Regional Information Report 5J12-09

Western Alaska Salmon Stock Identification Program Technical Document 4: Status of the SNP Baseline for Chum Salmon

by

James R. Jasper,

Nick DeCovich,

and

William D. Templin

May 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, Special Publications and the Division of Commercial Fisheries Regional Reports. 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

REGIONAL INFORMATION REPORT 5J12-09

WESTERN ALASKA SALMON STOCK IDENTIFICATION PROGRAM TECHNICAL DOCUMENT 4: STATUS OF THE SNP BASELINE FOR CHUM SALMON

by
James R. Jasper, Nick DeCovich, William D. Templin
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

May 2012

The Regional Information Report Series was established in 1987 and was redefined in 2006 to meet the Division of Commercial Fisheries regional need for publishing and archiving information such as project operational plans, area management plans, budgetary information, staff comments and opinions to Board of Fisheries proposals, interim or preliminary data and grant agency reports, special meeting or minor workshop results and other regional information not generally reported elsewhere. Reports in this series may contain raw data and preliminary results. Reports in this series receive varying degrees of regional, biometric and editorial review; information in this series may be subsequently finalized and published in a different department reporting series or in the formal literature. Please contact the author or the Division of Commercial Fisheries if in doubt of the level of review or preliminary nature of the data reported. Regional Information Reports are available through the Alaska State Library and on the Internet at http://www.adfg.alaska.gov/sf/publications/.

Note: This document serves as a record of communication between the Alaska Department of Fish and Game Commercial Fisheries Division and the Western Alaska Salmon Stock Identification Program Technical Committee. As such, these documents serve diverse ad hoc information purposes and may contain basic, uninterpreted data. The contents of this document have not been subjected to review and should not be cited or distributed without the permission of the authors or the Commercial Fisheries Division.

Note: The appearance of product names or specific company names is not an Alaska Department of Fish and Game recommendation for or implied endorsement. The Alaska Department of Fish and Game, in accordance with State of Alaska ethics laws, does not favor one group over another through endorsement or recommendation.

James R. Jasper, Nick DeCovich, William D. Templin
Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory,
333 Raspberry Road, Anchorage, Alaska 99518-1565

This document should be cited as:
Jasper, J. R., N. DeCovich, and W. D. Templin. 2012. Western Alaska Salmon Stock Identification Program Technical Document 4: Status of the SNP baseline for chum salmon. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-09, Anchorage.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Uncertainty about the magnitude, frequency, location, and timing of the nonlocal harvest of sockeye and chum salmon in Western Alaska fisheries was the impetus for the Western Alaska Salmon Stock Identification Project (WASSIP). The project was designed to use genetic data in mixed stock analysis (MSA) to reduce this uncertainty. A baseline of allele frequencies is required for use in mixed stock analysis to estimate the stock of origin of harvested fish. The single nucleotide polymorphism (SNP) baseline for chum salmon *Oncorhynchus keta* to be used for MSA in WASSIP is in a state of perpetual improvement. We collected baseline samples from spawning populations or obtained them from existing agency archives from throughout the range of chum salmon in the Pacific Rim. We constructed a baseline that was current through the 2008 collection season by screening available collections for 53 SNPs. A total of 16,036 individuals from 202 collections representing 153 populations were genotyped. The data used in this project was generated by multiple projects; therefore overall quality control statistics on successfully assayed genotypes are not available, but an example analysis of 3,886 individuals from 38 populations had an overall failure rate of less than 3%, and an error rate of less than 1%. We tested populations for conformance to Hardy-Weinberg expectations and gametic linkage disequilibrium, estimated heterozygosities and F_{ST}. Five locus pairs were found to have significant linkage disequilibrium and were removed, leaving the final suite of 53 SNP markers. Observed heterozygosity was lower than expected heterozygosity at every nuclear marker and over-all F_{ST} was 0.092. Population structure visualized at fine- and broad-scale levels with trees of genetic distances was concordant with past analyses. Simulations using the current baseline indicate that 12 of the 17 regions can be distinguished from each other with a high degree of accuracy (mean >90%). This preliminary analysis indicates that there is genetic similarity within fine-scale groups of populations in coastal western Alaska, which have been historically difficult to differentiate based on genetic marker. As yet there is not sufficient difference to produce reliable efforts from MSA, but increased resolution in the future based on an increased number of SNP markers should allow WASSIP to better distinguish among populations and regions in future MSAs.

Key words: Western Alaska Salmon Stock Identification Project, WASSIP, chum salmon, *Oncorhynchus keta*, mixed stock analysis, single nucleotide polymorphism, genetic baseline

INTRODUCTION

Under the Western Alaska Salmon Stock Identification Program (WASSIP), mixed stock analysis (MSA) to estimate the relative stock contribution of catches will be accomplished using the single nucleotide polymorphism (SNP) baseline for chum salmon. Original MSA analyses of harvests in this area were accomplished with a coastwide baseline of allozyme data that was developed in a multilaboratory effort (Kondzela et al. 2002, Seeb et al. 2004), but this baseline has been replaced with ones based on newer markers, which provide improved resolution and greater laboratory efficiency. A coastwide microsatellite baseline has been recently completed (Beacham et al. 2009), however, early in the process the decision was made to pursue a baseline using SNP markers. This decision was based on the automatic standardization of SNP markers, high throughput capabilities available through the infrastructure in the Alaska Department of Fish and Game (ADF&G) laboratory, relative genotyping costs, and the ability to access more of the genome than is available through microsatellites. The baseline of SNP markers has been in a state of continual development for more than five years and through the WASSIP project it is expected that it will become a fully functioning, coastwide replacement for the previous allozyme baseline.

The suite of SNP markers screened for the baseline has changed through time and will continue to grow or change as more markers become available. We currently screen for 60 nuclear and 3 mitochondrial markers, but the WASSIP Advisory Panel has requested that 96 SNP markers be incorporated into the baseline to improve the precision and accuracy of stock composition estimates. To meet this request, we are contracting the development of at least 33 novel SNP

markers that are targeted to differentiate among chum salmon populations spawning within western Alaska and the Alaska Peninsula drainages (Dann et al. 2012b). These new SNP markers will be assessed after screening a fraction of the baseline and the best-performing SNP markers will be added to the baseline during the winter of 2009/2010.

Here we present the current state of the chum salmon baseline based on samples collected through the 2008 collection season and genotyped for the currently available SNP markers. This analysis is not as developed as the analysis of the sockeye baseline (Dann et al. 2012a) for several reasons. First, much of the sockeye baseline needed to be analyzed and tested in preparation for ongoing MSA applications in the Bristol Bay and North Peninsula fisheries. Second, improvements to the chum salmon baseline are generally hindered by the lack of resolution among population groups in western Alaska. The resolving power of the current set of SNP markers is demonstrated in this document, but it will be more efficient to hold more in-depth analyses of population structure until after the new SNP markers have been developed and applied.

METHODS

TISSUE SAMPLING

Baseline samples for SNP analyses were collected from spawning populations or obtained from existing agency archives from throughout the range of chum salmon in the Pacific Rim. Many of the available samples were available from the samples used in the published survey of allozyme variation (Seeb et al. 2004). Target sample size for baseline collections was 100 individuals across all years for each population to achieve acceptable precision for the allele frequency estimates (Allendorf and Phelps 1981; Waples 1990a).

LABORATORY ANALYSIS

Assaying genotypes

Genomic DNA was extracted using a DNeasy® 96 Tissue Kit by QIAGEN® (Valencia, CA). Baseline population samples were genotyped using uniplex SNP genotyping performed in 384-well reaction plates and also by using the 48.48 array (Fluidigm Corporation) where 48 of the 52 markers were assayed in sets of 48 fish and the remaining markers were assayed on the 384-well platform. Laboratory methods followed the 5′ nuclease methods described in Seeb et al. (2009). 30 assays originated from Elfstrom et al. (2007), sixteen from Smith et al. (2005a), and 7 from Smith et al. (2005b). With either platform, genotypes from generally 384 fish were visualized using the GeneMapper (uniplex platform; Applied Biosystems) and BioMark (array platform; Fluidigm Corporation) software programs and scored for each marker by 2 people simultaneously. Scores were entered and archived in the Gene Conservation Laboratory Oracle database, LOKI.

Quality control

Three measures were taken to ensure quality control of the baseline data:

1. Regenotyping of samples: 8% of each collection was regenotyped for all markers to ensure that genotypes were reproducible, to identify laboratory errors, and to measure rates of inconsistencies during repeated analyses on the uniplex and array platforms. We

report error rates for a representative baseline project which consisted of 38 baseline collections comprising 3,886 individuals (~ 24% of current baseline).

2. Exclusion of individuals with an excessive dropout rates: A threshold of 80% scorable loci per individual was established and all individuals that did not meet this threshold were excluded from further analyses. This threshold was set to exclude individuals with poor quality DNA. Poor quality DNA leads to lower reproducibility and therefore adds error to the allele frequency estimates. The value of 80% was chosen based upon the observation that many individuals with high quality DNA had some dropouts, but generally less than 20% of markers, while those with poor quality DNA had higher dropout rates. As a result, there was little difference in which individuals were excluded from analysis when picking the threshold as long as it was within the 70% to 90% range.

   This rule (referred to as the "80% rule") will also be used for samples from fishery harvests to decrease errors and estimate variances caused by poor quality DNA and missing data. This approach is an attempt to balance the benefits from better data with the loss of power to accurately and precisely estimate stock proportions due to smaller sample sizes. One other potential disadvantage of this approach is the potential to introduce another form of bias if fish that are removed from analyses are not randomly distributed in the mixture. Heterogeneity in sample removal may introduce bias in subsequent estimates of stock proportions when samples with quality genotypic data are not representative of the entire harvest being sampled. We anticipate that bias will only be a concern if significant proportions of mixtures are excluded.

3. Finally, we searched for suspected duplicate fish within collections by identifying pairs of individuals that had identical multilocus genotypes at 38 or more loci. If suspected duplicates were found, the second individual in each matching pair was removed from further analyses.

STATISTICAL ANALYSIS

Heterozygosity and F_{ST}

Genotypic data were retrieved from LOKI database and were used to calculate allele frequencies. Observed heterozygosity, expected heterozygosity, and F_{ST} (Weir and Cockerham 1984) were calculated for all markers using the program GDA v1.1 (Lewis and Zaykin 2001).

Gametic disequilibrium

All pairs of nuclear markers were tested for gametic disequilibrium within each collection using GDA. We defined a pair of markers to be significantly out of gametic equilibrium if tests for gametic disequilibrium were significant ($P < 0.01$) for greater than half of all collections. When gametic linkage was significant, the SNP with the lowest F_{ST} in the pair was dropped. All mtDNA markers were combined into a single locus. Markers that did not exhibit gametic disequilibrium with any other markers, retained markers from marker pairs that exhibited gemetic disequilibrium, and the combined mtDNA markers were defined as loci for the remaining analyses.

Pooling collections into populations

Collections taken at the same location at similar calendar days in different years were pooled as suggested by Waples (1990b). Jasper et al. (2012a) has a more detailed investigation of temporal

variation among collections taken in different years at the same site and calendar time. Samples taken at the same location, but at substantially different calendar days, and samples taken from geographically proximate locations were tested for homogeneity using a chi-square test of allele frequency distributions across all loci. Groups of collections that demonstrated homogeneity ($P > 0.01$, not corrected for multiple tests) were pooled. The pooled and the remaining unpooled collections were defined as populations in further analyses. Our protocol was to drop populations from further analyses if they were represented by sample sizes of less than 30 fish. Due to the difficulty of obtaining individuals for baseline collections, this threshold is much smaller than that used for sockeye salmon and allows for more complete representation of populations in this preliminary analysis. When the baseline is completed, we expect to use a higher threshold.

Hardy-Weinberg equilibrium

Genotype distributions within collections were tested for deviation from Hardy-Weinberg expectation (H-W) using GDA v1.1. These tests were repeated once collections were pooled into populations. For H-W, critical values ($\alpha = 0.05$) were adjusted for multiple tests within markers among collections and multiple tests across markers within collections (Rice 1989).

Population structure visualization

Genetic distances between populations were measured using pair-wise F_{ST} (Weir and Cockerham 1984) calculated from the 53 SNP loci. Pair-wise F_{ST}'s were chosen instead of Cavalli-Sforza and Edwards (CSE) chord distances, which are subject to bias, because sample sizes were nonuniform and relatively small in some instances. To visualize genetic population structure, F_{ST} distances were plotted as a tree using the unweighted pair group method with arithmetic mean (UPGMA) algorithm. Two tree plots were produced: 1) all baseline populations and 2) restricted to populations from Western Alaska and the Alaska Peninsula (WAAP).

Baseline evaluation for MSA

Reporting groups were delineated based on geographic regions that were thought to be both identifiable and applicable for MSA analyses of mixtures sampled under the WASSIP program. During estimation of stock composition, 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 then assessed the potential of the baseline to identify these reporting groups for MSA applications with simulations. For the simulations, we generated 400 fish based on the population-specific allele frequencies from all the populations within each reporting group (i.e., 100% simulations). This process was repeated 1,000 times, and the mean and central 90% of the distribution of estimates were reported as the estimate and the 90% confidence interval. Simulated mixtures were analyzed using SPAM version 3.7b (Debevec et al. 2000; ADF&G 2001). A critical level of 90% correct allocation was used to determine if the reporting group was acceptably identifiable (e.g., Seeb et al. 2000). ONCOR (Kalinowski 2007) was not used for this analysis because it does not accommodate mtDNA loci. Baseline and mixture genotypes were randomly generated from the baseline allele frequencies assuming Hardy-Weinberg equilibrium.

RESULTS

TISSUE SAMPLING

A total of 16,036 individuals from 202 collections representing 153 putative populations (Table 1; Figures 1 and 2) have been genotyped at 53 SNPs. This baseline represents an increase of 34 populations to the 119 population baseline presented by the ADF&G Gene Conservation Laboratory in its proposal to Arctic-Yukon-Kuskokwim Sustainable Salmon Initiative for WASSIP funding in 2008 (reviewed in Weir et al. 2012). Collection sites ranged from Korea to Puget Sound, Washington. The most comprehensive representation in the baseline is from the western Alaska portion of the species range, i.e., populations from rivers draining into the Bering Sea and areas adjacent to the Bering Sea (Figure 1).

LABORATORY ANALYSIS

Assaying genotypes

While 61 SNP markers were available, some of these markers were excluded from this analysis because they were either not screened for the complete set of populations, were found to be out of Hardy-Weinberg equilibrium, or were linked to other markers that were included in the analysis. These issues resulted in a reduced set of 53 chum salmon SNP markers used in this analysis (Table 2); 2 mitochondrial DNA (mtDNA) and 51 nuclear DNA (nDNA).

Quality control

The data used in this project were generated by multiple projects; therefore overall quality control statistics are not available at this time. As an example of the quality control process we present the results from a recent analysis in which 3,886 individuals from 38 populations were analyzed. The overall failure rate for successfully assaying genotypes for this project was < 3%. The quality control checks employed demonstrated an error rate of < 1%. The quality control checks revealed pairs of individuals in some populations that had identical multilocus genotypes. Several populations had individuals with duplicate genotypes that were found to match at 38 or more SNPs, a strong indication that the tissues sampled were actually from the same individual. The second individual in the matching pair was removed from the analysis. All other genotype matches found involved 15 or fewer SNPs, an occurrence that is much more probable by chance when surveying 61 SNPs. These individuals were not removed from the baseline.

STATISTICAL ANALYSIS

Locus selection and gametic disequilibrium

For this project, 61 SNPs were originally surveyed. An initial review of these data found that two markers had no data for more than 80 populations, one was significantly out of H-W equilibrium after correcting for multiple tests, and five locus pairs were found to have significant linkage disequilibrium. After removing both the loci with incomplete data, the one out of H-W equilibrium, and removing the locus with the lowest overall F_{ST} in each linked pair, we arrived at the final suite of 53 SNPs, two mitochondrial and 51 nuclear SNPs, used in this analysis (Table 2).

Heterozygosity and F_{ST}

Observed heterozygosity, expected heterozygosity, and F_{ST} for each of the nuclear markers are included. Observed heterozygosity was lower than expected heterozygosity at every nuclear marker with the averages of 0.271 and 0.300, respectively. Observed heterozygosities ranged widely from 0.017 to 0.474. The F_{ST} estimate over all markers was 0.092 and the individual values ranged from 0.019 to 0.441.

Pooling collections into populations

The 202 collections were pooled to represent the 153 populations by combining collections taken from similar locations over multiple years and from nearby sites that exhibited genetic homogeneity. The average sample size per population was 79 fish. Within WAAP, the smallest population sample size was 46 fish (Goodnews River–North Fork).

Hardy-Weinberg equilibrium

Significant departures from H-W were not found in any of the 51 nuclear SNP markers after correcting for multiple tests (Table 2). Likewise, while almost all populations showed a significant departure from H-W at one or more loci, no population was found to be significantly out of H-W when correcting for multiple tests across loci.

Population structure visualization

Genetic relationships among baseline populations are shown schematically in the UPGMA trees (Figures 3 and 4). On the tree with the whole Pacific Rim baseline (Figure 3), the deepest structure was found between Japan/Korea and all other populations. The Russian populations appear much lower on the tree as a single group associated with the Alaska Peninsula and Gulf of Alaska populations (excluding Washington/Idaho). At this scale there is a strong clustering of populations by region, even within western Alaska. A closer look (Figure 4) shows that while there is intermixing of populations from the Norton Sound, Yukon Alaska Summer, Bristol Bay, and the Kuskokwim Summer groups, generally populations first cluster with populations from the same group before combining with populations from other groups.

Baseline evaluation for MSA

Based on the genetic structure revealed above, 17 reporting groups were delineated based on geographic regions and genetic similarity (Table 1, Figures 1 and 2). Because the WASSIP project is mainly interested in the fisheries of WAAP, 13 of the reporting groups were defined for western Alaska drainages and run times. Populations from outside this area were pooled into four groups. Greater resolution is available within these groups, but this resolution is not necessary for our purposes here.

Simulations using the current baseline indicate that 12 of the 17 regions can be distinguished from each other with a high degree of accuracy (mean >90%, Table 3). Not surprisingly, the regions that fail to be highly distinguishable are the regions included in the large, intermixed cluster seen in Figure 4.

DISCUSSION

This preliminary analysis presents a baseline that is 25% larger than the version previously reviewed by the WASSIP Technical Committee (Weir et al. 2012). The main areas in which

populations were added to the baseline were in Norton Sound, the Yukon River, and Washington State. Given that the main interest for WASSIP is in the harvest from the nearshore marine waters of the western Alaska region, the baseline should be at its most developed in the areas most likely to contribute to these harvests. In most western Alaska fisheries, the expectation is that the majority of the catch will come from "local" western Alaska stocks. However, studies of chum salmon harvests in fisheries along the Alaska Peninsula (Seeb and Crane 1999, Seeb and Crane 2004) have shown that both Asian and eastern Gulf of Alaska stocks can periodically contribute to these harvests. Efforts to augment the baseline further have been halted pending the development of the new set of SNP markers and the eventual increase to 96 SNPs for baseline analysis.

The structure of chum salmon on a coastwide scale has been explored repeatedly and the patterns seen in this analysis show similar results. Japanese and Korean chum salmon populations are the most divergent set in the baseline as seen previously with allozymes (Seeb and Crane 1999) and with microsatellites (Beacham et al. 2009). The location of Russian populations of chum salmon as a single group associated with the Alaska Peninsula and Gulf of Alaska populations (Figure 3) was also not unexpected; similar association between these regional groups was noted in Seeb and Crane (1999) using allozyme loci and was proposed as a possible source of bias through misallocation.

One of the chief areas of concern for distinguishing fine-scale groups of populations is in coastal western Alaska (Norton Sound, Yukon River–Summer run, Kuskokwim River–Summer run, and Bristol Bay). These populations have historically been difficult to differentiate based on genetic markers, yet some means to separate these populations is important for management. This preliminary analysis indicates that there is genetic similarity within these fine-scale groups, but as yet there is not sufficient difference to produce reliable estimates from MSA. Our expectation is that a concerted effort to discover and use SNPs that distinguish populations within this area will eventually allow MSA applications to distinguish among drainages.

ACKNOWLEDGEMENTS

The Technical Document series served as a record of communication between the Alaska Department of Fish and Game Commercial Fisheries Division and the Western Alaska Salmon Stock Identification Program (WASSIP) Technical Committee during the implementation of the program. The authors would like to thank the WASSIP Technical Committee and Advisory Panel for their constructive input on each of the documents throughout the project. The authors would also like to thank Erica Chenoweth who coordinated and prepared the Technical Document series for publication and Publication Specialists Amy Carroll and Joanne MacClellan for implementing the series into Regional Information Reports.

FUTURE ANALYSES

1. Increase sample sizes for collections for which we have existing tissues to be genotyped.
2. Incorporate collections gathered through the 2009 field collection season into baseline analyses.
3. Assess the suite of developing SNPs (see Dann et al. 2012b) for utility in describing genetic variation within the WASSIP study area and for accurately and precisely estimating stock proportions in mixture samples from area fisheries.

4. The corrections for multiple tests resulted in low power to detect significant departures from H-W, so we will examine the number of departures from H-W by marker and by population prior to correcting for multiple tests to assess any patterns in departures from H-W.
5. Perform proof tests with either 200 or 400 fish in reporting groups where adequate numbers of fish exist. This process will also allow us to test the behavior of the baseline in the Bayesian mixed stock analysis model.
6. Investigate the presence and utility of loci identified as under selection.
7. Investigate diversity within and among regions using log-likelihood ratios (*G* statistics), AMOVA, and Nei's gene diversity analysis.
8. For new levels of hierarchy, compare levels of heterogeneity using Fisher's *F*-test to better understand how diversity is distributed in the baseline.
9. Examine the distribution of allelic richness by region and ascertainment region to assess ascertainment bias.
10. Repeat simulations using ONCOR without the mtDNA loci. This will allow the assessment of the baseline using the ideas proposed in Anderson et al. (2008).
11. Utilize statistical methods developed for estimating small proportions to increase the performance of MSA through decreased bias and increased precision. These methods might include the use of informative priors when using Bayesian methods for GSI and the use of a stratified estimate protocol (Jasper et al. 2012b).
12. Investigate the utility of reducing the range of the baseline to include only those populations that are likely to be present in WASSIP mixtures.
13. Assess the possibility of sex linked/associated markers amongst increasing suite of SNPs.

REFERENCES CITED

ADF&G (Alaska Department of Fish and Game). 2001. SPAM Version 3.5: Statistics Program for Analyzing Mixtures. Alaska Department of Fish and Game, Commercial Fisheries Division, Gene Conservation Lab. Available for download from http://www.cf.adfg.state.ak.us/geninfo/research/genetics/software/spampage.php (Accessed May 16, 2012).

Anderson E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for estimating the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 65:1475–1486.

Beacham, T. D., J. R. Candy, K. D. Le, and M. Wetklo. 2009. Population structure of chum salmon (*Onchorhyncus keta*) across the Pacific Rim, determined from microsatellite analysis. Fishery Bulletin 107:244–260.

Dann, T. H., A. Barclay and C. Habicht. 2012a. Western Alaska Salmon Stock Identification Program Technical Document 5: Status of the SNP baseline for sockeye salmon. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-10, Anchorage.

Dann, T. H., J. R. Jasper, H. A. Hoyt, H. Hildebrand, and C. Habicht. 2012b. Western Alaska Salmon Stock Identification Program Technical Document 6: Selection of the 96 SNP marker set for sockeye salmon. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-11, Anchorage.

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 91:509–510.

Elfstrom, C. M., C. T. Smith, and L. W. Seeb. 2007. Thirty-eight single nucleotide polymorphism markers for high-throughput genotyping of chum salmon. Molecular Ecology Notes 7(6):1211–1215.

Jasper, J. R., C. Habicht, and W. D. Templin. 2012a. Western Alaska Salmon Stock Identification Program Technical Document 2: Investigation of temporal variation in sockeye and chum salmon baselines. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-07, Anchorage.

Jasper, J. R., C. Habicht, W. D. Templin. 2012b. Western Alaska Salmon Stock Identification Program Technical Document 3: Estimating small proportions. Alaska Department of Fish and Game, Regional Information Report 5J12-08, Anchorage.

Kalinowski, S. T., K. R. Manlove, and M.L. Taper. 2007. ONCOR a computer program for genetic stock identification. Montana State University, Bozeman. Available at http://www.montana.edu/kalinowski/Software/ONCOR.htm (Accessed May 16, 2012).

Kondzela, C. M., P. A. Crane, S. Urawa, N. V. Varnavskaya, V. Efremov, X. Luan, W. D. Templin, K. Hayashizaki, R. L. Wilmot, and L. W. Seeb. 2002. Development of a comprehensive allozyme baseline for Pacific Rim chum salmon. NPAFC Doc. 629. http://www.npafc.org/new/publications/Documents/PDF%202002/629%28USA%29.pdf (Accessed May 16, 2012).

Lewis, P. O. and D. Zaykin. 2001. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0. URL http://darwin.eeb.uconn.edu/eeb348/archives/000359.html (Accessed April 20, 2012).

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

Seeb, L. W. and P. A. Crane. 1999. Allozymes and mitochondrial DNA discriminate Asian and North American populations of chum salmon in mixed-stock fisheries along the south coast of the Alaska Peninsula. Transactions American Fisheries Society 128:88–103.

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.

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.

REFERENCES CITED (Continued)

Seeb, J. E., C. E. Pascal, R. Ramakrishnan, and L. W. Seeb. 2009. SNP genotyping by the 5'-nuclease reaction: advances in high throughput genotyping with non-model organisms. A. Komar, editor Methods in Molecular Biology, Single Nucleotide Polymorphisms, 2nd Edition. Humana Press.

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., J. Baker, L. Park, L. W. Seeb, C. Elfstrom, S. Abe, and J. E. Seeb. 2005b. Characterization of 13 single nucleotide polymorphism markers for chum salmon. Molecular Ecology Notes 5:259–262.

Waples, R. S., 1990a. Conservation genetics of Pacific salmon. II. Effective population size and the rate of loss of genetic variability. Journal of Heredity. 81:267–276

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.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Weir, B. S., R. S. Waples and T. P. Quinn. 2012. Western Alaska Salmon Stock Identification Program Technical Document 1: Technical committee comments and review of the original proposal. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-06, Anchorage.

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.

TECHNICAL COMMITTEE REVIEW AND COMMENTS

Unedited comments by the WASSIP Technical Committee on documents discussed at 23 September 2009 meeting of the WASSIP Advisory Panel.

Document 4: Status of the SNP baseline for chum salmon

Table 2: results shown in the *P*-HWE column are suspicious. If the test is valid (and Type I error rate is close to the nominal alpha), then the *P* values for conditions where the null hypothesis is satisfied should show an even distribution across the range 0-1. Most of these values are skewed toward very high values, suggesting that the test is strongly biased against finding statistical significance.

TABLES

Table 1.–Baseline collection information organized geographically by reporting group and subdivided by population. Each line contains an individual collection with associated collection name, collection year, and sample size. Some collections were pooled based on geographic proximity and tests of homogeneity (see text for methods).

Region	Population	Collection	Year	N
Japan/Korea				
	Chitose River	Chitose River - early	2003	79
	Chitose River	Chitose River - late	2003	80
	Gakko River	Gakko River - early	2003	79
	Kushiro River	Kushiro River	1998	79
	Sasauchi River	Sasauchi River	1990	78
	Shari River	Shari River	2001	77
	Shibetsu River	Shibetsu River	2003	78
	Shinzunai River	Shinzunai River	2002	80
	Tokachi River	Tokachi River	2002	79
		Tokachi River	1990	80
	Tokoro River	Tokoro River	2005	100
	Tokushibetsu River	Tokushibetsu River	2004	80
	Tsugaruishi River	Tsugaruishi River	1999	80
	Yurappu River	Yurappu River - early	1997	80
	Yurappu River	Yurappu River - late	1997	80
	Namdae River	Namdae River - Female	2005	96
		Namdae River - Male	2005	96
Russia				
	Amur River	Amur River - summer	1997	60
		Amur River - summer	2001	99
	Anadyr River	Anadyr River - early	2000	28
		Anadyr River - early	1993	31
	Apuka River	Apuka River	2002	49
	Bistraya River	Bistraya River	1998	69
	Bolshaya River	Bolshaya River	1997	96
	Kamchatka	Kamchatka - early	2003	50
		Kamchatka - early	1990	50
	Palana River	Palana River	1998	95

-continued-

Table 1. Page 2 of 8.

Region	Population	Collection	Year	N
	Pymta	Pymta	1993	50
	Tigil River	Tigil River	2002	44
Kotzebue Sound				
	Noatak River	Noatak River - above hatchery	1991	95
	Kelly Lake	Kelly Lake - Noatak River	1991	95
	Kiana River	Kiana River	2004	95
	Kobuk River	Kobuk River	2005	95
		Kobuk - Salmon River	1991	95
	Selby Slough	Selby Slough	1994	95
Seward Peninsula				
	Agiapuk River	Agiapuk River	2005	94
	American River	American River	2004	95
Norton Sound				
	Eldorado River	Eldorado River	2005	94
	Fish River	Fish River	2004	95
	Kwiniuk River	Kwiniuk River	2004	189
	Niukluk River	Niukluk River	2004	95
	Nome River	Nome River	2005	190
	Pikmiktalik River	Pikmiktalik River	2005	95
	Pilgrim River	Pilgrim River	1994	90
		Pilgrim River	2005	94
	Shaktoolik River	Shaktoolik River	2005	95
	Snake River	Snake River	1993	35
		Snake River	1995	58
		Snake River	2005	95
	Solomon River	Solomon River	1993	2
		Solomon River	1996	5
		Solomon River	1995	65
	Unalakleet River	Unalakleet River	1992	48
		Unalakleet River	2004	95
	Ungalik River	Ungalik River	2005	54

-continued-

Table 1. Page 3 of 8.

Region	Population	Collection	Year	N
Yukon Alaska Summer				
	Black River	Black River	2006	95
	Andreafsky River	West Fork Andreafsky River	1993	94
		East Fork Andreafsky River	1993	95
		Andreafsky River - East Fork weir	2004	94
	Atchuelinguk River	Atchuelinguk River	1989	51
	Anvik River	Swift River	1992	94
		Yellow River	1992	80
		Otter Creek	1993	96
		Beaver Creek	1993	95
		Beaver Creek	1992	15
	California Creek	California Creek	1997	93
	Gisasa River	Gisasa River	1994	95
	Innoko River	Innoko River	1993	86
	Kaltag River	Kaltag River	1992	93
	Melozitna River	Melozitna River	2003	94
	Nulato River	Nulato River	1994	95
	Rodo River	Rodo River	1989	73
	Tolstoi Creek	Tolstoi Creek	1997	95
	Chulinak	Chulinak	1989	92
	Clear Creek	Clear Creek	1995	94
	Melozitna River	Melozi Hot Springs Creek	1994	95
	Tozitna River	Tozitna River	2003	95
	Koyukuk River	Henshaw Creek weir - early	2004	94
		Huslia River - Early	1993	95
Yukon Alaska Fall/Middle				
	Big Salt River	Big Salt River	2001	71
	Black River	Black River	1995	95
	Bluff Cabin	Bluff Cabin	1992	95
	Chandalar River	Chandalar River	2001	95
	Chena River	Chena River	1994	95

-continued-

Table 1. Page 4 of 8.

Region	Population	Collection	Year	N
	Delta River	Delta River	1992	95
		Delta River	1994	95
	Koyukuk River	Henshaw Creek weir - late	1995	62
	Kantishna River	Kantishna River	2001	94
	Salcha River	Salcha River	2001	85
	Sheenjek River	Sheenjek River	1992	96
	Tanana River	Tanana River Mainstem	1993	48
	Toklat River	Geiger Creek	1994	95
		Sushana River	1994	95
Yukon Canada				
	Fishing Branch	Fishing Branch	1994	95
	Porcupine River	Old Crow	2007	92
	Big Creek	Big Creek	1995	95
	Donjek River	Donjek River	1994	73
	Kluane River	Kluane River	2001	93
		Kluane River	2007	33
	Minto Slough	Minto Slough	1989	92
	Pelly River	Pelly River	1993	84
	Tatchun Creek	Tatchun Creek	1992	93
	Teslin River	Teslin River	1992	93
Kuskokwim Bay				
	Goodnews River	Goodnews River - North Fork	2006	46
		Goodnews Weir	1991	100
	Kanektok River	Kanektok River	1994	95
Kuskokwim Summer				
	Holokuk River	Holokuk River	1995	48
		Holokuk River	2007	62
	Tuluksak River Weir	Tuluksak River Weir	2007	198
	Kasigluk River	Kasigluk River	1994	70
		Kisaralik River	1994	95
	Kogrukluk River	Kogrukluk River	1992	44

-continued-

Table 1. Page 5 of 8.

Region	Population	Collection	Year	N
		Kogrukluk River	1993	50
	Kwethluk River	Kwethluk River	2007	198
		Kwethluk River	1994	96
	Aniak River	Aniak River	1992	94
	George River	George River	1996	95
		George River	2007	289
	Nunsatuk River - (Set A)	Nunsatuk River	1994	96
	Oskawalik River - (Set D)	Oskawalik River	1994	58
	Stony River	Stony River - early	1994	95
		Stony River - late	1994	56
		Necons River	2006	6
		Necons River	2007	127
	Tatlawiksuk River	Tatlawiksuk River weir	2007	298
	Takotna River	Takotna River - 4th of July Creek	1994	95
Kuskokwim Fall				
	South Fork Kuskokwim	South Fork Kuskokwim	1995	95
	Big River	Big River	1996	95
Nunivak Island				
	Dahloongamiut River	Dahloongamiut River	2006	95
Bristol Bay				
	Togiak River	Togiak River	1993	95
	Mulchatna River	Mulchatna River	1994	95
	Stuyahok River	Stuyahok River	1992	31
		Stuyahok River	1993	56
	Big Creek	Big Creek	1993	80
	Pumice Creek	Pumice Creek	1993	95
	Whale Mountain Creek	Whale Mountain Creek	1993	95
North Peninsula				
	Frosty Creek	Frosty Creek	1992	95
	Ilnik River	Ilnik River	2002	50
	Joshua Green	Joshua Green	1994	98

-continued-

Table 1. Page 6 of 8.

Region	Population	Collection	Year	N
	Lawrence Valley	Lawrence Valley	1992	95
	Meshik River	Meshik River	1992	78
	Moller Bay	Moller Bay	1998	95
	North of Cape Seniavin	North of Cape Seniavin	2001	54
	Plenty Bear Creek	Plenty Bear Creek	1993	92
South Peninsula				
	Alagogshak River	Alagogshak River	1993	88
	Canoe Bay Creek	Canoe Bay Creek	1992	94
	Little John Lagoon	Little John Lagoon	1992	80
	Volcano Bay	Volcano Bay	1996	42
	Volcano River	Volcano River	1992	64
Western Gulf of Alaska				
	American River	American River	1992	95
	Big Sukoi	Big Sukoi	1992	95
	Sturgeon River	Sturgeon River	1992	71
	McNeil River	McNeil River Lagoon	1994	60
		McNeil River	1996	49
	Chunilna River	Chunilna River	1993	87
	Susitna River	Susitna River	1996	95
	Lake Creek	Lake Creek	1996	95
	Olsen Creek	Olsen Creek	1995	95
	WHN Hatchery	WHN Hatchery	1992	87
Eastern Gulf of Alaska				
	Chilkat River	Chilkat River	2006	93
	DIPAC Hatchery	DIPAC Hatchery	2006	95
	Hidden Falls Hatchery	Hidden Falls Hatchery	2006	95
	Long Bay	Long Bay	1991	66
		Long Bay	1992	95
	Taku River	Taku River - fall	2006	93
	Disappearance	Disappearance	1998	95
	Fish Creek	Fish Creek - early	1988	50

-continued-

Table 1. Page 7 of 8.

Region	Population	Collection	Year	N
		Fish Creek - late	1988	50
	Karta River	Karta River	2006	56
	North Arm Creek	North Arm Creek	2006	95
	Nekite River	Nekite Channel	1989	48
		Nekite River	1989	48
	Big Mission Creek	Big Mission Creek - fall	2003	47
		Big Mission Creek - fall	2002	47
	Dewatto River	Dewatto River - fall	1998	16
		Dewatto River - fall	1998	63
	Dosewallips River	Dosewallips River - summer	2003	47
		Dosewallips River - summer	2000	46
	Elwha River	Elwha River	2004	95
	Hamma Hamma River	Hamma Hamma River - summer	2001	16
		Hamma Hamma River - summer	2001	47
		Hamma Hamma River - summer	2003	48
	Jimmy Creek	Jimmy Creek - summer	2000	46
		Jimmy Creek - summer	2001	49
	Lilliwaup River	Lilliwaup River - fall	2005	45
		Lilliwaup River - fall	2006	48
	Lilliwaup River	Lilliwaup River - summer	2002	43
		Lilliwaup River - summer	2001	48
	Lower Skagit River	Lower Skagit River - fall	1998	91
	Mounts Creek	Mounts Creek - winter	1998	48
	Nisqually River Hatchery	Nisqually River Hatchery	2004	95
	North Creek	North Creek - fall	1994	47
		North Creek - fall	1998	48
	Quilcene	Quilcene - summer	2001	47
		Quilcene Bay - summer	1997	16
	Sherwood Creek	Sherwood Creek - summer	1994	95
	Skamokawa Creek	Skamokawa Creek - fall	2000	3
		Skamokawa Creek - fall	2001	4

-continued-

Table 1. Page 8 of 8.

Region	Population	Collection	Year	N
		Skamokawa Creek - fall	2002	72
	Union River	Union River - summer	2000	16
		Union River - summer	2004	42
		Union River - summer	2003	53

Table 2.–Fifty-three SNPs used in the current ADF&G chum salmon baseline, including observed heterozygosity (H_O)**,** expected heterozygosity (H_E), F_{ST} and measures of conformance to Hardy-Weinberg Equilibrium (*P*-HWE). Superscripts preceding SNP names indicate sets which were pooled into a single locus.

Published Name	F_{ST}	H_E	H_O	***P*-HWE**	**Citation**
Oke_PPA2-635	0.120	0.415	0.362	0.984	Elfstrom et al. 2007
Oke_AhR1-278	0.019	0.084	0.083	1.000	Elfstrom et al. 2007
Oke_AhR1-78	0.042	0.490	0.470	0.840	Elfstrom et al. 2007
Oke_arf-319	0.043	0.336	0.322	0.947	Smith et al. 2005a
Oke_U401-220	0.052	0.343	0.327	0.996	Elfstrom et al. 2007
Oke_CKS-389	0.074	0.404	0.370	0.980	Smith et al. 2005b
Oke_copa-211	0.197	0.116	0.094	1.000	Smith et al. 2005a
Oke_ctgf-105	0.045	0.218	0.206	1.000	Elfstrom et al. 2007
Oke_DM20-548	0.068	0.496	0.464	0.108	Smith et al. 2005b
Oke_eif4ebp2-64	0.077	0.156	0.144	1.000	Smith et al. 2005a
Oke_FARSLA-242	0.187	0.138	0.112	1.000	Elfstrom et al. 2007
Oke_GHII-3129	0.104	0.357	0.320	1.000	Elfstrom et al. 2007
Oke_GnRH-527	0.100	0.246	0.227	1.000	Smith et al. 2005b
Oke_GPDH-191	0.052	0.470	0.449	0.607	Smith et al. 2005a
Oke_GPH-78	0.070	0.221	0.205	0.999	Elfstrom et al. 2007
Oke_GPH-105	0.070	0.496	0.458	0.627	Elfstrom et al. 2007
Oke_hnRNPL-239	0.057	0.088	0.082	1.000	Elfstrom et al. 2007
Oke_HP-182	0.055	0.369	0.354	0.950	Elfstrom et al. 2007
Oke_HSP90BA-299	0.033	0.017	0.017	1.000	Elfstrom et al. 2007
Oke_hsc71-199	0.073	0.079	0.072	1.000	Smith et al. 2005a
Oke_il-1racp-67	0.057	0.319	0.297	1.000	Smith et al. 2005a
Oke_IL8r-272	0.063	0.223	0.207	1.000	Smith et al. 2005b
Oke_KPNA2-87	0.136	0.159	0.138	1.000	Elfstrom et al. 2007
Oke_MAPK1-135	0.070	0.170	0.159	1.000	Elfstrom et al. 2007
Oke_MARCKS-362	0.202	0.498	0.401	0.999	Elfstrom et al. 2007
Oke_Moesin-160	0.038	0.105	0.102	1.000	Smith et al. 2005a

-continued-

Table 2. Page 2 of 2.

Published Name	F_{ST}	H_E	H_O	*P*-HWE	Citation
Oke_ras1-249	0.110	0.454	0.407	0.954	Elfstrom et al. 2007
Oke_RFC2-618	0.217	0.365	0.287	1.000	Smith et al. 2005a
Oke_RH1op-245	0.110	0.097	0.083	1.000	Smith et al. 2005a
Oke_serpin-140	0.070	0.499	0.456	0.352	Smith et al. 2005a
Oke_TCP1-78	0.129	0.213	0.182	0.954	Elfstrom et al. 2007
Oke_Tf-278	0.165	0.380	0.315	0.788	Elfstrom et al. 2007
Oke_Tsha1-196	0.067	0.342	0.313	0.949	Smith et al. 2005a
Oke_u1-519	0.125	0.329	0.286	1.000	Smith et al. 2005b
Oke_u202-131	0.082	0.114	0.105	1.000	Smith et al. 2005a
Oke_u212-87	0.106	0.091	0.079	1.000	Smith et al. 2005a
Oke_u216-222	0.040	0.208	0.198	1.000	Smith et al. 2005a
Oke_u217-172	0.049	0.492	0.474	0.998	Smith et al. 2005a
Oke_u200-385	0.101	0.500	0.446	0.990	Smith et al. 2005a
Oke_U302-195	0.112	0.306	0.286	0.495	Elfstrom et al. 2007
Oke_U503-272	0.019	0.109	0.102	0.956	Elfstrom et al. 2007
Oke_U502-241	0.441	0.381	0.216	1.000	Elfstrom et al. 2007
Oke_U504-228	0.098	0.482	0.439	0.547	Elfstrom et al. 2007
Oke_U505-112	0.034	0.438	0.415	0.008	Elfstrom et al. 2007
Oke_U506-110	0.171	0.257	0.203	0.567	Elfstrom et al. 2007
Oke_U507-286	0.059	0.491	0.471	0.402	Elfstrom et al. 2007
Oke_U509-219	0.051	0.496	0.470	0.155	Elfstrom et al. 2007
Oke_U510-204	0.032	0.316	0.305	1.000	Elfstrom et al. 2007
Oke_U511-271	0.064	0.166	0.151	1.000	Elfstrom et al. 2007
Oke_U514-150	0.053	0.214	0.203	1.000	Elfstrom et al. 2007
Oke_U305-130	0.048	0.473	0.444	0.632	Elfstrom et al. 2007
Oke_Cr386[a]	NA	NA	NA	NA	Smith et al. 2005b
Oke_ND3-69[a]	NA	NA	NA	NA	Smith et al. 2005b

[a] These SNPs were combined into a single haplotype.

Table 3.–Mean reporting group allocations of simulated mixtures of chum salmon from the baseline of 52 SNP loci using SPAM. Each set of mixtures (N=400) was created from a single reporting region based on allelic frequencies for that region. The results reported are the mean, standard deviation (SD) and bounds of the middle 90% (CI) of correct allocations from 1,000 bootstrap iterations.

Reporting Group	Mean	SD	90% CI
Japan/Korea	0.998	0.003	(0.992, 1.000)
Russia	0.985	0.008	(0.969, 0.996)
Kotzebue Sound	0.941	0.025	(0.893, 0.978)
Seward Peninsula	0.908	0.034	(0.850, 0.960)
Norton Sound	0.758	0.068	(0.637, 0.860)
Yukon Alaska Summer	0.725	0.082	(0.586, 0.855)
Yukon Alaska Fall/Middle	0.907	0.036	(0.846, 0.961)
Yukon Canada Fall	0.933	0.031	(0.874, 0.980)
Kuskokwim Bay	0.685	0.071	(0.565, 0.802)
Kuskokwim Summer	0.645	0.094	(0.476, 0.785)
Kuskokwim Fall	0.935	0.027	(0.888, 0.975)
Nunivak Island	0.972	0.020	(0.933, 1.000)
Bristol Bay	0.697	0.066	(0.588, 0.804)
North Peninsula	0.941	0.025	(0.897, 0.977)
South Peninsula	0.920	0.031	(0.864, 0.967)
Western Gulf of Alaska	0.947	0.021	(0.908, 0.979)
Eastern Gulf of Alaska	0.988	0.008	(0.973, 0.999)

FIGURES

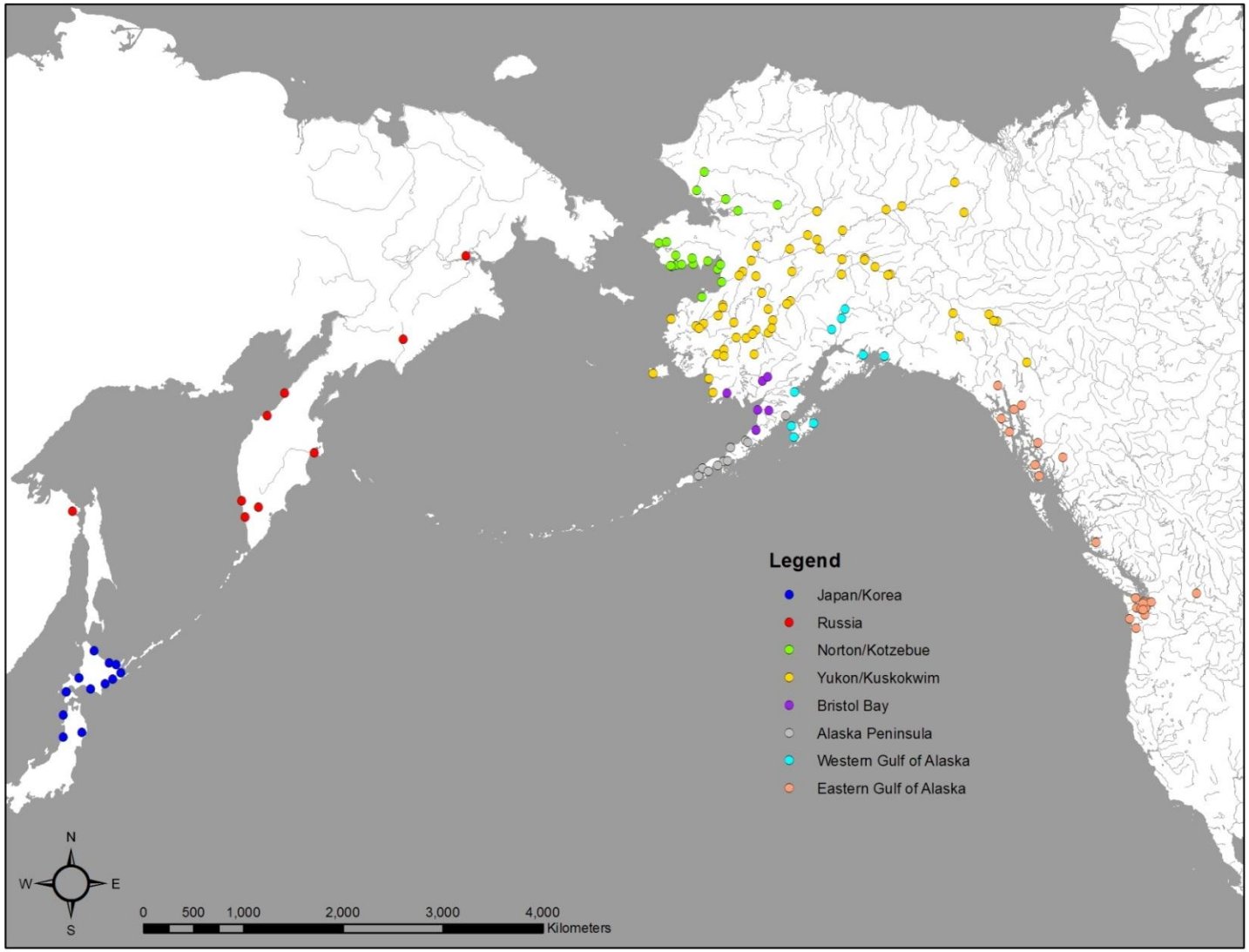

Figure 1.–Map of coast-wide chum salmon sample locations. Colored dots represent each of 8 reporting regions. For clarity, 13 groups in western Alaska are combined into 4 broad-scale groups.

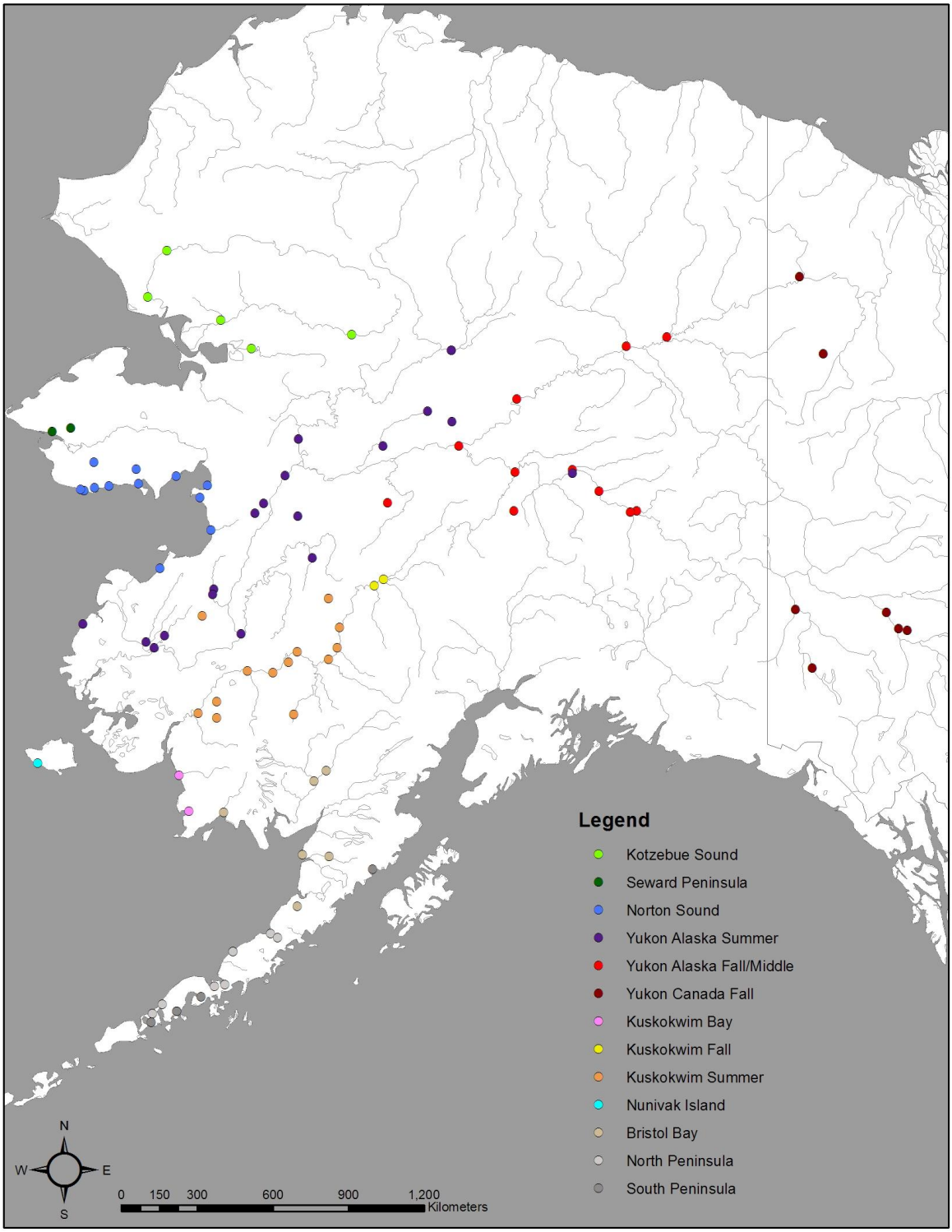

Figure 2.–Map of Western Alaska chum salmon sample locations for which data from 62 SNP loci have been collected and are used in the existing baseline. Colored dots represent each of 13 reporting regions.

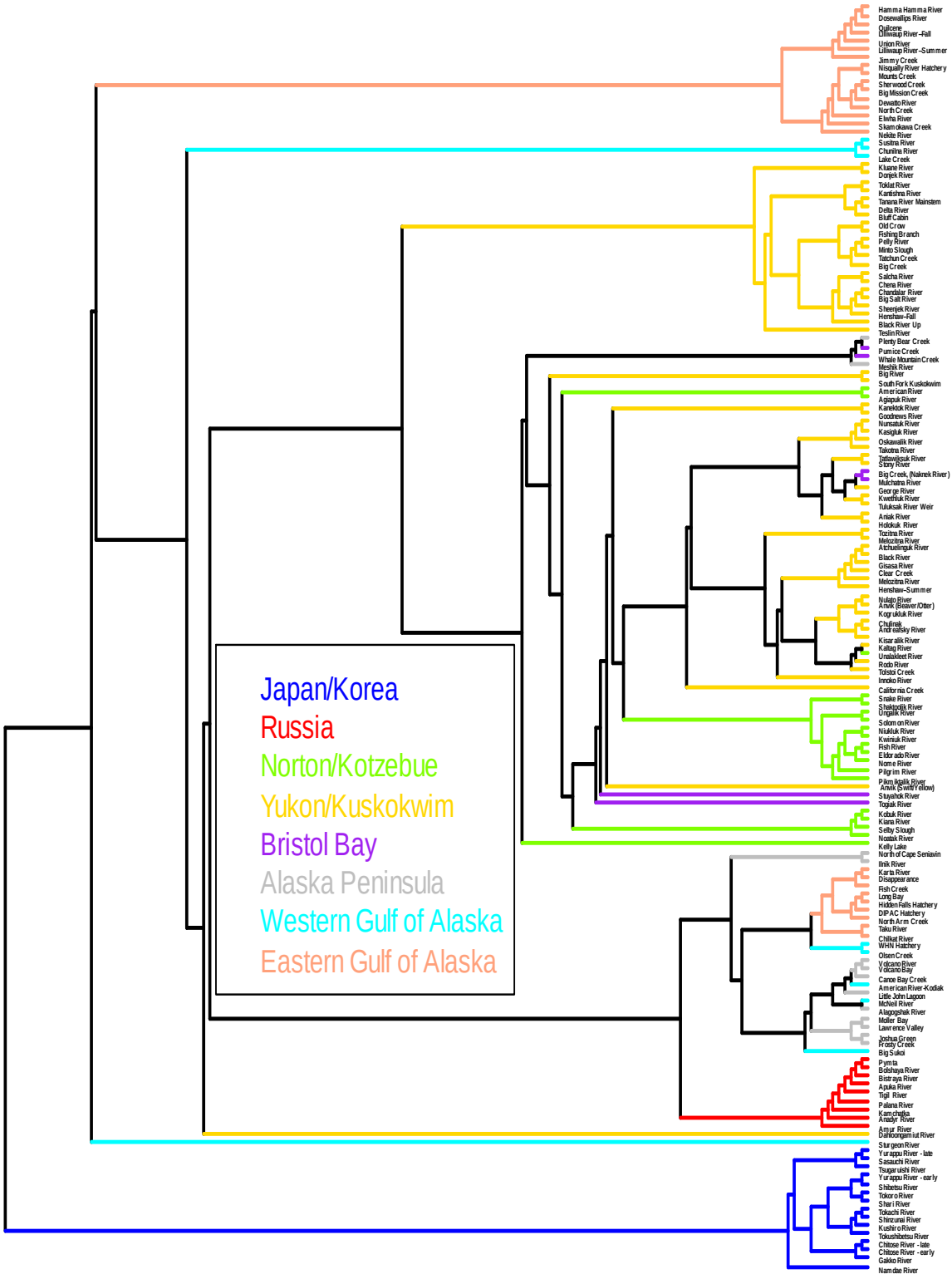

Figure 3.–Unweighted pair-group method (UPGMA) tree of pair-wise F_{ST} among the 153 populations included in the coast-wide 53 SNP baseline.

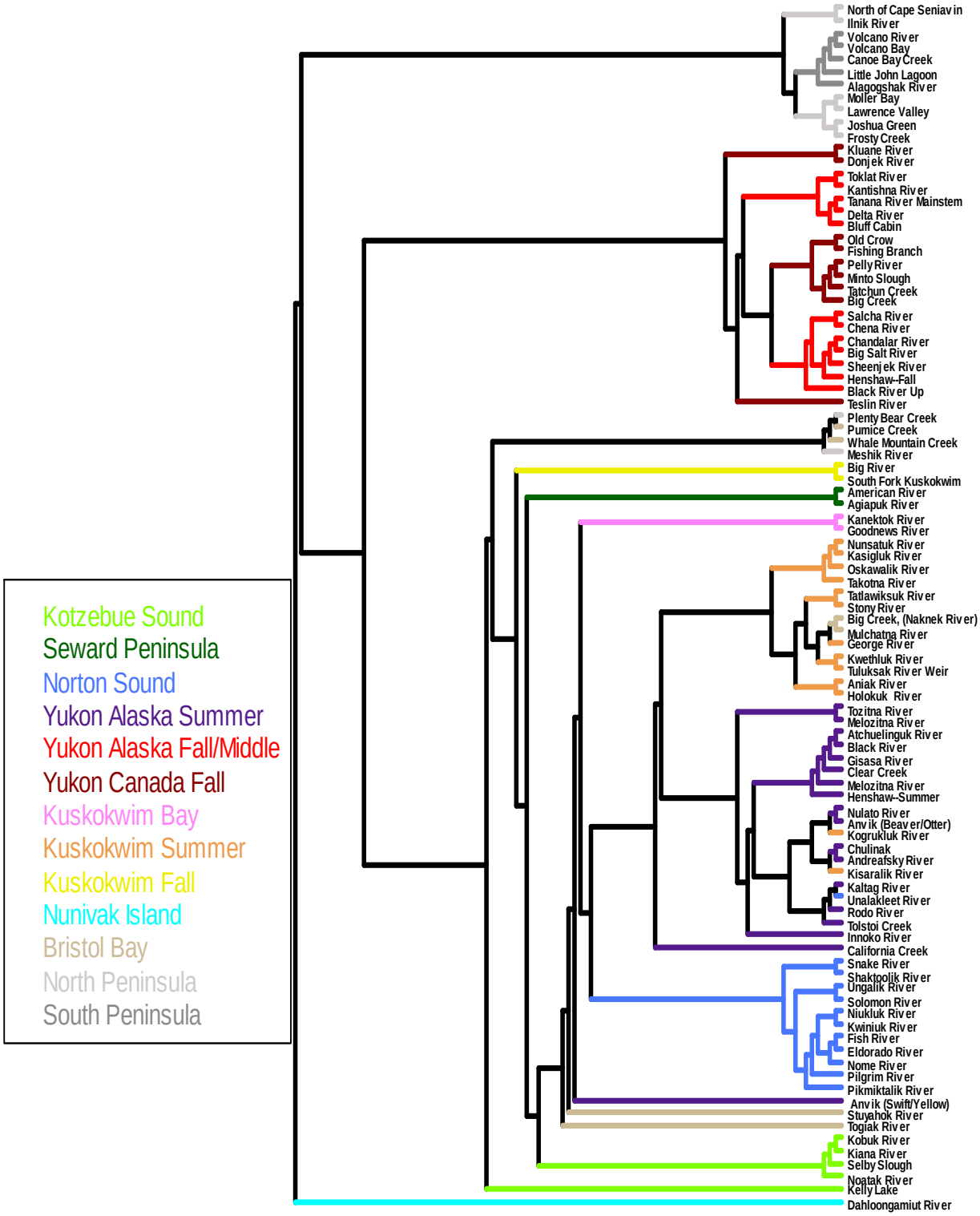

Figure 4.–Unweighted pair-group method (UPGMA) tree of pair-wise F_{ST} among the 95 populations included in the Western Alaska portion of the coastwide 53 SNP baseline.