Transactions of the American Fisheries Society 134:207–217, 2005

[Note]

Use of the 5′-Nuclease Reaction for Single Nucleotide Polymorphism Genotyping in Chinook Salmon

CHRISTIAN T. SMITH* AND JAMES E. SEEB

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

PIPER SCHWENKE

National Marine Fisheries Service, Conservation Biology, 2725 Montlake Boulevard East, Seattle, Washington 98122, USA

LISA W. SEEB

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

Abstract.—Migratory and stock composition studies of Chinook salmon *Oncorhynchus tshawytscha* require genetic markers by which a large number of individuals can be processed in a relatively short time. Given the multijurisdictional geographic range of this species, it is further desirable that genetic markers and the corresponding data be transportable across laboratories. We developed 10 single nucleotide polymorphism (SNP) genotyping assays in Chinook salmon based on the 5′-nuclease reaction. Using these assays, a single technician with two thermal cyclers can generate thousands of genotypes per day. The genotyping assays described here are easy to standardize across laboratories, and the resulting genotype data are readily combined with those collected by means of any other sequence detection platform. The rapid rate at which genotyping may be done using these markers and the fact that SNP data are standardized across laboratories and platforms much more readily than are data from other genetic marker classes suggest that SNPs will become an increasingly important tool for mixture studies of Chinook salmon and other salmonids.

Chinook salmon *Oncorhynchus tshawytscha* spawn in rivers around the North Pacific basin southward to approximately 40°N. As is typical for many salmonids, Chinook salmon from throughout the range may form large aggregates during their oceanic and coastal migrations (Myers et al. 1987). The ability of fishery managers to identify the natal origins of fish comprising these aggregates is desirable both for fishery management and for the protection of individual stocks or populations. In recent years, mixed-stock analysis techniques based on genetic markers have been successfully applied to estimate the origins of mixtures taken in nearshore fisheries and to delineate the oceanic migratory route of individual stocks or aggregates (e.g., Seeb et al. 2004). The allele frequency baselines required for such applications must often be composed of samples representing broad geographic sections of the species' range. Given the interjurisdictional nature of these analyses, the genetic markers chosen for such studies should ideally be accessible through relatively rapid and transportable laboratory assays.

Both allozymes and microsatellite DNA have been successfully used in stock composition analyses of Chinook salmon (Nelson et al. 2001; Beacham et al. 2003; Guthrie and Wilmot 2004). A challenge in using genetic markers for which the underlying DNA sequence is not ascertained during genotyping has been a lack of standardization of alleles and thus the inability of managers to combine data across laboratories. The geographic range of Chinook salmon encompasses multiple political and management jurisdictions, making collaboration among the respective management agencies and transparent analyses crucial to effective management. Given the lack of platform-independent reference standards for most allozyme and microsatellite alleles, merging of data generated in different laboratories entails time and monetary costs. In the case of allozymes, laboratories from throughout the North Pacific basin collaborated to standardized a database consisting of 33 loci in 254 Chinook salmon populations (Teel et al. 2000). In the case of microsatellites, no database for Chinook salmon has been successfully

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

Received February 4, 2004; accepted June 24, 2004

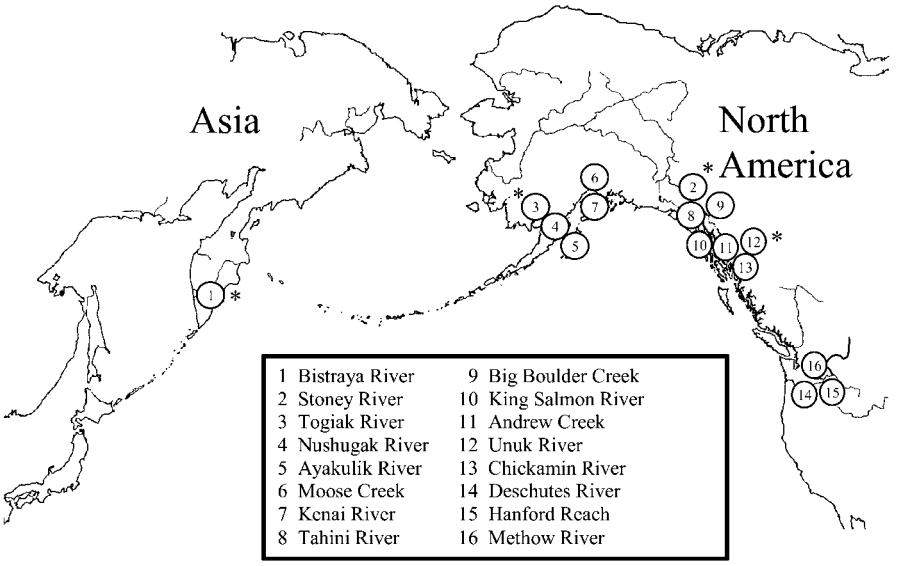

FIGURE 1.—Map showing the 16 locations at which Chinook salmon were sampled. Asterisks denote collections from which initial individuals were analyzed by DNA sequencing. Sample sizes are listed in Table 3.

shared among laboratories, although efforts are currently under way in the Pacific Northwest.

Single nucleotide polymorphisms (SNPs) are a class of genetic marker based on single DNA base differences between individuals at defined positions in the genome. The five possible alleles for an SNP are the four DNA bases (adenine [A], guanine [G], cytosine [C], and thymine [T]) and deletion (the absence of the defined genomic position due to mutation). Genotyping SNP loci involves either performing DNA sequencing or applying an assay that allows inference of the allele present (a range of these techniques is described in Kwok 2003). Theoretical considerations that make SNPs appealing as tools for conservation and population studies have been the subject of several recent reviews (Brumfield et al. 2003; Morin et al. 2004; Schlotterer 2004). Two aspects of SNP data that make these markers particularly appealing for the purpose of fishery management are (1) the rapid throughput rate of the available assays relative to those employing other genetic markers and (2) the fact that since each SNP allele is unambiguously related to a DNA sequence, no standardization is required to combine or compare data across laboratories.

The 5′-nuclease reaction is a high-throughput SNP genotyping method in which the 5′-exonuclease activity of DNA polymerase is used to digest allele-specific probes in the course of the polymerase chain reaction (PCR; Holland et al. 1991). The use of fluorogenic probes in a real-time PCR machine further allows amplification and genotyping to take place simultaneously (Lee et al. 1993), obviating the need for electrophoresis. Prior studies have demonstrated that this method is both rapid and accurate (Ranade et al. 2001). Here we describe the development of ten 5′-nuclease SNP genotyping assays for use with Chinook salmon.

Methods

Genomic DNA was extracted from 1,251 Chinook salmon representing 16 collections from throughout the species' range (Figure 1) using a DNeasy 96 kit (QIAGEN, Valencia, California).

Ten loci known to contain SNPs in Chinook salmon based on previous publications (Table 1) were examined. The polymerase chain reaction primers and conditions used to amplify and sequence these loci followed those in the respective publications (Table 1) and are available on request. After initial amplification, each locus was cleaned using QIAquick columns (QIAGEN), sequenced using the Applied Biosystems (ABI, Foster City, California) BigDye 3.1 DNA sequencing kit, purified using DyeEx 2.0 columns (QIAGEN), and analyzed on an ABI377 DNA sequencer. The sequences were aligned and examined for SNPs using SeqMan (DNASTAR) and the SNP Pipeline (Buetow et al. 1999). Oligonucleotide probes and PCR primers for use in 5′-nuclease reactions were

TABLE 1.—Chinook salmon loci surveyed for single nucleotide polymorphisms (SNPs). Assay names were assigned according to the convention of the National Center for Biotechnology Information's database of SNPs; ADFG refers to the laboratory where the analysis was done, *Ots* to the species in question (Chinook salmon), and the remaining letters to the gene in question.

Assay	Number of sequences examined	Sequence length (base pairs)	Number of SNPs	Assay target details
ADFG|*Ots*_Prl2	40	865	3	Transition corresponding to base 1,282 of Xiong et al.'s (1992) alignment, which falls within intron 2 of the prolactin 2 gene (position 3,688 in GenBank accession number S66606). The other two SNPs observed at this locus were an adenine–thymine transversion at position 3,572 and a cytosine–thymine transition at position 3,675 of S66606.
ADFG|*Ots*_P53	40	578	3	Synonymous transition in the P53 gene as described by Park et al. (1996; position 369 of GenBank accession numbers AF223793–AF223818 and AF071574).
ADFG|*Ots*_GH2	40	1,040	1	Originally noted in Chinook salmon by Park et al. (1995), this transversion corresponds to base 1,252 of McKay et al.'s (1996) alignment of growth hormone II (position 1,252 of GenBank accession number OTU28157). The nonsynonymous change in exon 5 replaces a glutamine with a leucine molecule.
ADFG|*Ots*_MHC2	40	258	7	This transversion corresponds to position 77 in the amino acid sequence of the major histocompatibility complex (MHC) class II β1 locus published by Kim et al. (1999; position 38 of the following GenBank accession numbers: ''A'' allele, AF041010–AF041011, OTU80301, AY100007–AY1000011, OTU34719; ''G'' allele, OTU80299, AY100006, OTU34718, OTU80300, OTU34720). The change is nonsynonymous, resulting in an amino acid replacement of phenylalanine with valine.
ADFG|*Ots*_Ots2	40	274	1	Transition at position 180 of the clone sequence (GenBank accession number AF107030) of the *Ots2* microsatellite flanking region (Banks et al. 1999) noted by Blankenship et al. (2002).
ADFG|*Ots*_C3N3	347	368	2	Synonymous transversion in mitochondrial DNA cytochrome *c* oxidase subunit III (position 10,409 of GenBank accession number AF392054).
ADFG|*Ots*_MHC1	8	222	6	Transition corresponding to position 37 of the MHC class I B locus exon 3 (alpha2 domain) alignment presented by Miller and Withler (1998); allele ''A'' corresponds to *Onts B*6*, *3*, and *5* (GenBank accession numbers AF104585, AF104587), and allele ''G'' corresponds to *Onts-B4*, *1*, and *2* (GenBank accession numbers AF104586, AF104583, AF104584). Those authors observed no departures from Mendelian inheritance at this locus and suggested that this might be a pseudogene based on a stop codon in allele *B*4*.
ADFG|*Ots*_Tnsf	36	1,970	11	Synonymous transition in the transferrin gene described by Ford et al. (1999). The ''T'' allele corresponds to position 489 of GenBank accession numbers AF114874, AF223681, AF223693, AF223699, AF223753, AF223761, AF223781. The ''C'' allele corresponds to the same position in accession numbers AF114869, AF114880, AF114886, AF114891, AF114896, AF223687, AF223705, AF223711, AF223717, AF223723, AF223729, AF223735, AF223741, AF223747, AF223760, AF223774, AF223788).
ADFG|*Ots*_SL	333	566	2	Transition in the 3′ untranslated region of the somatolactir precursor gene described by Ford (1998). The ''A'' allele corresponds to position 1,317 of GenBank accession numbers AF223833, AF223836, AF223839, AF223842, AF223845, AF223848, AF223851, AF223854, AF223866, and AF223890 and to position 1,316 of AF223884; the ''C'' allele corresponds to AF223857, AF223860, AF223863, AF223869, AF223872, AF223875, AF223878, AF223881 and AF223887).
ADFG|*Ots*_P450	310	370	4	Transversion in the 5′ untranslated region of cytochrome p450A described by Ford (1998; position of 495 of GenBank accession number AF059710).

TABLE 2.—Primer and probe sequences, annealing temperatures (T_a [°C]), and numbers of cycles used in SNP genotyping assays in Chinook salmon. Each probe was labeled with either VIC or 6FAM on its 5′ end and bore a minor groove binder and a nonfluorescent quencher on its 3′ end. See Table 1 for more details on the individual assays.

Assay	PCR primer and probe sequences (5′–3′)	T_a	Number of cycles	F_{ST}
ADFG|*Ots*_C3N3	CCGGATTCCATGGCCTACAC	60	45	0.299
	GCCAAAATGATGTTCGGATGTAAAGT			
	VIC-CTAGAAAGGTTGATCCAATAA			
	FAM-AAAGGTTGAGCCAATAA			
ADFG|*Ots*_Prl2	CCTGGTCTGTTTGTGATCAAGATG	60	44	0.077
	GGTTAACTCAAATAGAACATACTCTGACACA			
	VIC-ATGTATTGTTCATTTAATG			
	FAM-TGTATTGTTCGTTTAATG			
ADFG|*Ots*_P53	GGAACTTCCTCTCCCGTTCTG	60	45	0.038
	GCACACACACGCACCTCAA			
	VIC-CTGGGTCGGCGCT			
	FAM-TGGGTCGACGCTC			
ADFG|*Ots*_GH2	GCGTACTGAGCCTGGATGACA	62	45	0.192
	CCCCCAGGTTCTGGTAGTAGTTC			
	VIC-TGACTCTCAGCATCT			
	FAM-TGACTCTCTGCATCTG			
ADFG|*Ots*_MHC1	GTCCACATTCTCCAGTACATGTATGG	62	40	0.205
	CAAACCCCTCTGTCTGTTCAGT			
	VIC-CATCATCCCGTGAGCAG			
	FAM-TCATCATCCCATGAGCAG			
ADFG|*Ots*_MHC2	GTCCTCAGCTGGGTCAAGAG	62	40	0.327
	GTAGTGGAGAGCAGCGTTAGG			
	VIC-CTGGAGCGTTTCTGTA			
	FAM-CTGGAGCGTGTCTGTA			
ADFG|*Ots*_Ots2	CCTTTTAAACACCTCACACTTAGAGAAAT	62	45	0.126
	TGTAAAGATGACAGTCTACTATCCTGGTTT			
	VIC-CTGAAGCGTAGTTAAG			
	FAM-CACTGAAGCATAGTTAA			
ADFG|*Ots*_Tnsf	GCCAATACGGGTTCTGAACTGT	60	45	0.434
	CGGAATAGTCATAGTAGGGCTCGTT			
	VIC-TGCTCCAGATCTC			
	FAM-TGCTCCAGGTCTC			
ADFG|*Ots*_SL	AATATTGGCTTTCTGAGAATGCATTTGG	60	45	0.315
	CCAAGATACTTCCTTTAACTTCTCTGTCA			
	VIC-TCAAAGATATGATTCAATTAA			
	FAM-AAGATATGGTTCAATTAA			
ADFG|*Ots*_P450	TGAGCGAGATTTATCAAACTGTCAAAGA	60	45	0.301
	CCCAAGCGGGAGAACTTACAG			
	VIC-CCCCGAAGTACTTTT			
	FAM-CCCGAAGAACTTTT			

designed for one SNP in each locus using either Primer Express (ABI) or Assays-by-Design (ABI).

The genotyping assays were named according to the convention of the National Center for Biotechnology Information's database of SNPs (dbSNP), in which the laboratory identifier precedes the locus identifier. For example, the name of the first assay in Table 1 (ADFG|*Ots*_Prl2) indicates that the laboratory performing the assay is associated with the Alaska Department of Fish and Game, that the genetic material is from a Chinook salmon, and that the material is part of the gene coding for prolactin 2. The assays were performed in 384-well reaction plates, 2 wells in each plate serving as negative controls (no template) and 2 wells as positive controls (one for each allele). Each reaction was conducted in a 5-μL volume consisting of 0.15 μL template DNA in 1× TaqMan PCR cocktail (ABI), 900 nM of each PCR primer, and 200 nM of each probe (Table 2). Pipetting into the 384-well plates was done with a BioRobot RapidPlate (QIAGEN). Thermal cycling was performed on either an ABI7900 real-time sequence detection system or a DNA Engine Tetrad (MJ Research, Waltham, Massachusetts) as follows: an initial denaturation of 10 min at 95°C was followed by 40–45 cycles of 92°C for 15 s and an annealing−extension temperature (Table 2) for 1 min. All cycling was conducted at a ramp speed of 1°C per second. Scoring of individual genotypes was performed with Sequence Detection Software 2.1 (ABI) to generate scatter plots that graphically

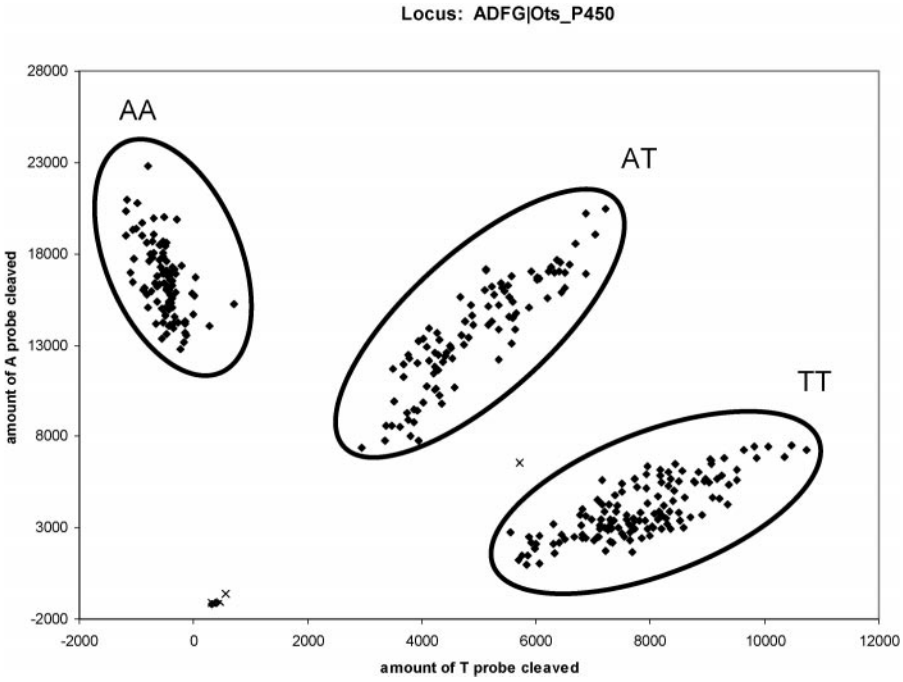

FIGURE 2.—Example of the data produced by a nuclear single nucleotide polymorphism genotyping assay (in this case for locus *Ots*_P450). Each diamond represents an individual fish whose genotype is determined by its position with respect to the two axes, which indicate the amount of each allele-specific probe (adenine [A] or thymine [T]) cleaved during the course of the assay. The ellipses indicate clusters of single genotypes. While mitochondrial assays produce two clusters, nuclear assays such as this one produce three clusters. The ×s represent unreadable samples (due to air bubbles, failed polymerase chain reactions, etc.); the diamonds in the lower left corner represent the negative controls (in which no template DNA was added).

depicted the amount of each allele-specific probe that bound to the PCR product of each individual (Figure 2).

The program GENEPOP (Raymond and Rousset 1997) was used to test for differences between observed genotype frequencies and those expected under Hardy–Weinberg equilibrium. Statistical significance levels were corrected for multiple simultaneous comparisons as described by Rice (1989; $\alpha = 0.05/16$ collections $= 0.003$). Nei's (1987) estimators of the genetic differentiation indices F_{IS} and F_{ST} (G_{ST}) were calculated with the program FSTAT (Goudet 2001). The program PHYLIP (Felsenstein 1995) was used to calculate genetic distances (Cavalli-Sforza and Edwards 1967) between all collections and to perform an analysis involving the unweighted pair-group method with arithmetic means (UPGMA) based on those distances. The portion of the observed genetic variation accounted for within collections as well as that accounted for by dividing the collections into ocean type (the Hanford Reach and Deschutes River collections) and stream type (all others) were assessed via analysis of molecular variance (AMOVA; Excoffier et al. 1992) based on F_{ST} (Weir and Cockerham 1984) as performed by ARLEQUIN (Schneider et al. 1999). The significance of these partitions was tested using 16,000 permutations.

Results

On average, the DNA sequences that we examined revealed an SNP approximately every 160 base pairs (bp). The criteria by which loci were chosen for sequencing preclude any general inferences regarding the frequency of SNPs within the Chinook salmon genome; however, our observation is fairly close to the rate of one SNP per 200–500 bp observed as an average across a wide range of loci and taxa (Brumfield et al. 2003).

The thermal cycling times for the present assays were under 2 h. After amplification, it took approximately 5 min to read and analyze a 384-well plate. Using two 384-well thermal cycler blocks three times per 7.5-h day, a single technician was able to generate and score 2,280 genotypes.

TABLE 3.—Observed allele frequencies and F_{IS} estimates for Chinook salmon collections taken from the Bistraya River (Bist), Stoney River (Ston), Togiak River (Togi), Nushugak River (Nush), Ayakulik River (Ayak), Moose Creek (Moos), Kenai River (Kena), Tahini River (Tahi), Big Boulder Creek (Big), King Salmon River (King), Andrew Creek (Andr), Unuk River (Unuk), Chickamin River (Chik), Deschutes River (Desc), Hanford Reach (Hanf), and Methow River (Meth). Numbers in parentheses are sample sizes. The F_{IS} values in bold italics denote collections and loci exhibiting genotypic ratios significantly different from those expected under Hardy–Weinberg equilibrium. No F_{IS} estimates are given for the mitochondrial single nucleotide polymorphism *Ots*_C3N3. Allele names correspond to the four standard nucleotide bases (i.e., adenine [A], guanine [G], cytosine [C], and thymine [T]). See Table 1 for additional details on the loci studied.

Locus	Allele	Bist (94)	Ston (95)	Togi (91)	Nush (95)	Ayak (93)	Moos (46)	Kena (92)
*Ots*_GH2	A	0.937	0.411	0.843	0.789	0.785	0.907	0.783
	T	0.063	0.589	0.157	0.211	0.215	0.093	0.217
	F_{IS}	−0.062	−0.169	−0.096	−0.071	0.007	−0.091	−0.14
*Ots*_Prl2	A	0.659	0.394	0.549	0.668	0.337	0.772	0.528
	G	0.341	0.606	0.451	0.332	0.663	0.228	0.472
	F_{IS}	−0.122	−0.088	−0.077	0.147	−0.016	−0.162	−0.042
*Ots*_Tnsl	A	0.017	0.106	0.102	0.129	0.081	0.045	0.056
	G	0.983	0.894	0.698	0.871	0.919	0.955	0.944
	F_{IS}	−0.011	−0.002	−0.108	−0.143	0.063	−0.032	−0.053
*Ots*_Ots2	A	0.112	0.037	0.12	0.101	0.02	0.028	0.101
	G	0.888	0.963	0.88	0.899	0.98	0.974	0.839
	F_{IS}	0.056	−0.033	−0.011	−0.107	−0.014	−0.013	0.069
*Ots*_MHC1	A	0.594	0.632	0.606	0.536	0.287	0.197	0.424
	G	0.406	0.368	0.394	0.464	0.713	0.803	0.576
	F_{IS}	0.265	0.01	0.006	0.002	−0.066	0.1	0.071
*Ots*_P53	A	0.426	0.271	0.38	0.489	0.563	0.463	0.309
	G	0.574	0.729	0.62	0.511	0.437	0.538	0.691
	F_{IS}	0.26	−0.098	0.008	0.184	0.025	0.258	0.085
*Ots*_MHC2	G	0.973	0.165	0.114	0.126	0.075	0	0.029
	T	0.027	0.835	0.886	0.874	0.925	1	0.971
	F_{IS}	−0.022	0.04	−0.004	−0.044	−0.076	NA	0.387
*Ots*_P450	A	0.128	0.234	0.188	0.183	0.192	0.104	0.319
	T	0.872	0.766	0.812	0.617	0.808	0.896	0.681
	F_{IS}	−0.141	−0.063	−0.225	0.142	−0.082	0.349	0.094
*Ots*_SL	A	0.781	0.853	0.59	0.697	0.434	0.458	0.819
	G	0.219	0.147	0.41	0.303	0.566	0.542	0.181
	F_{IS}	0.053	0.084	−0.213	−0.111	0.203	−0.049	0.03
*Ots*_C3N3	G	1	1	1	1	1	1	1
	T	0	0	0	0	0	0	0

In the mitochondrial assay (*Ots*_C3N3), a single probe was cleaved in each tube, indicating the haplotype of the individual. In nuclear assays, a single probe was cleaved in reactions containing homozygous templates while both probes were cleaved in reactions containing heterozygous templates (Figure 2).

Significant departures from Hardy−Weinberg equilibrium were detected in one or more collections for *Ots*_Ots2, *Ots*_MHC1, and *Ots*_MHC2 (Table 3). In each case the number of observed heterozygotes was less than expected under Hardy–Weinberg equilibrium conditions.

Allele frequencies for the SNP loci varied widely across the range of Chinook salmon (Table 3). For example, *Ots*_MHC2 exhibited a sharp allele frequency difference (≥41%) between the Bistraya River and North American collections, whereas *Ots*_Tnsf revealed a major transition between central and southeast Alaska (>27% difference between the Kenai River and Tahini River samples), and *Ots*_C3N3 revealed a transition within southeast Alaska (>21% difference between the King Salmon River and Andrew Creek samples). These steep allele frequency clines observed in different parts of the species' range in different SNPs resulted in F_{ST} estimates between 0.038 (*Ots*_P53) and 0.434 (*Ots*_Tnsf) (Table 2). The UPGMA tree (Figure 3) illustrated the large allele frequency differences observed among different collections taken within the Columbia River basin (Methow River samples exhibited allele frequencies that were more than 70% different from those of Hanford Reach and Deschutes River samples for *Ots*_MHC1, *Ots*_P450, and *Ots*_SL). The Alaskan collections clustered together and appeared less divergent from the Asian collection (Bistraya River) than they were from the Columbia River collections (Hanford Reach, Deschutes River, and Methow River). Within the Alaskan cluster, King

TABLE 3.—Extended.

Locus	Allele	Tahi (56)	Big (22)	King (94)	Andr (05)	Unuk (91)	Chik (50)	Desc (94)	Hanf (94)	Meth (49)
*Ots*_GH2	A	0.58	0.625	0.994	0.867	0.918	0.92	0.989	1	0.959
	T	0.42	0.375	0.006	0.133	0.082	0.08	0.011	0	0.041
	F_{IS}	−0.128	−0.319	0	−0.056	−0.064	−0.077	−0.005	NA	−0.032
*Ots*_Prl2	A	0.778	0.571	0.614	0.425	0.44	0.406	0.528	0.667	0.323
	G	0.222	0.429	0.386	0.575	0.56	0.594	0.472	0.333	0.677
	F_{IS}	−0.17	−0.036	−0.118	0.147	−0.289	0.104	0.136	0.106	0.2
*Ots*_Tnsl	A	0.278	0.297	0.161	0.311	0.429	0.432	0.921	0.973	0.771
	G	0.722	0.703	0.839	0.689	0.571	0.568	0.079	0.027	0.229
	F_{IS}	−0.152	−0.258	−0.104	0.01	0.158	−0.286	−0.08	−0.022	−0.169
*Ots*_Ots2	A	0.329	0	0.145	0.149	0.151	0.242	0.365	0.337	0.489
	G	0.671	1	0.855	0.851	0.849	0.758	0.635	0.663	0.511
	F_{IS}	0.474	NA	−0.071	0.101	−0.171	−0.127	***0.749***	***0.794***	0.184
*Ots*_MHC1	A	0.458	0.47	0.032	0.585	0.478	0.3	0.258	0.143	0.968
	G	0.542	0.53	0.968	0.415	0.522	0.7	0.742	0.857	0.032
	F_{IS}	−0.249	−0.141	***0.796***	0.041	0.028	−0.038	0.051	0.018	−0.022
*Ots*_P53	A	0.318	0.321	0.519	0.511	0.5	0.612	0.312	0.346	0.326
	G	0.682	0.679	0.481	0.489	0.5	0.388	0.688	0.654	0.674
	F_{IS}	0.129	0.199	0.008	0.226	0.039	−0.367	0.103	0.107	−0.077
*Ots*_MHC2	G	0.164	0.074	0.053	0.13	0.225	0.557	0.391	0.457	0.217
	T	0.836	0.926	0.947	0.87	0.775	0.443	0.609	0.543	0.783
	F_{IS}	***0.814***	0.365	−0.061	−0.143	0.038	−0.553	0.093	−0.046	−0.011
*Ots*_P450	A	0.166	0.205	0.368	0.287	0.366	0.458	0.946	0.932	0.074
	T	0.814	0.795	0.632	0.713	0.644	0.542	0.064	0.068	0.926
	F_{IS}	−0.217	−0.235	−0.032	−0.035	0.054	0.255	−0.051	−0.067	−0.07
*Ots*_SL	A	0.775	0.773	0.208	0.651	0.73	0.59	0.052	0.062	0.989
	G	0.225	0.227	0.792	0.349	0.27	0.41	0.948	0.948	0.011
	F_{IS}	−0.169	0.066	−0.12	0.187	−0.014	−0.023	−0.049	−0.049	0
*Ots*_C3N3	G	1	1	0.968	0.776	0.704	0.615	0.533	0.567	1
	T	0	0	0.032	0.224	0.296	0.385	0.467	0.433	0

Salmon River fish appeared to be the most distinct, reflecting the large allele frequency differences between this collection and those adjacent to it (>45% for both *Ots*_MHC1 and *Ots*_SL).

Analysis of molecular variance revealed that 58.2% of the observed variation was among individuals within the collections. Dividing the collections into ocean type (Hanford Reach and Deschutes River) and stream type (all others) accounted for 32.3% of the observed variation. The remaining 9.5% was among the collections within each of these two groups. Each of these portions was highly significant ($P < 0.01$). Repeating the analysis with a third group that consisted of either Columbia River stream-type (Methow River) or Asian fish (Bistraya River) did not increase the percentage of variation among groups.

Discussion

The 5′-nuclease reaction allowed SNP genotyping to be completed without any electrophoresis or other post-PCR handling steps. The lack of an electrophoretic component meant that thermal cycling was the rate-limiting step in these assays. With additional thermal cycler blocks, these rates could be substantially increased. Even more important in making the present assays rapid relative to those for other markers, however, was the relative simplicity of the raw data and therefore the level to which scoring could be automated (Figure 2). Using the present SNP assays, one can quite comfortably score 10,000 or more genotypes per day.

The departures from Hardy–Weinberg expectations observed at three of the SNP loci examined here (Table 2) suggest that (1) the respective collections are not representative of randomly mating populations, (2) the individual loci violate Hardy–Weinberg assumptions, such as that of selective neutrality, or (3) the present assays exhibit significant genotyping errors. The possibility that the

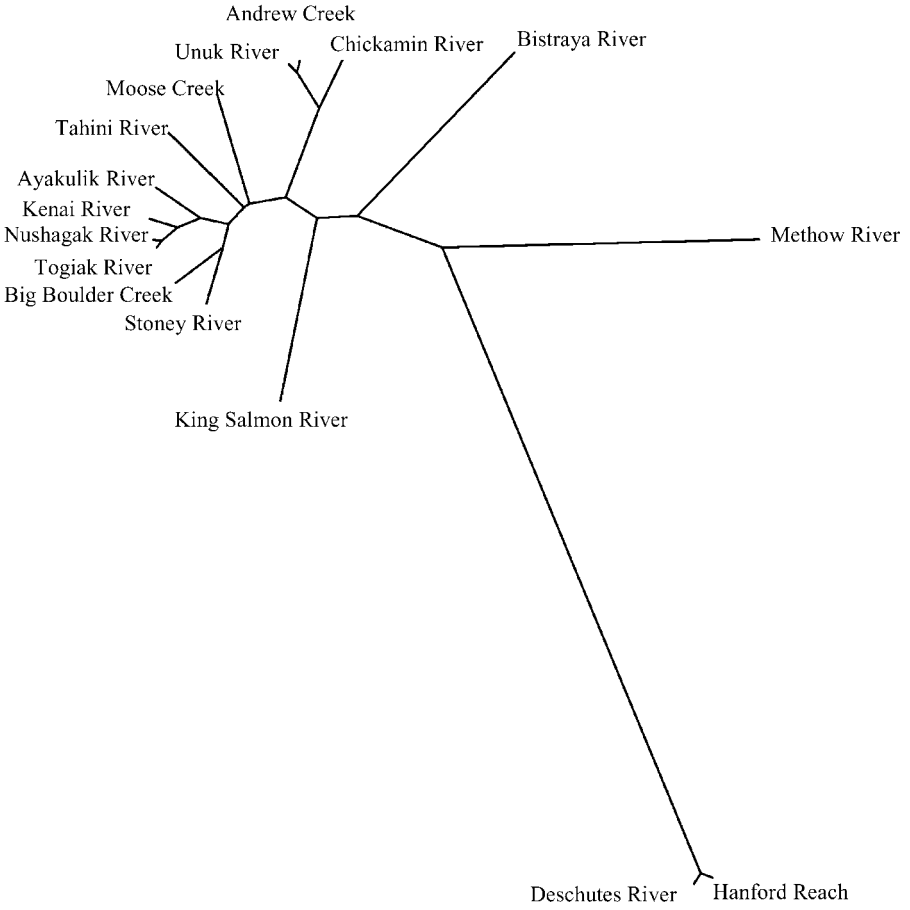

FIGURE 3.—Dendrogram produced by the unweighted pair-group method with arithmetic mean based on genetic distances for 10 single nucleotide polymorphism loci between 16 collections of Chinook salmon. See Figure 1 for collection sites.

collections do not represent populations suggests that skepticism should be exercised in using the present data for analyses at fine geographic scales. Owing to the broad physical distances between collections in the present study, this possibility will not greatly impact the analysis presented here. Given that the assays produced genotypes identical to those produced by DNA sequencing for the individuals for which we had sequence data, and given the previously documented accuracy of the 5′-nuclease reaction in larger studies (e.g., Ranade et al. 2001), the possibility that genotyping errors caused the departures seems unlikely.

The possibility that natural selection is driving the allele frequency differences between populations for some of these SNPs will have implications for the way in which data from these loci should be analyzed. The role of natural selection in determining allele frequencies for most SNPs is expected to be extremely small (Kimura 1968). In cases in which a larger role is suggested, however, the assumptions behind the analyses need to be carefully examined. Since several of the SNP loci presented here are located in or near genes that may be subject to natural selection and two of the SNPs actually code for amino acid changes (*Ots*_MHC2 and *Ots*_GH2), it seems likely that selection has played a role in shaping the frequencies for some of them. Population genetic analyses founded on Wright's (1951) island model and consequently on the assumption that allele frequency differences between populations are driven by a balance between migration and genetic drift may not be appropriate for data collected for these loci. For the purpose of assigning unknown fishery or high-seas samples to established baselines, how-

ever, the use of loci subject to natural selection is valid and in some cases may provide resolution beyond that available via neutral markers (e.g., Beacham et al. 2001).

The population structure resulting from the present SNP data exhibits several aspects that are concordant with those of previously published allozyme data. The primary split observed in an allozyme study of Chinook salmon from throughout British Columbia (Teel et al. 2000) was between the two races of Chinook salmon that were summarized by Healy (1991): ocean type and stream type. The primary split indicated by both the UPGMA analysis and the AMOVA of the present SNP data is also the distinction between the ocean-type collections (Hanford Reach and Deschutes River) and the stream-type collections (all others) (Figure 3). Beyond this split, the joining of collections in the UPGMA analysis largely reflects relative geographic positioning. Two exceptions to this pattern are (1) the relative distinctiveness of King Salmon River collections from other southeast Alaska collections and (2) the pairing of Stoney River (Yukon River drainage) with Big Boulder Creek (southeast Alaska) collections. Again, both the genetic distinctiveness of King Salmon River collections relative to those of other systems in Alaska and the genetic similarity between upper Yukon River collections and those of the coastal rivers of southeast Alaska have previously been described on the basis of allozyme data (Gharrett et al. 1987; Guthrie and Wilmot 2004). The F_{ST} estimates based on the SNPs presented here (range, 0.038–0.434; overall, 0.229) are comparable to those based on 29 allozyme loci from comparable populations (range, 0.0−0.455; overall, 0.090; Alaska Department of Fish and Game, unpublished data), suggesting a high potential for the utility of these SNPs in complex mixture analyses. A full comparison of the information in these SNPs relative to that provided by allozyme and microsatellite loci is pending completion of analysis of the latter in a comparable set of collections.

A present limitation of SNP markers in studies of Chinook salmon and most other species is that relatively few loci have been characterized. Although techniques for identifying large numbers of SNPs are well described (e.g., Osman et al. 2003; Werner et al. 2004), these techniques have not been applied to the vast majority of species for reasons of time and cost. Studies requiring the finest level of resolution, such as those involving individual stock assignment or inference of pairwise relationships between individuals, require large numbers of alleles to succeed and will thus be better served at present by markers such as microsatellites and amplification fragment length polymorphisms (Campbell et al. 2003; Glaubitz et al. 2003). Because only a small number of SNPs are presently available for Chinook salmon, studies to which SNP data alone provide adequate resolution for mixture analyses will be limited to those of regional groupings or more divergent genetic lineages. As additional SNPs are described, the range of applications for SNP data will increase.

The assays described here are simple to standardize across laboratories because they render only a single PCR product detectable, eliminate the multitude of potential interlaboratory inconsistencies associated with electrophoresis (e.g., Wattier et al. 1998; Davison and Chiba 2003), and have a small potential number of alleles per locus. Further, the data collected in the present assays directly reflect underlying DNA sequences and are thus readily combined with data collected across hardware and chemistry platforms as well as laboratories. This portability, in combination with the relatively rapid rate at which SNP data may be generated and the relatively high among-population diversity observed here, suggests that SNPs will become an increasingly important tool for complex mixture and migratory studies of Chinook salmon.

Acknowledgments

We are grateful for the excellent technical assistance provided by N. DeCovich. Tissue samples were generously provided by B. Heard (NOAA–Fisheries), S. Narum (Columbia River Intertribal Fish Commission), N. Varnavskaya (Kamchatka Research Institute of Fisheries and Oceanography), T. Lundrigan (University of Washington School of Fisheries) and ADFG field and laboratory staff. This manuscript was improved by suggestions provided by P. Moran (NOAA–Fisheries) and three reviewers. Funding was provided by the Southeast Alaska Sustainable Fisheries Fund and the Chinook Technical Committee of the Pacific Salmon Commission. Single-nucleotide polymorphism research in our laboratory was enabled by National Marine Fisheries Service Grant NA96FW0196 and North Pacific Research Board Grant R0303.

References

Banks, M. A., M. S. Blouin, B. A. Baldwin, V. K. Rashbrook, H. A. Fitzgerald, S. M. Blankenship, and D.

Hedgecock. 1999. Isolation and inheritance of novel microsatellites in Chinook salmon (*Oncorhynchus tshawytscha*). Journal of Heredity 90:281–288.

Beacham, T. D., J. R. Candy, K. J. Supernault, T. Ming, B. Deagle, A. Schulze, D. Tuck, K. H. Kaukinen, J. R. Irvine, K. M. Miller, and R. E. Withler. 2001. Evaluation and application of microsatellite and major histocompatibility complex variation for stock identification of coho salmon in British Columbia. Transactions of the American Fisheries Society 130: 1116–1149.

Beacham, T. D., J. R. Candy, K. J. Supernault, M. Wetklo, B. Deagle, K. Labaree, J. R. Irvine, K. M. Miller, R. J. Nelson, and R. E. Withler. 2003. Evaluation and application of microsatellites for population identification of Fraser River Chinook salmon (*Oncorhynchus tshawytscha*). Fishery Bulletin 101: 243–259.

Blankenship, S. M., B. May, and D. Hedgecock. 2002. Evolution of a perfect simple-sequence-repeat locus in the context of its flanking sequence. Molecular Biology and Evolution 19:1943–1951.

Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single-nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18:249–256.

Buetow, K. H., M. N. Edmonson, and A. B. Cassidy. 1999. Reliable identification of large numbers of candidate SNPs from public EST data. Nature Genetics 21:323–325.

Campbell, D., P. Duchesne, and L. Bernatchez. 2003. AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. Molecular Ecology 12:1979–1991.

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

Davison, A., and S. Chiba. 2003. Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. Molecular Ecology Notes 3:321–323.

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

Felsenstein, J. 1995. PHYLIP, version 3.6. University of Washington, Seattle.

Ford, M. J. 1998. Testing models of migration and isolation among populations of Chinook salmon (*Oncorhynchus tschawytscha*). Evolution 52:539–557.

Ford, M. J., P. J. Thornton, and L. K. Park. 1999. Natural selection promotes divergence of transferrin among salmonid species. Molecular Ecology 8:1055–1061.

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.

Glaubitz, J. C., O. E. Rhodes, and J. A. DeWoody. 2003. Prospects for inferring pairwise relationships with single-nucleotide polymorphisms. Molecular Ecology 12:1039–1047.

Goudet, J. 2001. FSTAT, version 2.9.3: a program to estimate and test gene diversities and fixation indices. Available: http://www.unil.ch/izea/softwares/fstat.html.

Guthrie, C. M., and R. L. Wilmot. 2004. Genetic structure of wild Chinook salmon populations of southeast Alaska and northern British Columbia. Environmental Biology of Fishes 69:81–93.

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.

Holland, P. M., R. D. Abramson, R. Watson, and D. H. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5′−3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proceedings of the National Academy of Science of the USA 88:7276–7280.

Kim, T. J., K. M. Parker, and P. W. Hedrick. 1999. Major histocompatibility complex differentiation in Sacramento River Chinook salmon. Genetics 151: 1115–1122.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature (London) 217:624–626.

Kwok, P.-Y., editor. 2003. Single-nucleotide Polymorphisms: methods and protocols. Humana Press, Totowa, New Jersey.

Lee, L. G., C. R. Connell, and B. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Research 21:3761–3766.

McKay, S. J., R. H. Devlin, and M. J. Smith. 1996. Phylogeny of Pacific salmon and trout based on growth hormone type-2 and mitochondrial NADH dehydrogenase subunit 3 DNA sequences. Canadian Journal of Fisheries and Aquatic Sciences 53:1165–1176.

Miller, K. M., and R. E. Withler. 1998. The salmonid class I MHC: limited diversity in a primitive teleost. Immunological Reviews 166:279–293.

Morin, P. A., G. Luikart, R. K. Wayne, and the SNP workshop group. 2004. SNPs in ecology, evolution, and conservation. Trends in Ecology and Evolution 19:208–216.

Myers, K. W., C. K. Harris, C. M. Knudsen, R. V. Walker, N. D. Davis, and D. W. Rogers. 1987. Stock origins of Chinook salmon in the area of the Japanese mothership salmon fishery. North American Journal of Fishery Management 7:450–474.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nelson, R. J., M. P. Small, T. D. Beacham, and K. J. Supernault. 2001. Population structure of Fraser River Chinook salmon (*Oncorhynchus tshawytscha*): an analysis using microsatellite DNA markers. Fishery Bulletin 99:94–107.

Osman, A., B. Jordan, P. A. Lessard, N. Muhammad, M. R. Haron, N. M. Riffin, A. J. Sinskey, C. Rha, and D. E. Housman. 2003. Genetic diversity of *Eurycoma longifolia* inferred from single-nucleotide polymorphisms. Plant Physiology 131:1294–1301.

Park, L. K., P. Moran, and D. A. Dightman. 1995. A polymorphism in intron D of the Chinook salmon growth hormone 2 gene. Animal Genetics 26:285.

Park, L. K., P. Moran, and D. A. Dightman. 1996. A Chinook salmon PCR−RFLP marker in the *p53* locus. Animal Genetics 27:121–128.

Ranade, K., M.-S. Chang, C.-T. Ting, D. Pei, C.-F. Hsiao, M. Olivier, R. Pesich, J. Hebert, Y.-D. I. Chen, V. Dzau, D. Curb, R. Olshen, N. Risch, D. R. Cox, and D. Botstein. 2001. High-throughput genotyping with single-nucleotide polymorphisms. Genome Research 11:1262–1268.

Raymond, M., and F. Rousset. 1997. GENEPOP, version 3.1b: an updated version of GENEPOP (version 1.2)—population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248–249.

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

Schlotterer, C. 2004. The evolution of molecular markers: just a matter of fashion? Nature Reviews Genetics 5:63–69.

Schneider, S., D. Roessli, and L. Excoffier. 1999. Arlequin, version 2. 0: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Seeb, L. W., P. A. Crane, C. M. Kondzela, R. L. Wilmot, S. Urawa, N. V. Varnavskaya, and J. E. Seeb. 2004. Migration of Pacific Rim chum salmon on the high seas: insights from genetic data. Environmental Biology of Fishes 69:21–36.

Teel, D. J., G. B. Milner, G. A. Winans, and W. S. Grant. 2000. Genetic population structure and origin of life history types in Chinook salmon in British Columbia, Canada. Transactions of the American Fisheries Society 129:194–209.

Wattier, R., C. R. Engel, P. Saumitou-Laprade, and M. Valero. 1998. Short-allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus *Gv1ct* in *Gracilaria gracilis* (Rhodophyta). Molecular Ecology 7:1569–1573.

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.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323–354.

Xiong, F., R. A. Chin, and C. L. Hew. 1992. A gene encoding Chinook salmon (*Oncorhynchus tshawytscha*) prolactin: gene structure and potential cis-acting regulatory elements. Molecular Marine Biology Biotechnology 1:155–164.