Transactions of the American Fisheries Society 137:96–106, 2008

DOI: 10.1577/T07-011.1

[Note]

Identification of Novel Single-Nucleotide Polymorphisms in Chinook Salmon and Variation among Life History Types

NATHAN R. CAMPBELL AND SHAWN R. NARUM*

Columbia River Inter-Tribal Fish Commission, 3059F National Fish Hatchery Road, Hagerman, Idaho 83332, USA

Abstract.—Single-nucleotide polymorphisms (SNPs) are simple base substitutions or small indels in otherwise conserved regions of nuclear or mitochondrial DNA. Assays detecting these genetic markers have been a powerful tool for conserving and managing fish populations. In nonmodel species, however, available DNA sequence data are limited and inventive techniques must be employed in screening for SNPs. In this study, we used expressed sequence tags from rainbow trout *Oncorhynchus mykiss* to design primers for amplification of genomic DNA fragments from Chinook salmon *O. tshawytscha*. The regions flanking the repeat sequence of published microsatellite loci in Chinook salmon were also chosen to examine nucleotide variation. Thirty-two individuals representing the majority of the species' range from Russia to California were sequenced at 19 loci to identify SNPs. We discovered 58 polymorphisms during the screen and chose 13 for development into 5′ exonuclease assays (TaqMan assays). The genotypes from the TaqMan assays were compared with sequence data from individuals from the ascertainment panel to confirm proper allele designations. To evaluate the power of these markers to differentiate populations and life history types, we typed 91 individuals representing the major lineages of Chinook salmon with 11 validated assays. Significant allele frequency differences were observed among life history types of Chinook salmon at seven SNP loci, demonstrating the power of these markers to differentiate lineages.

Chinook salmon *Oncorhynchus tshawytscha* are the largest of the seven species of Pacific salmon. They spawn in drainages of the North Pacific Ocean in both North America and Asia. Several populations of this species have vanished over the past century due to human influence on spawning habitat (Nehlsen et al. 1991). Their economic importance to the fishing industry and cultural significance to indigenous peoples has given rise to extensive conservation efforts. Such efforts include hatchery supplementation, harvest management, and habitat improvement. Conservation genetics research allows evaluation of the effects of management efforts on populations and provides information to guide future decisions.

The two most distinct forms of Chinook salmon, known as the ocean and stream types, are characterized by differences in juvenile out-migration, adult run timing, and spawning location. Ocean-type Chinook salmon typically have a brief juvenile period in freshwater before migrating to the ocean as subyearlings and return as adults in summer and fall to spawn in relatively warm main-stem rivers (hence, they are also called "summer run" and "fall run"; Healey 1991). Stream-type Chinook salmon have a prolonged juvenile phase in freshwater and out-migrate as yearlings; the adults tend to make longer spawning migrations to small, cold tributaries in the spring (hence, they are also termed "spring run"; Healey 1991).

Population genetics studies in Pacific salmon, such as those related to population differentiation, parentage analysis, and genetic stock identification, rely on a panel of genetic markers. Such markers have included allozymes, microsatellites, mitochondrial DNA (mtDNA), and single-nucleotide polymorphisms (SNPs). In recent years, the most commonly used genetic marker for Pacific salmon has been microsatellites (Bentzen et al. 1991; Nelson et al. 1998; Nelson et al. 2001). The use of microsatellites over roughly the last 10 years has illuminated fine-scale population structure among Chinook salmon in numerous regions and helped shape fisheries management programs (Beacham et al. 2006). However, combining or sharing genotype data between laboratories requires costly and arduous standardization and maintenance efforts (Seeb et al. 2007). These technical issues and the potential for reduced costs have prompted many researchers to explore alternative types of genetic markers such as SNPs (Smith et al. 2005a).

Single-nucleotide polymorphisms are defined simply as single base substitutions in a sequence of DNA. A looser definition of the term, however, also includes small deletion and insertion polymorphisms (see the National Center for Biotechnology Information's Web site, www.ncbi.nlm.nih.gov/projects/SNP/). Reported to occur approximately once in every 300 base pairs for Chinook salmon (Smith et al. 2005a), SNPs are the most common type of genetic sequence variation (Morin et al. 2004). Screening for SNPs in salmonid species has become attractive in recent years because of a rapidly growing collection of expressed sequence

* Corresponding author: nars@critfc.org

Received January 11, 2007; accepted July 27, 2007
Published online January 14, 2008

tags (EST) and bacterial artificial chromosome (BAC) clone sequences from rainbow trout *O. mykiss* and Atlantic salmon *Salmo salar* that are available on public databases (e.g., www.ncbi.nlm.nih.gov/ and that of the Institute for Genomic Research [TIGR], www.tigr.org/). The genetic similarity between rainbow trout and other *Oncorhynchus* species allows the use of these DNA sequences for the design of polymerase chain reaction (PCR) primers capable of amplifying and sequencing orthologous regions in closely related species. Once a suitable variable-SNP site is discovered, high throughput genotyping assays can be designed to discriminate the two alleles.

Owing to the biallelic nature of SNPs, a much larger panel of loci is required for population genetics studies than is the case with multiallelic microsatellites (Kalinowski 2002; Morin et al. 2004; Anderson and Garza 2006). The advantages of using more loci with fewer alleles per locus include less propensity for homoplasy, better representation of the genome, and easily standardized data sets. The obvious disadvantage is the increase in genotyping assays required. However, many high-throughput SNP detection methods have been developed that allow researchers to rapidly process samples (reviewed in Syvanen 2001). Assays such as the 5′ exonuclease assay (TaqMan; Applied Biosystems) and the single base extension assay (SBE) allow rapid analysis of a single locus per reaction. Custom multiplex assay systems are also offered by several distributors. However, further screening of the Chinook salmon genome for variable SNP sites will be required to make the transition from microsatellites to SNP markers that take advantage of available high throughput technology.

In this study, we investigated regions of the Chinook salmon genome to identify novel SNP sites that are variable in the major lineages of the species. We targeted a combination of known EST regions and microsatellite flanking regions to screen for variations. The ascertainment panels were large and covered a broad geographic range because verified markers need to be applicable in studies throughout the species' range. We screened a diverse panel of Chinook salmon genomes at 25 loci for SNP variations and developed thirteen 5′ exonuclease (Taqman) assays. Verified assays were also used for genotyping 91 individuals of the two life history types of Chinook salmon to determine the utility of these markers for differentiating lineages (supplemental data available at www.critfc.org/tech_rep.html).

Methods

Primer design.—Sixteen rainbow trout ESTs were selected from the GenBank and TIGR databases for primer design. Nine microsatellite loci were also chosen for amplification and sequencing of DNA flanking the repeat region. Primers were designed using the database sequences and the Primer3 program (frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to generate a product of about 400 base pairs (bp; the primers are listed in Table 1). Because primer design was based on mRNA sequences, small fragment sizes allowed some flexibility in case an intron region was amplified in the genomic DNA. Primers were designed to yield products from both the translated region and the 3′ untranslated region (UTR). Each set of primers was then tested by polymerase chain reaction (PCR) using Chinook salmon genomic DNA under the following conditions: 1× Thermophilic DNA polymerase buffer (Promega), 2 mM $MgCl_2$, 0.25 mM deoxynucleotide triphosphates, 0.1 mg/mL bovine serum albumin, and 1 unit/reaction ampliTaq polymerase (Applied Biosystems) with 2 μL genomic DNA extract in 12 μL total volume. The thermal cycler conditions for PCR were typically 94°C for 1 min; 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min/kilobase; and a cool down to 4°C (performed on MJ Research thermal cyclers). Amplified fragments were analyzed by agarose gel electrophoresis for clean amplification of a single product. Primer pairs that produced a clean fragment were selected for amplification and sequencing of the ascertainment panel. In some cases we changed the PCR conditions to eliminate nonspecific products; in these cases, only the annealing temperature and extension time were changed (Table 1). Primer sets that amplified the 3′ UTR were chosen preferentially because we expected they were more likely to contain polymorphisms than the translated region.

Single-nucleotide polymorphism discovery.—Ascertainment samples were chosen from populations representing the major lineages of the species, which range from Russia to California (supplemental data available at www.critfc.org/tech/tech_rep.html). Samples include individuals from populations in California (Feather River Hatchery [spring and fall runs], Butte Creek, Battle Creek, Sacramento River, and Eel River), Oregon (Sandy River fall run, Applegate River fall run, and South Fork Umpqua River), Washington (South Fork Hoh River, Calawah River, and North Fork Hatchery [Washington Department of Fish and Wildlife], Carson Hatchery strain, Cowlitz Hatchery spring run), Canada (McQueston River), Idaho (Rapid and Salmon rivers), Alaska (King Salmon, Tahini, Nelson, Ayakulik, and Togiak rivers, Moose Creek, and Chignik Lagoon), and Russia (Bistraya River). For samples for which tissue was available, fresh extractions were performed using the Qiagen DNeasy 96 kit.

TABLE 1.—List of forward (fwd) and reverse (rev) primers used for polymerase chain reaction (PCR) and sequencing reaction at all Chinook salmon loci sequenced, the annealing temperature (Ta) used for each primer pair, the results of each PCR reaction, and sequenced bases and number and type of variations observed in the screening; NA = not applicable.

Locus	PCR	sequencing primers	Product size	Sequence size	Sequence direction	Ta (°C)	Introns	Number of		
								SNPs	Indels	Variable repeats
Cox-1	Fwd:	CCCATCTGTTCCCCTGAGTA	400	396	Fwd + rev	55	1	8	3	0
	Rev:	CGAAACCACACTGCTTTCAA								
Hsp-70	Fwd:	CCGGTAGGGAAATCCTTGTC	450	369	Fwd:	55	1	3	0	0
	Rev:	TGCAATGTCCAACAATGAAA								
Myostatin 1a	Fwd:	TCATGGAGGTGACGATTTCA	520	491	Fwd:	55	0	2	0	0
	Rev:	TCTATTGCACCGTGTTCTGC								
Myostatin 1b	Fwd:	CTTTGGCTGGGACTGGATTA	400	341	Fwd:	64	0	1	0	0
	Rev:	AGCGAATGGACAGGTGTTTC								
Nkef	Fwd:	TCCAAGCAGCAGTAAGACGA	400	317	Fwd:	55	0	1	0	0
	Rev:	GGAGTTTTCGCCACATGTTT								
MyoD	Fwd:	CCATGACCCCATCTACCAAG	500	442	Fwd + rev	65	1	6	0	1 (GT...)
	Rev:	AAGCGCCTCAGGCTACATTA								
Nramp-α	Fwd:	AACGACTTCCTCAACGTGCT	1,400	665	Fwd:	59	2	2	0	0
	Rev:	ACCACTCCTCCAACGTGTTC								
Hepcidin	Fwd:	AGGTGCTAACGGAGGAGGTT	550	NA	Fwd + rev	55	NA	NA	NA	NA
	Rev:	GGTTGCAGTGAGAAAGACATCA								
AldolaseB	Fwd:	AAACGACCACCACGTCTACC	800	667	Fwd:	63.5	2	8	0	1 (TC...)
	Rev:	GGGAATGTCAACATGGAAGG								
Asp-AT	Fwd:	TTCTGATCCGGCCAATCTAC	800	558	Rev:	59	3	3	0	1 (GT...)
	Rev:	ACCTTTGACATGGTGGCTTC								
Cpsbp	Fwd:	AAGACGATGCTTCCTCTCCA	800	551	Fwd:	63.5	1	3	1	1[a]
	Rev:	TGTGACTGTGGACCAGGTTG								
Gdh	Fwd:	TGCCTGTAGTCATTCCATCG	600	466	Fwd:	58	0	1	1	0
	Rev:	GCCCTAAGGATTCAGCACAA								
OtsG474	Fwd:	AAGGGCTTTGAGGATCAGAA	400	300	Fwd + rev	55	NA	2	0	1 (ATCT...)
	Rev:	AAACGGAGAGGCGAATATGA								
OtsG68/208	Fwd:	CCAAGTGCAAGATCTGACCTC	400	294	Fwd + rev	55	NA	1	0	1 (AGAT...)
	Rev:	CCCATGTTTCATGCTCACAC								
Ots213	Fwd:	CCATTTGGTTCCGCTCTTTA	400	284	Fwd:	58	NA	1	0	1 (ATAC...)
	Rev:	GGGCTGTTTTTGTGCTTTTC								
Ogo4	Fwd:	GTCGTCACTGGCATCAGCTA	350	276	Rev:	58	NA	2	0	1 (GT...)
	Rev:	TCCAGGGTATTGTTCTAGACTGG								
Ots3	Fwd:	GCTCCAACAATCCCAGTTCT	180	69	Fwd + rev	55	NA	0	0	1 (CT...)
	Rev:	AGTGCTGTCCAAAGGTGAGC								
OtsG311	Fwd:	GTTCTTCCCTTGATGCAGGA	450	180	Fwd:	58	NA	1	0	2[b]
	Rev:	CCCCATCCATTGTTCTATCG								
OtsG249	Fwd:	CTCGCTTGGTTATGGAGGAG	450	294	Fwd:	65.5	NA	1	0	1 (AGAT...)
	Rev:	TGCTGACCTGTGAGTCCAAG								
OtsG78b	Fwd:	CCATTTGCACATCGTTACAA	500	293	Fwd:	58	NA	2	0	1 (ATAG...)
	Rev:	CCCTAGCAACAGCAGCCTAC								

[a] 41-bp minisatellite.
[b] AGGC ... ATAG ....

The remaining samples were collected from frozen DNA archives. A total of 47 samples were initially chosen as a template for amplification; the number of samples was later reduced to 32. Redundant and poor-quality samples were removed from the ascertainment panel. Five loci were screened using the 47-sample panel, and 14 were screened using the 32-sample panel. Each locus was amplified using AmpliTaq polymerase (Applied Biosystems) and MJ research thermal cyclers following standard or optimized PCR conditions from Table 1. Following PCR, 4 μL of each reaction was examined by agarose gel electrophoresis for clean amplification. The remaining PCR products were then treated with Exonuclease 1 (NEB) and shrimp alkaline phosphatase (Amersham) before sequencing. Sequencing reactions were performed using the Big Dye version 1.1 sequencing kit (ABI) and MJ research thermal cyclers. The reactions were cleaned up by ethanol–EDTA–NaAcetate precipitation before separation by capillary electrophoresis using a 3730 instrument (ABI).

Following electrophoresis, chromatogram data were analyzed by first using the Sequencing Analysis Program version 5.4 (ABI) to assign base calls. Then, Sequencher version 4.6 (Gene Codes) was used to align and edit the data. Sequencing data were used in blast search queries (www.ncbi.nlm.nih.gov/BLAST/) to determine whether the correct sequence had been amplified. Sequence alignments against rainbow trout EST sequences and the GeneBuilder program (www.itb.cnr.it/sun/webgene/)

TABLE 2.—Descriptions of all validated forward (fwd) and reverse (rev) primer assays. Assays are named after the single-nucleotide polymorphism (SNP) locus they are designed to detect. Assay names ending in "*x*" denote assays designed to detect indels; names ending in "*m*" were designed to detect multinucleotide polymorphism variations (AAC/TAT). All probes were labeled with VIC or 6FAM fluorescent dyes on their 5′ end and a nonfluorescent quencher and minor groove binder on their 3′ end. The primer–probe sequences for failed assays (*Ots*_hsp70–249 and *Ots*_cpsbp-89) are available on request; TR = translated region; MF = microsatellite flanking.

Assay		Primers	Probes	Gene region
*Ots*_cox1–241				
fwd primer	Fwd:	CACTGAACTGTAAGCCATTGTGATT	VIC-CACTACGGTAAGACCAT	3′ UTR
rev primer	Rev:	GTAAATGTAGTATACAGTATAGGCATCGTAGGT	FAM-CACTACAGTAAGACCAT	
*Ots*_nkef-192				
fwd primer	Fwd:	CATTTAGCAGACACTCTTATCTTAGTGTCA	VIC-AATAGGCCGACATCAA	3′ UTR
rev primer	Rev:	CGAATGTCCACCTCAGATGTTACAA	FAM-AAATAGGCCAACATCAA	
*Ots*_Ots474–480				
fwd primer	Fwd:	TCCTTGTGTCTAAAGGGCTTTGAG	VIC-CAGAATTAGCTTTGGACATT	MF
rev primer	Rev:	GGGCTTGCTAGTCCTAAACAGATC	FAM-TCAGAATTAGCTTTTGACATT	
*Ots*_myo1a-384				
fwd primer	Fwd:	CTCCCCCCTGGACTTTGG	VIC-ACAGATCCATCCACCACT	3′ UTR
rev primer	Rev:	GCTCTATTGCACCGTGTTCTG	FAM-AGATCCAGCCACCACT	
*Ots*_nramp-321				
fwd primer	Fwd:	GGCCATCTTTCAGGACGTACAG	VIC-CTAACGGGCATGAACGA	TR (silent)
rev primer	Rev:	GCATGCTCTGCAATACGTTGAG	FAM-ACGGGCATGAATGA	
*Ots*_gdh-81x				
fwd primer	Fwd:	CTTTTCTGAATTAGTGCTGTGCTTGT	VIC-TGTTACGGGACATACT	3′ UTR
rev primer	Rev:	CCAACTTCTTCAACTCTGTCAGTGA	FAM-TCTGTTACGGACATACT	
*Ots*_aspat-196				
fwd primer	Fwd:	CCTGAACAGGTACACACAAACGA	VIC-CACACCCACTCTTTAT	Intron
rev primer	Rev:	TCCAACTGATGAATATGACCAACATGAAT	FAM-CACACCCAGTCTTTAT	
*Ots*_myoD-364				
fwd primer	Fwd:	GTGTGTGTGTGTGTGTGTCATC	VIC-TCATCTTTTGTTATTTCCTTG	3′ UTR
rev primer	Rev:	TTTACACATATACAAAAATGGTCCTCTATTGTCAT	FAM-ATCTTTTGTTCTTTCCTTG	
*Ots*_Ots311-101x				
fwd primer	Fwd:	AAATGAGGCCGTCCTTTACACT	VIC-CTGAGATCACTTTGAGCAC	MF
rev primer	Rev:	GCAATACAAGCCCTTGATAATGAAGT	FAM-ACTGAGATCACTGAGCAC	
*Ots*_aldb-177M				
fwd primer	Fwd:	GCGATCAGGTGACGCTAAAATGA	VIC-CCAAATTGCTTAACCC	Intron
rev primer	Rev:	AGGAAGGTGATGCCTGAGAGA	FAM-CCAAATTGCTTTATCC	
*Ots*_Ots78B–71				
fwd primer	Fwd:	GCACATCGTTACAACACTGTATATAGACA	VIC-ACAGAAGATCCAAAATAGTAA	MF
rev primer	Rev:	AACAATACAAATAAACAGTCAACATCACACTC	FAM-AGAAGATCCAAAATAATAA	

were used to annotate the coding regions, UTRs, and introns. Annotated genomic sequence data generated using EST-derived primers were submitted to GenBank (www.ncbi.nlm.nih.gov/Genbank/index.html). Observed SNPs and small indels were recorded and minor allele frequencies were calculated for each. All observed SNP sites were submitted to the database dbSNP (www.ncbi.nlm.nih.gov/projects/SNP/). When multiple SNP sites were observed within a single gene, the polymorphism with the highest frequency minor allele was chosen for assay design.

Single-nucleotide polymorphism assays.—Custom Taqman allelic discrimination assays (5′ exonuclease assays) were ordered through Applied Biosystems by using their File Builder software to submit sequence data. Primer and probe mix (40×) for each submitted SNP was designed and produced by Applied Biosystems. A list of the primers and probes for each designed SNP assay is given in Table 2. Assays were performed in 384 well plates using approximately 25 ng of Chinook salmon genomic DNA and 5 μL of 1× TaqMan Universal PCR Master Mix (No AmpErase UNG, ABI) with 1× primer–probe mixture. Amplification was performed using MJ research thermal cyclers with 384-well blocks and standard two-step cycling for 50 cycles. Endpoint reads of each reaction were collected using an Applied Biosystems 9700HT instrument. Data were analyzed using the allelic discrimination function of Applied Biosystems SDS v.2.1 software. All assays were validated by comparing allele calls with sequencing data for agreement. Validated assays were used to genotype 27 spring-run, 32 summer-run, and 32 fall-run Chinook salmon collected at Bonneville Dam in 2004 as a preliminary test of the power of these markers to differentiate fish based on run timing. Fisher's exact test was implemented in GENEPOP (Raymond and Rousset 1995) to determine statistical significance of allele frequency differences at each locus between spring, summer, and fall run Chinook salmon. Hardy–Weinberg and linkage

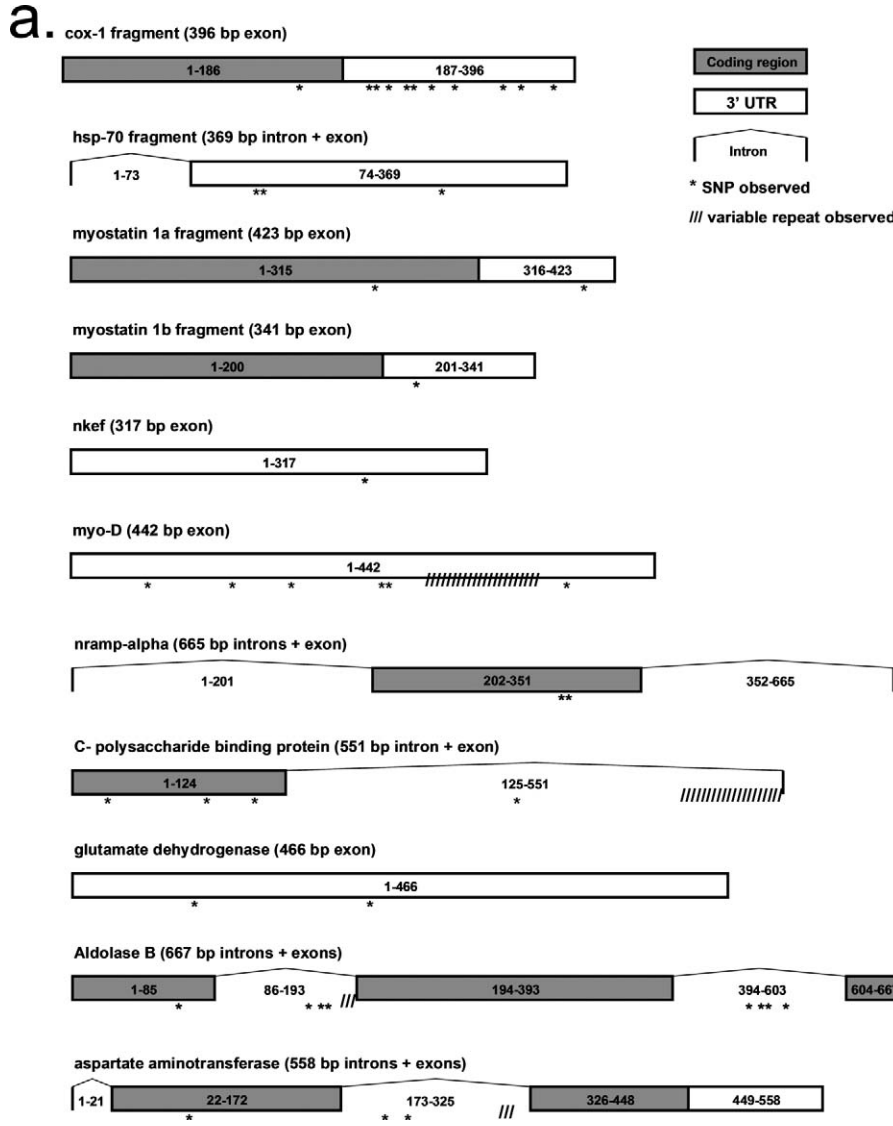

FIGURE 1.—Graphical representation of all sequenced Chinook salmon loci. Panel **(a)** presents models of each sequenced fragment from a known gene region. The models are broken down into coding, 3′ untranslated, and intron regions, and all variations observed in screening are marked. Nonsynonymous single-nucleotide polymorphism (SNP) sites were observed only in the C-polysaccharide binding protein locus. Panel **(b)** presents models of each sequenced fragment from each microsatellite locus, showing sequenced DNA regions in relation to the repeated sequence. The repeated sequence was not included in the numbering of sequenced base pairs (bp).

disequilibrium tests were also completed in GENEPOP. Because multiple comparisons were involved, corrections in both tests were made against type I error via the Bonferroni method (Rice 1989).

Results

Primer Design

Fifty-seven primer pairs were tested against Chinook salmon genomic DNA for amplification of 25 loci (Table 3). Twenty-nine pairs produced single bands by agarose gel electrophoresis at 19 loci under standard PCR conditions (data not shown). For loci with multiple working primer pairs, those that amplified the 3′ region were chosen preferentially for screening. Another six primer pairs at the remaining six loci were chosen for optimization of PCR conditions. Of the 25 loci, 4 failed to produce a single band by agarose gel analysis under any tested PCR condition. Amplified

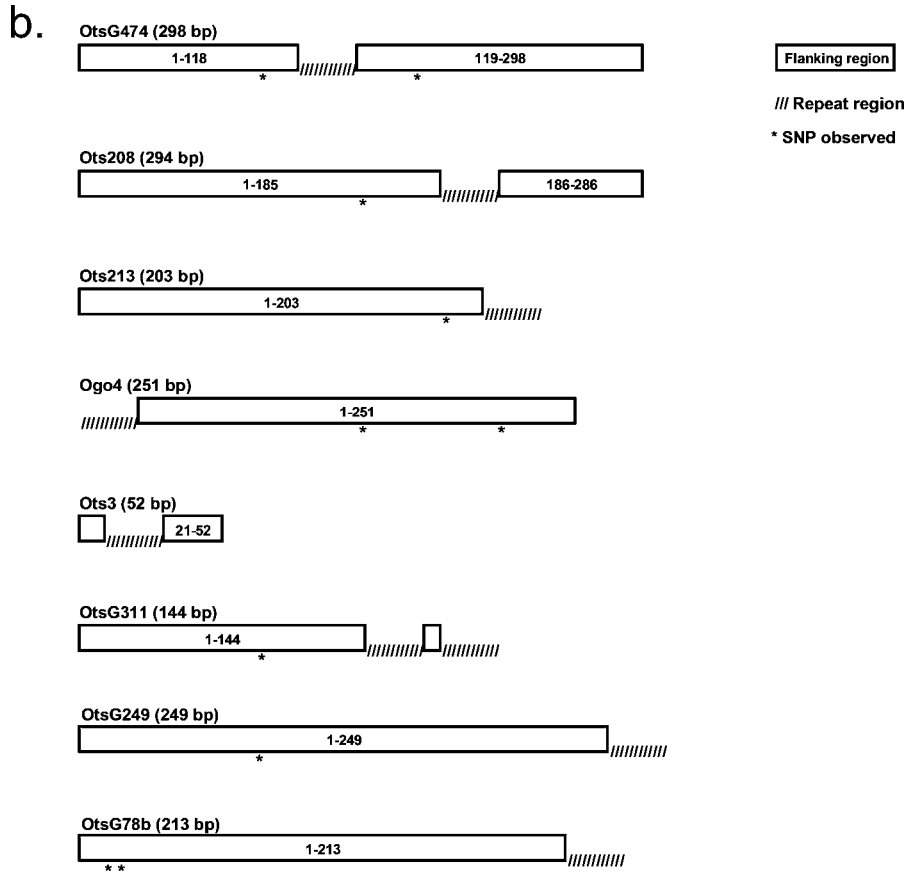

FIGURE 1.—Continued.

products from microsatellite regions often produced two bands attributable to a significant difference in the repeat size between the two alleles. Interestingly, two bands were also observed in the amplified product from the *myoD* locus. After sequencing of this locus, a polymorphic dinucleotide microsatellite repeat was discovered (Figure 1). Of the 11 gene fragments amplified, 4 yielded products significantly larger than predicted by the rainbow trout EST sequence data, indicative of intron regions.

Single-Nucleotide Polymorphism Discovery

Sequencing data were obtained at 19 loci, approximately 32 samples per locus producing usable sequences. A total of 770 sequences were produced during the screen and compiled into 19 alignments for screening. We screened 7,253 bases of consensus sequence for variations, with an SNP occurring once in approximately 134 bp. Genotypes of each individual from the screen were recorded for each variation observed along with the minor allele frequency for each SNP. The screen revealed a total of 58 previously unreported variations including four microsatellite–minisatellite repeats (Tables 1, 3).

Blast searches against consensus sequences confirmed amplification of all targeted gene fragments. Annotation of consensus sequence was performed before submission to GenBank and dbSNP (Table 3). A graphical view of all genomic fragments amplified, including SNP locations, is shown in Figure 1. As predicted, intron regions were revealed in samples where larger than expected fragments were produced by PCR. Of the 43 SNPs from gene regions, 24 were located in the 3′ UTR, compared with 10 in introns, and 9 in coding regions. Evaluation of the nine SNP sites located in coding regions revealed that six of them maintained synonymous amino acid sequence in the translated protein, and three (*Ots*_cpsbp-35, *Ots*_cpsbp-89, and *Ots*_cpsbp-del) were nonsynonymous mutations. Surprisingly, all nonsynonymous mutations were from the same gene, C-polysaccaride binding protein, coding for R/H, T/M, and RKSCN/*

TABLE 3.—Accession numbers for all Chinook salmon loci used in single-nucleotide polymorphism (SNP) screening. Accession numbers for DNA sequences used in primer design are shown with the targeted gene region. GenBank accession numbers for DNA sequences were generated during the screen, except where not applicable (NA). Sequenced tagged site (STS) numbers were generated as a reference to all reported sequence variations. Variations observed at each locus are listed by name and dbSNP accession number. Minor allele frequencies are shown for each SNP. GenBank and dbSTS accession numbers are not shown for loci that did not produce a DNA sequence. The dbSTS numbers at microsatellite loci were not generated but were instead cross-referenced to the original GenBank number.

		Accession number				
Locus	Primer design	GenBank	dbSTS	SNP site	dbSNP	Minor allele frequency (q)
Cox-1	AJ299018	EF042591	BV686590	*Ots*_cox1_156	ss65917702	0.18
				*Ots*_cox1_216	ss65917703	0.12
				*Ots*_cox1_219	ss65917704	0.02
				*Ots*_cox1_223	ss65917705	0.38
				*Ots*_cox1_229	ss65917706	0.45
				*Ots*_cox1_230a	ss65917707	[a]
				*Ots*_cox1_230b	ss65917708	[a]
				*Ots*_cox1_241	ss65917709	0.4
				*Ots*_cox1_250i	ss65917710	0.11
				*Ots*_cox1_302	ss65917711	0.27
				*Ots*_cox1_328	ss65917712	0.12
				*Ots*_cox1_378	ss65917713	0.32
Hsp-70	AB176854	EF042592	BV686591	*Ots*_hsp70_124	ss65917714	0.04
				*Ots*_hsp70_127	ss65917715	0.4
				*Ots*_hsp70_249	ss65917716	0.39
Myostatin 1a	AF273035	EF042593	BV686592	*Ots*_myo1a_198	ss65917717	0.02
				*Ots*_myo1a_384	ss65917718	0.08
Myostatin 1b	AF273036	EF042594	BV686593	*Ots*_myo1b_227	ss65917719	0.01
Nkef	AF250195	EF042595	BV686594	*Ots*_nkef_192	ss65917720	0.3
MyoD	Z46924	EF042596	BV686595	*Ots*_myoD_60	ss65917721	0.08
				*Ots*_myoD_124	ss65917722	0.03
				*Ots*_myoD_172	ss65917723	0.05
				*Ots*_myoD_211	ss65917724	0.27
				*Ots*_myoD_214	ss65917725	0.08
				*Ots*_myoDu	ss65978922	[b]
				*Ots*_myoD_364	ss65917726	0.4
Myosin	Z48794	NA	NA			
Nramp-α	AF048760	EF042597	BV686596	*Ots*_nramp_321	ss65917727	0.5
				*Ots*_nramp_327	ss65917728	0.5
Hepcidin	AF281354	NA	NA			
Hsf-1	AB062548	NA	NA			
AldolaseB	TC18597[c]	EF042598	BV686597	*Ots*_aldB_66	ss65917729	0.09
				*Ots*_aldB_177	ss65917730	0.12
				*Ots*_aldB_179	ss65917731	0.12
				*Ots*_aldB_182	ss65917732	0.02
				*Ots*_aldBu	ss65978923	-
				*Ots*_aldB_485	ss65917733	0.04
				*Ots*_aldB_492	ss65917734	0.02
				*Ots*_aldB_504	ss65917735	0.05
				*Ots*_aldB_548	ss65917736	0.04
Cpsbp	AF281345	EF042599	BV686598	*Ots*_cpsbp_35	ss65917737	0.09
				*Ots*_cpsbp_89	ss65917738	0.19
				*Ots*_cpsbp_del	ss65917739	0.07
				*Ots*_cpsbp_299	ss65917740	0.08
				*Ots*_cpsbp_mini	ss65978924	[b]
Gdh	AJ556997	EF042600	BV686599	*Ots*_gdh_81	ss65917741	0.28
				*Ots*_gdh_157	ss65917742	0.13
Asp-AT	TC9251[c]	EF042601	BV686600	*Ots*_aspAT_79	ss65917743	0.12
				*Ots*_aspAT_196	ss65917744	0.12
				*Ots*_aspAT_213	ss65917745	0.09
				*Ots*_aspATu	ss65978925	[b]
Calmodulin	TC25558[c]	NA	NA			
D6-fad	AY458652	NA	NA			
OtsG474	AF393200	AF393200	NA	*Ots*_ots474_80	ss65917746	0.07
				*Ots*_ots474_129	ss65917747	0.03
OtsG311	AF393194	AF393194	NA	*Ots*_otsG311d	ss65917748	0.17
OtsG78b	AF393188	AF393188	NA	*Ots*_otsG78b_66	ss65917749	0.14
				*Ots*_otsG78b_71	ss65917750	0.14
OtsG249	AF393192	AF393192	NA	*Ots*_ots249_151	ss65917751	0.39
OtsG68/208	AF393187	AF393187	NA	*Ots*_ots208_145	ss65917754	0.02
Ots213	AJ534363	AJ534363	NA	*Ots*_ots213_292	ss65917755	0.05

TABLE 3.—Continued.

Locus	Primer design	Accession number		SNP site	dbSNP	Minor allele frequency (q)
		GenBank	dbSTS			
Ogo4	AF009796	AF009796	NA	*Ots*_ogo4_109	ss65917752	0.05
				*Ots*_ogo4_222	ss65917753	0.03
Ots9	AF107037	AF107037	NA			
Ots3	AF107031	AF107031	NA			

[a] Contains three possible alleles at this locus. Minor allele frequency not calculated.
[b] Repeat region. Minor allele frequency data not reported.
[c] Accession number from the Institute for Genomic Research database.

amino acid variations, respectively (standard single-letter abbreviations for amino acids).

Assay Development

Thirteen SNP loci were chosen, based on minor allele frequency data and distribution of alleles, for development into 5′ exonuclease allelic discrimination assays (Table 2). Each assay was tested against 24 samples from the ascertainment panel to determine the accuracy of the allele calls. When the allele calls from assay data were compared with those made by sequence analysis, only assay *Ots*_hsp70–249 showed deviation from sequence data. This assay was retested using both higher and lower annealing temperatures with the same results (data not shown). The sequencing data for samples not in agreement was found to be of high quality, and an explanation for the incongruity could not be established. Assay *Ots*_hsp70–249 was therefore precluded from subsequent testing.

The 12 validated assays were then genotyped in 91 individuals representing the major lineages of Chinook salmon. Genotypes were recorded for each individual at each tested SNP locus (supplemental data available at critfc.org/tech_rep.html). Assay *Ots*_cpsbp-89 showed an excess of heterozygotes, and subsequent examination of sequence data revealed the reverse primer was designed over a deletion. For these reasons, this locus was removed from the data set. Allele frequencies at each locus were recorded for spring-, summer-, and fall-run fish, as a preliminary measure of the ability of these assays to distinguish between the major lineages of Chinook salmon.

Significant heterozygote deficiencies at the 11 validated loci were observed in the spring-run collection at *Ots*_gdh-81x ($P = 0.0072$) and the summer-run collection at *Ots*_nkef-192 ($P = 0.0035$) and *Ots*_nramp-321 ($P = 0.000$). However, because our collections were a mixture of unknown populations, separated only by run timing, these results were not unexpected. Pairwise exact tests of allele frequencies reveal significant differences (Bonferroni corrected critical value of $P \leq 0.0167$) between the spring run (stream type) and summer and fall runs (ocean type) at seven loci. Statistical significance between summer-run and fall-run allele frequencies was observed only at *Ots*_cox1–241 (Figure 2). However, *Ots*_nramp-321 showed a *P*-value of 0.0182, just above the Bonferroni corrected critical value but well below the less conservative Benjamini and Yekutieli correction of $P \leq 0.02727$ (Narum 2006). The assay for *Ots*_Ots474–480 was fixed for the guanine allele in all individuals tested (variation at this locus was observed in Oregon and California ascertainment samples only).

Run timing has been used in this study to differentiate ocean-type and stream-type fish. However, there are rare populations of late-running stream-type Chinook salmon that exist in the Columbia River drainage (Snake River). Therefore, there is a very small possibility that a few samples from our summer-run collection had the stream-type life history, although our data do not suggest that this is the case.

Discussion

Primer Design

Genetic sequences from related species have been used for primer design and amplification on numerous occasions in many diverse applications (Aitken et al. 2004; Frieson et al. 1999; Primmer et al. 2002). The technique of using primers designed based on rainbow trout EST sequences for amplification of DNA from other *Oncorhynchus* species has been successful in other studies (e.g., Smith et al. 2005). Despite the successful use of this technique, failure to amplify any product by PCR or to produce the expected ortholog when applied to another species is a common outcome. In our case, 5 of 16 target loci based on rainbow ESTs failed to produce sequencing data in Chinook salmon. The failure rate of this technique is likely not due to sequence dissimilarity alone, but also factors such as primer secondary structure, nonspecific binding, and gene duplication. In particular, gene duplication must be considered when dealing with salmonids, as duplication events have been reported as high as 50% for enzyme encoding loci (Allendorf and Thorgaard

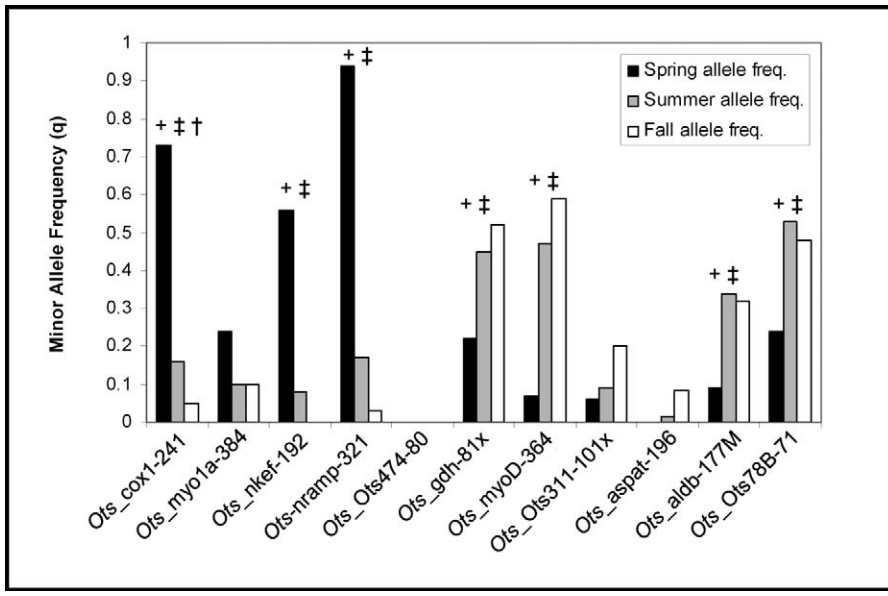

FIGURE 2.—Allele frequencies by Chinook salmon life history types (spring, summer, and fall runs). Validated assays from the screen were tested for their ability to distinguish fish by run timing. Differences in allele frequency were tested for statistical significance by Fisher's exact test; a single plus sign denotes a significant difference between the spring and summer runs, a double plus sign a significant difference between the spring and fall runs, and a dagger a significant difference between the summer and fall runs.

1984). Even when clean sequencing data are obtained for a given primer set, the possibility that the sequence represents four separate alleles and not two must be considered.

Single-Nucleotide Polymorphism Discovery

An important consideration in SNP discovery is ascertainment bias, whereby sample selection influences the SNPs observed in the screen by either a deficiency or overabundance of relevant populations (Carlson et al. 2003; Morin et al. 2004). We attempted to address this concern by including 32–48 individuals representing a broad geographic area. In this way, the representation of SNP variation across multiple populations could be considered similar. However, unless a very large number of samples are included, complete representation of genomic diversity across the species' range cannot be achieved.

An interesting result from the screen was the discovery of variable repeats linked to functional genes. Four variable repeat regions were observed in the 11 gene regions sequenced. The most interesting of these was a highly variable dinucleotide microsatellite repeat observed in the 3′ region of the *myoD* gene. Microsatellite repeats in Chinook salmon are often associated with heterochromatin or are in unknown regions of the genome and therefore classified as neutral genetic markers. However, although not functional at the protein level, variable repeats could influence protein expression at the levels of transcription and translation. For instance, microsatellite repeats existing in an intron may be a binding site for transcription factors (Janaki Ramaiah and Parnaik 2006) or be involved in mRNA splicing (Hui et al. 2005), whereas those in the 3′ UTR may be functional in mRNA stability (Chen et al. 2007). Therefore, the true influence of selection on microsatellite loci may not be realized.

Single-Nucleotide Polymorphism Assays

Thirteen 5′ exonuclease assays were developed based on the SNP variations observed in the screen. Assay *Ots*_hsp70–249, however, failed to accurately determine the proper allele, and a design flaw was discovered in assay *Ots*_cpsbp-89. Also, a null allele in assay *Ots*_ots78b-71 has been detected during a related study. The null allele appears to occur at very low frequencies in the Columbia River but at much higher frequencies in regions of northern British Columbia and southeast Alaska (unpublished data). Overall, the success rate of the assays developed here (~85%) is significantly higher than the approximately 50% rate reported by Smith et al. (2005a). However, three of our assays were designed in microsatellite regions expected

to be diploid, so this probably influenced the result. Moreover, several loci either failed to amplify a clean product or failed to provide clean sequencing data, which may be indicative of gene duplication (Smith et al. 2005a). This may have lead to an enrichment of single locus gene regions from which to choose SNP sites for assay design.

Also of note, assay *Ots*_aldB-177M was designed to detect a multinucleotide polymorphism site (reported as AAC/TAT). This site could also be interpreted as two separate SNP sites separated by a single base (A/T and C/T), although all samples sequenced in the ascertainment panel showed complete linkage between the two polymorphisms. The later testing of this assay on 91 samples showed no anomalies in the sample set, which is indicative of segregation of the two SNPs. Therefore, it is our contention that these two SNP sites represent a single mutation event and that our assay is designed to detect the only possible genotypes. However, the assay design for this site may be an important consideration in the event of future irregular results derived by this assay.

Preliminary testing of the validated assays using the spring, summer, and fall runs collected at Bonneville Dam was intended to illustrate differences in allele frequencies as a qualitative assessment of each locus. Although seven loci showed statistical significance between the stream-type (spring-run) and ocean-type (summer- and fall-run) fish, four of the loci (*Ots*_myo1a–384, *Ots*_Ots474–80, *Ots*_Ots311-101x, and *Ots*_aspat-196) showed no statistical significance between any of the three collections. However, due to our efforts to eliminate regional bias in our ascertainment panel, this result is not surprising. In the case of loci without significant differences in allele frequencies, the frequency of the minor allele is quite low across all samples or, as in the case of *Ots*_Ots474–80, was not observed in any individual. There may however be populations both outside and within the Columbia River basin that display higher frequencies of these alleles, and these assays may be quite useful when used on larger sample sizes for finer scale differentiation (Smith et al. 2005b).

Applications of Single-Nucleotide Polymorphisms for Conservation Genetics

Recent studies have demonstrated the utility of SNP markers to address questions pertinent to evolution and conservation genetics (Morin et al. 2004; Seddon et al. 2005). The novel SNPs reported in this study, along with previously published assays (e.g., Smith et al. 2005a), provide a suite of markers to utilize in future research. Simulations suggest that a large number of SNPs are necessary to provide similar statistical power as microsatellites (Kalinowski 2002; Anderson and Garza 2006), indicating additional SNP discovery is necessary. However, there may be applications in which a relatively small number of SNPs are adequate to address specific research questions (e.g., Smith et al. 2005b). In cases where genetic distance among collections is large, as few as 5–10 SNP loci may differentiate species or major lineages within species. For example, we observed significant differences in allele frequencies at seven SNPs among life history types of Chinook salmon that could be utilized to differentiate these types in mixed aggregates. Further, multilocus genotypes of multiple marker types such as microsatellites and SNPs can be combined, and this approach offers the potential for custom panels of loci with the greatest power to address specific questions. In any case, additional genetic markers offer further options to improve design of molecular research.

Acknowledgments

We are grateful to Ken Overturf for providing EST sequences for target genes from rainbow trout. Christian Smith and Ruth Phillips provided insight regarding the discovery and ascertainment approach. Supplemental tables of ascertainment samples and genotypes for each locus are available online at www.critfc.org/tech/tech_rep.html. This research was supported by funding from the Pacific Salmon Commission, Chinook Technical Committee.

References

Aitken, N., S. Smith, C. Schwartz, and P. A. Morin. 2004. Single-nucleotide polymorphism (SNP) discovery in mammals: a targeted gene approach. Molecular Ecology 13:1423–1431.

Allendorf, F. W., and G. H. Thorgaard. 1984. Tetraploidy and the evolution of salmonid fishes. Pages 1–53 *in* B. J. Turner, editor. Evolutionary genetics of fishes. Plenum, New York.

Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567–2582.

Beacham, T. D., J. R. Candy, K. L. Jonsen, J. Supernault, M. Wetklo, L. Deng, K. M. Miller, and R. E. Withler. 2006. Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim using microsatellite variation. Transactions of the American Fisheries Society 135:861–888.

Bentzen, P., A. S. Harris, and J. M. Wright. 1991. Cloning of hypervariable minisatellite and simple-sequence microsatellite repeats for DNA fingerprinting of important aquacultural species of salmonids and tilapia. Pages 243–262 *in* T. Burke, G. Dolf, A. J. Jeffreys, and R. Wolff, editors. DNA fingerprinting: approaches and applications. Birkhauser-Verlag, Basel, Switzerland.

Carlson, C. S., M. A. Eberle, M. J. Rieder, J. D. Smith, L. Kruglyak, and D. A. Nickerson. 2003. Additional SNPs and linkage-disequilibrium analyses are necessary for

whole-genome association studies in humans. Nature Genetics 33:518–521.

Chen, T. M., P. L. Kuo, C. H. Hsu, S. J. Tsai, M. J. Chen, C. W. Lin, and H. S. Sun. 2007. Microsatellite in the 3′ untranslated region of human fibroblast growth factor 9 (FGF9) exhibits pleiotrophic effect on modulating FGF9 protein expression. Human Mutation 28:98.

Friesen, V. L., B. C. Congdon, M. G. Kidd, and T. P. Birt. 1999. Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. Molecular Ecology 8:2147–2149.

Healey, M. C. 1991. Life history of Chinook salmon (*Oncorhynchus tshawytscha*). Pages 311–395 *in* C. Groot and L. Margolis, editors. Pacific salmon life histories. UBC Press, Vancouver.

Hui, J., L. H. Hung, M. Heiner, S. Schreiner, N. Neumuller, G. Reither, S. A. Haas, and A. Binderief. 2005. Intronic CA repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. EMBO (European Molecular Biology Organization) Journal 24:1988–1998.

Janaki Ramaiah, M., and V. K. Parnaik. 2006. An essential GT motif in the lamin A promoter mediates activation by CREB binding protein. Biochemical and Biophysical Research Communications 348:1132–1137.

Kalinowski, S. T. 2002. How many alleles per locus should be used to estimate genetic distances? Heredity 88:62–65.

Morin, P. A., and 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.

Narum, S. R. 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. Conservation Genetics 7:783–787.

Nehlsen, W., J. E. Williams, and J. A. Lichatowich. 1991. Pacific salmon at the crossroads: stocks at risk from California, Oregon, Idaho, and Washington. Fisheries 16(2):4–20.

Nelson, R. J., T. D. Beacham, and M. P Small. 1998. Microsatellite analysis of the population structure of a Vancouver Island sockeye salmon (*Oncorhynchus nerka*) stock complex using nondenaturing gel electrophoresis. Molecular Marine Biology and Biotechnology 7:312–319.

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. U.S. National Marine Fisheries Service Fishery Bulletin 99:94–107.

Primmer, C. R., T. Borge, J. Lindell, and G. P. Saetre. 2002. Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Molecular Ecology 11:603–612.

Raymond, M., and F. Rousset. 1995. 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.

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

Seeb, L. W., A. Antonovich, M. A. Banks, T. D. Beacham, M. R. Bellinger, S. M. Blankenship, M. Campbell, N. A. Decovish, J. C. Garza, C. M. Guthrie III, T. A. Lundrigan, P. Moran, S. R. Narum, J. J. Stephenson, K. J. Supernault, D. J. Teel, W. D. Templin, J. K. Wenburg, S. F. Young, and C. T. Smith. 2007. Development of a standardized DNA database for Chinook salmon. Fisheries 32:540–552.

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

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

Syvanen, A. C. 2001. Accessing genetic variation: genotyping single-nucleotide polymorphisms. Nature Reviews Genetics 2:930–942.