**Fishery Manuscript Series No. 13-12**

# Genetic Baseline of Kenai River Chinook Salmon for Mixed Stock Analyses, 2013

**by**

**Serena D. Rogers Olive**

**Andrew W. Barclay**

**Timothy R. McKinley**

**and**

**William D. Templin**

**December 2013**

**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, $\chi^2$, 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. 13-12*

# GENETIC BASELINE OF KENAI RIVER CHINOOK SALMON FOR MIXED STOCK ANALYSES, 2013

By
Serena D. Rogers Olive, Andrew W. Barclay,
Division of Commercial Fisheries, Gene Conservation Laboratory, Anchorage

Timothy R. McKinley,
Division of Sport Fish, Soldotna

and
William D. Templin
Division of Commercial Fisheries, Gene Conservation Laboratory, Anchorage

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

December 2013

This investigation was partially financed by the Federal Aid in Sport Fish Restoration Act (16 U.S.C. 777-777K) under Grant F12AF00029, Project F-10-28, Job No. S-2-5a, and the Pacific Coastal Salmon Recovery Fund/Southeast Sustainable Salmon Fund (NOAA Grant No. NA04NMF4380162).

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.

*Serena D. Rogers Olive, Andrew W. Barclay, William D. Templin,*
*Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory*
*333 Raspberry Road, Anchorage, AK 99518, USA*

*and*

*Timothy R. McKinley,*
*Alaska Department of Fish and Game, Division of Sport Fish,*
*43961 Kalifornsky Beach Road, Suite B, Soldotna, AK 99669, USA*

*This document should be cited as:*
*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.*

# TABLE OF CONTENTS

# LIST OF TABLES



# LIST OF FIGURES

# ABSTRACT

The Kenai River supports two temporal runs of Chinook salmon. Previous studies detected genetic differences between these early and late Chinook salmon runs using microsatellite markers. Here we examine a baseline of 11 populations using 42 single nucleotide polymorphism markers to determine population structure, which provides insight into potential identifiable units (reporting groups) for mixed stock analysis. As with earlier studies, we confirmed significant genetic difference between the temporal runs. This new baseline also includes larger sample sizes for several populations (lower Kenai River mainstem, Quartz Creek, and Grant Creek), enabling a more fine-scale analysis of reporting groups. The baseline was examined for variation at three levels of population groupings: 1) broad-scale (2 groups); 2) mid-scale (3 groups); and 3) fine-scale (6 groups). Baseline tests indicated that there is sufficient variation within population groups for all three levels for Kenai River Chinook salmon genetics applications.

Key words: Chinook salmon, Kenai River, Cook Inlet, *Oncorhynchus tshawytscha*, single nucleotide polymorphism, SNP, genetic baseline, population structure, mixed stock analysis, MSA

# INTRODUCTION

## BACKGROUND

Chinook salmon *Oncorhynchus tshawytscha* returning to the Kenai River in Southcentral Alaska are harvested in several nearshore marine and inriver fisheries. Harvest of both early- and late-run Chinook salmon in Cook Inlet primarily occurs in inriver sport and commercial set gillnet fisheries, although marine sport, commercial drift gillnet, subsistence and personal use fisheries also harvest these fish (Eskelin and Miller 2010; Perschbacher 2012). However, because the majority of early-run Chinook salmon enter the Kenai River before the commercial fishery opens each year, harvest of these fish in the commercial set gillnet fishery is much smaller than the late-run (Eskelin et al. 2013; McKinley et al. *In prep*). Since the Kenai River is close to major population centers and has trophy-sized fish, the inriver sport fishery for Chinook salmon is both popular and economically important to communities on the Kenai Peninsula.

Chinook salmon enter the Kenai River in two temporal runs, which are managed by the Alaska Department of Fish and Game (ADF&G) under separate management plans adopted by the Alaska Board of Fisheries. The early run is managed using stipulations in the *Kenai River and Kasilof River Early-Run King Salmon Management Plan* (5 AAC 57.160), and the late run is managed using the *Kenai River Late-Run King Salmon Management Plan* (5 AAC 21.359). The two runs are defined by ecological differences (mainly run-timing and general spawning destination), but genetic differences have also been demonstrated (Begich et al. 2010). Chinook salmon returning to the Kenai River from mid-May through June 30 are designated as early-run and spawn primarily in Kenai River tributaries (Burger et al. 1983; Bendock and Alexandersdottir 1992; ADF&G 1998). The average annual total run of the early run from 1986 to 2012 was estimated most recently to be 13,877 Chinook salmon (McKinley and Fleischman 2013). Chinook salmon designated as late-run return to the Kenai River from July 1 to mid-August and are more numerous; the average annual total run from 1986 to 2012 was 58,899 (Fleischman and McKinley 2013). Late-run Chinook salmon spawn primarily in the mainstem of the Kenai River (Burger et al. 1983; Hammarstrom et al. 1985, Bendock and Alexandersdottir 1992).

Genetic analysis of samples from Chinook salmon captured in test gillnets as they entered the river and passed the sonar at river mile (RM) 8.5 have shown that some mainstem spawners (considered to be from the late run) enter the Kenai River prior to July 1, the date used to change management from the early-run plan to the late-run plan. Likewise, some tributary

spawners (considered to be from the early run) pass the sonar site throughout July (McKinley et al. *In prep*). Management for two separate runs based on a date is a practical compromise that does not completely correspond to the underlying structure of the populations (McKinley et al. *In prep*). For the purposes of this report, *early-run* and *late-run* are terms that will refer to distinctions that are important for fishery management and genetically identifiable units are referred to as *reporting groups*.

Both the early and late runs are monitored inseason and postseason. The runs are monitored in-season by a sonar system located at RM 8.5 (Miller et al. 2012). Age, sex, and length of the inriver run is estimated by a drift gillnetting program conducted near the sonar site (Eskelin 2007). Additionally, the magnitude and age-sex-length composition of the sport harvest are estimated by a creel survey (Eskelin 2007). Postseason, ADF&G also tallies information from other fisheries that are known to harvest Kenai River Chinook salmon, including 1) a personal use dipnet fishery at the mouth of the Kenai River (Dunker 2013), 2) educational fishery gillnet harvests from the Kenaitze Indian Tribe (Shields 2006), 3) the proportion of Chinook salmon harvested in the commercial set gillnet fishery on the east side of Cook Inlet (east-side set net; Eskelin 2013), 4) the proportion of Chinook salmon harvested in the commercial drift gillnet fishery on the east side of Cook Inlet (Upper Cook Inlet drift), and 5) the proportion of Chinook salmon harvested in the sport troll fishery on the east side of Cook Inlet (Deep Creek Marine). In addition to supporting inseason management, research and management programs are the foundation for long-term quantitative stock assessment of Kenai River Chinook salmon.

Despite these efforts, estimating the abundance of Chinook salmon returning to spawn in either the mainstem or tributaries of the Kenai River is still problematic. For instance, although each run is managed as a separate breeding group, the degree and variability of overlap in the run timing and inriver harvest timing of tributary- and mainstem- spawning Chinook salmon is not known, nor is the composition of Deep Creek Marine and Upper Cook Inlet drift harvests estimated directly. The accuracy of assessing the abundance of mainstem- and tributary-spawning Kenai River Chinook salmon would be substantially improved if these factors were known.

The genetic baseline for Kenai River Chinook salmon has been developed and updated repeatedly over the past 20 years, beginning with mitochondrial DNA (mtDNA) and allozymes using protein electrophoresis (Adams et al. 1994). This analysis identified genetic differences between tributary- and mainstem-spawning Chinook salmon in the Kenai River drainage. Following this, microsatellites were used to quantify genetic differences among populations within each spawning type as well as to provide better estimates of stock composition in samples taken at the sonar site (Begich et al. 2010). More recently, populations in the Kenai River were included in a larger-scale baseline describing genetic variation in Chinook salmon populations in all of Upper Cook Inlet (Barclay et al. 2012). This last version of the baseline used single nucleotide polymorphisms (SNPs) as genetic markers and was primarily concerned with describing broad-scale genetic variation and the potential for mixed stock analysis of samples taken from the marine waters of Cook Inlet. The application of the baseline for analysis of samples taken within the Kenai River was not fully developed, tested, or described.

This report presents an update of the genetic baseline used for studies of Chinook salmon within the Kenai River. In it we describe additional details: 1) population structure using archived and

new collections and 42 SNP loci, and 2) estimated precision and accuracy of stock compositions for three levels of reporting groups.

## DEFINITIONS

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

*Admixed Individuals*. Individuals with progenitors from two or more previously separated populations.

*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.

*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 a measure of variability in a sample.

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

*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.

*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.

*Polymerase Chain Reaction* (*PCR*). A method to amplify DNA sequences, which can be used to generate millions of copies of the DNA.

*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 and rearing 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 one or more identifiable units in a genetic baseline to which portions of a mixture are allocated during mixed stock analyses.

*Run.* The number of salmon in a stock surviving to adulthood and returning to their natal streams in a calendar year. A run is composed of both harvested adult salmon and the escapement to spawning areas. A run can designate the annual return of fish in a calendar year. With the exception of pink salmon, a run is composed of several age classes because individuals from a given brood year mature at different times (from 5 AAC 39.222(f)).

*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

Tissue samples (axillary process) suitable for genetic analyses (hereafter *genetic samples*) were collected and preserved in 95% ethanol. Samples were either placed in 2 ml cryovials (Begich et al. 2010) or placed collectively into 125 to 500 ml containers, with 1 or more containers for each collection site for each year.

Baseline genetic samples were collected from spawning aggregations of Chinook salmon by ADF&G personnel using gillnets, beach seines, or hook-and-line gear (Table 1; Figure 1). Chinook salmon captured on the Kenai River mainstem were considered baseline if they were in spawning condition. Target sample size for each baseline population was 95 individuals across all years to achieve acceptable precision to estimate allele frequency (Allendorf and Phelps 1981; Waples 1990).

## LABORATORY ANALYSIS

### Assaying Genotypes

Genomic DNA was extracted using a DNeasy® 96 Tissue Kit by QIAGEN® (Valencia, CA). Fluidigm® 96.96 Dynamic Arrays[1] were used to screen 42 SNP markers (Table 2). The Fluidigm® 96.96 Dynamic Array contains a matrix of integrated channels and valves housed in an input frame. On each side of the frame are 96 inlets, one side to accept sample DNA and the other to accept assays for a unique SNP marker. An IFC Controller HX (Fluidigm) was used to for mixing the sample DNA and assays under pressure to create 9,216 separate reactions. Each reaction consists of a mixture of 4 µl of assay mix (1 × DA Assay Loading Buffer [Fluidigm], 10 × TaqMan® SNP Genotyping Assay [Applied Biosystems], and 2.5 × ROX [Invitrogen]) and 5 µl of sample mix (1 × TaqMan® Universal Buffer [Applied Biosystems]), 0.05 × AmpliTaq® Gold DNA Polymerase [Applied Biosystems], 1 × GT Sample Loading Reagent [Fluidigm], and 60 to 400 ng/µl DNA) combined in a 7.2 nl chamber. Thermal cycling was performed on an Eppendorf IFC Thermal Cycler as follows: 70°C for 30 min for Hot-Mix step, initial denaturation of 10 min at 96°C followed by 40 cycles of 96°C for 15 s and 60°C for 1 min. The

[1] Available from http://www.fluidigm.com (Accessed December 20, 2013).

Dynamic Arrays were read on a Fluidigm® EP1 System after amplification and scored using Fluidigm® SNP Genotyping Analysis software.

Assays that failed to amplify on the Fluidigm system were reanalyzed on the Applied Biosystems platform. Each reaction on this platform was performed in 384-well reaction plates in a 5 µl volume consisting of 5 to 40 ng/µl of template DNA, 1 × TaqMan® Universal PCR Master Mix (Applied Biosystems), and 1 × TaqMan® SNP Genotyping Assay (Applied Biosystems). Thermal cycling was performed on a Dual 384-Well GeneAmp® PCR System 9700 (Applied Biosystems) as follows: an initial denaturation of 10 min at 95°C followed by 50 cycles of 92°C for 1 s and annealing/extension temperature for 1 min. The plates were scanned on an Applied Biosystems Prism 7900HT Sequence Detection System after amplification and scored using Applied Biosystems Sequence Detection Software 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. A single-locus genotype was considered a failure when it could not be satisfactorily scored for a fish.

Quality control (QC) measures were instituted to identify laboratory errors and to determine the reproducibility of genotypes. In this process, 8% of every extraction plate is re-extracted and re-analyzed for all markers by staff not involved in 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. Because baseline collections were genotyped during several separate laboratory analyses, we report QC results for a single recent laboratory baseline analysis as a representative of the QC on the entire dataset.

## STATISTICAL ANALYSIS

### Data Retrieval and Quality Control

Genotypes were imported from LOKI[2] into R[3] using the *RODBC* package.[4] All subsequent analyses were performed in *R* unless otherwise noted.

---

[2] ADF&G Gene Conservation Laboratory Oracle database. Alaska Department of Fish and Game, Division of Commercial Fisheries. URL not publicly available.

[3] The R Project for statistical computing. ISBN 3-900051-07-0. Available from http://www.r-project.org/ (Accesssed December 23, 2013).

[4] RODBC: ODBC database access. R package version 1.3-2. Available from http://CRAN.R-project.org/package=RODBC (Accessed August 15, 2013).

Three statistical analyses were performed to confirm the quality of the data. First, individuals missing substantial genotypic data were removed following the 80% rule (Dann et al. 2009) which requires individuals included in the analysis to have complete genotypes for at least 80% of the loci surveyed. The inclusion of individuals with poor quality DNA might introduce genotyping errors into the baseline and reduce the accuracy of population structure analyses.

Second, markers that were identified as being invariant were excluded. A marker was considered invariant if no individuals or very few individuals (1 or 2) in only one collection possessed the alternate allele.

Third, individuals with duplicate genotypes were identified and removed from further analysis. 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 (fish collected within the same year at the same location), we tested for conformance of genotype frequencies to Hardy-Weinberg expectations (HWE) using Monte Carlo simulation with 10,000 iterations in the *Adegenet* package (Jombart 2008). We combined probabilities for each collection across loci and for each locus across collections using Fisher's method (Sokal and Rohlf 1995) and tested for significance after correcting for multiple tests with the sequential Bonferroni method (familywise critical value $\alpha = 0.05$; Rice 1988). We removed collections and loci that violated HWE from subsequent analyses.

### *Pooling collections into populations*

When appropriate, collections taken from the same geographic location and sampled at similar calendar dates were pooled following the recommendations of Waples (1990). 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 tributary or contiguous section of the mainstem (i.e., lower mainstem and upper mainstem). 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 to create populations. Finally, we tested populations for conformance to HWE following the same protocol described above to ensure that pooling of collections into populations was appropriate, and that tests for linkage disequilibrium would not result in falsely positive results due to departure from HWE.

### *Linkage disequilibrium*

We tested for linkage disequilibrium between each pair of nuclear markers in each population to ensure that subsequent analyses would be based on independent markers. We used the program *Genepop* version 4.0.11 (Rousset 2008) with 100 batches of 5,000 iterations for these tests. We summarized the frequency of significant linkage disequilibrium between pairs of

SNPs ($P < 0.05$). Pairs were considered to be linked if they exhibited significant association between alleles in more than half of all populations.

For many types of analyses, linked markers can provide additional information for assessing population structure. We used a method developed by Jasper and Templin (2012) that uses $f_{ORCA}$ (Rosenberg 2005) as a measure for assessing whether linked SNP pairs should be combined into composite haploid markers. 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 ($\Delta = f_{ORCA\text{-}pair} - \max(f_{ORCA\text{-}single1}, f_{ORCA\text{-}single2})$. This difference ($\Delta$) was our test statistic. Since we did not know the distribution of $\Delta$, we conducted a sampled randomization test (Sokal and Rohlf 1995). We randomly selected 1,000 SNP pairs, calculated $\Delta$ for each pair to empirically define the test statistic distribution, and set the 90$^{th}$ quantile of the distribution as a critical value ($\Delta_{90}$). We then either combined linked SNPs into composite, haploid markers if $\Delta$ was greater than this critical value or dropped the SNP with the lower $f_{ORCA}$ value if $\Delta$ was less than the critical value.

## Analysis of Genetic Structure

### *Visualization of genetic distances*

We took 2 approaches to visualizing genetic distances among collections. The first approach was to construct a consensus neighbor-joining tree based on pairwise $F_{ST}$ from 1,000 bootstrap iterations in which loci were resampled with replacement to assess the stability of tree nodes. The consensus tree was constructed and then colored and visualized using TREEVIEW (Page 1996). We also plotted Cavalli-Sforza Edwards distances (Cavalli-Sforza and Edwards 1967) in a multidimensional scaling plot using the package *rgl*[5] to provide further insight into the genetic structure of these collections.

### *Hierarchical log-likelihood tests*

We examined genetic diversity within the Kenai River drainage with a 3-level hierarchical log-likelihood ratio (G) analysis based on allele frequencies at 38 SNP loci. Populations were grouped hierarchically into 6 fine-scale groups based specifically on drainage and geographic features established previously in an analysis reported by Begich et al. (2010; Tables 1 and 3). Chinook salmon spawning within the mainstem of the Kenai River were separated into upper and lower population groups: 1) *Upper Mainstem* (Juneau Creek and mainstem spawning locations above Skilak Lake), and 2) *Lower Mainstem* (from mainstem spawning locations below Skilak Lake). The tributaries were also split into upper and lower population groups. The upper tributaries that enter the Kenai River above Skilak Lake were chosen as a single group: 3) *Quartz* (populations from Quartz, Crescent, and Grant creeks), and 4) *Russian*. The lower tributary groups included 5) *Killey* (populations from Benjamin Creek and Killey River), and 6) *Funny* (populations from Funny River and Slikok Creek; Table 1). We tested for homogeneity of allele frequencies within groups, among groups within regions, and between regions. To compare levels of heterogeneity among regions and groups, scaled G-statistics (G′; Goudet et al. 1996) were calculated by dividing G by degrees of freedom.

[5] 3D visualization device system (OpenGL). R Package version 0.91. Available from http://CRAN.R-project.org/package=rgl (Accessed December 13, 2013).

### *Inferring population structure and genetic differentiation*

We inferred genetic population structure in two ways. First, we used the Bayesian clustering program BAPS 5.2.[6] While STRUCTURE (Pritchard et al. 2000) is the most commonly used program for inferring population structure, BAPS is more reliable when $F_{ST}$ is small (Latch et al. 2006). This analysis was completed in two steps. In the first step, we applied the *cluster of groups of individuals* mixture model, where all individuals are fully assigned to an inferred cluster. The mixture analysis was run 10 times for each value of $K = \{1,2,3,\ldots 9\}$. The clusters identified in the initial BAPS run were then used to perform the admixture analysis where individual admixture proportions are estimated based on the most likely number of clusters previously identified. The admixture analysis included 50 individuals for 100 iterations and was repeated using 200 reference individuals for 10 iterations as suggested in the BAPS manual (Corander et al. 2009).

Second, we used the HIERFSTAT package (Goudet 2005) implemented in R version 2.1 to calculate *F*-statistics (Weir 1996) and observed heterozygosities for each locus and population.

### *Evaluating reporting groups for mixed stock analysis*

Mixed stock analysis uses genetic markers to estimate stock origin of fish caught in a mixed stock fishery (Shaklee et al. 1999). The Bayesian mixed stock analysis method implemented in BAYES (Pella and Masuda 2001) was used to run repeated proof tests to evaluate 3 sets of reporting groups. The division of reporting groups was based on previous groupings in Begich et al. (2010). The first evaluation was for a 2 reporting group system where the baseline populations were divided into 2 broad-scale reporting groups (mainstem and tributaries). Next we tested for a 3 reporting group mid-scale system, which consisted of the same mainstem group, but divided the tributaries into upper and lower groups. In the last test, we divided the baseline into 6 fine-scale groups (these are the same as the groups chosen for the hierarchical G-test with the exception of grouping Grant Creek with Russian River); Grant Creek did not have a large enough sample size (n = 55) to be considered a reporting group on its own. Analysis of genetic distances showed that this population was highly divergent from the *Quartz* group—where it had originally been placed in the hierarchical G-test. Since Grant Creek and Russian River are the two most divergent populations, we grouped them together for the purpose of MSA. These two populations were not genetically similar to each other, nor to any other population, and were not expected to contribute significantly to fishery samples. Grouping these two populations reduces bias in the MSA. Each test was repeated 10 times, using 5 chains of 20,000 iterations with 200 individuals per group. Sample size for each fine-scale group was 100 individuals, with the exception of the *Russian* and *Quartz* groups (each were assigned 50 individuals) because they had fewer than 300 samples.

# RESULTS

## TISSUE SAMPLING

A total of 2,216 Chinook salmon were sampled in spawning areas throughout the Kenai River drainage. Of the total samples collected, 2,205 were selected to be genotyped. The final

[6] University of Helsinki Bayesian Statistics Group. Available from http://www.helsinki.fi/bsg/software/index (Accessed December 23, 2013).

numbers of individuals genotyped and used for this baseline are found in Table 1. Samples collected between 2003 and 2006 were previously reported in Begich et al. (2010) and the SNP genotypes were reported in Templin et al. (2011). Between 2007 and 2011, an additional 14 collections were made from unrepresented and underrepresented areas of the drainage and reported in Barclay et al. (2012). Quartz Creek and Grant Creek were updated with additional samples in this baseline, for a total of 31 collections from the Kenai River drainage. Individuals from all locations, except Crescent Creek and the *Upper Mainstem*, were collected in multiple years (Table 1). The mainstem samples were not collected at a single location, but over several river miles of the upper and lower Kenai River mainstem. Sampling from the *Lower Mainstem* took place from RM 12.2 to 36.0 and RM 39.8 to 47.9. *Upper Mainstem* sampling took place from RM 65.8 to RM 79.7. Target sample sizes of 95 fish were met at 10 locations.

## LABORATORY ANALYSIS

### Assaying Genotypes

A total of 2,205 fish collected over spawning areas were selected for analysis and assayed for 42 SNP markers (Table 2). Of the 42 markers, 4 departed significantly from HWE or were linked with other markers and were removed from further analysis.

### Laboratory Failure Rates and Quality Control

For the 31 collections in the baseline, the overall failure rate for genotypes at the complete marker set (42 SNP markers) was 3.52%. The laboratory analysis of collections for this baseline has occurred as part of several projects, so the results from the most recent baseline project were used as a measure for the combined dataset. This project included 1,950 individuals and the discrepancy rate was 0.19%; therefore, the overall error rate was estimated to be 0.09%.

## STATISTICAL ANALYSIS

### Data Retrieval and Quality Control

For all analyzed collections, only a single SNP marker was found to be invariant in all but 2 individuals (Table 2). This marker was removed from further analyses. Based upon the 80% rule, 4.8% of individuals (n = 106) were removed from the baseline collections for not having high-quality DNA. Based on the 95% of loci criterion for detecting duplicate individuals, 0.45% of individuals (n = 10) were removed from baseline collections as duplicate individuals. No duplicate individuals were detected in 24 of the 31 baseline collections (77%).

### Baseline Development

#### *Hardy-Weinberg expectations*

Over all nuclear markers and collections, 15 of 1,271 tests deviated significantly from HWE ($P < 0.01$) without adjusting for multiple tests. These were spread over 14 markers, and no markers were out of HWE in more than 2 of the 31 collections. No collections departed from HWE at more than 3 of the 41 markers. After adjusting for multiple tests, all collections conformed to HWE.

#### *Pooling collections into populations*

A total of 11 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). Dave's Creek, a small tributary in the Quartz Creek drainage, pooled with the temporal Quartz Creek collections to form a single population. The *Quartz* population will refer to both of these sets of collections from this point forward. All collections pooled into populations met the minimum requirements of Fisher's method ($P > 0.01$) with the exception of the collections from Slikok Creek ($P = 0.0011$). However, the Slikok collections were collected across multiple years (2004, 2005, and 2008), and differences may have been due to temporal fluctuation of allele frequencies, so we chose to pool them anyway. Over all variant markers and populations, 14 of 451 tests did not conform to HWE ($P < 0.01$) without adjusting for multiple tests. These were spread over 10 markers and no markers were out of HWE in more than 4 of the 11 populations. After adjusting for multiple tests, 1 SNP marker (*Ots_il-1racp-166*) did not conform to HWE and was removed from further analyses. No population was out of HWE at more than 3 of 41 markers. After adjusting for multiple tests, all populations conformed to HWE.

### *Linkage disequilibrium*

In the tests for linkage disequilibrium, 2 SNP pairs showed evidence of linkage at $P < 0.05$ (*Ots_HSP90B-385* & *Ots_HSP90B-100*, for 72% of populations and *Ots_FGF6A* & *Ots_FGF6B*, in 100% of populations). The $f_{ORCA}$ 90% critical value ($\Delta_{90}$) was 0.044, which was greater than the $\Delta$ for *Ots_HSP90B-100* & *Ots_HSP90B-385* ($\Delta = 0.001$) and less than the $\Delta$ for *Ots_FGF6A & Ots_FGF6B* ($\Delta = 0.047$). An examination of the correlation coefficient *r* of alleles at the two *HSP90B* SNP's did not reveal a useful pattern of linkage across reporting groups, so the SNP with the lowest $f_{ORCA}$ value (*Ots_HSP90B-385)* was dropped from further analysis. The two *FGF6* SNPs were combined to form a single locus (*Ots_FGF6A_FGF6B).*

## Analysis of Genetic Structure

### *Visualization of genetic distances*

Genetic relationships among baseline populations were visualized in an neighbor-joining tree (Figure 2) and multidimensional scaling plots (Figure 3). In general, the neighbor-joining tree indicates populations clustering with other populations within the same area except for the 2 most genetically distinct populations (longer genetic branches): Russian River and Grant Creek. Five of 8 nodes were well supported (> 50% of bootstrap trees). The multidimensional scaling plots, as with the tree, show Grant Creek and Russian River as genetically distinct from other populations. The geographic extents of the 3 reporting groups were colored to show the distributions of the upper tributary, lower tributary, and mainstem populations (Figure 3b).

### *Hierarchical log-likelihood tests*

In the analysis of genetic heterogeneity, significant variation was found within and among both broad- and fine-scale population groups (Table 3). Greater among-group heterogeneity was found in the tributaries particularly among the upper tributaries (G' = 25.19). A test of heterogeneity within group could not be conducted on the *Russian* and *Lower Mainstem* groups because they only included 1 population each.

### *Inferring population structure and genetic differentiation*

Replicate runs of the clustering analysis in BAPS resulted in identical results, identifying 7 population clusters (log likelihood of the model = -61,964.4; posterior probability = 1; Figure 4). As with the previous analyses, BAPS clustered neighboring tributary populations with each other

(Figure 2) with the exception of Russian River and Grant Creek. The admixture analysis based on 100 simulations from posterior allele frequencies revealed 1.1% of all individuals had significant admixture ($P \leq 0.05$), whereas 2,005 individuals (96%) with the maximum probability of P = 1 were not significantly admixed.

Observed heterozygosities for SNP loci remaining in the baseline analysis ranged from 0.007 (*Ots_PGK-54*) to 0.500 (*Ots_IGF-I.1-76*; Table 2). Observed heterozygosity averaged over all loci was 0.245. Estimates of $F_{ST}$ ranged between 0.008 (*Ots_S7-1*and *Ots_GH2*) and 0.093 (*Ots_NOD1*) and for loci remaining in the baseline.

### *Evaluating reporting groups for mixed stock analysis*

Correct allocations for all 10 draws for both the 2 and 3 reporting group proof tests were well above 90% (Figures 5a and b). Most draws for the 6 reporting groups were also above 90% except *Quartz* (1 draw), *Lower Mainstem* (2 draws), and *Funny* (2 draws; Figure 6). The draws that were below 90% for these groups were all above 85% with the exception of *Funny* that had a single draw at 75.4%. However, the variance around the 6 groupings is much larger than for the 2 and 3 group tests.

# DISCUSSION

## COMPARISONS TO PREVIOUS FINDINGS

This is the most comprehensive analysis of baseline samples collected to test for fine-scale population structure among Chinook salmon returning to the Kenai River. Previous studies included a smaller set of collections and/or used other genetic markers to assess population structure. In one of the earliest studies, Adams et al. (1994) used mitochondrial DNA and allozyme markers to discriminate between mainstem and tributary spawning Chinook salmon in the Kenai and Kasilof rivers. That study was based on 400 samples representing 4 populations. In the first broad-scale study of genetic structure containing Cook Inlet populations, Crane et al. (1996) found significant heterogeneity among and within populations from the Susitna, Kenai, and Kasilof rivers using allozyme loci based on 496 samples representing 6 populations. The next study which focused on the Kenai River drainage was based on 977 samples representing 9 populations and used 13 microsatellites. That study found differences adequate for MSA within the drainage among 4 groups of populations (Lower Kenai River tributaries, Kenai River mainstem, Killey River, and Quartz Creek; Begich et al. 2010). Results from the fine-scale reporting group analysis in this study include the 4 groups identified by Begich et al. (2010). The most recently published study focused on broad-scale genetic structure of Chinook salmon populations around the Cook Inlet, but used SNP loci as genetic markers (Barclay et al. 2012). That study found differences adequate for MSA within the Kenai River and between the Kenai River, Kasilof River, Lower Kenai Peninsula rivers, and the Susitna River drainages and was based on 5,279 samples representing 30 populations.

Genetic relationships among populations within the current study agree with previous findings. Genetic relationships between tributary (majority of the early-run) and mainstem (majority of the late-run) populations in the Kenai River are similar to those first described in Adams et al. (1994); Kenai River tributary populations are distinct from the mainstem populations. The groups that this study found useful for MSA within the Kenai River are similar to the groups resulting from microsatellite analysis (where populations can be clustered into at least 3 groups useful for MSA), 2 tributary groups, and a single mainstem group.

## NEW FINDINGS

Relationships among populations that were not previously examined provide additional insight into population structure within the Kenai River. This is the first fine-scale analysis of population structure within the Kenai River to include representation of spawning Chinook salmon from Dave's and Grant creeks and the Upper Mainstem spawning locations. The number of samples available from other locations were also increased to levels considered adequate to truly represent the Chinook salmon spawning at these locations, including Quartz and Juneau creeks and Russian River. Many of these additional populations continue to cluster with other populations from the same tributary, but Russian River and Upper Mainstem clusters have been identified as potential separate reporting groups for MSA in the Kenai River. While some of these patterns of population structure were identified in broad-scale analyses included in Barclay et al. (2012), the results reported here confirm their utility as reporting groups in Kenai River MSA.

## USE IN GENETIC STUDIES

This report describes the underlying population structure and potential application of the Kenai River Chinook salmon baseline for genetic studies on Chinook salmon in Kenai River. Here we provide more in-depth population structure than was reported in Barclay et al. (2010) and demonstrate the accuracy and precision that can be expected from the baseline for genetics applications. This baseline is being used in studies to estimate the escapement and inriver run timing of Chinook salmon in the Kenai River.

# ACKNOWLEDGEMENTS

This study required the efforts of a large number of dedicated people. Funding for this work was provided by grants from the Alaska Sustainable Salmon Fund (AKSSF). Most importantly, we would like to acknowledge the work of the people in the ADF&G Gene Conservation Laboratory including: Wei Cheng, Zac Grauvogel, Heather Liller, Tara Harrington, and Judy Berger. Study design and field logistics for sample collections were completed by many dedicated staff that performed this task in addition to many other duties. We would like to thank the Soldotna Division of Sport Fish staff who, over the past 10 years, have collected enough baseline samples to make this study possible. Specifically we would like to thank Robert Begich, Jeff Perschbacher, Adam Reimer, Jerry Strait, Jenny Cope, Tony Eskelin, Jason Pawluk, Rob Massengill, Christy Cupp, Tom Johnson, Jeremiah Batsin, Tom Rhyner, Tye Wyatt, Traye Turner, Julianne Pettey, and Patti Berkhahn, who is now with the Divison of Habitat and Resotration. Finally, we would like to thank W. Stewart Grant, Christopher Habicht, Eric Volk, T. Mark Willette, and Jack Erickson for providing editorial review that increased the clarity of this report.

# 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(Supplement 1):172–181.

ADF&G (Alaska Department of Fish and Game). 1998. Catalog of waters important for spawning, rearing or migration of anadromous fishes, Southcenteral region. Alaska Department of Fish and Game, Habitat Division, Juneau.

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.

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., 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.

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.

Bendock, T. N., and M. Alexandersdottir. 1992. Mortality and movement behavior of hooked-and-released chinook salmon in the Kenai River recreational fishery, 1989-1991. Alaska Department of Fish and Game, Fishery Manuscript No. 92-2, Anchorage.

Burger, C. V., D. B. Wangaard, R. L. Wilmot, and A. N. Palmisano. 1983. Salmon investigations in the Kenai River, Alaska, 1979-1981. Alaska Field Station, National Fishery Research Center, Fish and Wildlife Service, U.S. Department of the Interior, Anchorage.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic Analysis: Models and Estimation Procedures. Evolution 21(3):550–570.

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.

Dunker, K. 2013. Upper Cook Inlet salmon personal use harvest monitorying (2013-2015). Alaska Department of Fish and Game, Regional Operational Plan ROP.SF.2A.2013.03, Anchorage.

Eskelin, A. 2007. Chinook salmon creel survey and inriver gillnetting study, lower Kenai River, Alaska, 2005. Alaska Department of Fish and Game, Fishery Data Series No. 07-87, Anchorage.

Eskelin, T., and J. D. Miller. 2010. A qualitative evaluation of parameters used to assess Kenai River king salmon, 1986-2010. Alaska Department of Fish and Game, Special Publication No. 10-18, Anchorage.

Eskelin, T., A. W. Barclay, and A. Antonovich. 2013. Genetic stock identification and age, sex, and length composition of Chinook salmon in upper Cook Inlet, Alaska, 2010-2013. Alaska Department of Fish and Game, Fishery Data Series No. 13-63, Anchorage.

Fleischman, S. J., and T. R. McKinley. 2013. Run reconstruction, spawner-recruit analysis, and escapement goal recommendation for late run Chinook salmon in the Kenai River. Alaska Department of Fish and Game, Fishery Manuscript Series No. 13-02, Anchorage.

# REFERENCES CITED (Continued)


Goudet, J., M. Raymond, T. DeMeeus, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144:1933–1940.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. Molecular Ecology Notes 2:184–186.

Hammarstrom, S. L., L. Larson, M. Wenger, and J. Carlon. 1985. Kenai Peninsula chinook and coho salmon studies. Alaska Department of Fish and Game. Federal Aid in Fish Restoration. Annual Performance Report, 1984-1985, Project F-9-17, 26 (G-II-L), Juneau.

Jasper, J. R., and W. D. Templin. 2012. Western Alaska Salmon Stock Identification Program Technical Document 10: Optimal rate of correct assignemnt with backward elimination locus selection. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Imformation Report 5J12-19, Anchorage.

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

Latch, E. K., G. Dharmarajan, J. C. Glaubitz, and O. E. Rhodes Jr. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics 7:295–302.

McKinley, T. R., A. Barclay, and J. Jasper. *In prep*. Inriver run timing and sport harvest timing for tributary and mainstem spawning Chinook salmon in the Kenai River, Alaska. Alaska Department of Fish and Game, Fishery Manuscript Series No., Anchorage.

McKinley, T. R., and S. J. Fleischman. 2013. Run reconstruction, spawner-recruit analysis, and escapement goal recommendation for early-run Chinook salmon in the Kenai River. Alaska Department of Fish and Game, Fishery Manuscript Series No. 13-03, Anchorage.

Miller, J. D., D. L. Burwen, and S. J. Fleischman. 2012. Estimates of Chinook salmon passage in the Kenai River using split-beam sonar, 2008-2009. Alaska Department of Fish and Game, Fishery Data Series No. 12-73, Anchorage.

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.

Page, R. D. 1996. TreeView: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12:357–358.

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

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

Perschbacher, J. 2012. Chinook salmon creel survey and inriver gillnetting study, lower Kenai River, Alaska, 2011. Alaska Deartment. of Fish and Game, Fishery Data Series No. 12-84, Anchorage.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155(2):945–959.

Rice, W. R. 1988. Analyzing tables of statistical tests. Evolution 43(1):223–235.

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.

Rosenberg, N. A. 2005. Algorithms for selecting informative marker panels for population assignment. Journal of Computation 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.

Shaklee, J. B., T. D. Beacham, L. Seeb, and B. A. White. 1999. Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fisheries Research 43:45–78.

# REFERENCES CITED (Continued)

Shields, P. 2006. Upper Cook Inlet commercial fisheries annual management report, 2005. Alaska Department of Fish and Game, Fishery Management Report Number 06-42, Anchorage.

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

Smith, C. T., W. D. Templin, J. E. Seeb, and L. W. Seeb. 2005b. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook salmon caught in Yukon River fisheries. North American Journal of Fisheries Management 25:944–953.

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.

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

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(Supplement 1): 226–246.

Waples, R. S. 1990. Temporal changes of allele frequency in Pacific Salmon: implications for mixed-stock fishery analysis. Canadian Journal of Fisheries and Aquatic Sciences 47(5):9.

Weir, B. S. 1996. Genetic Data Analysis II. 2nd Edition. Sinauer, Sunderland, MA.

# TABLES

Table 1.–Tissue collections of Chinook salmon throughout the Kenai River drainage, including the year sampled, number of samples collected (Initial), the number removed having less than 80% loci (Missing), the number removed for matching at 95% loci (Duplicates), the final number of individuals analyzed (Final), and the assigned reporting group for each collection. Map numbers correspond to populations in Figure 1.

| | Reporting Groups | | | | | | Number of Individuals | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Map No. | Population Groups | Broad-scale (2 Groups) | Mid-scale (3 Groups) | Fine-scale (6 Groups) | Location | Year(s) Collected | Initial | Missing | Duplicate | Final |
| | | Mainstem | Mainstem | | | | | | | |
| | | | | Upper Mainstem | | | | | | |
| 1 | 1 | | | | Upper mainstem[a] | 2009 | 200 | 9 | 0 | 191 |
| 2 | 1 | | | | Juneau Creek | 2005-2007 | 147 | 3 | 3 | 141 |
| | | | | Lower Mainstem | | | | | | |
| 3 | 2 | | | | Lower mainstem[a] | 2003/2004/2006/2011 | 393 | 13 | 0 | 380 |
| | | Tributary | Upper | | | | | | | |
| | | | | Quartz | | | | | | |
| 4 | 3 | | | | Quartz/Dave's Creek | 2006-2011 | 139 | 5 | 3 | 131 |
| 5 | 3 | | | | Crescent Creek | 2006 | 165 | 0 | 1 | 164 |
| | | | | Russian | | | | | | |
| 6 | 3 | | | | Grant Creek | 2011/2012 | 55 | 0 | 0 | 55 |
| 7 | 4 | | | | Russian River | 2005-2008 | 214 | 0 | 0 | 214 |
| | | | Lower | | | | | | | |
| | | | | Killey | | | | | | |
| 8 | 5 | | | | Benjamin Creek | 2005/2006 | 206 | 2 | 0 | 204 |
| 9 | 5 | | | | Killey River | 2005/2006 | 266 | 10 | 2 | 254 |
| | | | | Funny | | | | | | |
| 10 | 6 | | | | Funny River | 2005/2006 | 220 | 0 | 1 | 219 |
| 11 | 6 | | | | Slikok Creek | 2004/2005/2008 | 200 | 64 | 0 | 136 |
| | | | | | | Total | 2,205 | 106 | 10 | 2,089 |

[a] These sites were sampled over several river miles within the mainstem of the Kenai River.

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 Kenai River Chinook salmon. These summary statistics are based upon 11 populations within the Kenai River drainage.

| Assay Name | Source[a] | $H_O$ | $F_{IS}$ | $F_{ST}$ |
|---|---|---|---|---|
| *Ots_AsnRS-60* | b | 0.410 | 0.033 | 0.034 |
| *Ots_E2-275* | b | 0.339 | 0.003 | 0.039 |
| *Ots_ETIF1A* | c | 0.459 | 0.007 | 0.055 |
| *Ots_FARSLA-220* | d | 0.349 | -0.001 | 0.046 |
| *Ots_FGF6A*[b] | a | 0.365 | 0.020 | 0.081 |
| *Ots_FGF6B*[b] | a | 0.322 | 0.024 | 0.033 |
| *Ots_GH2* | e | 0.296 | 0.009 | 0.008 |
| *Ots_GPDH-338* | b | 0.107 | -0.001 | 0.031 |
| *Ots_GPH-318* | d | 0.132 | 0.026 | 0.026 |
| *Ots_GST-207* | d | 0.045 | 0.041 | 0.049 |
| *Ots_GTH2B-550* | a | 0.446 | -0.022 | 0.016 |
| *Ots_hnRNPL-533* | d | 0.306 | 0.006 | 0.026 |
| *Ots_HSP90B-100* | d | 0.243 | 0.045 | 0.038 |
| *Ots_HSP90B-385*[c] | d | 0.050 | -0.007 | 0.051 |
| *Ots_IGF-I.1-76* | b | 0.500 | -0.041 | 0.046 |
| *Ots_Ikaros-250* | b | 0.137 | -0.008 | 0.049 |
| *Ots_il-1racp-166* | b | 0.471 | -0.068 | 0.034 |
| *Ots_ins-115* | b | 0.036 | -0.009 | 0.013 |
| *Ots_LEI-292* | d | 0.071 | 0.022 | 0.011 |
| *Ots_LWSop-638* | b | 0.058 | -0.051 | 0.023 |
| *Ots_MHC1* | e | 0.416 | 0.036 | 0.057 |
| *Ots_MHC2* | e | 0.030 | 0.029 | 0.011 |
| *Ots_NOD1* | a | 0.460 | -0.004 | 0.093 |
| *Ots_P450* | e | 0.364 | 0.025 | 0.017 |
| *Ots_P53* | e | 0.417 | 0.015 | 0.022 |
| *Ots_PGK-54* | a | 0.007 | -0.013 | 0.012 |
| *Ots_Prl2* | e | 0.490 | -0.001 | 0.020 |
| *Ots_PSMB1-197*[d] | d | --- | --- | --- |
| *Ots_RAG3* | a | 0.247 | 0.034 | 0.062 |
| *Ots_S7-1* | a | 0.218 | 0.026 | 0.008 |
| *Ots_SClkF2R2-135* | b | 0.413 | -0.003 | 0.059 |
| *Ots_SERPC1-209* | d | 0.123 | 0.105 | 0.051 |
| *Ots_SL* | e | 0.411 | -0.043 | 0.035 |
| *Ots_SWS1op-182* | b | 0.456 | -0.038 | 0.015 |
| *Ots_TAPBP* | c | 0.248 | -0.005 | 0.025 |
| *Ots_Tnsf* | e | 0.196 | -0.021 | 0.010 |

-continued-

Table 2.–Page 2 of 2.

| Assay Name | Source[a] | $H_O$ | $F_{IS}$ | $F_{ST}$ |
|---|---|---|---|---|
| *Ots_u202-161*[e] | [b] | 0.037 | -0.006 | 0.039 |
| *Ots_u211-85* | [b] | 0.187 | -0.031 | 0.022 |
| *Ots_U212-158* | [b] | 0.045 | 0.038 | 0.070 |
| *Ots_u6-75* | [b] | 0.129 | -0.039 | 0.017 |
| *Ots_unk526* | [a] | 0.181 | 0.014 | 0.042 |
| *Ots_Zp3b-215* | [b] | 0.064 | 0.013 | 0.016 |
| *Ots_FGF6A_FGF6B* | [a] | --- | 0.000 | 0.056 |
| Average/Overall[f] | | 0.245 | 0.002 | 0.038 |

[a] Marker sources: a) Anna Elz, Anna.Elz@noaa.gov, Northwest Fisheries Science Center, personal communication; b) Smith et al. 2005a; c) Jennifer DeKoning, dekoning@vancouver.wsu.edu, Washington State University, Vancouver, personal communication; d) Smith et al. 2007; e) Smith et al. 2005b.

[b] These SNPs were combined into haplotypes and treated together as a single locus.

[c] Removed from further analyses due to linkage.

[d] Removed from further analyses for being invariant in all but two individuals.

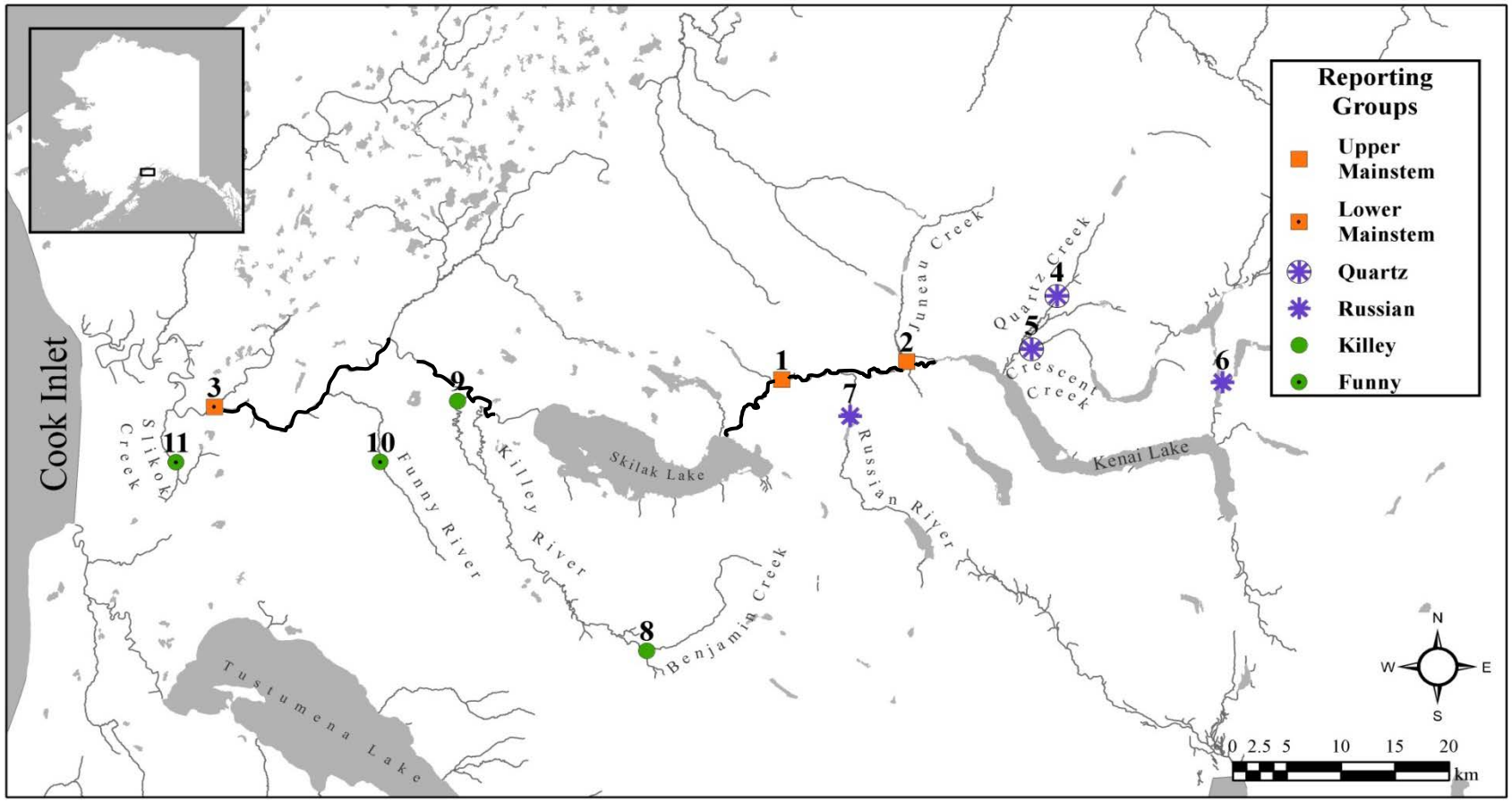
[e] Removed from further analyses for not conforming to HWE.

[f] Calculated from the 38 loci selected for further analysis.

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 that populations are homogeneous is true, is provided for inferring significance. The scaled G statistic (G′) is provided for comparing levels of heterogeneity. Population group numbers found on Table 1 correspond to the population groups below.

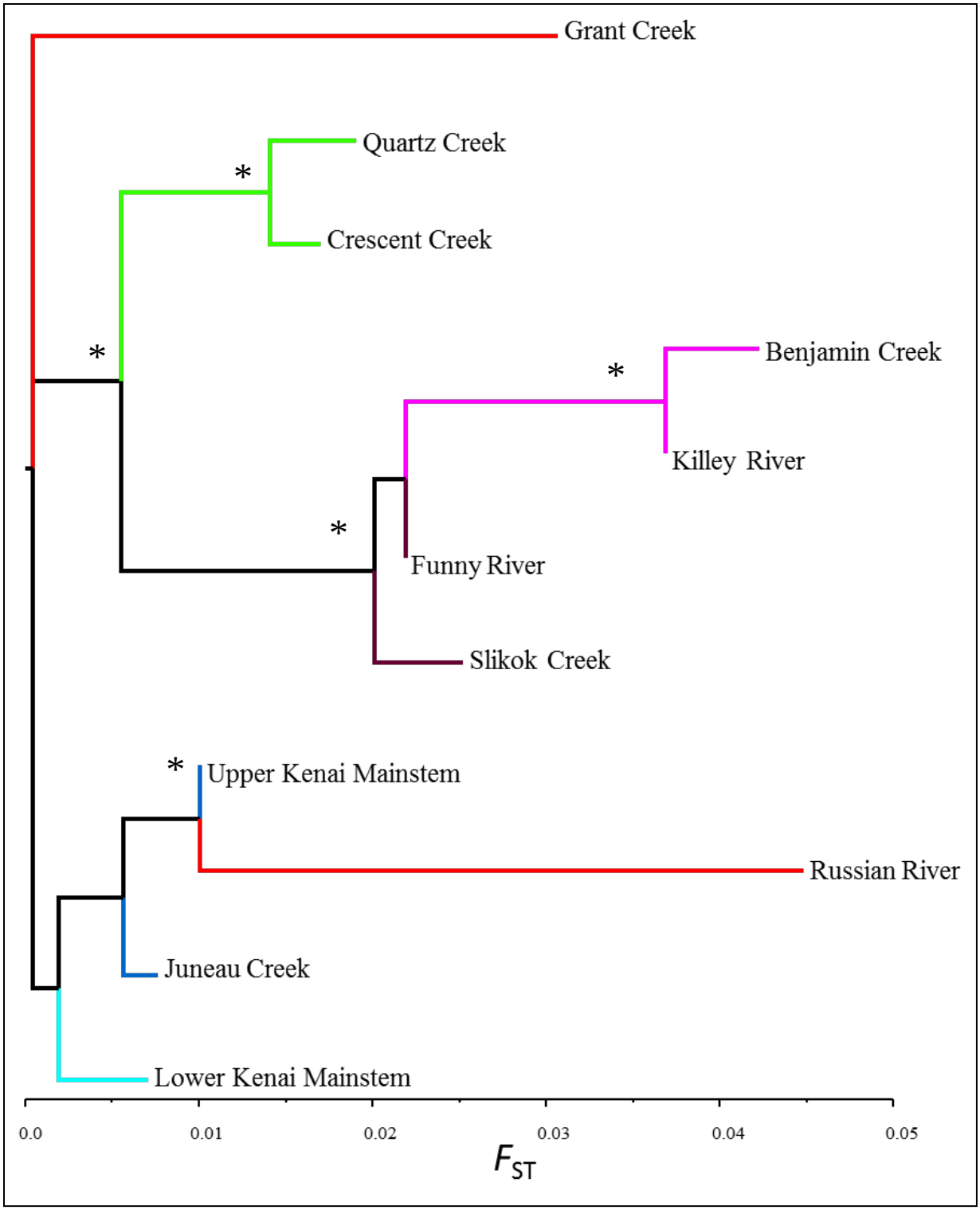
| Population groups | DF | G | G' | P |
|---|---|---|---|---|
| Total Kenai | 380 | 4962 | 13.06 | <0.001 |
| Between Regions | 38 | 828 | 21.80 | <0.001 |
| Within Regions | 342 | 4133 | 12.09 | <0.001 |
| Kenai Mainstem | 76 | 464 | 6.10 | <0.001 |
| Among Group | 38 | 353 | 9.28 | <0.001 |
| Within Group | 38 | 111 | 2.92 | <0.001 |
| Upper Mainstem | 38 | 111 | 2.92 | <0.001 |
| Lower Mainstem | --- | --- | --- | --- |
| Kenai Tributaries | 266 | 3670 | 13.80 | <0.001 |
| Among Group | 38 | 1556 | 40.95 | <0.001 |
| Within Group | 228 | 2114 | 9.27 | <0.001 |
| Upper | 114 | 1393 | 12.22 | <0.001 |
| Among Tributaries | 38 | 957 | 25.19 | <0.001 |
| Within Tributaries | 76 | 436 | 5.73 | <0.001 |
| Quartz/Crescent/Grant | 76 | 436 | 5.73 | <0.001 |
| Russian | --- | --- | --- | --- |
| Lower | 114 | 721 | 6.32 | <0.001 |
| Among Tributaries | 38 | 500 | 13.15 | <0.001 |
| Within Tributaries | 76 | 221 | 2.91 | <0.001 |
| Funny/Slikok | 38 | 110 | 2.89 | <0.001 |
| Benjamin/Killey | 38 | 111 | 2.92 | <0.001 |

# FIGURES




*Note:* Thick lines below Skilak Lake correspond to the lower mainstem sampling areas and the dark lines between Kenai Lake outlet and the inlet to Skilak Lake corresponds to the upper mainstem sampling areas.

Figure 1.–Sampling locations and reporting group designation for Chinook salmon originating from Kenai River drainage, Alaska. Numbers correspond to map numbers on Table 1. The broad-scale reporting groups are represented by circles and asterisks (tributaries) and squares (mainstem). The mid-scale reporting groups are represented by circles (lower tributaries), asterisks (upper tributaries), and squares (mainstem). The fine-scale reporting groups are represented by the 6 symbols in the legend. The thick lines indicate the sections of the mainstem where sampling took place.




*Note:* Colors denote reporting groups as in Figures 3a, 4, and 6. 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 2.–Consensus neighbor-joining tree based on $F_{ST}$ between Chinook salmon populations sampled from mainstem spawning areas and spawning drainages of the Kenai River, Alaska (see Table 1 for collection details).

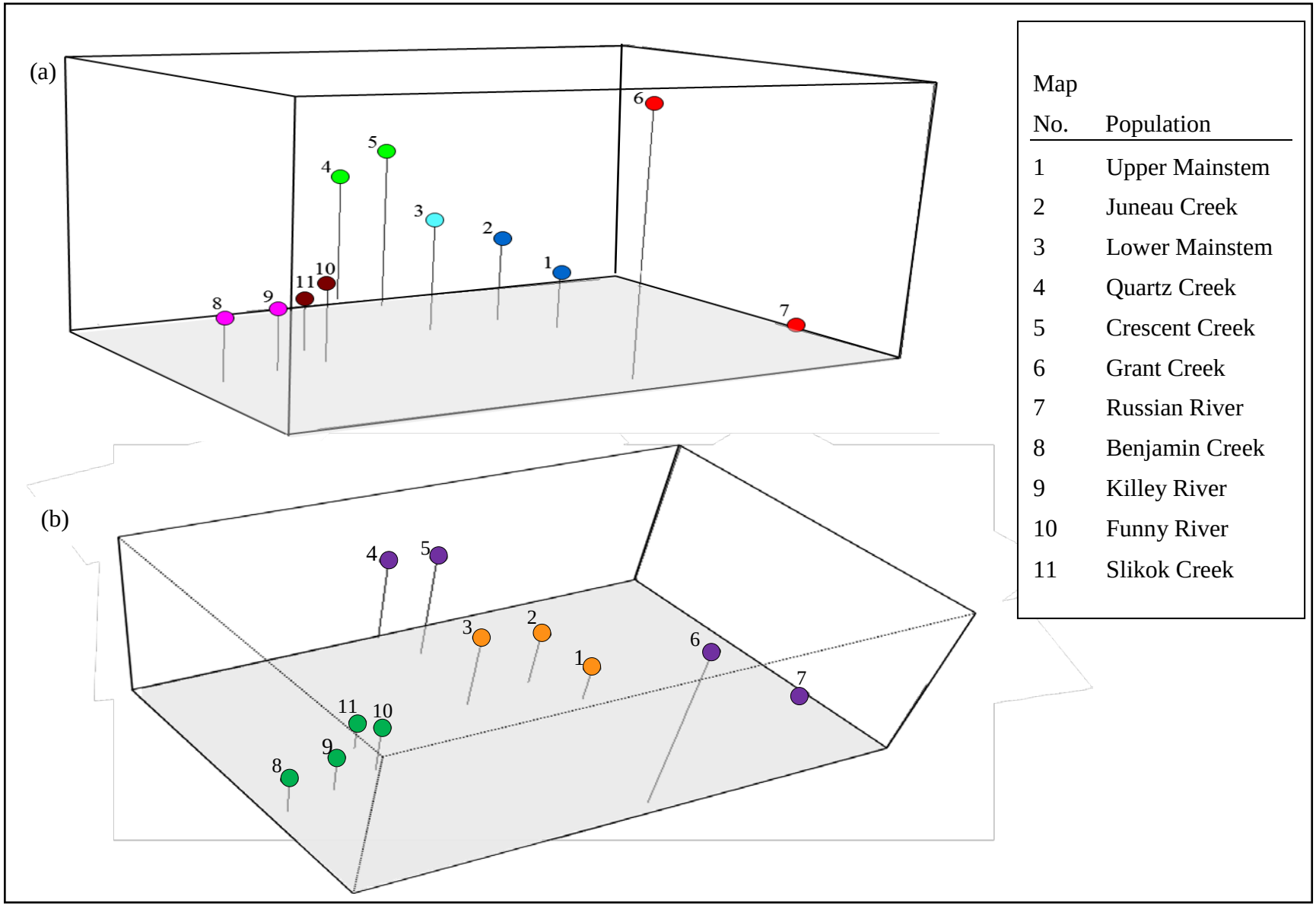



Figure 3.–Multidimensional scaling plots based on Cavalli-Sforza Edwards genetic distances between samples of Chinook salmon from Kenai River, Alaska. The colors in (a) correspond to those in Figures 1 and 2 and colors in (b) represent the Kenai River drainage as 3 reporting groups: Upper Kenai River tributaries (purple), Kenai River mainstem (orange) and Lower Kenai River tributaries (green). The multidimensional scaling plot in (b) has been re-oriented to better represent the clustering differences between the mainstem and tributary populations.

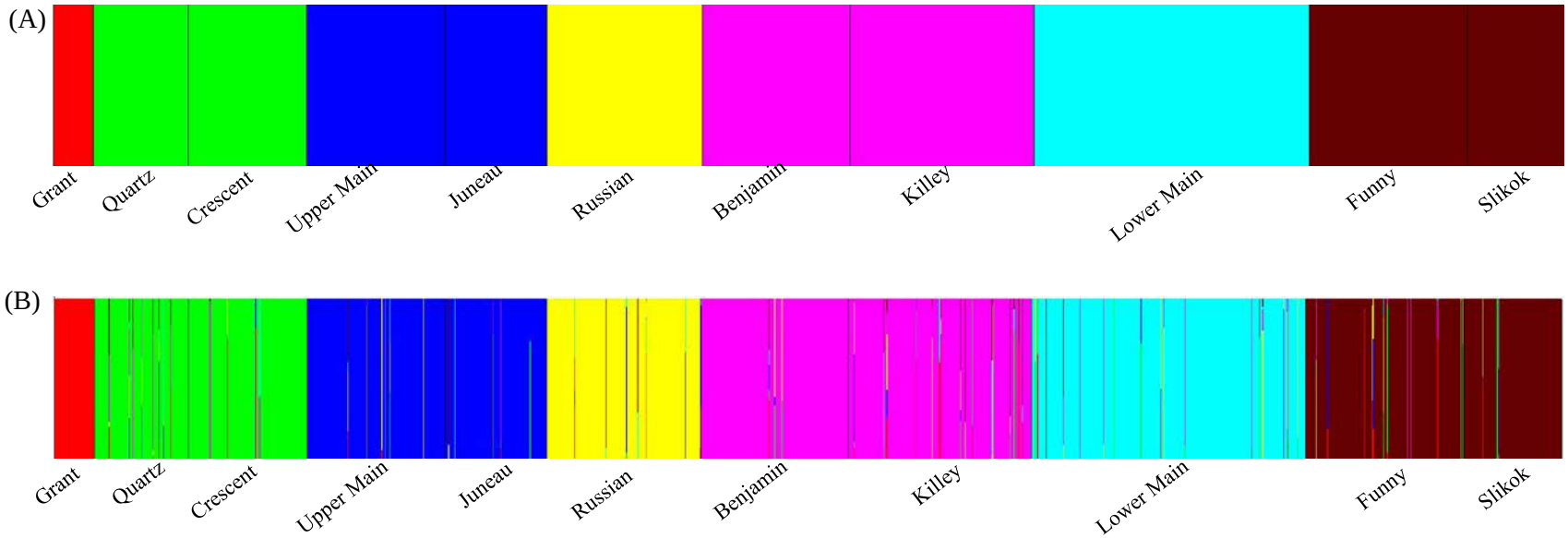



Figure 4.–Genetic assignment results based on Bayesian methods in the program BAPS under the population mixture analysis of clustering of groups of individuals (A) and population admixture (B) where vertical bars with several colors represent admixed individuals.

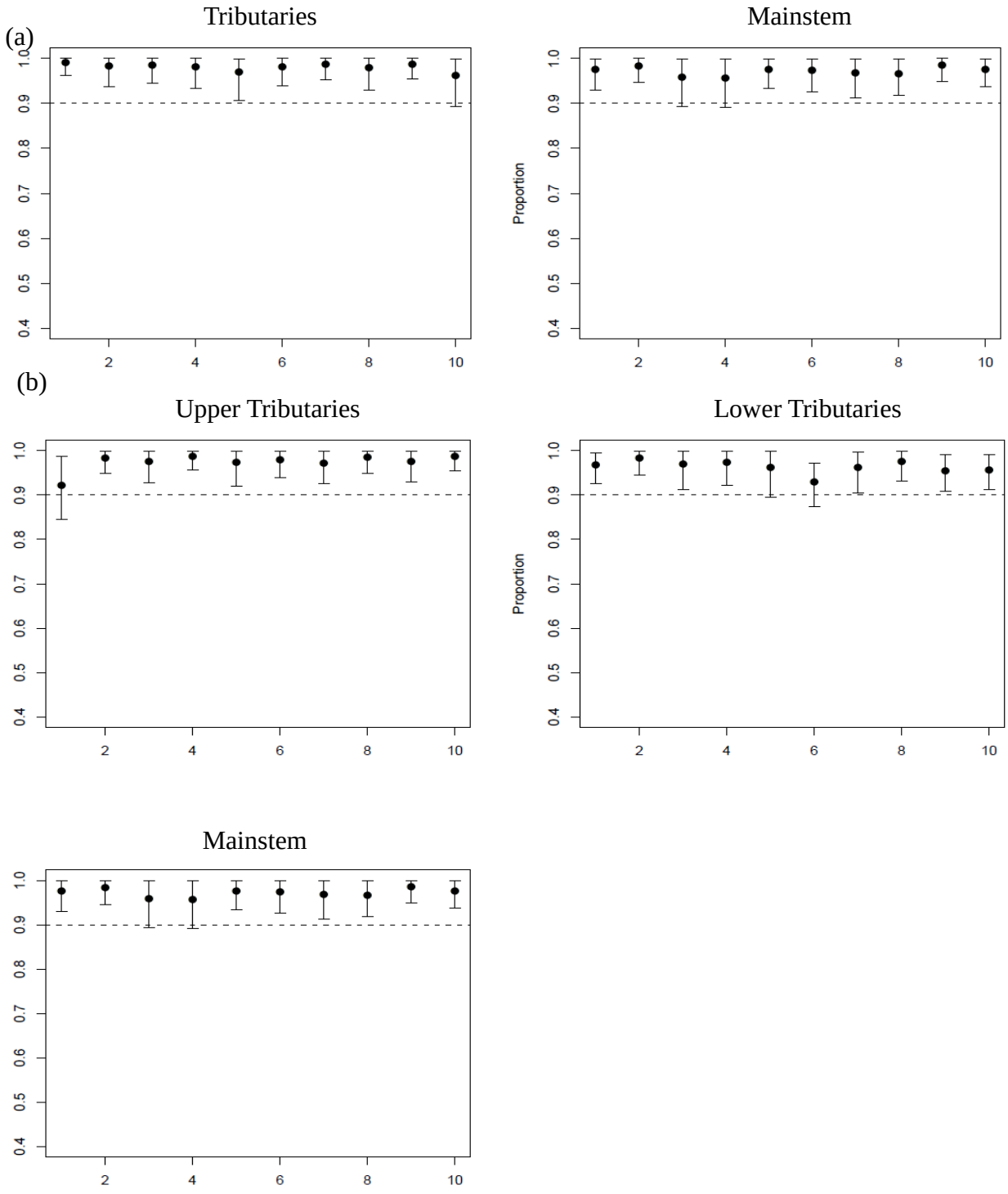



Figure 5.–Results of repeated proof testing for 2 reporting groups (a) and 3 reporting groups (b) including 95% credibility intervals. The Upper Tributaries reporting group contains the collections of Grant Creek, Crescent Creek, Quartz Creek, and Russian River. The Lower Tributaries reporting group contains the populations of Killey River, Benjamin Creek, Funny River, and Slikok River.

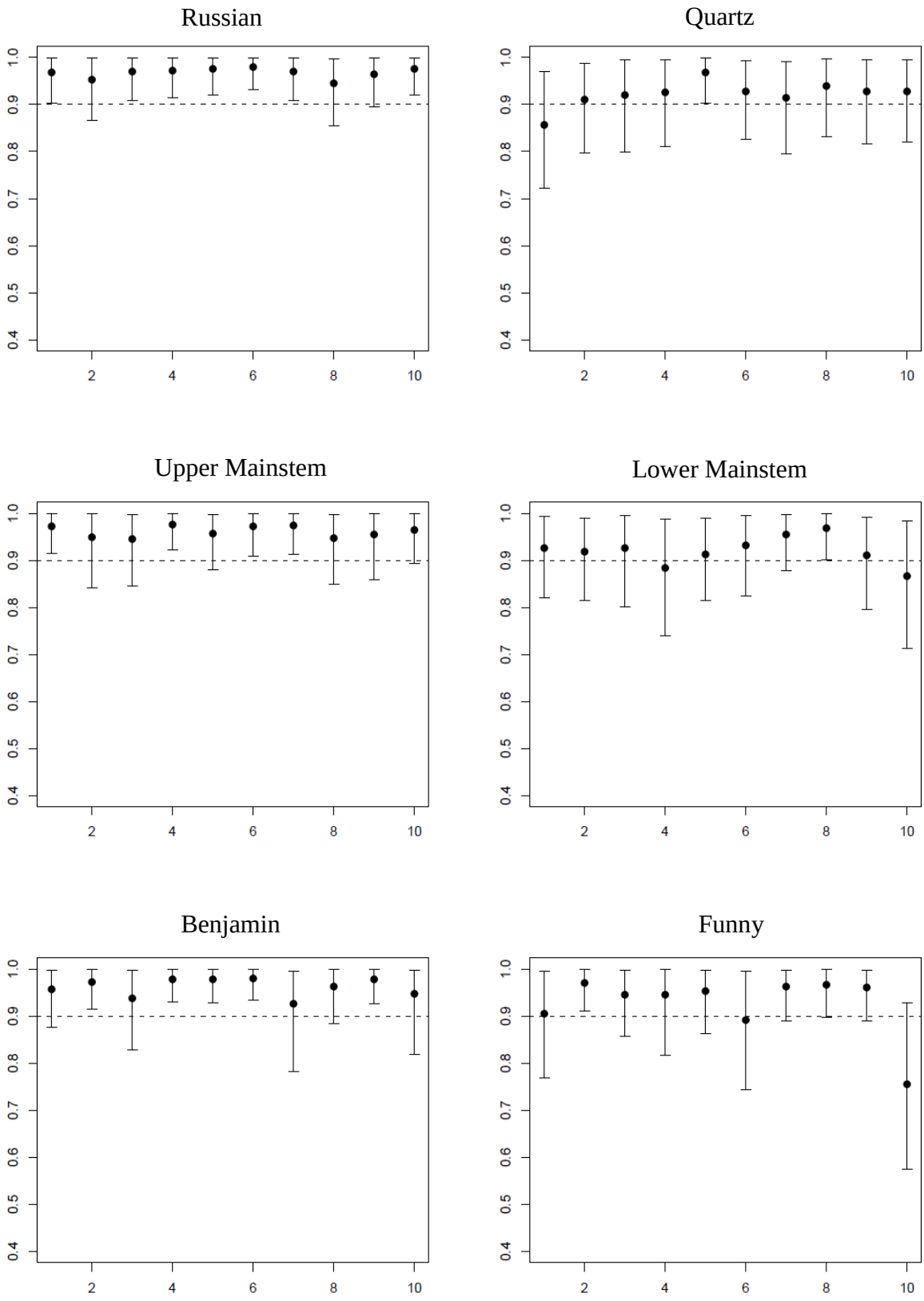



Figure 6.–Results of repeated proof testing for 6 reporting groups. The points represent each draw with 95% credibility intervals for each point.