Molecular Ecology Notes (2006) doi: 10.1111/j.1471-8286.2006.01320.x

PRIMER NOTE

Characterization of 19 single nucleotide polymorphism markers for coho salmon

C. T. SMITH,* L. PARK,† D. VANDOORNIK,† L. W. SEEB* and J. E. SEEB*
*Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, Alaska 99518, USA, †National Marine Fisheries Service, 2725 Montlake Boulevard, Seattle, Washington 98112, USA

Abstract

We report 39 single nucleotide polymorphisms (SNPs) observed in 23 nuclear DNA sequences in coho salmon *Oncorhynchus kisutch*. High-throughput genotyping assays based on the 5′-nuclease reaction were developed for 17 of these nuclear SNPs and for two previously published mitochondrial DNA SNPs. Minor allele frequency differences (Δq) among collections were between 5.2% and 51.2%, resulting in per locus F_{ST} estimates of 0.00–0.24 with an average of 0.09.

Keywords: 5′-nuclease reaction, coho salmon, *Oncorhynchus kisutch*, SNP, *Taq*Man

Received 22 November 2005; revision received 8 December 2005; accepted 8 January 2006

Coho salmon *Oncorhynchus kisutch* are anadromous, spawning in streams and tributaries of rivers that drain into the north Pacific Ocean from Asia and North America. Coho salmon are harvested in freshwater and near-shore areas that often do not coincide with spawning sites and thus contain mixtures of individuals from many populations. The sustainability of these harvests depends on the ability of resource managers to estimate the origins of the fish harvested, a task increasingly done using genetic data. Genetic markers requiring electrophoresis of proteins (Utter *et al*. 1970), coding DNA (Beacham *et al*. 2001) and microsatellite DNA (Smith *et al*. 1998) have been previously described in coho salmon. Here we describe 19 single nucleotide polymorphism (SNP) markers for use in this species.

DNeasy 96 tissue kits (QIAGEN) were used to extract DNA from 347 coho salmon fin tissue samples that had been preserved in 95% EtOH for one to nine years prior to the present work. These samples included 16 individuals from Kamchatka, 48 from the Kuskokwim River, 54 from Bristol Bay, 85 from Cook Inlet, 48 from south-east Alaska, and 96 from Washington/Oregon. Each collection was taken from a single population, except the Washington/Oregon collection, which contained eight representatives from each of 12 populations. DNA from 37 of these individuals (10 from the Kuskokwim River, 10 from Bristol Bay, 7 from Cook Inlet, and 10 from southeast Alaska) were amplified and sequenced in both directions using 25 sets of previously described polymerase chain reaction (PCR) primers and methods (Table 1; Smith *et al*. 2005). PHRED, PHRAP, CONSED (http://www.phrap.org) and POLYPHRED (Nickerson *et al*. 1997) were used to call bases, assemble sequences into contigs, view assemblies and identify SNPs.

Twenty-three of the 25 primer sets tested produced clean sequence data (Table 1). Five of the 23 loci exhibited no variation among the samples examined and 17 loci exhibited one to four SNPs each. The remaining locus (u202_2) contained 10 SNPs and one 7-bp indel, suggesting that the observed variation was likely paralogous rather than allelic. In total, 8966 bp were sequenced and 39 SNPs and one 7-bp indel were observed (dbSNP ss49845898–ss49845937). Consensus sequences were compared to GenBank nucleotide and amino acid databases using BLASTN and BLASTX. All loci but three had significant (*e*-values $\leq 10^{-5}$) matches to the expected loci; the primer sets ZNF330_1, arf_1, and GPDH_3 produced sequences that yielded no significant hits.

We used Assays-by-Design (Applied Biosystems) to create primers and allele-specific probes for use in 5′-nuclease reaction (Holland *et al*. 1991) genotyping assays for 21 SNPs from 16 of the sequenced regions (Table 1). We also created primers and probes for two previously published mitochondrial DNA (mtDNA) control region SNPs (Smith *et al*. 2001).

Genotyping assays were performed in 384-well reaction plates in 5 μL volumes consisting of 1.0 μL template DNA

Correspondence: Christian T. Smith, Fax: +907 267 2442; E-mail: christian_smith@fishgame.state.ak.us

© 2006 Blackwell Publishing Ltd

Table 1 Amplification and sequencing of 25 loci in coho salmon. Names and sequences of each nested primer set are listed. Locus identifiers are the names of the best BLAST match for each sequence or, if no match was observed (indicated by *), are the names of the expected products. Numbers of transition (ti), transversion (tv) and insertion/deletion (indel) SNPs observed in each PCR product and numbers of genotyping assays developed (number developed/number attempted) are listed. Within the samples examined in this study, LD was observed between three pairs of markers indicated by superscript letters (see text for details)

Primer set/sequences (5′–3′)	Locus identifier	Sequence length	SNPs (ti/tv/indel)	Genotyping assays
arf_1 CAAGAGAGATCTTCCCAGTGC ACTCTCCGCTGTCCTTCATC GCTTGACGAGAAGCAACTCA TTGGTGTTGTGGTGGTGAAT	arf*	394	1/0/0	1/1
AsnRS_2 CCAGAGAGGCTGATGACTGA GTTCAGCAGCCAGGTTAGGA CATCGGTGAGACCATCCTG GGGACAGGTCCCATACTTCC	—	Failed		
BAMBI_2 CCAAAGAGGTGTGGTTCAGG GGGTGTGGTGGGACATAAAG CTGGAGGCCTCATCTTGGT AAGTTATACGAACTCCAGCTTCC	BAMBI	285	0/0/0	
copa_2 ATGTGGACTGGCTGGAGACT GCTGTGGGTCTTCGTCTTTC TCACACAGCCTCTTCTGCTG CAGCAGTGCTACCAGCTCAC	copa	561	0/0/0	
E2 CATTCAATAAATCTGGGTGTACCA ATAGCGGCAGTCTCCAAATG TGAAAAGAAAAACAGAACCAAACA CAGGTGCCATGTTGTTCACT	E2	245	1/0/0	1/1
eif4ebp2_2 CAACCACATAAAACCCCACA CCAAGTTGGCCGTACAGTTA ACAGCCACCGGTGAGGAT CATACATCACACCCAGTGACG	eif4ebp2	201	1/0/0	1/1
GnRH_2 TTCAAGAGGCACTTTTCAAGA TATGGGGCATCCATTTCTTC GGGTGTTGGCATAAAGGGTA GGAGATCAAAGCGTTAACTTCA	GnRH	419	0/1/0	1/1
GPDH_3 GGAAAAGCTGTGCAGGAGTT CCACAAATTGTAAAATTGTTAAATGAA GTCTGCAGAACCACCCTGA TTGTCAAGTGTAGTTGGTTTAGTTGA	GPDH*	383	1/2/0	2/2
HGFA_4 ATGGAAGCGCCATAGACAAC GGAATGACAGCATGTCAGCTA GTTACCTGCAGGGAGGAGTG GCGGTCCAATTGAGAGACAC	HGFA	488	1/0/0	1[a]/1
IGF-I.1_1 TACTGAGGAGTGGCCTCCAT GTCCCAGGCTGTTTCACAAT AAGGCCAGATTGTTGGAGTG CCGTGATTGAGAGTCCCATAG	IGF-I.1	346	1/1/0	1/1
Ikaros CCCCAACCTCCTCTGACAT CACAGAGGTTGCACTCGAAG	Ikaros	391	0/0/0	

Table 1 *Continued*

Primer set/sequences (5′–3′)	Locus identifier	Sequence length	SNPs (ti/tv/indel)	Genotyping assays
CTCCATGTACCCCCTCCAC TGGTGTACATCACATGGTCTAGG				
il-1racp_2 CTGGTGCTGTTCGTGGTCTA TGTCGAAGATGCAGACGGTA GCTCTATCGCTCCTGGTTTG TAGCCCAGCTCATTCTCCAG	il-1racp	329	0/2/0	1/2
LWSop_1 GGCGACAAAACCAAGATACAA CCCAGACCCAGGAGAAGATA CCTTTACCTTCACCAACAGCA TGACACTTCCAAAGGGCTTG	LWSop	782	0/2/0	1[c]/2
preproinsulin_2 CATTGTGTTTCCCTCATCCA TGACAGCACTGCTCTACAATACC TACCATCACCATGGCCTTCT GCAACTTCGTTCTCCTTTGC	ins	493	3/0/0	2[b,b]/2
RFC2_1 CGGTACAGGAAAGACCACCA ATGATCTGCTCGTCCCTCAG AGCCATGAAAGATGCTGTCC ACTGCACAACGTGACTGGAT	—	Failed		
SClkF2R2_1 GAGTTCTGCTGTCACATGCTC CAGACTCGGTCCTCAAAAGC CCCAAAGAGCCTCCTGTGTA CCAAGCCATTTGGAGTTGAG	SClkF2R2	230	0/1/0	1/1
serpin_2 AAAGGCCTGGATTATACCACAA GTGCCTCTCTCATCCACCTC CAGCTAAAAGATTCCTCAACTGG ATCACTGCCTTGTGCAACAC	serpin	403	0/2/0	2[a]/2
SWSop_1 TGGGAAAGGACTTCCATCTG ACCCAGGTGAAGCCAACA TAAGGTCAGCCCATTTGAGG CAAATTTGAAGGTGCCGAAG	SWS1op	519	0/1/0	1[c]/1
u202_2 CCCCCATATAGTGCATTGCT TCATTCCAATCCCCATTAAA GGGCCCTTAGGTATTGGTGT GAGCGGGTGGAAAGACTGT	u202	310	7/3/1	0/1
u211_2 TCAGTTGAACATAAATGAAGATTTG CCAACTGGGGTTTGAGATGT AAGGGACTCCGAACTTCCAG AGTAATGGGTGCCGTTGTTC	u211	255	1/0/0	
u212_2 GCTGTTTGCTGACGTGAATG AGAGAGCAGCTGGTGGTTGT AATGAGCTCCTTGTCAAATGC AAAGGCTAGAGGCACAGCTA	u212	191	0/0/0	
u4 CAGCATTGCTTTCCAAGGAT TCCATGCTGCATATTCATGTT TCAAACATATATGCACTGGGTTC TTGGGTATTTCAAGTTGATTTTCA	u4	371	3/1/0	0/1

Table 1 *Continued*

Primer set/sequences (5′–3′)	Locus identifier	Sequence length	SNPs (ti/tv/indel)	Genotyping assays
u6 CAATCTAGTTTAACATTGAAATGACG TGTCATGAAACATTATGGAAATGA GAGGAAATGGCTCATGTGAAA GCAAACCAGACACTATGAAAAGC	u6	438	0/1/0	1/1
ZNF330_1 AGGTGTTCAAACAGGACAAGG CGATGACACGTTCTTCCAGT CAAGTGTGGCCATGAGACC CAGAGGCTCCATAACCTCCA	ZNF330*	403	1/0/0	
Zp3b_2 TTATGGAGACGGAGCTGGTT CTGCCTCCAGGTGTATGGTT TCAGATTTGAATCGGCTGTG CCCAGGAAGTAGGCATTTGA	Zp3b	529	0/0/0	
Total		8966	21/17/1	17/21

(*c*. 20 ng) in 1× *Taq*Man PCR cocktail (proprietary PCR buffer, thermal-stable DNA polymerase exhibiting 5′-nuclease activity, and a passive reference dye) (Applied Biosystems), 900 nM of each PCR primer and 200 nM of each probe (Table 2). Thermal cycling was conducted at ramp rates at 1 °C per second on an ABI 9700 (Applied Biosystems) as follows: an initial denaturation of 10 min at 95 °C was followed by 50 cycles of: 92 °C for 15 s and annealing/extension temperature (62 °C for Ots_GPDH-146, Ots_il-1racp-176, Ots_LWSop-554, Ots_serpin-32,

Table 2 Nineteen SNP markers in coho salmon. Marker names consist of the species identifier (*Oki* for *Oncorhynchus kisutch*), the locus identifier ('Cr' indicates mtDNA control region, all others are from Table 1) and a number that indicates the nucleotide position in the GenBank reference sequence that the genotyping assay targets. Number of individuals successfully genotyped (*N*), expected (H_E; assuming panmixia) and observed (H_O) heterozygosities, maximum difference in minor allele frequencies between collections (Δq) and F_{ST} estimates are listed for each marker

Marker name/ reference sequence	Oligonucleotide sequences (5′–3′)*	*N*	H_E	H_O	Δq	F_{ST}
Oki_Cr-209 AF318028	F: CAAGGTTTACATTAAGCAAACACGAGAT R: GCCCATCTTAGTTGGAGTTGTGTTA VIC-ACCATTCGGTTGGTTAT FAM-AACCATTCGTTTGGTTAT	347	0.00	0.00	0.05	0.03
Oki_Cr-296 AF318028	F: CTAAGATGGGCTCCGTCTTTACC R: AAATCGTTGGTCGGTTCTTACTACATT VIC-CAGCATCGGTCCTAC FAM-CAGCATCAGTCCTAC	347	0.00	0.00	0.13	0.13
Oki_arf-115 DQ274166	F: TCCTTGAATGATCCATTTGTCATATTGGT R: CCATTTCATTACCCCTGGGAAGATT VIC-ATCACATACGAATGAAG FAM-ATCACATACAAATGAAG	344	0.20	0.18	0.30	0.09
Oki_E2-87 DQ274158	F: GAACCCATTGTCCCTAGGAAAGG R: CCACTGCCCCCATATTGCT VIC-ACAATCTAGCGTTAACTC FAM-ACAATCTAGCATTAACTC	342	0.20	0.20	0.25	0.04
Oki_eif4ebp2-58 DQ274171	F: GCTCAGTTTGAGATGGACATCTGA R: GGAACCAGGTTTGTTTCTTCTTCTG VIC-ACCACACCGGAACAA FAM-CCACACCAGAACAA	347	0.10	0.07	0.31	0.24
Oki_GnRH-151 DQ274159	F: AGATTTCCATTTGTCAGCCATTACTCT R: ATCTGAAGATGCTAAATTTTCTCCTGGTAA VIC-CCAAAGTATCTGAAATAA FAM-CAAAGTATGTGAAATAA	346	0.02	0.02	0.07	0.06

Table 2 *Continued*

Marker name/ reference sequence	Oligonucleotide sequences (5′–3′)*	N	H_E	H_O	Δq	F_{ST}
Oki_GPDH-146 DQ274168	F: CCTATGGGTAGGTTGGGATTGGA R: CTGCTGCCCTCTGATGGTATG VIC-CCTAGACCTGTACCTATAG FAM-CTAGACCTGTACTTATAG	344	0.10	0.11	0.06	0.00
Oki_GPDH-188 DQ274168	F: GGGAAGCTGATCCTAGACCTGTA R: GCACTGCTGCCCTCTGAT VIC-AGTAGCAGCATCATAC FAM-CAGTAGCAGCCTCATAC	344	0.14	0.15	0.23	0.07
Oki_HGFA-311 DQ274167	F: AGAGTCACCCTGTCAGTCACA R: GACGGAGTACGGTACGATATCAC VIC-ACATTTTCAAACCATTCAAT FAM-CATTTTCAAACCGTTCAAT	342	0.50	0.43	0.48	0.08
Oki_IGF-I.1-163 DQ274156	F: GCCCTTCTCCCCCTATTGC R: TGGAAGAAGTTTGGAACCAGCAA VIC-CAGCTCTAACAAGAGTC FAM-AGCTCTAACAGGAGTC	339	0.50	0.43	0.22	0.01
Oki_il-1racp-176 DQ274175	F: CTCCAACATATTAAACCTGACCTGGAA R: AAGCAGAAACCCAGTAAGAAGGAAA VIC-ATTATTAGGGTCAGGAGTC FAM-TTATTAGGGACAGGAGTC	340	0.28	0.25	0.51	0.21
Oki_ins-167 DQ274155	F: TTTTACAACCCAAAGAGAGATGTGGAT R: TCGAGCAGCTATGAAGTCTTTTGAA VIC-CCCTTACAGGCAAGAC FAM-CCCCTTATAGGCAAGAC	342	0.36	0.31	0.33	0.05
Oki_ins-323 DQ274155	F: TGTGTAGAAATGGCATGTAGGAAACT CCTGTATCTCATAATTGCCTTCACATTC VIC-TGTTGTGGAACAATGTA FAM-TTGTGGAGCAATGTA	346	0.12	0.10	0.15	0.07
Oki_LWSop-554 DQ274160	F: TTGGCTACTTCATTCTGGGACATC R: CGCAAGTGGAGACAGTGTATCC VIC-CAATGTGTGTATTTGAG FAM-CAATGTGTGTCTTTGAG	343	0.29	0.22	0.43	0.18
Oki_SClkF2R2-120 DQ274178	F: TGCATTTATGGGCACAAACCAAA R: CGTTTTCATTATATTGATTTAGGAAAGATTTGACTTTGA VIC-CAGCTATTTGTGTACTACAT FAM-CAGCTATTTGTGTTCTACAT	344	0.32	0.34	0.24	0.04
Oki_serpin-130 DQ274165	F: GTGAAGGCAAGTTTGGCATGTTA R: GCAAAGACCCAGGCCTGAA VIC-CTAAGAACTGACCGCTGACA FAM-ATCTAAGAACTGACCTCTGACA	336	0.47	0.37	0.48	0.12
Oki_serpin-328 DQ274165	F: AAGGAACTCTCATACATAGGAGAACGT R: GCGTCAAGTGTATTTTGTGGTGAAA VIC-CATAGGTTTTCTAAGTGTTTCA FAM-TAGGTTTTCTAAGTCTTTCA	345	0.17	0.15	0.18	0.06
Oki_SWS1op-38 DQ274161	F: AGGGTTTGCAGATGACCACATATC R: GGGTTGGTGTCAGCTTGGT VIC-CTGTCCTTGCCTTTGAGA FAM-TGTCCTTGCATTTGAGA	346	0.44	0.34	0.38	0.11
Oki_u6-257 DQ274157	F: CTTAAGATTGTTATATGAAAAAGTAAAGTAAAAGTAAAGTATTATACCA R: CGATGCACTTTACATCCAACATATTCA VIC-TTTAGAATGTTATAATGATAACAAG FAM-AATGTTATAATGAGAACAAG	318	0.48	0.42	0.23	0.03

*Each allele-specific probe was labelled with either VIC or 6FAM on its 5′-end and a minor groove binder and a nonfluorescent quencher on its 3′ end.

Ots_SWS1op-38; 60 °C for all others) for 1 min. Following amplification, all plates were read and scored on an ABI 7900HT real-time PCR instrument using SEQUENCE DETECTION SOFTWARE 2.1 (Applied Biosystems).

Both assays targeting mtDNA SNPs were successful, as were 17 of 21 assays based on the sequence data described here. The remaining four assays yielded monomorphic products. The 17 nuclear SNPs reside across 14 loci (Table 1) among which the degree of physical linkage is unknown.

GENEPOP version 3.2a (Raymond & Rousset 1997) was used to calculate F_{ST} and to test for departures from Hardy–Weinberg equilibrium (HWE). LINKDOS (Black & Krafsur 1985) was used to test for linkage disequilibrium (LD) and to partition variance into within- and among- collection components (Ohta 1982). No significant ($\alpha = 0.05$) departures from HWE were detected. Several locus pairs exhibited significant ($\alpha = 0.05/6$ collections $= 0.008$) LD within one collection, and three locus pairs exhibited LD in multiple collections (Oki_HGFA-311 + Oki_serpin-130; Oki_ins-167 + Oki_ins-323; Oki_LWSop-554 + Oki_SWS1op-38). For every locus combination tested $D'_{IS} > D'_{ST}$, suggesting that drift is responsible for the variance in allele combinations among collections. While LD was not observed among pairs of SNPs within some individual sequence reads (GPDH, serpin, mtDNA), these pairs are clearly physically linked and population analyses using these markers will need to include some form of haplotype estimation (e.g. Excoffier & Slatkin 1995).

Acknowledgements

This work was funded by the state of Alaska. DNA sequencing was performed by Polymorphic DNA Technologies Inc. The authors are grateful to A. Surridge for providing comments and suggestions that improved this manuscript.

References

Beacham TD, Candy JR, Supernault KJ *et al*. (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.

Black IVWC, Krafsur (1985) A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. *Theoretical and Applied Genetics*, **70**, 491–496.

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Molecular Biology and Evolution*, **12**, 921–927.

Holland PM, Abramson RD, Watson R, Gelfand DH (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 Sciences, USA*, **88**, 7276–7280.

Nickerson DA, Tobe VO, Taylor SL (1997) POLYPHRED: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research*, **25**, 2745–2751.

Ohta T (1982) Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **79**, 1940–1944.

Raymond M, Rousset F (1997) GENEPOP: population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Smith CT, Elfstrom CM, Seeb JE, Seeb LW (2005) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology*, **14**, 4193–4203.

Smith CT, Koop BF, Nelson RJ (1998) Isolation and characterization of coho salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids. *Molecular Ecology*, **7**, 1614–1615.

Smith CT, Nelson RJ, Wood CC, Koop BF (2001) Glacial biogeography of North American coho salmon (*Oncorhynchus kisutch*). *Molecular Ecology*, **10**, 2775–2785.

Utter FM, Ames WE, Hodgins HO (1970) Transferring polymorphism in coho salmon (*Oncorhynchus kisutch*). *Journal of Fisheries Research Board of Canada*, **27**, 2371–2373.