Molecular Ecology (2005) **14**, 4193–4203 doi: 10.1111/j.1365-294X.2005.02731.x

Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon

CHRISTIAN T. SMITH, CARITA M. ELFSTROM, LISA W. SEEB and JAMES E. SEEB
Gene Conservation Laboratory, Division of Commercial Fisheries, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK 99518, USA

Abstract

Single nucleotide polymorphisms (SNPs) are a class of genetic markers that are well suited to a broad range of research and management applications. Although advances in genotyping chemistries and analysis methods continue to increase the potential advantages of using SNPs to address molecular ecological questions, the scarcity of available DNA sequence data for most species has limited marker development. As the number and diversity of species being targeted for large-scale sequencing has increased, so has the potential for using sequence from sister taxa for marker development in species of interest. We evaluated the use of *Oncorhynchus mykiss* and *Salmo salar* sequence data to identify SNPs in three other species (*Oncorhynchus tshawytscha*, *Oncorhynchus nerka* and *Oncorhynchus keta*). Primers designed based on *O. mykiss* and *S. salar* alignments were more successful than primers designed based on *Oncorhynchus*-only alignments for sequencing target species, presumably due to the much larger number of potential targets available from the former alignments and possibly greater sequence conservation in those targets. In sequencing ~89 kb we observed a frequency of 4.30×10^{-3} SNPs per base pair. Approximately half (53/101) of the subsequently designed validation assays resulted in high-throughput SNP genotyping markers. We speculate that this relatively low conversion rate may reflect the duplicated nature of the salmon genome. Our results suggest that a large number of SNPs could be developed for Pacific salmon using sequence data from other species. While the costs of DNA sequencing are still significant, these must be compared to the costs of using other marker classes for a given application.

Keywords: 5′-nuclease reaction, chinook salmon, chum salmon, *Oncorhynchus*, SNP, sockeye salmon

Received 7 May 2005; revision accepted 4 August 2005

Introduction

Pacific salmon of the genus *Oncorhynchus* spawn in lakes and rivers that drain into the North Pacific Basin from large geographic regions of both the Asian and North American continents. Most *Oncorhynchus* species have anadromous life history forms wherein individuals emerge in rivers and lakes and subsequently cover hundreds or thousands of kilometres in migrations through portions of the Pacific Ocean and Bering Sea and back to freshwater to spawn in their natal river. The great geographic distances that they migrate, the diverse environmental conditions that they experience during their migrations, the well-documented cases of parallel life history evolution (e.g. Taylor *et al*. 1996), and the 'keystone' role that these species play in transporting marine-derived nutrients to terrestrial environments (e.g. Reimchen *et al*. 2003) are all factors that have made salmon of interest to ecologists and evolutionary biologists. The considerable cultural and economic importance of these species to human communities throughout the North Pacific have further motivated numerous studies of the biology of Pacific salmon (reviewed by Groot & Margolis 1991; Quinn 2005).

The family Salmonidae is thought to have descended from a tetraploid ancestor between 25 and 100 million years ago (Ma) (Allendorf & Thorgaard 1984). Approximately 50% of protein loci examined in salmonid species continue to exhibit duplicate gene expression, possibly due

Correspondence: Christian T. Smith, Fax: + 907 2672442, E-mail: christian_smith@fishgame.state.ak.us

to selective advantages for these species (Bailey *et al*. 1978; Allendorf & Thorgaard 1984). The genera *Salmo (*including Atlantic salmon) and *Oncorhynchus* (including Pacific salmon) are thought to have diverged from one another approximately 18–20 Ma (McKay *et al*. 1996; McPhail 1997). A broad comparison of expressed sequence tags (EST) between *Salmo salar* (Atlantic salmon) and *Oncorhynchus mykiss* (rainbow trout) revealed approximately 6% sequence divergence between the two (Rise *et al*. 2004).

Present management strategies for Pacific salmon species by nations throughout the North Pacific are based to varying degrees on genetic mixed-stock analysis (MSA) techniques (Utter & Ryman 1993) and may require collection of thousands of genotypes per year in each nation. Further, international treaties and conventions such as the Pacific Salmon Treaty between the United States and Canada (www.psc.org) and the North Pacific Anadromous Fish Commission, between those two nations and Russia, South Korea and Japan (www.npafc.org), require genotyping of thousands of additional individuals per year for management and research. The marker classes most commonly used to meet these requirements have been allozymes and microsatellites. Allozymes have provided valuable insights over several decades (e.g. Seeb *et al*. 2004), but their use has subsided due to the relatively flexible sample formats, high throughput rates and improved resolution available via DNA markers. Microsatellite loci have over a decade of use in salmonids (Bentzen *et al*. 1991), are readily available for all Pacific salmon species and have proven very powerful in population structure and mixture studies (Nelson *et al*. 1998, 2003b). Further, laboratory protocols for high-throughput microsatellite analyses were available long before comparably efficient analysis protocols were available for other classes of DNA markers (e.g. Olsen *et al*. 1996). A limitation of microsatellite data with regards to multijurisdictional research and international treaties has been the fact that microsatellite data generated in one laboratory are not readily combined with or compared to data collected in another. This is because, based on the most common methods of microsatellite analyses, raw data are somewhat dependant on specific aspects of hardware and chemistry installed in a given laboratory (Wattier *et al*. 1998; Haberl & Tautz 1999) and on the environment in the laboratory in which data are collected (Davison & Chiba 2003). While standardization of defined microsatellite alleles and loci among defined laboratories is clearly technically possible, the costs of such efforts increase with the number of alleles per locus and the number laboratories involved. The result for Pacific salmon has been the development of laboratory-specific genetic baselines that are often redundant in that each laboratory must genotype samples from every population. Among state and federal laboratories in the United States and Canada, there are at least six overlapping baselines each for *Oncorhynchus keta* (chum salmon) and *Oncorhynchus nerka* (sockeye salmon), several of which contain overlapping data that researchers cannot presently combine or compare.

The term 'single nucleotide polymorphism' (SNP) has been used in recent years to describe small DNA sequence variants. While the term SNP suggests that only single-base substitutions should be included, various definitions have included indels, small (2–5) base pair substitutions and tandem repeat variation (e.g. dbSNP; www.ncbi.nlm.nih.gov/projects/SNP/). The ubiquity of SNPs throughout genomes examined to date has rendered them appealing as markers for a broad range of applications including fine-scale mapping and association and population genetic studies. Because SNPs are generally bi-allelic, laboratory analyses and scoring of raw data are highly amenable to automation. Population studies of both model and nonmodel organisms have utilized SNP data for over two decades (several examples in Avise 2004). Due to the expense and difficulty of sequencing loci in many individuals, most early studies were done using restriction fragment length polymorphism (RFLP) analyses. The minimal amount of laboratory equipment required to conduct RFLP analyses means that genotyping can be performed nearly anywhere, and this technique continues to serve several SNP genotyping projects today (e.g. Nelson *et al*. 2003a; Ryynänen & Primmer 2004; Gharrett *et al*. 2005). The development of several high-throughput genotyping technologies in recent years (summary in Kwok 2003) has broadened the range of research questions to which SNPs may be applied. These technologies have recently been used in a range of nonmodel organisms including cows (Werner *et al*. 2004), cyanobacteria (Batley & Hayes 2003), trees (Osman *et al*. 2003), wolves (Seddon *et al*. 2005), scallops (Elfstrom *et al*. 2005), and salmon (Sato *et al*. 2004; Smith *et al*. 2005). While SNP genotyping assays can be used to generate data rapidly and inexpensively, the costs of SNP discovery and of genotyping assay development must be considered in evaluating the potential utility of SNPs for any given application (Schlötterer 2004).

In species for which little DNA sequence data are available, SNP discovery strategies have included sequencing random genomic library clones (Primmer *et al*. 2002; Werner *et al*. 2004), sequencing polymerase chain reaction (PCR) products generated via degenerate oligonucleotide-primed PCR (DOP-PCR, e.g. Osman *et al*. 2003) or amplified fragment length polymorphism (AFLP) bands (e.g. Meksem *et al*. 2001; Brugmans *et al*. 2003; Nicod & Largiader 2003), and developing PCR and sequencing primers based on DNA sequence from sister taxa (e.g. Lyons *et al*. 1997; Primmer *et al*. 2002; Aitken *et al*. 2004). The last approach is particularly appealing for SNP discovery in species for which little DNA sequence data are available but for which sequence data are available for closely related taxa. The number and diversity of species being targeted for large-scale sequencing

Fig. 1 Map showing collections sites for *Oncorhynchus tshawytscha* (t), *Oncorhynchus* (k), and *Oncorhynchus nerka* (n). Ten individuals were sampled from each site.

has increased rapidly in recent years (lists of genome and EST sequencing projects available from EMBL and NCBI), and fish in particular are increasingly represented (Cossins & Crawford 2005). Although sequence data are scarce for most species of Pacific salmon, a wealth of EST data have recently been provided by studies of *O. mykiss* (Rexroad *et al*. 2003) and *S. salar* (Rise *et al*. 2004). We describe the use of a comparative approach to SNP discovery in *Oncorhynchus tshawytscha* (chinook salmon), *O. nerka* and *O. keta* and the validation of 53 of these SNPs through the use of high-throughput genotyping assays.

Methods and materials

SNP discovery

Two approaches were used to identify sequences that were conserved among species. The first approach utilized sequences from the genus *Oncorhynchus* as follows: All available sequences for the three species of interest were obtained from GenBank (786 sequences in *Oncorhynchus nerka*, 216 sequences in *Oncorhynchus keta* and 1573 sequences in *Oncorhynchus tshawytscha*). These sequences were compared with published *Oncorhynchus mykiss* sequences using MEGABLAST (Altschul *et al*. 1990; Zhang *et al*. 2000). Sequences from the three species of interest and the matching rainbow trout sequences were assembled using SEQMAN II (DNASTAR). Contiguous sequence reads (contigs) four or more accessions deep were identified as candidates for subsequent PCR primer design.

The second approach involved aligning *O. mykiss* sequences with *Salmo salar* sequences. All sequences in GenBank for both species were compared using BLASTN (Altschul *et al*. 1990, 1997) with an e-value cut-off of $\leq 10^{-5}$. The results of this comparison were then filtered for unique accessions from each species (to reduce redundancy based on multiple hits on single accessions) and for matches greater than 500 nucleotides long (to facilitate amplification and sequencing of the corresponding loci). Finally, in order to reduce the number of matches that represented duplications of one another (i.e. regions that were represented by multiple reads in both species), we used SEQMAN II to assemble the sequences in our filtered alignment.

Target regions of ~500 bp conserved between multiple sequences from each of the above assemblies were chosen for amplification and sequencing in *O. tshawytscha*, *O. keta* and *O. nerka*. Fifty-seven targets were chosen from the *Oncorhynchus*-only assembly, and 59 targets were chosen from the *O. mykiss* × *S. salar* assembly (Table 1). All primers designed based on the *O. mykiss* × *S. salar* alignment were based on contigs that were two or more accessions deep in each species.

Ten individuals of each species were sampled from five spawning sites (Fig. 1), yielding a total of 50 individuals of each species. DNA was extracted from fin tissue from each individual using a DNeasy Tissue Kit (QIAGEN).

Table 1 Amplification and sequencing success rates observed in three *Oncorhynchus* species with PCR primers that were designed using two different strategies: The first row lists results for primers that were designed using DNA sequence from the three target species plus *Oncorhynchus mykiss* homologues. The second row lists results for primers that were designed using DNA sequence from *O. mykiss* and *Oncorhynchus salar*

			No. of PCR products sequenced:	
Origin of sequence for primer design	No. of primer pairs tested	No. of PCR products	in at least one species	in all three species
Target species + *O. mykiss*	57	43	29 (51%)	18 (32%)
O. mykiss + *S. salar*	59	59	51 (86%)	36 (61%)
Totals:	116	102	80	54

Nested PCR primers were designed for the 116 target loci using PRIMER3 (Rozen & Skaletsky 2000; Table S1, Supplementary material). Initial PCR conditions were identical for all primer pairs: approximately 10 ng genomic DNA were amplified in 5-μL reactions consisting of 1.2 pmol each external primer, 1× Ampli*Taq* Gold PCR Mastermix (Applied Biosystems; AB) (15 mM Tris-HCl, pH 8.05, 50 mM KCl, 200 μM each dNTP, 2.5 mM $MgCl_2$, 0.05 U/μL DNA polymerase and proprietary stabilizers) and 0.08% glycerol. The thermal cycling profile included an initial denaturation of 94 °C for 4 min; 40 cycles of 94 °C for 20 s, 55 °C for 25 s and 72 °C for 1 min; and a final extension of 72 °C for 7 min. One microlitre of product from the initial reaction was amplified using the internal primers under reaction conditions identical to those used for external primers. PCR products were purified using Sephadex G-50 in spin-columns as described by Sambrook & Russell (2001), sequenced in both directions using the BigDye 3.1 Cycle Sequencing Kit (AB), and analysed on an ABI 3730 capillary DNA sequencer (AB).

PHRED (Ewing & Green 1998; Ewing *et al*. 1998), PHRAP (www.phrap.org), and CONSED (Gordon *et al*. 1998) were used to call bases, assemble sequences into contigs, and view assemblies, respectively. Contigs having either < 100 high-quality bases (PHRED quality value ≥ 20) or containing less than 10 reads were discarded. POLYPHRED (Nickerson *et al*. 1997) was run on the remaining contigs using stringent conditions (the flags: –score 90 and –quality 50 were used to prevent low-quality sequence from being considered, and the –source option was used to compare forward and reverse sequences of each PCR product) in order to identify high-quality discrepancies between sequence reads. The contigs were then revisited using CONSED, and each SNP call made by the software was either confirmed or rejected based on visual inspection of the chromatograms. Substitutions and indels involving one or two nucleotides were labelled as SNPs, but larger indels and tandem repeat variation were not. Consensus sequences for all contigs were trimmed where the ends fell below PHRED quality values of 20, except where visual inspection of the traces revealed additional high-quality sequence data. Trimmed sequences were compared to GenBank nonredundant nucleotide and amino acid sequence databases using BLASTN and BLASTX, respectively. In order to minimize reporting false matches, we followed the convention of Rexroad *et al*. (2003) and Rise *et al*. (2004) and only reported hits with e-values ≤ 10^{-5}. BLASTX hits were reported for sequences that did not have a BLASTN hit with an e-value < 10^{-5}. Sequences without significant matches to either database were designated 'u', for unknown, followed by a number.

The script tagRepeats.perl distributed with the PHRAP package was used to screen the contigs for sequences that had been identified as repetitive in GenBank salmonid accessions. Assemblies that revealed fixed heterozygotes (i.e. for which all or nearly all individuals examined showed two overlapping peaks) at multiple nucleotide positions were labelled 'duplicated'.

Genotyping assay development

A subset of the intraspecific SNPs observed in the sequence data were targeted for development of genotyping assays based on the 5′-nuclease reaction (Holland *et al*. 1991). We targeted 1–3 SNPs in each contig that (i) did not fall in suspected repetitive or duplicated regions, and (ii) did not fall on sequences not amenable to the 5′-nuclease reaction (i.e. SNPs near the ends of contigs and those in microsatellites). We used Assays-by-Design (AB) to design primers and allele-specific probes for 33 SNPs in *O. tshawytscha*, 37 SNPs in *O. keta* and 31 SNPs in *O. nerka* (Table 2). The assays were performed in 384-well reaction plates, with two wells in each plate as negative controls (no template). The templates for each reaction were the 50 individuals of each species that had been sequenced. Reactions were conducted in a 5-μL volume consisting of 0.15 μL template DNA in 1× TaqMan PCR cocktail (AB) 900 nM each PCR primer, and 200 nM each probe (Table S2, Supplementary material). Thermal cycling was performed on an ABI7900 real-time sequence detection system as follows: an initial denaturation of 10 min at 95 °C was followed by 50 cycles of 92 °C for 15 s and at annealing temperature (Table S2) for 1 min. All cycling was conducted at a ramp speed of 1 °C per second. SEQUENCE DETECTION software 2.1 (AB) was used to plot the fluorescence being emitted by the two allele-specific probes against each other as a scatter plot.

We predicted whether the SNP targeted by each assay was in a coding region and whether or not it was synonymous by comparing the sequences to homologous peptides (≤ 10^{-5} BLASTX matches) where possible using CSNPER (Kim *et al*. 2003). Where no homologs could be found, these predictions were made by searching sequences for known compositional features using GENSCAN (Burge & Karlin 1997).

Results

SNP discovery

Alignment of the *Oncorhynchus* sequences revealed 155 contigs that were four or more reads deep. Alignment of the *Oncorhynchus mykiss* × *Salmo salar* sequences revealed 151 586 matches spread across 39 981 unique accessions in both species. Of those, 12 988 were matches which spanned ≥ 500 nucleotides in length. Multiple alignment of the 12 988 sequences collapsed them into 6897 contigs containing reads from both species.

Between the alignment strategies described here, we were able to sequence approximately 69% (80/116) of the regions targeted (Table 1). Primers designed based on

Table 2 Eighty-eight contigs spread across 57 loci in three species of Pacific salmon. Locus IDs are based on the best GenBank BLAST match. The number of SNPs observed in each contig is listed along with the number of successful (+) and failed (–) genotyping assays designed. The species (Sp.) in which each match was observed is listed along with the corresponding e-value

	Oncorhynchus tshawytscha				*Oncorhynchus keta*				*Oncorhynchus nerka*			
Locus ID*	bp†	SNP‡	e-value§	Sp.¶	bp†	SNP‡	e-value§	Sp.¶	bp†	SNP‡	e-value§	Sp.¶
Vtg[d]	710[1]		0	*Omy*	189[1]		e-89	*Omy*				
	525[2]	6–	0	*Omy*	462[2]	2–	0	*Omy*	512[2]	2	0	*Omy*
IL-8R[d]	340[1]		e-174	*Omy*					346[1]		e-178	*Omy*
	822[2]	2	e-148	*Omy*	355[2]	1	e-178	*Omy*	276[2]	1	e-178	*Omy*
ins[d]	908[1]	19–	0	*Oke*					566[1]		0	*Oke*
	516[2]	5+	0	*Oke*	526[2]	13–	0	*Oke*	519[2]	13+	0	*Oke*
RH2 opsin[d]	421[1]	3	e-156	*Oki*	530[1]		e-176	*Oke*				
	476[2]		e-129	*One*	491[2]		e-132	*One*	491[2]		e-132	*One*
POMC	474	2–	0	*Oke*	474	3–	0	*Oke*	473		0	*Oke*
IGF-I.1	316	1+	e-180	*Ots*	372	2	0	*Ots*	370		0	*Ots*
hsc71	359		e-165	*Omy*	342	2+	e-174	*Omy*	339	1–	e-176	*Omy*
u6	454	2– +							463			
	433											
OMLYRNA	310		e-117	*Omy*								
Sycn	658		e-25[x]	*Mmu*	515	9–	e-24[x]	*Rno*	475	9–	e-24[x]	*Rno*
LHB	286	4–	e-144	*Ots*	350	4	e-173	*Oki*	371	4–	e-178	*Oki*
E2	432	1+	e-135	*Omy*	409	1	e-109	*Omy*	411	1+	e-129	*Omy*
GnRH[d]					402[1]	5–	0	*Omy*	419[1]	6–	0	*One*
	439[2]	4+	0	*Ots*	420[2]	4–	0	*One*	449[2]	9–	0	*One*
CaR	879		0	*Ssa*	487		0	*Ssa*	474		0	*Ssa*
LWSop	771	4+	e-161	*Ogo*					778		e-161	*One*
SWS1op	539	1+	e-165	*Oke*								
RH1op[d]	409		0	*Ots*	440	3– +	0	*Ogo*	461	5–	0	*Ogo*
u12	444				548		10	–	456	9+		
	381											
u4[d]	381	3+							381	1–		
CYP1[d]	439	7	0	*Omy*	390	8	0	*Omy*	399	9	0	*Omy*
Ikaros[d]	550	1+	0	*Omy*	644	3–	0	*Omy*	444	3–	0	*Omy*
Moesin	520[1]		e-15	*Gga*	270[1]		e-15	*Gga*				
	302[2]				298[2]	1+			309[2]			
u200[d]	876	5–			840	4+			853			
ZNF330	451	1+			437	7	e-5	*Hsa*				
	422	1										
TniUPP	435	1–	e-5[x]		322		e-6[x]	*Tni*				
serpin	442		e-15[x]		429	3+	e-14[x]	*Tni*	412	1+	e-14[x]	*Tni.*
AsnRS[d]	736[1]	2	e-15	*Omy*								
	507[2]	1+	e-6	*Rno*					651[2]	3–	e-17	*Gga*
u201	199[1]				192[1]							
	223[2]				222[2]	2–			224[2]			
arf	380[1]	3+			382[1]	1			361[1]			
	480[1]	3			875[1]	2+	e-6	*Mmu*	428[1]		e-6	*Mmu*
	144[2]				327[2]				320[2]	2		
	318[3]				318[3]	3			319[3]	1+		
u22	294	2			247				283	1–		
C4[d]	823[2]		e-61	*Omy*								
	801[5]	2	e-65	*Omy*	771[5]		e-61	*Omy*	740[5]		e-61	*Omy*
	540[6]	1–	e-52	*Omy*	520[6]		e-55	*Omy*	536[6]	3–	e-54	*Omy*
HGFA	368[1]				145[2]				114[2]			
	507[2]	2+	e-18	*Ssa*	504[2]	4	e-15	*Ssa*	510[2]		e-15	*Ssa*
	473[4]		e-137	*Omy*	551[4]	4+	e-138	*Omy*	509[4]	1++	e-135	*Omy*
RFC2	672	1– +	e-15	*Hsa*	684	4+	e-14	*Hsa*	679	3+	e-14	*Has*
Tsha1	580		e-73	*Omy*	343	2+	e-13	*Omy*	426		e-13	*Omy*

Table 1 *Continued*

	Oncorhynchus tshawytscha				*Oncorhynchus keta*				*Oncorhynchus nerka*			
Locus ID*	bp†	SNP‡	e-value§	Sp.¶	bp†	SNP‡	e-value§	Sp.¶	bp†	SNP‡	e-value§	Sp.¶
u26[d]					274	1–			330	5–		
GPDH									546[2]	5+	e-53	*Gmo*
	401[3]	1+			407[3]	3+			428[3]	3+		
IL-1b					360	3–	e-83	*Omy*				
ctss-prov	328		e-26	*Fhe*	314		e-133	*Omy*	383	6	e-52	*Dre*
	319		e-44	*Dre*	334		e-53	*Omy*				
DEAD[d]	207		e-35	*Dre*	169		e-6	*Pma*	240		e-23	*Omy*
					496		e-20	*Pma*				
					235		e-21	*Pma*				
CKmito1[d]	498		e-38	*Cac*	506		e-36	*Cac*	501		e-35	*Cac*
copa	336		e-5	*Dre*	350	1+	e-5	*Dre*	339		e-5	*Dre*
eif4ebp2	192		e-12	*Dre*	205	1+	e-12	*Tni*	216		e-12	*Tni*
u30					364							
u202	382[1]				379[1]				374[1]			
	349[2]	2+			341[2]	1+			348[2]			
F10[d]	647	1							626	1–	e-12[x]	*Tru*
u211[d]	243[1]				239[1]				259[1]			
	265[2]	1+			293[2]	3–			258[2]			
MGC75633	279		e-13	*Xtr*	286		e-13	*Xtr*	268		e-13	*Xtr*
u216					225[1]	3			220[1]	1		
	360[2]	4			359[2]	2+			358[2]	1+		
	350[3]	1			353[3]	5			359[3]			
Zp3b	235[1]		e-7[x]	*Ola*	236[1]	1	e-7[x]	*Ola*	233[1]	1+	e-7[x]	*Ola*
	520[2]	3+	e-6[x]	*Dre*	508[2]	6+	e-6[x]	*Dre*	518[2]	2	e-6[x]	*Dre*
il-1racp	347	2+	e-46	*Ssa*	343	8+	e-46	*Ssa*	345		e-48	*Ssa*
u217	458	1–			446	4 – +			447			
RIKEN	615	2	e-16[x]	*Rno*					623		e-16[x]	*Rno*
BAMBI	310		e-35	*Oar*	308	1+	e-32	*Oar*	312		e-35	*Oai*
DRB1	438	4–	e-10[x]	*Hsa*	426	2	e-10[x]	*Hsa*	433		e-9[x]	*Hsa*
u212	173[1]				168[1]				168[1]			
	206[2]	1+			203[2]	1+			203[2]			
u218	213				215				213			
SClkF2R2[d]	232[1]	1+	e-110	*Oke*	233[1]		e-122	*Oke*	229[1]	1–	e-116	*Oke*
					273[2]		e-140	*Oke*				
	269[3]		e-132	*Oke*	268[3]	2+	e-137	*Oke*	247[3]		e-137	*Oke*

*d, duplicated locus.

†superscript numbers following contig lengths indicate the primer set for that locus that was used to amplify that region (Table S1, Supplementary material).

‡+/– indicate number of SNPs validated/not validated by 5′-nuclease reaction.

§x, BLASTX result.

¶*Cac*, *Chaenocephalus aceratus*; *Dre*, *Danio rerio*; *Fhe*, *Fundulus heteroclitus*; *Gga*, *Gallus gallus*; *Gmo*, *Gadus morhua*; *Hsa*, *Homo sapiens*; *Mmu*, *Mus musculus*; *Oar*, *Ovis aries*; *Ogo*, *Oncorhynchus gorbuscha*; *Oke*, *Oncorhynchus keta*; *Oki*, *Oncorhynchus kisutch*; *Ola*, *Oryzias latipes*; *Omy*, *Oncorhynchus mykiss*; *One*, *Oncorhynchus nerka*; *Ots*, *Oncorhynchus tschawytscha*; *Pma*, *Pagrus major*; *Rno*, *Rattus norvegicus*; *Ssa*, *Salmo salar*; *Tni*, *Tetraodon nigroviridis*; *Tru*, *Takifugu rubripes*; *Xtr*, *Xenopus tropicalis*.

O. mykiss × *S. salar* alignments had a significantly ($\chi^2 = 10.10$, d.f. = 1, $P = 0.001$) higher success rate (61% of primer pairs produced sequence in all three species) than those designed based only on *Oncorhynchus* sequences (32% of primer pairs produced sequence in all three species).

The 80 PCR products that were sequenced yielded 68–78 contigs per species that were spread across 57 target regions (Table 2). Average sequence length was 410 bp resulting in a total of approximately 89 kb. Contig depth at the points SNPs were identified had an average value of 63 sequence reads; the read depth for each individual SNP is documented in the corresponding dbSNP record. Nearly 35% of the contigs examined had no ($e < 10^{-5}$) matches to named nucleotide or protein sequences. Over half (54%) of the contigs examined contained SNPs. The overall frequency was 4.30×10^{-3} SNPs per base pair (Fig. 2), which

Table 3 Numbers of transitions, transversions, and indels observed in three species of Pacific salmon

	bp	Transition	Transversion	Ratio	Indel	SNP freq ($\times 10^{-3}$)
Oncorhynchus tshawytscha	34 337	67	45	1.49	2	3.32
Oncorhynchus keta	27 139	73	77	0.95	5	5.71
Oncorhynchus nerka	27 571	55	56	0.98	3	4.13

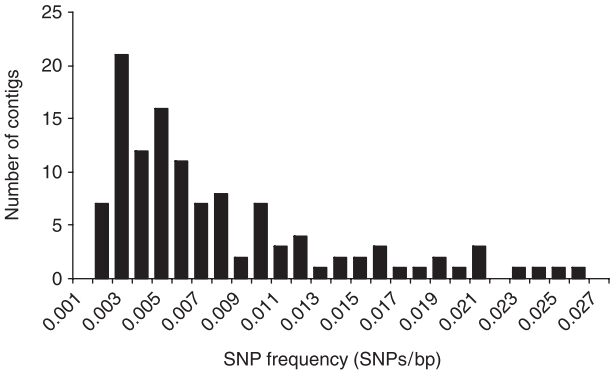

Fig. 2 Frequencies of SNPs observed in three species of Pacific salmon.

translates to approximately one SNP every 239 bp. The rate was higher in *Oncorhynchus keta* (1/175 bp) than in either of the other species (1/242 bp in *Oncorhynchus nerka* and 1/301 bp in *Oncorhynchus tshawytscha*; Table 3). Reverse strand sequence was available to corroborate both alleles at over half of all documented SNPs (57% in *O. tshawytscha*, 76% in *O. keta*, and 69% in *O. nerka*). The observed transition/transversion ratio (ti/tv) was similar for *O. keta* and *O. nerka* (0.95 and 0.98, respectively) and was somewhat higher in *O. tshawytscha* (1.49; Table 3). Sequences were deposited in GenBank (Accession nos DQ025536–DQ025751), and all observed SNPs were documented in dbSNP (ss48398580–ss48398969).

Overlapping sequence in all three species was obtained for 52 of the contigs examined here. Alignments of the consensus sequences from each revealed 4.0% sequence divergence between *O. tshawytscha* and *O. keta*, 3.7% between *O. tshawytscha* and *O. nerka*, and 2.2% between *O. keta* and *O. nerka*. ANOVA indicated that percentage sequence identity was not equal among all pairs of species ($P = 0.032$), and subsequent *t*-tests confirmed that sequence identity was higher between *O. keta* and *O. nerka* than it was between *O. tshawytscha* and *O. keta* ($t = 3.77$, d.f. = 51, $P < 0.000$), or *O. tshawytscha* and *O. nerka* ($t = 3.29$, $P = 0.002$).

Three of the loci examined (POMC, IL-1b, and DEAD) contained stretches of sequence that PHRAP flagged as bearing similarity to *Hpa*I repetitive elements.

Several of the contigs exhibited features that suggested the sequences might represent duplicated regions or paralogs. One of the initial features examined was heterozygosity among sequence reads. Some of the sequences contained nucleotide positions that were heterozygous in every individual. Nine loci (Vtg, IL-8R, ins, RH2op, GnRH, RH1op, CYP1, Ikaros and DEAD) exhibited such heterozygous 'excess' at multiple SNPs in a contig and were thus labelled 'duplicated'. Another criterion we used to label a sequence as 'duplicated' was fixed or near-fixed heterozygosity based on validation assays. Thirteen loci (RH1op, Vtg, ins, CYP1, CKmito1, GnRH, u26, u211, IL-8R, LWSop, SClkF2R2, u4 and AsnRS) exhibited fixed heterozygosity, and another two loci (C4 and F10) exhibited nearly fixed heterozygosity. In total, 18/57 (~32%) of the regions examined in this study met one or more of our criteria for being labelled 'duplicated'.

Genotyping assay development

Approximately half (53/101) of the 5′-nuclease assays were successful in discriminating alleles at the intraspecific SNPs that they were designed to target, and thus in validating the respective polymorphisms (Table S2). The most common reason for assays being classified as failures (41/48 failed assays) was a single fluorescence pattern for all individuals (a single genotype indicated for all samples). Other reasons for assays being classified as failures included the lack of detectable PCR product (4 assays) and the presence of product that departed from our expectations based on the DNA sequence data (e.g. more than three genotypes indicated; 3 assays). The 5′-nuclease reactions that were successful in discriminating alleles included 31 that targeted transitions, 21 that targeted transversions, and one that targeted an indel.

Five of the SNPs were predicted to be synonymous, two were predicted to be nonsynonymous, and the remainder fell outside predicted exons. The frequency of the minor allele (*q*) at the validated SNPs ranged from 1% to 49%, with an average of 17% (Table S2).

Discussion

SNP discovery

The use of conserved primers for characterizing SNPs across multiple target species presents a potential economy

for SNP development in nonmodel species (e.g. Aitken *et al*. 2004). Primers designed based on alignments of *Oncorhynchus mykiss* and *Salmo salar* were more successful in producing sequence data in the target species than were primers designed based on only *Oncorhynchus* sequences. We expect that, should these primers be applied to other *Oncorhynchus* species (e.g. *O. kisuteh*, *O. gorbuscha*, *O. mykiss*, and *O. masu*), then this pattern would persist and that more than half of the primers presented here would provide high-quality sequence data in those species. The use of these primers in nonsalmonid fishes may also be productive; however, we would advise alignment of the primers with sequence data from additional species (e.g. *Danio rerio*) prior to testing this hypothesis.

Allendorf & Thorgaard (1984) observed that approximately 50% of enzyme-encoding loci were retained as functional duplicates within the salmon genome. The persistence of duplicate loci in the salmonid lineage, despite millions of years of evolution since the tetraploidization event, is thought to be related to selective advantages resulting from subfunctionalization of specific genes (review in Prince & Pickett 2002). These selective advantages would be influenced by environmental factors suggesting that any given set of paralogs may have diverged in one species but not in another. In addition to the loci that we labelled as potentially duplicated due to excessive heterozygosity, some of the loci presented here are similar to regions known to be retained as duplicates (e.g. opsin genes; Dann *et al*. 2004). Further, some portion of the primers with which we were not able to obtain clean sequence data may have failed by virtue of the fact that they amplified multiple loci. Because the sequences presented here were not matched to specific chromosomal locations and because the reads are relatively short, our ability to distinguish between paralogs and regions exhibiting similarity for other reasons is limited. Our observation that 32% of the contigs we examined met one or more of our criteria for being labelled duplicated, considered in light of the probability of additional duplicated loci based on the above factors, seems reasonably close to the prediction of 50% that one might expect based on allozyme data. Genetic mapping of the loci presented here would clearly improve our understanding of duplication in the salmonid genome.

The frequencies with which SNPs were observed in the species examined here (from 1/175 bp in *O. keta* to 1/301 bp in *O. tshawytscha*) seem high compared to the frequency of SNPs reported in *S. salar* (1/586 bp; Hayes *et al*. 2004), but were well within the range of SNP frequencies observed in other species (summarized by Brumfield *et al*. 2003). The difference between the rates observed here and those reported for *S. salar* may reflect the fact that the fish examined here represent diverse wild populations rather than inbred aquaculture strains, and it may reflect the relatively large ascertainment sample used here (average contig depth of 63 reads relative to, for example, the species summarized by Brumfield *et al*. 2003).

The ti/tv ratio that we observed in *O. tshawytscha* (1.49) and the rate reported previously for *S. salar* (1.37; Hayes *et al*. 2004) are both higher than that reported for the nearest model organism, *D. rerio* (1.20; Stickney *et al*. 2002). The ti/tv ratios that we observed in *O. keta* and *O. nerka* (0.95 and 0.98, respectively), conversely, are lower than that reported for *D. rerio*. The difference that we observed between the species examined here may simply be an artefact of the specific loci that were examined. It should be noted, however, that if we disregarded the three loci with the most transversions from both *O. keta* and *O. nerka*, the ti/tv ratios in these species (1.05 in *O. keta* and 1.16 in *O. nerka*) remained lower than that reported for *D. rerio*. Given the small phylogenetic distance between *O. nerka* and *O. keta* relative to that between either of these species and *O. tshawytscha* (Utter & Allendorf 1994; Phillips *et al*. 2004), the similar ti/tv ratios observed in the former two species is interesting. Examination of *O. gorbuscha* (related to *O. nerka* and *O. keta*) and *O. kisutch* (related to *O. tshawytscha*) could aid in the interpretation of this result, as could examination of additional loci in the present species.

Genotyping assay development

The proportion of 5′-nuclease assays that were successful in discriminating alleles at the SNPs that they were designed to target (~50%) was lower than that reported for humans (80–92%; Morin *et al*. 1999; T. Dodge, Applied Biosystems, personal communication). Most of the assays designated as 'failures' in the present work exhibited a single fluorescence pattern for all individuals suggesting that either: (i) the putative SNP being targeted was a technical artefact rather than a real polymorphism, (ii) the region being targeted by primers and probes exists across multiple loci, or (iii) the probes used were not sufficiently sensitive to bind or not bind based on a single-base difference. The raw data from the 5′-nuclease assays did not facilitate distinction between these possibilities. The third possibility should account for ~10% of all assays failing based on the results of Morin *et al*. (1999). The first possibility probably accounts for some of our failures, but it is expected to be minimal based on the stringent conditions that we used to call SNPs. The second possibility seems likely to be more of an issue in salmon than in species that have not undergone recent genome duplication events. Although predicting assay success rates based on the 50% duplication estimate of Allendorf & Thorgaard (1984) is subject to errors based on a large number of factors, our results are surprisingly consistent with their estimate.

Assays in which more than three genotype clusters were detected may represent other duplication situations in which the SNP and its flanking region are being retained as duplicate loci. The multiple clusters observed in such assays may reflect cases in which the sequencing primers were specific to target a single locus, but the assay primers or probes are not.

The observation that only 39% of the *O. nerka* assays succeeded while 54% and 64% of *O. keta* and *O. tshawytscha* assays succeeded is of interest. The total number of SNPs observed in *O. nerka* was identical to that observed in *O. tshawytscha* so differences in the number of assays that worked could not be attributed to differences in the number of SNPs between species. The maintenance of duplicate loci via subfunctionalization relies to a large extent on drift and thus on effective population size. Species with small effective population sizes are subject to higher rates of genetic drift and are thus more likely to retain duplicate loci due to subfunctionalization (Prince & Pickett 2002). The relatively stringent freshwater requirements of the *O. nerka* life history limit gene flow in this species relative to other salmonids (Quinn 2005). While it is tempting to speculate that relatively few assays worked in *O. nerka* because that species retains a higher proportion of identity between paralogs, making it less likely that a set of PCR primers and probes will anneal to a single locus, mapping data and preferably a larger number of loci will be needed to test this hypothesis.

Two factors that should be considered when using SNPs (or any molecular markers) in population studies are ascertainment bias and the possibility of natural selection acting on allele frequencies of the markers. Ascertainment bias results in the selection of loci that have intermediate allele frequencies in the ascertainment sample, and loci that show increased variation in populations represented in the ascertainment collections relative to those not represented. Large and broadly sampled ascertainment collections afford one possibility for minimizing ascertainment bias. Statistical corrections have further been proposed, allowing researchers to correct for ascertainment bias when estimating several parameters including allele frequencies (Nielsen *et al*. 2004), migration (Wakeley *et al*. 2001), linkage disequilibrium (Nielsen & Signorovitch 2003), and population growth (Polanski & Kimmel 2003). In the case of selection, each locus will need to be considered individually. Although our *in silico* analyses suggested that only a few of the SNP genotyping assays presented here code for nonsynonymous substitutions, the fact that most of the PCR primers were developed based on ESTs and that a few had significant matches to genes thought to be under selective influence (e.g. eif4ebp2; Richter & Sonenberg 2005) would caution against the assumption of neutrality of these markers. Examining population data collected using the present assays for outlier loci (Beaumont & Nichols 1996; also see review by Luikart *et al*. 2003) may allow inferences to be made regarding the impacts of natural selection on allele frequencies at these loci in those populations. Natural selection may limit the utility of these markers for estimating population genetic parameters such as migration and effective population size, but could also render them very useful in the context of association and quantitative trait studies as well as in mixture analyses.

Conclusion

Our results suggest that the use of *Oncorhynchus mykiss* and *Salmo salar* sequence for SNP development in Pacific salmon is a straightforward although costly endeavour. Starting with 116 loci targeted in this study enabled us to sequence 68–78 contigs per species, from which we were able to design 12–21 genotyping assays per species. The number of SNPs required for population studies will vary considerably based on the specific objectives of the study, distinctiveness of the taxonomic units of interest, information content of the specific markers available, and the desired probability level (Seddon *et al*. 2005). While nine SNPs were sufficient for addressing international treaty objectives for *Oncorhynchus tshawytscha* in the Yukon River (Smith *et al*. 2005), the factors above will likely necessitate larger numbers of loci for analyses of salmon collected on the high seas and for projects in other species. Assuming we can generalize the number of SNP loci needed for a project to approximately that observed in cows (80; Werner *et al*. 2004), and assuming that the efficiency with which SNPs are converted to genotyping assays does not improve beyond that observed here, then approximately 500 loci would need to be sequenced in a species in order to discover an adequate SNP panel. The number of sequences presently in GenBank for several species, including *O. mykiss* and *S. salar*, are clearly sufficient for this purpose. While the costs of the required DNA sequencing are significant, for Pacific salmon these must be compared to the costs of redundant genotyping presently being done using non-standardized markers and to the costs of using genotyping technologies that are, relative to what is available, increasingly labour intensive.

Acknowledgements

This work was funded by grants from the North Pacific Research Board, the Southeast Sustainable Salmon Fund, the Pacific Salmon Commission's Yukon River Salmon Agreement and Northern Boundary Restoration & Enhancement Funds, and by the State of Alaska. The authors would like to express gratitude to C. Hawryshyn (UVIC) for providing unpublished opsin gene sequences. DNA sequencing was performed by Polymorphic DNA Technologies Incorporated. Comments made by L. Bernatchez and two anonymous referees were used to improve this manuscript substantially.

Supplementary material

The supplementary material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2731/MEC2731sm.htm

Table S1 Primers used to amplify loci in *Oncorhynchus tshawytscha, Oncorhynchus keta*, and *Oncorhynchus nerka*.

Table S2 Primers and probes used for 53 5′-nuclease reactions in Pacific salmon. Each probe was designed with either VIC or 6FAM on its 5′-end and a nonfluorescent quencher and minor groove binder on its 3′-end. Annealing temperatures (in °C) are given for each assay. The frequency of the minor allele (q) among 50 individuals in the ascertainment sample is given for each SNP.

References

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

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. Turner BJ), pp. 1–53. Plenum Press, New York.

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Altschul S, Madden T, Schaffer A *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Avise JC (2004) *Molecular Markers, Natural History and Evolution*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Bailey GS, Poulter RTM, Stockwell PA (1978) Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. *Proceedings of the National Academy of Sciences, USA*, **75**, 5575–5579.

Batley J, Hayes PK (2003) Development of high throughput single nucleotide polymorphism genotyping for the analysis of *Nodularia* (Cyanobacteria) population genetics. *Journal of Phycology*, **39**, 248–252.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **263**, 1619–1626.

Bentzen P, Harris AS, Wright JM (1991) Cloning of hypervariable minisatellite and simple sequence microsatellite repeats for DNA fingerprinting of important aquacultural species of salmonids and tilapia. In: *DNA Fingerprinting: Approaches and Applications* (eds Burke T, Dolf G, Jeffreys AJ, Wolff R), pp. 243–262. Birkhauser-Verlag, Basel, Switzerland.

Brugmans B, van der Hulst RGM, Visser RGF, Lindhout P, van Eck HJ (2003) A new and versatile method for the successful conversion of AFLP markers into single locus markers. *Nucleic Acids Research*, **31**, e55.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, **268**, 78–94.

Cossins AR, Crawford DL (2005) Fish as models for environmental genomics. *Nature Reviews Genetics*, **6**, 324–333.

Dann SG, Allison WE, 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**, 1–13.

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

Elfstrom CM, Gaffney PM, Smith CT, Seeb JE (2005) Characterization of 12 single nucleotide polymorphisms in weathervane scallop. *Molecular Ecology Notes*, **5**, 406–409.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, **8**, 186–194.

Ewing B, Hillier L, Wendl M, Green P (1998) Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Research*, **8**, 175–185.

Gharrett AJ, Matala AP, Peterson EL *et al.* (2005) Two genetically distinct forms of rougheye rockfish are different species. *Transactions of the American Fisheries Society*, **134**, 242–260.

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Research*, **8**, 195–202.

Groot C, Margolis L, eds (1991) *Pacific Salmon Life Histories*. UBC Press, Vancouver.

Haberl M, Tautz D (1999) Comparative allele sizing can produce inaccurate allele size differences for microsatellites. *Molecular Ecology*, **8**, 1347–1349.

Hayes B, Laerdahl J, Lien S *et al.* (2004) Detection of single nucleotide polymorphisms (SNPs) from Atlantic salmon expressed sequence tags (ESTs). *Proceedings of the 55th Annual Meeting of the European Association for Animal Production*. Bled, Slovenia.

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.

Kim H, Schmidt CJ, Decker KS, Emara MG (2003) A double-screening method to identify reliable candidate non-synonymous SNPs from chicken EST data. *Animal Genetics*, **34**, 249–254.

Kwok P-Y (2003) *Single Nucleotide Polymorphisms — Methods and Protocols*. Humana Press, Totowa, New Jersey.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lyons LA, Laughlin TA, Copeland NG *et al.* (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, **15**, 47–56.

McKay SJ, Devlin RH, Smith MJ (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.

McPhail JD (1997) The origin and speciation of *Oncorhynchus* revisited. In: *Pacific Salmon and Their Ecosystems: Status and Future Options* (eds Stouder DJ, Bisson PA, Naiman RJ), pp. 29–38. Chapman & Hall, New York.

Meksem K, Ruben E, Hyten D, Triwitayakorn K, Lightfoot DA (2001) Conversion of AFLP bands into high-throughput DNA markers. *Molecular Genetics and Genomics*, **265**, 207–214.

Morin PA, Saiz R, Monjazeb A (1999) High-throughput single nucleotide polymorphism genotyping by fluorescent 5′ exonuclease assay. *BioTechniques*, **27**, 538–522.

Nelson RJ, Beacham TD, Small MP (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 RJ, Stoehr M, Cooper G, Smith CT, Mehl H (2003a) High levels of chloroplast genetic variation differentiate coastal and interior Douglas-fir (*Pseudotsuga menziesii*) lineages in southern British Columbia. *Forest Genetics*, **10**, 153–157.

Nelson RJ, Wood CC, Cooper G, Smith C, Koop B (2003b) Population structure of sockeye salmon of the central coast of British Columbia: implications for recovery planning. *North American Journal of Fisheries Management*, **23**, 703–720.

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.

Nicod J-C, Largiader CR (2003) SNPs by AFLP (SBA): a rapid SNP isolation strategy for non-model organisms. *Nucleic Acids Research*, **31**, e19.

Nielsen R, Hubisz MJ, Clark AG (2004) Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics*, **168**, 2373–2382.

Nielsen R, Signorovitch J (2003) Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology*, **63**, 245–255.

Olsen JB, Wenburg JK, Bentzen P (1996) Semiautomated multilocus genotyping of Pacific salmon (*Oncorhnychus* spp.) using microsatellites. *Molecular Marine Biology and Biotechnology*, **5**, 259–272.

Osman A, Jordan B, Lessard PA *et al.* (2003) Genetic diversity of *Eurycoma longifolia* inferred from single nucleotide polymorphisms. *Plant Physiology*, **131**, 1294–1301.

Phillips RB, Matsuoka MP, Konkol NR, Mckay S (2004) Molecular systematics and evolution of the growth hormone introns in the Salmoninae. *Environmental Biology of Fishes*, **69**, 433–440.

Polanski A, Kimmel M (2003) New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. *Genetics*, **165**, 427–436.

Primmer CR, Borge T, Lindell J, Saetre GP (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.

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics*, **3**, 827–837.

Quinn TP (2005) *The Behavior and Ecology of Pacific Salmon and Trout*. American Fisheries Society in association with University of Washington Press, Bethesda.

Reimchen TE, Mathewson DD, Hocking MD, Moran J, Harris D (2003) Isotopic evidence for enrichment of salmon-derived nutrients in vegetation, soil, and insects in riparian zones in coastal British Columbia. In: *Nutrients in Salmonid Ecosystems: Sustaining Production and Biodiversity* (ed. Stockner J), pp. 59–69. American Fisheries Society Symposium 34, Bethesda.

Rexroad CEI, Lee Y, Keele JW *et al.* (2003) Sequence analysis of a rainbow trout cDNA library and creation of a gene index. *Cytogenetic and Genome Research*, **102**, 347–354.

Richter JD, Sonenberg N (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*, **433**, 477–480.

Rise ML, Von Schalburg KR, Brown GD *et al.* (2004) Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Research*, **14**, 478–490.

Rozen S, Skaletsky H (2000) PRIMER3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, New Jersey.

Ryynänen HJ, Primmer CR (2004) Primers for sequence characterization and polymorphism detection in the Atlantic salmon (*Salmo salar*) growth hormone 1 (*GH1*) gene. *Molecular Ecology Notes*, **4**, 664–667.

Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sato S, Moriya S, Azumaya T *et al.* (2004) *Genetic Stock Identification of Chum Salmon in the Central Bering Sea and Adjacent North Pacific Ocean by DNA Microarray During the Early Falls of 2002 and 2003*. North Pacific Anadromous Fisheries Commission Document 793. available at www.npafc.org/.

Schlötterer C (2004) The evolution of molecular markers — just a matter of fashion? *Nature Reviews Genetics*, **5**, 63–69.

Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology*, **14**, 503–511.

Seeb LW, Crane PA, Kondzela CM *et al.* (2004) Migration of Pacific Rim chum salmon on the high seas: insights from genetic data. *Environmental Biology of Fishes*, **69**, 21–36.

Smith CT, Templin WD, Seeb JE, Seeb LW (2005) Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of US and Canadian Chinook salmon caught in Yukon River fisheries. *North American Journal of Fisheries Management*, **25**, 944–953.

Stickney HL, Schmutz J, Woods IG *et al.* (2002) Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays. *Genome Research*, **12**, 1929–1934.

Taylor EB, Foote CJ, Wood CC (1996) Molecular genetic evidence for parallel life-history evolution within a Pacific salmon (sockeye salmon and kokanee, *Oncorhynchus nerka*). *Evolution*, **50**, 401–416.

Utter FM, Allendorf FW (1994) Phylogenetic relationships among species of *Oncorhynchus*: a consensus view. *Conservation Biology*, **8**, 864–867.

Utter FM, Ryman N (1993) Genetic markers and mixed stock fisheries. *Fisheries*, **18**, 11–21.

Wakeley J, Nielsen R, Liu-Cordero S, Ardlie K (2001) The discovery of single-nucleotide polymorphisms — and inferences about human demographic history. *American Journal of Human Genetics*, **69**, 1332–1347.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (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.

Werner FAO, Durstewitz G, Habermann FA *et al.* (2004) Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. *Animal Genetics*, **35**, 44–49.

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

The authors presently work in the State of Alaska Department of Fish and Game genetics laboratory. Their shared interests are in using molecular tools to improve understanding of the evolution and population biology of fish and marine invertebrate species and in the application of that understanding toward sustainable management policies. Key aspects of their recent work include investigations of gene flow in scallop, crab, and rockfish species and the collaborative development of laboratory-independent genetic baselines for Pacific salmon species.
