Molecular Ecology Notes (2006) **6**, 1255–1259 doi: 10.1111/j.1471-8286.2006.01507.x

PRIMER NOTE

Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon

CARITA M. ELFSTROM, CHRISTIAN T. SMITH and JAMES E. SEEB
Gene Conservation Laboratory, ADF & G, 333 Raspberry Road, Anchorage, Alaska 99518, USA

Abstract

We characterize 32 single nucleotide polymorphism genotyping assays for resolving genotypic variation in sockeye salmon *Oncorhynchus nerka* in the Pacific Rim. These assays are based on the 5′-nuclease reaction and thus facilitate high-throughput genotyping with minimal optimization time. Minor allele frequency differences (Δq) among collections were between 4.7% and 97.9%, resulting in per locus F_{ST} estimates of 0.02–0.71 with an average of 0.22.

Keywords: SNP, 5′-nuclease assay, *Oncorhynchus nerka*, PCR, sockeye salmon, *Taq*Man

Received 14 May 2006; revision accepted 20 June 2006

Studies of the ecological genetics of fish sometimes require the analysis of thousands of individuals at many loci (e.g. Seeb *et al*. 1999). Admixture analyses of complex fisheries also may require analysis of thousands of individuals (Seeb & Crane 1999), sometimes with rapid turnaround (Seeb *et al*. 2000). Assays used to conduct such studies need to be rapid and inexpensive.

Single nucleotide polymorphism (SNP) markers have been identified using restriction fragment length polymorphism (RFLP) analysis, DNA sequencing and various conformation polymorphism analyses in several species of fishes (e.g. Bickham *et al*. 1995). While these techniques have been used to develop and assay several informative SNPs, broad-scale application of these markers was limited by time and cost. Here we describe 32 comparatively inexpensive and high-throughput SNP genotyping assays that resolve genetic variation in sockeye salmon (*Oncorhynchus nerka*) inhabiting the Pacific Rim.

We identified SNPs in 29 nuclear and three mitochondrial DNA (mtDNA) sequences (Table 1). Potential loci were selected using previously published primer sets designed from alignments of two species DNA sequences (Lyons *et al*. 1997; Smith *et al*. 2005), primers designed for other salmon species (Xiong *et al*. 1992; Park *et al*. 1995, 1996; Ford *et al*. 1999; Gharrett *et al*. 2001; Moran 2002; Dann *et al*. 2004), primers designed for sockeye salmon (Bickham *et al*. 1995; Miller *et al*. 2001), and one novel primer pair based on sequence obtained using DNA Walking SpeedUp Premix Kit (SeeGene, Inc.) (One_U301, F-CCTTACCAACAGTATGTGCCA, R-AGTTGCCATACTACTACTGGCTA) (Table 1). DNeasy 96 Tissue Kits (QIAGEN) was used to extract DNA from 50 individuals, 10 individuals collected at five geographical locations (Kamchatka, Bristol Bay, Kodiak/Afognak Islands, Southcentral Alaska and Southeast Alaska). These individuals were sequenced in both directions and sequences were aligned and screened for SNPs using SEQUENCHER 4.5 (GeneCodes Corporation).

We used OLIGO 6 (Molecular Biology Insights, Inc.) and PRIMER EXPRESS (Applied Biosystems) or Assay-By-Design (Applied Biosystems) to create primers and allele-specific probes for use in 5′-nuclease reaction (Holland *et al*. 1991). SNP genotyping was performed on 1146 sockeye salmon (Table 2) in 384-well reaction plates. Each reaction was conducted in a 5-μL volume consisting of 0.10-μL template DNA in 1 × *Taq*Man Universal Buffer (Applied Biosystems), 900 nM of each polymerase chain reaction (PCR) primer, and 200 nM of each probe. Thermal cycling was performed on a Dual 384-Well GeneAmp PCR System 9700 as follows: an initial denaturation of 10 min at 95 °C followed by 50 cycles of: 92 °C for 15 s and annealing/extension temperature for 1 or 1.5 min (Table 1). Cycling was conducted at a ramp speed of 1 °C per second. The plates were read on an ABI PRISM 7900HT Sequence Detection System after amplification and scored using SEQUENCE DETECTION software 2.2 (Applied Biosystems).

GENEPOP version 3.2a (Raymond & Rousset 1997) was used to calculate F_{ST} (Weir & Cockerham 1984), expected and observed heterozygosities, and to test for departures from Hardy–Weinberg equilibrium (HWE). LINKDIS (Black

Correspondence: James E. Seeb, Fax: + 907 2672442; E-mail: jim_seeb@fishgame.state.ak.us

Table 1 Thirty-two SNP markers in sockeye salmon. Marker names consist of the species identifier (*One* for *O. nerka*) and the locus identifier, and a number indicating the nucleotide position in the GenBank reference sequence that the genotyping assay targets. Source of sequencing primers, annealing temperature (T_a), number of individual 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

Locus name/marker name/GenBank Accession/source of sequencing primers	Oligonucleotide sequences (5′–3′)*	T_a (°C)	N	H_E	H_O	Δq	F_{ST}
Acyl-coenzyme A-binding protein	F: GAGGTGTGGGCTGACCA	60	1117	0.38	0.42	0.53	0.11
One_ACBP-79	R: TCGACCGCTGGCAGTG						
DQ386287	VIC-CAGAGGTCATGGTTCTA						
(Smith *et al*. 2005)	FAM-CAGAGGTCATAGTTCTA						
Aldolase B	F: TTCTCAGTTGTCAATTTCTCTCTTTCACA	60	1097	0.28	0.27	0.19	0.02
One_ALDOB-125	R: TTCAGCCATGTCAATTGGAATGTG						
DQ386280	VIC-CCGGACTTGTTTTAACA						
Smith (2005)	FAM-CCGGACTTGTTGTAACA						
Aldolase B	F: CCCGTGCCGGACTTGTT	60	1108	0.27	0.28	0.22	0.03
One_ALDOB-135	R: TCAGCCATGTCAATTGGAATGTGA						
DQ386280	VIC-ACAGCACGAAATTA						
Smith (2005)	FAM-ACAGCACAAAATTA						
Chaperonin containing TCP1, subunit 3	F: AGTCTCAAGCATTTGCTGCAGATA	60	1115	0.02	0.02	0.10	0.08
One_CCT3-282	R: GGCTAATAGGCTGCTGTTGAAAGT						
DQ386289	VIC-ATTCAAGATGAATCTC						
Smith (2005)	FAM-CATTCAAGATTAATCTC						
Cytochrome oxidase *c*	F: CATAGTAATGCCTGCTGCTAGGA	60	1126	N/A	N/A	0.91	0.30
One_CO1	R: CCACTTTTTGTTTGAGCTGTGCTAA						
AY353068-AY353070	VIC-ACTTCTACTACTTTCCC						
(Gharrett *et al*. 2001)	FAM-ACTTCTACTACTCTCCC						
Connective tissue growth factor	F: AAGGACAGAAACATATATGCGTATATTCAATGT	60	1117	0.04	0.04	0.10	0.05
One_ctgf-301	R: CTGTCTTTCGTCCCCTCTTTAGG						
DQ386288	VIC-TGATGGATGTGTAGGGC						
Smith (2005)	FAM-TGATGGATGTTTAGGGC						
Cytochrome *b*	F: CCTGGGAGATCCAGACAATTTTA	60	1126	N/A	N/A	0.98	0.70
One_Cytb_17	R: CGTAAGCGAAAAGGAAGTATCACTCT						
AY353063-AY353065	VIC-CAACCCGCTAGTTAC						
Bickham *et al*. 1995)	FAM-AACCCGCTGGTTAC						
Cytochrome *b*	F: CCTGGGAGATCCAGACAATTTTA	60	1125	N/A	N/A	0.85	0.37
One_Cytb_26	R: CGTAAGCGAAAAGGAAGTATCACTCT						
AY353063–AY353065	VIC-TTTGATATGAGGTGGAGTAA						
Bickham (1995)	FAM-TGATATGAGGTGGGGTAA						
Growth hormone 2	F: GGCATCAACCTGCTCATCGA	60	1126	0.25	0.21	0.42	0.12
One_GHII-2165	R: TGCACAAAGTGCGGCAC						
U14535	VIC-CACAAATGGAAATTGA						
(Park *et al*. 1995)	FAM-CACAAATGGTAATTGA						
Glycoprotein hormone-alpha-subunit	F: CAAGAAGAATCAAGAGAAAGAGAGATGGT	60	1114	0.44	0.35	0.76	0.18
One_GPH-414	R: CCTAGTGTCATGCACATAACGTGTA						
DQ386289	VIC-AAGAACTAGAATGGAACAGA						
Smith (2005)	FAM-AAGAACTAGAATGAAACAGA						
Heat-shock protein	F: ACAGCGAAACTATTGATTTAAGGCTCAT	60	1102	0.39	0.29	0.74	0.25
One_hsc71–220	R: CGCAGGTAAATCACTGATCATGTTT						
DQ386293	VIC-ATTGGCCACAGCGC						
Moran (2002)	FAM-ATTGGCAACAGCGC						
*Hpa*I repeat sequence	F: TGTTGTTCCTAGGCTGTCATTGAAA	60	1111	0.48	0.43	0.53	0.09
One_*Hpa*I-71	R: CCCTGCGTATTACTAAGGCCATATTTATT						
DQ386294	VIC-TCAGTTAAGAACTAATTCT						
(Lyons *et al*. 1997)	FAM-AGTTAAGAACAAATTCT						
*Hpa*I repeat sequence	F: CCTGAGTTGTGTTCAATGGGCATAA	61	1108	0.31	0.19	0.62	0.35
One_*Hpa*I-99	R: TGGGTCATGTTCATTAGAGCACAAA						
DQ386281	VIC-AACGGAAGAAACCCCTCAA						
Smith (2005)	FAM-AACGGAAGAAACTCCTCAA						

Table 1 *Continued*

Locus name/marker name/GenBank Accession/source of sequencing primers	Oligonucleotide sequences (5′–3′)*	T_a (°C)	N	H_E	H_O	Δq	F_{ST}
Karyopherinalpha 2 (RAG cohort 1, importin-alpha 1)	F: AGTGTCCTCCCATACAGTTCGA	61	1110	0.38	0.26	0.92	0.29
One_KPNA-275	R: GGTGGTTTGGTCAGAGTTTCCA						
DQ386282	VIC-CCCTGACCAACATC						
Smith (2005)	FAM-CCCTAACCAACATC						
Karyopherin-alpha 2 (RAG cohort 1, importin-alpha 1)	F: TGGGCCCTGGGAAACATC	60	1111	0.38	0.26	0.92	0.28
One_KPNA-422	R: CCATAGCCACTTTCGATACAGGTAA						
DQ386282	VIC-CTGGTATGAGAAGGCACA						
Smith (2005)	FAM-TGGTATGAGGAGGCACA						
Leukocyte elastase inhibitor	F: ACAGCGCATCCCCATAATGG	61	1110	0.50	0.41	0.64	0.15
One_LEI-87	R: GCCTTTGTGGAGGTCAACGA						
DQ386279	VIC-ACTCGCCACCTCTGT						
Smith (2005)	FAM-TCGCCGCCTCTGT						
Major histocompatibilty complex, class II, beta	F: CAGCACTGTGGGGAGGTA	54[a]	1126	0.27	0.23	0.45	0.22
One_MHC2–091	R: TGTTAAGGAGCCCTGCTCA						
AY386254-AY386258	VIC-CACTGAGTATGGTGTGAAG						
(Miller *et al*. 2001)	FAM-CACTGAGCATGGTGTG						
Major histocompatibilty complex, class II, beta	F: CAGCACTGTGGGGAGGTA	54[a]	1126	0.32	0.20	0.72	0.35
One_MHC2–109	R: TGTTAAGGAGCCCTGCTCA						
AY386254–AY386258	VIC-ATGCAAAAGCATGGAAC						
Miller (2001)	FAM-ATGCAGAAGCATGGAA						
Major histocompatibilty complex, class II, beta	F: GCATGGTGTGAAGAATGCA	54[a]	1127	0.48	0.32	0.80	0.32
One_MHC2–190	R: TGTTAAGGAGCCCTGCTCA						
AY386254–AY386258	VIC-CTGCTATCGACTACAG						
Miller (2001)	FAM-CGCTGCTATCTACTACA						
Major histocompatibilty complex, class II, beta	F: GAGCAGGGCTCCTTAACA	54[a]	1106	0.50	0.29	0.88	0.38
One_MHC2–251	R: GGTCTTGACTTGMTCAGTCA						
U34707, U34711	VIC-CAGGCCCCTGACTG						
Miller (2001)	FAM-ACAGGCCTCTGACTGA						
SNPstr	F: CCATAGTGTATCACACAATACTCATGTCT	60	1124	0.31	0.23	0.63	0.25
One_Ots213–181	R: TCTATCATCTGCAAATCTGTGTACTAGACT						
DQ386285	VIC-CTTTGAATTAAAAACATTTTT						
Smith (2005)	FAM-CTTTGAATTAAAAACTTTTTT						
p53 tumour suppression gene	F: GACAATCTTAAAGCGGTGGTCTTG	60	1125	0.16	0.11	0.51	0.31
One_p53–534	R: AACCTTTATCAGCCATCATCCAACT						
DQ386284	VIC-ATGTCCAAAGATCTGG						
(Park *et al*. 1996)	FAM-AATGTCCAAATATCTGG						
Prolactin II	F: ACCTCTCTCTCTCTCAGGACTCTCA	62	1125	0.50	0.43	0.57	0.08
One_Prl2	R: GAGGAGGTGTGACACATAGATGGA						
AY353071–AY353072	VIC-ACCAATGGGACGAGTG						
(Xiong *et al*. 1992)	FAM-CCACCAATTGGACGAG						
Recombination-activating gene	F: AGCTCACACATACAACAAATATGATCTAATGT	60	1117	0.11	0.09	0.36	0.20
One_RAG1-103	R: GTGAACTGCATCTTTGAACAAATGC						
DQ386290	VIC-CGAATCTCAACAATAAGT						
Moran (2002)	FAM-CTCGAATCTCAACTATAAGT						
Recombination-activating gene, 3′UTR	F: AGATAAAGATGGTTTCAAAGTCACCCA	60	1113	0.14	0.12	0.24	0.09
One_RAG3-93	R: GGGCTGCCATCTAAAAAATATTGCT						
DQ386291	VIC-CATTTTGGACTTCGGGACC						
Moran (2002)	FAM-CATTTTGGACTTTGGGACC						
RH2 Opsin	F: GCTGCTAGGTCAAACTCGAAGAG	60	1117	0.02	0.02	0.05	0.02
One_RH2op-395	R: CAGCCTTGTTCAACCCCATTATCTA						
DQ386277	VIC-TGGGAACATCATTTTTTAA						
(Dann *et al*. 2004)	FAM-TTGGGAACATAATTTTTTAA						
Stanniocalcin	F: CAACACAACATCAACATCATTAATAAACATTCTG	61	1104	0.48	0.31	0.91	0.35
One_STC-410	R: AACATCCCCGTTTTGACCACTTAT						

Table 1 *Continued*

Locus name/marker name/GenBank Accession/source of sequencing primers	Oligonucleotide sequences (5′–3′)*	T_a (°C)	N	H_E	H_O	Δq	F_{ST}
DQ386278	VIC-CCGATGGGTATATTATTATA						
Moran (2002)	FAM-CCGATGGGTATATTGTTATA						
SNPstr	F: CACACCTGAGGCACAAGCT	60	1127	0.48	0.33	0.85	0.32
One_STR07	R: GTATGTCTACCAGAGAGGTCAAGGA						
DQ386286	VIC-ACGCACACTGTCCTT						
Smith (2005)	FAM-ACGCACACTCTCCTT						
Transferrin	F: AGCAGGTGTAAGCATGTGTACTT	60	1123	0.48	0.33	0.95	0.34
One_Tf_ex11–750	R: CCTGCTCTGCCTCAACAATGTTAA						
DQ267489	VIC-CAGGGTCGCTGCAC						
(Ford *et al*. 1999)	FAM-CCAGGGTCACTGCAC						
Transferrin	F: GCCCTTAGCACTTCAGTTGCA	60	1122	0.16	0.13	0.34	0.19
One_Tf_in3–182	R: CAGACAGAAACCATTTGATCCGATTC						
DQ267488	VIC-AACAGAAAGTCTACACTTT						
Ford (1999)	FAM-ACAGAAAGTCTGCACTTT						
FK506-like gene	F: AGCCAGTAGCCGATAATGTTTGTC	60	1121	0.20	0.19	0.31	0.07
One_U301-92	R: CCCCTCCCAAATTGCTAGCT						
DQ267490	VIC-CCATGGATTAAAATATTT						
Original sequence	FAM-CCATGGATTAAACTATTT						
Vimentin	F: TTCTGGGTGGACTCATTGATCAC	60	1116	0.23	0.21	0.29	0.08
One_VIM-569	R: ATGCGTTATACCTGTAATCTGCAAGT						
DQ386292	VIC-AAGTGTTTCCATACTCACTATA						
Moran (2002)	FAM-AAGTGTTTCCATATTCACTATA						

*, Each allele-specific probe was labelled with either VIC or 6-FAM on its 5′ end and a minor groove binder and a nonfluorescent quencher on its 3′ end; †, Annealing time 1.5 min.

Table 2 Population of sockeye salmon, geographical location, and year of collection

Population	Location	Latitude	Longitude	Year
Vichenkiya River	Kamchatka Peninsula	51.45	157.10	2000
Kamchatka River	Kamchatka Peninsula	56.23	162.50	1998
Salmon Lake	Norton Sound, Alaska	64.90	–165.08	2001
Glacial Lake	Norton Sound, Alaska	64.84	–165.70	2004
Gibralter River	Bristol Bay, Alaska	59.44	–154.85	2000
Ugashik Lake	Bristol Bay, Alaska	57.57	–157.00	2000
Skilak Lake	Cook Inlet, Alaska	60.47	–150.51	1992
Russian River	Cook Inlet, Alaska	60.49	–150.00	1992
Red Bay Lake	Southeast Alaska	56.26	–133.32	1993
McDonald Lake	Southeast Alaska	55.57	–132.71	2001
Baker Lake	Puget Sound, Washington	48.46	–121.47	1996
Cedar River	Puget Sound, Washington	47.30	–122.13	1994

& Krafsur 1985) was used to test for linkage disequilibrium (LD) and to partition variance into within- and among-collection components (Ohta 1982). Significance levels for all tests were corrected for multiple simultaneous tests (Rice 1989). Several populations showed significant departure ($\alpha = 0.05$) from HWE for one locus. One population (Kamchatka River, Russia) showed significant departure for two loci (*One_GHII-2165* and *One_MHC2_190*). Our sample from this population is poorly documented and may in fact contain individuals from multiple spawning sites. Significant LD was observed for SNPs within loci (i.e. within aldolase B, karyopherin-alpha 2 and major histocompatibility complex), but LD was not detected among SNPs from different sequences.

Data collected using the present assays directly reflect underlying DNA sequences and are thus readily combined with data collected across hardware and chemistry platforms and across laboratories. This transportability of data

as well as the rapid rate at which SNP data are generated renders these markers well suited to mixture and migratory studies of sockeye salmon.

Acknowledgements

This work was funded equally by three sources: State of Alaska, North Pacific Research Board, and Pacific Salmon Commission Northern Fund. DNA sequencing was performed by Polymorphic DNA Technologies, ACGT, Inc. and MWG, Inc.

References

Bickham JW, Wood CC, Patton JC (1995) Biogeographic implications of cytochrome *b* sequences and allozymes in sockeye (*Oncorhynchus nerka*). *Journal of Heredity*, **86**, 140–144.

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

Dann SG, Allison WT, Levin DB, Taylor JS, Hawryshyn CW (2004) Salmonid opsin sequences undergo positive selection and indicate an alternate evolutionary relationship in *Oncorhynchus*. *Journal of Molecular Evolution*, **58**, 400–412.

Ford MJ, Thornton PJ, Park LK (1999) Natural selection promotes divergence of transferrin among salmonid species. *Molecular Ecology*, **8**, 1055–1061.

Gharrett AJ, Gray AK, Brykov V (2001) Phylogeographic analysis of mitochondrial DNA variation in Alaskan coho salmon, *Oncorhynchus kisutch*. *Fishery Bulletin*, **99**, 528–544.

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.

Lyons LA, Laughlin TF, Copeland NG *et al.* (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, **1**, 47–56.

Miller KM, Kaukinen KH, Beacham TD, Withler RE (2001) Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica*, **111**, 237–257.

Moran P (2002) Current conservation genetics: building an ecological approach to the synthesis of molecular and quantitative genetic methods. *Ecology of Freshwater Fish*, **11**, 30–55l.

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.

Park LK, Moran P, Dightman DA (1995) A polymorphism in intron D of the chinook salmon growth hormone 2 gene. *Animal Genetics*, **4**, 285.

Park LK, Moran P, Dightman DA (1996) A chinook salmon PCR-RFLP marker in the p53 locus. *Animal Genetics*, **2**, 127–128.

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

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Seeb LW, Crane PA (1999) Allozymes and mitochondrial DNA discriminate Asian and North American populations of chum salmon in mixed-stock fisheries along the south coast of the Alaska Peninsula. *Transactions of the American Fisheries Society*, **128**, 88–103.

Seeb JE, Habicht C, Templin WD *et al.* (1999) Allozyme and mtDNA variation describe ecologically important genetic structure of even-year pink salmon inhabiting Prince William Sound, Alaska. *Ecology of Freshwater Fish*, **8**, 122–140.

Seeb LW, Habicht C, Templin WD *et al.* (2000) Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the *Exxon Valdez* oil spill. *Transactions of the American Fisheries Society*, **129**, 1223–1249.

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

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

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