Yentna River fish wheel escapement mixtures, most of the assignment was to both the JCL and SusYen reporting groups; however, all 3 years had a portion of assignment to West reporting group mixtures: 1% in 2005, 12% in 2008, and 6% in 2010.

DISCUSSION

This manuscript updates the previously reported baseline of Barclay et al. (2010a) and reports new analyses of genetic data collected from sockeye salmon originating from all major systems in UCI. The updates include the addition of new samples that provide better representation of existing populations, 10 new populations, and provide additional test mixtures for assessing the baseline for MSA.

BASELINE DEVELOPMENT

The pattern of similarity between populations revealed by this baseline is similar to the pattern revealed by other baselines (Habicht et al. 2007b; Barclay et al. 2010a), including those based on other marker types (Seeb et al. 2000; Allendorf and Seeb 2000). Straying among spawning areas is usually higher within drainages than among drainages (Wood et al. 1994) which can result in similarity among salmon spawning within a drainage and higher differentiation among salmon spawning in different drainages. The populations from the Kenai and Kasilof rivers form a large cluster with internal structure. All markers surveyed have shown little genetic heterogeneity among populations spawning in the Kasilof River drainage (Burger et al. 1997), although phenotypic diversity was observed by Woody et al. (2000). While Burger et al. (1995) detected a distinct late run of river spawners at the outlet of Tustumena Lake, no outlet spawners were included in either the allozyme or SNP baselines. Within the Kenai River drainage 4 main groups were found: 1) between Skilak and Kenai lakes and Skilak Lake outlet; 2) Hidden Lake; 3) Ptarmigan, Quartz, Moose, Railroad, and Johnson creeks and Tern Lake; and 4) Upper Russian Lake.

Variation was also found among populations within the remaining regions: Susitna and Yentna rivers, Knik Arm, Northeast Cook Inlet, Turnagain Arm, and West Cook Inlet. Unlike the Kenai and Kasilof drainages, there are no large nursery lakes that support multiple tributary-spawning populations within these regions. These systems tend to have a number of isolated smaller lakes. The close affinity of the Yentna and Susitna slough spawners may indicate common ancestry and a high level of historical gene flow similar to the *river-type* sockeye salmon described by Gustafson and Winans (1999).

DIFFERENCES IN THE BASELINE BETWEEN ANALYSES

Currently, SNPs have been screened on 69 populations in UCI with an average of 143 individuals per population. The major difference between this baseline analysis and the previous analysis (Barclay et al. 2010a) is that it contains genotypes from an additional 51 SNP markers.

This baseline analysis also contains new collections from Crescent (3; including 1 new population), West (3), SusYen (6; including 2 new populations), JCL (4), Fish (2), KTNE (5; including 3 new populations), and Kenai (5; including 1 new population). To reduce the cost of adding additional SNP markers, some collections previously reported in Barclay et al. (2010a) were analyzed for a reduced number of individuals and other collections were dropped for well-represented populations.

The method for pooling collections into populations differs between this analysis and the previous analysis in that a step-wise protocol was used and only geographically proximate collections were pooled. Because of this, a couple of previously defined populations in Barclay et al. (2010a) were broken into several populations (i.e., the Tustumena Lake population was broken out into 4 populations and the Between Kenai and Skilak lakes population was broken into 2 populations).

The MPBTS population, which was included in SusYen in the previous analysis, was moved to the JCL reporting group because it was genetically similar to the Larson Lake populations (Figure 5). This similarity was supported by a Fisher's exact test of allele frequency homogeneity between the MPBTS and Larson Lake East Shore populations where the test indicated no difference between the populations ($P = 0.48$). During initial baseline development, a proof test was performed for JCL while the MPBTS population was still included in SusYen and BAYES results revealed a large proportion of JCL misallocating to the MPBTS population (data not shown).

BASELINE PERFORMANCE

The performance of the baseline to correctly allocate fish during proof tests (Table 4) demonstrates that the variation among reporting groups in the baseline is adequate to produce highly accurate estimates of stock composition; however, allocations of some of the new escapement samples (Table 5) indicate that the baseline may not be complete for the Yentna River. Prior to this report, escapement tests from the Yentna, Susitna, Kenai, and Kasilof rivers had high correct allocations (Barclay et al. 2010a), indicating that the baselines adequately characterized the genetic variation among populations for MSA. This report again confirmed earlier findings for these samples, showing high correct allocations for all samples. However, this report contains new escapement samples that indicate the existence of additional important genetic variation in the Yentna River that is not represented in the baseline. Between 6% and 12% of the Yentna River fish wheel mixtures misallocated to the West reporting group in the 2010 (6%) and 2008 (12%) samples, respectively (Table 5). This contrasts with the 0% misallocation to the West reporting group from the SusYen proof test (Table 4). Additional field collection and laboratory analysis of spawning aggregates in the Yentna River should be conducted to close this gap. Once new collections are incorporated in the baseline, the 2008 and 2010 escapement samples from the Yentna River can be used to verify that the gap is closed. These new results show that continuous baseline testing is important to ensure that baselines are performing well for MSA and to provide insights into potential biases in MSA estimates.

ACKNOWLEDGEMENTS

The baseline augmentation from 45 markers to 96 markers would not have occurred without the impetus and funding from the Western Alaska Salmon Stock Identification Program (WASSIP; http://www.adfg.alaska.gov/index.cfm?adfg=wassip.main). As part of WASSIP, a

baseline for sockeye salmon from Cape Sucking to Kotzebue Sound using 96 markers was developed, which included all baseline data used in this report. We would like to acknowledge the work of the people in the department's Gene Conservation Laboratory that developed the WASSIP baseline. Additionally, we would like to acknowledge the department projects which paid for the collections we used as baseline test mixtures and the work of department staff that collected the samples.

Laboratory and statistical analyses were funded by WASSIP and the State of Alaska. Test mixture collections were funded by the Alaska Sustainable Salmon Fund, *Exxon Valdez* Oil Spill Trustee Council, and the State of Alaska.

REFERENCES CITED

Allendorf, F. W., and S. R. Phelps. 1981. Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences 38:1507–1514.

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 54:640–651.

Baker, T. T., A. C. Wertheimer, R. D. Burkett, R. Dunlap, D. M. Eggers, E. I. Fritts, A. J. Gharrett, R. A. Holmes and R. L. Wilmot. 1996. Status of Pacific salmon and steelhead in Southeastern Alaska. Fisheries 21:6–18.

Barclay, A. W., C. Habicht, W. D. Templin, H. A. Hoyt, T. Tobias, and T. M. Willette. 2010a. Genetic stock identification of Upper Cook Inlet sockeye salmon harvest, 2005–2008. Alaska Department of Fish and Game, Fishery Manuscript No. 10–01, Anchorage.

Barclay, A. W., C. Habicht, T. Tobias, and T. M. Willette. 2010b. Genetic stock identification of Upper Cook Inlet sockeye salmon harvest, 2009. Alaska Department of Fish and Game, Fishery Data Series No. 10–93, Anchorage.

Burger, C. V., J. E. Finn, and L. Holland-Bartels. 1995. Pattern of shoreline spawning by sockeye salmon in a glacially turbid lake: evidence for subpopulation differentiation. Transactions of the American Fisheries Society 124:1–15.

Burger, C. V., W. J. Spearman, and M. A. Cronin. 1997. Genetic differentiation of sockeye salmon subpopulations from a geologically young Alaskan lake system. Transactions of the American Fisheries Society 126:926–938.

Dann, T. H., C. Habicht, J. R. Jasper, H. A. Hoyt, A. W. Barclay, W. D. Templin, T. T. Baker, F. W. West, and L. F. Fair. 2009. Genetic stock composition of the commercial harvest of sockeye salmon in Bristol Bay, Alaska, 2006–2008. Alaska Department of Fish and Game, Fishery Manuscript Series No. 09-06, Anchorage. http://www.sf.adfg.state.ak.us/FedAidPDFs/fms09-06.pdf (Accessed November 26, 2012).

Dann, T. H., C. Habicht, J. R. Jasper, E. K. C. Fox, H. A. Hoyt, H. L. Liller, E. S. Lardizabal, P. A. Kuriscak, Z. D. Grauvogel, and W. D. Templin. 2012. Sockeye salmon baseline for the Western Alaska Salmon Stock Identification Program. Alaska Department of Fish and Game, Special Publication No. 12-12, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/SP12-12.pdf (Accessed November 26, 2012).

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 6:1255–1259.

Grant, W. S., G. B. Milner, P. Krasnowski, and F. M. Utter. 1980. Use of biochemical genetic variants for identification of sockeye salmon (*Oncorhynchus nerka*) stocks in Cook Inlet, Alaska. Canadian Journal of Fisheries and Aquatic Sciences 37:1236–1247.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–511.

Goudet, J. 2011. hierfstat: Estimation and tests of hierarchical F-statistics. R package version 0.04-6. http://CRAN.R-project.org/package=hierfstat.

Gustafson, R. G., and G. A. Winans. 1999. Distribution and population genetic structure of river- and sea-type sockeye salmon in western North America. Ecology of Freshwater Fish 8:181–193.

Habicht, C., L. W. Seeb, and J. E. Seeb. 2007a. Genetic and ecological divergence defines population structure of sockeye salmon populations returning to Bristol Bay, Alaska, and provides a tool for admixture analysis. Transactions of the American Fisheries Society 136:82–94.

Habicht, C., W. D. Templin, T. M Willette, L. F. Fair, S. W. Raborn, and L. W. Seeb. 2007b. Post-season stock composition analysis of Upper Cook Inlet sockeye salmon harvest, 2005–2007. Alaska Department of Fish and Game, Fishery Manuscript No. 07-07, Anchorage. http://www.sf.adfg.state.ak.us/FedAidPDFs/fms07-07.pdf

Habicht, C., L. W. Seeb, K. W. Myers, E. V. Farley Jr., and J. E. Seeb. 2010. Summer–Fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. Transactions of the American Fisheries Society 139:1171–1191.

REFERENCES CITED (Continued)

Hauser, L., M. Baird, R. Hilborn, L. W. Seeb, and J. E. Seeb. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. Molecular Ecology Resources 11:150–161.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. doi: 10.1093/bioinformatics/btn129

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 62:2143–2158.

Lewis, P. O., and D. Zaykin. 2001. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0. URL http://lewis.eeb.uconn.edu/lewishome/software.html (Accessed March, 2009; site currently discontinued).

Melton, L. 2003. On the trail of SNPs. Nature 422:917–923.

Narum, S. R., M. Banks, T. Beacham, R. Belllinger, M. Campbell, J. DeKoning, A. Elz, C. Guthrie, C. Kozfkay, K. Miller, P. Moran, R. Phillips, L. Seeb, C. Smith, K. Warheit, S. Young, and J. C. Garza. 2008. Differentiating salmon populations at broad and fine geographic scales with microsatellites and single nucleotide polymorphisms. Molecular Ecology (17)15:3464–3477.

NRC (National Research Council). 1996. Upstream: Salmon and Society in the Pacific Northwest. Committee on Protection and Management of Pacific Northwest Salmonids. National Academy Press, Washington, D.C.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Pearcy, W. 1992. Ocean ecology of north pacific salmonids. University of Washington Press, Seattle.

Pella, J., and M. Masuda. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. Fishery Bulletin 99:151–167.

Pella, J. J., and G. B. Milner. 1987. Use of genetic marks in stock composition analysis. Pages 247–276 [*In*]: N. Ryman and F. Utter, editors. Population Genetics and Fishery Management. Washington Sea Grant, University of Washington Press, Seattle.

R Development Core Team. 2010. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. www.r-project.org/ (Accessed December 14, 2012).

Ranade, K., M.–S. Chang, C.–T. Ting, D. Pei, C.–F. Hsiao, M. Olivier, R. Pesich, J. Hebert, Y.–D. I. Chen, V. Dzau, D. Curb, R. Olshen, N. Risch, D. R. Cox, and D. Botstein. 2001. High-throughput genotyping with single nucleotide polymorphisms. Genome Research 11:1262–1268.

Ricker, W. E. 1958. Maximum sustained yields from fluctuating environments and mixed stocks. Journal of the Fisheries Research Board of Canada 15:991–1006.

Rosenberg, N. A. 2005. Algorithms for selecting informative marker panels for population assignment. Journal of Computational Biology 12(9):1183-1201.

Rousset, F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8(1):103–106.

Seeb, L. W., P. A. Crane, C. M. Kondzela, R. L. Wilmot, S. Urawa, N. V. Varnavskaya, and J. E. Seeb. 2004. Migration of Pacific Rim chum salmon on the high seas: Insights from genetic data. Environmental Biology of Fishes 69:21–36.

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 129:1223–1249.

REFERENCES CITED (Continued)

Seeb, L. W., W. D. Templin, K. E. Tarbox, R. Z. Davis, and J. E. Seeb. 1997. Kenai River sockeye salmon restoration, Restoration Project 96255-2 Final Report, Exxon Valdez Oil Spill Trustee Council, Anchorage.

Smith C. T., A. Antonovich, W. D. Templin, C. D. 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 SNPs to STRs and allozymes. Transactions of the American Fisheries Society 136:1674–1687.

Smith, C. T., C. M. Elfstrom, J. E. Seeb, and L. W Seeb. 2005a. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14:4193–4203.

Smith, C. T., W. D. Templin, J. E. Seeb, and L. W. Seeb. 2005b. 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 25:944–953.

Sokal, R. R. and F. J. Rohlf. 1995. Biometry. 3rd Edition. Freeman, San Francisco, CA.

Templin, W. D., R. L. Wilmot, C. M. Guthrie III, and L. W. Seeb. 2005. United States and Canadian Chinook salmon populations in the Yukon River can be segregated based on genetic characteristics. Alaska Fisheries Research Bulletin 11:44–60.

Thomas, A., B. O'Hara, U. Ligges, and S. Sturtz. 2006. Making BUGS Open. R News 6 (1):12–17.

Wang, J., M. Lin, A. Crenshaw, A. Hutchinson , B. Hicks , M. Yeager , S. Berndt, W.–Y. Huang, R. B. Hayes, S. J. Chanock, R. C. Jones, and R. Ramakrishnan. 2009. High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. BMC Genomics 10:561.

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 of mixed stock fishery analysis. Canadian Journal of Fisheries and Aquatic Sciences 47:968–976.

Warnes, G. R. 2010. gplots: Various R programming tools for plotting data. R package version 2.8.0. http://CRAN.R-project.org/package=gplots

Weir, B. 1996. Genetic Data Analysis (second edition). Sinauer Associates, Inc., Sunderland, MA.

Wilmot, R. L., and C. V. Burger. 1985. Genetic differences among populations of Alaskan sockeye salmon. Transactions of the American Fisheries Society 114:236–243.

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 44:866–881.

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.

Woody, C. A., J. Olsen, J. Reynolds, and P. Bentzen. 2000. Temporal variation in phenotypic and genotypic traits in two sockeye salmon populations, Tustumena Lake, Alaska. Transactions of the American Fisheries Society 129:1031–104.

TABLES AND FIGURES

Table 1.–Tissue collections of sockeye salmon in the Upper Cook Inlet genetic baseline including the year sampled and the number of individuals (*N*) analyzed from each collection and their assigned reporting group for genetic stock identification.

Map No. [a]	Pop. No.	Collection No.	Reporting Group	Location	Sample Year	*N*
1	1	1	Crescent	Wackton Creek	2009	94
2	2	2		Pyramid Creek	1994	48
2	2	3			2009	95
3	3	4		Crescent Lake Outlet	1994	47
3	3	5			2009	95
4	4	6	West	Little Jack Creek	2006	95
5	5	7		Packers Lake	1992	95
				South Fork Big River		
6	6	8		site 1	2007	123
7	6	9		site 2	2009	48
8	7	10		Wolverine Creek	1993	95
9	8	11		Black Sand Creek	2007	95
10	9	12		Farros Lake	2007	95
11	10	13		McArthur River	1993	95
12	11	14		Chilligan River	1992	48
12	11	15			1994	48
13	12	16		Chackachatna River Slough	2008	95
14	13	17		Coal Creek	2009	48
15	13	18		West Fork Coal Creek	2009	47
16	14	19	SusYen	Susitna River Sloughs	1995	50
16	14	20			1996	6
17	14	21			1997	94
18	15	22		Byers Lake	1993	48
18	15	23			2007	95
19	16	24		Spink Creek	2008	95
20	17	25		Swan Lake	2006	95
20	17	26			2007	47
20	17	27			2009	48
21	18	28		Stephan Lake	1993	48
21	18	29			2007	95
22	19	30		Sheep River	2008	95
			JCL	Larson Lake		
23	20	31		Lake Outlet	1993	95
23	20	32			2011	125
24	21	33		East Shore	2006	95
24	21	34			2011	90
25	22	35	JCL	Mama and Papa Bear Lakes	1997	50
25	22	36			2007	54
26	22	37		Talkeetna River Slough	1997	79
27	23	38	SusYen	Birch Creek	2007	95
28	24	39		West Fork Yentna River	1992	96
28	24	40			1993	100
29	25	41		Kichatna River	2007	95

-continued-

Table 1.–Page 2 of 3.

Map No. [a]	Pop. No.	Collection No.	Reporting Group	Location	Sample Year	*N*
30	25	42		Johnson Creek	2009	95
31	26	43		Whiskey Lake	2006	58
31	26	44			2009	47
32	26	45		Hewitt Lake	2006	65
33	27	46		Moose Creek	2007	95
34	28	47		Puntilla Lake	2006	95
35	29	48		Trimble River	2007	47
35	29	49			2009	48
36	30	50		Red Salmon Lake	2006	95
37	31	51		Hayes River	2008	48
37	31	52			2009	47
38	32	53		Skwentna River Slough	2007	108
39	32	54		Canyon Creek	2007	65
40	33	55		Shell Lake	1993	48
40	33	56			2006	95
40	33	57			2009	95
41	34	58	JCL	Judd Lake	1993	96
41	34	59			2006	94
41	34	60			2009	95
42	35	61	SusYen	Trinity Lake	1992	48
42	35	62			2009	95
43	36	63	JCL	Chelatna Lake	1993	95
43	36	64			2006	95
43	36	65			2009	95
44	37	66	Fish	Meadow Creek	1994	94
45	37	67		Little Meadow Creek	2009	142
46	38	68		Big Lake Outlet	2011	188
47	39	69	KTNE	Nancy Lake	1993	95
47	39	70			2010	95
48	40	71	KTNE	Cottonwood Creek	1993	95
49	40	72		Wasilla Creek	1998	71
50	41	73		Eska Creek	2006	95
51	42	74		Jim Creek	1997	95
51	42	75			2011	65
52	43	76		Bodenburg Creek	2006	95
53	44	77		Eagle River	2011	95
54	45	78		Sixmile Creek	2008	95
55	46	79		Williwaw Creek	2006	39
55	46	80			2007	69
56	47	81		Carmen Lake	2010	95
57	48	82		Chickaloon River	2010	95
58	49	83		Swanson River	1997	95
59	50	84		Daniels Lake	1995	95
60	51	85		Bishop Creek	1993	95
61	52	86	Kenai	Railroad Creek	1997	48

-continued-

Table 1.–Page 3 of 3.

Map No. a	Pop. No.	Collection No.	Reporting Group	Location	Sample Year	*N*
62	52	87		Johnson Creek	1997	88
63	53	88		Moose Creek	1993	47
63	53	89			1994	95
64	54	90		Ptarmigan Creek	1992	47
64	54	91			1993	95
65	55	92		Tern Lake	1992	48
65	55	93			1993	48
66	56	94		Quartz Creek	1993	94
				Kenai River, between Skilak and Kenai lakes		
67	57	95		site 1	1994	47
68	57	96		site 2	1994	48
69	57	97		site 3	1994	143
70	58	98		site 4	1993	95
71	58	99		site 5	1994	48
72	58	100		site 6	1994	95
73	58	101		Lower Russian River	1993	95
			Kenai	Upper Russian River Early		
74	59	102		Goat Creek	1992	96
74	59	103			1997	95
74	59	104			2009	95
				Upper Russian River Late		
74	60	105		Goat Creek	2009	95
75	61	106		Bear Creek	2009	95
76	62	107		Upper Lake South Shore	1999	95
76	62	108			2009	95
77	63	109		Upper Lake Outlet	1999	95
77	63	110			2009	95
78	64	111		Hidden Lake	1993	95
78	64	112			2008	95
79	65	113		Skilak Lake Outlet	1992	96
79	65	114			1994	95
79	65	115			1995	48
			Kasilof	Tustumena Lake		
80	66	116		site A	1994	48
81	66	117		site B	1994	48
82	67	118		Seepage Creek	1994	95
83	67	119		Glacier Creek	1994	95
84	68	120		Moose Creek	1992	96
85	68	121		Bear Creek	1992	95
86	69	122		Nikolai Creek	1992	95

[a] Map numbers correspond to sampling sites on Figure 1, population numbers represent all the collections that contribute to a single population.

Table 2.–Location, capture method, date(s) sampled, and sample size (*N*) for tissue collection of sockeye salmon sampled for genetic studies operated within 6 of the major drainages into Upper Cook Inlet.

Map No.[a]	Location	Capture Method	Date	*N*
87	Crescent Lake (lake spawners, two locations)	seine	1995	95
88	Crescent River (river km 2.5)	fish wheel	7/2–28/1992	95
89	Susitna River (river km 116)	fish wheel	7/26/1992	190
89			7/15–8/12/2008	253
90	Yentna River (river km 6.5)	fish wheel	7/15/1992	190
90			7/7–8/15/2010	1045
91	Fish Creek	weir	1993	95
91			2008	190
92	Kenai River (river km 30.6)	fish wheel	7/11–20/2005	190
93	Kenai River (river km 72.4)	inclined-plane trap	5/26–6/1/2010	419
94	Kasilof River (river km 11.3)	fish wheel	7/11–20/2005	190

[a] Map numbers correspond to test mixture sampling sites on Figure 1.

Table 3.– Source, observed heterozygosity, F_{IS}, and F_{ST} for the 96 single nucleotide polymorphisms (SNPs) used to analyze the population genetic structure of Upper Cook Inlet, Alaska sockeye salmon. These summary statistics are based upon the 69 populations within Upper Cook Inlet (Table 1).

Assay	Source[a]	H_O	F_{IS}	F_{ST}
One_ACBP-79	A	0.438	0.016	0.106
One_agt-132	B	0.400	-0.011	0.117
One_aldB-152	C	0.357	-0.010	0.112
One_apoe-83	B	0.314	-0.011	0.101
One_c3-98	B	0.146	-0.152	0.078
One_CD9-269	B	0.294	0.005	0.096
One_cetn1-167	B	0.418	0.009	0.149
One_CFP1	D	0.215	0.015	0.094
One_cin-177	C	0.449	0.014	0.075
One_CO1[b]	A	---	0.000	0.332
One_ctgf-301	A	0.064	0.002	0.034
One_Cytb_17[b]	A	---	0.000	0.781
One_Cytb_26[b]	A	---	0.000	0.298
One_E2-65	A	0.334	0.011	0.153
One_gdh-212	C	0.375	0.012	0.130
One_GHII-2165	A	0.216	0.013	0.154
One_ghsR-66	C	0.417	-0.002	0.133
One_GPDH-201	A	0.467	0.005	0.053
One_GPDH2-187[c]	A	0.161	0.023	0.089
One_GPH-414	A	0.403	-0.005	0.066
One_HGFA-49	A	0.234	0.024	0.084
One_HpaI-71	A	0.351	0.017	0.120
One_HpaI-99	A	0.127	-0.010	0.136
One_hsc71-220	A	0.308	0.012	0.156
One_Hsp47	D	0.341	0.014	0.123
One_IL8r-362	A	0.083	-0.014	0.172
One_KCT1-453	B	0.141	0.023	0.101
One_KPNA-422	A	0.279	0.019	0.113
One_LEI-87	A	0.443	0.018	0.089
One_lpp1-44	B	0.439	0.001	0.110
One_metA-253	C	0.030	-0.001	0.036
One_MHC2_190[d]	A	0.290	0.033	0.359
One_MHC2_251[d]	A	0.333	0.026	0.303
One_Mkpro-129	C	0.426	-0.002	0.150

-continued-

Table 3.–Page 2 of 3.

Assay	Source[a]	H_O	F_{IS}	F_{ST}
One_ODC1-196	B	0.420	0.005	0.154
One_Ots208-234	C	0.123	0.009	0.051
One_Ots213-181	A	0.238	0.005	0.056
One_p53-534	A	0.049	0.007	0.239
One_pax7-248	C	0.207	0.013	0.116
One_PIP	D	0.397	0.015	0.143
One_Prl2	A	0.461	0.005	0.094
One_rab1a-76	B	0.213	0.012	0.143
One_RAG1-103	A	0.089	-0.012	0.116
One_RAG3-93	A	0.086	-0.017	0.068
One_redd1-414	C	0.394	0.011	0.168
One_RFC2-102	A	0.300	0.028	0.143
One_RFC2-285	A	0.078	0.021	0.093
One_rpo2j-261	C	0.313	0.006	0.127
One_sast-211	C	0.079	0.038	0.043
One_spf30-207	C	0.363	0.003	0.131
One_srp09-127	C	0.022	0.000	0.026
One_ssrd-135	C	0.445	0.005	0.093
One_STC-410	A	0.339	0.006	0.208
One_STR07	A	0.375	0.000	0.143
One_SUMO1-6	C	0.216	0.019	0.045
One_sys1-230	C	0.423	0.002	0.119
One_taf12-248	C	0.007	0.010	0.050
One_Tf_ex11-750	A	0.367	0.002	0.162
One_Tf_in3-182	A	0.048	-0.001	0.085
One_tshB-92	C	0.131	-0.015	0.174
One_txnip-401	C	0.004	0.026	0.033
One_U1003-75	B	0.260	0.024	0.145
One_U1004-183	B	0.281	-0.004	0.392
One_U1009-91	B	0.282	-0.012	0.141
One_U1010-81	B	0.063	0.026	0.051
One_U1012-68	B	0.170	0.041	0.115
One_U1013-108	B	0.229	0.028	0.114
One_U1014-74	B	0.266	0.000	0.077
One_U1016-115	B	0.440	0.025	0.086
One_U1024-197	B	0.127	0.027	0.103
One_U1101	B	0.272	0.012	0.068

-continued-

Table 3.–Page 3 of 3.

Assay	Source[a]	H_O	F_{IS}	F_{ST}
One_U1103	B	0.023	-0.018	0.155
One_U1105	B	0.314	0.021	0.178
One_U1201-492	B	0.409	0.002	0.073
One_U1202-1052	B	0.372	0.023	0.088
One_U1203-175	B	0.446	0.000	0.082
One_U1204-53	B	0.323	0.018	0.099
One_U1205-57	B	0.064	0.025	0.275
One_U1206-108	B	0.309	0.012	0.059
One_U1208-67	B	0.369	-0.009	0.077
One_U1209-111	B	0.203	0.016	0.073
One_U1210-173	B	0.194	0.024	0.067
One_U1212-106	B	0.417	0.017	0.107
One_U1214-107	B	0.086	-0.020	0.093
One_U1216-230	B	0.332	0.018	0.105
One_U301-92	A	0.272	0.002	0.084
One_U401-224	A	0.465	0.001	0.071
One_U404-229	A	0.048	0.031	0.073
One_U502-167	A	0.048	-0.009	0.070
One_U503-170	A	0.168	-0.013	0.110
One_U504-141	A	0.377	0.020	0.077
One_vamp5-255	C	0.349	-0.006	0.103
One_vatf-214	C	0.061	0.027	0.192
One_VIM-569	A	0.203	0.016	0.131
One_ZNF-61	A	0.302	0.005	0.152
One_Zp3b-49	A	0.117	0.011	0.414
One_CO1_Cytb17-26	A	---	0.000	0.394
One_MHC2_190-251	A	---	0.000	0.259
Average/Overall[e]		0.260	0.007	0.130

[a] A) Gene Conservation Laboratory of the Alaska Department of Fish and Game; B) International Program for Salmon Ecological Genetics at the University of Washington; C) Hagerman Genetics Laboratory of the Columbia River Inter-Tribal Fish Commission; and D) Molecular Genetics Laboratory at the Canadian Department of Fisheries and Oceans (Habicht et al. 2010).

[b] These SNPs were combined into haplotypes and treated together as a single mtDNA locus, *One_CO1_Cytb17-26*.

[c] These SNPs were dropped due to linkage.

[d] These SNPs were combined into haplotypes and treated together as a single locus, *One_MHC2_190-251*.

[e] Overall F_{ST} was calculated from the 92 loci selected for further analysis.

Table 4.–Allocation proportions, standard deviation (SD), and 90% credibility interval (CI) for mixtures of known fish (n = number of fish) removed from the baseline populations that contribute to each reporting group (proof tests).

Mixture Origin		Reporting Group							
n		Crescent	West	JCL	SusYen	Fish	KTNE	Kenai	Kasilof
Crescent									
100	Proportion	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	SD	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Lower 90% CI	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Upper 90% CI	1.00	0.03	0.01	0.01	0.01	0.01	0.01	0.01
West									
200	Proportion	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
	SD	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Lower 90% CI	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00
	Upper 90% CI	0.00	1.00	0.01	0.00	0.00	0.00	0.01	0.01
JCL									
200	Proportion	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00
	SD	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
	Lower 90% CI	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00
	Upper 90% CI	0.00	0.00	1.00	0.02	0.00	0.00	0.00	0.00
SusYen									
200	Proportion	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
	SD	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	Lower 90% CI	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00
	Upper 90% CI	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00
Fish									
100	Proportion	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00
	SD	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	Lower 90% CI	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.00
	Upper 90% CI	0.01	0.01	0.01	0.01	1.00	0.01	0.01	0.01
KTNE									
200	Proportion	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00
	SD	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
	Lower 90% CI	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.00
	Upper 90% CI	0.00	0.00	0.00	0.00	0.02	1.00	0.00	0.00
Kenai									
200	Proportion	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
	SD	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00
	Lower 90% CI	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00
	Upper 90% CI	0.00	0.00	0.02	0.00	0.01	0.01	1.00	0.01
Kasilof									
200	Proportion	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	Lower 90% CI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98
	Upper 90% CI	0.00	0.00	0.00	0.00	0.00	0.00	0.01	1.00

Note: Shaded cells show correct allocations.

Note: Proportions for a given mixture may not sum to 1 due to rounding error.

Table 5.–Stock composition estimates, standard deviation (SD), 90% credibility interval (CI), sample size (n) for mixtures of sockeye salmon sampled for genetic analysis from Crescent, Susitna, Yentna, Kenai, and Kasilof rivers and Fish Creek (escapement mixtures).

Mixture Origin			Reporting Group							
Year	n		Crescent	West	JCL	SusYen	Fish	KTNE	Kenai	Kasilof
Crescent fish wheel and lake										
1992/1995	188	Proportion	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		SD	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Lower 90% CI	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Upper 90% CI	1.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Susitna fish wheel										
1992	189	Proportion	0.00	0.00	0.38	0.61	0.00	0.01	0.00	0.00
		SD	0.00	0.01	0.04	0.04	0.00	0.01	0.00	0.00
		Lower 90% CI	0.00	0.00	0.32	0.54	0.00	0.00	0.00	0.00
		Upper 90% CI	0.00	0.01	0.44	0.67	0.01	0.02	0.00	0.00
2008	253	Proportion	0.00	0.00	0.55	0.43	0.00	0.00	0.01	0.00
		SD	0.00	0.00	0.03	0.03	0.00	0.00	0.01	0.00
		Lower 90% CI	0.00	0.00	0.50	0.38	0.00	0.00	0.00	0.00
		Upper 90% CI	0.00	0.01	0.61	0.48	0.00	0.01	0.02	0.00
Yentna fish wheel										
2005	190	Proportion	0.00	0.01	0.56	0.42	0.00	0.00	0.00	0.00
		SD	0.00	0.01	0.04	0.04	0.00	0.00	0.00	0.00
		Lower 90% CI	0.00	0.00	0.50	0.36	0.00	0.00	0.00	0.00
		Upper 90% CI	0.01	0.04	0.62	0.48	0.00	0.01	0.00	0.00
2008	347	Proportion	0.00	0.12	0.37	0.51	0.00	0.00	0.00	0.00
		SD	0.00	0.02	0.03	0.03	0.00	0.00	0.00	0.00
		Lower 90% CI	0.00	0.09	0.32	0.46	0.00	0.00	0.00	0.00
		Upper 90% CI	0.00	0.16	0.42	0.55	0.00	0.00	0.00	0.00
2010	1043	Proportion	0.00	0.06	0.59	0.35	0.00	0.00	0.00	0.00
		SD	0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.00
		Lower 90% CI	0.00	0.04	0.56	0.33	0.00	0.00	0.00	0.00
		Upper 90% CI	0.00	0.07	0.62	0.38	0.00	0.00	0.00	0.00
Fish Creek Weir										
1993	95	Proportion	0.00	0.00	0.00	0.00	0.97	0.02	0.00	0.00
		SD	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00
		Lower 90% CI	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.00
		Upper 90% CI	0.01	0.01	0.01	0.01	1.00	0.08	0.01	0.01
2008	190	Proportion	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00
		SD	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00
		Lower 90% CI	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.00
		Upper 90% CI	0.02	0.01	0.01	0.00	1.00	0.01	0.01	0.01
Kenai River fish wheel										
2005	190	Proportion	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
		SD	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
		Lower 90% CI	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00
		Upper 90% CI	0.00	0.01	0.01	0.00	0.00	0.01	1.00	0.02
Kenai River inclined-plane trap										
2010	420	Proportion	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
		SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Lower 90% CI	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
		Upper 90% CI	0.00	0.00	0.01	0.01	0.00	0.00	1.00	0.00
Kasilof River fish wheel										
2005	189	Proportion	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.98
		SD	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
		Lower 90% CI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.97
		Upper 90% CI	0.01	0.01	0.00	0.01	0.00	0.00	0.03	1.00

Note: Shaded cells show correct allocations.

Note: Proportions for a given mixture may not sum to 1 due to rounding error.

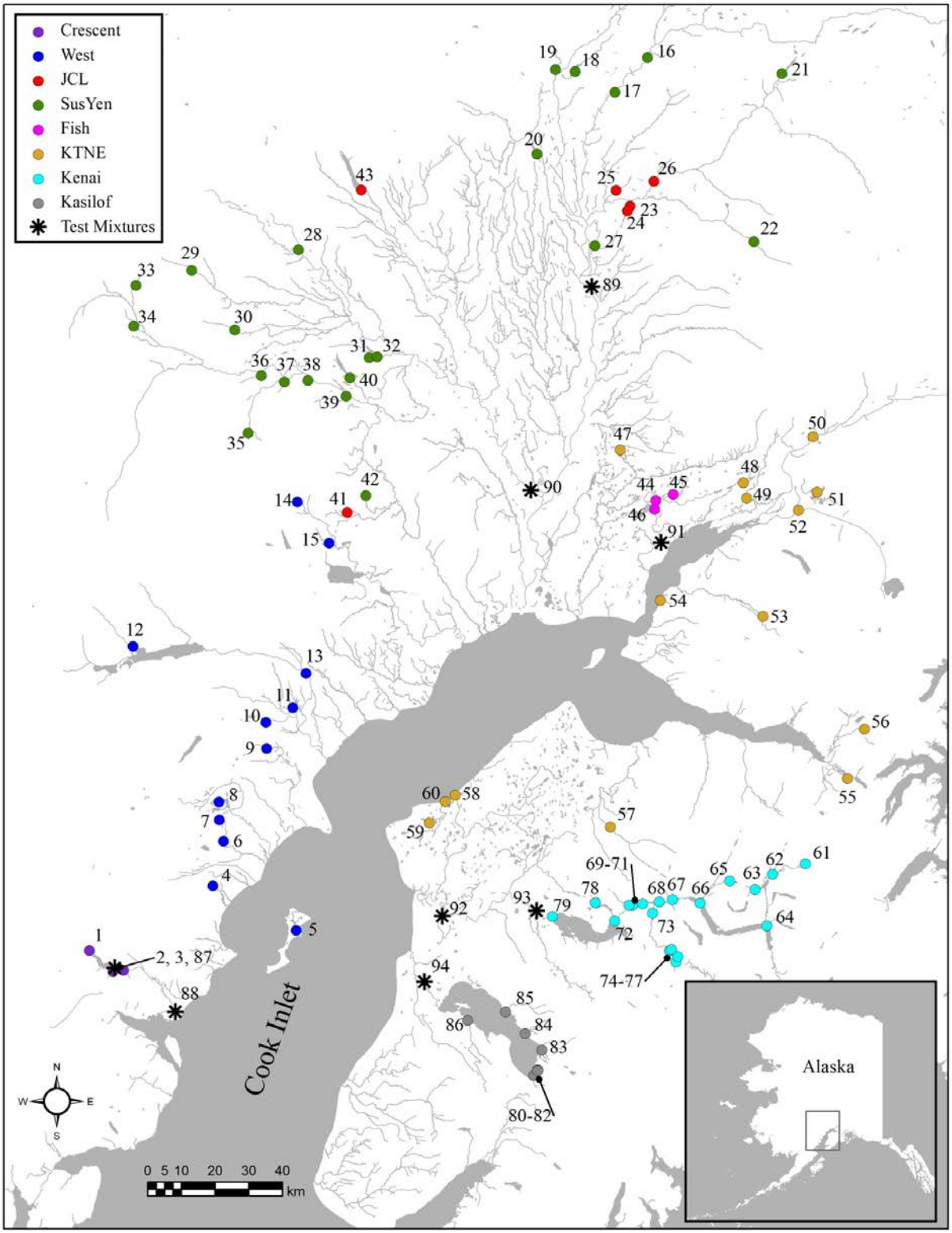

Figure 1.–Sampling locations for sockeye salmon in Upper Cook Inlet, Alaska, 1992–2011 used to compile the genetic baseline.

Note: Numbers correspond to map numbers on Tables 1 and 2.
Note: Colors for each reporting group and symbols for escapement mixture samples are indicated in the legend.

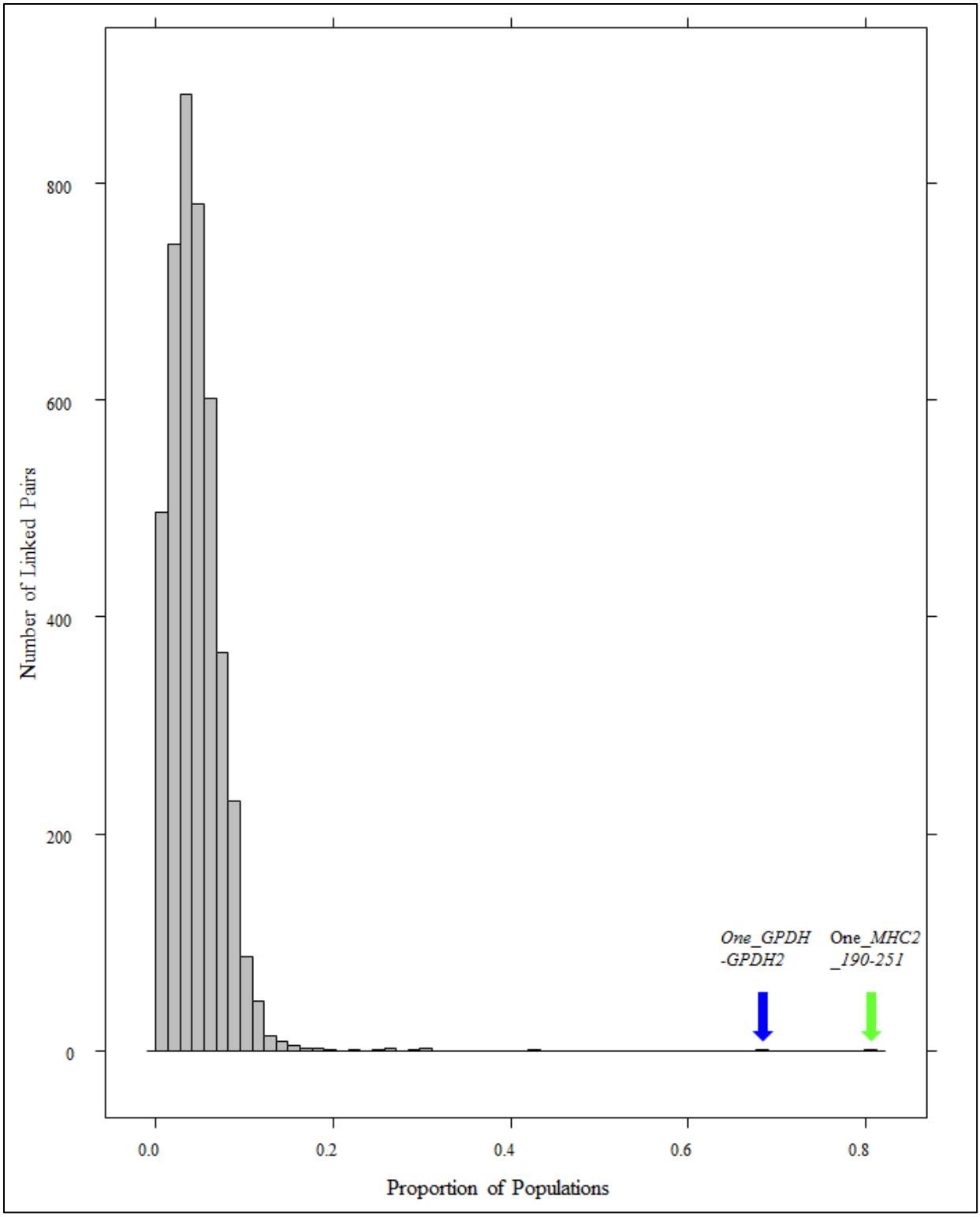

Figure 2.– The proportion of UCI populations showing significant linkage ($P < 0.05$) for the 4,278 SNP pairs.

Note: Arrows point to the SNP pairs that were significantly linked in >50% of populations.

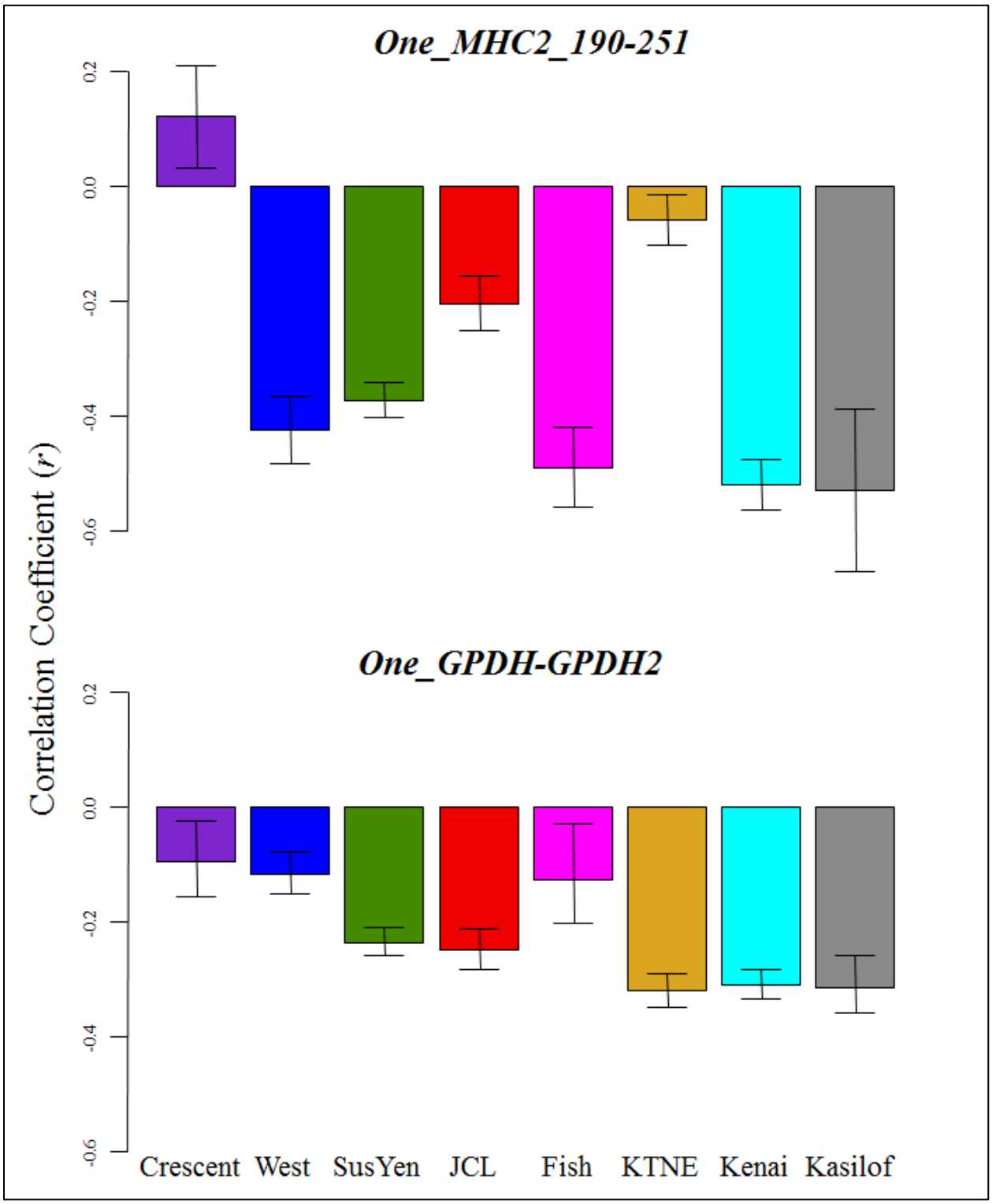

Figure 3.– Correlation coefficient (*r*) by reporting group for linked SNP pairs *One_MHC2_190-251* and *One_GPDH-GPDH2*.

Note: See text for details.

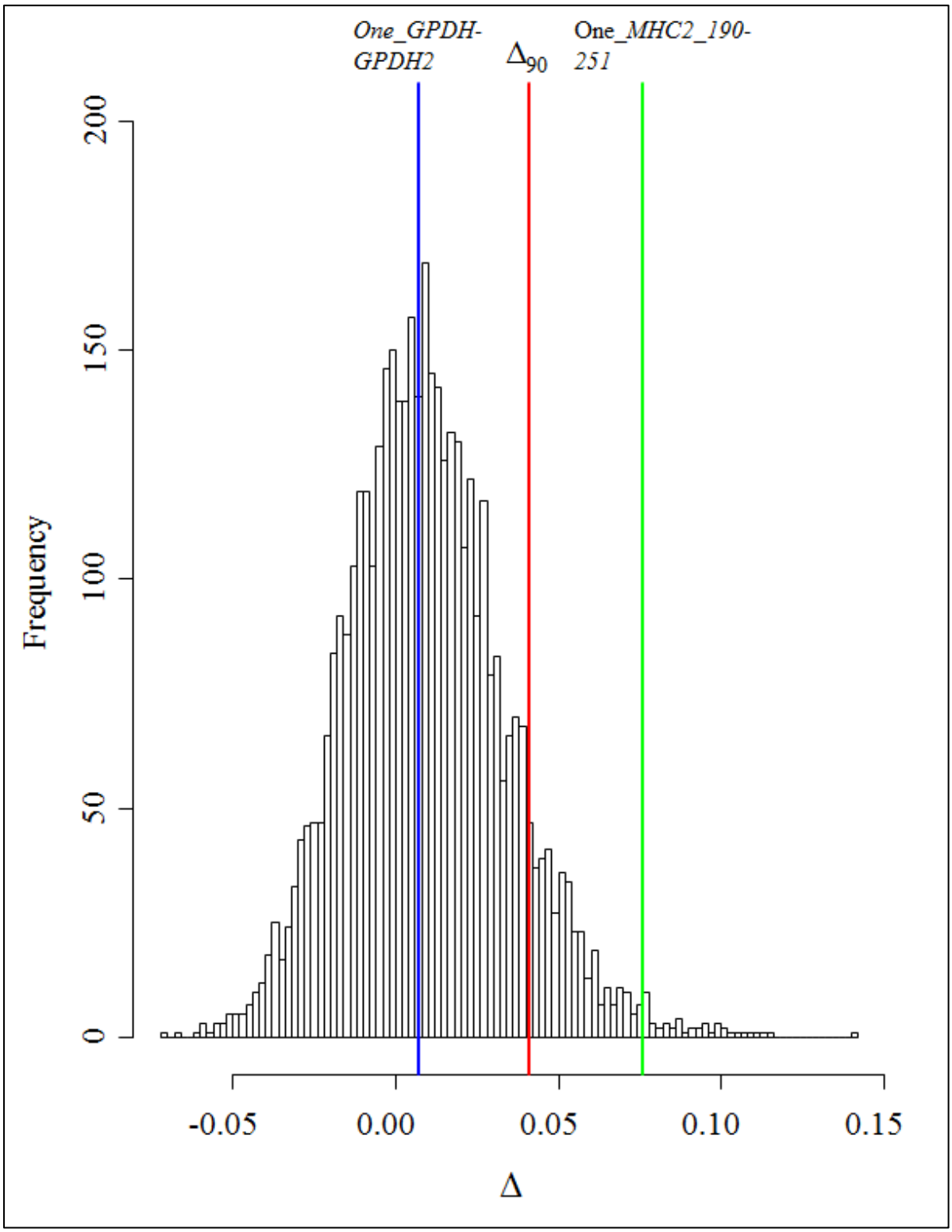

Figure 4.– The distribution of Δ for 4,278 SNP pairs with Δ_{90} in red and the Δ values for the *One_GPDH-GPDH2* pair in blue and the *One_MHC2_190-251* pair in green.

Note: See text for details.

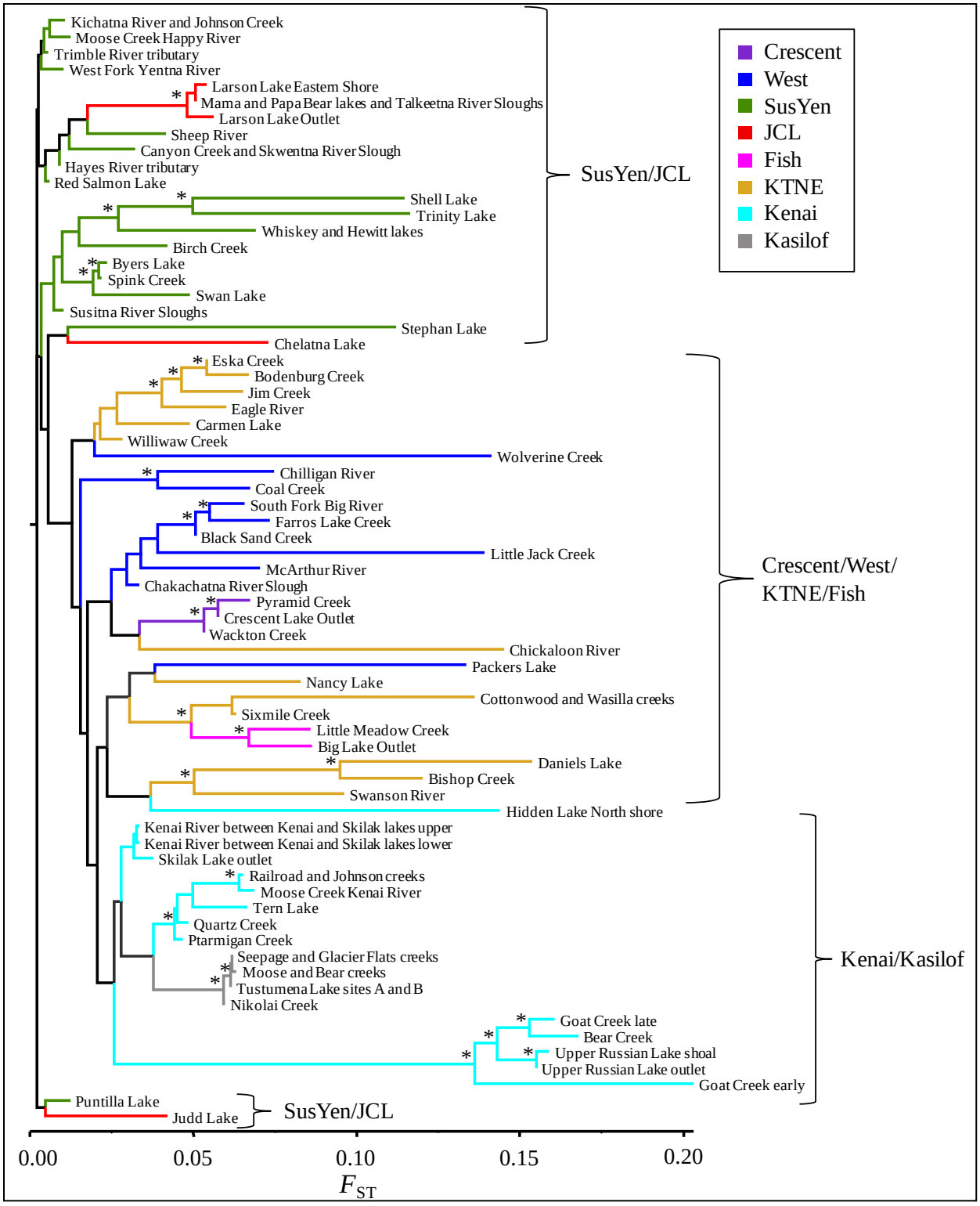

Figure 5.–Consensus Neighbor-Joining tree based on F_{ST} between sockeye salmon populations sampled from spawning areas in drainages of Upper Cook Inlet, Alaska.

Note: See Table 1 for collection details.
Note: Population clusters within reporting groups are noted.
Note: Colors denote reporting groups as in Figure 1. Bootstrap consensus nodes occurring in >50% of trees are marked with an asterisk.