Fishery Manuscript Series No. 15-01

Genetic Baseline for Upper Cook Inlet Chinook Salmon: 42 SNPs and 7,917 Fish

by

Andrew W. Barclay

and

Christopher Habicht

March 2015

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. 15-01

GENETIC BASELINE FOR UPPER COOK INLET CHINOOK SALMON: 42 SNPS AND 7,917 FISH

by

Andrew W. Barclay and Christopher Habicht
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

March 2015

This report was prepared under award #NA09NMF4380373 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, administered by the Alaska Department of Fish and Game. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce.

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.

Note: Product names used in this publication are included for completeness but do not constitute product endorsement.

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

and

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. 2015. Genetic baseline for Upper Cook Inlet Chinook salmon: 42 SNPs and 7,917 fish. Alaska Department of Fish and Game, Fishery Manuscript Series No. 15-01, Anchorage.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Chinook salmon stocks support important commercial, sport, personal use, subsistence, and educational fisheries in Cook Inlet, Alaska with annual harvests that can exceed 90,000 fish. Many populations contribute to these harvests, which often occur in areas where stocks intermingle, so the exploitation and productivity of individual stocks are not well known. This lack of knowledge compromises both protection of stocks from overharvest and utilization of stocks with fish in excess of spawning needs. Genetic analysis, using single nucleotide polymorphism (SNP) technology, can discriminate among discrete fish stocks in fishery samples when adequate stock structure exists. Here, we update a previously reported baseline of 30 populations using 46 SNP markers with an additional 25 populations and a subset of 42 SNP markers to determine population structure and test potential reporting groups for mixed stock analysis (MSA). After nearly doubling the number of populations in the baseline, southern region stocks continue to have greater genetic diversity than northern region stocks. Visualization of baseline population structure revealed 3 genetically similar groups of populations in the northern region (NorthWestCI, MatSu, and KnikTurnagain) of Cook Inlet and 2 in the southern region (KenaiKasilof and SKenaiPen); baseline performance for MSA was tested for these groups by randomly drawing 10 samples of 200 fish from each group. Test mixture samples from southern region groups had the correct allocations greater than 90% and correct allocations in northern region groups varied by draw and ranged 79.5–97.6%. Future studies employing this baseline for MSA will need to conduct new baseline evaluation tests for population groups chosen to answer study-specific questions.

Key words: Chinook salmon, Cook Inlet, *Oncorhynchus tshawytscha*, single nucleotide polymorphism, SNP, mixed stock analysis, MSA

INTRODUCTION

BACKGROUND

Populations of Chinook salmon (*Oncorhynchus tshawytscha*) support important fisheries in the Upper Cook Inlet management area (UCI; Figure 1). Total harvests of Chinook salmon in all fisheries in UCI averaged approximately 76,000 fish during the years 2000 to 2009 (Begich and Pawluk 2010; Bosch 2010; Oslund and Ivey 2010; Shields 2010; Szarzi et al. 2010). Most harvests occur during the homeward migration from the open ocean or in the lower reaches of river drainages, areas where stocks are mixed. Without stock-specific harvest information, the exploitation and productivity of any single stock cannot be estimated, limiting management for sustained yield by the Alaska Department of Fish and Game (ADF&G) under the policy for the management of sustainable salmon fisheries (Alaska Administrative Code 5 AAC 39.222). Understanding the return-per-spawner relationship (through the development of brood tables) for each stock provides the basis for the escapement goals and subsequent management strategy. Appropriate escapement goals allow for an adequate escapement of fish into natural streams to produce sustained yield without foregoing harvest of the surplus.

Genetic baselines provide the basis for successful mixed stock analysis (MSA) using genetic markers (e.g., Crane et al. 2000; Seeb et al. 2000; Beacham et al. 2009; Habicht et al. 2010). These genetic baselines illuminate the population structure and guides in the delineation of reporting groups (stocks) for MSA. The population structure of Chinook salmon on a broad geographic scale is well understood, but until recently, understanding of the population structure within Cook Inlet has been incomplete. Early studies examining population structure of Cook Inlet Chinook salmon populations were limited to Kenai and Kasilof rivers (Adams et al. 1994; Begich et al. 2010; Rogers Olive et al. 2013) and broad-scale analyses with a few representative populations (Crane et al. 1996; Teel et al. 1999; Templin et al. 2011). The most comprehensive information on population structure of Chinook salmon in Cook Inlet appears in Barclay et al. (2012), in which stock structure was investigated using a more complete set of 30 populations. This study found that there are 2 regional genetic groups among Cook Inlet Chinook salmon

populations, with little divergence in the northern region (west Cook Inlet, Yentna River, Susitna River, Knik Arm, and Turnagain arm populations) and higher divergence in the southern region (Kenai River, Kasilof River, and southern Kenai Peninsula populations). The study also found congruence with findings from previous studies with limited baselines. Here we present a more comprehensive analysis of UCI Chinook salmon population structure using 55 populations, providing better representation of Chinook salmon spawning in western and northern UCI drainages. We also include simulated MSA results testing reporting groups delineated using the population structure for applications within UCI.

DEFINITIONS

Definitions of commonly used genetic terms are provided here to better understand the methods, results, and interpretation of this study.

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

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

F_{ST}. Fixation index is an estimate of 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, sometimes 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 (*HWE*). Genotype frequencies expected from a given set of allele frequencies for a locus. Fit to HWE genotypic proportions assumes random mating, no mutation (the alleles remain unchanged), no migration or emigration (no exchange of alleles between populations), infinitely large population size, and no selective pressure for or against the alleles.

Heterozygosity. The proportion of individuals in a population that have 2 different allele forms (are heterozygous) at a particular marker. Average heterozygosity can be used as measure of variability in a sample.

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

Linked markers. Genetic markers showing linkage disequilibrium, or physical linkage on a chromosome.

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

Population. A locally interbreeding group of spawning individuals that do not interbreed with individuals in other spawning aggregations, and that may be uniquely adapted to a particular spawning habitat. This produces isolation among populations and may lead to the appearance of unique attributes (Ricker 1958) that result in different productivity rates (Pearcy 1992; NRC 1996). This population definition is analogous to *spawning aggregations* described by Baker et al. (1996) and *demes* described by the NRC (1996).

Reporting group. A group of populations in a genetic baseline to which portions of a mixture are allocated during mixed stock analysis.

Single nucleotide polymorphism (*SNP*). DNA nucleotide variation (A, T, C, or G) at a single nucleotide site. SNPs can differ among individuals or within an individual between homologous nucleotide sites on paired chromosomes.

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) that occur within the same geographic area and are managed as a unit (from 5 AAC 39.222(f)).

METHODS

TISSUE SAMPLING

Baseline

Tissue samples suitable for genetic analyses (hereafter *genetic samples*) were collected and were subsequently frozen (heart, muscle, liver, and eye; samples collected prior to 2003) or preserved in 95% ethanol (axillary fin). Frozen tissues were placed 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 genetic samples were collected from spawning aggregates of Chinook salmon by ADF&G personnel using gillnets, beach seines, or hook-and-line gear (Table 1; Figure 1). Target sample size for each baseline aggregate was 95 individuals across all years to achieve acceptable precision to estimate allele frequency (Allendorf and Phelps 1981; Waples 1990a).

Test Mixture

Test mixture genetic samples were collected from adult Chinook salmon captured within the Susitna River during their homeward migration by a Sport fish division drift gillnetting project (Yanusz et al. 2013).

LABORATORY ANALYSIS

Assaying Genotypes

DNA extraction and genotyping generally followed the methods described in detail in Barclay et al. (2012). Briefly, genomic DNA was extracted from tissue samples using a DNeasy® 96 Tissue Kit by QIAGEN® (Valencia, CA). Fluidigm® 192.24 and 96.96 Dynamic Arrays (San Francisco, CA) were used to screen 39 SNP markers; this differs from the methods of Barclay et al. (2012) where only the 96.96 Dynamic Arrays were used. The Dynamic Arrays were read on a Fluidigm® EP1™ System or BioMark™ 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. The plates were scanned on an Applied Biosystems Prism 7900HT Sequence Detection System after amplification and scored using Applied Biosystems' Sequence Detection Software 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 locus for a fish could not be satisfactorily scored.

Quality control (QC) measures were instituted to identify laboratory errors and to determine the reproducibility of genotypes. In this process, 8 of every 96 fish (1 row per 96-well plate) were reanalyzed for all markers by staff not involved with the original analysis. Laboratory errors found during the QC process were corrected, and genotypes were corrected in the database. Inconsistencies not attributable to laboratory error were recorded, but original genotype scores were retained 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 were estimated as half the rate of inconsistencies.

STATISTICAL ANALYSIS

Data Retrieval and Quality Control

We retrieved genotypes from LOKI and imported them into *R*[1] with the *RODBC* package (Ripley 2013). 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. First, we identified SNP markers that were invariant in all individuals or that had very few individuals with the alternate allele in only 1 collection. We excluded these markers from further statistical analyses. Second, we identified individuals that were missing substantial genotypic data because they likely had poor quality DNA. 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. 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 95% of screened loci. The sample with the most missing genotypic data from each duplicate pair was removed from further analyses. If both samples had the same amount of genotypic data, the first sample was removed from further analyses.

Baseline Development

Hardy–Weinberg expectations

For each locus within each 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

[1] *R* Development Core Team. 2014. *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

locus across collections using Fisher's method (Sokal and Rohlf 1995) and removed collections and loci that violated HWE from subsequent analyses after correcting for multiple tests with Bonferroni's method ($\alpha = 0.05$ / number of collections).

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 that might represent the same population. We defined collections as being *geographically proximate* if they were within the same river. 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 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 if they exceeded 50 samples. 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 4.1.4 (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 program *R2jags* 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 et al. 2013).

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 MSA. We used f_{ORCA} as our measure of information, which 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 1995). 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

Temporal variation

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 among-years component of variance and the assessment of its magnitude relative to the among-population component of variance. This analysis was conducted using the software package *GDA* (Lewis and Zaykin 2001).

Hierarchical log-likelihood tests

We examined genetic diversity within Cook Inlet with a 3-level hierarchical log-likelihood ratio (G) analysis. Populations were grouped hierarchically into 8 fine-scale groups based on drainage and geographic features (Table 1; Figure 1):

1) *West* (populations from streams draining into western UCI, south of the Susitna River),
2) *Yentna* (Yentna River populations),
3) *Susitna* (Susitna River populations),
4) *Knik* (populations from streams draining into Knik Arm),
5) *Turnagain* (populations from streams draining into Turnagain Arm),
6) *Kenai* (Kenai River populations),
7) *Kasilof* (Kasilof River populations), and
8) *SKenaiPen* (populations from Kenai Peninsula streams, south of the Kasilof River).

These were further grouped into 2 broad-scale regions:

1) Northern (West, Susitna, Yentna, Knik, and Turnagain groups) and
2) Southern (Kenai, Kasilof, and SKenaiPen groups).

We tested for homogeneity of allele frequencies within groups, among groups within regions, and between regions. To compare levels of heterogeneity between regions and groups, scaled G-statistics (G′) were calculated by dividing G by degrees of freedom.

Visualization of genetic distances

We took 2 approaches to visualizing genetic distances among collections. Both approaches were based on pairwise F_{ST} estimates from the final set of independent markers with the package *hierfstat* (Goudet 2013). The first approach was to construct 1,000 bootstrapped neighbor-joining (NJ) 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). While these trees provided insight into the variability of the genetic structure of these collections, pairwise distances visualized in 3

dimensions were more intuitive. In the second approach, we ran principal coordinates analysis (PCA; Gower 1966) for pairwise F_{ST} and plotted the results of the first 3 principal components.

Baseline Evaluation for Mixed Stock Analysis

We used the results from the visualizations to delineate 5 reporting groups that might perform adequately for MSA (Table 1; Figure 1). We assessed the accuracy and precision for MSA using these reporting groups with proof tests and a test mixture of known-origin fish. Methods for these tests followed those used by Habicht et al. (2012). These reporting groups were made up of one or more of the previously-defined fine scale groups (*italics*) and/or a subset of these groups as follows:

1) NorthWestCI (*West, Yentna*, and western Susitna River populations),
2) MatSu (Eastern Susitna River and Matanuska River populations),
3) KnikTurnagain (*Knik* and *Turnagain*),
4) KenaiKasilof (*Kenai* and *Kasilof*),
5) SKenaiPen (*SKenaiPen*).

Proof tests

In the 100% proof tests, mixtures were created by randomly sampling 100 or 200 fish from the baseline for a single reporting group and then rebuilding the baseline without the sampled fish. These tests provide a measure of the potential accuracy and precision possible for designated reporting groups, as well as a means to understand the direction of bias when estimating stock proportions.

The stock composition of the proof test mixtures was estimated using the software package BAYES (Pella and Masuda 2001). BAYES employs a Bayesian algorithm to estimate the most probable contribution of the baseline populations to explain the combination of genotypes in the mixture sample. We followed a BAYES protocol similar to the protocol reported in Barclay and Habicht (2012). However, instead of running 5 independent Markov Chain Monte Carlo chains, we ran 1 chain with 40,000 iterations and discarded the first 20,000 iterations. The prior parameters for each reporting group were defined to be equal (i.e., a *flat* prior). Within each reporting group, the population prior parameters were divided equally among the populations within that reporting group. Stock proportion estimates and the 90% credibility intervals for each proof test mixture were calculated by taking the mean and 5% and 95% quantiles of the posterior distribution from the single chain output.

Proof tests were repeated 10 times for each reporting group using a different mixture and baseline each time. These tests provided an indication of the power of the baseline for MSA assuming that all populations were represented in the baseline.

Known-origin mixture

Using the set of individuals sampled from Susitna River fish wheels in 2012 (Yanusz et al. 2013), we estimated the stock composition of this mixture using the same reporting groups used in the proof tests (see the Methods section *Tissue Sampling*, *Test Mixture*; Figure 1). This was the most challenging test because fish may have originated from populations not represented in the baseline. This may also be the most realistic test of baseline performance because the

proportion of fish from each population in the mixture was more likely to be in proportion to the relative run strength of each population within the drainage for that year.

RESULTS

TISSUE SAMPLING

Baseline

A total of 11,030 genetic samples were collected from spawning populations of Chinook salmon throughout UCI (Table 1; Figure 1). Prior to 2008, a total of 33 collections were made within UCI, with the majority coming from Kenai Peninsula streams; details for these collections are described in Barclay et al. (2012). Between 2008 and 2011, an additional 79 collections were made focusing on unrepresented and underrepresented areas of UCI. The 172 collections were taken at 79 locations throughout UCI drainages; individuals from 47 of these locations were collected in multiple years. Target sample sizes of 95 fish were met at 42 locations.

Test Mixture

A total of 104 genetic samples were collected from adult Chinook salmon in the lower Susitna River in 2012 (Yanusz et al. 2013; Figure 1).

LABORATORY ANALYSIS

Assaying Genotypes

A total of 8,125 fish collected over spawning areas and 104 fish collected by the lower Susitna River netting project were selected for analysis and assayed for 42 SNP markers (Table 1). Samples not included in the analysis were from locations likely representing populations with insufficient sample size among collections (<50 individuals), locations with a large sample size among collections (>95 individuals), and/or are scheduled to be analyzed at a later date.

Laboratory Failure Rates and Quality Control

For all 125 collections in the baseline, the overall failure rate for genotypes at the 42 SNP markers was 2%. The most recent UCI baseline laboratory project included 845 individuals (~12% of the current baseline). The discrepancy rate for this project was 0.04%; therefore, the overall error rate was 0.02%.

STATISTICAL ANALYSIS

Data Retrieval and Quality Control

For all analyzed collections, 1 SNP was invariant among all individuals, and no SNPs had very few individuals with the alternate allele in only 1 collection (Table 2). The invariant marker was removed from further analyses. Using the criterion for sufficiently complete genotypes, 2.37% of individuals were removed from the baseline collections. Based on the criterion for detecting duplicate individuals, 0.03% of individuals were removed from baseline collections as duplicate individuals. No duplicate individuals were detected in 109 of the 125 baseline collections (87%).

Baseline Development

Hardy–Weinberg expectations within collections

Over all nuclear SNPs and collections, 34 of 5,375 tests deviated significantly from HWE ($P < 0.01$) without adjusting for multiple tests. These were spread over 21 SNPs, and no SNPs were out of HWE in more than 4 of the 125 collections. No collections departed HWE at more than 3 of the 43 SNPs. After adjusting for multiple tests, all collections conformed to HWE.

Pooling collections into populations and HWE within populations

A total of 55 populations were identified after pooling collections taken at 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). Over all variant SNPs and populations, 53 of 2,200 tests did not conform to HWE ($P < 0.01$) without adjusting for multiple tests. These were spread over 32 SNPs, and no SNPs were out of HWE in more than 7 of the 55 populations. After adjusting for multiple tests, 1 SNP (*Ots_il-1racp-166*) did not conform to HWE and was removed from further analyses. No population was out of HWE at more than 4 of 40 SNPs. After adjusting for multiple tests, all populations conformed to HWE.

Linkage disequilibrium

In the tests for linkage disequilibrium, 1 SNP pair, *Ots_FGF6A* and *Ots_FGF6B,* showed significant linkage ($P < 0.05$) in all populations. Correlation coefficients r between the first alphabetical allele in the linked pair of SNPs varied across reporting groups and ranged from 0.82 to 0.97. The 90% critical value of the f_{ORCA} difference distribution Δ_{90} was 0.026, which was greater than Δ for the linked pair. *One_FGF6A* was dropped from further analysis because it had a lower f_{ORCA} value than *One_FGF6B*.

Analysis of Genetic Structure

Temporal variation

Forty-one populations were included in the analysis of temporal variation of allele frequencies. Allele frequencies for all populations appeared to be temporally stable. Within populations, 37 pairs of collections were 1–6 years apart and 4 were 13–17 years apart (Table 1). The 3-level ANOVA indicated that the ratio of variation among temporal collections to the variation among populations was 5.3%.

Hierarchical log-likelihood test

In the analysis of genetic heterogeneity, grouping populations into 8 fine-scale groups and 2 broad-scale regions, significant variation was found within each group, among groups, and between regions (Table 3). Within regions, the Southern region had more genetic heterogeneity ($G' = 14.39$) than the Northern region ($G' = 4.79$). Within the Northern region, *Susitna* ($G' = 5.41$) had greatest genetic heterogeneity among populations, followed by the *Knik* ($G' = 3.58$), *Yentna* ($G' = 2.77$), *Turnagain* ($G' = 2.04$), and *West* ($G' = 1.95$) population groups. Within the Southern region, the *Kasilof* ($G' = 23.23$) group had the greatest genetic heterogeneity followed by the *Kenai* ($G' = 12.40$) and *SKenaiPen* ($G' = 1.59$) groups, which had the lowest genetic heterogeneity of the 8 groups. Greater among-group heterogeneity was found in the Southern region.

Visualization of genetic distances

When an NJ tree was used to visualize genetic relationships, baseline populations formed 3 major clusters (Figure 2). The first 2 clusters, at the bottom of the tree, included *West*, *Yentna*, and *Susitna* group populations. In these clusters, populations generally clustered with other populations from the same reporting group or with geographically proximate populations. With the exception of 1 *West* population (Straight Creek), these 2 clusters include all baseline populations west of the Susitna River mainstem. The third cluster, at the top of the tree, included all populations from the *Knik*, *Turnagain*, *Kenai*, *Kasilof*, and *SKenaiPen* groups and all *Susitna* group populations from tributaries draining into the Susitna River from the east and above the Deshka River. Within this cluster, Knik Arm (excluding Moose Creek) and Turnagain Arm populations formed their own cluster. Moose Creek (Matanuska River) showed more similarity to populations from the *Susitna* group than other Knik Arm populations.

Kenai Peninsula populations (*Kenai*, *Kasilof*, *SKenaiPen* groups) formed 2 genetically distinct clusters with relatively deep genetic structure (long genetic branches); Lower Kenai River tributary, Kasilof River tributary, and *SKenaiPen* populations formed 1 cluster and Upper Kenai River tributary, Kenai River mainstem, and Kasilof River mainstem populations formed another cluster. Kasilof River populations grouped more closely to other Kenai Peninsula populations than to each other; Kasilof River mainstem grouped closely with Kenai River mainstem and Crooked Creek grouped with Slikok Creek and *SKenaiPen* populations.

Several populations appeared to be more genetically distinct (on longer genetic branches): Deshka River; Chulitna River (Middle and East forks); Russian River; Grant Creek; and Kasilof River mainstem. Of the 17 well supported nodes (>50% of bootstrap trees), 9 occurred in the Southern region and 8 occurred in the Northern region. In the Northern region nearly all well supported nodes occurred in the *Susitna* group (7 of 8 significant nodes).

When F_{ST} among baseline populations was analyzed using PCA and the first 3 principal coordinates (PC1, PC2, and PC3; Figures 3 and 4) were plotted, PC1 and PC2 showed clustering of populations into Northern and Southern regions (Figure 3). As with the tree, the PCA also shows the Chulitna (Middle and East forks), Deshka, and Russian rivers as genetically distinct from other populations (Figures 3 and 4). Fine-scale PCA plots of each region showed similar genetic relationships among populations to those in the NJ tree (Figure 4). In the Northern region, *West* and *Yentna* populations, as well as *Susitna* populations west of the Susitna River mainstem, formed a tight cluster. Populations in the *Knik* and *Turnagain* groups clustered closely with the exception of Moose Creek (*Knik* group), which clustered closely with Susitna River populations. In the Southern region, Crooked Creek (*Kasilof* group) clustered closely with Slikok Creek (*Kenai* group). Although the Kasilof River mainstem clustered with Middle and Lower Kenai River mainstem populations with high bootstrap values in the NJ tree (Figure 2), they appeared genetically distinct from one another in the PCA plots (Figures 3 and 4). The long branch length for the Kasilof River mainstem population in the NJ tree is congruent with the relationship in the PCA.

Baseline Evaluation for Mixed Stock Analysis

Proof tests

Correct allocations for all 50 repeated proof tests ranged from 73.1% to 99.1% (Figure 5). The SKenaiPen reporting group had highest correct allocations (>98%) and the least variation

between repeats. KenaiKasilof had the next highest correct allocations with repeats ranging from 90.4% to 97.1%. NorthWestCI, MatSu, and KnikTurnagain reporting groups had the most variable results, with correct allocations ranging from 79.5% to 97.6% for NorthWestCI, from 73.1% to 93.5% for MatSu, and from 80.6% to 97.2% for KnikTurnagain.

Known-origin mixture

In the analysis of the known-origin mixture, the combined allocation of reporting groups containing Susitna River populations (NorthWestCI and MatSu) was greater than 96%. Less than 3% of the escapement mixture allocated to the KnikTurnagain reporting group. The combined allocation of the Kenai Peninsula reporting groups (KenaiKasilof and SKenaiPen) was less than 2%.

DISCUSSION

This report expands and updates a previously reported baseline of Chinook salmon in UCI (Barclay et al. 2012) and provides the first analysis of MSA potential. The expansion includes the addition of new samples that provide better representation of 13 existing populations and representation of 25 new populations. The updates include a re-analysis of population structure with the expanded baseline. The first analyses of MSA performance were based on reporting groups delineated through the population structure analysis and provide insights into the likely MSA potential of alternate reporting groups in UCI.

EXPANDED BASELINE

Adequate representation of populations is a prerequisite for applying genetic data to MSA applications (Utter and Ryman 1993). Adequate representation depends on the population structure, with lower representation needed when the structure is organized by regions (Wood 1989). In previous analyses, the population structure for some regions (e.g., Susitna River) indicated high heterogeneity (Barclay et al. 2012). Better representing these highly heterogeneous regions is critical to determining the potential of the baseline for MSA.

In the Northern region, this study added 23 new populations to the 15 Northern region populations reported in Barclay et al. (2012) study (Table 1). Among the Northern region groups, *Susitna* increased by the most populations (12 populations), followed by *Yentna* (4 populations) and *Turnagain* (3 populations). The *West* and *Knik* groups each increased by 2 populations. The previous study found little segregation between groups in the Northern region.

These additions provide a more complete population structure in the Northern region, especially for the Susitna River. The *Susitna* populations distributed into 3 clusters: 1) a group of highly heterogeneous populations draining into the Chulitna River, 2) more homogeneous populations draining into the Susitna River from the east, and 3) the Deshka River and Sucker Creek that drain into the lower Susitna River from the west (Figure 2). The *Yentna* populations are distributed into 2 clusters but with low heterogeneity within river (Figure 2, Table 3). The *Turnagain* and *Knik* populations cluster on 1 branch with low heterogeneity. Finally *West* populations show the lowest levels of heterogeneity (Table 3) and are distributed with short branches on clusters containing *Yentna* and *Susitna* populations.

In the Southern region, this study added 3 new populations to the 15 reported in Barclay et al. (2012), including 2 *Kenai* populations and 1 *SKenaiPen* population. With the exception of the 3

new populations, we observed no changes to the previously reported population structure within this region.

COMPARISONS TO PREVIOUS FINDINGS OF POPULATION STRUCTURE

Patterns of genetic diversity are similar to those noted in the Barclay et al. (2012) study, where genetic diversity was high in the Southern region ($G' = 14.57$) and low in the Northern region ($G' = 6.37$). However, with the more than doubling of the number of populations in the Northern region, the genetic diversity observed in this region was further reduced ($G' = 4.79$; Table 3). This reduction in genetic diversity could be explained by the addition of many populations that are genetically similar to other populations within regional groups and across regional groups (Figure 2).

The patterns of population structure revealed in this study build upon the patterns observed previously by Barclay et al. (2012) and are congruent with studies with more limited datasets (e.g., Adams et al. 1994; Begich et al. 2010; Rogers Olive et al. 2013; Templin et al. 2011). A comparison of population structure patterns to studies with limited datasets is discussed in Barclay et al. (2012); the same patterns hold true for this study.

Relationships among populations that were not examined in Barclay et al. 2012 provide additional insight into population structure within UCI. This study includes additional population samples from all fine-scale groups except *Kasilof*, with the majority of samples coming from northern Cook Inlet Chinook salmon populations (*West, Yentna, Susitna, Knik, and Turnagain*; Table 1).

The most notable finding of this study is the geographic pattern of population structure in the Northern region (Figures 1 and 2). We found low levels of heterogeneity among the *West* and *Yentna* populations and some affinity between these populations and populations draining into the lower and upper Susitna River from the east. This pattern led us to group all these populations together into the NorthWestCI reporting group for MSA testing. The affinity between Moose Creek, Matanuska River, and tributaries draining into the Susitna River from the east led us to group these populations into the MatSu reporting group. Finally, the aggregation of all the remaining *Knik* and *Turnagain* populations into a single cluster led us to group these into the KnikTurnagain reporting group.

The lack of genetic heterogeneity among collections within *West* and *Yentna* were surprising given the heterogeneity of the habitats. Some of these populations travel large distances from the ocean to their spawning grounds (Peter's Creek, >120 km) relative to others (Straight Creek, <30 km), some migrate through cold and silty rivers fed by glaciers (Coal Creek), while others migrate through warm and tannic waters fed by rain and ground water (Chuitna River).

POTENTIAL FOR MSA

Delineating reporting groups for MSA is dictated by the issue to be resolved, the expected composition of the mixture, and the genetic structure of the underlying populations (Habicht et al. 2012). In this report we only incorporated the population structure in delineating reporting groups that might perform well in MSA applications. These results can be used in the future as stakeholders bring forth issues to be solved. These proof tests and the underlying population structure identified in this report can be used to provide insights into what alternative reporting groups might perform well to answer stakeholder questions. Alternate reporting groups will need

to be tested on a case by case basis, depending on study objectives and the potential composition of the mixed stock sample being analyzed.

The proof tests using reporting groups delineated using population structure show promise for use of MSA for Chinook salmon to resolve issues in Cook Inlet (Figure 5). The consistency in performance of the proof tests across replicates is likely due to the homogeneity among populations within the reporting group, as each replicate consists of a random set of individuals within the reporting group. For example, the SKenaiPen reporting group allocated above 98% in every replicate (Figure 5), and all the populations within this reporting group were clustered together above a significant node (Figure 2). NorthWestCI reporting group, on the other hand, showed high variation among replicates (0.79–0.98; Figure 5) and the underlying populations were distributed across multiple clusters (Figure 2).

The proof tests used here might be optimistic because mixture samples constructed for these tests were made up of populations from single reporting groups. Proof tests performed using these 100% mixtures often produce much more optimistic results due to the way the Bayesian algorithm is informed by the composition of the mixture. For example, if the majority of fish in a sample come from a single reporting group, the likelihood of BAYES assigning a fish to that reporting group increases. Once stakeholder issues are identified, proof tests can be done with mixture samples composed of samples from multiple reporting groups in proportion to the expected composition of mixed stock sample from a given fishery and time.

There may be other reporting groups that will perform well, especially for questions where the baseline and/or the composition can be restricted. For example, a reporting group consisting of a single or combination of populations from the Chulitna River might perform well on stock mixtures of migrating fish collected in the lower Susitna River. Alternatively, the Deshka River might perform well as a reporting group for a similar mixture. Within the Kenai River, reporting groups consisting of single or combinations of populations from tributary spawners might perform well on stock mixtures of migrating fish collected in the lower Kenai River or in fisheries in salt water near the mouth of the Kenai River.

On the other extreme, this baseline may not be appropriate for fishery mixtures captured in Lower Cook Inlet. Lower Cook Inlet fishery mixtures may include stocks from Lower Cook Inlet and/or from outside Cook Inlet (Welch et al. 2014). Therefore, fisheries outside UCI should include Chinook salmon stocks from outside of UCI. A rangewide baseline for Chinook salmon is currently available (Templin et al. 2011) and could be combined with this baseline to analyze fish captured in Lower Cook Inlet.

INFLUENCE OF HATCHERY STOCKS ON MSA

Within Cook Inlet there are 4 hatchery populations that are permitted to be released at locations other than their natal stream: Ninilchik River, Crooked Creek, Ship Creek, and Deception Creek. All of these populations are represented in this baseline. However, nonnatal stream released (NNSR) hatchery fish that are released into areas outside their reporting groups could complicate interpretation of MSA results. Fortunately, there are a few methods that could be implemented to partition these fish appropriately.

One method would use coded wire tag (CWT) or otolith data. Some hatchery Chinook salmon released into Cook Inlet are tagged with CWTs, and all fish are otolith marked. If tagging methods were implemented using codes to identify release site of fish, then tagged hatchery fish

sampled for MSA could be used to determine the proportion NNSR fish in a mixed-stock sample. With this information, the MSA estimate for a reporting group containing hatchery populations could be adjusted to account for hatchery fish released at other locations. However, currently neither the CWT nor otolith marking resolve stocking location within Cook Inlet.

A second method would use genetic data to determine which fish in a mixture are of NNSR origin. This method, known as parentage-based tagging (Anderson and Garza 2005), could use the same genetic data already collected on fish sampled for MSA. This method requires that all parents from each brood year are sampled and genotyped for the same panel of markers included in the MSA genetic baseline. In addition, stocking programs would need to track contribution families. Once a parental genotype database has been developed their progeny are effectively *tagged* and they can then be assigned back to their parents. With this method all genetically tagged NNSR hatchery fish in a mixed-stock sample could identified. Continuous sampling and genotyping of brood stocks each year would be required in order to have a complete parental genotype database for all potential brood years in a mixed-stock sample. This sampling effort has begun and the department has genetic samples from the majority of broodstock parents starting in 2010.

ACKNOWLEDGEMENTS

This study required the efforts of a large number of dedicated people. Most importantly, we would like to acknowledge the work of the people in the ADF&G Gene Conservation Laboratory, including: Serena Rogers Olive, Heather Liller, Heather Hoyt, Paul Kuriscak, Christy Cupp, and Zac Pechacek, who oversaw DNA extractions, screened SNPs, and ensured quality control of laboratory data; Judy Berger for archiving samples; Eric Lardizabal for database support; and Matt Bowes, Jason Fox, Nathan Shoutis, Chase Jalbert, Cassey McConnell, and Jed Konsor for spearheading the 2011–2013 field collections. We would like to thank Don Reeves, Kiche and Kaasan Braendel, Darly Lescanec, Leslie and Gabrielle Vail, Chase Jalbert, Clint McBride, Nick Cassara, Matt Warnke, and Ingrid Jensen for their sampling efforts during 2008–2010; Carol Kerkvliet and her crews for sampling southern Kenai Peninsula streams; and Robert Begich, Tim McKinley, Jeff Perschbacher, Adam Reimer, Jerry Strait, Jenny Cope, Tony Eskelin, Jason Pawluk, Rob Massengill, Christy Cupp, Tom Johnson, Jeremiah Batsin, Tom Rhyner, Patti Berkahn, Tye Wyatt, Traye Turner, and Julianne Pettey for collecting baseline samples in the Kenai and Kasilof rivers. We also would like to thank Tim McKinley and Richard Yanusz for reviewing this report and the people involved in the collection of samples analyzed in this study, collected over an 18-year period prior to the inception of this project. Field collections and laboratory and statistical analyses were funded by State of Alaska and Alaska Sustainable Salmon Fund project numbers 44517 “West Cook Inlet Chinook Baseline” and 45864 “Northern Cook Inlet Chinook GSI.” Field collections in 2013 were funded by the Alaska Energy Authority for the Susitna–Watana Hydroelectric Project. This report was prepared under award #NA09NMF4380373 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, administered by the Alaska Department of Fish and Game.

REFERENCES CITED

Adams, N. S., W. J. Spearman, C. V. Burger, K. P. Currens, C. B. Schreck, and H. W. Li. 1994. Variation in mitochondrial DNA and allozymes discriminates early and late forms of Chinook salmon (Oncorhynchus tshawytscha) in the Kenai and Kasilof Rivers, Alaska. Canadian Journal of Fisheries and Aquatic Sciences 51 (S1): 172–181.

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.

Anderson, E. C., and J. C. Garza. 2005. A description of full genotyping. Report submitted to the Pacific Salmon Commission, Vancouver, British Columbia. http://swfsc.noaa.gov/publications/FED/00675.pdf

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 escapements in southern Alaska. Fisheries 21(10): 6–18.

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.

Barclay, A. W., C. Habicht, R. A. Merizon, and R. J. Yanusz. 2012. Genetic baseline for Upper Cook Inlet Chinook salmon: 46 SNPs and 5,279 fish. Alaska Department of Fish and Game, Fishery Manuscript Series No. 12-02, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/FMS12-02.pdf

Beacham, T. D., J. R. Candy, C. Wallace, S. Urawa, S. Sato, N. V. Varnavskaya, K. D. Le, and M. Wetklo. 2009. Microsatellite stock identification of chum salmon on a Pacific Rim basis. North American Journal of Fisheries Management 29(6):1757–1776.

Begich, R. N., and J. A. Pawluk. 2010. 2008-2010 Recreational fisheries overview and historical information for North Kenai Peninsula: fisheries under consideration by the Alaska Board of Fisheries, February 2011. Alaska Department of Fish and Game, Fishery Management Report No. 10-51, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/FMR10-51.pdf

Begich, R. N., W. D. Templin, A. W. Barclay, and L. W. Seeb. 2010. Development of microsatellite genetic markers for Kenai River Chinook salmon. Alaska Department of Fish and Game, Fishery Data Series No. 10-38, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/FDS10-38.pdf

Bosch, D. 2010. Area management report for the recreational fisheries of Anchorage, 2009 and 2010. Alaska Department of Fish and Game, Fishery Management Report No. 10-53, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/Fmr10-53.pdf

Crane, P. A., W. D. Templin, D. M. Eggers, and L. W. Seeb. 2000. Genetic stock identification of Southeast Alaska chinook salmon fishery catches; a report of the Alaska Department of Fish and Game pursuant to National Oceanic and Atmospheric Administration Award No. NA87FPO408. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J00-01, Anchorage.

Crane, P. A., W. D. Templin, and L. W. Seeb. 1996. Genetic stock identification of Alaska Chinook salmon: a report of the Alaska Department of Fish and Game pursuant to National Oceanic and Atmospheric Administration Award No. NA46FD0356. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J99-17, Juneau.

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.adfg.alaska.gov/FedAidPDFs/FMS09-06.pdf

Goudet, J. 2013. hierfstat: Estimation and tests of hierarchical F-statistics. R package version 0.04-10. http://www.r-project.org, http://www.unil.ch/popgen/softwares/hierfstat.htm.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325–338.

REFERENCES CITED (Continued)

Habicht, C., L. W. Seeb, K. W. Myers, E. V. Farley, and J. E. Seeb. 2010. Summer–fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. Transactions of the American Fisheries Society 139(4): 1171–1191.

Habicht, C., W. D. Templin, N. A. DeCovich, and J. R. Jasper. 2012. Western Alaska Salmon Stock Identification Program Technical Document 15: Chum salmon reporting group evaluations using simulated fishery mixtures. Alaska Department of Fish and Game, Division of Commercial Fisheries Regional Information Report 5J12-22, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/RIR.5J.2012.22

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

Lewis, P. O., and D. Zaykin. 2001. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0. http://lewis.eeb.uconn.edu/lewishome/software.html (Accessed March 10, 2009; site currently discontinued).

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

Oslund, S., and S. Ivey. 2010. Recreational fisheries of Northern Cook Inlet, 2009-2010: Report to the Alaska Board of Fisheries, February 2011. Alaska Department of Fish and Game, Fishery Management Report No. 10-50, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/FMR10-50.pdf

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2): 289–290.

Pearcy, W. G. 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.

Ricker, W. E. 1958. Maximum sustained yields from fluctuating environments and mixed stocks. Journal of the Fisheries Research Board of Canada 15(5): 991–1006.

Ripley, B. 2013. RODBC: ODBC Database Access. R package version 1.3-10. http://CRAN.R-project.org/package=RODBC.

Rogers Olive, S. D., A. W. Barclay, T. R. McKinley, and W. D. Templin. 2013. Genetic baseline of Kenai River Chinook salmon for mixed stock analyses, 2013. Alaska Department of Fish and Game, Fishery Manuscript Series No. 13-12. Anchorage.

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., 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.

Shields, P. 2010. Upper Cook Inlet commercial fisheries annual management report, 2010. Alaska Department of Fish and Game, Fishery Management Report No. 10-54, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/FMR10-54.pdf

Smith, C. T., A. Antonovich, W. D. Templin, C. M. Elfstrom, S. R. Narum, and L. W. Seeb. 2007. Impacts of marker class bias relative to locus-specific variability of 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.

REFERENCES CITED (Continued)

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 U.S. 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.

Szarzi, N. J., C. M. Kerkvliet, B. J. Failor, and M. D. Booz. 2010. Recreational fisheries in the Lower Cook Inlet Management Area, 2008-2010, with updates for 2007. Alaska Department of Fish and Game, Fishery Management Report No. 10-38, Anchorage. http://www.adfg.alaska.gov/FedAidpdfs/Fmr10-38.pdf

Teel, D. J., P. A. Crane, C. M. Guthrie III, A. R. Marshall, D. M. Van Doornik, W. D. Templin, N. V. Varnavskaya, and L. W. Seeb. 1999. Comprehensive allozyme database discriminates chinook salmon around the Pacific Rim. (NPAFC document 440). Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, Alaska USA 99518.

Templin, W. D., J. E. Seeb, J. R. Jasper, A. W. Barclay, and L. W. Seeb. 2011. Genetic differentiation of Alaska Chinook salmon: the missing link for migratory studies. Molecular Ecology Resources 11(S1): 226–246.

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

Utter, F., and N. Ryman. 1993. Genetic markers and mixed stock fisheries. Fisheries, 18(8): 11–21.

Warnes, G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeler, M. Schwartz, and B. Venables. 2013. Various R programming tools for plotting data. R package version 2.12.1. http://CRAN.R-project.org/package=gplots

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(5): 968–976.

Weir, B. 1996. Genetic data analysis. 2nd edition. Sinauer Associates, Inc, Sunderland, MA.

Welch, D. W., A. D. Porter, and P. Winchell. 2014. Migration behavior of maturing sockeye (Oncorhynchus nerka) and Chinook salmon (O. tshawytscha) in Cook Inlet, Alaska, and implications for management. Animal Biotelemetry 2(1):35. doi:10.1186/s40317-014-0018-3

Wood, C. C. 1989. Utility of similarity dendrograms in stock composition analysis. Canadian Journal of Fisheries and Aquatic Sciences. 46: 2121-2128.

Yanusz, R. J., P. Cleary, S. Ivey, J. W. Erickson, D. J. Reed, R. A. Neustel, and J. Bullock. 2013. Distribution of spawning Susitna River Chinook *Oncorhynchus tshawytscha* and pink salmon *O. gorbuscha*, 2012 Susitna-Watana Hydroelectric Project (FERC No.14241). Prepared by Alaska Department of Fish and Game Division of Sport Fish for Alaska Energy Authority.

TABLES AND FIGURES

Table 1.–Tissue collections of Chinook salmon throughout Upper Cook Inlet, including the year collected, number of samples collected (N), the number of individuals analyzed from each collection included in the baseline (N_a), and source of the collection. Map numbers correspond to sampling sites on Figure 1; unique population numbers represent all the analyzed collections that contribute to a single population. Tests of homogeneity were performed on groups of populations (Group) and proof tests for MSA were performed on a fewer set of groups of populations (Reporting Group).

Map No.	Pop No.	Group[a]	Reporting Group[b]	Location	Year Collected	N	N_a[c]	Source[d]
1	---	1	1	Crescent River	2010	3	---	A
2	1	1	1	Straight Creek	2010	105	95	B
3	---	1	1	Nikolai Creek	2012	33	---	B
3	---	1	1		2013	47	---	C
4	2	1	1	Chuitna River	2008	20	20	A
4	2	1	1		2009	122	122	A
5	3	1	1	Coal Creek	2009	42	42	A
5	3	1	1		2010	35	35	A
5	3	1	1		2011	43	43	A
6	4	1	1	Theodore River	2010	34	34	A
6	4	1	1		2011	55	55	B
6	4	1	1		2012	104	104	D
6	---	1	1		2013	47	---	D
7	5	1	1	Lewis River	2011	47	47	A
7	5	1	1		2012	42	42	D
8	---	2	1	Clearwater Creek	2012	26	---	B
9	6	2	1	Red Creek	2012	29	29	B
9	6	2	1		2013	82	82	E
10	---	2	1	Happy River	2012	18	---	B
11	---	2	1	Red Salmon Creek	2012	12	---	B
12	7	2	1	Hayes River	2012	5	5	B
12	7	2	1		2013	45	45	E
13	8	2	1	Canyon Creek	2012	31	31	B
13	8	2	1		2013	61	61	E
14	9	2	1	Talachulitna River	1995	58	58	C
14	9	2	1		2008	74	74	A
14	9	2	1		2010	48	48	A
15	---	2	1	Lake Creek	2008	1	---	A
16	10	2	1	Sunflower Creek	2009	53	53	A
16	10	2	1		2011	74	74	A
17	11	2	1	Peters Creek	2009	27	27	A
17	11	2	1		2010	6	6	A

-continued-

Table 1.–Page 2 of 6.

Map No.	Pop No.	Group[a]	Reporting Group[b]	Location	Year Collected	N	N_a[c]	Source[d]
18	---	3	2	Kosina Creek	2012	10	---	B
18	---	3	2		2013	3	---	E
19	12	3	2	Portage Creek	2009	15	15	A
19	12	3	2		2010	10	10	A
19	12	3	2		2011	116	116	A
19	12	3	2		2013	25	25	E
20	13	3	2	Indian River	2013	81	81	E
21	14	3	1	Middle Fork Chulitna River	2009	72	72	A
21	14	3	1		2010	104	97	A
21	---	3	1		2013	61	---	E
22	15	3	1	East Fork Chulitna River	2009	5	5	C
22	15	3	1		2010	2	2	C
22	15	3	1		2011	6	6	C
22	15	3	1		2013	64	64	E
23	---	3	1	Honolulu Creek	2013	31	---	E
24	---	3	1	Pass Creek	2013	33	---	E
25	16	3	1	Byers Creek	2013	55	55	E
26	17	3	1	Spink Creek	2013	56	56	E
27	18	3	1	Troublesome Creek	2013	71	71	E
28	19	3	1	Bunco Creek	2013	103	103	E
29	---	3	2	Upper Talkeetna no name A	2013	25	---	E
30	20	3	2	Upper Talkeetna no name B	2013	71	71	E
31	21	3	2	Stephan Lake weir	2008	19	19	A
32	21	3	2	Prairie Creek	1995	52	52	C
32	21	3	2		2008	98	98	A
32	---	3	2		2013	33	---	E
33	22	3	2	East Fork Iron Creek	2013	57	57	E
34	23	3	2	Disappointment Creek	2013	64	64	E
35	24	3	2	Chunilna Creek	2009	50	50	A
35	24	3	2		2012	79	79	B
35	---	3	2		2013	5	---	E
36	---	3	2	Sheep Creek	2013	24	---	E

-continued-

Table 1. –Page 3 of 6.

Map No.	Pop No.	Group[a]	Reporting Group[b]	Location	Year Collected	N	N_a[c]	Source[d]
37	---	3	2	North Fork Kashwitna River	2013	12	---	E
38	25	3	2	Montana Creek	2008	33	33	A
38	25	3	2		2009	155	155	A
38	25	3	2		2010	30	30	A
38	---	3	2		2013	213	---	D
39	26	3	2	Little Willow Creek	2013	55	55	E
40	27	3	2	Willow Creek	2005	74	74	C
41	27	3	2	Deception Creek	2009	122	100	C
41	---	3	2		2013	245	---	C
42	28	3	1	Moose Creek - Deshka River	1995	51	51	C
42	28	3	1		2012	52	52	B
43	28	3	1	Deshka River weir	2005	200	200	D
44	29	3	1	Sucker Creek	2011	91	91	A
44	29	3	1		2012	53	53	B
45	---	3	1	Wolverine Creek	2011	1	---	A
46	30	4	3	Little Susitna River	2009	3	3	A
46	30	4	3		2010	122	122	A
47	---	4	3	Kings River	2013	4	---	E
48	---	4	3	Granite Creek	2013	30	---	E
49	31	4	2	Moose Creek - Matanuska River	1995	20	20	C
49	31	4	2		2008	33	33	F
49	31	4	2		2009	22	22	F
49	31	4	2		2012	80	80	B
50	---	4	3	South Fork Eagle River	2009	1	---	C
50	32	4	3		2011	4	4	B
50	32	4	3		2012	68	68	B
51	32	4	3	Meadow Creek	2009	6	6	C
52	33	4	3	Ship Creek	2009	311	280	C
52	---	4	3		2013	52	---	C
53	34	5	3	Campbell Creek	2010	3	3	B
53	34	5	3		2011	33	33	A
53	34	5	3		2012	75	75	B

-continued-

Table 1. –Page 4 of 6.

Map No.	Pop No.	Group[a]	Reporting Group[b]	Location	Year Collected	N	N_a[c]	Source[d]
54	---	5	3	Rabbit Creek	2011	8	---	A
55	---	5	3	Bird Creek	2009	2	---	C
55	---	5	3		2011	35	---	A
55	---	5	3		2012	5	---	B
56	---	5	3	Carmen River	2003	5	---	C
56	35	5	3		2011	19	19	A
56	35	5	3		2012	31	31	B
56	---	5	3		2013	23	---	C
57	---	5	3	Granite Creek	2011	1	---	A
58	---	5	3	Canyon Creek	2013	22	---	C
59	36	5	3	Resurrection Creek	2010	24	24	B
59	36	5	3		2011	61	61	A
59	36	5	3		2012	13	13	B
60	37	5	3	Chickaloon River	2008	2	2	C
60	---	5	3		2009	1	---	C
60	37	5	3		2010	65	65	B
60	37	5	3		2011	63	63	B
61	38	6	4	Grant Creek	2011	23	23	B
61	38	6	4		2012	36	32	B
61	---	6	4		2013	33	---	G
62	39	6	4	Quartz Creek	2006	35	34	H
62	39	6	4		2008	34	34	H
62	39	6	4		2009	41	41	H
62	39	6	4		2010	4	4	H
62	39	6	4		2011	13	13	H
63	39	6	4	Dave's Creek	2007	8	8	H
63	39	6	4		2008	5	5	H
64	40	6	4	Crescent Creek	2006	165	165	H
65	41	6	4	Juneau Creek	2005	32	32	H
65	41	6	4		2006	91	91	H
65	41	6	4		2007	24	24	H
66	42	6	4	Russian River	2005	24	24	H

-continued-

Table 1. –Page 5 of 6.

Map No.	Pop No.	Group[a]	Reporting Group[b]	Location	Year Collected	N	N_a[c]	Source[d]
66	42	6	4	Russian River	2006	16	16	H
66	42	6	4		2007	84	83	H
66	42	6	4		2008	91	91	H
67	43	6	4	Upper Kenai River mainstem	2009	200	200	H
68	44	6	4	Benjamin Creek	2005	56	56	H
68	44	6	4		2006	150	150	H
69	45	6	4	Killey River	2005	68	68	H
69	45	6	4		2006	190	190	H
70	46	6	4	Funny River	2005	37	37	H
70	46	6	4		2006	183	183	H
71	47	6	4	Middle Kenai River mainstem	2003	80	80	H
71	47	6	4		2004	39	39	H
71	47	6	4		2006	183	183	H
72	48	6	4	Lower Kenai River mainstem	2010	37	37	H
72	48	6	4		2011	90	90	H
73	49	6	4	Slikok Creek	2004	48	48	H
73	49	6	4		2005	100	95	H
73	49	6	4		2008	58	57	H
74	50	7	4	Middle Kasilof River mainstem	2005	273	273	H
75	50	7	4	Lower Kasilof River mainstem	2005	144	49	H
76	---	7	4	Crooked Creek	1992	95	---	H
76	51	7	4		2005	212	212	H
76	---	7	4		2009	332	---	H
76	51	7	4		2011	374	96	H
76	---	7	4		2012	58	---	H
76	---	7	4		2013	470	---	H
77	52	8	5	Ninilchik River weir	2006	190	162	I
77	---	8	5		2009	93	---	I
77	52	8	5		2010	50	50	I
77	---	8	5		2011	49	---	I
77	---	8	5		2012	34	---	I
77	---	8	5		2013	22	---	I

-continued-

Table 1.–Page 6 of 6.

Map No.	Pop No.	Group[a]	Reporting Group[b]	Location	Year Collected	N	N_a[c]	Source[d]
78	53	8	5	Deep Creek	2009	100	100	I
78	53	8	5		2010	99	99	I
78	---	8	5		2011	50	---	I
79	54	8	5	Stariski Creek	2011	57	56	I
79	54	8	5		2012	50	50	I
80	55	8	5	Anchor River weir	2006	200	200	I
80	---	8	5		2009	10	---	I
80	55	8	5		2010	50	50	I
80	---	8	5		2011	50	---	I
80	---	8	5		2012	50	---	I

[a] Group: 1) *West*; 2) *Yentna*; 3) *Susitna*; 4) *Knik*; 5) *Turnagain*; 6) *Kenai*; 7) *Kasilof*; *SKenaiPen*.

[b] Reporting Group: 1) NorthwestCI; 2) MatSu; 3) KnikTurnagain; 4) KenaiKasilof; 5) SKenaiPen.

[c] --- indicates no samples were included in the baseline analysis.

[d] Collection sources: A) AKSSF project 45864; B) AKSSF project 44517; C) Miscellaneous projects; D) Sport Fish Division weir; E) Susitna-Watana Hydroelectric Project; F) Chickaloon Village Environmental Stewardship Department; G) Grant Creek Hydroelectric Project; H) Sport Fish Division – Soldotna; I) Sport Fish Division – Homer.

Table 2.–Source, observed heterozygosity (H_o), F_{IS}, and F_{ST} for 42 single nucleotide polymorphisms (SNPs) used to analyze the population genetic structure of Upper Cook Inlet Chinook salmon. These summary statistics are based upon the 55 populations within Upper Cook Inlet detailed in Table 1.

Assay Name	Source[a]	H_o	F_{IS}	F_{ST}
Ots_AsnRS-60	B	0.371	-0.007	0.038
Ots_E2-275	B	0.305	-0.002	0.022
Ots_ETIF1A	C	0.483	0.006	0.021
Ots_FARSLA-220	D	0.350	0.005	0.022
*Ots*_FGF6A[b]	A	0.421	0.003	0.032
Ots_FGF6B	A	0.393	0.004	0.023
Ots_GH2	E	0.304	-0.013	0.023
Ots_GPDH-338	B	0.079	-0.017	0.025
Ots_GPH-318	D	0.124	0.003	0.033
Ots_GST-207	D	0.058	0.014	0.034
Ots_GTH2B-550	A	0.460	-0.002	0.015
Ots_hnRNPL-533	D	0.322	-0.008	0.025
Ots_HSP90B-100	D	0.217	0.018	0.025
Ots_HSP90B-385	D	0.012	-0.007	0.055
Ots_IGF-I.1-76	B	0.458	-0.006	0.095
Ots_Ikaros-250	B	0.115	0.007	0.026
Ots_il-1racp-166[c]	B	0.478	-0.070	0.037
Ots_ins-115	B	0.038	-0.015	0.011
Ots_LEI-292	D	0.033	0.035	0.042
Ots_LWSop-638	B	0.056	0.002	0.018
Ots_MHC1	E	0.429	0.011	0.026
Ots_MHC2	E	0.026	-0.008	0.017
Ots_NOD1	A	0.454	0.011	0.067
Ots_P450	E	0.286	-0.005	0.021
Ots_P53	E	0.444	-0.001	0.018
Ots_PGK-54	A	0.030	0.011	0.015
Ots_Prl2	E	0.473	0.013	0.033
Ots_PSMB1-197[d]	D	0.000	-0.002	0.002
Ots_RAG3	A	0.290	0.001	0.040

-continued-

Table 2.–Page 2 of 2.

Assay Name	Source [a]	H_o	F_{IS}	F_{ST}
Ots_S7-1	A	0.177	0.016	0.022
Ots_SClkF2R2-135	B	0.374	0.001	0.078
Ots_SERPC1-209	D	0.185	0.071	0.043
Ots_SL	E	0.477	-0.026	0.065
Ots_SWS1op-182	B	0.455	-0.021	0.010
Ots_TAPBP	C	0.322	-0.006	0.030
Ots_Tnsf	E	0.201	0.001	0.021
Ots_u202-161	B	0.080	-0.005	0.029
Ots_u211-85	B	0.243	-0.006	0.029
Ots_U212-158	B	0.038	-0.004	0.030
Ots_u6-75	B	0.094	-0.018	0.040
Ots_unk526	A	0.233	0.016	0.056
Ots_Zp3b-215	B	0.067	0.020	0.056
Average/Overall		0.245	0.001	0.036

[a] Marker sources: A) Northwest Fisheries Science Center-NOAA (Unpublished); B) Smith et al. 2005a; C) Washington State University Vancouver (Unpublished); D) Smith et al. 2007; E) Smith et al. 2005b.

[b] This was removed because it is linked with *Ots_FGF6B* and using the paired genotypes did not add significant information.

[c] These were removed because they did not conform to Hardy–Weinberg expectations.

[d] These were removed because they were invariant.

Table 3.–Hierarchical log-likelihood ratio (G) analysis of population structure based on allele frequencies at 38 SNP loci. The probability of the statistic (P), assuming the null hypothesis is true, is provided for inferring significance. The scaled G statistic (G′) is provided for comparing levels of heterogeneity.

	DF	G	G'	*P*
Upper Cook Inlet	2,106	21,151	10.04	<0.001
Between Regions	39	4,887	125.31	<0.001
Within Regions	2,067	16,264	7.87	<0.001
Northern	1,404	6,723	4.79	<0.001
Among Groups	156	1,635	10.48	<0.001
Within Groups	1,248	5,088	4.08	<0.001
West	156	304	1.95	<0.001
Susitna	663	3,586	5.41	<0.001
Yentna	195	541	2.77	<0.001
Knik	117	418	3.58	<0.001
Turnagain	117	239	2.04	<0.001
Southern	663	9,541	14.39	<0.001
Among Groups	78	3,131	40.14	<0.001
Within Groups	585	6,410	10.96	<0.001
Kenai	429	5,318	12.40	<0.001
Kasilof	39	906	23.23	<0.001
SKenaiPen	117	186	1.59	<0.001

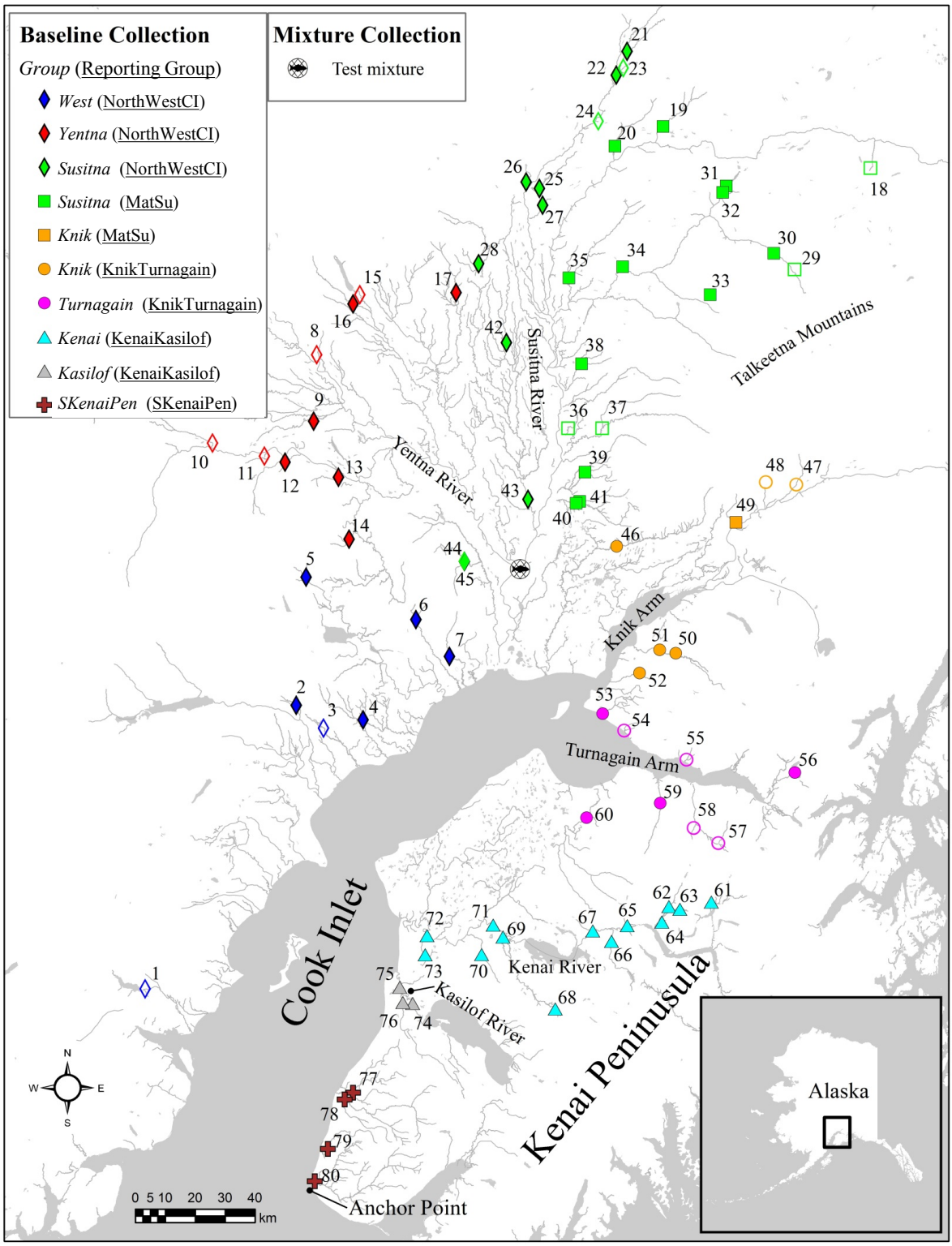

Figure 1.–Sampling locations for Chinook salmon originating for Upper Cook Inlet, Alaska, 1992–2013.

Note: Numbers correspond to map numbers on Table 1. Colors correspond to the 8 regional groups used for the hierarchical log-likelihood test and shapes correspond to the 5 reporting groups used in the mixed stock analysis tests, shapes without fill indicate collections that were not included in the baseline due to inadequate sample sizes.

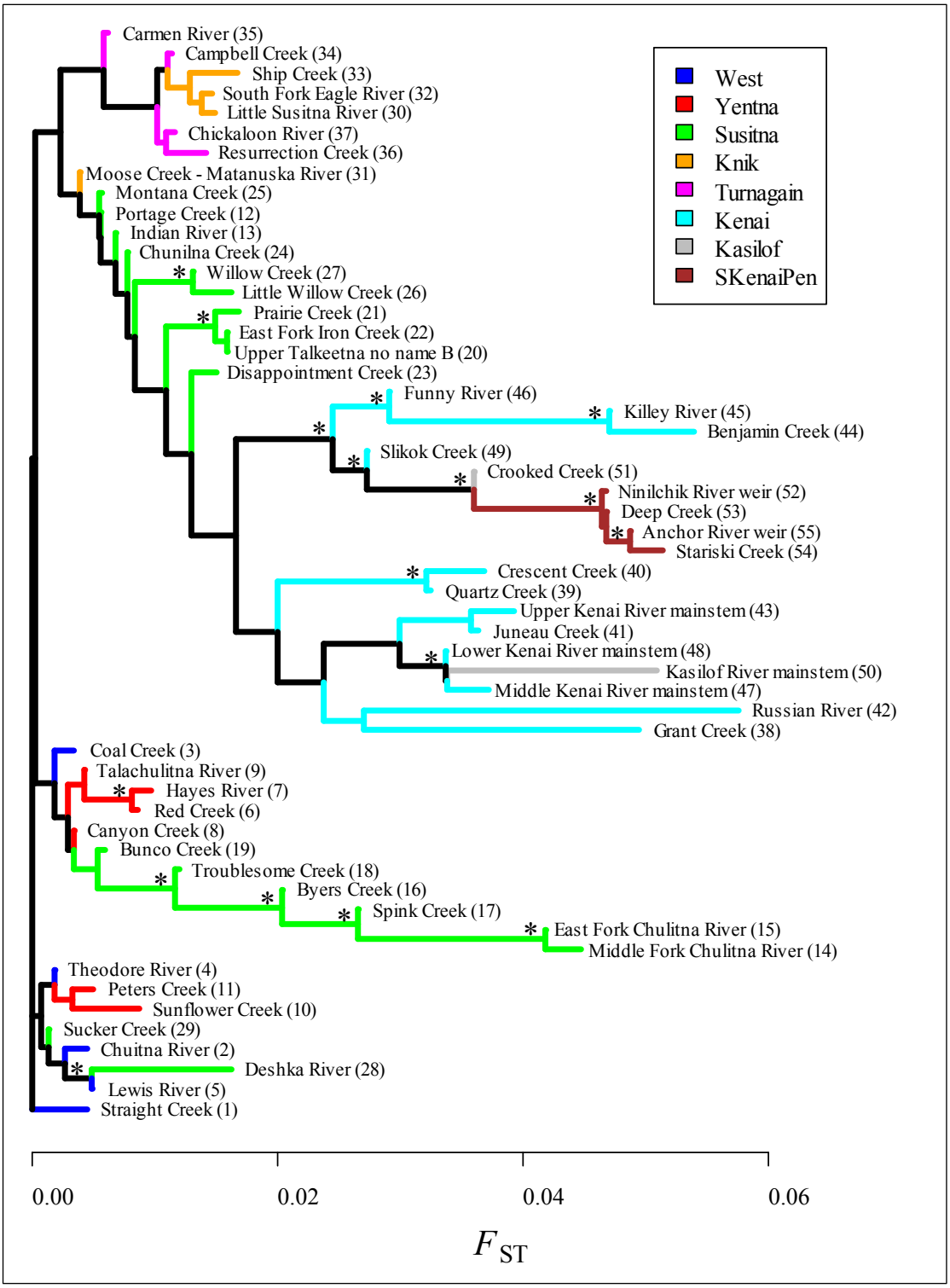

Figure 2.–Consensus neighbor-joining (NJ) tree based on F_{ST} between Chinook salmon populations sampled from spawning areas in drainages of Upper Cook Inlet, Alaska (see Table 1 for collection details).

Note: Colors denote groups as in Figures 1, 3, and 4. Numbers in parentheses correspond to unique population numbers on Table 1. Bootstrap consensus nodes occurring in >50% of trees are marked with an asterisk.

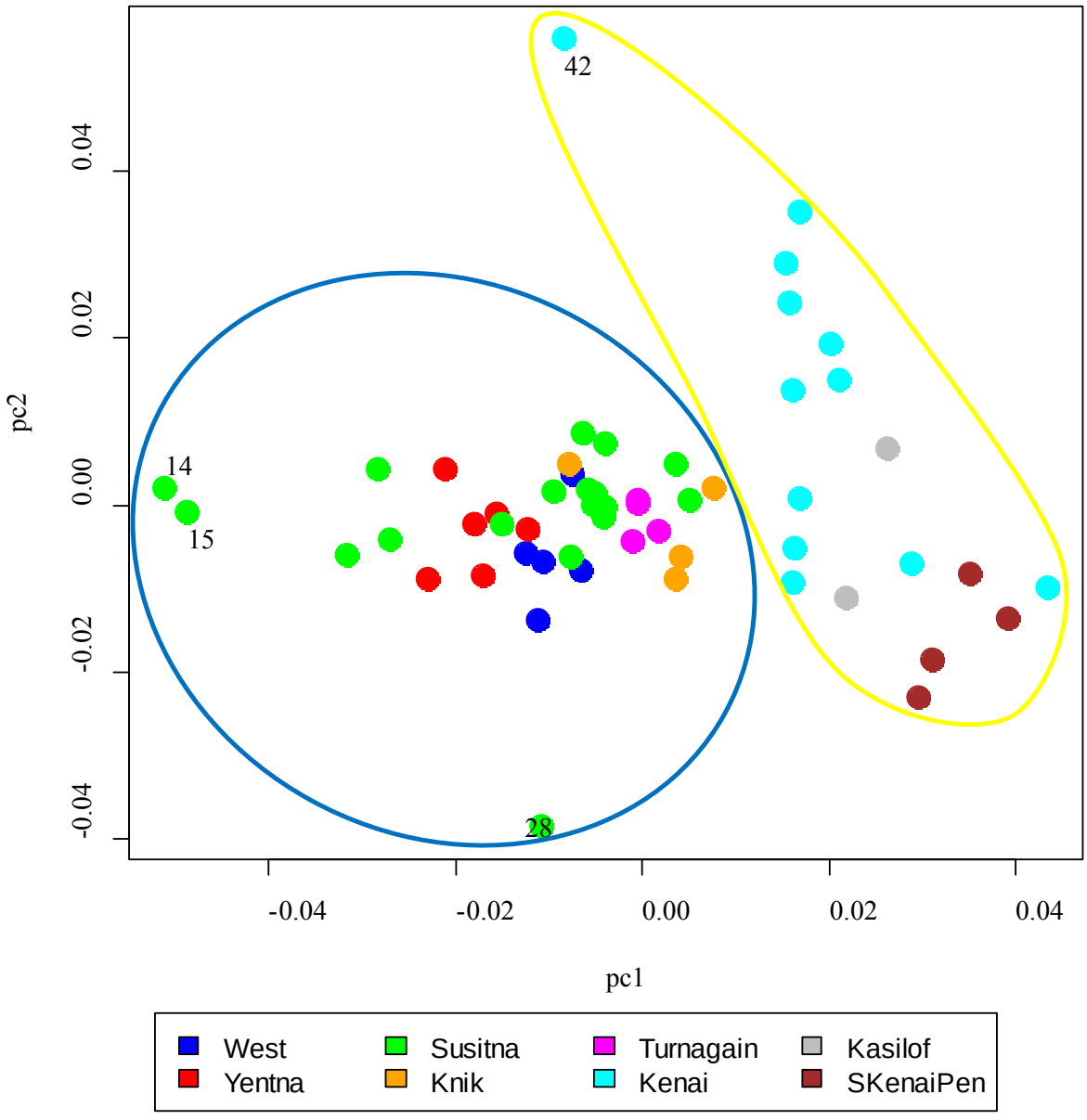

Figure 3.–Principal coordinates 1 (pc1) and 2 (pc2) based on F_{ST} between Chinook salmon populations sampled from spawning areas in drainages of Upper Cook Inlet, Alaska (see Table 1 for collection details).

Note: Group colors correspond to those in Figures 1, 2, and 4 and numbers on outlier populations correspond to unique population numbers on Table 1. Northern regions are outlined in blue; Southern regions are outlined in yellow.

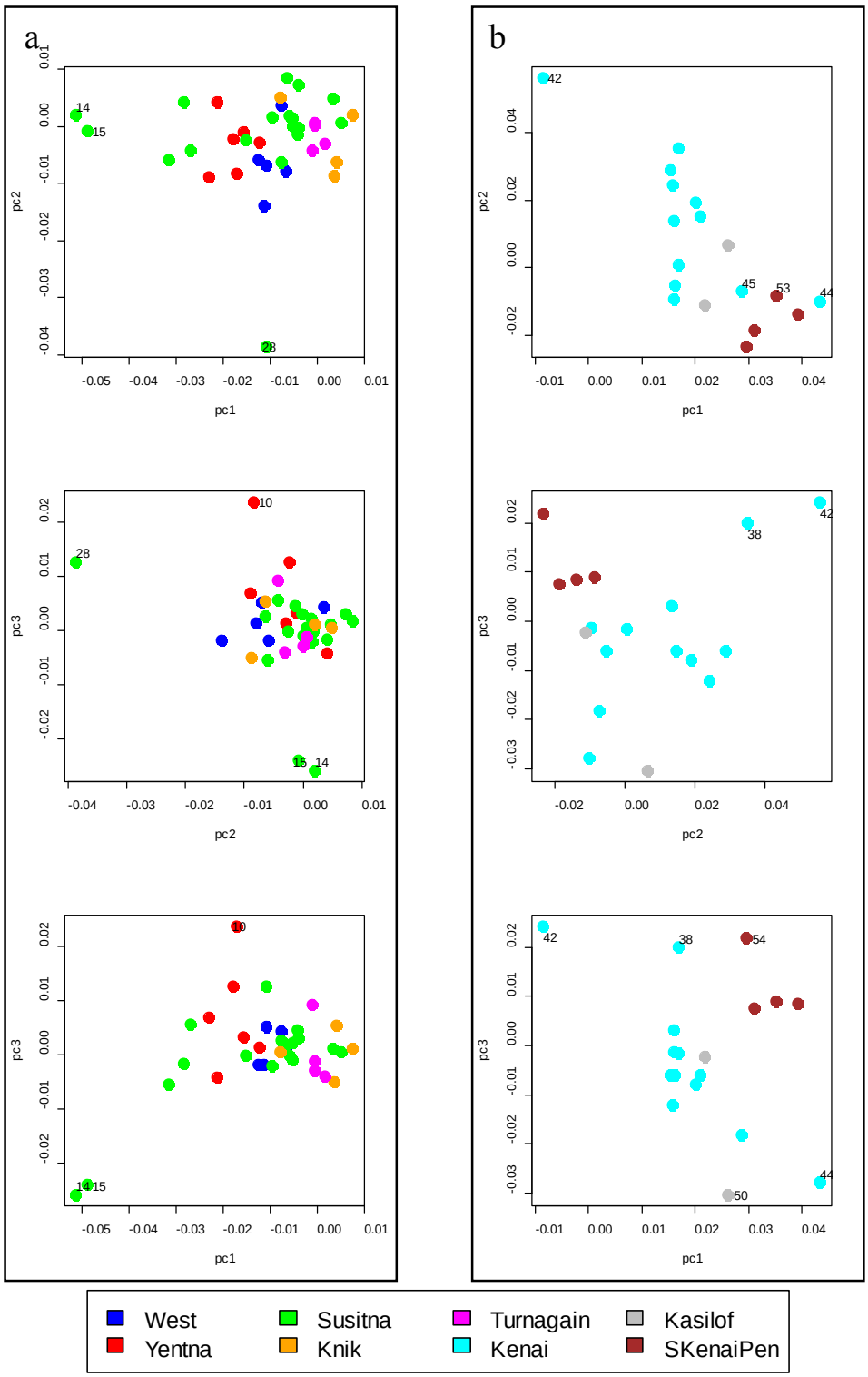

Figure 4.–Plots of principal coordinates (pc) 1–3 from a principal coordinated analysis based on F_{ST} between Chinook salmon populations from the a) Northern region and b) Southern region of Cook Inlet, Alaska.

Note: Group colors correspond to those in Figures 1–3 and numbers correspond to unique population numbers on Table 1.

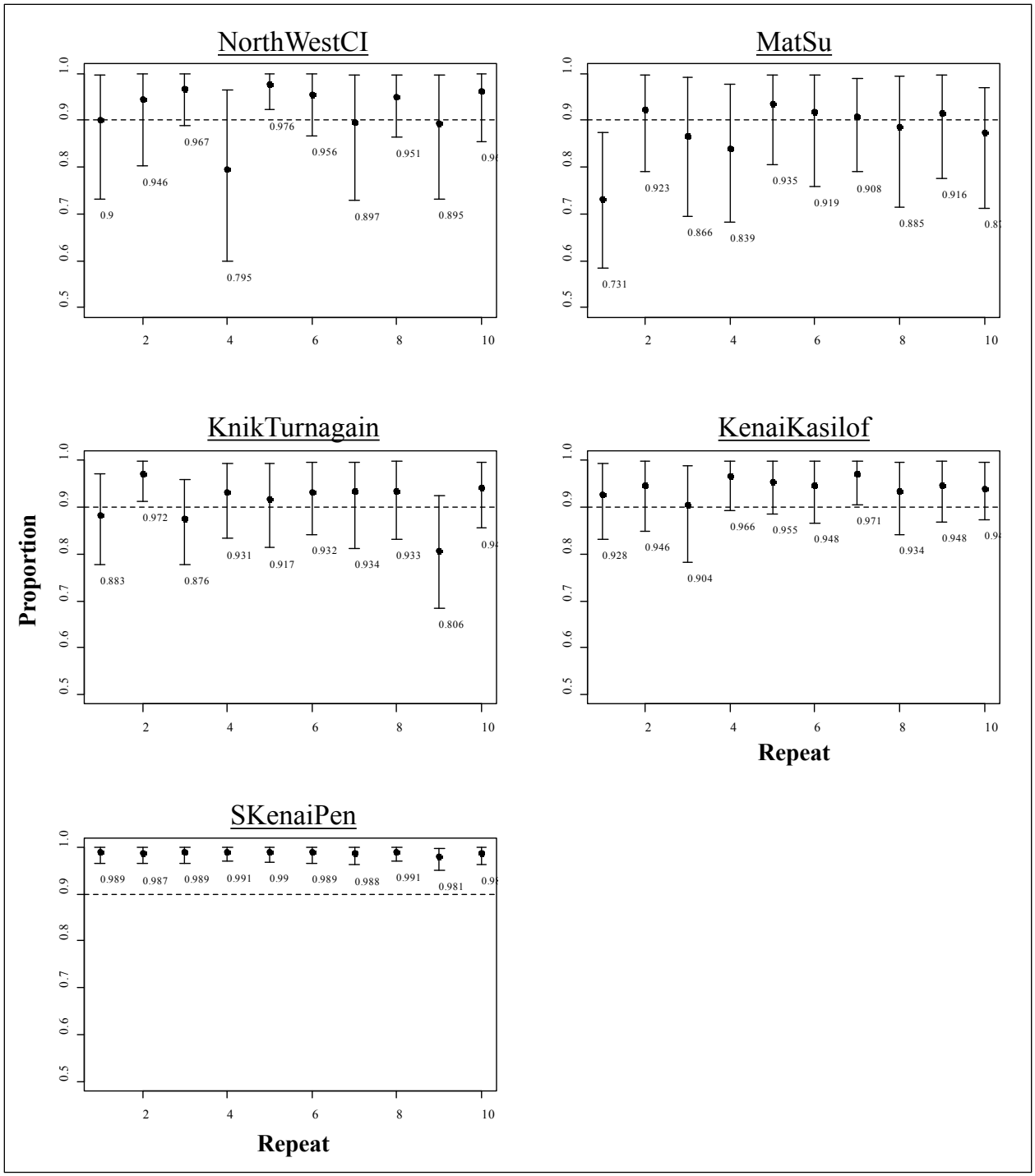

Figure 5.–Results of repeated proof testing for 5 reporting groups. The points represent the correct allocation from each repeat with 95% credibility intervals for each point. Point estimates for each repeat are included below the lower credibility interval.