

Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms

S. R. NARUM,* M. BANKS,† T. D. BEACHAM,‡ M. R. BELLINGER,† M. R. CAMPBELL,§ J. DEKONING,¶ A. ELZ,** C. M. GUTHRIE III,†† C. KOZFKAY,§ K. M. MILLER,‡ P. MORAN,** R. PHILLIPS,¶ L. W. SEEB,‡‡,*** C. T. SMITH,‡‡,††† K. WARHEIT,§§ S. F. YOUNG§§ and J. C. GARZA¶¶
 *Columbia River Inter-Tribal Fish Commission, Hagerman Fish Culture Experiment Station, 3059-F National Fish Hatchery Road, Hagerman, ID 83332, USA, †Marine Fisheries Genetics, Coastal Oregon Marine Experiment Station, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365-5229, USA, ‡Department of Fisheries and Oceans, Pacific Biological Station, Nanaimo, BC, Canada V9T 6N7, §Idaho Department of Fish and Game, Eagle Fish Genetics Laboratory, 1800 Trout Road, Eagle, ID 83616, USA, ¶Washington State University, Vancouver, EL5 230F, 14204 NE Salmon Creek Ave., Vancouver, WA 98686-9600, USA, **National Marine Fisheries Service, Northwest Fisheries Science Center, 2725 Montlake Blvd. E., Seattle, WA 98112, USA, ††National Marine Fisheries Service, Auke Bay Laboratory, Alaska Fisheries Science Center, 11305 Glacier Highway, Juneau, AK 99801, USA, ‡‡Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, AK 99518, USA, §§Washington Department of Fish and Wildlife, PO Box 600, Olympia, WA 98501, USA, ¶¶National Marine Fisheries Service, Southwest Fisheries Science Center, 110 Shaffer Road, Santa Cruz, CA 95060, USA

Abstract

Single nucleotide polymorphisms (SNPs) are appealing genetic markers due to several beneficial attributes, but uncertainty remains about how many of these bi-allelic markers are necessary to have sufficient power to differentiate populations, a task now generally accomplished with highly polymorphic microsatellite markers. In this study, we tested the utility of 37 SNPs and 13 microsatellites for differentiating 29 broadly distributed populations of Chinook salmon ($n = 2783$). Information content of all loci was determined by I_n and G'_{ST} , and the top 12 markers ranked by I_n were microsatellites, but the 6 highest, and 7 of the top 10 G'_{ST} ranked markers, were SNPs. The mean ratio of random SNPs to random microsatellites ranged from 3.9 to 4.1, but this ratio was consistently reduced when only the most informative loci were included. Individual assignment test accuracy was higher for microsatellites (73.1%) than SNPs (66.6%), and pooling all 50 markers provided the highest accuracy (83.2%). When marker types were combined, as few as 15 of the top ranked loci provided higher assignment accuracy than either microsatellites or SNPs alone. Neighbour-joining dendrograms revealed similar clustering patterns and pairwise tests of population differentiation had nearly identical results with each suite of markers. Statistical tests and simulations indicated that closely related populations were better differentiated by microsatellites than SNPs. Our results indicate that both types of markers are likely to be useful in population genetics studies and that, in some cases, a combination of SNPs and microsatellites may be the most effective suite of loci.

Keywords: assignment tests, Chinook salmon, differentiation, genetic markers, microsatellites, SNPs

Received 15 December 2007; revision received 2 May 2008; accepted 28 May 2008

Correspondence: S. R. Narum, Fax: 208-837-6047;

Email: nars@critfc.org

***Present address: School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98632, USA.

†††Present address: US Fish and Wildlife Service, Abernathy Fish Technology Center, 1440 Abernathy Creek Rd., Longview, WA 98632, USA.

Introduction

Microsatellites have been the molecular marker of choice in ecological and conservation genetics studies in the last decade due to their high variability and power to resolve population structure. However, complicated mutation models (i.e. two-phase mutation model; Di Rienzo *et al.* 1994), high incidence of homoplasy (Estoup *et al.* 1995;

Garza & Freimer 1996; Hedrick 1999), high potential error rate and low genotyping throughput (Miller *et al.* 2002; Hoffman & Amos 2005) have led researchers to consider alternative marker types. With increasing genomic information available for nonmodel organisms, single nucleotide polymorphisms (SNPs) have begun to see increased use as genetic markers for population genetic studies. These polymorphisms represent the most abundant variation in the genome of most organisms, and are spread throughout the entire genome at high density. Furthermore, mutation rates, mutation models and error rates for SNPs are generally well understood, providing a foundation for estimating parameters that are important in ecological and conservation genetics. In addition, the vast majority of SNPs have only two alleles and this small number of alleles allows rapid genotyping with low error rates (Morin *et al.* 2004). A potential caveat of SNPs as population genetic markers is ascertainment bias in SNP discovery which occurs when markers are chosen based on polymorphism level, SNPs are identified from an ascertainment panel with few individuals, or markers are utilized for populations not included in the discovery panel (Luikart *et al.* 2003; Clark *et al.* 2005). Another complication is that SNP allele frequencies may be affected by natural selection if markers are linked or associated with functional genes. However, recent studies (e.g. Smith *et al.* 2007) and literature reviews (e.g. Luikart *et al.* 2003) indicate that these issues are not significant problems for most population genetic applications (but see Wilding *et al.* 2001). Despite the appealing attributes of SNPs, uncertainty remains whether available bi-allelic SNP markers have similar power to differentiate populations as highly polymorphic microsatellite markers in ecological and conservation genetics studies of nonmodel organisms.

Several recent theoretical studies have addressed the potential utility of SNPs and microsatellites for a wide variety of applications, including estimating genetic variation, pedigree reconstruction, and population structure (see Morin *et al.* 2004 for review). Consensus exists about the need for more SNP than microsatellite markers, especially for applications such as parentage and kinship studies. Kalinowski (2002) demonstrated that, at strictly neutral loci, the coefficient of variation of F_{ST} was determined by the number of alleles, not the number of loci, with similar values for either many markers with few independent alleles, or few markers with many alleles. However, in empirical studies, natural selection and ascertainment bias can complicate the relationship between number of alleles and relative resolving power of different molecular markers.

The application of SNPs in conservation genetics research has been limited to studies that have investigated a small number of populations (e.g. Seddon *et al.* 2005) or few loci (e.g. Bensch *et al.* 2002). These empirical studies have

compared genetic signals of SNPs with microsatellites, but the small number of populations or loci has limited broad inference about comparative power of the two marker types for differentiating populations. Some studies have also attempted to utilize a combination of neutral markers and loci known or believed to be under selection, to distinguish signals of adaptive divergence from genetic drift in natural populations (e.g. Heath *et al.* 2006). Single nucleotide polymorphisms are particularly well suited to studies attempting to identify natural selection, due to their abundance in expressed sequence tags (e.g. Bouck & Vision 2007; Ryyänen *et al.* 2007). However, if natural selection in populations causes significant deviations from the expectations of neutral divergence due to genetic drift, or if genotype frequencies from SNPs do not conform to Hardy–Weinberg proportions, this will violate the assumptions of many statistical analyses that are heavily utilized in population genetics (e.g. that genetic distance estimates reflect evolutionary time since divergence between populations). Thus, more thorough empirical work is needed to evaluate the utility of SNPs for studying population structure.

In this study, we genotyped 37 SNP and 13 microsatellite markers in 29 Chinook salmon (*Oncorhynchus tshawytscha*) populations from a broad geographical range to evaluate the power of both types of markers to differentiate populations and reconstruct phylogeographical relationships. We examined information content of each locus, pairwise differences in population allele frequencies, phylogenetic clustering patterns, and population assignment accuracy with both the SNPs and microsatellites separately, as well as with all 50 loci combined. Additionally, loci with low information content were systematically removed in repeated assignment tests to determine the effect of reducing the number of loci on assignment success. The populations evaluated in this study provided the opportunity to compare these markers across varying levels of population divergence, including between reproductively isolated life history types, among regions separated by large geographical distance (up to 2782 km), and among populations at fine geographical scale (less than 3 km). Patterns of population differentiation at individual loci encompassing a wide range of heterozygosity were also evaluated to provide general recommendations for population genetic studies.

Methods and materials

Sampling and genetic data collection

A total of 2783 tissue samples were taken from Chinook salmon from rivers and hatcheries on the Pacific coast of North America between 35°N and 60°N (Fig. 1) to represent 29 wild and hatchery-reared populations. Fin,

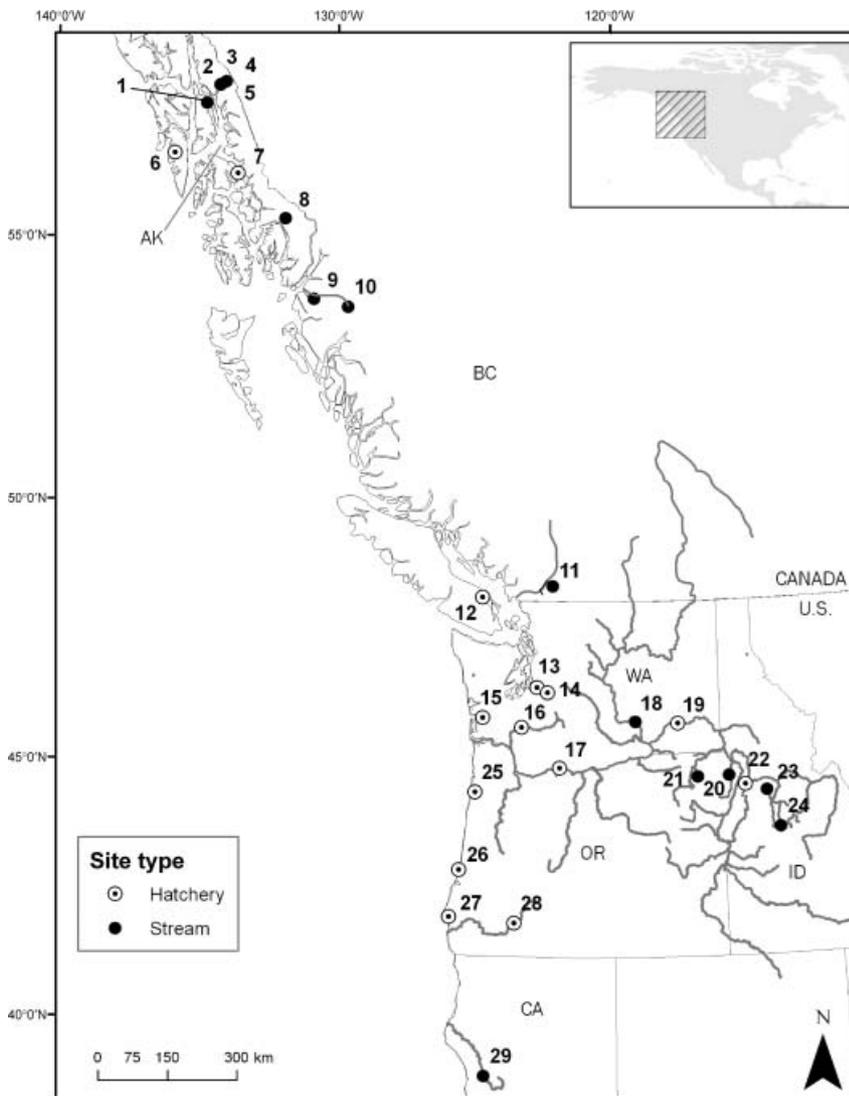


Fig. 1 Map of sample sites for Chinook salmon from the Pacific coast of North America. Sites are numbered as follows: (1) King Salmon H., (2) Kowatua Cr., (3) Nakina R., (4) Upper Nahlin R., (5) Little Tatsamenie R. su, (6) Andrew Medveje H., (7) Andrew Crystal H., (8) Cripple Cr., (9) Kateen R., (10) Little Kitsumkalum R., (11) Harrison R. f, (12) Nanaimo H. f, (13) Soos H., (14) Cle Elum H., (15) Forks Cr. H., (16) Cowlitz H. sp, (17) Spring Cr. H. tule, (18) Hanford Reach f, (19) Lyons Ferry H. f, (20) Imnaha R. sp, (21) Minam R. sp, (22) Rapid R. H. sp, (23) Secesh R. sp, (24) Johnson Cr. sp, (25) Nestucca H. f, (26) Umpqua H. sp, (27) Elk H. f, (28) Cole Rivers H. sp, (29) Eel R. f. Site name abbreviations include 'H.' for hatcheries, 'R.' for rivers, and 'Cr.' for creeks. Where appropriate, adult migration timing follows population name with the following abbreviations (sp = spring, su = summer, f = fall).

opercle, or organ tissues were collected and frozen, stored in ethanol, or dried on blotter paper for preservation.

DNA was extracted from all samples with QIAGEN DNeasy filter-based protocols and arrayed in 96- or 384-well plates for high-throughput genotyping. Template DNA was amplified via polymerase chain reaction (PCR) at a total of 13 microsatellite loci (Table 1) including 9 tetranucleotide and 4 dinucleotide loci. Microsatellite loci were amplified and genotyped following conditions in Seeb *et al.* (2007). Fluorescently labeled PCR products were separated with either Applied Biosystems 377, 3730 or 3100 Genetic Analyzers and scored with GENEMAPPER or GENOTYPER software.

Template DNA was also amplified for 37 SNP loci (Table 1) with at least two wells in each plate as negative (no-template) controls. One locus (*Ots_C3N3*) was mitochondrial with haploid genotypes, and all other loci were

nuclear with diploid genotypes. Reactions consisted of 5 or 10 μ L of 1 \times TaqMan PCR Mastermix (Applied Biosystems), 900 nM of each PCR primer, and 200 nM of each probe. Thermal cycling was performed as follows: an initial denaturation of 10 min at 95 $^{\circ}$ C was followed by 50 cycles of: 92 $^{\circ}$ C for 15 s and 60 $^{\circ}$ C for 1 min. Annealing temperatures were lowered to 58 $^{\circ}$ C for two assays (*Ots_MHC2*, *Ots_AsnRS-60*), and raised to 62 $^{\circ}$ C for four assays (*Ots_MHC1*, *Ots_TCL1*, *Ots_P53*, and *Ots_u6-75*) in order to optimize genotyping results. Following amplification, end-point reads of all plates were performed on either AB7900, AB7500 or AB7300 real-time sequence detection system instruments. Scoring of individual genotypes was performed using SEQUENCE DETECTION SOFTWARE version 2.1 (Applied Biosystems) to generate scatter plots that graphically depicted the amount of each allele-specific probe that bound to the PCR product of each individual.

Table 1 Summary of 50 genetic markers (13 microsatellites and 37 single nucleotide polymorphisms) studied in 29 populations of Chinook salmon. The column 'Deviate NE' indicates loci that deviated from neutral expectations ('+' directional selection; '-' balancing selection)

Locus	Number of alleles	H_E	F_{IS}	F_{ST}	G'_{ST}	In	Deviate NE	Source*
Microsatellite markers								
<i>Ogo2</i>	21	0.764	0.001	0.079	0.336	0.364		1
<i>Ogo4</i>	19	0.778	-0.010	0.109	0.492	0.518	+	1
<i>Ok1100</i>	42	0.936	-0.001	0.025	0.373	0.356		2
<i>Omm1080</i>	53	0.946	0.020	0.029	0.493	0.525		3
<i>Ots201b</i>	52	0.923	-0.006	0.037	0.462	0.475		4
<i>Ots208b</i>	47	0.943	0.008	0.028	0.466	0.450		4
<i>Ots211</i>	35	0.926	0.002	0.038	0.498	0.486		4
<i>Ots212</i>	34	0.866	0.018	0.058	0.426	0.415		4
<i>Ots213</i>	46	0.931	0.007	0.038	0.531	0.513		4
<i>Ots3M</i>	19	0.712	0.007	0.118	0.412	0.386	+	5
<i>Ots9</i>	9	0.581	0.002	0.074	0.179	0.164		5
<i>OtsG474</i>	16	0.487	-0.005	0.187	0.368	0.376	+	6
<i>Ssa408</i>	35	0.874	0.040	0.059	0.463	0.435		7
Single nucleotide polymorphism markers								
<i>Ots_arf-188</i>	2	0.012	0.025	0.000	0.000	0.014		8
<i>Ots_AsnRS-60</i>	2	0.382	0.007	0.067	0.109	0.030	-	8
<i>Ots_C3N3</i>	2	0.241	1.000	0.467	0.619	0.265	+	9
<i>Ots_CYP17</i>	2	0.066	-0.020	0.129	0.138	0.060		2
<i>Ots_E2-275</i>	2	0.407	-0.002	0.188	0.320	0.104		8
<i>Ots_E9BAC</i>	2	0.001	-0.004	0.000	0.000	0.001		2
<i>Ots_ETIF1A</i>	2	0.394	-0.016	0.216	0.361	0.117		2
<i>Ots_FGF6A</i>	2	0.333	-0.012	0.177	0.269	0.087		2
<i>Ots_GnRH-271</i>	2	0.110	0.024	0.187	0.211	0.064		8
<i>Ots_GPDH-338</i>	2	0.081	-0.012	0.049	0.053	0.025	-	8
<i>Ots_GTH2B-550</i>	2	0.316	0.018	0.365	0.538	0.227	+	2
<i>Ots_HGFA-446</i>	2	0.030	0.107	0.000	0.000	0.038		8
<i>Ots_IGF-I.1-76</i>	2	0.190	-0.035	0.185	0.229	0.088		8
<i>Ots_Ikaros-250</i>	2	0.045	-0.037	0.025	0.026	0.033		8
<i>Ots_MHC1</i>	2	0.363	-0.031	0.279	0.441	0.161		9
<i>Ots_MHC2</i>	2	0.299	0.051	0.219	0.314	0.131		9
<i>Ots_NOD1</i>	2	0.311	-0.013	0.295	0.432	0.153		2
<i>Ots_P450</i>	2	0.270	0.043	0.464	0.639	0.284	+	9
<i>Ots_P53</i>	2	0.414	0.038	0.105	0.181	0.050		9
<i>Ots_PGK-54</i>	2	0.275	-0.043	0.236	0.327	0.140		2
<i>Ots_Prl2</i>	2	0.452	-0.003	0.101	0.187	0.053		9
<i>Ots_RAG3</i>	2	0.330	0.028	0.335	0.504	0.188	+	2
<i>Ots_RFC2-558</i>	2	0.304	0.048	0.267	0.386	0.158		8
<i>Ots_SClkF2R2-135</i>	2	0.437	-0.055	0.094	0.169	0.035	-	8
<i>Ots_S7-1</i>	2	0.461	0.016	0.072	0.136	0.049		2
<i>Ots_SL</i>	2	0.265	0.026	0.474	0.649	0.301	+	9
<i>Ots_SWS1op-182</i>	2	0.432	-0.033	0.139	0.247	0.079		8
<i>Ots_TAPBP</i>	2	0.233	0.064	0.420	0.552	0.219	+	2
<i>Ots_TCL1</i>	2	0.470	0.034	0.068	0.129	0.035	-	2
<i>Ots_Tnsf</i>	2	0.352	0.039	0.270	0.421	0.148		9
<i>Ots_u202-161</i>	2	0.329	0.024	0.321	0.482	0.190	+	8
<i>Ots_u211-85</i>	2	0.269	0.027	0.465	0.640	0.285	+	8
<i>Ots_u212-158</i>	2	0.159	0.004	0.021	0.026	0.022	-	8
<i>Ots_u4-92</i>	2	0.203	-0.021	0.202	0.255	0.094		8
<i>Ots_u6-75</i>	2	0.190	-0.019	0.087	0.108	0.040		8
<i>Ots_unk526</i>	2	0.237	-0.027	0.044	0.059	0.027		2
<i>Ots_ZNF330-181</i>	2	0.055	-0.030	0.076	0.080	0.040		8

'+' locus above the 99% confidence interval of F_{ST} distribution vs. heterozygosity (F_{DIST2}); '-' locus below the 99% confidence interval of F_{ST} distribution vs. heterozygosity (F_{DIST2}). In , measure of informativeness for assignment (Rosenberg *et al.* 2003). *Source for running conditions, primer and probe sequences: (1) Olsen *et al.* 1998, (2) unpublished, (3) Rexroad *et al.* 2001, (4) Greig *et al.* 2003, (5) Banks *et al.* 1999, (6) Williamson *et al.* 2002, (7) Cairney *et al.* 2000, (8) Smith *et al.* 2005a, and (9) Smith *et al.* 2005b.

Statistical analysis

Deviation from Hardy–Weinberg equilibrium was evaluated at each locus and population with the Markov chain Monte Carlo (MCMC) approximation of Fisher's exact test implemented in GENEPOP version 3.3 (1000 batches with 1000 iterations; Raymond & Rousset 1995). Tests for linkage disequilibrium between all pairs of loci were also performed using the MCMC approximation of the exact test in GENEPOP. Because multiple comparisons were involved, correction against type I error was made in both tests with the Bonferroni method (Rice 1989).

To estimate genetic diversity of each population sample, unbiased heterozygosity (H_E), observed heterozygosity (H_O), and total number alleles were estimated for all microsatellite and SNP loci in HP-RARE (Kalinowski 2005). Estimates of allelic richness (AR) using rarefaction were completed with HP-RARE using a sample size of 38 individuals.

To evaluate information content of the 50 markers included in this study (13 microsatellites and 37 SNPs), we estimated ' ln ', a measure of informativeness for assignment (Rosenberg *et al.* 2003). To estimate ln , a single allele is randomly selected from each locus for assignment to one population. As another measure of information content of each locus for differentiating populations, we calculated G'_{ST} following Hedrick (2005) to provide standardized estimates of genetic distance for each marker. Values of ln and G'_{ST} were utilized to compare information content among marker types and to rank all loci. The lowest ranked loci were systematically removed in repeated assignment tests to determine the effect of reducing the number of loci on assignment success. We also evaluated whether removing loci with low variation ($H_E < 0.05$) improved assignment results.

In order to estimate the ratio of SNPs to microsatellites, loci of each marker type were randomly resampled 100 times with replacement (www.random.org) for 1, 25, 50, 75, and 100% of the available markers (1, 3, 7, 10, and 13 microsatellites; 1, 9, 19, 28, and 37 SNPs). Ratio of mean ln for each random proportion was estimated and compared to ratio of mean ln for highest ranked markers at each quantile.

Individual assignment tests were performed with the Bayesian allele frequency estimation method (Rannala & Mountain 1997) in GENECLASS2 (Piry *et al.* 2004) with the leave-one-out option. Individuals were self-classified to the 29 reference populations with three different sets of genotypes: (i) 13 microsatellites only, (ii) 37 SNPs only, and (iii) all 50 markers. To evaluate how reducing loci influenced assignment accuracy, tests were repeated by incrementally decreasing the number of loci by five (removing the least informative loci as ranked by ln and G'_{ST} over all populations).

Pairwise values of the variance in allele frequencies among population samples (F_{ST} ; Weir & Cockerham 1984) were estimated for the microsatellites, SNPs, and all markers combined in GENEPOP. MCMC approximations of exact tests were performed in GENEPOP (1000 batches and 1000 iterations) to determine significance of pairwise genic differentiation between all collections. Significance levels were adjusted for multiple tests with two methods in order to compare results from each procedure: (i) standard Bonferroni correction (Rice 1989), and (ii) a modified version of the False Discovery Rate referred to as the BY-FDR (Benjamini & Yekutieli 2001) that is expected to provide a large increase in power to identify differentiated populations relative to the Bonferroni method (Narum 2006). In addition to pairwise F_{ST} for all loci, we estimated global F_{ST} values for each locus across all populations. Since loci with high heterozygosity (i.e. microsatellites) have low maximum F_{ST} values (Hedrick 1999), we also calculated standardized estimates of genetic distance for each marker (G'_{ST} ; Hedrick 2005) for all 50 markers using MICRO-SATELLITE ANALYSER version 4.05 (Dieringer & Schlötterer 2003).

In order to infer the degree of relatedness between sample collections, pairwise chord distances (Cavalli-Sforza & Edwards 1967) were calculated between all population samples with the POPULATIONS software package (Langella 2001). These genetic distances were then used to construct neighbour-joining trees of sample populations for 1000 bootstrap replicates. A consensus dendrogram was then constructed with the program TREEVIEW (Page 1996). Three dendrograms were constructed, one each with microsatellites and SNPs separately, and one with all markers combined.

In order to identify loci at which allele frequencies were likely affected by natural selection, the method of Beaumont & Nichols (1996) was used to identify 'outlier loci' from a plot of heterozygosity vs. F_{ST} . This was done using the program FDIST2 (<http://www.rubic.rdg.ac.uk/~mab/software.html>) with simulation to independently generate a distribution of F_{ST} , based on 50 000 replicates, for the microsatellites (stepwise mutation model) and the SNPs (infinite alleles model). The simulation results were then plotted to represent the median, and the 0.005 and 0.995 quantiles (between which 99% of the data points are expected to lie). Loci lying above or below these quantiles may be under directional or balancing selection, respectively, in some populations. The FDIST2 analysis was done iteratively to avoid an upward bias in quantiles (potentially resulting in type I error for balancing selection) by removing outlier loci above the 0.995 quantile in the initial run (Beaumont, University of Reading, Reading, UK, personal communication). In subsequent runs of FDIST2, three microsatellite loci and eight SNPs with significant deviations from neutral expectations (Table 1) were excluded from simulations to estimate the distribution of F_{ST} .

Table 2 Genetic diversity estimates and self-assignment accuracy for 29 populations of Chinook salmon with 13 microsatellite loci, 37 single nucleotide polymorphisms (SNPs), and all 50 markers combined for assignment

Population	<i>n</i>	Expected heterozygosity		Observed heterozygosity		Allelic richness		Self-assignment success (%)		
		Microsatellites	SNPs	Microsatellites	SNPs	Microsatellites	SNPs	Microsatellites	SNPs	Combined
King Salmon H.	44	0.742	0.251	0.726	0.243	7.4	1.83	100.0	95.5	100.0
Kowatua Cr.	144	0.817	0.274	0.811	0.272	12.1	1.81	43.8	45.1	54.2
Nakina R.	58	0.818	0.286	0.815	0.278	12.1	1.83	25.9	12.1	27.6
Upper Nahlin R.	81	0.810	0.279	0.808	0.277	11.9	1.83	44.4	28.4	51.9
Little Tatsamenie R. su	94	0.822	0.264	0.797	0.262	11.9	1.78	34.0	67.0	51.1
Andrew Medveje H.	76	0.826	0.310	0.823	0.302	12.6	1.86	46.1	30.3	44.7
Andrew Crystal H.	207	0.811	0.315	0.810	0.298	11.7	1.85	74.4	30.0	77.8
Cripple Cr.	144	0.840	0.322	0.835	0.307	13.2	1.87	62.5	47.2	81.3
Kateen R.	95	0.846	0.329	0.834	0.316	12.9	1.85	65.3	55.8	78.9
Little Kitsumkalum R.	91	0.834	0.303	0.841	0.282	13.2	1.83	53.8	74.7	83.5
Harrison R. f	94	0.845	0.237	0.857	0.217	13.7	1.77	87.2	78.7	96.8
Nanaimo H. f	95	0.841	0.262	0.836	0.254	11.4	1.78	92.6	76.8	100.0
Soos H.	93	0.814	0.267	0.836	0.274	12.0	1.73	93.5	91.4	100.0
Cle Elum H.	96	0.824	0.285	0.813	0.280	11.7	1.82	90.6	78.1	99.0
Forks Cr. H.	96	0.855	0.294	0.851	0.284	12.9	1.78	92.7	67.7	97.9
Cowlitz H. sp	96	0.865	0.244	0.856	0.238	13.1	1.75	92.7	89.6	100.0
Spring Cr. H. tule	96	0.832	0.167	0.829	0.170	12.3	1.65	88.5	92.7	99.0
Hanford Reach f	96	0.873	0.225	0.871	0.204	14.9	1.76	69.8	92.7	94.8
Lyons Ferry H. f	90	0.867	0.233	0.837	0.221	13.8	1.73	64.4	95.6	96.7
Imnaha R. sp	96	0.779	0.230	0.789	0.228	11.3	1.76	70.8	41.7	78.1
Minam R. sp	95	0.792	0.242	0.784	0.237	12.0	1.73	70.5	49.5	78.9
Rapid R. H. sp	94	0.759	0.231	0.769	0.225	10.2	1.74	80.9	42.6	80.9
Secesh R. sp	96	0.775	0.227	0.757	0.229	10.6	1.68	76.0	78.1	88.5
Johnson Cr. sp	96	0.772	0.227	0.767	0.247	10.4	1.69	87.5	100.0	99.0
Nestucca H. f	88	0.822	0.282	0.811	0.297	11.8	1.79	87.5	87.5	95.5
Umpqua H. sp	95	0.864	0.297	0.838	0.271	13.8	1.83	65.3	45.3	76.8
Elk H. f	93	0.837	0.262	0.825	0.255	11.5	1.71	88.2	73.1	96.8
Cole Rivers H. sp	91	0.847	0.264	0.837	0.254	12.3	1.82	73.6	70.3	82.4
Eel R. f	53	0.767	0.208	0.756	0.193	11.1	1.74	96.2	94.3	100.0

H., hatcheries; R., rivers; Cr., creeks; Where appropriate, adult migration timing follows population name with the following abbreviations: sp, spring; su, summer; f, fall.

Simulations to evaluate statistical power of microsatellites and SNPs to differentiate populations were completed with POWSIM (Ryman & Palm 2006). This analysis simulated multiple populations that have diverged to predefined 'true' levels of divergence, and determined the power of a data set with sample sizes, number of loci, and allele frequencies equal to the present study to differentiate populations.

Simulations included true levels of divergence (F_{ST}) that ranged from 0.0001 to 0.001, and significance for both Fisher's exact test and chi-squared tests from 1000 replicates were reported.

Results

Of 1421 tests for deviation from Hardy-Weinberg equilibrium (29 populations at 49 loci; *Ots_C3N3* excluded), 9 were statistically significant (all with heterozygote deficiencies) following Bonferroni correction (critical value = 0.00004). The nine significant heterozygote deficits were distributed between microsatellite loci, with two deviations at *OMM1080*

(Elk H. and Nestucca H.), two at *Ssa408* (Kowatua Cr. and Cle Elum H.), and one at *Ots212* (Lyons Ferry H.) and the SNP loci with two deviations at *Ots_u212-158* (King Salmon R. and Umpqua H.), and two at *Ots_MHC2* (Cripple Cr. and Rapid River H.). Of 35 525 pairwise tests for linkage disequilibrium, only 51 were significant and deviations were not consistently found either among specific loci pairs or within specific populations.

All microsatellite and SNP loci were polymorphic in at least 1 of 29 populations, with a minimum H_E of 0.001 for SNP locus *Ots_E9BAC* and a maximum H_E of 0.946 for microsatellite locus *OMM1080* (Table 1). The number of observed alleles for the microsatellite loci ranged from 9 (*Ots9*) to 53 (*OMM1080*), and 2 alleles were present for every SNP (Table 1). Trends in population genetic diversity differed depending upon the loci (Table 2), as microsatellites provided higher estimates of H_E and AR for large natural populations (i.e. Hanford Reach). In contrast, H_E and AR of the SNP markers were consistently higher in populations from northern latitudes than in those from California and the Columbia River.

Proportion	No. of markers		Randomly resampled		Highest ranked	
	Microsatellites	SNPs	Mean ratio	Median ratio	Mean ratio	Median ratio
1%	1	1	4.1	6.9	1.7	1.7
25%	3	9	4.1	5.5	2.2	2.3
50%	7	19	4.0	5.0	2.8	3.1
75%	10	28	3.9	5.0	3.4	3.7
100%	13	37	3.9	5.0	3.9	5.0

Table 3 Mean and median ratio of informativeness of single nucleotide polymorphisms (SNPs) to microsatellites from randomly resampled (100 iterations with replacement) and highest ranked (*In*) loci in Chinook salmon

Information content of each locus as determined with *In* revealed that microsatellites were generally more informative for assignment than SNPs (Table 1). The highest 12 markers ranked with *In* were microsatellites, with only one microsatellite (*Ots9*) ranking below any SNPs. Eight SNP markers ranked higher than *Ots9*, and the remaining 29 SNPs ranked below this microsatellite locus with few alleles (nine). However, when the 50 loci were ranked by G'_{ST} , the 6 highest values and 7 of the highest 10 values were for SNPs (Table 1). Loci in the highest 50% of G'_{ST} values included 13 SNPs and 12 microsatellites, with the majority of the lower 50% of G'_{ST} values comprised of SNPs (24 of 25), indicating greater variation in G'_{ST} for SNPs than for microsatellites. Within marker class, both measures of information content provided similar ranks for SNPs and were highly correlated ($r^2 = 0.96$), but were less similar for microsatellites ($r^2 = 0.85$). We also found that average informativeness of random microsatellites was approximately four times greater than randomly sampled SNPs (Table 3). However, the ratio of SNPs to microsatellites was consistently lower when the highest ranked markers (*In*) were included at each quantile (Table 3). For example, with 50% of available loci, the randomly resampled ratio of SNPs to microsatellites was equal to 4.0, but the ratio of ranked SNPs to ranked microsatellites was only 2.8.

Accuracy of average individual assignment to reference population was higher for the microsatellites (73.1%) than for the SNPs (66.6%), but success was highest when all markers were combined (83.2%). Results of multilocus assignment tests were displayed as a matrix of population assignment percentage that corresponds to genetic relationships determined from neighbour-joining dendrograms (Fig. 2a–c). Self-assignments to reference populations lie on a diagonal line with the highest assignment percentage expected to the reference population. Mis-assignments created a pattern of blocks that corresponded to closely related populations in the neighbour-joining dendrograms. Both microsatellites and SNPs frequently assigned 1–10% of individuals to incorrect populations (Fig. 2a, b), but mis-assignments were greatly reduced when all 50 markers were used in assignment tests (Fig. 2c). Examining mis-assignments by geographical region, we observed that

populations in the southeastern Alaska/northern British Columbia region consistently had the lowest assignment success with all three sets of loci (Fig. 2a–c).

Assignment tests were repeated with incremental decreases in the number of markers with the least informative loci removed first, as determined by *In* and G'_{ST} values. Assignment success remained high when 25–50 markers were utilized (range of 79.9–83.2%), and locus compositions were very similar when ranked with either *In* and G'_{ST} (Fig. 3). Results indicated that average assignment accuracy with the 25 most informative markers (*In* = 79.9%; G'_{ST} = 81.1%) was nearly equivalent to that for all 50 markers (83.2%; Fig. 3). When using only the 15 markers with highest G'_{ST} (8 SNPs plus 7 microsatellites), assignment accuracy (73.4%) was still higher than with either all microsatellites (73.1%) or all SNPs (66.6%) alone. Assignment accuracy with the 15 markers with highest *In* was slightly better at 75.1%, but composition of loci differed greatly with *In* vs. G'_{ST} (Fig. 3). Assignment accuracy decreased quickly with less than 15 markers with both ranking procedures, but reducing loci based on *In* rather than G'_{ST} appeared to be more effective for achieving high assignment success with less than 15 loci (Fig. 3).

Between marker classes, 9 of the most informative microsatellite loci ranked by *In* (Table 1) had nearly equal assignment success as all 37 SNPs (67.0% to 66.7%, respectively). In terms of alleles, SNPs with 37 independent alleles were equivalent to 312 independent microsatellite alleles (from the 9 most informative loci). While previous studies (e.g. Liu *et al.* 2005) have shown that inclusion of less informative loci may reduce inference of population structure, we observed equal average assignment success as the full panel of 37 SNPs (66.6%) when 1–4 SNP loci with H_E less than 0.05 were excluded from assignment tests. However, examination of assignment results of single populations showed that some had increased or decreased success when SNPs with low variation were removed. Removal of 1–2 SNPs (*Ots_E9BAC*, *Ots_arf-188*) had no effect, but removal of a third (*Ots_HGFA-446*) changed assignment success for some populations by a range of –1.3% to 4.5%, and removal of a fourth (*Ots_Ikaros-250*) had an effect that ranged from –4.2 to 4.5%. These results suggest that inclusion of markers with overall low variability may be

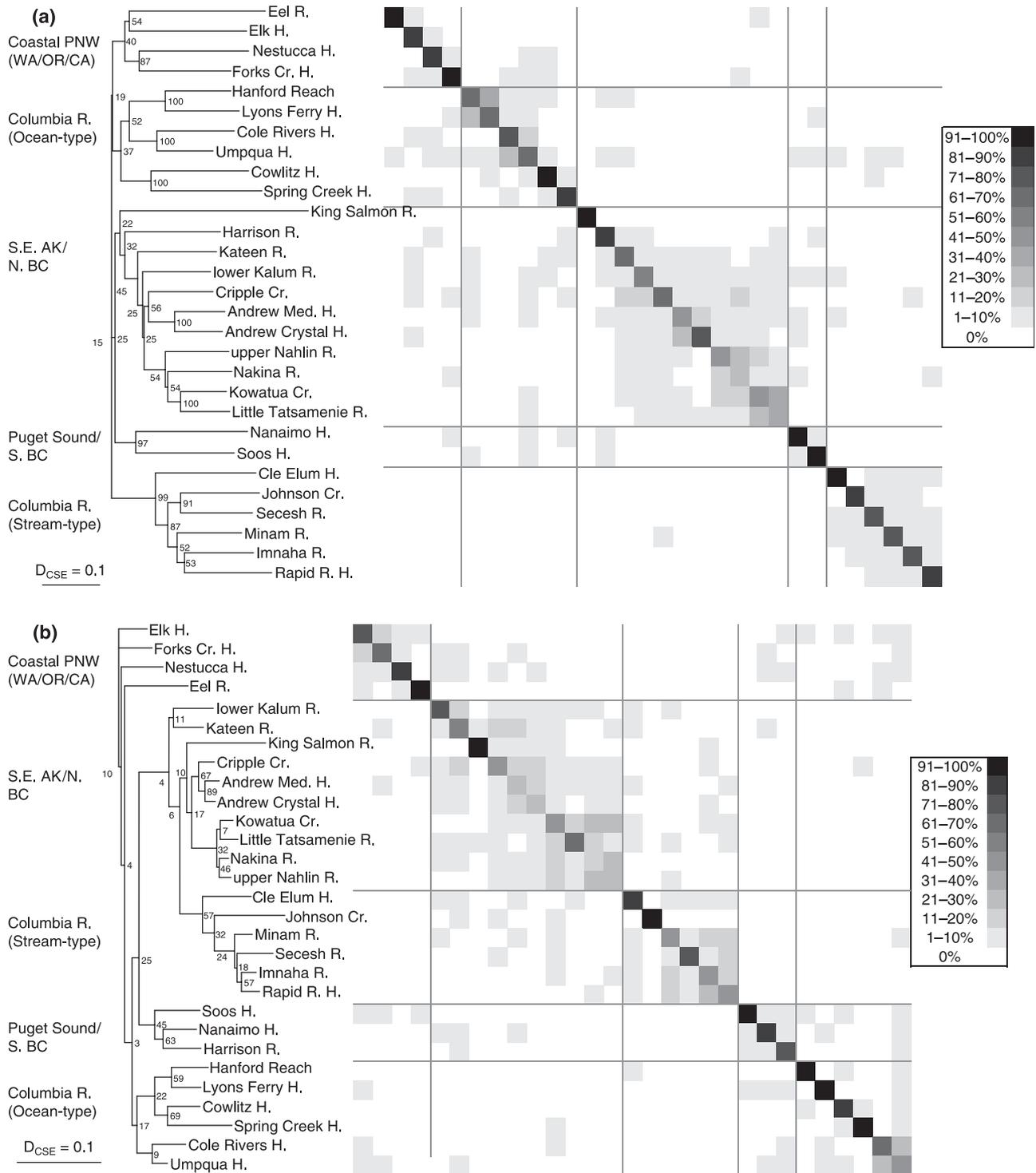


Fig. 2 Chord distance (D_{CSE}) neighbour-joining dendrograms and self-assignment matrices of populations of Chinook salmon from North America as determined with (a) 13 microsatellites, (b) 37 SNPs, and (c) all 50 markers combined. The diagonal represents the percentage of self-assigned individuals from a population and shaded blocks above and below the diagonal indicate percentage of mis-assignments to populations corresponding with the dendrogram. Grey grid lines correspond to regional clusters in the neighbour-joining dendrogram. Shading scale at the right of each figure depicts percentage assignment in 10% increments.

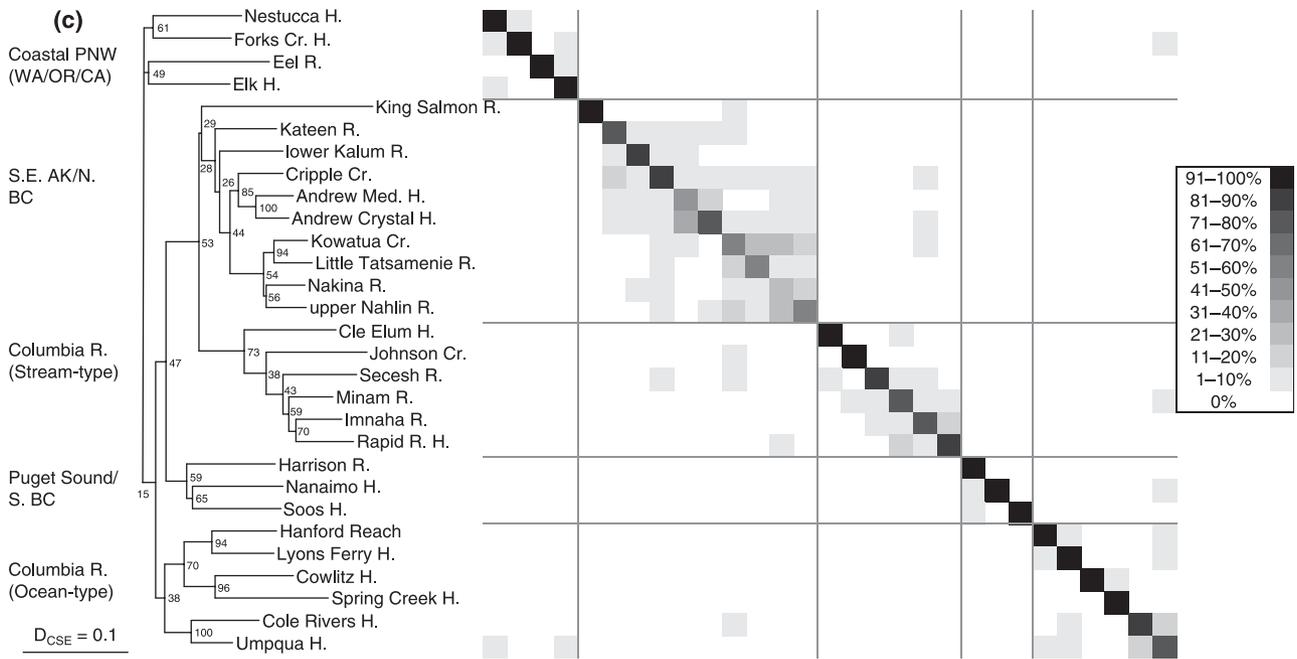


Fig. 2 Continued

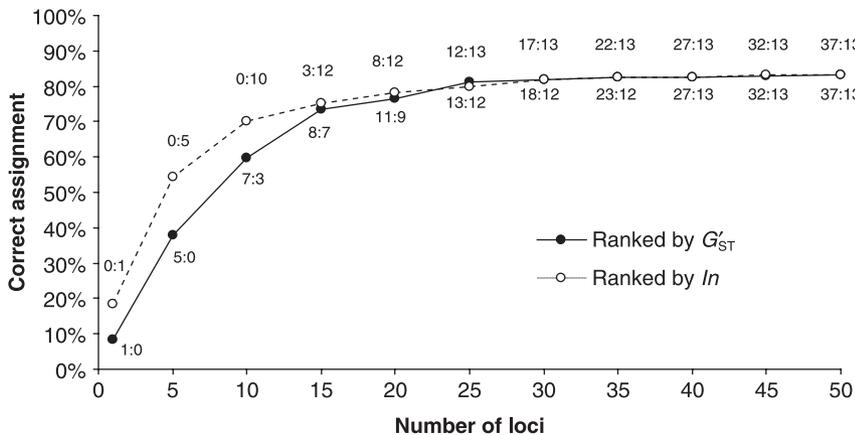


Fig. 3 Percentage of correct self-assignment with decreasing numbers of markers from a set of 37 SNPs and 13 microsatellites. Starting with all 50 loci, markers were eliminated incrementally based on lowest In or G'_{ST} values shown in Table 1. Number of SNPs: number of microsatellites is included above (In) and below (G'_{ST}) each point.

detrimental to assignment success for specific populations that are fixed at these loci, but may be beneficial to other populations that display slight variation.

Tests for population differentiation with microsatellite loci were highly significant among all 406 pairwise combinations at the BY-FDR corrected critical value ($P \leq 0.0076$), but one comparison (Kowatua Cr. vs. Little Tatsamenie R.) was not significant with Bonferroni correction ($P \leq 0.0001$). With SNPs, all but 3 of 406 tests were significant with BY-FDR correction (southeastern Alaska populations Kowatua Cr. vs. Nakina R., Nakina R. vs. Upper Nahlin R., and Crystal H. vs. Medvejie H.) and a fourth comparison was not significant with Bonferroni correction (Kowatua Cr. vs. Upper Nahlin R.). Differentiation with all 50 markers was significant at all but 1 of 406 pairwise tests (Nakina R. vs.

Upper Nahlin Cr.) for both multiple test procedures. For microsatellites, there was a strong positive correlation ($r^2 = 0.72$) between heterozygosity and ability to differentiate populations (Fig. 4). In contrast, for SNP loci, intermediate heterozygosities (~ 0.25) were best for pairwise population differentiation (Fig. 4). Global F_{ST} estimates across all populations indicated that SNP loci generally had higher divergence (average $F_{ST} = 0.192$) than microsatellite loci (average $F_{ST} = 0.067$; Table 1). However, maximum F_{ST} estimates for microsatellites were limited by higher heterozygosity (average $H_E = 0.821$) than for SNPs (average $H_E = 0.263$). Estimates of G'_{ST} (Hedrick 2005) provided a more comparable measure of divergence between marker types, and average G'_{ST} across populations was higher for microsatellites (0.423) than for SNPs (0.277).

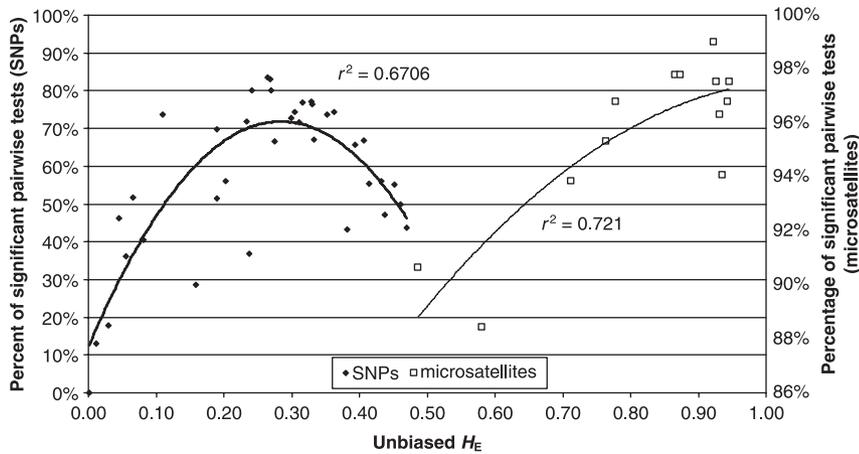


Fig. 4 Percentage of significant pairwise exact tests of allele frequency distributions of Chinook salmon populations versus heterozygosity for 13 microsatellite loci and 37 SNP markers.

The three neighbour-joining dendrograms created with microsatellites, SNPs, and all markers combined provided trees with similar clustering patterns (Fig. 2). Populations generally clustered by geographical region, with five major groups evident in each of the three dendrograms (southeastern Alaska/northern British Columbia, Puget Sound/southern British Columbia, Columbia River ocean-type, Columbia River stream-type and coastal Pacific Northwest). Slight differences in clustering patterns among trees were seen for one population (Harrison R.) that clustered with Puget Sound/southern British Columbia, its region of origin, for both SNPs and combined trees (Fig. 2b, c), but with the southeastern Alaska/northern British Columbia region with microsatellites (Fig. 2a). Even though the genetic relationships of populations were generally consistent among the three trees, bootstrap support for internal branches was consistently higher for the tree produced with microsatellites than with SNPs or with the combined markers. When branch relationships were identical among dendrograms, the microsatellite tree had higher bootstrap values than SNPs in all but one case, and higher values than the combined markers in nine cases.

Tests for outlier loci that may be under natural selection found 20 loci with values outside of their respective 99% confidence intervals for expected F_{ST} (Fig. 5a, b), 4 microsatellites (3 loci above and 1 below), and 16 SNP loci (8 loci below and 8 above). Most of the divergence at these outlier loci (see Table 1 for list) was due to differences between reproductively isolated populations of Columbia River salmon from the life history types known as ocean- and stream-type. Despite high divergence among populations, none of these loci had consistent deviations from Hardy-Weinberg equilibrium.

Simulations of statistical power of each marker type indicated that microsatellites were more powerful than SNPs for differentiating populations (Table 4). Pairwise differentiation based on microsatellites was significant in 100% of the 1000 replicates with both chi-squared and

Table 4 Simulations for statistical power of microsatellites and single nucleotide polymorphisms (SNPs) to differentiate populations at varying levels of true divergence (F_{ST}) with sample sizes, number of loci, and allele frequencies equal to the present study. Results are provided for both chi-squared and Fisher's exact tests for the proportion of simulations out of 1000 that were significant with a critical value of 0.05

True F_{ST}	Microsatellites		SNPs	
	Chi-squared	Fisher	Chi-squared	Fisher
0.0001	0.844	0.789	0.221	0.271
0.0002	1.000	1.000	0.515	0.538
0.0003	1.000	1.000	0.785	0.791
0.0004	1.000	1.000	0.949	0.944
0.0005	1.000	1.000	0.984	0.980
0.0006	1.000	1.000	0.998	0.998
0.0007	1.000	1.000	1.000	0.999
0.0008	1.000	1.000	1.000	1.000

Fisher's exact tests at a true F_{ST} as low as 0.0002. At the same level of divergence of 0.0002, pairwise differentiation based on SNPs was only significant in slightly more than half of the 1000 replicates (chi-squared = 51.5% and Fisher's exact test = 53.8%). The suite of SNPs were significant for a high proportion of tests (> 94%) with true divergence down to 0.0004, but were not significant 100% of the time for both tests until F_{ST} was equal to 0.0008.

Discussion

Both microsatellite and SNP markers were useful for differentiating most population samples in this study (> 99% of pairwise tests), but closely related populations were generally better distinguished with microsatellites than with SNPs. As confirmed by simulations, this is a result of the additional power provided by the greater total number of independent alleles for the microsatellites than

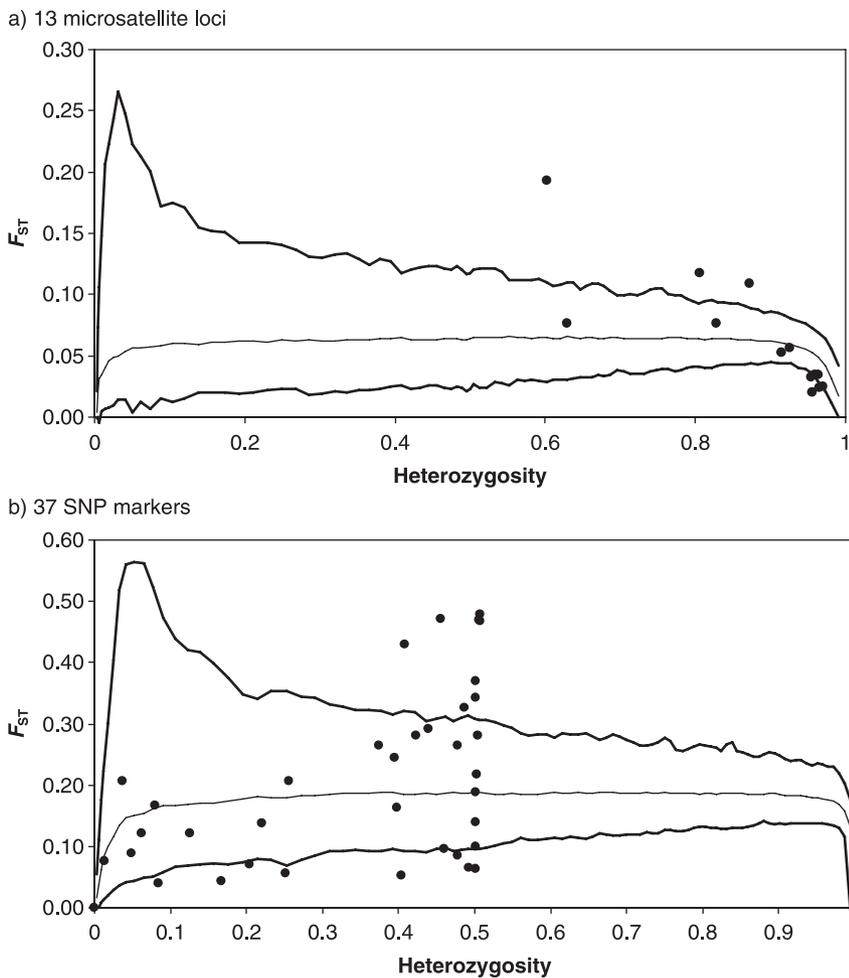


Fig. 5 F_{ST} values plotted against heterozygosity (F_{DIST2} ; Beaumont & Nichols 1996) for (a) 13 microsatellite loci (stepwise mutation model), and (b) 37 SNP markers (infinite alleles model). The dashed line represents the median and solid lines represent the 99% confidence interval boundaries. See list in Table 1 for loci outside these boundaries ('under selection').

for the SNPs (415 and 37, respectively) used to distinguish populations (Kalinowski 2004). However, there were four populations where the SNP markers had at least 20% higher assignment accuracy than microsatellites (Kowatua Cr., Little Tatsamenie R., Hanford Reach, and Lyons Ferry H.; Table 2). In these cases, assignment accuracy was low for microsatellites (< 70%) due to high levels of misassignment to neighbouring populations. This may be a reflection of SNP loci in these specific populations that have relatively divergent allele frequencies, due to natural selection or ascertainment bias, and that provided additional power to differentiate populations with high gene flow between them.

Genetic relationships among populations on bootstrap consensus trees were consistent with all three sets of markers, demonstrating that reliable population structure can be determined with microsatellites, SNPs, or a combination of the two types of loci. Furthermore, the dendrograms were consistent with population structure of the species as determined in previous studies (e.g. Waples *et al.* 2004; Beacham *et al.* 2006). The addition of a mitochondrial SNP (*Ots_C3N3*) may have improved resolution of the dendro-

gram, as previous studies have demonstrated that combining mitochondrial and nuclear markers is desirable for determining phylogenetic relationships (e.g. Gaines *et al.* 2005; Fouquet *et al.* 2007). It was also apparent from Fig. 3 that using only the most informative loci, without consideration of marker type, will allow accurate differentiation of populations without employing all available markers.

Both theoretical (e.g. Kuhner *et al.* 2000) and empirical (e.g. Clark *et al.* 2005; Smith *et al.* 2007) studies have demonstrated the bias in estimates of population genetic diversity from SNP markers as a function of ascertainment bias. In our study, SNPs that were primarily ascertained from the northern portion of the species range (Smith *et al.* 2005a, b) were more diverse and informative in Chinook salmon populations from Alaska and British Columbia than markers ascertained from southern regions. Since most of the SNP markers in this study were ascertained in the northern portion of the range, this created a bias in overall population diversity estimates with the SNPs, whereas H_E and AR estimates with the microsatellites were more consistent with patterns of genetic diversity observed in previous

studies of the species (e.g. Waples *et al.* 2004; Beacham *et al.* 2006). While problematic for genetic diversity estimates and any applications assuming a random genomic sample, ascertainment bias of SNPs can be exploited to provide additional power for population assignment and differentiation in the geographical region or phylogenetic lineage of ascertainment (Luikart *et al.* 2003; Morin *et al.* 2004).

Several markers with outlier F_{ST} values (Fig. 5) showed evidence for either directional or balancing selection, but contemporary selection coefficients were likely slight within and among populations included in this study. More than half of the outlier loci (11 of 20) were markers that were distinctive for ocean- and stream-type life history types of Chinook salmon (i.e. *Ots_SL*, Ford 1998; *OtsG474*, Narum *et al.* 2004). The large F_{ST} values at these loci most likely reflect palaeogeological isolation (Waples *et al.* 2004) and, therefore, ancient rather than contemporary selection among life history types. Nine outlier loci had patterns of diversity consistent with the operation of balancing selection (Table 1), which is difficult to detect by comparing F_{ST} and heterozygosity even when selection coefficients are 20 times greater than migration rate (Beaumont & Balding 2004). This suggests that the application of tests for balancing selection in this study are either not highly reliable (despite iterative analysis to reduce upward bias), or selection coefficients were very large. Populations of Chinook salmon demonstrate a pattern of isolation by distance rather than a symmetrical island model assumed with *FDIST2*, but this method has been effective for evaluating adaptive loci in this species (e.g. Heath *et al.* 2006) and this migration model is not expected to greatly affect our results (see Beaumont & Nichols 1996) nor would we expect different results with a Bayesian approach (Beaumont & Balding 2004). Balancing selection should result in an excess of heterozygotes within populations, but no loci had a significantly higher than expected proportion of heterozygotes in any population in this study. Furthermore, one SNP locus (i.e. *Ots_MHCII*) in this study has been demonstrated to be under balancing selection in some Chinook salmon populations (Miller *et al.* 1997; Heath *et al.* 2006), yet this locus did not have significant results for deviation from neutral expectations or Hardy–Weinberg equilibrium. Thus, it is likely that contemporary selection coefficients at these loci are small in these Chinook salmon populations relative to migration and genetic drift, and our results may be a circumstance of some loci overly influencing the mean distribution of F_{ST} . However, the signature of historical balancing selection on genetic markers in salmonid populations can persist for a long time, even in the absence of contemporary selection (e.g. Aguilar & Garza 2006), so it is also possible that the positive tests for reduced divergence at least partially reflect ancient selective pressures.

Although lack of conformance to neutral expectations for loci in this study would violate assumptions of several

statistical tests employed, none of the populations had more than a single locus that deviated from the expectations of Hardy–Weinberg equilibrium. Recent studies that remove loci believed to be under selection from multilocus data sets have demonstrated that these patterns are unlikely to alter inference about population structure (Heath *et al.* 2006; Smith *et al.* 2007). However, if consistent deviations from Hardy–Weinberg equilibrium were observed at a locus or in a population, this may introduce bias to some statistical tests.

This study empirically demonstrates the utility of SNP markers for population genetic studies of salmon, but these markers are not yet available for many other nonmodel species. While SNPs are abundant in the genome of most organisms (occurring every ~200–500 bp; Morin *et al.* 2004), validating and designing assays for these markers can be significantly more complicated and expensive than for microsatellites, due to the large amount of sequencing necessary to identify and validate SNPs. In addition, ascertainment bias is a concern for both types of markers when ascertainment involves small sample sizes and/or a fraction of the species' geographical range or phylogenetic diversity are screened for polymorphism (reviewed in Luikart *et al.* 2003). However, highly polymorphic microsatellites are less likely to suffer from strong ascertainment bias than bi-allelic SNPs because polymorphism is more likely to be widespread in multi-allelic microsatellites. Identification and validation of SNPs can be further complicated in organisms with ancestrally duplicated genomes (e.g. salmonids and sturgeon) and those with incomplete genome sequence (e.g. Smith *et al.* 2005a). Conversion of validated SNPs into functional genotyping assays can also be time consuming, and many polymorphisms may lie in genomic regions incompatible with specific laboratory methodology (i.e. 5' exonuclease assay) resulting in reduced numbers of usable loci. Despite the additional considerations necessary when developing SNP markers, their potential advantages, such as reduced genotyping error rate, high-throughput potential, lower assay costs, and high genomic density likely justify the development of these loci in many nonmodel organisms.

In the present study, we found mean ratios of random SNPs to microsatellites in Chinook salmon (from 3.9 to 4.1) to be within the range of ratios shown in humans (e.g. Rosenberg *et al.* 2003; Liu *et al.* 2005). While evaluation of marker types in this study was limited to a specific suite of microsatellites and SNPs, we found that selecting markers based on information content is likely to reduce the ratio of SNPs to microsatellites. Extensive simulations have been completed in humans with hundreds of microsatellites and thousands of SNPs (Rosenberg *et al.* 2003; Liu *et al.* 2005). Rosenberg *et al.* (2003) resampled subsets of each marker type from 377 microsatellites and 8714 SNPs and found random microsatellites to be more informative than random

SNPs with mean ratios of 2.8 to 4.3 for humans among African, European, and East Asian continental groups. Similarly in tests of two races of humans (black non-Hispanic and white), Liu *et al.* (2005) determined that mean informativeness of random microsatellites is 2.5 to 6.3 times that of random SNPs from a set of 328 microsatellites and 15 840 SNPs. However, these two studies in humans found contradictory results regarding which marker type constituted the majority of the most informative markers, with Liu *et al.* (2005) finding SNPs in the majority, and Rosenberg *et al.* (2003) found microsatellites to be predominant. These studies evaluate the informativeness of both SNPs and microsatellites from tremendous numbers of loci, but they were limited by few populations, unbalanced sample sizes, and that different populations were analysed with each marker type. The present study includes many fewer loci from a nonmodel organism, but marker types were evaluated for differentiation of several identical populations at broad and fine geographical scale, as well as among distinct life history types (i.e. races).

In conclusion, both the microsatellite and SNP markers evaluated in this study were highly effective for determining genetic relationships and differentiating populations. Both types of markers have positive and negative attributes and when both are available for a study organism, the choice between them will best be determined by the hypotheses being tested, as well as laboratory-specific costs, and technical capabilities. However, given the advances in human genetics using high density maps with > 25 000 SNPs, and the unprecedented insight into population structure, natural selection and disease (e.g. Sabeti *et al.* 2006) that has resulted from their use, SNPs are likely to become the marker of choice for many studies of nonmodel organisms as more genomic sequence becomes available. In the meantime, combining the most informative markers available is likely to be the most powerful approach for differentiating populations for ecological and conservation genetics.

Acknowledgements

We are grateful for laboratory contributions from Vanessa Jacobson. Figure 1 was provided by David Graves. Funding for this research came from the Chinook Technical Committee of the Pacific Salmon Commission to the Genetic Analysis of Pacific Salmonids consortium.

References

- Aguilar A, Garza JC (2006) A comparison of variability and population structure for major histocompatibility complex and microsatellite loci in California coastal steelhead (*Oncorhynchus mykiss* Walbaum). *Molecular Ecology*, **15**, 923–937.
- Banks MA, Blouin MS, Baldwin BA *et al.* (1999) Isolation and inheritance of novel microsatellites in Chinook salmon. *Journal of Heredity*, **90**, 281–288.
- Beacham TD, Jonsen KL, Supernault J, Wetklo M, Deng L, Varnavskaya N (2006) Pacific Rim population structure of Chinook salmon as determined from microsatellite analysis. *Transactions of the American Fisheries Society*, **135**, 1604–1621.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society London. Series B, Biological Sciences*, **263**, 1619–1626.
- Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Annals of Statistics*, **29**, 1165–1188.
- Bensch S, Åkesson S, Irwin DE (2002) The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology*, **11**, 2359–2366.
- Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology*, **16**, 907–924.
- Cairney M, Taggart JB, Hoyheim B (2000) Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. *Molecular Ecology*, **9**, 2175–2178.
- Cavalli-Sforza LL, Edwards WWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **32**, 550–570.
- Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, **15**, 1496–1502.
- Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences, USA*, **91**, 3166–3170.
- Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.
- Estoup A, Tailliez C, Cornuet JM, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Molecular Biology and Evolution*, **12**, 1074–1084.
- Ford MJ (1998) Testing models of migration and isolation among populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolution*, **52**, 539–557.
- Fouquet A, Vences M, Salducci MD *et al.* (2007) Revealing cryptic diversity using molecular phylogenetics and phylogeography in frogs of the *Scinax ruber* and *Rhinella margaritifera* species groups. *Molecular Phylogenetics and Evolution*, **43**, 567–582.
- Gaines CA, Hare MP, Beck SE, Rosenbaum HC (2005) Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 533–542.
- Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Research*, **6**, 211–217.
- Greig C, Jacobson DP, Banks MA (2003) New tetranucleotide microsatellites for fine-scale discrimination among endangered Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology Notes*, **3**, 376–379.
- Heath DD, Shrimpton JM, Hepburn RI, Jamieson SK, Brode SK, Docker MF (2006) Population structure and divergence using microsatellite and gene locus markers in Chinook salmon (*Oncorhynchus tshawytscha*) populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 1370–1383.
- Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

- Hoffman JL, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, **14**, 599–612.
- Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? *Heredity*, **88**, 62–65.
- Kalinowski ST (2004) Genetic polymorphism and mixed stock fisheries analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 1075–1082.
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.
- Kuhner MK, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics*, **156**, 439–447.
- Langella O (2001) Populations 1.2.24: population genetic structure (individuals or populations distances, phylogenetic trees). Available: <http://bioinformatics.org/~tryphon/populations/>.
- Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, **6**(S1), 1471–2156.
- 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.
- Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, **160**, 357–366.
- Miller KM, Withler RE, Beacham TD (1997) Molecular evolution at MHC genes in two populations of Chinook salmon *Oncorhynchus tshawytscha*. *Molecular Ecology*, **6**, 937–954.
- Morin PA, Luikart G, Wayne RK, SNP Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.
- Narum SR (2006) Beyond Bonferroni: Less conservative analyses for observation genetics. *Conservation Genetics*, **7**, 783–787.
- Narum SR, Powell MS, Talbot AJ (2004) A distinctive microsatellite locus that differentiates ocean-type from stream-type Chinook salmon in the interior Columbia River Basin. *Transactions of the American Fisheries Society*, **133**, 1051–1055.
- Olsen JB, Bentzen P, Seeb JE (1998) Characterization of seven microsatellite loci derived from pink salmon. *Molecular Ecology*, **7**, 1083–1090.
- Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357–358.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences, USA*, **94**, 9197–9221.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Rexroad CE III, Coleman RL, Martin AM, Hershberger WK, Killefer J (2001) Thirty-five polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics*, **32**, 317–319.
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.
- Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics*, **73**, 1402–1422.
- Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600–602.
- Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of the efficiency of single nucleotide polymorphisms (SNPs) and microsatellites for the estimates of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *Journal of Heredity*, **98**, 692–704.
- Sabeti PC, Schaffner SF, Fry B *et al.* (2006) Positive natural selection in the human lineage. *Science*, **312**, 1614–1620.
- 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, Antonovich A, Banks MA *et al.* (2007) Development of a standardized DNA database for Chinook salmon. *Fisheries*, **32**, 540–552.
- Smith CT, Elfstrom CM, Seeb JE, Seeb LW (2005a) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology*, **14**, 4193–4203.
- Smith CT, Seeb JE, Schwenke P, Seeb LW (2005b) Use of the 5'-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. *Transactions of the American Fisheries Society*, **134**, 207–217.
- Smith CT, Antonovich A, Elfstrom CM, Narum SR, Seeb LW (2007) Impacts of marker class bias relative to locus-specific variability on population inferences in Chinook salmon; a comparison of SNPs to STRs and allozymes. *Transactions of the American Fisheries Society*, **136**, 1674–1687.
- Waples RS, Teel DJ, Myers JM, Marshall AR (2004) Life-history divergence in Chinook salmon: historical contingency and parallel evolution. *Evolution*, **58**, 386–403.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.
- Williamson KS, Cordes JF, May B (2002) Characterization of microsatellite loci in Chinook salmon (*Oncorhynchus tshawytscha*) and cross-species amplification in other salmonids. *Molecular Ecology Notes*, **2**, 17–19.

All contributing authors are actively involved in molecular genetics research of Pacific salmon. Authors are participants in a consortium, Genetic Analysis of Pacific Salmonids (GAPS), which was created with the intent of developing collaborative science to address population structure and genetic stock identification.
