North American Journal of Fisheries Management 25:944–953, 2005

DOI: 10.1577/M04-143.1

[Article]

Single Nucleotide Polymorphisms Provide Rapid and Accurate Estimates of the Proportions of U.S. and Canadian Chinook Salmon Caught in Yukon River Fisheries

CHRISTIAN T. SMITH,* WILLIAM D. TEMPLIN, JAMES E. SEEB, AND LISA W. SEEB

Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, Alaska 99518, USA

Abstract.—As anadromous Chinook salmon *Oncorhynchus tshawytscha* bound for U.S. and Canadian spawning grounds migrate through the U.S. portion of the Yukon River, they are targeted by several fisheries. To fulfill treaty obligations between the two countries, fishery managers need to know what portion of fish caught in the United States are of Canadian origin. Allozyme markers have been used to assign individuals in mixed fishery samples to U.S. and Canadian portions of the Yukon River; however, these markers are limited by sampling difficulties and by the number of available loci. Microsatellite DNA markers have been considered as an alternative; however, microsatellite data are not readily transportable among laboratories or countries. Here we present the use of single nucleotide polymorphism (SNP) markers that combine the ease of sampling and large potential number of loci of other DNA markers with universally transportable data. Simulations and analyses of known fish suggest that the SNP baseline can be used to assign fish to country of origin with more than 95% accuracy. Assignments based on SNP data are largely concordant with those based on allozyme data. The SNP baseline described here may be used to provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook salmon caught in Yukon River fisheries.

Chinook salmon *Oncorhynchus tshawytscha* are anadromous; individuals migrate hundreds or even thousands of kilometers through freshwater to arrive at natal spawning sites. Upriver migrations occur at times of the year that are characteristic of each region and river and can consist of tens or even hundreds of thousands of individuals swimming through constricted bodies of freshwater over a period of several weeks (Healy 1991). Estuarine and in-river fisheries target these runs for sport, commercial, and subsistence purposes. In many cases, these runs are composed of a mixture of individuals from different stocks bound for multiple tributaries or localities within a river. Variations in environmental conditions and magnitudes of spawning habitats within drainage systems result in differing abundances of these components and thus varying amounts of the fishing pressure they can withstand. To ensure adequate escapement to all the individual management units within a drainage, knowledge of the spatial and temporal composition of the run as it is being targeted by each fishery is highly desirable.

* Corresponding author: christian_smith@fishgame.state.ak.us

Received August 13, 2004; accepted January 27, 2005
Published online July 20, 2005

The Yukon River is the third largest river in North America, originating in British Columbia and draining through the Yukon Territory and Alaska before emptying into the Bering Sea. The watershed covers approximately 840,000 km^2, of which more than a third is located in Canada. Chinook salmon spawn in tributaries of the Yukon River from near the river's mouth to more than 3,200 km upstream (Healy 1991). Subsistence fisheries targeting Chinook salmon in the Yukon River have played a critical role in the survival and culture of people inhabiting this region for thousands of years. Sport and commercial fisheries have also become very important to communities surrounding this drainage over the past century. Because many of these fisheries are located in Alaska but many of the fish being caught are of Canadian origin, the sustainability of these fisheries depends on cooperation between management authorities of United States and Canada.

Effective December 2002, the Yukon River Salmon Agreement was included as an annex of the Pacific Salmon Treaty between the United States and Canada. Under this agreement, both nations agreed that harvest sharing of salmon stocks would be managed according to the principles of precautionary abundance-based management and that each side would manage their fisheries both

to ensure that sufficient fish are available to meet escapement requirements and, whenever possible, to provide for subsistence and commercial harvests. A bilateral Joint Technical Committee was conceived to coordinate research that would further these ends. One of the key pieces of information required by the Joint Technical Committee is the proportion of the harvest in U.S. fisheries that is bound for Canadian spawning grounds.

Genetic mixture analyses provide one means by which fishery managers may assess the contributions of putative populations to mixed stock samples (e.g., Utter et al. 1987; Shaklee et al. 1999; Seeb et al. 2000). Genetic markers are first examined in spawning populations to develop a comprehensive baseline. Once a baseline is developed, mixture samples from fisheries or other collections are examined. Finally, statistical methods are used to determine the most likely composition of the mixture, given the baseline data. Early genetic surveys revealed significant variation among Yukon River Chinook salmon (Gharrett et al. 1987; Beacham et al. 1989), suggesting the potential for genetic mixture analyses of this group. An allozyme baseline was subsequently constructed and used to provide managers with estimates of the Canadian proportion of U.S. test fisheries (Templin et al. 2005). Limitations of allozyme markers include the requirement of lethal sampling and cryopreservation and a limited number of loci. The advent of polymerase chain reaction (PCR) analysis opened the way for a large number of potential DNA markers that do not require lethal sampling or cryopreservation of members of the test population. Microsatellite DNA baselines are presently under construction to study a range of aspects of the biology of Chinook salmon on the Yukon River, as well as to determine the country of origin of fish caught in Alaskan fisheries. Although microsatellites can provide a very large amount of information, their use requires spending time and funds on standardization if data are to be compared or combined across laboratories (or across countries in the present context).

A single nucleotide polymorphism (SNP) is a polymorphism arising by change of a single nucleotide base between the DNA sequences of two taxa. Because each SNP generally has only two possible alleles, an average SNP will not provide as much information as an average microsatellite (Rosenberg et al. 2003). Nonetheless, advances in genotyping chemistries have compensated for this somewhat by allowing SNP data to be collected and scored very rapidly in comparison with other genetic markers. An attribute of SNPs that make them appropriate for multijurisdictional management issues, such as Yukon River Chinook salmon management, is the fact that SNP data are readily compared and combinable across laboratories. Here we describe the creation of a SNP baseline for Yukon River Chinook salmon and application of the baseline in determining the composition of fisheries to meet treaty obligations under the Yukon River Salmon Agreement.

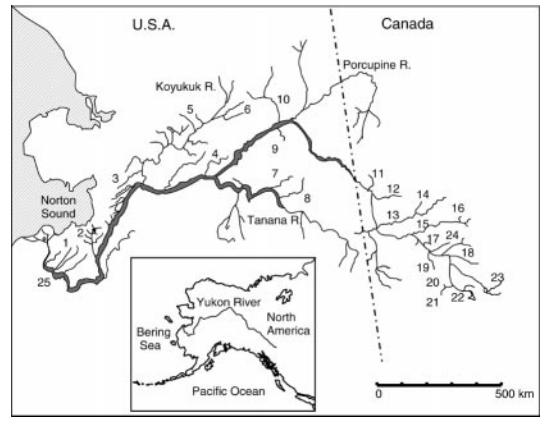

FIGURE 1.—Map of the Yukon River showing some of the major tributaries as well as all of the localities from which the present baseline and mixture samples were taken. Locality names are listed in Table 1.

Methods

Samples of spawning populations from the United States (Figure 1; Table 1) were collected as pieces of tissue about 5 mm^2 and stored in 95% solutions of ethanol. Samples from the Canadian portion of the Yukon River were provided to us as DNA that had been prepared by the Chelex protocol described by Small et al. (1998). In addition to the baseline samples, a sample consisting of DNA from 95 fish from Tatchun River plus Little Salmon River (Figure 1; Table 1) was provided by the Canadian Department of Fisheries and Oceans.

U.S. test fishery samples were collected during the summer of 2003 at Pilot Station, Alaska (approximately river km 196; Figure 1) as 5-mm^2 fin clips preserved in 95% ethanol at ambient temperature. The test fishery samples were divided into three subsamples based on the dates on which they were caught: period 1, June 10–14; period 2, June 15–20; and period 3, June 21–July 17. The lengths of the fishery periods were determined arbitrarily in an effort to compensate for the uneven rates at which fish were caught during the sam-

TABLE 1.—Allocation of Chinook salmon from the Yukon River to country, region, and locality of origin. Collection locality numbers correspond to those in Figure 1. All samples represent adult fish, except one consisting of eight juveniles taken from Mayo River in 2003 (denoted 2003j). Percent correct allocations for each collection are based on simulations for ''baseline'' collections and on fish of known origin for ''mixture'' collections. The origin of the test fishery samples was unknown, so correct allocations could not be calculated for those samples.

Collection	Year or period	*n*	Correct allocation		
			Country	Region	Locality
Baseline					
United States			0.96		
Lower U.S. river				0.94	
1. Andreafsky River	2003	203			0.85
2. Anvik River	2003	85			0.81
3. Gisasa River	2001	179			0.75
4. Tozitna River	2003	190			0.80
Middle U.S. river				0.94	
5. Henshaw Creek	2001	90			0.79
6. South Fork Koyukuk River	2003	54			0.79
7. Chena River	2001	187			0.91
8. Salcha River	2003	55			0.76
9. Beaver Creek	1997	94			0.82
10. Chandalar River	2003	111			0.80
Canada			0.97		
Canadian border				0.93	
11. Chandindu River	2001	84			0.68
12. Klondike River	1995	5			0.62
	2001	10			
	2003	68			
Pelly River drainage				0.85	
13. Stewart River	1997	99			0.45
14. Mayo River	1992	70			0.57
	1997	32			
	2003	30			
	2003j	8			
15. Pelly River	1997	112			0.76
16. Blind Creek	2003	94			0.77
Upper Yukon River				0.93	
17. Tatchun Creek	1987	27			0.69
	1996	105			
	1997	31			
18. Big Salmon River	1987	76			0.83
	1997	36			
19. Nordenskiold River	2003	56			0.83
23. Nisutlin River	1987	17			0.80
Takhini River drainage				0.93	
20. Stoney Creek	1993	94			0.68
21. Takhini River	2002	64			0.75
	1997	62			
	2003	33			
Whitehorse				0.88	
22. Whitehorse	1985	39			0.88
	1997	114			
	1997	39			
Mixtures					
United States					
Lower U.S. river					
1. Andreafsky River	2003	95	1.00	1.00	64
Canada					
Upper Yukon River					
17. Tatchun Creek and 24. Little Salmon River	1997	95	1.00	0.92	n/a
Test fishery samples					
United States					
Lower U.S. river					
25. Pilot Station	Period 1	227			
	Period 2	197			
	Period 3	163			

pling: period 2 was extended to 6 days to allow collection in this period to approach 200 individuals; period 3, which represents the tail end of the run, lasted several weeks. Genomic DNA was extracted from each tissue sample and purified from each Chelex DNA sample by using DNAeasy 96 columns (QIAGEN, Valencia, California).

SNP genotyping using the 5′-nuclease reaction.—Ten SNP loci (*Ots_Prl2*, *Ots_P53*, *Ots_GH2*, *Ots_MHC2*, *Ots_Ots2*, *Ots_C3N3*, *Ots_MHC1*, *Ots_Tnsf*, *Ots_SL*, and *Ots_P450*) were genotyped according to the methods described by Smith et al. (2005). Amplification was conducted with four 384-well thermal cycler blocks on a DNA Engine Tetrad (MJ Research, Waltham, Massachusetts) plus one 384-well thermal cycler block in an ABI7900 real-time PCR instrument (Applied Biosystems Inc., Foster City, California). End-point analysis of each 384-well plate was performed with an ABI7900. To score alleles, we used Sequence Detection Software 2.1 (Applied Biosystems Inc.). To ensure data integrity, each run was scored independently by two researchers.

Population structure.—GENEPOP 3.4 (Raymond and Rousset 1995a) was used to test for differences between observed genotypic frequencies in each baseline sample and those expected under Hardy–Weinberg equilibrium. Statistical significance levels were corrected for multiple simultaneous comparisons as described by Rice (1989); that is, $\alpha = 0.05/10$ loci $= 0.005$. For sites from which multiple samples were taken (Table 1), we tested for differences in allele frequencies among those samples by using the probability test for population differentiation (Raymond and Rousset 1995b), as implemented in GENEPOP. Statistical significance levels were again corrected for multiple simultaneous comparisons within each collection site (e.g., for a site from which three pairwise comparisons between samples were made, $\alpha = 0.05/3 = 0.017$). Samples taken from a single site and not exhibiting significant allele frequency differences from one another were pooled for subsequent analyses. GENEPOP was also used to estimate F_{ST} for each locus across populations and F_{IS} for each locus in each collection following the methods of Weir and Cockerham (1984). PHYLIP 3.6 (Felsenstein 1995) was used to calculate genetic distances (Cavalli-Sforza and Edwards 1967) between all samples and to perform a neighbor-joining analysis based on those distances.

Genetic stock identification.—To evaluate the potential utility of the SNP baseline for genetic stock identification, we considered three different types of mixture samples: (1) computer-generated (simulated) fish, (2) individuals of known population origin, and (3) test fishery samples (unknown origin). For all estimates of stock contribution, we used the conditional maximum likelihood model implemented in SPAM 3.7 (Debevec et al. 2000). Reporting regions were defined at three hierarchical levels: individual populations (collection localities), broader geographic regions (suggested by allozyme data [Templin et al., 2005] and by the present neighbor-joining analysis), and national origin. Individual population estimates were first calculated and then summed into larger reporting regions. Reporting regions with mean correct estimates of 90% or better are considered highly identifiable in fishery applications. The top level of this hierarchy (national origin) is directly relevant to the Yukon River Salmon Agreement and is thus the level we used to evaluate the practical utility of the SNP baseline.

For the purpose of assigning simulated fish, each mixture ($N = 400$) was composed entirely (100%) of the reporting group under study, and all stocks in the reporting group contributed equally to the mixture. Average estimates of mixture proportions and 90% confidence intervals for the reporting group were derived from 1,000 simulations in which baseline and mixture genotypes were randomly generated from the baseline allele frequencies, assuming a Hardy–Weinberg equilibrium among the genotypes. The 1,000 estimates for a region were sorted from lowest to highest, taking the 51st and 950th values in the sequence as the lower and upper bounds, respectively, of the 90% confidence interval for that region. To examine the effect of unequal contributions to simulated mixtures, we repeated the above simulations but substituted a range (0.00, 0.10, 0.30, 0.50, 0.70, 0.90, and 1.00) of contributions from fish of U.S. origin.

We evaluated two mixtures consisting of individuals of known origin. The first, a collection of 95 Canadian fish that originated from the Little Salmon and Tatchun rivers, was not included in the baseline. The second was a randomly selected subsample of 95 individuals from the Andreafsky River, Alaska; in this case the baseline allele frequencies were recalculated without the selected individuals before mixture analysis.

The test fishery samples were those collected from Pilot Station (approximately river km 196; Figure 1).

We computed the 90% confidence intervals for all regional contribution estimates to real samples

TABLE 2.—Frequencies of one allele observed at each locus in samples of Yukon River Chinook salmon and the numbers of fish (*n*) upon which the estimates were based. Allele names indicate the variable bases (adenine [A], guanine [G], or thymine [T]) in the corresponding probe sets, as described by Smith et al. (2005).

	Ots_GH2		*Ots_Prl2*		*Ots_Tnsf*		*Ots_Ots2*		*Ots_MHC1*	
Collection and statistic	*n*	A	*n*	A	*n*	A	*n*	G	*n*	A
Andreafsky River	188	0.835	200	0.663	183	0.257	199	0.859	186	0.379
Anvik River	81	0.870	72	0.708	84	0.125	79	0.873	62	0.395
Gisasa River	170	0.715	174	0.601	174	0.172	174	0.891	169	0.485
Tozitna River	163	0.632	175	0.514	184	0.266	190	0.961	185	0.438
Henshaw Creek	90	0.383	89	0.360	89	0.354	89	0.989	88	0.426
South Fork Koyukuk River	50	0.460	48	0.438	52	0.413	52	0.971	52	0.317
Chena River	174	0.514	178	0.393	170	0.550	178	0.997	177	0.511
Salcha River	52	0.596	51	0.402	54	0.546	54	1.000	54	0.370
Beaver Creek	91	0.538	84	0.554	91	0.319	68	1.000	84	0.304
Chandalar River	101	0.421	105	0.338	108	0.157	106	0.986	100	0.470
Chandindu River	84	0.286	83	0.355	83	0.470	84	0.976	84	0.256
Klondike River	82	0.274	81	0.370	75	0.513	80	0.975	81	0.253
Stewart River	96	0.370	97	0.278	96	0.177	99	0.939	95	0.211
Mayo River	132	0.417	124	0.266	139	0.194	136	0.930	132	0.28
Pelly River	110	0.282	110	0.232	111	0.194	111	0.964	108	0.264
Blind Creek	93	0.306	94	0.362	94	0.191	94	0.957	94	0.202
Tatchun Creek	159	0.362	161	0.289	154	0.172	163	0.954	147	0.466
Big Salmon River	111	0.216	108	0.347	108	0.13	111	0.869	108	0.435
Nordenskiold River	56	0.223	56	0.259	56	0.125	56	0.866	55	0.436
Stoney Creek	94	0.404	90	0.394	93	0.108	93	0.962	94	0.633
Takhini River	155	0.452	155	0.477	155	0.155	156	0.971	158	0.674
Whitehorse	142	0.567	142	0.317	144	0.191	152	0.967	139	0.439
Nisutlin River	55	0.445	53	0.274	51	0.127	56	0.929	54	0.472
F_{ST}		0.132		0.073		0.094		0.035		0.061

by using 1,000 bootstrap replicates of the baseline and mixture genotypes. For each resample, contribution estimates were generated for all populations and summed to the regional level. Bounds for 90% confidence intervals were calculated as described above.

Results

SNP Genotyping Using the 5′-Nuclease Reaction

Using five 384-well thermal cycler blocks four times a day, we were able to process more than 7,000 genotypes. Comparison of the number of samples run (Table 1) with the number of samples on which allele frequencies were calculated (Table 2) reveals that the PCR failure rate per marker ranged from 2.0% for *Ots_MHC2* to 5.8% for *Ots_Prl2*. Because the scoring of SNP genotypes is largely automated, scoring took approximately 5 min per 384-well plate. The initial baseline took 3 weeks to complete.

Population Structure

Nine of the 10 SNPs examined were polymorphic in the Yukon River; the exception, mitochondrial DNA SNP *Ots_C3N3*, was fixed for a single haplotype in all collections. The Anvik River, Alaska, sample exhibited a significant ($P = 0.001$) departure from genotypic frequencies expected under Hardy–Weinberg equilibrium. Inspection of the data revealed that this anomaly resulted from a homozygote excess at *Ots_Ots2* ($F_{IS} = 0.547$). No other departures from Hardy–Weinberg expectations were detected. Probability tests for population differentiation did not reveal any significant differences between samples taken from any locality in different years. Multiple samples collected from each locality were therefore pooled for that locale for further analyses. For the nine polymorphic loci observed in this study the F_{ST} estimates ranged from 0.017 at *Ots_P450* to 0.132 at *Ots_GH2* (Table 2).

The neighbor-joining analysis divided the Chinook salmon samples among seven major branches (Figure 2): lower U.S. river (Andreafsky, Anvik, Gisasa, and Tozitna rivers), middle U.S. river (Henshaw Creek, South Fork Koyukuk River, Chena River, Salcha River, Beaver Creek, and Chandalar River), Canada border (Chandindu and Klondike rivers), Pelly River drainage (Stewart, Mayo, and Pelly rivers and Blind Creek), upper Yukon River (Tatchun Creek and Big Salmon, Nordenskiold, and Nisutlin rivers), Takhini River drainage (Stoney Creek and Takhini River), and Whitehorse. Each of these branches contained ei-

TABLE 2.—Extended.

Collection and statistic	*Ots_P53*		*Ots_MHC2*		*Ots_P450*		*Ots_SL*		*Ots_C3N3*	
	n	A	*n*	T	*n*	A	*n*	A	*n*	G
Andreafsky River	193	0.578	195	0.974	203	0.232	203	0.714	203	1.000
Anvik River	77	0.494	85	0.988	81	0.278	80	0.756	85	1.000
Gisasa River	170	0.544	179	0.972	175	0.240	173	0.815	174	1.000
Tozitna River	187	0.468	189	0.944	189	0.185	174	0.764	189	1.000
Henshaw Creek	87	0.598	89	1.000	90	0.161	90	0.717	89	1.000
South Fork Koyukuk River	52	0.779	51	1.000	53	0.245	51	0.902	51	1.000
Chena River	173	0.749	183	1.000	173	0.124	179	0.824	185	1.000
Salcha River	51	0.667	54	1.000	54	0.241	54	0.769	50	1.000
Beaver Creek	91	0.621	94	0.984	86	0.267	90	0.772	91	1.000
Chandalar River	101	0.564	107	0.981	107	0.248	107	0.734	111	1.000
Chandindu River	84	0.286	83	1.000	83	0.307	83	0.825	83	1.000
Klondike River	73	0.274	81	0.988	80	0.262	80	0.825	83	1.000
Stewart River	96	0.448	95	0.974	99	0.217	96	0.896	98	1.000
Mayo River	137	0.442	138	0.986	134	0.205	139	0.896	137	1.000
Pelly River	107	0.495	111	0.932	110	0.214	111	0.851	111	1.000
Blind Creek	93	0.489	94	0.957	94	0.362	94	0.910	90	1.000
Tatchun Creek	163	0.390	163	0.988	155	0.245	160	0.953	159	1.000
Big Salmon River	110	0.400	110	0.995	109	0.330	110	0.905	112	1.000
Nordenskiold River	56	0.446	56	1.000	55	0.164	56	0.964	56	1.000
Stoney Creek	93	0.274	94	0.835	93	0.231	94	0.851	94	1.000
Takhini River	153	0.301	157	0.866	154	0.214	158	0.864	159	1.000
Whitehorse	141	0.468	149	0.849	136	0.364	139	0.917	147	1.000
Nisutlin River	53	0.283	56	1.000	55	0.327	56	0.973	56	1.000
F_{ST}		0.070		0.063		0.017		0.040		0.000

ther U.S. or Canadian samples; no branch contained both.

Genetic Stock Identification

Simulated mixtures composed entirely of fish from each of the regions in this study were correctly assigned to locality 45–91% of the time, to region 85–94% of the time, and to country 96–97% of the time (Table 1). When the proportion of U.S. fish in the simulated mixtures was varied from 0.00 to 1.00, mean estimates differed from the true contributions by no more than 4%. The 90% confidence intervals were largest when the true contribution was 0.50 U.S. fish (40–60%) and smallest when the true contribution was 0.00 (0–8%). These simulations indicate that estimates based on the SNP baseline were 2% less accurate and had larger 90% confidence intervals than estimates based on the allozyme baseline of Templin et al. (2005; Table 3).

Using the SNP baseline, we correctly assigned fish samples of known origin to locality 64% of the time, to region 92–100% of the time, and to country 100% of the time.

Assignment of the test fishery samples to the lower, middle, and upper river regions indicated an increase in captures of lower-river fish and a decrease in captures of middle-river fish over the collection period (Figure 3). A notable difference in allocations made by the two marker types, however, is that SNP data assigned a much higher proportion of the fish from periods 2 and 3 to the middle U.S. river (Figure 3).

Discussion

Single nucleotide polymorphisms are an appealing management tool because of the rapid rate at which samples can be processed. Whereas comparable baselines for other genetic markers have taken several months to generate, the present baseline was completed in a matter of weeks. Advancing chemistries and hardware have allowed other markers to be determined quickly, but we are not aware of any technology that allows scoring of hypervariable loci such as amplification fragment length polymorphisms or microsatellites to be automated to the same extent as for SNPs.

We observed a significant departure from Hardy–Weinberg equilibrium at *Ots_Ots2* in the Anvik River sample. Repeating the genotyping of this sample yielded results identical to those observed the first time. Homozygote excesses relative to genotypic ratios expected under Hardy–Weinberg equilibrium may be indicative of the Wahlund Effect (i.e., the sample of individuals we genotyped may represent more than one population). Given

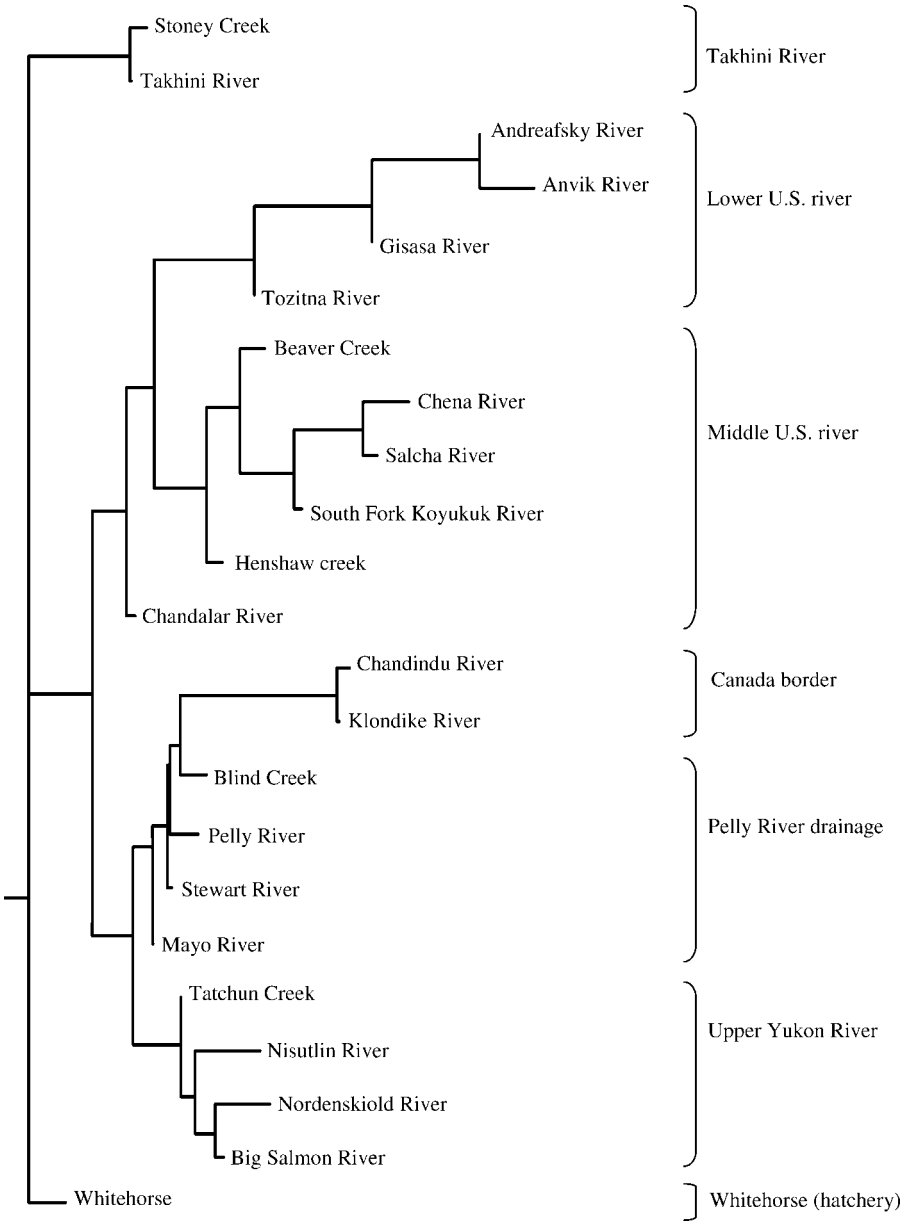

FIGURE 2.—Neighbor-joining dendrogram based on genetic distances for 10 single nucleotide polymorphism loci in 23 samples of Yukon River Chinook salmon.

that allele frequencies in the Anvik collection were similar to those observed in adjacent collections (Figure 2), along with the broad geographic scale of the management issue addressed here, the heterozygote excess observed in the Anvik collection is probably not a problem. The homozygote excess observed here, however, does warn that additional samples from this region should be taken before assessing population structure at a finer level, such as within the lower Yukon River.

The population structure revealed by the neighbor-joining analysis was largely concordant with that based on allozyme data (Templin et al. 2005) or on scale-pattern analyses (Alaska Department of Fish and Game, unpublished data). Both of these earlier data sets identified six regions

TABLE 3.—Estimated composition of simulated samples of Chinook salmon in terms of U.S. and Canadian stocks using the allozyme baseline and the single nucleotide polymorphism (SNP) baseline. Fish consist entirely of individuals from the region in question; simulations were generated as described in the text. The U.S. stock is broken into lower and middle rivers to highlight differences between the estimates based on the SNP baseline and those based on the allozyme baseline of Templin et al. (2005).

	Allozyme baseline		SNP baseline	
Simulated mixture	Mean	90% CI	Mean	90% CI
Region				
Lower U.S. River	0.97	(0.92–1.00)	0.94	(0.87–0.99)
Middle U.S. River	0.97	(0.94–1.00)	0.94	(0.87–1.00)
Canada	0.99	(0.97–1.00)	0.97	(0.91–1.00)
Country				
United States	0.98	(0.94–1.00)	0.96	(0.90–1.00)
Canada	0.99	(0.97–1.00)	0.97	(0.91–1.00)

within the Yukon River drainage among which Chinook salmon were highly distinguishable from one another. The major branches of the neighbor-joining dendrogram based on SNP data closely mirror these six regions. A seventh group in the present study is the Whitehorse sample. This divergence, coupled with the fact that Whitehorse is a hatchery population, led us to treat Whitehorse as a seventh group in the present analysis. The SNP data, in concordance with allozyme data, grouped both sets of the U.S. baseline samples onto a single branch in the neighbor-joining analysis. However, baseline coverage for both SNPs and allozymes is incomplete near the international border. Subpopulations adjacent to one another on the two sides of the international border may share higher levels of gene flow and thus be more difficult to distinguish from one another than subpopulations that are separated by larger geographic distances. As we acquire additional samples from the Canadian and U.S. portions of the Porcupine River (Figure 1) and from the Yukon main stem near the international border, our power to assign unknown fish using any baseline may quite possibly be reduced.

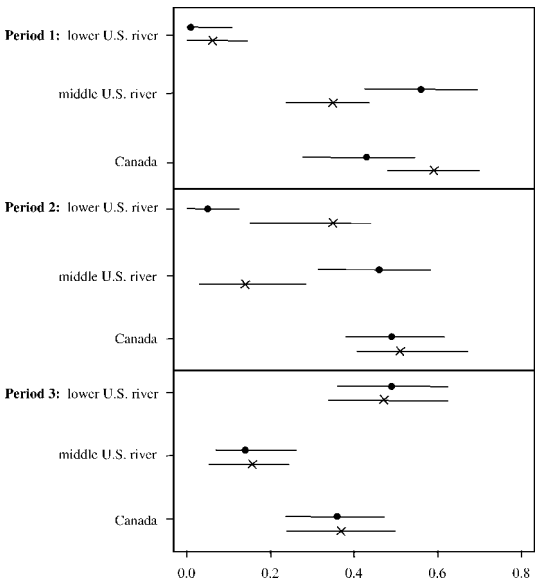

FIGURE 3.—Estimated composition of Pilot Station, Alaska, test fishery samples in terms of U.S. and Canadian stocks. The U.S. source is broken into a lower U.S. river and a middle U.S. river to highlight differences between estimates from the present single nucleotide polymorphism (SNP) baseline and those from the allozyme baseline. Dots indicate estimates based on SNPs, and ×s indicate estimates based on allozymes. Bars indicate 90% confidence intervals.

Mixture assignments for both simulated fish and fish of known origin indicated that the present SNP baseline can correctly assign Yukon River Chinook salmon to country of origin with more than 95% accuracy. Assignment of samples to the seven regions identified here was close to 90% accurate, suggesting that the level of accuracy for a given study must be considered before this baseline is used for that purpose. Correct assignment to individual localities was almost universally less than 90% accurate (Table 1), suggesting that the present baseline is not suitable for assignment at this level—whether because of a lack of resolution of the present markers or a lack of real genetic distinction between these localities. We are currently examining additional SNPs and microsatellite markers to improve our understanding of the biology of the Yukon River Chinook salmon population. In the context of management needs under the Yukon River Salmon Agreement, the present set of SNP loci appears sufficient to distinguish U.S. from Canadian stocks. Should finer resolution

be deemed necessary in the future based on changing management needs, treaty negotiations, or the addition of more baseline samples, then additional SNP loci will be added.

Allozyme and SNP baselines were concordant in their assignment of the Pilot Station, Alaska, test fishery samples to country of origin (Figure 3). Within the United States, however, the baselines differed in how they divided the Pilot Station fish between the lower and middle river sections in periods 1 and 2. Simulations suggested that the two baselines assigned to the middle and lower regions had similar accuracy and precision (Figure 3). A potential explanation for the discrepancy is that although the allozyme baseline has no samples between the Tanana River and the U.S.–Canadian border, the SNP baseline has two samples from this region (Beaver Creek and Chandalar River, both in the middle U.S. river), thus rendering the SNP baseline more likely to assign samples to this region successfully. This does not necessarily mean that the SNP estimates are more accurate. We speculate that estimates for both baselines might change had we been able to acquire more extensive samples from the middle U.S. river, including the Porcupine River in both countries. For the present, concordance among baselines regarding estimates of Canadian contributions (Figure 3) gives us confidence in using both methods in the context of the Yukon River Salmon Agreement (although allozyme sample requirements limit the future potential of that baseline).

We found the SNPs used here successfully differentiated the Canadian and U.S. samples we examined; however, additional loci will probably be needed before SNPs can be used for analyses of mixtures in which the number of regions to be distinguished is large. The number of SNPs that will be required for an application will depend on the number of taxa, the divergence between those taxa, the polymorphism of the SNP loci, and the desired probability level (Seddon et al. 2005). For inference of human ancestry, Rosenberg et al. (2003) suggested that one dinucleotid microsatellite contains the information of five to eight random SNPs. Empirical data from European dairy breeds revealed that 37 SNPs yielded discrimination power equivalent to 22 microsatellite loci (Werner et al. 2004), which suggests a very different information content ratio between the two marker types. Part of the difference between these two information ratios may be that the former is based on ''random'' SNPs and microsatellites. In practice, markers chosen for management applications may reflect significant ascertainment biases and accordingly be more informative than ''random'' loci (e.g., Bensch et al. 2002). The nine polymorphic SNPs examined here appear sufficient for the present management issue but would not be for examination of finer-scale population structure.

The SNP markers used in the present baseline could be used by any laboratory to produce mixture results identical to those described. Further, these data can be combined with other baseline data for the same loci, independent of chemistry or hardware and without any need for interlaboratory allelic standardization. This advantage stems directly from the fact that each allele included in the baseline is unambiguously related to a DNA sequence. In the context of international and multijurisdictional fisheries, reproducibility and transportability of data across laboratories and countries can be very advantageous. Moreover, the ability of scientists and managers to combine data across laboratories without having to spend time and money standardizing alleles at each laboratory represents another efficiency that stands to be gained by using SNPs.

Acknowledgments

This work was funded by NOAA Cooperative Agreement NA03NMF4380185 (Yukon River Salmon Agreement Studies). This manuscript was substantially improved in response to comments from three reviewers and the NAJFM editorial staff. The authors are grateful for technical assistance provided by N. DeCovich, A. Antonovich, J. Berger, A. Barclay, and D. Robinson. Baseline samples used for this work were collected under Federal Office of Subsistence Management Project 12-121, which involved collaboration among the U.S. Fish and Wildlife Service's Conservation Genetics Laboratory, the Canadian Department of Fisheries and Oceans' Pacific Biological Station, and the Alaska Department of Fish and Game (ADF&G) Gene Conservation Laboratory. We thank G. Sandone, J. Hilsinger, and D. Bergstrom, ADF&G, for their support on this project.

References

Beacham, T. D., C. B. Murray, and R. E. Withler. 1989. Age, morphology, and biochemical genetic variation of Yukon River Chinook salmon. Transactions of the American Fisheries Society 118:46–63.

Bensch, S., S. Akesson, and D. E. Irwin. 2002. The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. Molecular Ecology 11:2359–2366.

Cavalli-Sforza, L. L., and A. W. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. American Journal of Human Genetics 19: 233–257.

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(6):509–510.

Felsenstein, J. 1995. PHYLIP. University of Washington, Seattle.

Gharrett, A. J., S. M. Shirley, and G. R. Tromble. 1987. Genetic relationships among populations of Alaskan Chinook salmon (*Oncorhynchus tshawytscha*). Canadian Journal of Fisheries and Aquatic Sciences 44:765–774.

Healy, M. C. 1991. Life history of Chinook salmon. Pages 311–394 *in* C. Groot and L. Margolis, editors. Pacific salmon life histories. UBC Press, Vancouver.

Raymond, M., and F. Rousset. 1995a. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248–249.

Raymond, M., and F. Rousset. 1995b. An exact test for population differentiation. Evolution 49:1280–1283.

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

Rosenberg, N. A., L. M. Li, R. Ward, and J. K. Pritchard. 2003. Informativeness of genetic markers for inference of ancestry. American Journal of Human Genetics 73:1402–1422.

Seeb, L. W., C. Habicht, W. D. Templin, K. E. Tarbox, R. Z. Davis, L. K. Brannian, and J. E. Seeb. 2000. Genetic diversity of sockeye salmon (*Oncorhynchus nerka*) of Cook Inlet, Alaska, and its application to restoration of populations affected by the *Exxon Valdez* oil spill. Transactions of the American Fisheries Society 129:1223–1249.

Seddon, J. M., H. G. Parker, E. A. Ostrander, and H. Ellegren. 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Molecular Ecology 14:503–511.

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

Small, M. P., T. D. Beacham, R. E. Withler, and R. J. Nelson. 1998. Discriminating coho salmon (*Oncorhynchus kisutch*) populations within the Fraser River, British Columbia, using microsatellite DNA markers. Molecular Ecology 7:141–155.

Smith, C. T., J. E. Seeb, P. Schwenke, and L. W. Seeb. 2005. Use of the 5′-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. Transactions of the American Fisheries Society 134:207–217.

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

Utter, F., D. Teel, G. Milner, and D. McIsaac. 1987. Genetic estimates of stock compositions of 1983 Chinook salmon, *Oncorhynchus tshawytscha*, harvests off the Washington coast and the Columbia River. Fisheries Bulletin 85:13–23.

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

Werner, F. A. O., G. Durstewitz, F. A. Habermann, G. Thaller, W. Kramer, S. Kollers, J. Buitkamp, M. Georges, G. Brem, J. Mosner, and R. Fries. 2004. Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. Animal Genetics 35:44–49.