APPLICATION OF MOLECULAR MARKERS TO MIXED-STOCK ANALYSIS OF YUKON RIVER FALL CHUM SALMON

By

Blair G. Flannery

RECOMMENDED:

Advisory Committee Chair

Director, Fisheries Division

APPROVED:

Dean, School of Fisheries and Ocean Sciences

Dean of the Graduate School

April 7, 2004

Date

APPLICATION OF MOLECULAR MARKERS TO MIXED-STOCK ANALYSIS OF YUKON RIVER FALL CHUM SALMON

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Blair G. Flannery, B.S.

Fairbanks, Alaska

May 2004

Abstract

Country of origin provides the basis for allocating harvests of Yukon River chum salmon. The genetic divergence among Yukon River chum salmon populations adjacent to the international border as revealed by allozyme and microsatellite variation is insufficient to determine the country of origin of returning fish using mixed-stock analysis (MSA). Consequently, we investigated the resolution provided by alternative genetic markers in an attempt to detect levels of divergence that would be sufficient for MSA. We analyzed 10 Yukon River chum salmon populations for variation at 30 variable amplified fragment length polymorphism (AFLP) loci and for mitochondrial DNA (mtDNA) restriction site variation. We assessed these markers for their utility in MSA and, for mtDNA, phylogeographic analysis. The AFLP results show that MSA was most successful when mixtures were allocated to regions. The AFLP data were able to provide improved country of origin MSA estimates for the border populations with a 6.5% improvement for the Canadian populations over microsatellite analysis. No divergence in mtDNA haplotype frequency distributions was detected ($P>0.05$) within the Yukon River. Lack of mtDNA divergence likely resulted from a Pleistocene bottleneck that led to panmixia of the mtDNA genome.

Table of Contents

List of Figures

Chapter 1

Chapter 2

List of Tables

Chapter 1

Chapter 2

List of Appendices

Chapter 1

Chapter 2

Preface

Manuscripts based on this research have been submitted for publication in the journal *Transactions of the American Fisheries Society* (published by the American Fisheries Society, Bethesda, MD). The authorship on the papers is Blair G. Flannery, John K. Wenburg and Anthony J. Gharrett. Blair G. Flannery designed the experiments, conducted all laboratory and data analyses, and wrote the manuscript. Anthony J. Gharrett assisted in the design of the experiments and provided overall project management. John K. Wenburg assisted with the design of the experiments, project management, and procurement of funding. This thesis would not have been possible without the support of Steve Miller, William Spearman, Steve Klein, Dick Pospahala, Steve Klosiewski, Richard Hannon, LaVerne Smith, William Smoker, Andy Gray, and Richard Wilmot. I thank my wife Andrea Medeiros for formatting this thesis for publication.

The Conservation Genetics Laboratory, U.S. Fish and Wildlife Service provided the funding to support this research with a grant (FWS 70181-9-N059; USGS BRD 1434-HQ-99-RM-0016) administered by the Alaska Cooperative Fish and Wildlife Research Unit, Biological Resource Division, U.S. Geological Survey, University of Alaska Fairbanks.

General Introduction

The United States and Canada entered into negotiations concerning salmon (*Oncorhynchus* spp.) in the 1970's because of disputes over the allocation of the salmon resource. In 1985, the two countries signed the Pacific Salmon Treaty, which solidified the obligation these countries have to manage the Yukon River salmon resources and pertinent Pacific coast stocks. Under Article III of the treaty, the countries agreed to conserve salmon and to maintain maximum production. They also agreed that each country should receive the benefits of salmon produced in its own rivers.

Chum salmon (*O. keta*) of the Yukon River, a major transboundary drainage, are of particular interest and concern. In March of 2001, the countries reached an agreement that the United States would reduce its catch of Canadian origin fall chum salmon by 10 percent. Identification of the relative proportion of the chum salmon resource for these countries would simplify allocation and management.

The Yukon River flows more than 2000 miles through Canada and Alaska. Chum salmon spawn throughout the drainage in both countries and chum salmon are known, in general, to exhibit fidelity to natal spawning grounds (Salo 1991). There are two distinct runs of chum salmon: summer and fall. Summer chum salmon enter the river between early June and early July and spawn primarily in tributaries of the lower 500 miles of the Yukon River. Fall chum enter the river from mid-July to late August and spawn from the Tanana River confluence to the headwaters of the Yukon River (Buklis 1981). Chum salmon are an important resource for the subsistence fisheries in the United States and Canada as well as to the wildlife that depend upon them for food.

Management of Pacific salmon fisheries presumes that the appropriate harvest rate of each population is a function of their sizes and reproductive capacities. Consequently, appropriate management requires knowledge of the species' population structure and the ability to control the harvest of individual populations (Allendorf et al. 1987). Genetic analysis of a species can provide information about population structure. Using this information as a baseline, the composition of stock mixtures in harvests can be estimated

using a process called mixed-stock analysis (MSA) (Pella and Milner 1987). Previous genetic analyses of Yukon River chum salmon reported low resolution among populations adjacent to the international border. The similarities of the border populations resulted in MSA estimates for country of origin that had low accuracy and precision (Beacham et al. 1988; Wilmot et al. 1992; Scribner et al. 1998). The purpose of this research is to apply new molecular genetic techniques in an attempt to improve Yukon River chum salmon MSA estimates by country of origin. This thesis reports the results of the application of amplified fragment length polymorphism (AFLP) and mitochondrial DNA restriction site analysis to MSA of Yukon River fall chum salmon.

References

Allendorf, F.W., N. Ryman, and F.M. Utter. 1987. Genetics and Fishery Management. pp. 1-19. *In*: N. Ryman and F. Utter (ed.) Population Genetics and Fishery Management. University of Washington Press, Seattle, Washington.

Beacham, T.D., C.B. Murray, and R.E. Withler. 1988. Age, morphology, developmental biology, and biochemical genetic variation of Yukon River fall chum salmon, *Oncorhynchus keta*, and comparisons with British Columbia populations. Fishery Bulletin 86:663-674.

Buklis, L.S. 1981. Yukon and Tanana River fall chum tagging studies, 1976-1980. Alaska Department of Fish and Game Informational Leaflet 194. 85 pp.

Pella J.J. and G.B. Milner. 1987.Use of genetic marks. pp.247-276. *In*: N. Ryman and F. Utter (ed.) Population Genetics and Fishery Management. University of Washington Press, Seattle, Washington.

Salo, E.O. 1991. Life history of chum salmon. pp. 233-309. *In*: Groot, C. and L. Margolis (ed.) Pacific Salmon Life Histories. UBC Press, Vancouver, B.C.

Scribner, K.T., P.A. Crane, W.J. Spearman, and L.W. Seeb. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 55: 1748-1758.

Wilmot, R.L., R.J. Everett, W.J. Spearman, and R. Baccus. 1992. Genetic stock identification of Yukon River chum and chinook salmon – 1987-1990. Progress Report U.S. Fish and Wildlife Service, Anchorage, AK. 132 pp.

CHAPTER 1

Mitochondrial DNA variation within and among Yukon River chum salmon populations

Abstract

We surveyed 97% of the mtDNA genome for restriction site variation in Yukon River chum salmon to investigate the evolution of genetic variability in this large system, and to evaluate its potential for mixed-stock analysis (MSA). Initial exhaustive surveys of two populations identified 16 variable sites. We then surveyed 10 Yukon River populations, spanning over 2000 river miles, and four populations from other Alaskan locations for variation at these sites. Restriction site variation revealed 16 composite haplotypes; no divergence in mtDNA haplotype frequency distributions was observed among Yukon River populations. However, a haplotype tree showed two lineages, and nested clade analysis revealed significant relationships between the geographical distribution of haplotypes and their genealogy for a two-step clade that dominates in the Yukon River and the total cladogram. The demographic signal was consistent with isolation by distance. The absence of historic fragmentation and/or range expansion signals suggests that incomplete lineage sorting is responsible for paraphyly within the Yukon River. Mismatch analysis revealed that either an ancestral Yukon River population underwent a post-Pleistocene expansion followed by subdivision or that expansion and subdivision coincided. Although mtDNA does not appear useful for MSA of Yukon River chum salmon, it is useful for separating population history from contemporary processes and aids our understanding of Yukon River chum salmon evolution.

Introduction

Gender based asymmetric dispersal and gene flow make investigating mitochondrial DNA (mtDNA) essential when assessing population structure and phylogeography. In many species, females are philopatric whereas males disperse more widely; such situations may lead to panmixia of the nuclear genome among locales, but to subdivision in the mtDNA genome. This has serious management implications because demographics are female dependent, thus reestablishment from other locales in the event of extirpation would be improbable (Avise 2000).

Mitochondrial DNA is haploid, maternally inherited, does not undergo heterologous recombination, has a high mutation rate (5-10 times greater on average than nuclear genes), and has an effective population size (N_e) that is one-fourth of the nuclear genome (Gyllensten et al. 1985b; Brown et al. 1979). These attributes make mtDNA sensitive to founder events and bottlenecks and potentially more likely to reveal divergence among populations with low to moderate $N_{e(female)}$'s (Ferguson and Danzmann 1998). Gene genealogies can be subjected to nested clade analysis (NCA), which correlates geographic distribution and genealogical relationships, and pairwise site differences can be subjected to mismatch analysis; together they can provide insight into the demographic history underlying population structure (Templeton 1998, Rogers and Harpending 1992).

Low levels of genetic divergence have been observed among Yukon River chum salmon populations, especially those near the United States/Canada border (Beacham et al. 1988, Wilmot et al. 1992, Scribner et al. 1998), which creates problems for allocation among user groups. Previous genetic studies focused primarily on allozyme and microsatellite variation. The only previously published mtDNA study of Yukon River chum salmon that we are aware of evaluated only the variation in the ND5/ND6 region of mtDNA using seven restriction endonucleases (Scribner et al. 1998). Because there may be little evolutionary concordance among mtDNA regions within and among salmonid species (Churikov et al. 2001), the full potential of mtDNA variability has not yet been explored in Yukon River chum salmon.

The objectives of this study were to analyze mtDNA restriction site variability in Yukon River populations to aid our understanding of their evolution, and to evaluate the potential of mtDNA variation to resolve divergence among populations. The specific questions addressed were: (1) what are the most informative mtDNA regions for Yukon River chum salmon; (2) does the observed variation provide a means for distinguishing among chum salmon populations, especially those adjacent to the international border, within the Yukon River; and (3) what demographic and evolutionary processes influence the current population structure?

Methods

Sample collection

Tissue samples were collected from 10 chum salmon populations within the Yukon River and from four other Alaskan locations (Figure 1.1). Eight of the Yukon River populations represent major fall chum salmon producers, and two summer populations were included (Chulinak and South Fork Koyukuk) to represent summer chum salmon caught in mixed fisheries. Populations outside the Yukon River provided an indication of the total scope of mtDNA variation among Alaskan chum salmon populations.

DNA preparation and analysis

Total genomic DNA was extracted from heart or fin tissue (~25mg) by proteinase K and the Puregene™ DNA isolation kit (Gentra Systems Inc. Minneapolis, MN). Concentrations of DNA were determined by fluorometry.

Specific primers were used to amplify seven regions [12S/16S (annealing temperature --T_A = 50C), ND1/ND2 (T_A = 50C), COI/A8 (T_A = 52C), A8/ND3 (T_A = 50-52C), ND3/ND4 (T_A = 50-52C), ND5/ND6 (T_A = 49C), and CytB/Dloop (T_A = 50C)] by polymerase chain reaction (PCR), which included 97% of the mitochondrial genome (Gharrett et al. 2001). PCR-amplification was carried out with Stratagene Robocyclers in 50 µl volumes under the following conditions: 2.5 mM $MgCl_2$, 1X PCR buffer (20 mM Tris-HCl pH 8.0, 50 mM KCl), 200 µM for each dNTP, 0.2 µM for each primer, 1 unit of

Taq polymerase (Promega Madison, WI), and 5 µl of DNA (~ 250 ng). For the A8/ND3 and ND3/ND4 regions, 225 µM for each dNTP and 0.25 µM for each primer were used. Thermal cycling began with an initial cycle of: 95C for three minutes; followed by 32 cycles of: 95C for 1 min, 49-52C for 1 min (specific T_A's are listed above), and 72C for 2 min 30 sec; and completed with one final extension cycle of 72C for 5 minutes. The A8/ ND3 and ND3/ND4 regions were started at 50C for 5 cycles and then continued at 52C for the remaining 27 cycles. Each region was cut with 14 restriction endonucleases (*Ase* I, *Ban* II, *Bst*U I, *Dde* I, *Hha* I, *Hin*f I, *Mbo* I, *Msp* I, *Nla* IV, *Rsa* I, *Sau*96 I, *Scr*F I, *Sty* I, *Taq* I). Reactions were carried out in 15 µl volumes with 1X restriction buffer (provided by the manufacturer), 5 units of restriction enzyme, and 5 µl of PCR product. Fragments were separated by size in a 2.5% agarose/synergel matrix in 0.5X TBE (TBE is 0.045M Tris-borate and 0.001M EDTA pH 8). DNA in the gels was stained by ethidium bromide (50µg/ml) for 30 minutes and destained for 20 minutes by electrophoresis. A photograph was taken of the gels using an ultraviolet transilluminator (312 nm). Fragments smaller than 200 bp were resolved by electrophoresis through 8% polyacrylamide gels (29:1 acrylamide:bisacrylamide) in 2X TAE (TAE is 40 mM Tris-acetic acid and 1 mM EDTA, pH 8.0). Syber Green I™ (1:10,000 dilution; Molecular Probes, Eugene OR) was used to stain the DNA for 15 minutes. A photograph was taken of the gel on a transilluminator. Both 100 and 25 bp ladders were used to estimate fragment lengths. The software program ProRFLP™ (DNA ProScan Inc., Nashville, TN) was used to estimate the size of the bands.

Population surveys

Restriction site variation in mtDNA was surveyed: (1) to find the most informative sites both diagnostically and evolutionarily in Yukon River chum salmon (Churikov et al. 2001); (2) to identify a subset of the possible mtDNA region/enzyme combinations that could be run in a reasonable amount of time for a large MSA study; and (3) to assess the discriminatory powers of this technique for population identification.

All seven regions of the mtDNA genome were amplified from ten samples each from Big Creek and Chulinak River, which were chosen because they are separated by

nearly the full length of the Yukon River. Each amplified mtDNA region was subjected to restriction digest analyses using the suite of 14 restriction endonucleases. Subsequently, 10 samples each from Kobuk River, Joshua Green River, Port Dick Creek and an additional 10 samples from Chulinak River were analyzed using only the identified variable region-by-enzyme combinations (ND5/ND5 – *Ase* I, *Dde* I, *Cfo* I, *Rsa* I, *Sau*96 I; ND3/ND4 – *Dde* I, *Hinf* I, *Taq* I; A8/ND3 – *Ase* I, *Dde* I, *Nla* IV; COI/A8 – *Mbo* I, *Msp* I; 12S/16S – *Msp* I; CytB/Dloop – *Dde* I). Finally, 10 samples each from all the Yukon River populations and 10 additional samples for Chulinak and Big Creek were analyzed for RFLP variation using combinations that revealed major subdivisions in the gene genealogy (ND5/ND6 *Rsa* I; COI/A8 – *Mbo* I, *Msp* I).

Statistical analyses

Restriction sites were inferred for each enzyme from mtDNA fragment patterns that differed by a single site. Composite haplotypes based on all region-by-enzyme combinations were developed for each fish with the GENERATE program in REAP (McElroy et al. 1992). The program GROUP in REAP eliminated redundant haplotypes. The program D in REAP calculated nucleotide divergence among haplotypes. The program DA in REAP calculated haplotype and nucleotide diversity within populations and average net nucleotide divergence among populations. The program REDUCE in REAP eliminated monomorphic sites.

The binary codes produced by GENERATE and REDUCE in REAP were used to calculate the pairwise site differences among haplotypes, which facilitated the construction of a minimum-spanning tree showing the most parsimonious gene genealogy. The program PARSPROB 1.13 (http://bioag.byu.edu/zoology/crandall_lab/programs.htm) was used to determine if the haplotypes were parsimoniously connected at the 0.95 level.

Neighbor-joining trees (Saitou and Nei 1987) were constructed from the matrix of estimates of population net nucleotide divergence (from DA in REAP) using PHYLIP (Felsenstein 1995).

Tests of homogeneity of haplotype frequencies among the populations were conducted with the MONTE program in REAP (10,000 iterations). Where significant heterogeneity was observed, the haplotypic variation was partitioned hierarchically by analysis of molecular variance (AMOVA; Excoffier et al. 1992), based on the pairwise site differences among the haplotypes, with the significance of the φ-statistics determined by 17,000 randomizations using Arlequin 2.0 (Schneider et al. 2000). The population structure tested included 7 regions: 1) United States Yukon River summer-run, 2) United States Yukon River fall-run, 3) Canada Yukon River, 4) Northwest Alaska, 5) Aleutian Peninsula, 6) South Central Alaska, and 7) Southeast Alaska.

Nested clade analyses were conducted using GEODIS 2.0 (Templeton 1998, Posada et al. 2000); significance was determined from 17,000 randomizations to detect significant associations among geographic locations and haplotypes. Causes of significant associations were inferred from the key provided by Templeton (1998). To assist in rooting the cladogram, chum salmon data from Churikov et al. (2001) were combined with data from this study. These additional data were not included in the nested clade analyses because they will be reported elsewhere, and were used only to root the cladogram. A minimum-spanning tree of the combined data was constructed based on the variable region/enzyme combinations observed, which facilitated the development of a nested clade design. Ambiguities in the topology were resolved (Templeton et al. 1987, 1992), and outgroup weights were calculated for the haplotypes using a heuristic algorithm (Castelloe and Templeton 1994).

The Yukon River populations were tested for sudden population expansion with mismatch and intermatch distribution analyses (Rogers and Harpending 1992, Rogers and Jorde 1995) using Arlequin 2.0 with significance and 95% confidence intervals for the estimated demographic parameters determined by 1000 parametric bootstrap simulations.

Results

Comprehensive restriction site analysis of two Yukon River populations (Big Cr. and Chulinak) revealed seven haplotypes and surveyed a total of 432 sites of which 16 were

variable. Initially, 20 variable sites were identified but four were removed due to possible duplication of recognition sites between *Msp* I and *Nla* IV/*Scr*F I and overlaps between amplified regions. Variation was observed in all regions except for the ND1/ND2 region (Table 1.1). The ND5/ND6 region had the most variation with five variable sites that resolved four haplotypes and a nucleotide diversity, calculated using all monomorphic and polymorphic sites, of 1.52 substitutions per 1000 base pairs (bp). Between 8.1 and 12.9% of the nucleotides in each mtDNA region were surveyed. Overall, 1771.3 bp or 10.7% of the mitochondrial genome were examined, which revealed a haplotype diversity of 0.64 and a nucleotide diversity of 0.50 substitutions per 1000 bp. Haplotype distributions did not differ (P>0.05) between the two populations (Table 1.2). Composite haplotype frequencies ranged from 5 – 60%; five of the seven haplotypes were singletons. The haplotype genealogy revealed a star-like pattern with rarer haplotypes diverging radially from the more abundant central haplotype (Figure 1.2). Two mtDNA evolutionary lineages, E (haplotypes A-F) and G (haplotype G), previously identified by Churikov et al. (2001) were observed. Average nucleotide divergence between lineages, based on all monomorphic and polymorphic sites among haplotypes, was 1.96 substitutions per 1000 bp as compared to 1.04 substitutions per 1000 bp within lineages.

Results from the comprehensive survey were ambiguous in determining the best set of region-by-enzyme combinations to describe population structure; a total of 10 region-by-enzyme combinations showed variation, none of which could be eliminated without possibly omitting useful sites.

12S/16S	COI/COII	A8/COIII	ND3/ND4	ND5/ND6	Cytb/D-loop
Msp I	*Mbo* I	*Ase* I	*Dde* I	*Ase* I	*Dde* I
	Msp I	*Dde* I	*Hinf* I	*Dde* I	
		Nla IV	*Taq* I	*Cfo* I	
				Rsa I	
				*Sau*96 I	

These combinations were too numerous to run in a large MSA study, so an additional survey analyzing just these sites was performed to clarify the haplotype genealogy. Expanding the number of populations to include Kobuk, Joshua Green, and Port Dick Creek, but focusing only on the variable region/enzyme combinations, revealed eight new haplotypes for a total of 15 haplotypes (Table 1.3). A total of 87 restriction sites was detected of which 23 were variable. In this survey fewer enzymes were used. Consequently, a smaller portion of the genome was surveyed: a total of 356 nucleotides or 2.14% of the mitochondrial genome. The average number of restriction sites and nucleotides examined per haplotype was 72.94 and 297.87 (1.79% of mtDNA genome), respectively. Because the variable sites were the focus of the reduced survey, the magnitude of nucleotide diversity and divergence estimates increased, but the haplotype diversity remained about the same, as expected. Haplotype frequencies ranged from 1.6 – 55.7%; 11 out of 15 haplotypes were singletons. Haplotype frequency distributions did not differ ($P>0.05$) among the five populations (Table 1.3).

The haplotype genealogy revealed two star-like phylogenies consisting of the E and G lineages (Figure 1.2). The structure of the G lineage became clearer with the addition of haplotypes H, I, J, and O. The Kobuk River was the only population in which the G lineage was absent. Haplotypes F, K, L, M, and N were added to the E lineage. The NCA revealed significant associations ($P<0.05$) between the geographic distribution and the genealogical relationship among subclades within clades 2-1, and the total cladogram (Figure 1.3). The results identified the cause of the associations as restricted gene flow with isolation by distance (Figure 1.3, Table 1.5). Inclusion of the data from Churikov et al. (2001) improved the resolution of the gene genealogy, and allowed for a better estimation of the root than could be achieved by using data from this study alone (Figure 1.2). Haplotype W (C in Churikov et al. 2001) had the largest outgroup weight and nested in clade 3-2 (Table 1.6, Figure 1.3). Therefore, the root for the nested clade analysis of the present study was assigned to clade 3-2. Although one might argue that haplotype E is the root because of its high outgroup weight and multiplicity, the results of the NCA were unaffected by alternative rooting strategies.

The region-by-enzyme combinations that contributed most to defining the haplotype genealogy were analyzed to evaluate their potential for Yukon River chum salmon MSA (Figure 1.2: ND5/ND6 *Rsa* I; COI/COII *Mbo* I, *Msp* I). These sites distinguished the E, G, and C clades, which were the most abundant and divergent groups. A total of 17 restriction sites were surveyed with these combinations of regions and restriction enzymes; five variable sites revealed one new haplotype, P. Only 68 bp or 0.41% of the genome was covered with these enzymes. On average, 14.2 sites and 56.8 bp (0.34% of the genome) were surveyed. Haplotype frequencies ranged from 0.58% – 66.08%; haplotypes P and D were singletons. The haplotype genealogy displayed both E and G lineages with the majority (87%) belonging to the E lineage (Figure 1.2). Haplotype frequency distributions exhibited heterogeneity among the 14 populations ($P<0.01$, Table 1.4). However, no heterogeneity was observed among populations within the Yukon River. An AMOVA, based on the mutational differences among haplotypes, was used to partition the heterogeneity hierarchically into three levels of population structure: among regions (φ_{CT}), among populations within regions (φ_{SC}), and among populations (φ_{ST}). The regions were the United States Yukon River summer-run, United States Yukon River fall-run, Canada Yukon River, Northwest Alaska, Aleutian Peninsula, South Central Alaska, and Southeast Alaska. Overall AMOVA revealed that 88.31% of the variation occurred within and 11.69% was attributed to divergence among populations ($\varphi_{ST} = 0.1169$, $P<0.001$). The majority of the among population variation (11.62%) was accounted for by differences among regions ($\varphi_{CT} = 0.1162$, $P<0.05$). Differences among populations within regions did not account for detectable variation 0.07% ($\varphi_{SC} = 0.0007$, $P>0.05$). Neighbor-joining analysis clustered most of the Yukon River populations together (Figure 1.6). However, Chandalar, Teslin and Kluane were somewhat divergent.

Mismatch and intermatch distributions were unimodal and conformed to the sudden expansion model (Figure 1.5, Table 1.7). Mismatch and intermatch distributions for the more comprehensive mtDNA surveys showed the same patterns (results not shown). The intermatch distribution estimated τ, the time in generations since expansion measured in mutational units, at 0.97 (0-2.62). We did not attempt to translate τ into years

since expansion because most estimates of sequence divergence are based on interspecific comparisons, which can be quite different from intraspecific.

Discussion

Nucleotide diversity estimates differed among mtDNA regions, ranging from 0 to 1.52 substitutions per 1000 nucleotides. A similar pattern was observed in coho, pink, sockeye, and chum salmon (Gharrett et al. 2001, Churikov et al. 2001). All three studies confirmed that surveys of the entire genome are appropriate for estimating nucleotide diversities and divergences and choosing appropriate regions for surveying population divergence. The ND5/ND6 region was most informative because it had the highest nucleotide and haplotypic diversities. This contrasted with the results of Churikov et al. (2001) for chum salmon, and suggests that the mtDNA regions that are informative species wide may not be as informative in certain geographic regions. There was agreement that 12S/16S, ND1/ ND2, and CytB-Dloop were the least informative regions. Moreover, the two studies were in agreement that *Mbo* I and *Msp* I variation in the COI/COII region revealed the major genealogical break.

A star-like pattern was expected in the Yukon River chum salmon gene genealogy because of a presumed post-Pleistocene population expansion, but the large number of mutational steps separating the common haplotype (E) from the less common (D, F, G) was unexpected. It is unlikely that sufficient time has elapsed during the current interglacial interval to produce the divergence observed between the most distal tips of the haplotype tree. Large ancestral Yukon River population sizes that were not reduced during the last glacial maximum, secondary contact, or incomplete lineage sorting may explain our results (Avise et al. 1987). However, incomplete lineage sorting is the most likely cause of this variation as evidence points to a bottleneck in the history of Yukon River chum salmon and the absence of fragmentation/range expansion events does not support secondary contact. The high haplotypic diversity ($h = 0.64$) and low nucleotide diversity ($\pi = 0.05\%$) suggest that a bottleneck preceded rapid population growth that allowed for mutations to accumulate (Grant and Bowen 1998). This is further supported by the unimodal mismatch/intermatch distributions that show waves with only steep

faces. Such waves conform to the theoretical expected pattern of an expansion occurring at $\tau = 0$ (Rogers and Harpending 1992). Clearly, the expansion occurred in the recent past as measured in mutational time. While the wide confidence intervals of the estimated demographic parameters do not allow for much interpretation, the combination of the wave patterns and the estimate of the intermatch τ at 0.97 supports the hypothesis of a post-Pleistocene expansion. The occurrence in the gene genealogy of the E and G lineages was interesting because they may have sorted separately into the Pleistocene Beringian and Cascadian glacial refugia (Churikov et al. 2001, Figure 1.2). Ancestral populations that are isolated by geography or behavior are expected to undergo stochastic lineage sorting if there is polymorphism in the ancestral population (Avise et al. 1990). The populations would at first be polyphyletic for the separate mtDNA lineages, but eventually attain reciprocal monophyly after roughly 4 $N_{e(female)}$ generations (Avise et al. 1984). Potential reasons for the apparent paraphyly include secondary contact or insufficient time to produce monophyly. Again, incomplete lineage sorting is the probable cause for the apparent paraphyly as no evidence for historic fragmentation/range expansion events, required to support compete lineage sorting, was detected.

Examining ND5/ND6 *Rsa* I and COI/COII *Mbo* I and *Msp* I variation in all of the populations uncovered significant heterogeneity, but no heterogeneity was detected among populations within the Yukon River. The AMOVA revealed that populations were more divergent among regions than within regions and that ND5/ND6 *Rsa* I resolved population structure better than COI/COII *Mbo*I, and *Msp*I. Given the small sample sizes, these results were not surprising. However, two geographically distant populations, Chulinak and Big Creek had relatively larger sample sizes of 31 and 20, respectively, but did not differ.

Mitochondrial DNA restriction site analysis is a labor-intensive technique and requires definitive results to warrant its application for stock identification. Our results do not support application of mtDNA variation to MSA of Yukon River chum salmon. This conclusion is supported by a previous study of mtDNA variation in Yukon River chum salmon that surveyed *Rsa* I variation in ND5/ND6 in addition to other restriction

endonucleases (Scribner et al. 1998). In that study, variation in the ND5/ND6 region yielded inaccurate and imprecise MSA estimates because the data provided little resolution. These results suggest that future studies on MSA of Yukon River chum salmon should focus on the nuclear genome. The size of the nuclear genome and its ability to undergo recombination provide a nearly limitless number of possibilities for analysis. Haplotypic variation of mtDNA is, however, well suited for phylogenetic and phylogeographic analyses that may address questions about the contemporary structure. Such data may be useful for determining what combinations of contemporary gene flow, $N_{e(female)}$ and historical demographic processes are influencing population structure.

Limited mtDNA divergence in Yukon River chum salmon may have resulted from bottlenecks caused by Pleistocene glaciations. Glacial advances limited the habitat in which chum salmon could survive and likely resulted in the extirpation/reduction of populations causing the loss of mtDNA diversity. The nearly identical patterns of the mismatch and intermatch distributions (Figure 1.6) suggest that an ancestral Yukon River population expanded and then subdivided, or that expansion and subdivision coincided in the current interglacial period (Rogers and Jorde 1995). This signifies that glacial advances caused a severe bottleneck, which resulted in a panmictic mtDNA genome for Yukon River chum salmon. The lack of mtDNA divergence in Yukon River chum salmon is consistent with observations that northern species exhibit little mtDNA divergence because of glaciations as compared to species from unglaciated areas (Billington 2003). Mitochondrial DNA should be more affected by bottlenecks than nuclear DNA because of its smaller N_e. Furthermore, because mtDNA does not undergo heterologous recombination, mtDNA behaves as a single locus, as compared to the multilocus (multivariate) variation available at nuclear loci. Consequently, population bottlenecks can reduce mtDNA haplotypic variation in a population and limit the apparent divergence among populations whereas nuclear DNA is not affected as dramatically (Ferris and Berg 1987).

Augmenting this dearth of mtDNA divergence among Yukon River chum salmon populations is recurrent contemporary gene flow, indicated by NCA. It has been

suggested that chum salmon have high straying rates, although little direct experimental evidence exists (Johnson et al. 1997). The research that has been done on straying of chum salmon in other geographical regions estimated rates from 2.5 to 42.0% (Quinn 1993, Tallman and Healey 1994). Even low rates of gene flow can arrest divergence among populations (Kimura and Weiss 1964).

Although the application of mtDNA to MSA is not promising it has shed light on Yukon River chum salmon life history. The NCA and mismatch analyses of Yukon River chum salmon mtDNA revealed that a combination of historical and contemporary events has shaped its population structure. Such evidence explains why so little divergence has been observed among the populations when the rate of mtDNA evolution and mode of inheritance predict otherwise. The evidence of unbiased gene flow and mismatch/ intermatch distributions indicative of expansion and subdivision provides some assurance that reestablishment of extirpated populations is probable, although not in the lifetime of a fisheries manager. The utility of mtDNA, in combination with new statistical tests, to answer pertinent management and evolutionary questions reinforces the Avise et al. (1987) assertion that "mtDNA is not just another molecular marker."

Acknowledgments

This study was funded by the Conservation Genetics Laboratory, U.S. Fish and Wildlife Service through a grant (FWS 70181-9-N059; USGS BRD 1434-HQ-99-RM-0016) administered by the Alaska Cooperative Fish and Wildlife Research Unit, Biological Resource Division, U.S. Geological Survey, University of Alaska Fairbanks. We thank Steve J. Miller, William J. Spearman, Steve Klein, Steve Klosiewski, Richard Hannan, LaVerne Smith, and Richard Pospahala from the U.S. Fish and Wildlife Service; William W. Smoker from the University of Alaska Fairbanks; Richard L. Wilmot and Andrew K. Gray from the National Marine Fisheries Service for their support of this project. We thank Ted Otis and Fred Bue of the Alaska Department of Fish and Game for providing samples from Port Dick Creek and Kobuk River, respectively.

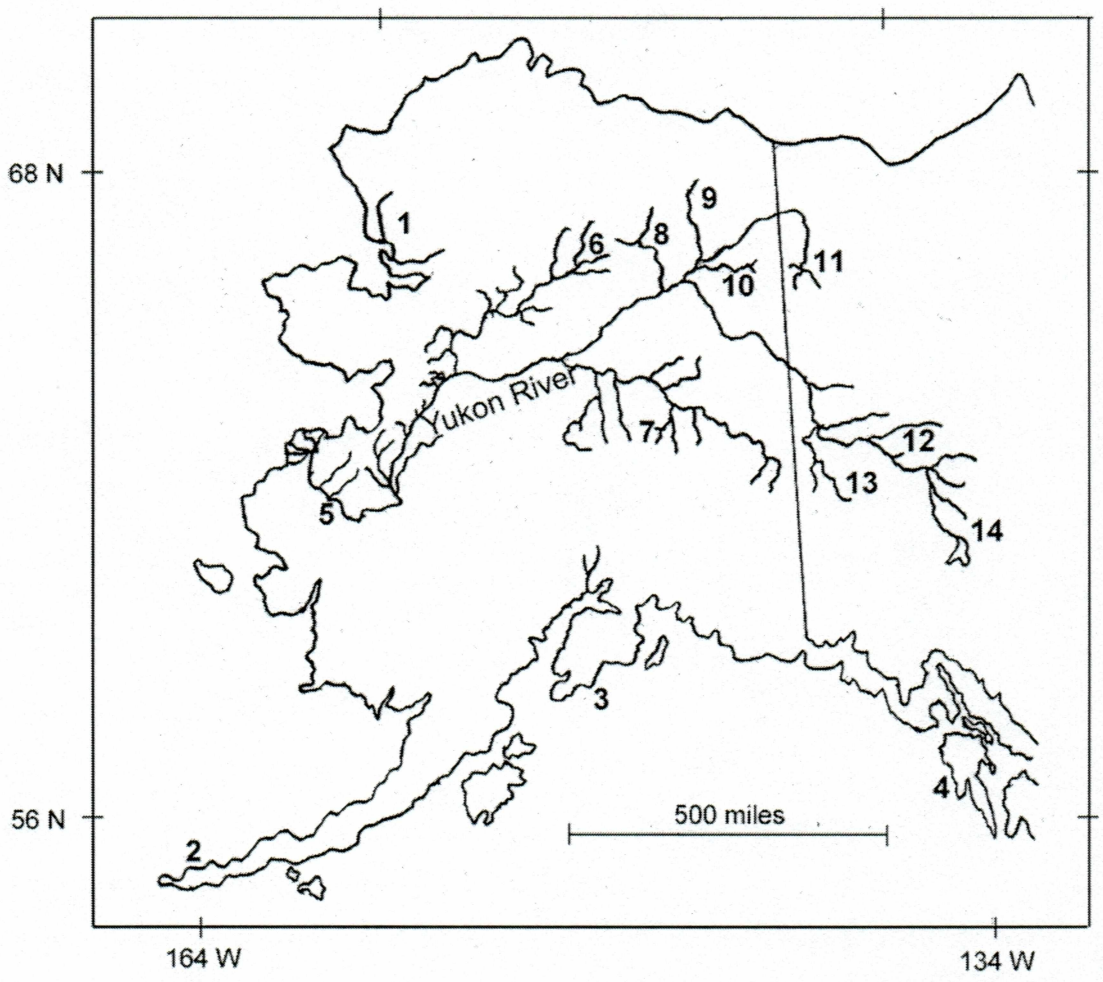

Figure 1.1. Sampling locations: 1 = Kobuk, 2 = Joshua Green, 3 = Port Dick Cr., 4 = W. Crawfish Inlet, 5 = Chulinak, 6 = South Fork Koyukuk, 7 = Delta, 8 = Chandalar, 9 = Sheenjek, 10 = Black, 11 = Fishing Branch, 12 = Big Cr., 13 = Kluane, 14 = Teslin.

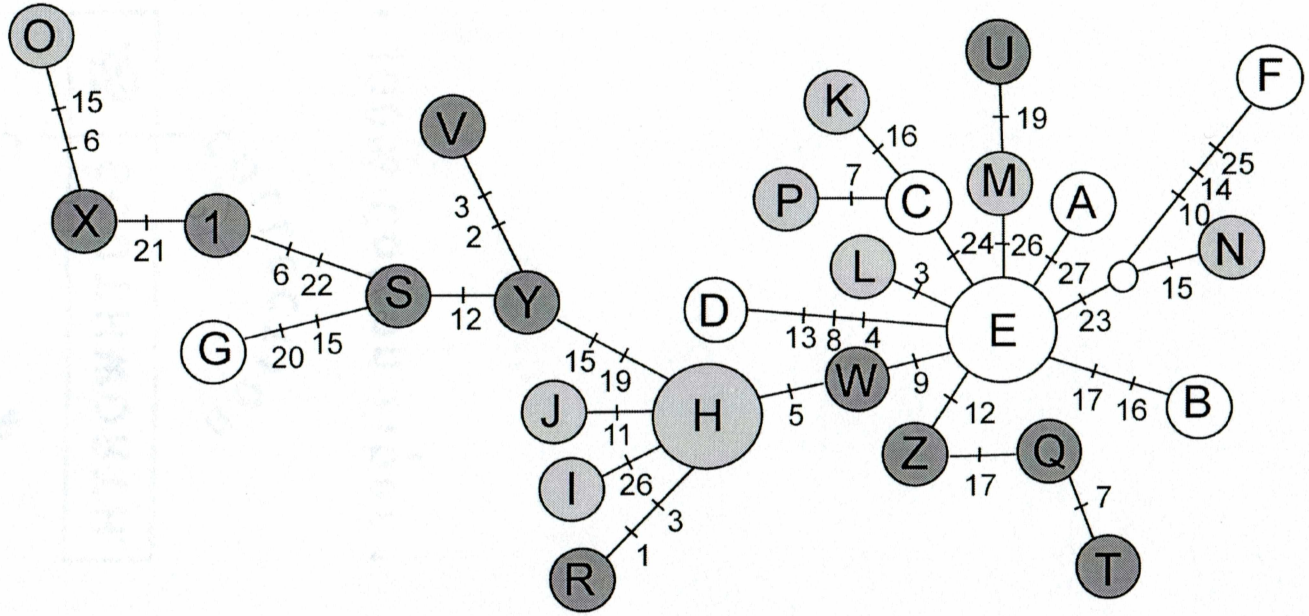

Figure 1.2. Minimum-spanning tree showing the most parsimonious haplotype connections. Tick marks represent mutational steps and the numbers refer to the restriction sites tabled in Appendix 1.1. The white circle represents haplotypes observed in the first survey; the light gray circle represents haplotypes observed in the second and third surveys. The dark gray circle represents haplotypes observed by Churikov et al. 2001.

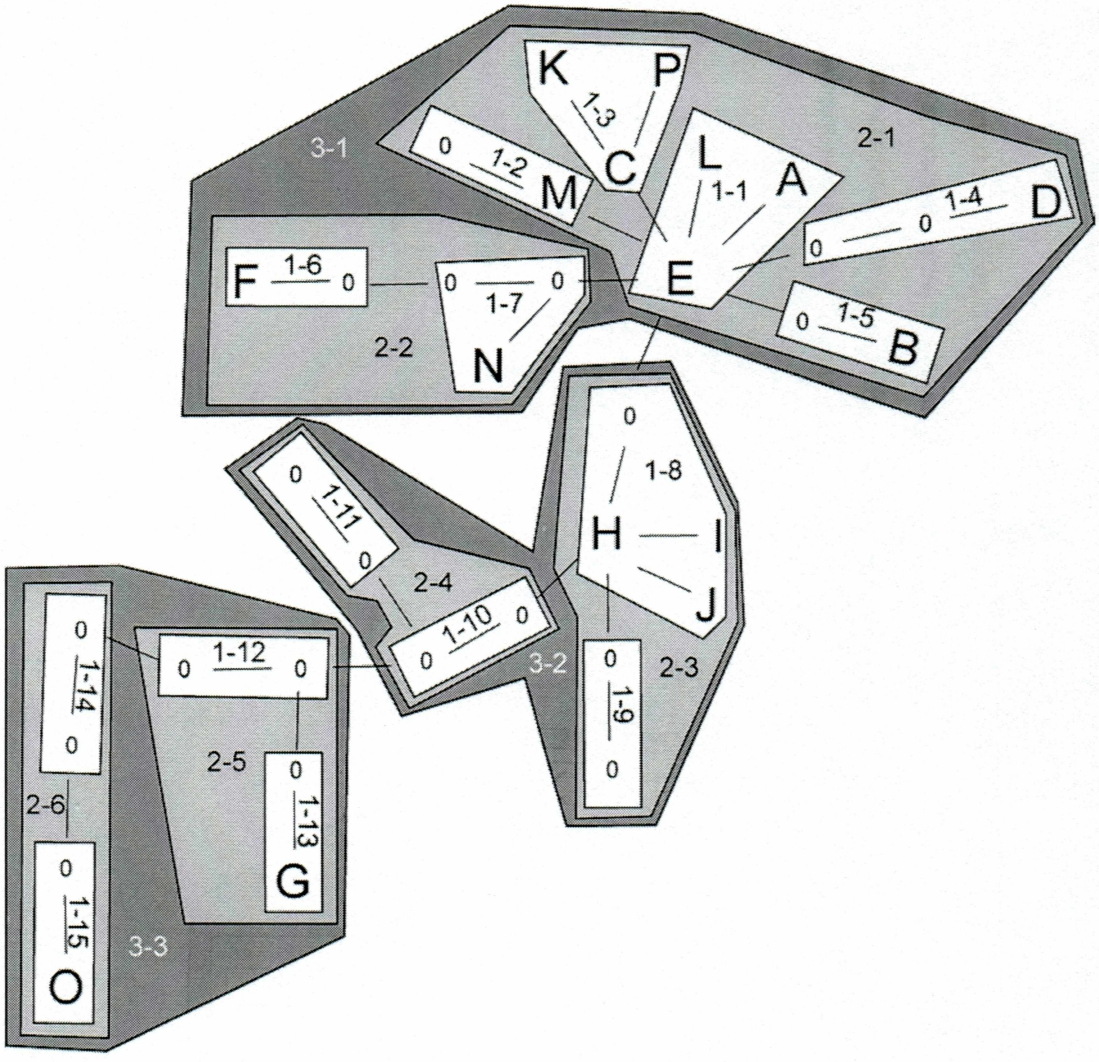

Figure 1.3. Nested clade design developed from the haplotype network. Only halpotypes found in the present study are shown, haplotypes observed by Churikov et al. 2001 are accounted for by leaving zeros in their place.

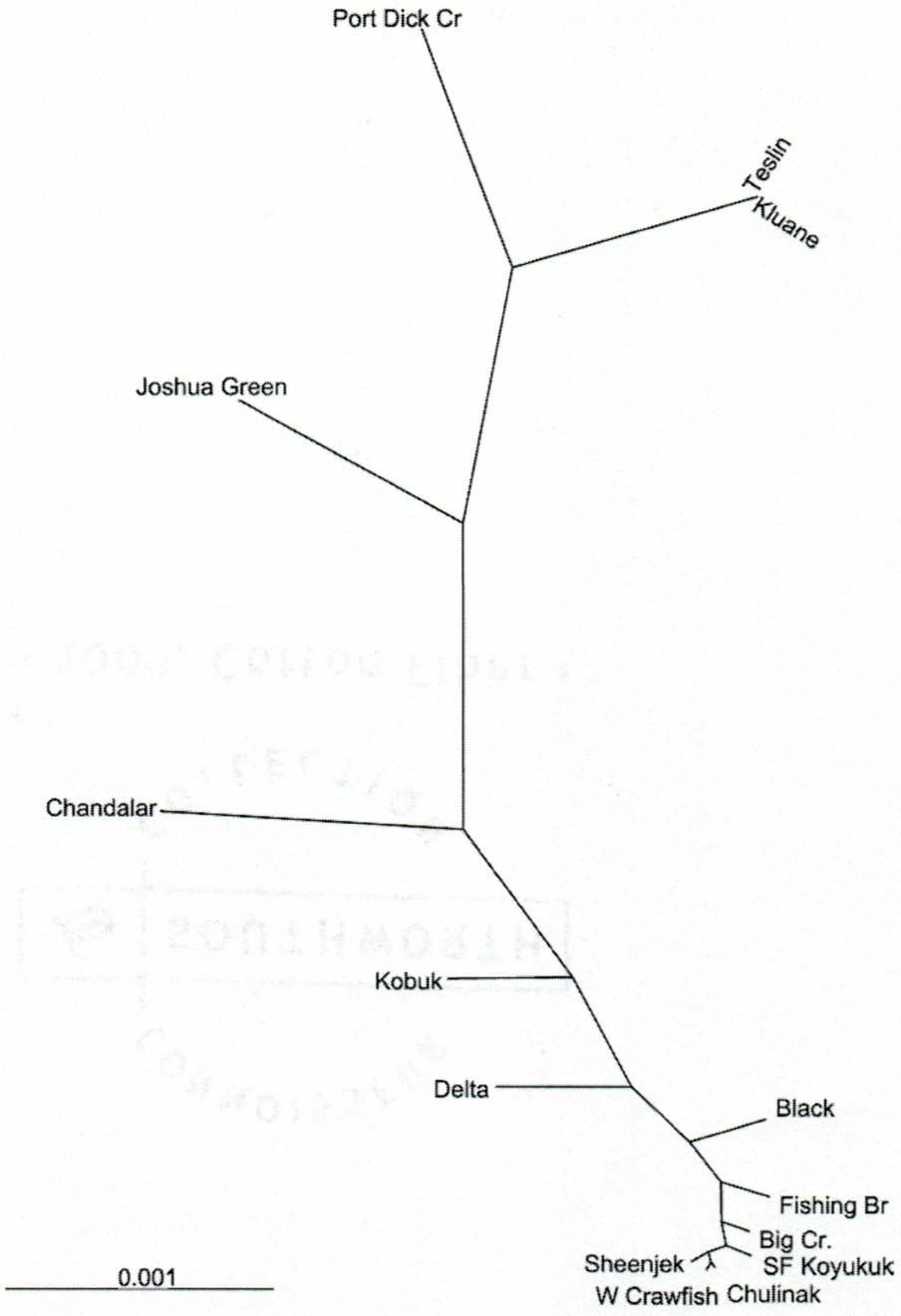

Figure 1.4. Unrooted neighbor joining tree constructed from population pairwise net nucleotide divergence estimates calculated from composite mtDNA halplotypes collected in the third survey.

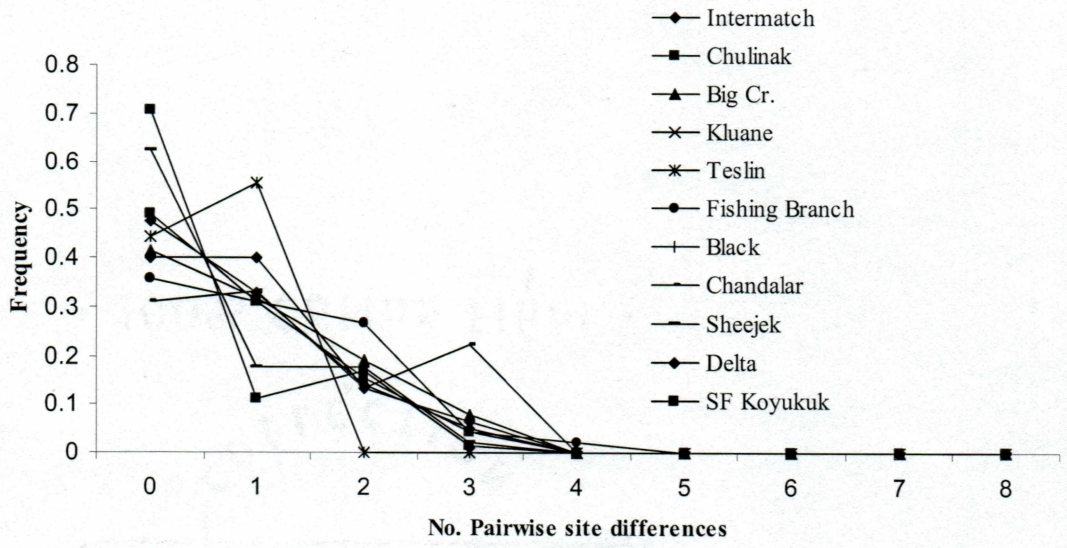

Figure 1.5. Distributions of mismatch and intermatch pairwise restriction site differences within and among ten Yukon River chum salmon populations.

Table 1.1. Restriction site variation in each of seven regions examined using 14 restriction endonucleases in the first survey. Fragment sizes estimated from *Oncorhychus mykiss* mtDNA sequence (Zardoya et al. 1995).

Region	Fragment size	Mean No. of sites	Mean No. of nucleotides	% coverage	No. of variable sites	No. of haplotypes	Haplotype Diversity	Nucleotide Diversity (per 1000)
12S/16S	2402	76.5	309.33	12.88	1	2	0.100±0.088	0.16
ND1/ND2	2689	77.0	324.00	12.05	0	1	0	0.00
COI/COII	2471	54.0	225.33	9.12	3	3	0.195±0.115	0.67
A8/COIII	2115	39.8	170.60	8.07	3	4	0.284±0.128	1.21
ND3/ND4	2331	52.3	218.00	9.35	3	3	0.195±0.115	0.69
ND5/ND6	2488	53.8	224.83	9.04	5	4	0.432±0.126	1.52
Cytb/D-loop	2599	68.5	284.67	10.95	1	2	0.100±0.088	0.18
Total	16600	423.1	1761.24	10.61	16	7	0.637±0.116	0.50

Table 1.2. Distribution of haplotypes and diversity measures for populations in the first suvey. No heterogeneity among population haplotype frequencies was observed (P_{MC}>0.05; 10,000 iterations) (Monte program in REAP; McElroy et al. 1992).

		E lineage Haplotype						G lineage Haplotype
Collection	n	A	B	C	D	E	F	G
Big Creek	10	0	0	3	0	6	0	1
Chulinak	10	1	1	0	1	6	1	0
Total	20	1	1	3	1	12	1	1
Average								

Table 1.3. Distribution of haplotypes for populations in the second survey. No heterogeneity among population haplotype frequencies was observed (P_{MC}>0.05; 10,000 iterations) (Monte program in REAP; McElroy et al. 1992).

		E lineage Haplotype										G lineage Haplotype				
Collection	n	E	K	L	C	M	N	D	A	F	B	G	H	I	J	O
Big Creek	10	6	0	0	3	0	0	0	0	0	0	1	0	0	0	0
Chulinak	21	10	0	0	1	3	1	1	1	1	1	0	1	0	0	1
Joshua Green	10	6	0	0	0	0	0	0	0	0	0	0	2	1	1	0
Kobuk	10	7	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Port Dick Cr.	10	5	0	0	0	0	0	0	0	0	0	0	5	0	0	0
Total	61	34	1	1	5	3	1	1	1	1	1	1	8	1	1	1

Table 1.4. Distribution of haplotypes for populations in the third survey. Heterogeneity among population haplotype frequencies was observed (P_{MC}<0.01; 10,000 iterations) (Monte program in REAP; McElroy et al. 1992).

		E lineage Haplotype				G lineage Haplotype
Collection	n	E	C	P	D	G
Black	10	6	3	0	0	1
Fishing Br.	10	6	2	1	0	1
Sheenjek	10	8	1	0	0	1
Chandalar	10	3	5	0	0	2
SF Koyukuk	10	7	2	0	0	1
Delta	10	6	3	0	0	1
Big Cr.	20	12	5	0	0	3
Teslin	10	5	5	0	0	0
Kluane	10	5	5	0	0	0
Chulinak	31	26	1	0	1	3
Kobuk	10	8	2	0	0	0
Joshua Green	10	6	0	0	0	4
Port Dick Cr.	10	5	0	0	0	5
W. Crawfish	10	10	0	0	0	0
Total	171	113	34	1	1	22

Table 1.5. Nested clade analysis showing only clades with significant results. D_C is the clade distance, D_N is the nested clade distance. L is significantly large. S is significantly small. NS is not significant. Templeton (1998) inference key was used to interpret the results.

Clade	Sub-clade	Tip/Interior	D_C	D_N
2-1	1-1	I	NS	L
	1-2	T	S	S
	1-3	T	S	S
	1-4	T	NS	NS
	1-5	T	NS	NS
		I-T contrast	L	L

Key: 1-2-3-4 No. Isolation by distance.

Clade	Sub-clade	Tip/Interior	D_C	D_N
Total Cladogram	3-1	T	NS	NS
	3-2	I	NS	L
	3-3	T	NS	S
		I-T contrast	NS	L

Key: 1-2-11-17-4 No. Isolation by distance.

Table 1.6. Halpotype outgroup weights

Haplotype	Multiplicity	Weight	Outgroup weight
E	52	66.00	0.215
H	21	24.00	0.078
I	1	0.50	0.002
J	1	0.50	0.002
K	1	0.50	0.002
L	3	1.50	0.005
C	5	58.00	0.189
M	3	56.00	0.182
N	1	0.50	0.002
O	1	0.50	0.002
D	1	0.50	0.002
A	1	0.50	0.002
F	1	0.50	0.002
B	1	0.50	0.002
G	1	0.50	0.002
Q	1	2.00	0.007
R	1	0.50	0.002
S	2	2.00	0.007
T	1	0.50	0.002
U	1	0.50	0.002
V	1	0.50	0.002

Table 1.6. *(Continued)*

Haplotype	Multiplicity	Weight	Outgroup weight
W	1	74.00	0.241
X	6	6.00	0.020
Y	1	3.00	0.010
Z	1	0.50	0.002
1	1	7.00	0.023
Total	111	307	1

Table 1.7. Observed polymorphic sites, mean site difference, time of expansion measured in mutational units, and population size before and after the expansion from the third survey (Schneider et al. 2000). Bracketed values are 95% confidence intervals derived from 1000 parametric bootstrap simulations. All estimates conform to the null hypothesis of sudden expansion (P>0.05).

Population	S	Mean site diff.	Tau	Theta0	Theta1
Intermatch	5 (3-13)	0.77 (0.13-1.77)	0.97 (0-2.62)	0 (0-0.70)	2.35 (0.59-4407)
Chulinak	4 (0-5)	0.49 (0-1.51)	3.06 (0.52-4.35)	0.09 (0-.049)	0.409 (0-209.159)
Big Creek	3 (1-9)	0.93 (0.10-2.17)	1.27 (0-3.71)	0.10 (0-1.20)	2.20 (0.16-4762)
Kluane	1 (1-9)	0.56 (0.20-1.80)	0.87 (0-1.76)	0 (0-1.43)	1680 (2.29-6842)
Teslin	1 (1-9)	0.56 (0.20-1.80)	0.87 (0-1.84)	0 (0-1.43)	1680 (2.29-6842)
Fishing Branch	4 (1-9)	1.07 (0.20-2.51)	1.36 (0.15-3.61)	0 (0-0.67)	4.41 (0.58-6634)
Black	3 (1-9)	0.87 (0.20-1.80)	0.91 (0-2.05)	0 (0-1.31)	1333 (2.64-5576)
Chandalar	3 (1-10)	1.27 (0.20-3.16)	2.06 (0.25-5.31)	0.04 (0-2.80)	2.82 (0.61-5372)
Sheenjek	3 (0-6)	0.60 (0-2.22)	2.79 (0.47-4.19)	0.01 (0-1.01)	0.64 (0-2758)
Delta	3 (1-9)	0.87 (0.20-1.80)	0.91 (0-1.88)	0 (0-1.51)	1333 (2.64-5528)
SF Koyukuk	3 (0-6)	0.76 (0-2.07)	1.11 (0.17-3.55)	0.01 (0-0.71)	1.62 (0-4714)

Appendix 1.1. Polymorphic restriction site information from this study and Churikov et al. 2001

	1	2	3	4	5	6	7	8	9
	12S/16S				COI/COII				
Haplotypes	*Msp* I	*Msp* I	*Msp* I	*Msp* I	*Mbo* I	*Mbo* I	*Msp* I	*Msp* I	*Msp* I
A	0	0	1	1	1	1	0	1	1
B	0	0	1	1	1	1	0	1	1
C	0	0	1	1	1	1	0	1	1
D	0	0	1	0	1	1	0	0	1
E	0	0	1	1	1	1	0	1	1
F	0	0	1	1	1	1	0	1	1
G	0	0	1	1	0	1	0	1	0
H	0	0	1	1	0	1	0	1	0
I	0	0	1	1	0	1	0	1	0
J	0	0	1	1	0	1	0	1	0
K	0	0	1	1	1	1	0	1	1
L	0	0	0	1	1	1	0	1	1
M	0	0	1	1	1	1	0	1	1
N	0	0	1	1	1	1	0	1	1
O	0	0	1	1	0	1	0	1	0
P	-	-	-	-	1	1	1	1	1
Q	0	0	1	1	1	1	0	1	1
R	1	0	0	1	0	1	0	1	0
S	0	0	1	1	0	1	0	1	0
T	0	0	1	1	1	1	1	1	1
U	0	0	1	1	1	1	0	1	1
V	0	1	0	1	0	1	0	1	0
W	0	0	1	1	1	1	0	1	0
X	0	0	1	1	0	0	0	1	0
Y	0	0	1	1	0	1	0	1	0
Z	0	0	1	1	1	1	0	1	1
1	0	0	1	1	0	0	0	1	0

Appendix 1.1. *(Continued)*

	10	11	12	13	14	15	16	17	18
	A8/ND3				ND3/ND4				
Haplotypes	*Ase* I	*Dde* I	*Dde* I	*Nla* IV	*Dde* I	*Dde* I	*Dde* I	*Hinf* I	*Taq* I
A	1	1	1	0	1	1	1	0	0
B	1	1	1	0	1	1	1	1	1
C	1	1	1	0	1	1	1	0	0
D	1	1	1	1	1	1	1	0	0
E	1	1	1	0	1	1	1	0	0
F	0	1	1	0	0	1	1	0	0
G	1	1	0	0	1	1	1	0	0
H	1	1	1	0	1	1	1	0	0
I	1	1	1	0	1	1	1	0	0
J	1	0	1	0	1	1	1	0	0
K	1	1	1	0	1	1	0	0	0
L	1	1	1	0	1	1	1	0	0
M	1	1	1	0	1	1	1	0	0
N	1	1	1	0	1	0	1	0	0
O	1	1	0	0	1	1	1	0	0
P	-	-	-	-	-	-	-	-	-
Q	1	1	0	0	1	1	1	1	0
R	1	1	1	0	1	1	1	0	0
S	1	1	0	0	1	0	1	0	0
T	1	1	0	0	1	1	1	1	0
U	1	1	1	0	1	1	1	0	0
V	1	1	1	0	1	0	1	0	0
W	1	1	1	0	1	1	1	0	0
X	1	1	0	0	1	0	1	0	0
Y	1	1	1	0	1	0	1	0	0
Z	1	1	0	0	1	1	1	0	0
1	1	1	0	0	1	0	1	0	0

Appendix 1.1. *(Continued)*

	19	20	21	22	23	24	25	26	27
				ND5/ND6				CYTB/DL	
Haplotypes	*Ase* I	*Dde* I	*Dde* I	*Dde* I	*Cfo* I	*Rsa* I	*Sau96* I	*Dde I*	*Dde* I
A	0	1	1	0	0	1	1	0	0
B	0	1	1	0	0	1	1	0	1
C	0	1	1	0	0	0	1	0	1
D	0	1	1	0	0	1	1	0	1
E	0	1	1	0	0	1	1	0	1
F	0	1	1	0	1	1	0	0	1
G	1	0	1	0	0	1	1	0	1
H	0	1	1	0	0	1	1	0	1
I	0	1	1	0	0	1	1	1	1
J	0	1	1	0	0	1	1	0	1
K	0	1	1	0	0	0	1	0	1
L	0	1	1	0	0	1	1	0	1
M	0	1	1	0	0	1	1	1	1
N	0	1	1	0	1	1	1	0	1
O	1	1	0	1	0	1	1	0	1
P	-	-	-	-	-	1	-	-	-
Q	0	1	1	0	0	1	1	0	1
R	0	1	1	0	0	1	1	0	1
S	1	1	1	0	0	1	1	0	1
T	0	1	1	0	0	1	1	0	1
U	1	1	1	0	0	1	1	1	1
V	1	1	1	0	0	1	1	0	1
W	0	1	1	0	0	1	1	0	1
X	1	1	0	1	0	1	1	0	1
Y	1	1	1	0	0	1	1	0	1
Z	0	1	1	0	0	1	1	0	1
1	1	1	1	1	0	1	1	0	1

Appendix 1.2. Restriction fragment sizes for seven PCR-amplified regions of chum salmon mtDNA produced by each of 14 endonucleases. Fragment sizes are in base pairs (bp). * indicates additional restriction patterns that were detected by Churikov et al. 2001, but were not observed in this study.

12S/16S						
Ase I	*Ban* II	*BstU* I	*Dde* I	*Hha* I	*Hinf* I	*Mbo* I
A	A	A	A	A	A	A
1400	1200	1250	405	1200	2250	860
1000	1200	460	295	390	150	560
		300	260	335		360
		265	200	195		220
		120	200	160		180
		40	160	130		90
			160			75
			130			70
			105			
			80			
			65			
			60			
			40			
			35			

Appendix 1.2. *(Continued)*

12S/16S										
		Msp I			*Nla* IV	*Rsa* I	*Sau96* I	*ScrF* I	*Sty* I	*Taq* I
A	B	C*	D*	F	A	A	A	A	A	A
	1010				750	770	630	850	2225	1050
				750	600	480	420	450	175	720
620				620	500	425	390	400		520
					375	280	205	325		110
			570		200	190	190	200		
450	450	450	450			170	185	150		
		420				80	170			
390		390	390	390			140			
300	300	300	300				70			
		200								
180	180	180	180	180						
176	176	176	176	176						
130	130	130	130	130						
126	126	126	126	126						
			50							

Appendix 1.2. *(Continued)*

ND1/ND2						
Ase I	*Ban* II	*BstU* I	*Dde* I	*Hha* I	*Hinf* I	*Mbo* I
A	A	A	A	A	A	A
1310	1240	790	785	1100	1390	1310
1090	695	710	495	790	705	405
225	345	640	200	385	300	375
	142	245	200	280	140	220
	110	122	175		95	200
	90	95	170			110
			155			
			125			
			90			

Appendix 1.2. *(Continued)*

ND1/ND2						
Msp I	*Nla* IV	*Rsa* I	*Sau96* I	*ScrF* I	*Sty* I	*Taq* I
A	A	A	C	A	A	A
1420	1290	575	1530	1256	1960	1310
560	496	335	300	430	305	990
515	319	285	280	292	215	310
120	273	270	260	205	140	
	270	240	160	177		
	72	215	90	166		
		180		123		
		170				
		110				
		85				
		83				
		30				

Appendix 1.2. *(Continued)*

COI/COII/A8

Ase I	*Ban* II	*BstU* I	*Dde* I	*Hha* I	*Hinf* I	*Mbo* I			*Msp* I			
A	A	A	A	A	A	A	B	C*	A	B	C	D
2230	1650	1575	480	1530	580		1020	1020	975	975		1655
240	710	895	380	830	490					830		
	160		370	195	490			655	680		680	
			365		250	600					665	
			335		185	420			550	550	550	550
			240		160	400	400	400			310	
			190		160	390	390		200	200	200	200
			120		150	275	275	275	150		150	150
			30		80	265	265					
						95	95	95				

Appendix 1.2. *(Continued)*

COI/COII/A8							
Nla IV		*Rsa* I	*Sau96* I	*ScrF* I		*Sty* I	*Taq* I
A	B	A	A	A	B	A	A
944	944	1050	1520		1377	1450	620
	856	330	310	801		910	490
500		230	300	576			360
356		220	185	421	421		280
188	188	205	145	303	303		265
143	143	170		207	207		210
141	141	140		134	134		140
110	110	70		80	80		98
99	99	50					

Appendix 1.2. *(Continued)*

A8/A6/COIII/ND3

Ase I		*Ban* II	*BstU* I	*Dde* I			*Hha* I	*Hinf* I		*Mbo* I
A	B	A	A	A	B	C	A	A	F	A
	2190	1100	2180			1000	1240	560		2030
1700		880		810	810		940			150
490		210			665			510	510	
				395		395				
				325	325	325		340	340	
				270		270				
				190	190			305	305	
				185	185	185			305	
									255	
								190	190	
								160	160	
								140	140	

Appendix 1.2. *(Continued)*

A8/A6/COIII/ND3								
Msp I	*Nla* IV		*Rsa* I	*Sau96* I	*ScrF* I	*Sty* I	*Taq* I	
A	A	B	A	A	A	A	A	B
940	607	607	725	600	1032	1090	1590	
615	490	490	505	595	466	560		1290
460	418		420	420	318	250	420	420
175	286	286	170	300	303	100		300
		223	150	175	87	100	195	195
	195	195	105	105		80		
	195	195	95					
		195						

Appendix 1.2. *(Continued)*

ND3/ND4L/ND4									
Ase I	*Ban* II	*BstU* I	*Dde* I				*Hha* I	*Hinf* I	
A	A	A	A	B	D	E	A	A	B
1970	1410	1530			1035		1020	1180	1180
380	620	530	920	920		920	710	540	
	320	290	555	555	555	555	620	375	375
									330
						515			210
			310	310	310			150	150
			205	205	205			120	120
				165					
			115			115			
			58	58	58	58			
			55	55	55	55			
			50		50	50			

Appendix 1.2. *(Continued)*

ND3/ND4L/ND4								
Mbo I	*Msp* I	*Nla* IV	*Rsa* I	*Sau96* I	*ScrF* I	*Sty* I	*Taq* I	
A	A	A	A	A	A	A	A	B
920	900	925	570	930	1289	1950		
420	890	455	490	420	1049	250	1120	1120
380	560	312	380	310		100	1110	
265		200	370	200		75		924
123		160	200	185				
95		160	150	160				
55		101	105	155				186
42			85				99	99
25							45	45

Appendix 1.2. *(Continued)*

ND5/ND6											
Ase I		*Ban* II	*BstU* I	*Dde* I				*Hha* I		*Hinf* I	*Mbo* I
A	B	A	A	A	B	C*	D	A	B	A	A
1450		1180	1630	575	575	575	575	1200		930	1420
	890	940	860				570		1000	750	415
800	800	200		360	360	360	360	795	795	280	200
	560	190			330			450	450	280	170
285	285			300			300		200	180	130
				290	290	290		125	125	80	110
				280	280	280					80
				220		220	220				
				200	200	200	200				
				160	160	160	160				
					160	160					
					140	140					
				110		110	110				

Appendix 1.2. *(Continued)*

ND5/ND6								
Msp I	*Nla* IV	*Rsa* I		*Sau96* I		*ScrF* I	*Sty* I	*Taq* I
A	A	A	B	A	B	A	A	A
1300	932		1775	730	730	930	1600	670
430	912	1600		720	720	648	940	580
420	305	395	395	510	510	643		410
320	200	225	225		470	266		395
	99	175		370				305
	77	115	115	100				165
		36	36	75	75			

Appendix 1.2. *(Continued)*

Cytb/D-loop

Ase I	*Ban* II	*BstU* I	*Dde* I			*Hha* I	*Hinf* I	*Mbo* I	*Msp* I	*Nla* IV
A	A	A	A	B	C	A	A	A	A	A
1600	1450	2100	1020		1020	890	1080	1320	630	841
600	1050	250		660		700	675	575	440	475
410	180	150			545	500	665	360	350	359
		105		360		240	255	305	335	298
			300	300		200		110	220	174
			250	250	250	150			185	161
			245	245					160	146
			240	240	240				135	123
			200	200	200				105	108
			110	110	110				65	
			95	95	95				40	
			55	55	55					
			55	55	55					
			30	30	30					

Appendix 1.2. *(Continued)*

Cytb/D-loop				
Rsa I	*Sau96* I	*ScrF* I	*Sty* I	*Taq* I
B	A	A	A	A
1030	1640	981	1060	960
535	810	376	790	950
355	210	297	735	740
340		211	110	95
225		155		
195		146		
15		144		
		123		
		110		
		74		

References

Avise, J.C. 2000. Phylogeography. Harvard University Press, Cambridge, Massachusetts. 464 pp.

Avise, J.C., and seven coauthors. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecological Systematics 18:489-522.

Avise, J.C., C.D. Ankney, and W.S. Nelson. 1990. Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. Evolution 44:1109-1119.

Avise, J.C., J.E. Neigel, and J. Arnold. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. Journal of Molecular Evolution 20:99-105.

Beacham, T.D., C.B. Murray, and R.E. Withler. 1988. Age, morphology, developmental biology, and biochemical genetic variation of Yukon River fall chum salmon, *Oncorhynchus keta*, and comparisons with British Columbia populations. Fishery Bulletin 86(4):663-674.

Billington, N. 2003. Mitochondrial DNA. pp. 59-100. *In*: E.M. Hallerman (ed.) Population genetics: principles and applications for fisheries scientists. American Fisheries Society, Bethesda, Maryland.

Brown, W.M., M. George, Jr., and A.C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences U.S.A. 76:1967-1971.

Castelloe, J. and A.R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Molecular Phylogenetics and Evolution 3:102-113.

Churikov, D., M. Matsuoka, X. Luan, A.K. Gray, V.A. Brykov, and A.J. Gharrett. 2001. Assessment of concordance among genealogical reconstructions from various mtDNA segments in three species of Pacific salmon (genus *Oncorhychus*). Molecular Ecology 10:2329-2339.

Eggers, D.M. 2001. Biological escapement goals for Yukon River fall chum salmon. Regional Information Report No. 3A01-10. Alaska Dept. of Fish and Game. Anchorage, AK. 77 pp.

Johnson O.W. and five coauthors. 1997. Status review of chum salmon from Washington, Oregon, and California. NOAA Technical Memorandum NMFS-NWFSC-32. 280 pp.

Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular inference from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Felsenstein, J. 1995. PHYLIP: phylogeny inference package, version 3.57c. University of Washington, Seattle, WA.

Ferguson, M.M., and R.G. Danzman. 1998. Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? Canadian Journal of Fisheries and Aquatic Sciences 55:1553-1563.

Ferris, S.D., and W.J. Berg. 1987. The utility of mitochondrial DNA in fish genetics and fishery management. pp. 277-299. In: N. Ryman and F. Utter (ed.) Population Genetics and Fishery Management. University of Washington Press, Seattle, Washington.

Gharrett, A.J., A.K. Gray, and V.A. Brykov. 2001. Phylogeographic analysis of mitochondrial DNA variation in Alaskan coho salmon, *Onchorhynchus kisutch*. Fishery Bulletin 99:528-544.

Grant, W.S. and B.W. Bowen. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. Journal of Heredity 89:415-426.

Gyllensten, U., D. Wharton, and A.C. Wilson. 1985. Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. Journal of Heredity 76:321-324.

Kimura, M. and G.H. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561-576.

McElroy, D.M., P. Moran, E. Birmingham, and I. Kornfield. 1992. REAP: an integrated environment for the manipulation and phylogenetic analysis of restriction data. Journal of Heredity 83:157-158.

Posada, D., Crandall, K.A., Templeton, A.R. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Molecular Ecology 9:487-488.

Quinn, T.P. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fisheries Research 18:29-44.

Rogers, A.R. and H.C. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9: 552-569.

Rogers, A.R. and L.B. Jorde. 1995. Genetic evidence on modern human origins. Human Biology 67:1-36.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Schneider, S., Kueffer, J.M., Roessli, D., and L. Excoffier. 2000. ARLEQUIN ver 2.0: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Scribner, K.T., P.A. Crane, W.J. Spearman, and L.W. Seeb. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 55: 1748-1758.

Tallman R.F. and M.C. Healey. 1994. Homing, straying, and gene flow among seasonally separated populations of chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 51:577-588.

Templeton, A.R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Molecular Ecology 7:381-397.

Templeton, A.R., E. Boerwinkle, and C.F. Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and the analysis of alcohol dehydrogenase activity in *Drosophila*. Genetics 117:343-351.

Templeton, A.R., K.A. Crandall, and C.F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619-633.

Wilmot, R.L., R.J. Everett, W.J. Spearman, and R. Baccus. 1992. Genetic stock identification of Yukon River chum and chinook salmon – 1987-1990. Progress Report U.S. Fish and Wildlife Service, Anchorage, AK. 132 pp.

Zardoya, R.A. Garrido-Pertierra, and J.M. Bautista. 1995. The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout *Oncorhynchus mykiss*. Journal of Molecular Evolution 41:942-951.

CHAPTER 2

Application of Amplified Fragment Length Polymorphisms (AFLP) to Mixed-Stock Analysis of Yukon River Fall Chum Salmon, *Oncorhynchus keta*

Abstract

To manage Yukon River fall chum salmon effectively and allocate catches equitably between the United States and Canada, harvest estimates for individual populations must be determined. This task is difficult because the harvest of these populations takes place before they segregate into spawning aggregations. Past studies of Yukon River chum salmon populations using allozymes, microsatellites, and mitochondrial DNA (mtDNA) restriction site analysis have been unable to produce mixed-stock analysis (MSA) estimates that were ≥ 80% accurate to the country of origin. Here we examine another genetic marker type, amplified fragment length polymorphisms (AFLP), in an attempt to increase our ability to distinguish between United States and Canadian fall Yukon River chum salmon. Ten chum salmon populations from the United States and Canada were analyzed at 30 variable AFLP loci. Results show that Yukon River chum salmon populations are structured by both run time and regional location. The MSA was most successful when mixtures were allocated to regions. The AFLP data were able to provide improved MSA estimates for the border populations by country of origin with a 6.5% improvement for the Canadian populations over microsatellite analysis. Furthermore, AFLP shows promise in MSA applications because, of all the markers tested, AFLP may be the least expensive, quickest to run, and most accurate. In general, the results for all marker types were similar, suggesting that the dominant factors influencing population genetic structure are geographically restricted gene flow, large effective population sizes (N_e), and historical demographic events. It does not appear to be the failure of a marker system that limits attaining higher accuracy in MSA.

Introduction

To manage a mixed stock fishery for Yukon River chum salmon (*Oncorhynchus keta*), populations must be identified and the proportion of each population in the catch must be determined (Larkin 1981). If fishery managers allow harvests based on the belief they are harvesting a single population when, in fact, a mixture of populations is being harvested, the result can be excessive exploitation of an individual population through differential harvest rates (Allendorf et al. 1987). Differential harvest can decrease overall production for the entire system.

Genetic analysis of Yukon River chum salmon based on allozymes, microsatellites, minisatellites, and mtDNA restriction site analysis revealed a population structure defined by run time and geographic region with greater divergence between run times and among regions than within run times and regions (Beacham et al. 1988; Wilmot et al. 1992; Beacham 1996; Spearman and Miller 1997; Scribner et al. 1998; Conservation Genetics Laboratory, U.S. Fish and Wildlife Service, unpublished data). Application of mixed-stock analysis (MSA) (e.g., Grant et al. 1980) to simulations of mixed fisheries showed that the most accurate and precise estimates were obtained for geographically defined regions that ignored political boundaries. Moreover, simulations showed that problems persist in achieving accurate and precise estimates of country of origin from mixtures of populations adjacent to the United States/Canada border; their genetic similarity reflects their geographic and possibly ecological similarity. The low genetic resolution of populations in this area has confounded resource management. Increased resolution would improve management decisions designed to meet allocation and harvest/conservation goals mandated in the Pacific Salmon Treaty. Because of the difficulty in identifying differences among the border populations, we explored a new approach.

The AFLP technique (Vos et al. 1995) allows access to thousands of anonymous loci, which may be most appropriate for detecting divergence among closely related populations (Bernatchez and Duchesne 2000). For MSA applications, a technique that rapidly surveys numerous variable anonymous loci distributed through out the genome

may yield resolution in population studies for which results from traditional methods are unsatisfactory. Disadvantages of this technique include dominant expression and other assumptions that must be made for analyses, including, homology of comigrating bands, Mendelian inheritance, and Hardy-Weinberg equilibrium. Fortunately these assumptions appear to hold generally and the bias introduced is probably negligible (Parker et al. 1998; Krauss 2000; Vekemans et al. 2002). Here we analyzed the variation at AFLP loci from populations of fall and summer Yukon River chum salmon and a reference sample (out-group) from Southeast Alaska. The study objectives were (1) to improve MSA estimates for Yukon River United States/Canada border chum salmon populations, (2) to determine if the most informative subsets of loci, as determined by relative and absolute divergence and variance estimates, improve MSA by amplifying the signal of divergence, and (3) to determine if AFLP markers reveal a picture of Yukon River chum salmon population structure that is concordant with other markers previously used.

Methods

Sample collection and DNA isolation

Between 1989 and 1996, 956 samples were collected from eleven populations (Figure 2.1). Eight of the populations represent the major producers of fall chum salmon for the Yukon River. Two summer populations (Chulinak, South Fork Koyukuk) were included to represent summer chum salmon in mixed fisheries. A genetically divergent chum salmon population from Southeast Alaska was included as a reference. Total genomic DNA was extracted from either heart or fin tissue (~25mg) with proteinase K and the Puregene™ DNA isolation kit (Gentra Systems Inc. Minneapolis, MN). Concentrations of DNA were quantified by fluorometry.

AFLP protocol

Methods generally followed those in Perkin Elmer's AFLP plant mapping protocol and included four steps: (1) digestion-ligation; (2) pre-amplification; (3) selective amplification; and (4) electrophoresis (Perkin Elmer 1997). Total genomic DNA (250 ng) was subjected to simultaneous digestion by restriction endonucleases *Eco*R I and

Mse I and ligation of adapters to the fragments (Vos et al. 1995). Pre-amplification polymerase chain reaction (PCR) was done in 20 µl volumes: 2.5 mM $MgCl_2$, 1X PCR buffer (20 mM Tris-HCl pH 8.0, 50 mM KCl), 200 µM for each dNTP, 0.3 µM for *Eco*R I + A (where *Eco*R I + A consists of the core and enzyme sequence plus the selective nucleotide: 5′-GACTGCGTACCAATTC+A-3′) and 0.3 µM for *Mse* I + C (5′-GATGAGTCCTGAGTAA +C-3′) primers, 0.016 units of Taq polymerase, and 4 µl of diluted digestion-ligation mix. PCR was performed in a Stratagene Robocyler® with an initial cycle of 2 minutes at 72°C, followed by 20 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 2 min, with one final cycle at 60°C for 30 min.

Selective PCR was done in 10 µl volumes; conditions were: 2.5 mM $MgCl_2$, 1X PCR buffer (20 mM Tris-HCl pH 8.0, 50 mM KCl), 200 µM for each dNTP, 0.05µM for *EcoR* I + AXX (X stands for any nucleotide) fluorescently labeled primer, 0.25 µM for Mse I + CXX primer, 0.008 units Taq polymerase, and 3 µl of diluted pre-amplification product. Thermal cycling conditions were: initial denaturation cycle of 94°C for 3 min, followed by 94°C for 1 min, 66-57°C for 1 min (touchdown phase), 72°C for 2 min 10 sec, then 20 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 2 min 10 sec, with a final single cycle of 60°C for 30 min.

Three to four µl of PCR product were electrophoresed on a denaturing 6% polyacrylamide gel for 50 to 105 minutes. An FMBIO II® (Hitachi Software Engineering America Ltd.) fluorescent imaging scanner was used to visualize the DNA fragments. The sizes of bands were estimated and scored by FMBIO® analysis software v8.0. All band scores were verified by visual inspection, independent of the software.

A pilot study was conducted to identify a subset of approximately 30 informative loci. Initially, two samples from each population were amplified by PCR using 64 primer combinations (Perkin Elmer AFLP plant kit) to identify primers that revealed frequency differences among the populations. Next, 40 DNA samples from each population were pooled into four groups of 10 and screened for all 64 primer combinations to discover those that might reveal large population specific frequency differences and/or unique loci. Finally, 10 individual samples from each of the populations were processed for 13

promising primer combinations. The loci were scored as binary (1, 0) phenotypic data and analyzed for variance among populations. The primer combinations that produced 30 unambiguously scored loci with the greatest variance in banding frequencies among populations were run on the entire sample set (Figure 2.1). To ensure that the results were reproducible, 10 random samples from each population were subjected to the entire AFLP process twice.

Statistical analyses

The hierarchy of Yukon River chum salmon populations analyzed was based on seasonal run timing and geographic location. The AFLP data can be interpreted as either phenotypic data showing the presence of a dominant allele, or transformed to estimate allele frequencies (Nei and Kumar 2000). Analyses were conducted on both phenotypic frequencies and estimated allele frequencies, depending on the test. Allele frequencies were estimated by a Bayesian procedure (Zhivotovsky 1999) that assumes a non-uniform prior distribution of fragment frequencies except for the gametic phase equilibrium test where a maximum likelihood approach was used.

Statistical allelic analyses

Gametic phase equilibrium was estimated by the program RAPDLD (Black 1997) and analyzed following the methods of Apostal et al. (1996). Average unbiased expected heterozygosities were calculated according to Nei (1978). The percentage of polymorphic loci at the 95% criterion was determined. Nonmetric multidimensional scaling (Kruskal 1964) of chord distances (Cavalli-Sforza and Edwards 1967; Wright 1978) calculated from allele frequencies was carried out in NTSYS 2.1 (Rohlf 2000). In addition, a minimum spanning tree (MST) constructed from the chord distance matrix was superimposed on the multidimensional scaling to detect distortions in the ordination process. The distances of the MST were ranked from 1-10 with 1 being the smallest and 10 being the largest. Saltery Cr., the reference population, was not included in any further analyses.

A hierarchical unbiased gene diversity analysis was used to partition the genetic variation due to population divergence, according to the following model: $H_T=H_S+D_{SR}+D_{RC}+D_{CT}$ (Chakraborty and Leimar 1987; Nei and Chesser 1983). The harmonic mean was conservatively based on the sample size (N) so the bias correction was slightly excessive. H_T is the overall average expected heterozygosity; H_S is the average expected heterozygosity within populations; $D_{SR} = H_R - H_S$ is the diversity attributed to divergence among populations within regions; $D_{RC} = H_C - H_R$ is the diversity attributed to divergence among regions within run times; $D_{CT} = H_T - H_C$ is the diversity attributed to divergence between run times. The relative proportions of the total diversity accounted for by the different levels of hierarchy were estimated by calculating coefficients of gene differentiation (G_{ST}-statistics). Effective migration rates (N_em) were estimated from G_{ST}-statistics, assuming a hierarchical island model at equilibrium using the formula: $N_em = ((1-G_{ST})-1)/4)/(g/(g-1))2$ (Zhivotovsky et al. 1994). Significance of the G_{ST}-statistics was inferred from log-likelihood ratio tests of homogeneity (Chakraborty and Leimar, 1987). To further evaluate the reliability of the data set, confidence intervals for G_{ST}-statistics and migration estimates were determined by bootstrapping over loci 1000 times.

Isolation by distance (IBD) was examined by standard linear regression of population pairwise matrices of $\log_{10}(N_em)$ on $\log_{10}$(distance in km), N_em was estimated from ¼$(1/G_{ST} - 1)$ (Slatkin 1993). Significance of the correlation between the two matrices was determined by the Mantel test (Mantel 1967) with 10,000 randomizations using Arlequin 2.0 (Schneider et al. 2000).

Statistical phenotypic analyses

Hierarchical log-likelihood ratio tests were conducted to determine the heterogeneity of phenotypic frequencies among populations within regions, among regions, and between run times (*G*-test, Sokal and Rohlf 1995). To avoid violating asymptotic assumptions, phenotypes with expected overall counts of less than four were excluded from analysis. The magnitude of heterogeneity among and within fall regions was compared using an approximate *F*-statistic (Smouse and Ward 1978): $F_{\text{df among, df within}} = (G_{\text{among}}/\text{df}_{\text{among}})/(G_{\text{within}}/$

df_{within}). Significance levels were adjusted to prevent type I error for multiple tests of the same hypothesis by dividing the alpha by the number of tests (Cooper 1968).

The MSA simulations were performed using the direct maximum likelihood method as implemented in the program SPAM 3.5 (Debevec et al. 2000). Parametric bootstrap resampling of both the baseline and mixture was repeated 1000 times to derive mean allocation estimates and to evaluate precision. Specifically, bootstrapping the baseline entails randomly and independently sampling phenotypes at each locus in each population from the baseline phenotypic frequency distribution to re-estimate the parameters (phenotype frequencies) of the frequency distribution based on the baseline sample sizes. Mixture bootstrapping consists of first randomly drawing N individuals (defined by the stated mixture sample size and the specified mixture composition) from the multinomial sample of individuals; and then creating the multilocus phenotypes for those individuals by randomly and independently sampling with replacement phenotypes at each locus using the relevant observed baseline phenotype frequencies, and an assumption of gametic phase equilibrium.

Artificially simulated mixtures (N = 400) representing 100% of each individual population were subjected to MSA to test baseline performance. Mean allocations to individual populations were then summed for geographically defined regions. For the United States/Canada border populations, additional summing to politically defined border regions was also done. The results of the AFLP analyses were compared to results from microsatellite analyses by conducting simulations using data from 11 microsatellite loci collected from the same individuals and populations (Conservation Genetics Laboratory, U.S. Fish & Wildlife Service, unpublished data). Additional MSA simulations were performed on the regions (biological and political) using equal proportions of each population from the region in the simulated mixtures. A randomization test was conducted using the Resample Stats Excel add-in (Resampling Stats, Inc. Arlington, VA) to determine if there were significant differences between AFLP and microsatellite MSA estimates for the border region by country of origin. The AFLP and microsatellite estimates were assumed to be independent samples and the difference between the mean

estimates was used as the test statistic. Estimates were randomized 1000 times between the AFLP and microsatellite categories.

An attempt to improve MSA estimates for the border populations was made by selecting only the most informative loci among these populations for analysis. This approach was taken because loci that are homogenous among populations may overwhelm the signal of divergence and prevent improved MSA estimates. A variety of criteria were used for determining the most informative loci, these included the highest *G*-test ratio, G_{ST}, D_{ST}, H_T, allelic variance, and phenotypic variance among the border populations. These criteria were chosen because no one metric fully describes population relationships as described by Wright (1978) who stated that the allelic variance, limiting variance (H_T), and fixation index must be analyzed when interpreting data. The other criteria were chosen because they are additional measures that detect divergence, the key to success for MSA. The 10 and six most informative loci for AFLP and microsatellites, respectively, were evaluated for each selection criteria in MSA simulations performed on the United States and Canadian border regions with equal contributions from the populations.

To determine the accuracy over a range of mixture scenarios, MSA was performed on simulated mixtures (N = 400) of Canadian chum salmon originating from the four Canadian populations ranging from 0% to 100% at 20% increments. Equal proportions of the chum salmon populations were used to create the mixtures. An accuracy graph plotted the mean estimated proportions against true proportions along with the 95% nonsymmetric confidence intervals.

Results

The two primer combinations chosen, amplified 117 AFLP loci of which 59 were variable. Data for 30 of these loci, identified in the pilot study, were collected for analyses.

Primer pairs	No. of fixed loci	No. of variable loci	No. of variable loci scored
E-ACC/M-CAC	22	40	17
E-AAG/M-CAC	36	19	13
Total	58	59	30

Gametic phase disequilibrium ($P<0.05$) was observed in 3%-7% of the pairwise tests within populations, which generally conforms to the expectation of the type I error rate. Two pairs of loci showed consistent linkage ($P<0.05$), one in six and the other in eight populations; the less informative locus from each pair was omitted from further analyses.

Average unbiased expected heterozygosities and percent polymorphic loci were similar among populations ranging from 27.8% to 31.6%, and 82.1% to 89.3%, respectively. There was a significant difference between the extremes of heterozygosities ($P=0.05$, Sign-test, Sokal and Rohlf 1995), but not for polymorphic loci ($P>0.05$, McNemar's test, Sokal and Rohlf 1995). However, differences among the populations of greater significance were measured by the *G*-test.

Nonmetric multidimensional scaling (Kruskal 1964) of chord distances (Cavalli-Sforza and Edwards 1967; Wright 1978) calculated from allele frequencies at 28 AFLP loci revealed a population genetic structure defined by run time and geographic region (Figure 2.2). The measure of stress for the analysis was low (0.02), signifying good compression of the data into two dimensions. Saltery, the reference population from Southeast Alaska, differed from the Yukon River populations. Within the Yukon River, populations were less divergent but formed three clearly defined groups. These groups included summer populations from the lower region (Chulinak, South Fork Koyukuk), and fall populations from the middle region (Delta), and the border region (Black, Sheenjek, Chandalar, Fishing Branch, Big Creek) (Figure 2.1). The upper Yukon River fall populations (Teslin and Kluane) did not cluster closely either with each other or with other populations. The minimum spanning tree revealed that the closest relationships were among the border populations and then between the summer populations. This

indicates that there was some distortion in plotting the populations in the first two dimensions but, in general, the scaling accurately described the estimated genetic relationships among the Yukon River chum salmon populations.

Significant overall heterogeneity was observed between run times and within and among regions (Table 2.1). However, levels of heterogeneity among fall regions were greater than levels within fall regions ($F_{42, 105} = 3.20$ $P<0.0001$). Populations of the border region had the lowest heterogeneity. The upper region, which includes the Kluane and Teslin populations, was the most heterogeneous. Out of 21 total loci and after correcting for multiple tests, significant heterogeneity was observed at: seven loci between summer and fall runs, one locus within the lower Yukon region, two loci within the border region, four loci within the upper region, five loci within fall run regions, and five loci among fall run regions (Appendix 2.1).

Variation at 28 loci was partitioned by gene diversity analysis (Chakraborty and Leimar 1987; Nei and Chesser 1983) to examine the magnitude of divergence among the populations and levels of hierarchy (Table 2.2). Overall gene diversity analysis revealed that 97.9% of the variation occurred within populations, and 2.1% among them. The majority of among-population diversity (1.3%) was accounted for by regional differences; and differences between run times and among populations within regions accounted for 0.2% and 0.6%, respectively. All values differed significantly from zero as inferred from the hierarchical *G*-tests (Chakraborty and Leimar 1987). Estimates of $N_e m$ were 4.6 to 43.5 migrants per generation; the greatest value was among populations within a region and the lowest value was among regions (Table 2.2).

The 95% bootstrap confidence interval for the between run-times coefficient of gene differentiation (G_{CT}; -0.046% to 0.4%) includes 0. This would lead to the interpretation that G_{CT} is not significant; however, tests of significance are most powerful and appropriate when done on "raw" data (i.e., phenotype or allele frequencies) rather than sample statistics (Hawkins et al. 2002). A possible explanation for this situation begins with the study design. Because the emphasis is on fall Yukon River chum salmon, the design is unbalanced and includes eight fall and two summer populations.

Furthermore, South Fork Koyukuk is a mid-river summer population, which is more closely related to fall populations because of a possible cline in Yukon River chum salmon (Crane et al. 2001). Therefore, the design is not the best for characterizing the summer and fall relationship. Whereas *G*-tests reveal that there are significant differences between run times, quantifying the magnitude of divergence and gene flow between the runs will require a better design.

Regression analysis and the Mantel test suggest that the Yukon River chum salmon populations exhibit a genetic structure that is consistent with an isolation by distance model (R^2=0.3163; Figure 2.3). The correlation between $N_e m$ and geographic distance is significant (P=0.002), which suggests that gene flow is geographically restricted and that these populations are approximately at migration-drift equilibrium (Slatkin 1993).

Estimates of the proportionate contributions of populations to a simulated mixture varied between 62.5% and 89.9% accuracy (S.E., 4.8% to 10.6%); with results for the populations adjacent to the international border ranging from 62.5% to 79.9% (S.E., 8.9% to 10.6%). Results improved considerably when estimates were summed over regions (Table 2.3). Misallocations were usually made to adjacent regions. For the border populations, summing to politically defined border regions reduced accuracy and precision as compared to those for the biological region (Table 2.4). Results from the simulations using the microsatellite data revealed that AFLP data were somewhat better at characterizing the Canadian border populations both individually and regionally, the randomization test revealed that the Canadian border region estimate for the AFLP data was significantly greater (P<0.0001) than the estimate for the microsatellite data (Table 2.4, 2.5). In contrast, the microsatellites had better individual estimates for the U.S. border populations; but the U.S. border region estimate was marginally lower (P=0.048) than the AFLP estimate (Table 2.4, 2.5).

Efforts to increase the accuracy and precision of estimates for the politically defined border regions by using only the most informative loci produced mixed results for the AFLP and microsatellite data. For the AFLP data, estimates for the Canada border

region improved (89.7% - 92.4%; S.E. 6.8% - 7.3%) at the expense of the United States border region (76.1% - 81.0%; S.E. 10.1% - 12.2%), except for the loci with the highest H_T (total expected heterozygosity). These loci produced estimates for both regions of 68.0±14.6%, which were less accurate and precise. The standard errors for the MSA estimates derived from subsets of loci were high, and the results were quite susceptible to changes in methods, suggesting that the differences among subsets of loci are not large. For the microsatellite data, estimates were not affected by locus selection. The estimates for the Canada border and United States border ranged from 78.8% - 79.9% (S.E. 4.5% - 5.6%) and 79.6% - 82.4% (S.E. 4.6% - 5.8%), respectively.

The MSA of simulated samples of Canadian chum salmon ranging from 0% to 100% at 20% increments revealed that accuracy was greatest when Canadian chum salmon comprised 40% of the mixture (Figure 2.4). Theory predicts that this point would be most accurate because there are four Canadian and six U.S. populations with the simulated mixtures comprised of equal proportions of each population; accuracy for MSA is greatest when all the populations contribute to the mixture equally (Wood et al. 1987). Estimates were within 11.4% of the actual value; Canadian chum salmon proportions at 0% and 20% expected were overestimated. This overallocation was distributed among the Canadian border populations and Teslin, with an underalloction of United States border populations, especially Sheenjek and Chandalar. Canadian chum salmon proportions were underestimated when the expected value ranged between 60% and 100% because of underallocation of fish to Fishing Branch and overallocation of fish to Sheenjek and Chandalar.

Discussion

Surveying Yukon River chum salmon populations for AFLP variation revealed temporal and spatial structuring that follows an isolation by distance pattern. Nested within run time, levels of heterogeneity were significantly greater among regions than within regions, suggesting that the Yukon River chum salmon populations form the following regional associations: Lower River, Middle River, Border area, and Upper River. The Upper River populations, Kluane and Teslin, were distinct from each other and from the

other populations. These results are consistent with those of previous studies (Beacham et al. 1988; Wilmot et al. 1992, 1994; Beacham 1996; Spearman and Miller 1997; Scribner et al. 1998 ; Conservation Genetics Laboratory, U.S. Fish & Wildlife Service, unpublished data). In addition, G_{ST} values for AFLP, allozyme (Wilmot et al. 1994), and microsatellite (Conservation Genetics Laboratory, U.S. Fish & Wildlife Service, unpublished data) data collected for the same populations were 0.021, 0.019, and 0.017, respectively. These values are also similar to those reported for allozyme data (albeit at different geographic scales) for the Yukon River 0.028 (Seeb and Crane 1999), Southeast Alaska/British Columbia 0.030 (Kondzela et al. 1994), British Columbia 0.023 (Beacham et al. 1987), Washington/British Columbia 0.028 (Phelps et al. 1994), and Japan/Russia 0.038 (Winans et al. 1994).

The level of AFLP divergence we observed produced the most accurate and precise MSA estimates for biogeographic regions in comparison to estimates for political regions or individual populations. Resolution among the border populations is low. Increased genetic resolution among the border populations is desirable for managing and allocating fish. Unlike past studies, simulations show that the AFLP technique was able to produce MSA estimates that were $\geq$ 80% accurate for the border populations by country of origin. However, while the improvement of the AFLP data set over the others was significant, it was not dramatic (Table 2.4, 2.5). Although all data sets gave estimates of $\geq$ 80% for the U.S. border populations, only the AFLP technique revealed sufficient divergence to provide an estimate $\geq$ 80% accurate for the Canadian border populations.

The amount of divergence revealed varies among loci (Ewens 1983); therefore, in an attempt to improve MSA estimates, subsets of informative loci were analyzed with MSA simulations in an effort to concentrate the signal of divergence. Different criteria were used in selecting informative loci because no single measure can capture the full spectrum of population structure. In various situations variance estimates may capture the total amount of divergence among populations better than F_{ST} or some analogue, which, although popular for estimating divergence, actually assesses progress towards allelic fixation (Wright 1978). However, this approach did not improve results, which supports

the observations of Scribner et al. (1998) who also failed to improve MSA estimates by using subsets of diverse loci, and who suggested that loci with large allelic diversity do not necessarily improve stock discrimination. In addition, it has been shown that two-allele systems with the common allele frequency at or below 0.8 provide the strongest resolution among stocks, and that multiple rare alleles provide little information for stock mixture analysis (Waples and Smouse 1990).

The results presented here were generally similar to those of previous studies; the different data sets provide a concordant picture of the genetic structure of Yukon River chum salmon. The genetic compositions of these populations are the net result of four factors acting separately or in combination: 1) gene flow; 2) limited genetic drift (large N_e); 3) limited time of separation; and 4) historical demographic events. First, geographically restricted gene flow is supported by indirect estimates of effective migration and isolation by distance analysis. Estimates of migration based on G_{ST} statistics assume drift-migration equilibrium, which the significant isolation by distance suggests (Slatkin 1993). This assumption is often not met for salmonids over long ranges; however, the Yukon River is a relatively small area, compared to the entire range of chum salmon, and the rates of gene flow observed were in line with estimates from other chum salmon studies (e.g., Beacham et al. 1987; Kondzela et al. 1994; Phelps et al. 1994) and with estimates for Yukon River chinook salmon (Gharrett and Zhivotovsky 2003). Moreover, meeting this assumption is not critical in situations where $4 < N_e m < 1$ (Hutchinson and Templeton 1999). Secondly, a large N_e, which reduces the rate of genetic drift, could be responsible for low G_{ST} values and is supported by Pleistocene refugial evidence, and current run sizes (Eggers 2001; Neigel 2002). However, this scenario suffers because Pleistocene glaciations severely reduced water levels, which would have led to the extirpation or reduction of populations, and census sizes are rarely indicative of N_e. Thirdly, a limited time of separation among the populations is possible because of probable habitat loss during the Pleistocene glaciations and the likely recolonization afterwards, although the significant isolation by distance indicates that recolonization has not occurred very recently (Slatkin 1993). Lastly, residual historic fragmentation or range

expansion events resulting from glacial processes of the Pleistocene and beyond could have influenced the current pattern (Templeton 1998).

For example, the divergence of the Kluane and Teslin populations may be related to the geologic history of the region. The Yukon-Kuskokwim region was a refuge during the last glacial maximum (14,000-20,000 years ago), except for the upper river portion, which was glaciated and now harbors the Kluane and Teslin populations. The minimum spanning tree indicates that Kluane and Teslin may have been recolonized from populations of the border region, and the genetic divergence may be the result of postglacial founder events or subsequent bottlenecks. Recolonization from the border region is much more probable than from Southeast Alaska, as has been speculated (Wilmot et al. 1994) because Lake Kluane drained via the Alsek River to the Gulf of Alaska, not the Bering Sea, until a headwater exchange occurred about 400 years ago. Southeast Alaskan chum salmon are adapted to short migrations up coastal rivers and it is doubtful that they could have survived the 2000 mile migration of the Yukon River.

The similarities among the populations appear to be, primarily, the result of gene flow, thus it is unlikely that other types of neutral genetic markers will tell a substantially different story. However, increasing the number of loci may increase resolution. Because considerable effort has been devoted to Yukon River chum salmon MSA with similar outcomes, we conclude that investigation of additional molecular techniques assaying neutral genetic markers is unwarranted. Although, additional work focusing on loci under selection (e.g., MHC, *Pan* I) and on techniques that offer, potentially, faster sample processing (e.g., micorarrays, SNPs) may be worthwhile for providing fisheries managers with more accurate and timely MSA estimates. In the meantime, the AFLP technique may be preferable because it can accurately apportion mixtures to the following Yukon River regions: Lower, Middle, Border United States, Border Canada, and Upper, can rapidly resolve thousands of anonymous loci, and may be the least expensive, although microsatellites performed nearly as well.

Mixed-stock analysis using AFLP or microsatellite markers offers managers a powerful tool for evaluating fisheries. Current management of Yukon River chum salmon

relies on sonar enumeration, test fishery catch per unit effort (CPUE), and subsistence harvest reports to make an initial assessment of run strength, little information concerning the contribution of regions and populations to harvests is incorporated. The fact that fall chum salmon usually enter the river in four to five pulses associated with offshore wind events or high tides complicates this assessment because the "wide fluctuations in pulse size and run timing have not yet been characterized" (JTC 2001). For example, in 2001 the first pulse of fall chum salmon was among the largest recorded at 109,000 fish; subsequent pulses were weak and the overall run was poor. Runs have been depressed since 1998 and subsistence and commercial fishing are restricted or closed when the total run size drops below 600,000 fish. If there is an indication that a region will exceed escapement goals, a directed fishery may be opened. Unfortunately, current inseason run assessment tools are generally inadequate to allow directed fisheries (JTC 2001). Region-specific pulses of chum salmon are however, detectable by MSA; this information should assist in making inseason assessments and regulating fisheries.

Acknowledgments

This study was funded by the Conservation Genetics Laboratory, U.S. Fish and Wildlife Service through a grant (FWS 70181-9-N059; USGS BRD 1434-HQ-99-RM-0016) administered by the Alaska Cooperative Fish and Wildlife Research Unit, Biological Resource Division, U.S. Geological Survey, University of Alaskas Fairbanks. We thank Steve J. Miller, William J. Spearman, Steve Klein, Richard Pospahala, Steve Klosiewski, Richard Hannan and LaVerne Smith from the U.S. Fish and Wildlife Service; William W. Smoker from the University of Alaska Fairbanks; and Richard L. Wilmot from the National Marine Fisheries Service for their support of this project. We thank Bernie May and Jeremy J. Agresti from the University of California, Davis for valuable training on the AFLP process. We thank Jud Monroe for reviewing an earlier version of this manuscript.

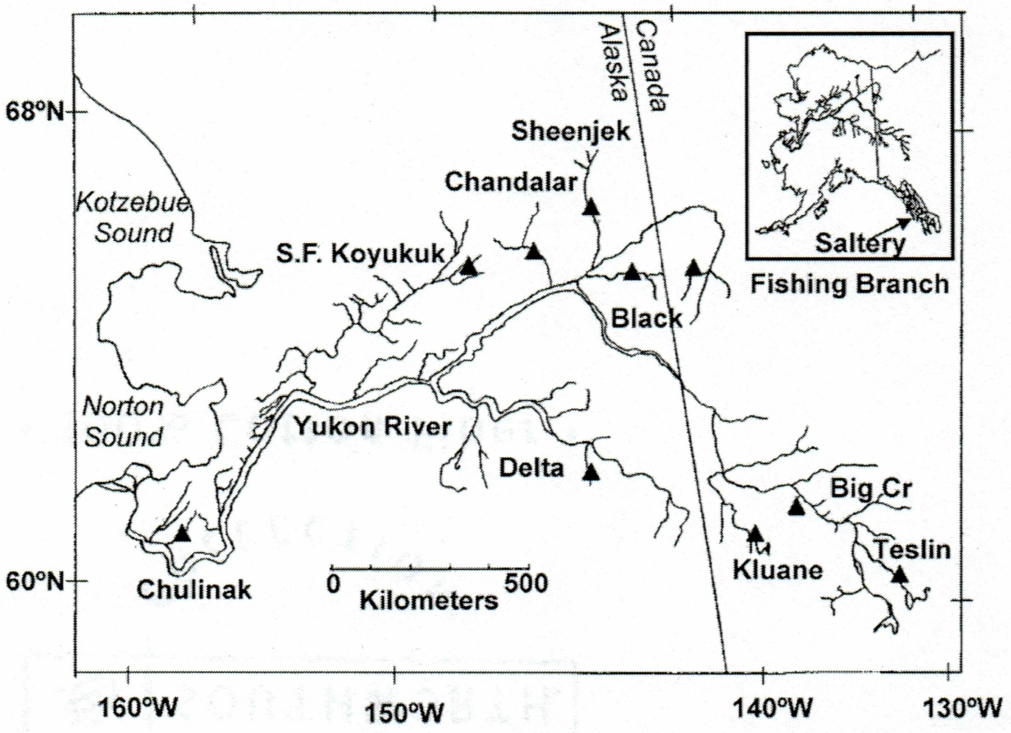

Figure 2.1. Sampling sites: Chulinak (N = 96), S.F. Koyukuk (N = 96), Delta (N = 80), Chandalar (N = 75), Sheenjek (N = 79), Black (N = 96), Fishing Branch (N = 96), Big Cr. (N = 96), Kluane (N = 96), Teslin (N = 96), Saltery Cr. (N = 50).

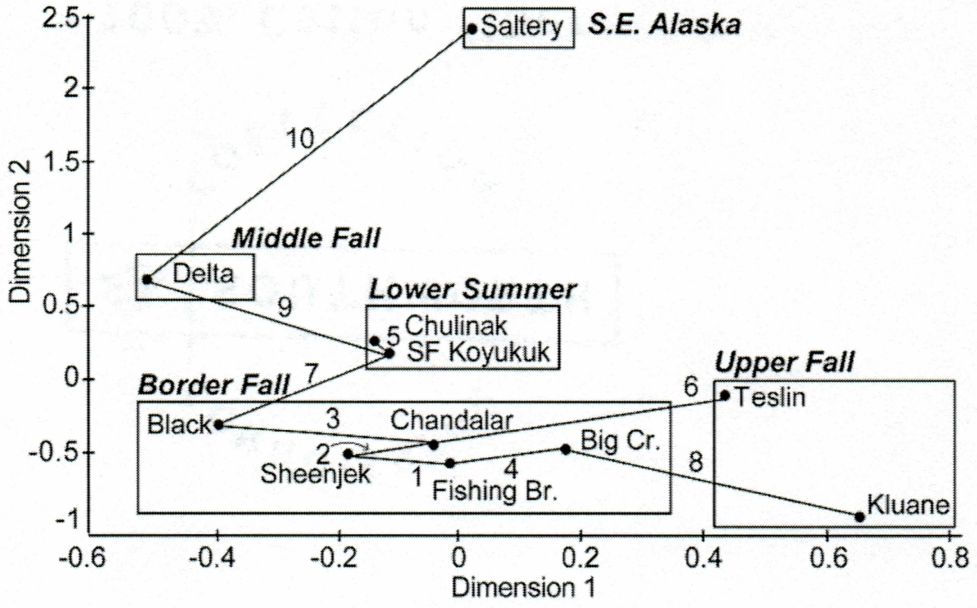

Figure 2.2. Nonmetric multidimensional scaling of chord distances (Cavalli-Sforza and Edwards, 1967; Wright, 1978) among chum salmon populations from the Yukon River and Southeast Alaska. Distances were estimated from allele frequencies at 28 AFLP loci. A minimum spanning tree is superimposed to detect distortions in the ordination process. The numbers 1-10 rank order the relative genetic distances from shortest to longest. Overall stress was 0.02.

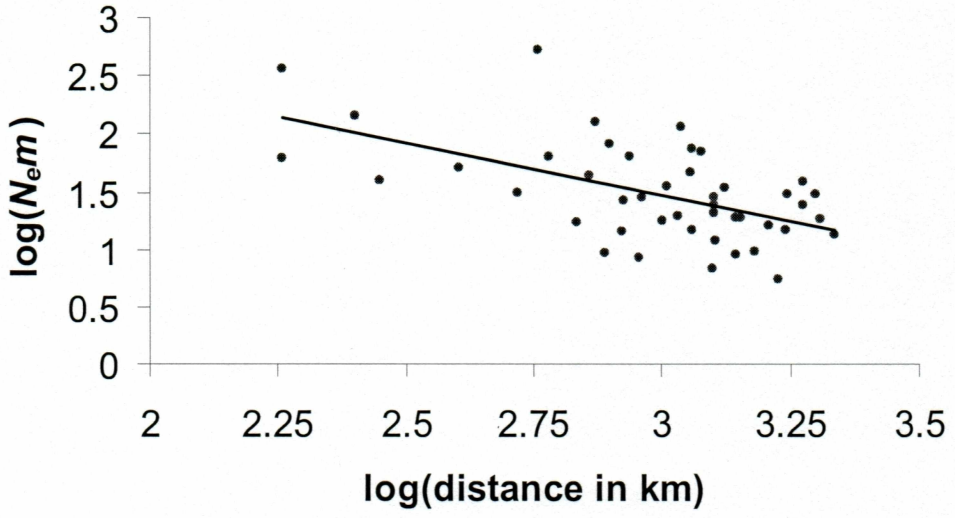

Figure 2.3. Regression analysis of $\log_{10}(N_em)$ on $\log_{10}$(geographic distance in km) separating populations. $N_em = ¼(1/G_{ST} - 1)$ assuming an island model at equilibrium. Linear regression was significant: y = -0.8983 + 4.1614; R^2=0.3163; Mantel test was P=0.002.

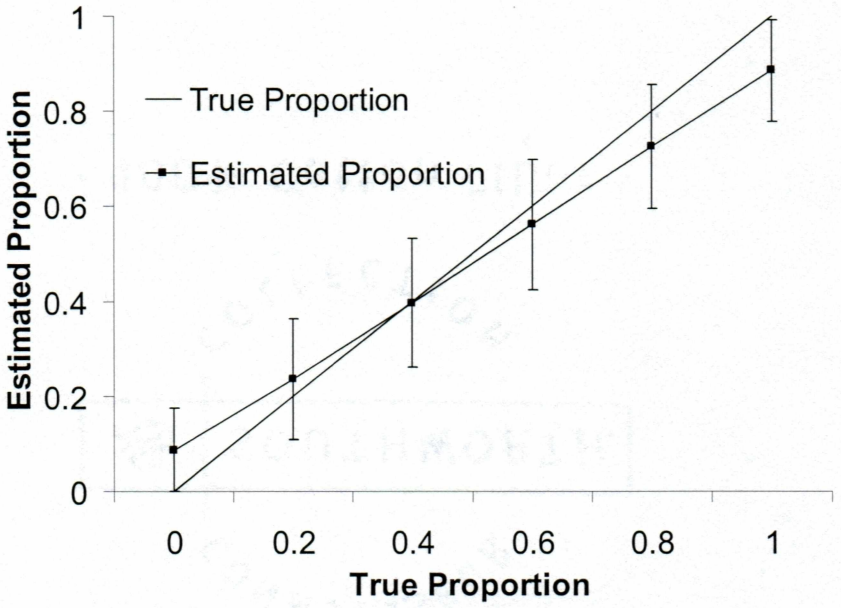

Figure 2.4. Estimated versus true proportions of Canadian-origin Yukon River chum salmon determined by MSA on artificially simulated mixtures (developed from AFLP baseline data) comprised of equal proportions of the Canadian chum salmon populations varying from 0% to 100% in 20% increments. Equal proportions of the United States chum salmon populations made up the difference in the mixtures that had less than 100% Canadian chum salmon. 1000 bootstrap resamples were used to derive the mean estimate. Error bars represent the 95% nonsymmetric confidence interval.

Table 2.1. Hierarchical tests of homogeneity using log-likelihood ratio analysis (Sokal and Rohlf 1995) of 21 AFLP phenotypic marker frequencies among populations within a region, among regions within a run time and between run times. No test indicates that a single population was collected for that region.

Source of variation	df	G
Summer		
Lower	21	46.27*
Fall		
Middle	No test	No test
Border	84	167.86**
Upper	21	97.36**
Within Fall	105	265.22**
Among Fall	42	339.80**
Total Fall	147	605.02**
Within Summer and Fall	126	311.48**
Between Summer and Fall	21	133.42**
Total	189	784.70**

*P<0.05; **P<0.001.

Table 2.2. Hierarchical gene diversity analysis and estimates of effective migration assuming a hierarchical island model at equilibrium for Yukon River chum salmon populations averaged over 28 AFLP loci based on allele frequency estimates.

Source	Gene diversity	Coefficient of gene differentiation (95% conf int)	$N_e m$ (95% conf int)
Within populations	$H_S = 0.298$	$H_S/H_T = 0.979$ (0.958 to 0.993)	
Among populations within regions	$D_{SR} = 0.002$	$G_{SR} = 0.006$* (0.003 to 0.009)	43 (27 to 87)
Among regions within run-times	$D_{RC} = 0.004$	$G_{RC} = 0.013$* (0.003 to 0.032)	5 (2 to 18)
Between run-times	$D_{CT} = 0.001$	$G_{CT} = 0.002$* (-4.6×10^{-4} to 0.004)	33 (-184 to 351)
Total gene diversity	$H_T = 0.304$	$G_{ST} = 0.021$* (0.007 to 0.040)	12 (6 to 34)

*$P<0.001$ inferred from hierarchical tests of homogeneity. 95% confidence intervals estimated by bootstrapping over loci 1000 times.

Table 2.3. Results of allocations to biologically defined regions from MSA of simulated mixtures developed from baseline data collected at 28 AFLP loci. In each simulation, a single source served as contributors to a mixture of 400 fish; estimates of contributions to the mixture were summed for the regions. Regions are represented by equal proportions of fish from populations in the region. The estimates for middle fall and Delta are based on the same simulation. The mixture and baseline were bootstrapped 1000 times. A 100% allocation to the region containing the individual populations would indicate perfect baseline performance.

Source of mixture		Mean (SE) percent by stock region			
Region	Baseline Population	Lower Summer	Middle Fall	Border Fall	Upper Fall
Summer		86.3 (5.9)	0.9 (1.3)	9.9 (5.4)	2.9 (2.9)
	Chulinak	90.0 (5.2)	0.8 (1.2)	7.1 (4.8)	2.1 (2.5)
	SF Koyukuk	83.4 (7.1)	1.2 (1.6)	11.5 (6.5)	3.9 (3.6)
Middle Fall		3.8 (3.3)	89.9 (4.8)	4.9 (3.7)	1.4 (2.0)
	Delta*	3.8 (3.3)	89.9 (4.8)	4.9 (3.7)	1.4 (2.0)
Border Fall		3.7 (3.0)	0.6 (0.8)	90.5 (4.9)	5.2 (4.0)
	Chandalar	2.5 (2.8)	0.4 (0.8)	85.4 (6.9)	11.7 (6.6)
	Black	10.4 (6.3)	1.9 (1.9)	86.1 (6.7)	1.6 (2.4)
	Sheenjek	1.2 (1.8)	0.5 (0.8)	95.7 (3.7)	2.6 (3.1)
	Fishing Br.	0.8 (1.3)	0.2 (0.4)	97.6 (2.6)	1.4 (2.0)
	Big Cr.	2.1 (2.3)	0.6 (0.9)	89.8 (6.0)	7.5 (5.7)
Upper Fall		2.6 (2.6)	0.3 (0.6)	13.9 (6.4)	83.2 (6.7)
	Teslin	2.7 (2.8)	0.7 (1.2)	15.6 (7.2)	81.0 (7.6)
	Kluane	1.7 (2.4)	0.0 (0.0)	8.0 (5.2)	90.3 (5.7)

* There is a single Middle Fall population.

Table 2.4. Results of allocations for the border populations to politically defined regions from MSA of simulated mixtures using baseline data from 28 AFLP loci. In each simulation, a single source served as contributors to a mixture of 400 fish; estimates of contributions to the mixture were summed for the regions. Regions are represented by equal proportions of fish from populations in the region. The mixtures and baselines were bootstrapped 1000 times. A 100% allocation to the region containing the individual populations would indicate perfect baseline performance.

Source of mixture		Mean (SE) percent by stock region				
Region	Baseline population	Lower Summer	Middle Fall	Border United States	Border Canada	Upper Fall
Border United States		4.6 (3.5)	0.8 (1.1)	82.8 (7.0)	7.3 (5.3)	4.5 (3.9)
	Chandalar	2.5 (2.8)	0.4 (0.8)	80.3 (8.0)	5.1 (4.8)	11.7 (6.6)
	Black	10.4 (6.3)	1.9 (1.9)	82.5 (7.6)	3.6 (3.8)	1.6 (2.4)
	Sheenjek	1.2 (1.8)	0.5 (0.8)	79.0 (8.6)	16.7 (7.9)	2.6 (3.1)
Border Canada		1.4 (1.8)	0.3 (0.7)	7.9 (5.8)	86.2 (6.6)	4.2 (3.8)
	Fishing Br.	0.8 (1.3)	0.2 (0.4)	10.2 (7.4)	87.4 (7.6)	1.4 (2.0)
	Big Cr.	2.1 (2.3)	0.6 (0.9)	4.7 (4.8)	85.1 (7.1)	7.5 (5.7)

Table 2.5. Results of allocations for the border populations to politically defined regions from MSA of simulated mixtures using baseline data from 11 microsatellite loci. In each simulation, a single source served as contributors to a mixture of 400 fish; estimates of contributions to the mixture were summed for the regions. Regions are represented by equal proportions of fish from populations in the region. The mixtures and baselines were bootstrapped 1000 times. A 100% allocation to the region containing the individual populations would indicate perfect baseline performance.

Source of mixture		Mean (SE) percent by stock region				
Region	Baseline population	Lower Summer	Middle Fall	Border United States	Border Canada	Upper Fall
Border United States		2.6 (1.6)	1.5 (1.5)	82.4 (5.0)	11.9 (4.4)	1.6 (1.3)
	Chandalar	2.8 (2.1)	1.2 (1.3)	82.7 (5.9)	11.3 (5.0)	2.0 (1.7)
	Black	2.4 (1.7)	1.4 (1.4)	83.3 (5.6)	11.1 (5.0)	1.8 (1.4)
	Sheenjek	2.3 (1.8)	2.0 (1.8)	86.4 (5.3)	8.1 (4.2)	1.2 (1.0)
Border Canada		2.2 (1.4)	1.7 (1.4)	14.4 (4.7)	79.7 (5.1)	2.0 (1.4)
	Fishing Br.	1.6 (1.4)	1.1 (1.2)	13.0 (4.9)	82.1 (5.5)	2.2 (1.6)
	Big Cr.	2.3 (1.9)	2.4 (1.9)	14.3 (5.1)	78.4 (5.9)	2.6 (1.8)

Appendix 2.1. Hierarchical tests of homogeneity using log-likelihood ratio analysis (Sokal and Rohlf 1995) of 21 AFLP phenotypic marker frequencies among populations with a region, among regions within a run time and between run times. No test indicates that a single population was collected for that region. Shaded area indicates that locus was significant ($P<0.05$) after correction for multiple tests (N/α).

	ACC CAC									
	54		63		92		123		162	
Source of variation	df	*G*	df	*G*	df	*G*	df	*G*	df	*G*
Summer										
Lower	1	1.58	1	6.73	1	0.77	1	14.00	1	0.32
Fall										
Middle		No test		No test		No test		No test		No test
United States/Canada	4	4.01	4	0.84	4	0.21	4	2.09	4	9.11
Upper	1	5.64	1	6.19	1	0.61	1	4.90	1	9.09
Within Fall	5	9.65	5	7.03	5	0.83	5	6.99	5	18.20
Among Fall	2	4.19	2	5.82	2	15.33	2	3.53	2	8.12
Total Fall	7	13.83	7	12.84	7	16.15	7	10.52	7	26.32
Between Summer and Fall	1	9.54	1	1.98	1	2.08	1	1.28	1	19.99
Total	9	24.96	9	21.56	9	19.01	9	25.79	9	46.63

Appendix 2.1. *(Continued)*

	ACC CAC									
	165		171		173		209		217	
Source of variation	df	*G*	df	*G*	df	*G*	df	*G*	df	*G*
Summer										
Lower	1	2.81	1	0.07	1	0.33	1	1.53	1	0.00
Fall										
Middle		No test		No test		No test		No test		No test
United States/Canada	4	2.94	4	5.38	4	3.14	4	12.80	4	32.83
Upper	1	12.19	1	0.04	1	0.02	1	0.25	1	0.87
Within Fall	5	15.13	5	5.42	5	3.16	5	13.05	5	33.70
Among Fall	2	8.52	2	13.61	2	8.67	2	7.38	2	23.84
Total Fall	7	23.64	7	19.03	7	11.83	7	20.43	7	57.54
Between Summer and Fall	1	12.72	1	0.87	1	2.05	1	0.03	1	24.93
Total	9	39.17	9	19.97	9	14.21	9	21.98	9	82.48

Appendix 2.1. *(Continued)*

	ACC CAC						AAG CAC					
	297		364		440		81		98		364	
Source of variation	df	*G*	df	*G*	df	*G*	df	*G*	df	*G*	df	*G*
Summer												
Lower	1	0.52	1	1.81	1	4.33	1	1.38	1	0.01	1	1.81
Fall												
Middle		No test		No test		No test		No test		No test		No test
United States/Canada	4	7.19	4	15.97	4	1.90	4	4.65	4	4.47	4	15.97
Upper	1	2.52	1	14.10	1	0.99	1	0.08	1	14.32	1	14.10
Within Fall	5	9.71	5	30.07	5	2.89	5	4.73	5	18.79	5	30.07
Among Fall	2	6.87	2	6.88	2	1.82	2	0.68	2	2.80	2	6.88
Total Fall	7	16.59	7	36.94	7	4.71	7	5.40	7	21.59	7	36.94
Between Summer and Fall	1	2.06	1	11.66	1	11.92	1	12.01	1	0.78	1	11.66
Total	9	19.17	9	50.41	9	20.95	9	18.80	9	22.39	9	50.41

Appendix 2.1. *(Continued)*

	AAG CAC											
	122		167		241		252		324		364	
Source of variation	df	*G*	df	*G*	df	*G*	df	*G*	df	*G*	df	*G*
Summer												
Lower	1	0.02	1	0.07	1	4.91	1	0.35	1	4.69	1	1.81
Fall												
Middle												
United States/Canada	4	5.09	4	8.60	4	10.49	4	10.65	4	19.26	4	15.97
Upper	1	0.83	1	20.19	1	0.47	1	3.95	1	0.07	1	14.10
Within Fall	5	5.92	5	28.80	5	10.97	5	14.60	5	19.33	5	30.07
Among Fall	2	7.23	2	23.46	2	6.20	2	178.55	2	1.89	2	6.88
Total Fall	7	13.14	7	52.26	7	17.16	7	193.15	7	21.22	7	36.94
Between Summer and Fall	1	8.99	1	0.15	1	4.74	1	0.27	1	0.69	1	11.66
Total	9	22.16	9	52.47	9	26.82	9	193.77	9	26.60	9	50.41

Appendix 2.1 (*Continued)*

Source of variation	AAG CAC 394 df	*G*
Summer		
Lower	1	0.05
Fall		
Middle		No test
United States/Canada	4	6.25
Upper	1	0.03
Within Fall	5	6.28
Among Fall	2	4.43
Total Fall	7	10.70
Between Summer and Fall	1	4.67
Total	9	15.42

References

Allendorf, F.W., N. Ryman, and F.M. Utter. 1987. Genetics and Fishery Management. pp. 1-19. *In*: N. Ryman and F. Utter (ed.) Population Genetics and Fishery Management. University of Washington Press, Seattle, Washington.

Apostal, B.L., W.C. Black IV, P. Reiter, and B.R. Miller. 1996. Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. Heredity 76: 325-334.

Beacham, T.D. 1996. The use of minisatellite DNA variation for stock identification of chum salmon, *Oncorhynchus nerka*. Fishery Bulletin 94:611-627.

Beacham, T.D., A.P. Gould, R.E. Withler, C.B. Murray, and L.W. Barner. 1987. Biochemical genetic survey and stock identification of chum salmon (*Oncorhynchus keta*) in British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 44:1702-1713.

Beacham, T.D., C.B. Murray, and R.E. Withler. 1988. Age, morphology, developmental biology, and biochemical genetic variation of Yukon River fall chum salmon, *Oncorhynchus keta*, and comparisons with British Columbia populations. Fishery Bulletin 86:663-674.

Bernatchez, L., and P. Duchesne. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Canadian Journal of Fisheries and Aquatic Sciences 57:1-12.

Black, W.C., IV. 1997. FORTRAN programs for the analysis of RAPD-PCR markers in populations. Colorado State University, Fort Collins, CO. 4 pp.

Cavalli-Sforza, L.L., and A.W.F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Evolution 21:550-570.

Chakraborty, R., and O. Leimar. 1987. Genetic variation within a subdivided population. pp. 89-120. *In*: N. Ryman and F. Utter (ed.) Population Genetics and Fishery Management. University of Washington Press, Seattle, Washington.

Cooper, D.W. 1968. The significance level in multiple tests made simultaneously. Heredity 23:614-617.

Crane, P.A., W.J. Spearman, L.W. Seeb. 2001. Yukon River chum salmon: report for genetic stock identification studies, 1992-1997. Regional Information Report No. 5J01-08. 28 pp.

Debevec, E.M., R.B. Gates, M. Masuda, J. Pella, J. Reynolds, and L.W. Seeb. 2000. SPAM (Version 3.2): statistics program for analyzing mixtures. Journal of Heredity 91:509-510.

Eggers, D.M. 2001. Biological escapement goals for Yukon River fall chum salmon. Regional Information Report No. 3A01-10. Alaska Dept. of Fish and Game. Anchorage, AK. 77 pp.

Ewens, WJ. 1983. The role of models in the analysis of molecular genetic data with particular reference to restriction fragment data. pp. 45-73. *In*: B.S. Weir (ed.) Statistical Analysis of DNA Sequence Data. Marcel Dekker, New York.

Gharrett, A.J., and L.A. Zhivotovsky. 2003. Migration. pp. 141-174. *In*: E.M. Hallerman (ed.) Population genetics: principles and applications for fisheries scientists. American Fisheries Society, Bethesda, Maryland.

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

Hawkins, S.L., and nine coauthors. 2002. Population structure of odd-broodline Asian pink salmon and its contrast to the even-broodline structure. Journal of Fish Biology 60:370-388.

Hutchinson, D.W., and A.R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift of the distribution of genetic variability. Evolution 53:1898-1914.

JTC (The United States and Canada Yukon River Joint Technical Committee). 2001. Yukon River salmon season review for 2001 and technical committee report. JTC, Whitehorse, Yukon. 158 pp.

Kondzela, C.M., C.M. Guthrie, S.L. Hawkins, C.D. Russell, and J.H. Helle. 1994. Genetic relationships among chum salmon populations in Southeast Alaska and Northern British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 51(Suppl. 1):50-64.

Krauss, S.L. 2000. Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. Molecular Ecology 9:1241-1245.

Kruskal, J.B. 1964. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:1-27.

Larkin, P.A. 1981. A perspective on population genetics and salmon management. Canadian Journal of Fisheries and Aquatic Sciences 38:1469-1475.

Mantel, N.A. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Nei, M. 1978. Estimation of heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Nei, M., and R.K. Chesser. 1983. Estimation of fixation indices and gene diversities. Annals of Human Genetics 47: 253-259.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York. 333 pp.

Neigel, J.E. 2002. Is F_{ST} obsolete? Conservation Genetics 3:167-173.

Parker, P.G., A.A. Snow, M.D. Schug, G.C. Booton, and P.A. Fuerst. 1998. What molecules can tell us about populations: choosing and using a molecular marker. Ecology 79:381-382.

Perkin-Elmer Corporation. 1997. AFLP plant mapping. Perkin-Elmer Corporation, Foster City, CA. 45 pp.

Phelps, S.R., L. LeClair, S. Young, and L. Blankenship. 1994. Genetic diversity patterns of chum salmon in the Pacific Northwest. Canadian Journal of Fisheries and Aquatic Sciences 51(Suppl. 1):65-83.

Rohlf, F.J. 2000. NTSYS-pc: numerical taxonomy and multivariate analysis systems. Exeter Software. Setauket, NY. 38 pp.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin version 2.000: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Scribner, K.T., P.A. Crane, W.J. Spearman, and L.W. Seeb. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 55: 1748-1758.

Seeb, L.W., and P.A. Crane. 1999. High genetic heterogeneity in chum salmon in western Alaska, the contact zone between northern and southern lineages. Transactions of the American Fisheries Society 128:58-87

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264-279.

Smouse, P.E., and R.H. Ward. 1978. A comparison of the genetic infrastructure of the Ye'cuana and the Yanomama: a likelihood analysis of genotypic variation among populations. Genetics 88:611-631.

Sokal, R.R., and F.J. Rohlf. 1995. Biometry. W.H. Freeman. New York. 887 pp.

Spearman, W.J., and S.J. Miller. 1997. Mixed-stock analysis of chum salmon (*Oncorhynchus keta*) from the Yukon River district 5 subsistence fishery. Alaska Fisheries Technical Report No. 40. U.S. Fish and Wildlife Service, 1011 E. Tudor Road, Anchorage, AK. 21 pp.

Templeton, A.R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Molecular Ecology 7:381-397.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Vekemans, X., T. Beauwens, M. Lemaire, and I. Roldan-Ruiz. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular Ecology 11:139-151.

Waples, R.S., and P.E. Smouse. 1990. Gametic disequilibrium analysis as a means of identifying mixtures of salmon populations. American Fisheries Society Symposium 7:439-458.

Wilmot, R.L., R.J. Everett, W.J. Spearman, and R. Baccus. 1992. Mixed-stock analysis of Yukon River chum and chinook salmon – 1987-1990. Progress Report U.S. Fish and Wildlife Service, Anchorage, AK. 132 pp.

Wilmot R.L., R.J. Everett, W.J. Spearman, R. Baccus, N.V. Varnavskaya, and S.V. Putivkin. 1994. Genetic stock structure of western Alaska chum salmon and a comparison with Russian far east stocks. Canadian Journal of Fisheries and Aquatic Sciences 51(Suppl. 1):84-94.

Winans, G.A., P.B. Aebersold, S. Urawa, and N. V. Varnavskaya. 1994. Determining continent of origin of chum salmon (*Oncorhynchus keta*) using mixed-stock analysis techniques: status of allozyme baseline in Asia. Canadian Journal of Fisheries and Aquatic Sciences 51(Suppl. 1):95-113.

Wood, C.C., S. McKinnell, T.J. Mulligan, and D.A. Fournier. 1987. Stock identification with the maximum-likelihood mixture model: sensitivity analysis and application to complex problems. Canadian Journal of Fisheries and Aquatic Sciences 44: 866-881.

Wright, S. 1978. Evolution and the genetics of populations: Volume 4, variability within and among natural populations. The University of Chicago Press, Chicago. 580 pp.

Zhivotovsky, L.A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology 8:907-913.

Zhivotovsky, L.A., A.J. Gharrett, A.J. McGregor, M.K. Glubokovsky, and M.W. Feldman. 1994. Gene differentiation in Pacific salmon (*Oncorhynchus sp.*): facts and models with reference to pink salmon (*O. gorbuscha*). Canadian Journal of Fisheries and Aquatic Sciences 51(Suppl. 1):223-232.

General Conclusion

Yukon River chum salmon exhibited sufficient variation with AFLP markers, but not mtDNA markers, to enable accurate and precise MSA estimates to country of origin for the border populations. Whereas mtDNA does not hold promise for MSA of Yukon River chum salmon, it does have unique properties that make it useful for addressing demographic questions that may allow a better understanding of chum salmon life history.

The paradigm of salmon management is to harvest populations individually; however, this model is not applicable to most fisheries as salmon populations do not segregate until on the spawning grounds. Harvesting mixtures of populations is the rule for salmon; hence, MSA utilizing AFLP markers offers managers a tool to prevent differential harvest rates. Moreover, MSA may allow for the prosecution of run surpluses previously unidentified because of inadequate inseason run assessment tools and the necessity for conservative management.