

## Impacts of Marker Class Bias Relative to Locus-Specific Variability on Population Inferences in Chinook Salmon: A Comparison of Single-Nucleotide Polymorphisms with Short Tandem Repeats and Allozymes

CHRISTIAN T. SMITH,\*<sup>1</sup> ANTON ANTONOVICH, WILLIAM D. TEMPLIN, AND CARITA M. ELFSTROM  
*Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory,  
333 Raspberry Road, Anchorage, Alaska 99518, USA*

SHAWN R. NARUM  
*Columbia River Inter-Tribal Fish Commission, 3059-F National Fish Hatchery Road,  
Hagerman, Idaho 83332, USA*

LISA W. SEEB  
*Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory,  
333 Raspberry Road, Anchorage, Alaska 99518, USA*

**Abstract.**—Single-nucleotide polymorphisms (SNPs) exhibit several attributes that make them appealing as a class of genetic markers for applications in ecology and evolution. Two commonly cited limitations of SNPs in this capacity are that ascertainment bias and natural selection may shape allele frequencies of these markers, thus biasing estimates of population structure. The impacts of ascertainment bias and selection on estimates of population parameters have been demonstrated in a few model species, but their impacts relative to locus-specific variability and other potential complications on structure inferences in wild populations are unclear. We examined 22 allozymes, 9 short tandem repeats (STRs), and 41 SNPs in approximately 1,300 Chinook salmon *Oncorhynchus tshawytscha* representing 16 collections. We used plots of the genetic differentiation index  $F_{ST}$  versus heterozygosity and sequence criteria to identify SNPs that might be under natural selection. We then calculated several measures of population structure based on the three marker sets and a subset of the SNPs from which loci identified as likely targets of natural selection had been removed. Correlation of genetic distances between collections was stronger between allozymes and SNPs than between either of these markers and STRs, suggesting that the influences of marker class bias (e.g., selection and ascertainment bias) were smaller than impacts of locus-specific effects. Divergence estimates between SNP ascertainment populations were not significantly higher when based on SNPs than when based on other markers. Overall divergence ( $F_{ST}$ ) was higher for SNPs than for allozymes; however, the choice of  $F_{ST}$  estimator influenced the relative values for STRs and SNPs. Estimates of within-population diversity based on allozymes and STRs correlated better with each other than with estimates based on SNPs; such estimates based on SNPs were relatively low for collections from populations outside the geographic coverage of the SNP ascertainment sample.

Population genetic analyses have increased our understanding of a broad range of evolutionary and ecological processes. Expanding numbers and classes of available genetic markers have afforded more flexibility in choosing tools for use in any given study (Schlötterer 2004). One class of genetic markers that has received recent attention for its potential in ecological and evolutionary studies is single-nucleotide polymorphisms (SNPs; Brumfield et al. 2003; Morin

et al. 2004). Several attributes of SNPs (including density in the genome, relatively well-understood evolutionary properties, and small number of alleles per locus) have led to the use of these markers for mapping and association studies in model genetic species (e.g., Kruglyak 1997). Consequently, several high-throughput SNP genotyping platforms and chemistries were developed over the past decade (reviewed by Kwok 2003). Despite optimism regarding the potential applications of SNPs in wild populations of nonmodel organisms, it is possible that SNP markers are detrimentally affected by ascertainment bias (Brumfield et al. 2003; Morin et al. 2004) and natural selection of alleles (Luikart et al. 2003; Heath et al. 2006).

Wright (1951) identified four forces that shape allele frequencies at a locus as selection, mutation, drift, and

\* Corresponding author: christian\_smith@fws.gov

<sup>1</sup> Present address: U.S. Fish and Wildlife Service, Abernathy Fish Technology Center, 1440 Abernathy Creek Road, Longview, Washington 98632, USA.

Received October 9, 2006; accepted May 31, 2007  
Published online November 29, 2007

migration ( $N_e m$ ). If one assumes that the influences of selection and mutation are very small relative to those of drift and migration, then allele frequencies can be explained in terms of balance between drift and migration. Population genetic analyses using allozymes, short tandem repeats (STRs), and other marker classes often make several assumptions, including negligible selection and mutation. Because SNP markers in nonmodel organisms are often developed based on expressed sequence tags or sequences of genes with known functions, neutrality assumptions for these may be more questionable than for markers developed by sequencing random shotgun clones from a genomic library. Measurement and modeling of selection are extremely complicated and impractical in many cases, so one technique that researchers use is to identify outlier loci and remove them prior to population analyses (Beaumont and Nichols 1996). However, the question remains as to how much influence selected loci will have on analyses of wild populations of an organism, especially if large numbers of neutral loci are also included in the analyses. Luikart et al. (2003) summarized results from studies of several species and found that inclusion of a small proportion of selected loci did not, in most cases, influence biological inferences or management recommendations. Parallel analyses of allozymes and STRs have indicated that the two marker classes reveal highly concordant population structures in sockeye salmon *Oncorhynchus nerka* (Allendorf and Seeb 2000) and chum salmon *O. keta* (Scribner et al. 1998). Divergence estimates between populations will be influenced by the evolutionary properties and locus-specific history of any genetic marker; however, the studies described above lead us to expect that analyses of reasonable numbers of either allozymes or STRs will lead to accurate inference of population structure.

Ascertainment bias is the effect of using genetic markers that are chosen or developed in a way that makes them nonrepresentative of genetic variation in the population(s) of interest to infer various aspects of population structure. In developing SNP markers, for example, researchers generally examine DNA sequences of several loci in a small number of individuals. This group of individuals is sometimes referred to as the "ascertainment panel" or "panel." After sequence analysis, the researcher will develop assays that genotype a nonrandom portion of the SNPs observed in the panel. Finally, the assays are applied to a larger and broader collection of individuals drawn from one or more populations. It is obvious from this process that (1) the number of individuals in the panel, (2) the selection of those individuals from the larger population, and (3) the choice of which observed SNPs are

targeted for assay development will potentially bias the amount and pattern of diversity observed in the broader population study.

Because the probability of SNP detection is proportional to the frequency of the SNP's minor allele in the panel, one effect of ascertainment bias is that low-frequency SNPs will be missed (not assayed; Nielsen and Signorovitch 2003). Polymorphisms that exist at low frequencies in the population(s) from which the panel was drawn will be missed, as will polymorphisms that exist exclusively outside the panel populations. Failure to detect or account for these polymorphisms will result in an underestimation of interpopulation divergence (and thus overestimation of migration), which should be increasingly pronounced in populations with increasing phylogenetic distance from the panel population(s).

While it seems likely that ascertainment bias and selection will influence population inferences in non-model species, it is unknown whether these sources of bias will be notably more pronounced in SNPs than other markers. Further, the influence of these sources of bias relative to marker-specific variation present in all classes of genetic markers is unclear. Assessment of these issues will require side-by-side comparisons of population inferences from SNPs and other markers on the same samples (Brumfield et al. 2003). By comparing data for allozymes, STRs, and SNPs in a common set of populations, it should be possible to test predictions that natural selection and ascertainment bias acting on a set of SNPs will lead to inference of a different structure than that inferred using allozymes or STRs. Here, we examine how patterns and levels of population divergence in Chinook salmon *O. tshawytscha* collections differ both broadly and with respect to panel populations when based on allozymes, STRs, and SNPs.

## Methods

Genomic DNA was extracted from 1,324 Chinook salmon representing 16 collections taken from a broad section of the species' range (Figure 1) using a DNeasy 96 Tissue Kit (QIAGEN, Valencia, California). The DNA eluted from the columns (1  $\mu$ L for each STR; 0.1  $\mu$ L for each SNP) was used as template for polymerase chain reaction (PCR) amplification.

Allozyme analysis for collections 1–13 was described by Teel et al. (1999). We did not have allozyme tissue samples from collections 14–16, so we substituted published allele frequencies for geographically proximal collections (Waples et al. 2004; available: [www.nwfsc.noaa.gov/publications/](http://www.nwfsc.noaa.gov/publications/)).

Nine STR loci were amplified in three PCR multiplexes (Table A.1): (1) primers *One7* (0.15  $\mu$ M), *Ots1* (0.15  $\mu$ M), and *Ots2* (0.04  $\mu$ M); (2) primers

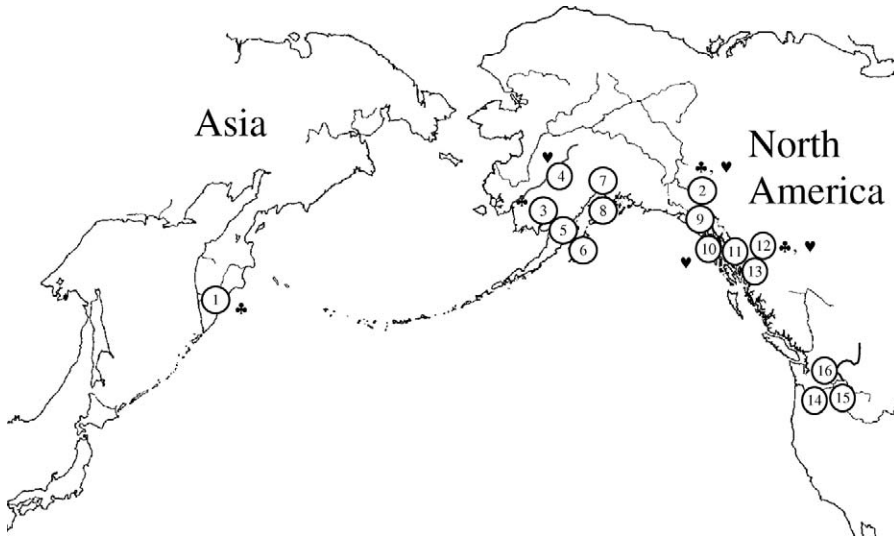


FIGURE 1.—Chinook salmon sample locations (sample sizes for short tandem repeats [STRs] and single-nucleotide polymorphisms [SNPs] in parentheses after each location): (1) Bistraya River, Russia (96, 96); (2) Stoney River, Alaska (91, 96); (3) Togiak River, Alaska (96, 96); (4) Kuskokwim River, Alaska (93, 95); (5) Nushagak River, Alaska (96, 96); (6) Ayakulik River, Kodiak Island, Alaska (90, 94); (7) Moose Creek, Alaska (50, 51); (8) Kenai River, Alaska (91, 96); (9) Tahini River, Alaska (52, 68); (10) King Salmon River, Alaska (90, 96); (11) Andrew Creek, Alaska (94, 96); (12) Unuk River, Alaska (88, 96); (13) Chickamin River, Alaska (50, 56); (14) Deschutes River, Oregon (92, 96); (15) Methow River, Washington (36, 50); and (16) Johnson Creek, Washington (30, 46). Clubs (♣) denote panel collections for 10 loci known a priori to be polymorphic in Chinook salmon (i.e., collections 1–3, and 12), and hearts (♥) denote panel collections for the remaining SNPs (i.e., collections 2, 4, 10, and 12).

*One102* (0.10  $\mu$ M), *Ots107* (0.06  $\mu$ M), and *uSat73* (0.06  $\mu$ M); and (3) primers *One13* (1.20  $\mu$ M), *One9* (0.04  $\mu$ M), and *Ots100* (0.25  $\mu$ M). Short tandem repeat amplification was carried out in 10- $\mu$ L reaction volumes consisting of 10 mM of tris-HCl, 50 mM of KCl, 0.2 mM of each deoxynucleotide triphosphate, and 0.5 units of *Taq* DNA polymerase (enzyme number 2.7.7.7, IUBMB 1992; Promega, Madison, Wisconsin) using an MJ research PTC-225 thermal cycler. The  $MgCl_2$  concentration was 2.0 mM in multiplexes 1 and 2 and 2.5 mM in multiplex 3. Reactions proceeded as follows: an initial denaturation of 92°C for 2 min; seven cycles of 92°C for 1 min, annealing temperature (multiplex 1 = 53°C; multiplex 2 = 54°C; multiplex 3 = 60°C) for 2 min, and 72°C for 20 s; and 20 cycles of 92°C for 30 s, annealing temperature for 2 min, and 72°C for 20 s. All thermal cycling was conducted at a ramp speed of 1°C/s. Amplification products were size fractionated on 5% denaturing polyacrylamide gels using an ABI377 DNA sequencer (Applied Biosystems, Foster City, California). Alleles were identified and sized using the internal lane sizing standard and local Southern sizing algorithm in GeneMapper (Applied Biosystems).

To identify novel SNP markers, we used the methods and ascertainment panels described by Smith

et al. (2005a; ascertainment collections are denoted by hearts in Figure 1).

Forty-one SNP genotyping assays (Table A.1) were applied to all samples. Genotyping assays were run in 384-well reaction plates; four wells in each plate served as negative (no-template) controls. Reactions consisted of 5  $\mu$ L of 1 $\times$  TaqMan PCR Mastermix (Applied Biosystems), 900 nM of each PCR primer, and 200 nM of each probe (probes for novel SNP assays are listed in Table 1). Thermal cycling was performed on an AB9700 thermal cycler using an initial denaturation of 10 min at 95°C followed by 50 cycles of 92°C for 15 s and 60°C for 1 min. All cycling was conducted at a ramp speed of 1°C/s. After amplification, end point reads of all plates were performed on an AB7900 real-time sequence detection system. Scoring of individual genotypes was performed using Sequence Detection Software (Applied Biosystems) to generate scatterplots that graphically depicted the amount of each allele-specific probe that bound to the PCR product of each individual.

*Departures from Hardy–Weinberg equilibrium and genotypic equilibrium.*—We used Genepop (Raymond and Rousset 1997) to perform exact tests (Guo and Thompson 1992) for genotypic ratios that departed from Hardy–Weinberg equilibrium (HWE) expecta-

TABLE 1.—Details for 10 novel Chinook salmon single-nucleotide polymorphism genotyping assays (F = forward; R = reverse). Each oligonucleotide probe was labeled on its 5' end with either VIC (Applied Biosystems) or 6-carboxyfluorescence; (6-FAM) and bore a minor groove binder and a nonfluorescent quencher on its 3' end.

Locus name, marker name, and GenBank accession number	Oligonucleotide sequences (5'–3')
Phenylalanine-tRNA synthetase-like gene, alpha subunit <i>Ots_FARSLA-220</i> (DQ908919)*	F: GTTCGTGGGATTGTTCAATGTTTCAT R: CTTGGACAGGCTCACATTACCATA VIC-CCTTGGATGGGATGTG FAM-CCTTGGATAGGATGTG
Pi-class glutathione S-transferase gene <i>Ots_GST-207*</i> (DQ908920)	F: GGAGAACATGCATCACCATTCAAG R: TCAGCAAACGAAGGCTATGTAGAAAT VIC-ATGAGAGAGTCTTTCTCTGTT FAM-ATGAGAGAGTCTTTTCTGTT
Pi-class glutathione S-transferase gene <i>Ots_GST-375*</i> (DQ908920)	F: CAGCCCGTCCAAAAATCAAG R: CAGGAATATCACTGTTTGCCATTGC VIC-TTTCTGTAGGCGTCAGAG FAM-TCTTGTAGGCATCAGAG
Heat-shock 90-kDa protein gene <i>Ots_HSP90B-100*</i> (DQ908921)	F: CACCTTAGTTCACGCAACATG R: CTGCGTGTATTGTAGTGGTGACA VIC-TCTATGGTGTGATTCATT FAM-TTCTATGGTGTAAATTCATT
Heat-shock 90-kDa protein gene <i>Ots_HSP90B-385*</i> (DQ908921)	F: CCCTCTCAGCCACCAGGTA R: CTAGGCTGGAGCTGACATCTC VIC-ACCCACGCCAACT FAM-AAACCCACACCAAACT
Leukocyte elastase inhibitor <i>Ots_LEI-292*</i> (DQ908922)	F: CACCTGAACCTCCACTGTGT R: GCTGTGACCTATGAGAAAATTGTG VIC-CATCATGTCAAGGCCTG FAM-ATCATGTCAAGCCTG
Proteasome (prosome, macropain) subunit, beta-type 1 gene <i>Ots_PSMB1-197*</i> (DQ908923)	F: AGAATGTCTAGAGTTGCCTTGAACC R: CCAATCCAACAGCACAAATATGACT VIC-AATAATACATCACTTTTTTCTC FAM-ATACATCACTATTTTTCTC
Antithrombin gene <i>Ots_SERPINC1-209*</i> (DQ908924)	F: CTAAGTCTCTCCTGCCTAATGTGGAT R: CCAAGATTGAGACTTACTATACATTTACAAGTACA VIC-CATTACAGCTTTTTTTC FAM-ATTCAGCATTTTTTC
Heterogeneous nuclear ribonucleoprotein L-like gene <i>Ots_hnRNPL-533*</i> (DQ914957)	F: TCTTTGATATTGAGCTCATAAAAGCAAGT R: TCCTTGTTCATCCATCAGGCATAAAA VIC-CATTTACCAGTTTCACACAC FAM-TTTACCAGTTCACACACAC
Glycoprotein hormone alpha-subunit-like gene <i>Ots_GPH-318*</i> (DQ914958)	F: GGTGATAACAGGTGTTGCACCAA R: TCAGGTGGTGGTGGACAAC VIC-ATCAAGCTGACGAACCA FAM-CAAGCTGACAAACCA

tions and Fisher's tests for genotypic linkage disequilibrium between each pair of loci across samples. Critical values for both tests were adjusted for multiple comparisons using the Bonferroni method (Rice 1989). Loci exhibiting significant genotypic disequilibrium were phased into haplotypes using the Excoffier-Laval-Balding algorithm (Excoffier et al. 2003) implemented in Arlequin (Excoffier et al. 2005). The resulting phased genotypes were used in all subsequent analyses.

*Detection of loci under the influence of natural selection.*—Two criteria were used to identify loci at which allele frequencies were potentially influenced by natural selection. The first was simply a determination

of whether an SNP was nonsynonymous. The second was to use the method of Beaumont and Nichols (1996) to identify "outlier loci" from a plot of heterozygosity versus  $F_{ST}$  (estimated as  $\beta$ ; Cockerham and Weir 1993). This was done in the program FDIST2 (M. A. Beaumont, University of Reading, UK) by generating a distribution of  $F_{ST}$  based on 20,000 replicates of the SNP and allozyme data and then plotting the 0.005 and 0.995 quantiles (between which 99% of the data points are expected to lie). Loci lying above or below these lines were designated as being under selection (Figure 2). Short tandem repeat loci were added to the figure for completeness; however, it was noted that (1) the null distribution was generated using an infinite allele

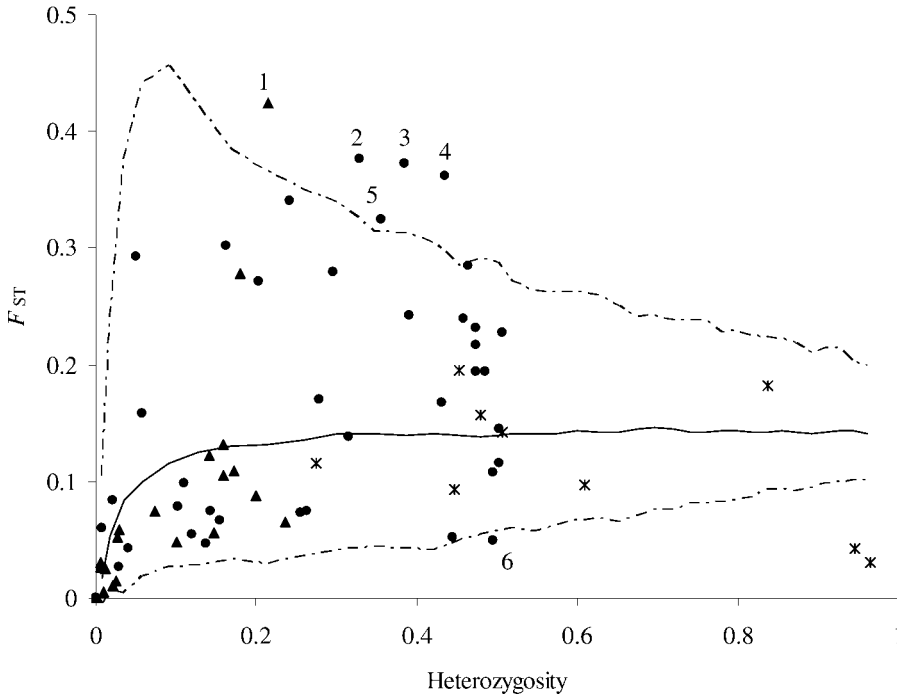


FIGURE 2.—Genetic differentiation index  $F_{ST}$  values estimated from 22 allozyme loci (▲), 9 short tandem repeat (STR) loci (\*), and 39 single-nucleotide polymorphism (SNP) loci (●) plotted against heterozygosity in Chinook salmon. Dashed lines represent 0.005 and 0.995 quantiles. Numbers denote loci outside the boundaries: (1) *sMEP-1\**, (2) *Ots\_u211-85\**, (3) *Ots\_FARSLA-220\**, (4) *Ots\_Tnsf\**, (5) *Ots\_MHC2\**, (6) *Ots\_P53\**; two (non-numbered) STR loci had heterozygosity values of about 1.

model (which may be less appropriate for these markers than a stepwise model), and (2) loci with very high heterozygosity values have low maximum  $F_{ST}$  values (Hedrick 1999).

*Comparison of structure based on different marker types with and without selected single-nucleotide polymorphisms.*—To evaluate the contribution of natural selection to population structure indicated by the present SNPs, we created a second set of SNP data from which we removed loci that met either of the above criteria. This set of putatively neutral SNPs (nSNPs) was included as a fourth data set in the neighbor-joining and divergence analyses outlined below.

The first aspect of concordance between marker sets examined was the pattern of intercollection divergence. Chord distances (Cavalli-Sforza and Edwards 1967) were calculated for each pair of collections for each marker type and were used to construct neighbor-joining (Saitou and Nei 1987) dendrograms using PHYLIP (Felsenstein 1989). Consistency among markers used to generate each dendrogram was assessed by using 1,000 bootstrap replicate data sets to generate chord distances and neighbor-joining dendrograms. Correlations between the pairwise chord distance matrices generated using each marker class

were calculated and tested using a permutation procedure (Mantel 1967). Our hypothesis was that greater correlation between allozymes and STRs than between either type and SNPs would be indicative of a bias acting on the SNP data.

The second aspect of concordance examined was the amount of intercollection divergence indicated by the different marker types. Overall  $F_{ST}$  among collections was estimated as the coancestry coefficient  $\theta$  (Weir and Cockerham 1984),  $\beta$  (Cockerham and Weir 1993), population differentiation index  $G_{ST}$  (Nei 1973), and standardized genetic differentiation index  $G'_{ST}$  (Hedrick 2005) based on allele frequencies observed for each marker. Mean values of each statistic were calculated for each marker type, and 95% confidence intervals (CIs) of each mean were calculated from the  $t$ -distribution based on locus-specific estimates. For each locus, we calculated the within-subpopulation inbreeding coefficient  $F_{IS}$  (following Weir and Cockerham 1984) and the overall inbreeding coefficient  $F_{IT}$  (as  $1 - \{[1 - F_{IS}] \times [1 - \theta]\}$ ). Partitioning of variance within and among collections for each marker type was performed using an analysis of molecular variance (AMOVA) framework (Excoffier et al. 1992). Differences among marker sets were assessed based on

nonoverlapping 95% CIs. As allozyme data for the southernmost three collections consisted of only collection allele frequencies, these collections were excluded from AMOVA.

*Definition of ascertainment panels.*—The SNPs examined in this study were developed using two different ascertainment panels (Figure 1). One ascertainment panel was used to develop 10 SNPs described by Smith et al. (2005b). This subset of SNPs is referred to here as SNPap1, and the ascertainment collections are denoted by clubs in Figure 1. The remaining assayed SNPs were developed using the ascertainment panel described by Smith et al. (2005a); this subset is designated SNPap2, and the ascertainment collections are denoted by hearts in Figure 1. A further complication is that the SNPs in SNPap1 were developed using sequences that were known a priori to be broadly polymorphic in Chinook salmon, while those in SNPap2 were chosen more arbitrarily (see Smith et al. 2005a for details).

*Effects of ascertainment bias on interpopulation inferences.*—To examine whether the SNP data made the ascertainment collections appear disproportionately divergent from one another, we ordered pairs of collections based on smallest to largest pairwise chord distance. This was done for each marker set (allozymes, STR, SNPap1, and SNPap2), giving us four columns that were each 120 ( $[16 \times 15]/2$ ) rows tall. Each column contained the same set of pairwise comparisons, but the order of pairs in each column was based on the chord distances generated using the corresponding marker set. Spearman's rank correlation coefficient ( $\rho_s$ ) was calculated as a measure of concordance between each pair of columns. We then removed all pairwise comparisons between panel collections and recalculated  $\rho_s$ . Overlap between 95% CIs was used to determine whether the inclusion versus exclusion of panel collections affected  $\rho$ . If ascertainment bias is inflating relative divergence among ascertainment panel collections, then we expect that (1)  $\rho_s$  will be higher between allozymes and STRs than between either of these markers and SNPs and (2) removal of the ascertainment collections will increase  $\rho_s$  between SNPs and the other two markers.

*Effects of ascertainment bias on intrapopulation inferences.*—A second aspect of ascertainment bias examined was whether within-collection diversity was upwardly biased for the panel collections based on SNPs. Within-collection diversity  $H_s$  (Nei 1973) was estimated for each collection based on each marker type (allozyme, STR, or SNP). Correlations ( $r^2$ ) were calculated for all pairs of marker types. If ascertainment bias is inflating within-collection diversity estimates for collections represented in the panel, then allozymes

and STRs will correlate better with each other than with SNPs.

## Results

Number of observed alleles, estimates of heterozygosity, and  $F$ -statistics for each locus are listed in Table A.1 (allele frequencies for allozymes, STRs, and SNPs are available at [www.genetics.cf.adfg.state.ak.us/publish/publish.php](http://www.genetics.cf.adfg.state.ak.us/publish/publish.php)). The STR data exhibited approximately twice as many alleles (167) as did allozymes (98) and SNPs (81). All markers exhibited multiple alleles, although nearly half of the allozymes (*sAAT-1,2\**; *mAAT-1\**; *GPIA\**; *HAGH\**; *LDHB-2\**; *LDHC\**; *sMDHA-1,2\**; *PEPD-2\**; and *PGDH\**) and one SNP (*Ots\_PSMB1-197\**) had major allele frequencies greater than 0.950 in all collections. While fixed allele frequency differences were not observed among collections, the SNP *Ots\_FARSLA-220\** exhibited a 0.983 frequency difference between two adjacent collections (14 and 15).

### *Departures from Hardy–Weinberg Equilibrium and Genotypic Equilibrium*

Tests for departures from HWE yielded significant ( $\alpha = 0.05$ ) results for seven of the nine STR loci examined. Six of these loci deviated from HWE for one to three collections each, and generally the deviations occurred in different collections for each locus (only collection 14 exhibited departures from HWE two loci). The seventh STR locus (*One13*) exhibited departures in every collection. One significant result involved a negative  $F_{IS}$  estimate (*Ots2* in collection 5), while the 24 other significant results involved positive  $F_{IS}$  estimates. No departures from HWE were detected at any of the allozyme or SNP loci.

Genotypic disequilibrium was not detected at any of the allozyme or STR loci. The two *Ots\_GST\** SNPs (*Ots\_GST-207\** and *-375\**) were significantly linked to each other in several collections, as were the two *Ots\_HSP90B\** SNPs (*Ots\_HSP90B-100\** and *-385\**). Phasing the alleles at these loci yielded three alleles at *Ots\_GST\** and four alleles at *Ots\_HSP90B\**. *Ots\_Prl2\** and *Ots\_SWS1op-182\** yielded a significant result in the King Salmon River (Alaska) collection but appeared unlinked in all other collections.

### *Detection of Loci under the Influence of Natural Selection*

Based on the plot of  $F_{ST}$  versus heterozygosity (Figure 2), we noted that four loci (*Ots\_FARSLA-220\**, *Ots\_MHC2\**, *Ots\_Tnsf\**, and *Ots\_u211-85\**) fell above the upper 0.995 quantile and that one locus (*Ots\_P53\**) fell below the lower 0.005 quantile. A single allozyme locus (*s-MEP-1\**) also fell above the 0.995 quantile.

TABLE 2.—Summary of genetic markers examined in 16 Chinook salmon collections (see Figure 1). Note that 41 single-nucleotide polymorphism (SNP) assays were run; however, two pairs of assays were phased to create haplotypes, which reduced the number of loci to 39. Neutral SNPs (*n*SNPs) were a subset of SNPs, excluding the loci exhibiting evidence of selection. Migration estimates ( $N_e m$ ) were calculated as  $([1/F_{ST}] - 1)/4$  (Wright 1951) based on the upper and lower 95% confidence limits of  $F_{ST}$ .

Variable <sup>a</sup>	Allozyme	Short tandem repeat	SNP	<i>n</i> SNP
Number of loci examined	22	9	39	31
Number of observed alleles	98	167	81	65
Mean $\theta$ (95% CI)	0.077 (0.043–0.111)	0.121 (0.083–0.159)	0.170 (0.134–0.206)	0.146 (0.116–0.177)
$N_e m$	2.0–5.6	1.3–2.8	1.0–1.6	1.2–1.9
Mean $\beta$ (95% CI)	0.078 (0.037–0.120)	0.118 (0.080–0.155)	0.170 (0.135–0.205)	0.148 (0.115–0.181)
$N_e m$	1.8–6.6	1.4–2.9	1.0–1.6	1.1–1.9
Mean $G_{ST}$ (95% CI)	0.076 (0.044–0.108)	0.120 (0.084–0.157)	0.167 (0.133–0.201)	0.144 (0.115–0.173)
$N_e m$	2.1–5.5	1.3–2.7	1.0–1.6	1.2–1.9
Mean $G'_{ST}$ (95%CI)	0.087 (0.049–0.125)	0.332 (0.230–0.435)	0.230 (0.184–0.275)	0.204 (0.162–0.245)
$N_e m$	1.7–4.9	0.3–0.8	0.7–1.1	0.8–1.3
Variance among collections (%)	8.38	11.19	13.30	12.02

<sup>a</sup> Variables are the coancestry coefficient ( $\theta$ ); genetic differentiation index ( $\beta$ ); population differentiation index ( $G_{ST}$ ); and standardized genetic differentiation index ( $G'_{ST}$ ).

Previous sequence analyses had indicated that three SNPs were nonsynonymous (*Ots\_GH2\**, *Ots\_HGFA\**, and *Ots\_Ikaros-250\**). We produced a subset of the SNP data (i.e., *n*SNPs) that excluded these eight loci (Table 2).

*Comparison of Structure Based on Different Marker Types with and without Selected Single-Nucleotide Polymorphisms*

Comparisons of pairwise population chord distance matrices resulted in significant ( $P < 0.001$ ) positive correlations in each case, suggesting broad similarities in patterns of collection divergence based on the different marker types. Correlation coefficients were slightly higher between allozymes and SNPs (0.791) than between allozymes and STRs (0.634) or SNPs and STRs (0.703). Removal of the eight SNPs designated as being under selection increased the correlation between STRs and SNPs (0.734) but decreased the correlation between allozyme and SNPs (0.773). The neighbor-joining dendrograms (Figure 3) provide a graphical representation of the relative similarities of collections based on the different marker sets. The SNP and allozyme dendrograms appear to share the most similarities. The STRs and SNPs placed the upper Yukon and Kuskokwim River drainages (Alaska; collections 2 and 4) together, but allozymes did not. Single-nucleotide polymorphisms paired the Bistraya River, Russia (collection 1), with Ayakulik River, Kodiak Island, Alaska (collection 6), but allozymes and STRs did not. The SNPs paired the Cook Inlet collections (7 and 8; Moose Creek and Kenai River, Alaska) together, but allozymes and STRs did not. When the eight SNPs were removed, the remaining *n*SNPs no longer paired the Cook Inlet collections. Inspection of the allele frequencies revealed that these

two collections shared similar allele frequencies for two of the selected SNPs (*Ots\_Tnsf\** and *Ots\_MHC2\**). The only pair of populations with 70% or greater bootstrap support in all data sets was the Methow River and Johnson Creek, Washington (collections 15 and 16).

Mean values of  $\theta$ ,  $\beta$ , and  $G_{ST}$  provided nearly identical estimates of  $F_{ST}$  (and thus,  $N_e m$ ) within each marker class, while estimates of  $G'_{ST}$  were generally higher (Table 2). Comparison of 95% CIs indicated that  $G'_{ST}$  was only significantly higher than the other estimators for STRs. All three estimators of  $F_{ST}$  varied depending on the marker class used. Mean values of  $\theta$ ,  $G_{ST}$ , and  $G'_{ST}$  were higher for SNPs than for allozymes (nonoverlapping 95% CIs), and  $G'_{ST}$  was higher for STRs than that for allozymes or *n*SNPs. Mean estimates of each statistic were lower for the *n*SNP subset than for all SNPs; however, 95% CIs overlapped for each estimator.

The AMOVA based on the 13 collections for which we had genotype data indicated that the mean (SE) percentage of variation among collections was 8.38% (1.71%) for allozymes, 11.19% (1.97%) for STRs, and 13.30% (1.58%) for SNPs. When the analysis was repeated with all 16 collections, the percentage of variation was 12.16% (1.97%) for STRs and 17.06% (1.84%) for SNPs. Although both point estimates were higher for SNPs than for STRs, the 95% CIs overlapped, and thus we did not observe significantly different patterns in the partitioning of variation for these marker types.

*Effects of Ascertainment Bias on Interpopulation Inferences*

The  $p_s$ -values between the orders of pairwise chord distances for the marker sets examined here revealed

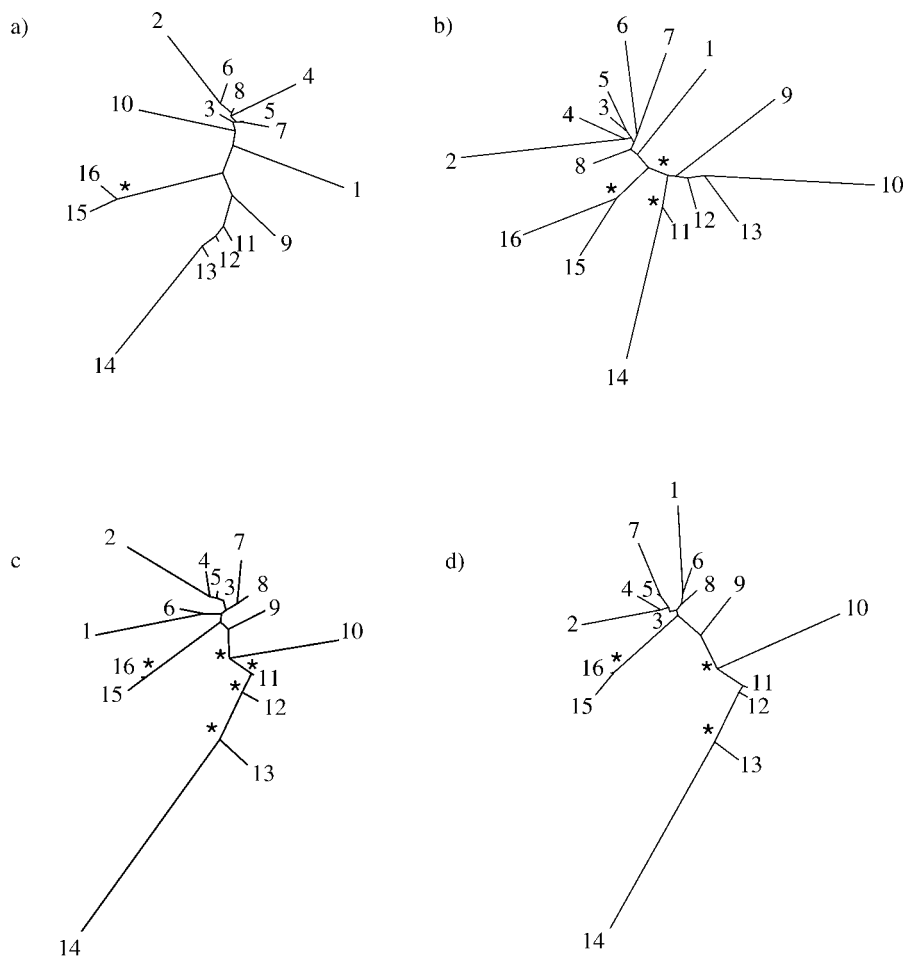


FIGURE 3.—Neighbor-joining dendrograms based on chord distances calculated from four sets of genetic markers in Chinook salmon (collection numbers are defined in Figure 1). The marker sets were (a) 22 allozymes, (b) 9 short tandem repeats (STRs), (c) 39 single-nucleotide polymorphisms (SNPs), and (d) 31 putatively neutral SNPs (a subset of those in [c]). Asterisks indicate nodes with 70% or greater bootstrap support.

TABLE 3.—Spearman's rank correlation coefficients ( $\rho_s$ ; with 95% CI) between orders of pairwise chord distances for genetic markers in 16 populations of Chinook salmon (STR = short tandem repeats; SNP<sub>ap1</sub> and SNP<sub>ap2</sub> = single-nucleotide polymorphism ascertainment panels 1 and 2, respectively) for all samples and for only those samples that were not included in ascertainment panel collections.

Marker pair	$\rho_s$	
	All samples	Ascertainment collections removed
Allozyme and STR	0.71 (0.53–0.88)	0.71 (0.53–0.88)
Allozyme and SNP <sub>ap1</sub>	0.78 (0.60–0.95)	0.78 (0.54–1.00)
Allozyme and SNP <sub>ap2</sub>	0.68 (0.50–0.86)	0.70 (0.46–0.95)
STR and SNP <sub>ap1</sub>	0.60 (0.42–0.77)	0.61 (0.37–0.86)
STR and SNP <sub>ap2</sub>	0.73 (0.55–0.91)	0.60 (0.36–0.84)

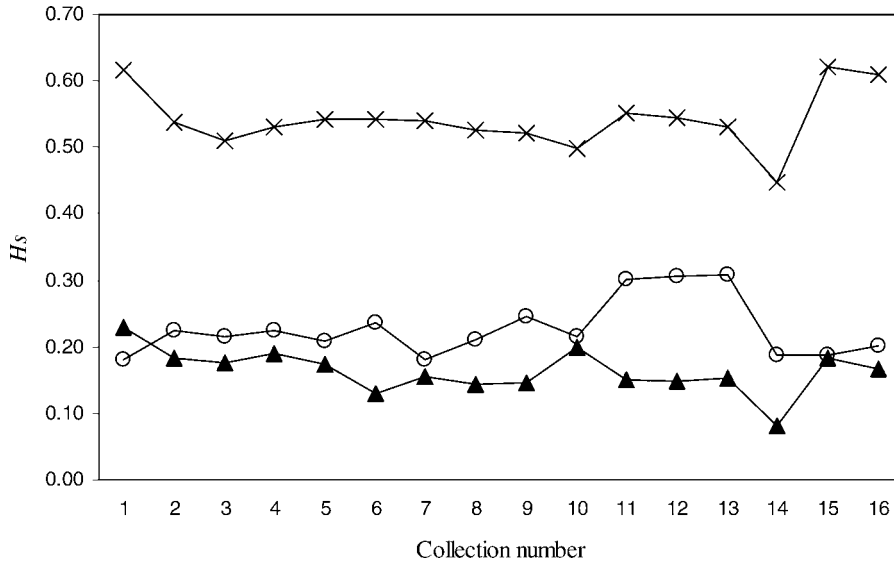


FIGURE 4.—Estimates of within-collection variation  $H_s$  for 16 Chinook salmon collections (see Figure 1) including those used in the SNPap2 discovery panel (2, 4, 10, and 12; see text). Three marker sets were used: allozymes (▲), short tandem repeats (STRs; ×), and single-nucleotide polymorphisms (SNPs; ○). Estimates based on STRs and allozymes correlated better with one another ( $r^2 = 0.34$ ) than with estimates based on SNPs (SNP–STR  $r^2 = 0.02$ ; SNP–allozyme  $r^2 = 0.04$ ).

little evidence of ascertainment bias (Table 3). The highest  $\rho_s$  observed was between allozymes and SNPap1 (0.78), and the lowest was between STRs and SNPap1 (0.60). Correlation between SNPs and the other two marker classes did not appear to improve when the SNP ascertainment collections were removed from this analysis. Overlap between 95% CIs suggested that none of the differences between marker classes or between comparisons with or without the ascertainment collections were significant (Table 3).

#### Effects of Ascertainment Bias on Intrapopulation Inferences

The plot of within-collection diversity (Figure 4) was examined for evidence that collections with high diversity based on allozymes and STRs were less diverse based on SNPs. We observed a stronger correlation between allozymes and STRs ( $r^2 = 0.34$ ) than between either of these marker classes and SNPs (allozyme–SNP  $r^2 = 0.04$ ; STR–SNP  $r^2 = 0.02$ ). Examination of Figure 4 suggests a discrepancy among the three marker classes as indicators of the relative diversity of collections outside the SNPap2 ascertainment range (i.e., outside of collections 2–12). To determine whether this caused the large discrepancy in  $r^2$  values, we recalculated  $r^2$  between paired marker classes based only on the ascertainment collections (2–12). In this case, the correlation between allozymes and STRs was reduced ( $r^2 = 0.28$ ) and the correlation of

SNPs with the other markers increased (allozyme–SNP  $r^2 = 0.11$ ; STR–SNP  $r^2 = 0.17$ ).

#### Discussion

The collections examined here were chosen to represent a broad sample of genetic diversity in Chinook salmon, so it is not surprising that considerable polymorphism was observed. The 0.983 allele frequency difference observed for the SNP *Ots\_FAR-SLA-220\** between collections 14 and 15 is extreme; however, an ancient lineage break has previously been described between the corresponding populations (Waples et al. 2004), and randomly amplified polymorphic DNA (Rasmussen et al. 2003) and STR (Narum et al. 2004) markers with similar resolving power for these groups have been reported.

#### Departures from Hardy–Weinberg Equilibrium and Genotypic Equilibrium

Departures from genotypic ratios expected under HWE can be caused by (1) departures from the evolutionary model assumed for HWE, (2) pooling of populations into one collection (i.e., Wahlund effect), or (3) technical errors leading to miscalled genotypes. Of the 70 markers examined here, 7 exhibited departures from genotypic frequencies expected under HWE conditions. All seven were STRs, and most departures were due to homozygote excess (indicated by positive  $F_{IS}$  estimates). Because the statistical power

to detect departures from HWE in a given sample is inversely related to the number of alleles at a marker (i.e., for a fixed sample size, a greater number of categories means lower counts per category) and because we only detected departures from HWE in the marker class with the largest number of alleles per locus, genotyping errors seem a likely explanation for our results. Null alleles (those with a mutation in the PCR priming site) and allelic dropout (preferential amplification of shorter alleles) are two common explanations for such results in STR data sets. We observed a genotypic linkage between *Ots\_Pr12\** and *Ots\_SWS1op-182\** in collection number 10 that was absent from all other collections; therefore, the linkage is probably attributable to population history or selection (discussion in Ohta 1982) rather than physical linkage.

#### *Detection of Loci under the Influence of Natural Selection*

Our power to detect which of the 38 nuclear SNP loci were under the influence of natural selection was limited. Based on a null hypothesis of neutrality and the fact that we started with 60 loci (38 nuclear SNPs + 22 allozymes), we would expect approximately one locus to lie outside the 0.005–0.995 quantiles. The model used here predicts that the four loci above the 0.995 quantile (*Ots\_FARSLA-220\**, *Ots\_MHC2\**, *Ots\_Tnsf\**, and *Ots\_u211\**) are subject to directional selection, while the locus under the 0.005 quantile (*Ots\_P53\**) is subject to balancing selection. Two of the four loci above the 0.995 quantile have previously been identified as targets of natural selection in Chinook salmon (Miller et al. 1997; Ford 2000). One potential problem with using the Beaumont and Nichols (1996) test on the present data set is that it assumes equal migration between populations, and violations of this assumption could lead to type I errors (i.e., labeling neutral loci as under selection). Because our objective was to remove loci that were potentially under selection, however, we do not suspect that this caused a problem for our downstream analyses. While it would be unsound to conclude that all loci outside the 99% quantiles or all loci coding nonsynonymous substitutions are strongly influenced by natural selection, we expect that exclusion of these loci removed the targets of the strongest selection and thus rendered our remaining SNP set (nSNPs) nearly neutral.

#### *Comparison of Structure Based on Different Marker Types with and without Selected Single-Nucleotide Polymorphisms*

The overall concordance that we observed in the chord distance matrices and neighbor-joining trees

based on the three marker types revealed a broad-scale population structure that was independent of the marker class used. The higher correlation between allozymes and SNPs than between either marker class and STRs suggests that the influence of biases acting only on SNPs (e.g., ascertainment bias and selection) was small relative to other sources of variation (e.g., locus sampling effects and marker class evolutionary rates) affecting the relative chord distances between collections. Anderson et al. (2005) observed that geographic distance in malarial parasites had a higher correlation with genetic distance when based on 10 nSNPs than when based on 10 selected SNPs. A comparable analysis here would be complicated by the unequal number of SNPs in each category; however, because we identified loci as selected based on  $F_{ST}$  outliers, we expect that population structure based solely on these loci would be biased. The selection vectors on these loci may be different between different collections, and the large number of unlinked loci examined was probably enough to override selection acting in a particular direction in one or a few collections.

The four  $F_{ST}$  estimators used to quantify divergence between collections indicated comparable levels of among-collection variation for the three marker classes. The difference in  $G_{ST}$  values between STRs and allozymes was probably an artifact of the large number of alleles and concomitant high heterozygosity at STR loci (Hedrick 1999). Although  $\theta$  and its simplified form,  $\beta$ , are analogous to Wright's (1951)  $F_{ST}$  and can be used to infer a balance between migration and drift,  $G'_{ST}$  is expected to reflect differences in mutation rate to a large extent and population inferences based on this statistic are less clear. We used  $G'_{ST}$  to facilitate comparison of divergence estimates regardless of the heterozygosity differences between markers; in this respect, it has proven useful. Of the three marker types examined here, SNPs provided the most consistent estimates of  $F_{ST}$  (and thus,  $N_e m$ ) across methods. Aside from limitations based on numbers of alleles, comparisons of  $F_{ST}$  based on SNPs and STRs are limited by the different evolutionary properties of the two marker classes. The results presented here are concordant with those of other salmonid studies that found comparable overall estimates of  $\theta$  based on allozymes and STRs despite a large amount of variation among specific loci (Scribner et al. 1996; Allendorf and Seeb 2000). The overlap in estimates based on STRs and SNPs suggest that such trends hold for SNPs as well.

Within- and among-collection variance proportions were similar for all three markers in the 13 collections and similar for STRs and SNPs within the entire data set (Table 1). Point estimates of among-collection

variance were highest for SNPs, but 95% CIs for all marker types overlapped. The 95% CIs also included the published estimates based on allozymes within British Columbia (8.7%; Teel et al. 2000) and south of British Columbia (12.3%; Utter et al. 1989).

#### *Effects of Ascertainment Bias on Intrapopulation Inferences*

An apparent effect of ascertainment bias was observed in our examination of within-collection diversity. In this case, a stronger correlation was observed between our two control data sets (allozymes and STRs) than between either set and SNPs. Studies of humans (Mountain and Cavalli-Sforza 1994) and fruit flies *Drosophila* spp. (Schlötterer and Harr 2002) have revealed that estimates of interpopulation diversity in ancient populations can be biased downward when based on SNPs that are ascertained in relatively derived panel populations. Given the relatively ancient divergence between collections 14 and 15–16 (Waples et al. 2004) and the fact that none of these populations were represented in the ascertainment panels, it is not surprising that the SNPs examined here did not reveal the greater diversity indicated by allozymes and STRs for collections 15–16 (Figure 4).

#### **Conclusions**

Biases affecting allele frequencies at a single marker seem unlikely to change the overall structure observed when the number of markers examined is large. Even biases affecting several markers (e.g., under the influence of natural selection) might be of little consequence as long as the selection vectors are different. This, or the possibility that selection coefficients were small relative to migration and drift for the present loci, may explain why selection did not appear to shape branching order (the relative placement of individuals in the neighbor-joining analysis), divergence estimates, or variance partitions. The impact of a few selection-influenced markers on population structure inferences cannot yet be generalized. However, empirical comparisons such as the one presented here advance our understanding. In any case, routine screening of loci for the potential influence of selection prior to structure estimation, followed by estimation of structure with and without the outlier loci, is easy to do and seems well advised (Luikart et al. 2003).

Because SNPs are often ascertained in batches, ascertainment bias probably acts in conjunction across several loci. The effects of ascertainment bias on inferences of some population parameters can be minimized by careful selection of the source and size of the ascertainment panel (Akey et al. 2003), and this

concept has been used to guide SNP discovery projects (e.g., Cappuccino et al. 2006). The fact that we did not observe any impact of ascertainment bias on branching order, overall divergence, variance partitioning, or estimation of divergence among panel collections may be partially due to our relatively large and diverse panel. Additional comparative data sets will be required before we can make general statements about the effect of ascertainment bias on branching order and divergence. However, assuming a large, diverse ascertainment panel is used, the effect will be negligible in many cases. For comparisons of panel populations with nonpanel populations, we expect the results to vary quite a bit depending on which nonpanel collections are examined; even large, diverse panels might not be able to prevent such variability. Our Chinook salmon data support the inadvisability of estimating within-population diversity based on SNPs developed in ascertainment panels from derived populations.

#### **Acknowledgments**

We are grateful for excellent technical assistance provided by Andy Barclay, Nick DeCovich, and Zac Grauvogel. We are also grateful to Anti Vasemägi for conversations during which he clarified several aspects of tests for natural selection. TaqMan primers and probes for *Ots\_FGF6A\** were provided by Linda Park (National Oceanic and Atmospheric Administration, Northwest Fisheries Science Center). Funding was provided by the Southeast Sustainable Salmon Fund and a Saltonstall–Kennedy Research Grant. This manuscript was improved based on comments from three anonymous reviewers. The views expressed here are those of the authors and not necessarily those of the U.S. Fish and Wildlife Service. Reference to trade names does not imply endorsement by the U.S. Government.

#### **References**

- Aebersold, P. B., G. A. Winans, D. J. Teel, G. B. Milner, and F. M. Utter. 1987. Manual for starch gel electrophoresis: a method for the detection of genetic variation. NOAA Technical Report NMFS 61.
- Akey, J. M., K. Zhang, M. Xiong, and L. Jin. 2003. The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium. *Molecular Biology and Evolution* 20:232–242.
- Allendorf, F. W., and L. W. Seeb. 2000. Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution* 54:640–651.
- Anderson, T. J., S. Nair, D. Sudimack, J. T. Williams, M. Mayxay, P. N. Newton, J. P. Guthmann, F. M. Smithuis, T. H. Tran, I. V. van den Broek, N. J. White, and F. Nosten. 2005. Geographical distribution of selected and

- putatively neutral SNPs in Southeast Asian malaria parasites. *Molecular Biology and Evolution* 22:2362–2374.
- Banks, M. A., M. S. Blouin, B. A. Baldwin, V. K. Rashbrook, H. A. Fitzgerald, S. M. Blankenship, and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in Chinook salmon (*Oncorhynchus tshawytscha*). *Journal of Heredity* 90:281–288.
- Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society London B* 263:1619–1626.
- Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18:249–256.
- Cappuccino, I., L. Pariset, P. Ajmone-Marsan, S. Dunner, O. Cortes, G. Erhardt, G. Lühken, K. Gutscher, S. Joost, I. J. Nijman, J. A. Lenstra, P. R. England, S. Zundel, G. Obexer-Ruff, A. Beja-Pereira, A. Valentini, and Econogene Consortium 2006. Allele frequencies and diversity parameters of 27 single nucleotide polymorphisms within and across goat breeds. *Molecular Ecology Notes* 6:992–997.
- Cavalli-Sforza, L. L., and A. W. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* 19:233–257.
- Cockerham, C. C., and B. S. Weir. 1993. Estimation of gene flow from *F*-statistics. *Evolution* 47:855–863.
- Estoup, A., P. Presa, F. Krieg, D. Vaiman, and R. Guyomard. 1993. (CT)<sub>n</sub> and (GT)<sub>n</sub> microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity* 71:488–496.
- Excoffier, L., G. Laval, and D. Balding. 2003. Gametic phase estimation over large genomic regions using an adaptive window approach. *Human Genomics* 1:7–19.
- Excoffier, L., G. Laval, and S. Schneider. 2005. ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47–50. Available: la-press.com. (October 2007).
- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
- Felsenstein, J. 1989. PHYLIP: phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
- Ford, M. J. 2000. Effects of natural selection on patterns of DNA sequence variation at the transferrin, somatolactin, and p53 genes within and among Chinook salmon (*Oncorhynchus tshawytscha*) populations. *Molecular Ecology* 9:843–855.
- Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Heath, D. D., J. M. Shrimpton, R. I. Hepburn, S. K. Jamieson, S. K. Brode, and M. F. Docker. 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, P. W. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633–1638.
- Hedrick, P. W. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* 53:313–318.
- IUBMB (International Union of Biochemistry and Molecular Biology). 1992. Enzyme nomenclature 1992. Academic Press, San Diego, California.
- Kruglyak, L. 1997. The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics* 17:21–24.
- Kwok, P. Y. 2003. Single nucleotide polymorphisms—methods and protocols. Humana Press, Totowa, New Jersey.
- Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4:981–994.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27:209–220.
- Miller, K. M., R. E. Withler, and T. D. Beacham. 1997. Molecular evolution at MHC genes in two populations of Chinook salmon *Oncorhynchus tshawytscha*. *Molecular Ecology* 6:937–954.
- Morin, P. A., G. Luikart, and R. K. Wayne. SNP workshop group. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19:208–216.
- Mountain, J. L., and L. L. Cavalli-Sforza. 1994. Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. *Proceedings of the National Academy of Sciences of the USA* 91:6515–6519.
- Narum, S. R., M. S. Powell, and A. J. Talbot. 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.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* 70:3321–3323.
- Nelson, R. J., and T. D. Beacham. 1999. Isolation and cross species application of microsatellite loci useful for study of Pacific salmon. *Animal Genetics* 30:228–229.
- Nelson, R. J., T. D. Beacham, and M. P. Small. 1998. Microsatellite analysis of the population structure of a Vancouver Island sockeye salmon (*Oncorhynchus nerka*) stock complex using nondenaturing gel electrophoresis. *Molecular Marine Biology and Biotechnology* 7:312–319.
- Nielsen, R., and J. Signorovitch. 2003. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology* 63:245–255.
- Ohta, T. 1982. Linkage disequilibrium with the island model. *Genetics* 101:139–155.
- Olsen, J. B., S. L. Wilson, E. J. Kretschmer, K. C. Jones, and J. E. Seeb. 2000. Characterization of 14 tetranucleotide microsatellite loci derived from sockeye salmon. *Molecular Ecology* 9:2185–2187.
- Rasmussen, C., C. O. Ostberg, D. R. Clifton, and R. J. Rodriguez. 2003. Identification of a genetic marker that discriminates ocean-type and stream-type Chinook

salmon in the Columbia River basin. Transactions of the American Fisheries Society 132:131–142.

Raymond, M., and F. Rousset. 1997. GENEPOP (version 3.1b), an updated version of GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248–249.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406–425.

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

Schlötterer, C., and B. Harr. 2002. Single nucleotide polymorphisms derived from ancestral populations show no evidence for biased diversity estimates in *Drosophila melanogaster*. Molecular Ecology 11:947–950.

Scribner, K. T., P. A. Crane, W. J. Spearman, and L. W. Seeb. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 55:1748–1758.

Scribner, K. T., J. R. Gust, and R. L. Fields. 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic applications. Canadian Journal of Fisheries and Aquatic Sciences 53:833–841.

Smith, C. T., C. M. Elfstrom, J. E. Seeb, and L. W. Seeb. 2005a. Use of sequence data from rainbow trout and

Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14:4193–4203.

Smith, C. T., J. E. Seeb, P. Schwenke, and L. W. Seeb. 2005b. Use of the 5'-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. Transactions of the American Fisheries Society 134:207–217.

Teel, D. J., P. A. Crane, C. M. Guthrie, A. R. Marshall, D. M. Van Doornik, W. D. Templin, N. V. Varnavskaya, and L. W. Seeb. 1999. Comprehensive allozyme database discriminates Chinook salmon around the Pacific Rim. Alaska Department of Fish and Game, Division of Commercial Fisheries, North Pacific Anadromous Fish Commission Document 440, Anchorage.

Teel, D. J., G. B. Milner, G. A. Winans, and W. S. Grant. 2000. Genetic population structure and origin of life history types in Chinook salmon in British Columbia, Canada. Transactions of the American Fisheries Society 129:194–209.

Utter, F. W., G. Milner, G. Stahl, and D. Teel. 1989. Genetic population structure of Chinook salmon, *Oncorhynchus tshawytscha*, in the Pacific Northwest. U.S. National Marine Fisheries Service Fishery Bulletin 87:239–264.

Waples, R. S., D. J. Teel, J. M. Myers, and A. R. Marshall. 2004. Life-history divergence in Chinook salmon: historic contingency and parallel evolution. Evolution 58:386–403.

Weir, B. S., and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323–354.

**Appendix: Chinook Salmon Loci Surveyed**

TABLE A.1.—Loci surveyed in 16 collections of Chinook salmon. The annealing temperature and MgCl<sub>2</sub> concentration associated with each short tandem repeat (STR) multiplex are given in Methods. Diversity and structure indices are within-collection diversity (*H<sub>s</sub>*), within-subpopulation inbreeding coefficient (*F<sub>IS</sub>*), overall inbreeding coefficient (*F<sub>IT</sub>*), coancestry coefficient (*θ*), genetic differentiation index (*β*), population differentiation index (*G<sub>ST</sub>*), and standardized differentiation index (*G'<sub>ST</sub>*).

Locus	STR multiplex	Number of alleles	<i>H<sub>s</sub></i>	<i>F<sub>IS</sub></i>	<i>F<sub>IT</sub></i>	<i>F<sub>ST</sub></i>				Source <sup>a</sup>
						<i>θ</i>	<i>β</i>	<i>G<sub>ST</sub></i>	<i>G'<sub>ST</sub></i>	
<b>Allozyme</b>										
<i>sAAT-1,2*</i>		4	0.018	0.004	0.025	0.021	0.025	0.023	0.024	1
<i>sAAT-3*</i>		4	0.150	0.004	0.118	0.114	0.088	0.110	0.130	1
<i>mAAT-1*</i>		5	0.011	0.005	0.013	0.008	0.005	0.011	0.012	1
<i>ADA-1*</i>		5	0.115	0.004	0.143	0.140	0.132	0.134	0.152	1
<i>sAH*</i>		5	0.147	0.004	0.124	0.120	0.109	0.115	0.136	1
<i>GPIA*</i>		4	0.017	0.004	0.023	0.019	0.010	0.021	0.021	1
<i>HAGH*</i>		5	0.028	0.004	0.059	0.055	0.058	0.055	0.056	1
<i>sIDHP-1*</i>		9	0.160	0.005	0.120	0.116	0.105	0.112	0.135	1
<i>LDHB-2*</i>		4	0.005	0.004	0.030	0.026	0.030	0.028	0.028	1
<i>LDHC*</i>		4	0.012	0.004	0.028	0.024	0.027	0.026	0.027	1
<i>sMDHA-1,2*</i>		4	0.001	0.001	0.004	0.004	0.000	0.004	0.004	1
<i>sMDHB-1,2*</i>		6	0.024	0.024	0.004	0.118	0.014	0.017	0.017	1
<i>mMDH-2*</i>		3	0.143	0.005	0.233	0.229	0.277	0.220	0.259	1
<i>sMEP-1*</i>		4	0.138	0.005	0.340	0.337	0.424	0.323	0.379	1
<i>sMEP-2*</i>		2	0.121	0.005	0.140	0.136	0.123	0.131	0.151	1
<i>MPI*</i>		5	0.202	0.004	0.081	0.077	0.065	0.075	0.095	1
<i>PEPA*</i>		5	0.064	0.004	0.082	0.078	0.075	0.076	0.081	1
<i>PEPB-1*</i>		5	0.149	0.005	0.063	0.058	0.056	0.058	0.069	1

TABLE A.1.—Continued.

Locus	STR multiplex	Number of alleles	$H_s$	$F_{IS}$	$F_{IT}$	$F_{ST}$				Source <sup>a</sup>
						$\theta$	$\beta$	$G_{ST}$	$G'_{ST}$	
<i>PEPD-2*</i>		3	0.001	0.005	0.021	0.016	0.000	0.019	0.019	1
<i>PGDH*</i>		3	0.001	0.004	0.005	0.001	0.000	0.005	0.005	1
<i>TPI-3*</i>		4	0.028	0.005	0.057	0.052	0.051	0.052	0.054	1
<i>TPI-4*</i>		5	0.076	0.004	0.057	0.053	0.048	0.054	0.058	1
<b>STR</b>										
<i>One7</i>	1	2	0.382	0.043	0.231	0.197	0.196	0.192	0.319	2
<i>Ots1</i>	1	8	0.55	0.039	0.133	0.098	0.097	0.099	0.229	3
<i>Ots2</i>	1	21	0.432	-0.024	0.137	0.157	0.142	0.154	0.278	3
<i>One102</i>	2	2	0.393	0.012	0.097	0.086	0.093	0.086	0.146	4
<i>Ots107</i>	2	36	0.898	0.048	0.091	0.045	0.043	0.048	0.497	5
<i>uSat73</i>	2	5	0.382	-0.006	0.155	0.160	0.157	0.156	0.259	6
<i>One13</i>	3	15	0.671	0.446	0.545	0.179	0.183	0.178	0.564	2
<i>One9</i>	3	3	0.222	0.130	0.249	0.136	0.115	0.134	0.174	2
<i>Ots100</i>	3	75	0.925	0.025	0.057	0.033	0.031	0.037	0.526	7
<b>Single-nucleotide polymorphism</b>										
<i>Ots_E2-275*</i>		2	0.39	-0.045	0.162	0.198	0.195	0.192	0.323	8
<i>Ots_arf-188*</i>		2	0.008	0.150	0.201	0.060	0.060	0.063	0.063	8
<i>Ots_AsnRS-60*</i>		2	0.418	0.028	0.079	0.052	0.052	0.055	0.097	8
<i>Ots_C3N3*</i>		2	0.107			0.309	0.302	0.303	0.342	9
<i>Ots_FARSLA-220*</i>		2	0.249	0.097	0.435	0.374	0.373	0.363	0.491	Table 1
<i>Ots_FGF6A*</i>		2	0.338	0.092	0.293	0.222	0.240	0.216	0.333	10
<i>Ots_E2-275*</i>		2	0.39	-0.045	0.162	0.198	0.195	0.192	0.323	8
<i>Ots_GH2*</i>		2	0.237	-0.068	0.118	0.174	0.170	0.169	0.225	9
<i>Ots_GnRH-271*</i>		2	0.048	-0.020	0.153	0.170	0.158	0.166	0.175	8
<i>Ots_GPDH-338*</i>		2	0.159	0.065	0.122	0.061	0.067	0.064	0.077	8
<i>Ots_GPH-318*</i>		2	0.234	0.098	0.168	0.078	0.073	0.080	0.106	Table 1
<i>Ots_GST-207*</i>		3 <sup>b</sup>	0.166	0.011	0.365	0.358	0.340	0.346	0.420	Table 1
<i>Ots_GST-375*</i>										
<i>Ots_HGFA-446*</i>		2	0.041	0.095	0.360	0.293	0.292	0.284	0.297	8
<i>Ots_hnRNPL-533*</i>		2	0.37	0.021	0.222	0.206	0.194	0.200	0.325	Table 1
<i>Ots_HSP90B-100*</i>		4 <sup>b</sup>	0.363	-0.048	0.184	0.222	0.231	0.215	0.346	Table 1
<i>Ots_HSP90B-385*</i>										
<i>Ots_IGF-1.1-76*</i>		2	0.356	-0.005	0.171	0.175	0.168	0.171	0.272	8
<i>Ots_Ikaros-250*</i>		2	0.109	0.051	0.146	0.101	0.098	0.101	0.114	8
<i>Ots_il-1racp-16*</i>		2	0.362	0.013	0.230	0.220	0.217	0.214	0.343	8
<i>Ots_LEI-292*</i>		2	0.046	0.040	0.076	0.038	0.043	0.042	0.044	Table 1
<i>Ots_MHC1*</i>		2	0.408	0.022	0.203	0.185	0.228	0.179	0.311	9
<i>Ots_MHC2*</i>		2	0.222	0.045	0.396	0.367	0.325	0.356	0.464	9
<i>Ots_ZNF330-181*</i>		2	0.021	-0.103	-0.006	0.088	0.084	0.088	0.090	8
<i>Ots_LWSop-638*</i>		2	0.098	0.050	0.119	0.073	0.078	0.074	0.083	8
<i>Ots_SWS1op-182*</i>		2	0.434	-0.024	0.095	0.116	0.108	0.114	0.207	8
<i>Ots_Ots2*</i>		2	0.258	0.077	0.199	0.132	0.139	0.131	0.179	9
<i>Ots_P450*</i>		2	0.308	-0.008	0.230	0.236	0.242	0.229	0.337	9
<i>Ots_P53*</i>		2	0.469	0.035	0.083	0.050	0.049	0.053	0.103	9
<i>Ots_PSMB1-197*</i>		2	0.001	0.001	0.001	0.000	0.000	0.005	0.005	Table 1
<i>Ots_Pri2*</i>		2	0.453	-0.021	0.074	0.093	0.116	0.092	0.174	9
<i>Ots_ins-115*</i>		2	0.03	-0.038	-0.013	0.024	0.026	0.028	0.029	8
<i>Ots_RFC2-558*</i>		2	0.143	0.035	0.301	0.276	0.271	0.267	0.314	8
<i>Ots_SC1kF2R2-135*</i>		2	0.431	0.015	0.148	0.135	0.144	0.132	0.239	8
<i>Ots_SERPC1-209*</i>		2	0.133	0.169	0.222	0.064	0.074	0.066	0.077	Table 1
<i>Ots_SL*</i>		2	0.336	-0.016	0.272	0.284	0.284	0.275	0.423	9
<i>Ots_Tnsf*</i>		2	0.265	-0.007	0.343	0.347	0.362	0.336	0.465	9
<i>Ots_u202-161*</i>		2	0.209	0.031	0.311	0.289	0.279	0.280	0.358	8
<i>Ots_u211-85*</i>		2	0.177	0.004	0.403	0.401	0.376	0.389	0.478	8
<i>Ots_U212-158*</i>		2	0.114	-0.027	0.032	0.057	0.055	0.059	0.067	8
<i>Ots_u4-92*</i>		2	0.131	-0.085	-0.037	0.044	0.047	0.047	0.055	8
<i>Ots_u6-75*</i>		2	0.251	0.030	0.100	0.072	0.074	0.073	0.100	8

<sup>a</sup> Sources for running conditions and oligonucleotide sequences are as follows: (1) Aebersold et al. (1987); (2) Scribner et al. (1996); (3) Banks et al. (1999); (4) Olsen et al. (2000); (5) Nelson and Beacham (1999); (6) Estoup et al. (1993); (7) Nelson et al. (1998); (8) Smith et al. (2005a); (9) Smith et al. (2005b); and (10) Linda Park, National Oceanic and Atmospheric Administration, Northwest Fisheries Science Center, personal communication.

<sup>b</sup> Number of unique haplotypes predicted by the Excoffier–Laval–Balding algorithm method (Excoffier et al. 2003). This number accounts for the blank cells below it.