Report on survey of microsatellite and single nucleotide (SNP) variation in northern British Columbia Chinook salmon and SNP variation in sockeye salmon

Terry D. Beacham, K. Jonsen, B. McIntosh, and C. W. Wallace

Department of Fisheries and Oceans

Pacific Biological Station

Nanaimo, BC

Canada V9T 6N7

February 2010

A project funded by the Northern Boundary Restoration and Enhancement Fund

Contents

Abstract

Variation at 14 microsatellite loci (STRs), one major histocompatibility complex (MHC) locus, and 49 single nucleotide polymorphism (SNPs) loci was surveyed in 44 populations of sockeye salmon (*Oncorhynchus nerka*) over 16 regions from southern and central British Columbia. Sequential addition of the five highest rated SNPs to the suite of 14 microsatellites provided the equivalent average accuracy when compared with the current suite of microsatellites and MHC. The six highest rated microsatellites provided the equivalent average stock identification resolution and individual assignment accuracy compared with the 46 highest rated SNPs. For regional stock compositions, 53-104 SNPs were projected to be required to provide accuracy and precision equivalent to the microsatellites. For population-specific stock compositions, 75-79 SNPs were projected to be required to provide accuracy and precision equivalent to the microsatellites. Equivalency in individual assignment accuracy to region was estimated to require 100 SNPs of the quality evaluated in the study, whereas equivalent accuracy in assignment to specific populations was estimated to require 124 SNPs. Applications that incorporate the existing power of a combined microsatellite-SNP approach are the best current technique available for sockeye salmon stock identification applications in southern British Columbia.

Variation at 29 microsatellite loci (STRs) and 73 single nucleotide polymorphism (SNPs) loci was surveyed in populations of Chinook salmon (*Oncorhynchus tshawytscha*) in northern British Columbia. Microsatellites with larger observed allele numbers displayed lower F_{ST} values, but these same loci with lower F_{ST} values also tended to provide more accurate estimates of stock composition. The observed number

of alleles was related to the power of the locus for providing accurate estimates of stock composition of simulated single-population samples. The number of microsatellites required to produce stock identification results equivalent to those produced from applying SNPs depended upon the specific task. Regional estimates of stock composition required at least 14 microsatellites, population-specific estimates of stock composition required six to nine microsatellites, and individual assignment to population and region required eight to 12 microsatellites. Options available for improving the accuracy and precision of stock composition estimates for 12-locus DFO microsatellite suite range include adding either four microsatellites or 25 SNPs to the existing suite. For the 13-locus GAPS microsatellites, either two microsatellites or 20-25 SNPs can be added to the existing suite. The addition of either sets of loci to both sets of microsatellites would provide generally equivalent results. When the enhanced microsatellite baselines were evaluated, 79-88 SNPs were estimated to be required for equivalency of regional accuracy and precision for the enhanced DFO microsatellites, and 68-88 SNPs were estimated to be required for the GAPS baseline. The enhanced microsatellite baselines were projected to require 179 SNPs and 166 SNPs, respectively, for equivalency of precision of population-specific estimates. The level of regional accuracy of individual assignment available from the enhanced DFO and GAPS suites of microsatellites was projected to require 90 and 82 SNPs, respectively. The level of individual assignment to specific populations available from the enhanced DFO and GAPS suites of microsatellites was projected to require 137 and 121 SNPs, respectively. A combined microsatellite and SNP approach to estimation of stock composition in fisheries

management applications may be possible for Chinook salmon stock identification applications in British Columbia.

Introduction

The first objective of project NEF 2008-I-15b was to survey variation at 45 SNP loci in 4,500 individuals from British Columbia sockeye salmon populations. In fact, during the conduct of the study, 49 SNPs were surveyed for 44 populations of sockeye salmon from central and southern British Columbia. An evaluation of the 49-SNP baseline was conducted for stock identification applications of sockeye salmon in British Columbia and compared with an existing microsatellite baseline. The results of these analyses are outlined in Appendix 1.

The second objective of the study was to survey variation at 53 SNPs for 1,200 Chinook salmon in northern and central British Columbia. During the conduct of the study, approximately 73 SNPs were surveyed for variation in 12 populations. The data were subsequently combined with those available from a concurrent survey of southern British Columbia populations, and an evaluation conducted stock identification applications of Chinook salmon in British Columbia and compared with existing microsatellite baselines. The results of these analyses are outlined in Appendix 2.

Appendix 1.

A comparison of stock and individual identification for sockeye salmon (*Oncorhynchus nerka*) in British Columbia provided by microsatellites (STRs) and single nucleotide polymorphisms (SNPs)

Abstract

Variation at 14 microsatellite loci (STRs), one major histocompatibility complex (MHC) locus, and 49 single nucleotide polymorphism (SNPs) loci was surveyed in 44 populations of sockeye salmon (*Oncorhynchus nerka*) over 16 regions from southern and central British Columbia. Sequential addition of the five highest rated SNPs to the suite of 14 microsatellites provided the equivalent average accuracy when compared with the current suite of microsatellites and MHC. The six highest rated microsatellites provided the equivalent average stock identification resolution and individual assignment accuracy compared with the 46 highest rated SNPs. For regional stock compositions, 53-104 SNPs were projected to be required to provide accuracy and precision equivalent to the microsatellites. For population-specific stock compositions, 75-79 SNPs were projected to be required to provide accuracy and precision equivalent to the microsatellites. Equivalency in individual assignment accuracy to region was estimated to require 100 SNPs of the quality evaluated in the study, whereas equivalent accuracy in assignment to specific populations was estimated to require 124 SNPs. Applications that incorporate the existing power of a combined microsatellite-SNP approach are the best current technique available for sockeye salmon stock identification applications in southern British Columbia.

Keywords: genetic stock identification, microsatellites, SNPs, sockeye salmon

Introduction

Management of salmon fisheries targeting Fraser River sockeye salmon (*Oncorhynchus nerka)* in southern British Columbia and Washington is reviewed and directed by the Fraser River Panel of the Pacific Salmon Commission (PSC), with the Panel composed of fishers, managers, and assessment staff of both Canada and the United States. There are two main sources of information required in the determination of weekly management decisions, namely the abundance of sockeye salmon that is present, and the stock composition of that abundance (Woodey 1987). Genetic stock identification (GSI) is an important part of the management process, and results are used to assess stock-specific timing and abundance patterns. Daily and weekly estimates of sockeye salmon abundance and stock composition from test fisheries are used to manage fisheries so that harvest and escapement goals can be achieved. Prior to 2009, GSI of sockeye salmon in fisheries in this region was conducted with a set of 14 microsatellites and one major histocompatibility complex (MHC) locus (Beacham et al. 2004a), with accuracy of estimates confirmed by a number of blind sample tests (Beacham et al. 2004b). Management of Fraser River sockeye salmon requires that stock composition estimates be available within an 8-24 hour period after sample arrival in the laboratory (Beacham et al. 2004a). Estimates of stock composition derived solely from microsatellites are available within an 8-hour period, but incorporation of MHC variation requires a 24-hour analysis. Rapid estimates of stock composition are required for incorporation into in-season management decisions, with estimates of stock composition based upon only microsatellites sometimes used in decisions, with subsequent incorporation of MHC variation into the estimates the following day generally confirming decisions reached the previous day. Final estimates of stock composition available the same day as sample delivery would be beneficial. Incorporation of markers for Fraser River sockeye salmon GSI applications require stock composition estimates to the individual population

or tributary in many cases, as well as identification of individuals to specific populations or tributaries (Cooke et al. 2006; Young et al. 2006; Macdonald et al. 2009).

Although application of the 14 microsatellites and the MHC locus meet current resolution and cost requirements, there may be other stock identification techniques available that may provide advantages of increased resolution at lower cost per individual. The most widely promoted current technique is the application of single nucleotide polymorphisms (SNPs). The benefits of applying SNPs relative to microsatellites were suggested to be ease of data standardization among laboratories, high throughput, high diversity among populations, low genotyping errors, and low cost of analysis per individual (Smith et al. 2005a, b). Indeed, a panel convened to evaluate the coded-wire tagging program in Pacific salmon recommended that "the Pacific Salmon Commission support an immediate evaluation of the coordinated transition for all salmon species from genetic stock identification (GSI) based on the use of microsatellite markers to GSI based on single nucleotide polymorphism (SNP) markers" (Pacific Salmon Commission 2005). As a functioning GSI program for sockeye salmon in Washington and British Columbia based upon microsatellites and MHC variation predated this report (Beacham et al. 2004a, b), but SNP variation had not been investigated for Washington or British Columbia sockeye salmon, it was not possible to conduct an immediate evaluation of a proposed transition for GSI applications from microsatellites to SNPs. Accordingly, there was no scientific consensus as to the preferred approach to apply in GSI of Pacific salmon. Subsequently, a report derived from two GSI-based workshops recommended that "the PSC should continue to support microsatellites as a demonstrated coast-wide tool for GSI for Chinook salmon and continue to support the development of SNPs. Both microsatellites and SNPs have demonstrated capabilities in estimation of stock composition. The value of SNPs has been clearly demonstrated at a regional level, but the effectiveness of a coast-wide application of SNPs remains to be explored." (Pacific Salmon Commission 2008). The report noted that "Additional comparisons of current and future costs and performance of microsatellites and SNPs will be required before an objective selection of a single approach might be justified."

SNPs are currently applied to estimate stock compositions of sockeye salmon fisheries in Alaska. A suite of 45 SNPs has been used to estimate stock compositions of sockeye salmon in Cook Inlet fisheries (Habicht et al. 2007), and the authors indicate that this same suite of SNPs has been used in GSI applications for management of Bristol Bay sockeye salmon, as well as in the Northern Boundary area (area adjacent to the international border) in southeast Alaska. In the Northern Boundary area, fisheries intercept sockeye salmon bound for both Southeast Alaskan and Canadian rivers, and managers and assessment staff from both the United States and Canada need to be satisfied that reliable estimates of stock composition are provided by GSI applications. The 14 microsatellites and MHC locus have been previously tested and demonstrated to provide high resolution in stock composition estimates for fisheries in the area (Beacham et al. 2005a), and this is the technology applied in GSI of Northern Boundary sockeye salmon fisheries on the Canadian side of the border. SNPs are applied on the American side of the border (in addition to scale pattern analysis), but a definitive test of stock identification resolution provided by SNPs has yet to be conducted.

Before conclusions can be drawn on the efficacy of microsatellites and SNPs for sockeye salmon GSI applications, comparisons between the two techniques must obviously be conducted. The focus of any evaluation would be the resolution of stock composition estimates provided by the two techniques and the cost per individual required to obtain the observed resolution. The key question to answer is how many SNPs must be used to provide stock composition estimates of equivalent quality both in terms of accuracy and precision when compared with estimates produced with an existing microsatellite baseline. Alternatively, if the resolution required for existing management applications is satisfactorily determined from a suite of SNPS, how many microsatellites must be analyzed in order to provide comparable resolution of stock composition both in terms of accuracy, precision, and individual assignment. Initial comparisons of the efficacy of microsatellites and SNPs for stock and individual identification of Chinook salmon (*O. tshawytscha*) were reported by Beacham et al. (2008) and Narum et al. (2008), but no comparison has been conducted for sockeye salmon.

In the current study, we evaluate the utility of using variation at 14 microsatellite loci and a MHC locus for stock identification applications for 44 populations of sockeye salmon ranging from central coast British Columbia to the Fraser River, with emphasis on Fraser River populations. We also evaluate the accuracy and precision of estimates of stock composition derived from a suite of 49 SNPs (the 45 SNPs outlined by Habicht et al. (2007) plus 4 SNPs indentified in our laboratory) from the same 44-population baseline. The key questions investigated were: 1) How many SNPs are required to replace the MHC locus in GSI applications for Fraser River sockeye salmon with no loss of average population resolution?; 2) How many microsatellites are required to provide equivalent stock identification resolution to that of the 49 SNPs?; and 3) How many SNPs are required to replace the current 14-microsatellite baseline used in GSI applications?

Materials and Methods

Collection of DNA samples and laboratory analysis

Tissue samples were collected from mature sockeye salmon from 44 populations ranging from the central coast of British Columbia to the Fraser River drainage, with the majority of the populations surveyed from the Fraser River (Table 1, Fig. 1). Samples were generally preserved in 95% ethanol and sent to the Molecular Genetics Laboratory at the Pacific Biological Station. Historical scale samples were occasionally accessed as well. DNA was extracted from the tissue or scale samples using a variety of methods, including a chelex resin protocol outlined by Small et al. (1998), a Qiagen 96-well Dneasy® procedure, or a Promega Wizard SV96 Genomic DNA Purification system. Once extracted DNA was available, variation at microsatellite, MHC, and SNP loci was surveyed. For the microsatellites, polymerase chain reaction (PCR) products at 14 microsatellite loci: *Ots2*, *Ots3* (Banks et al. 1999), *Ots100*, *Ots103, Ots107, and Ots108* (Beacham et al. 1998; Nelson and Beacham 1999), *Oki1* (two loci), Oki6, *Oki10, Oki16,* and

Oki29 (Smith et al. 1998), *One8* (Scribner et al. 1996), and *Omy77* (Morris et al. 1996) were initially size fractionated in denaturing polyacrylamide gels using an ABI 377 automated DNA sequencer, and genotypes were scored by Genotyper 2.5 software (Applied Biosystems, Foster City, CA) using an internal lane sizing standard. Later in the study, microsatellites were size fractionated in an ABI 3730 capillary DNA sequencer, and genotypes were scored by GeneMapper software 3.0 (Applied Biosystems, Foster City, CA) using an internal lane sizing standard. Allele scores derived from Genotyper or GeneMapper were verified by either one or two laboratory personnel. Repeatability of genotyping was evaluated by repeat PCR analysis and scoring of approximately 500 fish. Discrepancies in scoring were observed in 21 genotypes of 6720 genotypes scored across all loci, for a genotyping error rate of 0.31%. Allele identification between the two sequencers were standardized by analyzing approximately 600 individuals on both platforms and converting the sizing in the gel-based data set to match that obtained from the capillary-based set.

Genetic variation at the MHC class II *DAB*-β1 locus (Miller et al. 2001) was surveyed by denaturing gradient gel electrophoresis (DGGE). β1 alleles were separated by DGGE with the Bio-Rad (Hercules, CA) D Gene™ or D Code™ electrophoresis systems, with conditions determined by the methods of Miller et al. (1999). Fluorescently-multiplexed (FM)-DGGE (Miller et al. 2000) was used in the population survey, with scoring of genotypes conducted by one individual.

Variation was analyzed at 49 SNPS (46 nuclear, 3 mitochondrial), with the 45 SNPs outlined by Habicht et al. (2007) supplemented by an additional 4 SNPs. SNPs added by our laboratory were: *One-Hsp47* primer sequences F: CGT TCA AAT AAA TGC TGT TTG GCC TTT, R: GTG GTG TTC GGA TTT TTC CTG AAA, probe sequences: VIC-TTA TTG ACT ATG GCA CAT TG, FAM-TTG ACT ATG GCG CAT TG; *One-PIP1* primer sequences F: ACA GAG TCA GGA CTT GAT ATG TAC AGA**,** R: CCT GAC GAG GGT CTA CTA CAC T**,**

probe sequences: VIC-AAC ACA CAT TTC TCA ACA CA, FAM-ACA CAC ATT TTT CAA CAC A; *One-SEPP* primer sequences F: GGA TGG AGG GAG CAA AAC AGA, R: ACC CTC CCA GGA GTG TAT GG, probe sequences: VIC-AAA CGT TAA CAT AAT CAC C, FAM-AAA CGT TAA AAT AAT CAC C; and *One-C1259* primer sequences F: GCA TTT CTC AAC CCC ACA ACA C, R: CCT ATT TAG GGC TCT ATT TCT GTG AG, probe sequences: VIC-ACT GGG AAC TAT TTA AAA, FAM-CTG GGA ACT ATT AAA AA. The listing of all SNPs surveyed is outlined in Table 2. After PCR amplification, the plates (384-well) were read on an ABI Prism 7900HT Sequence Detection System by one individual using Sequence Detection software from ABI. Repeatability of genotyping was evaluated by repeat PCR analysis and scoring of genotypes, with discrepancies in scoring observed in 21 genotypes of 5320 genotypes scored across all loci, for a genotyping error rate of 0.39%.

Estimation of stock composition in single-population samples

Two software packages were utilized in estimation of stock composition of single-population mixtures: Statistical Package for the Analysis of Mixtures software program (SPAM version 3.7) (Debevec et al. 2000) and ONCOR (Kalinowski et al. 2007). Three mitochondrial (mt) DNA SNPs were analyzed in the survey, and as ONCOR is currently unable to analyze variation incorporating mitochondrial haplotypes, SPAM was used exclusively for those analyses incorporating mitochondrial SNPs. Genotypic frequencies were determined for each locus in each population and were used to estimate stock composition of simulated single-population samples. The Rannala and Mountain (1997) correction to baseline allele frequencies was used for SPAM analyses in order to avoid the occurrence of fish in the mixed sample from a specific population having an allele not observed in the baseline samples from that population. All diploid loci were considered to be in Hardy-Weinberg equilibrium (microsatellites (Beacham et al. (2006a), SNPs (Beacham et al. unpub. data)), and expected genotypic frequencies were

determined from the observed allele frequencies. Reported stock compositions for simulated single-population samples are the bootstrap mean estimate of each mixture of 200 fish analyzed, with mean and variance estimates derived from 100 bootstrap simulations. Each baseline population and simulated single-population sample was sampled with replacement in order to simulate random variation involved in the collection of the baseline and fishery samples. When ONCOR was used to estimate stock compositions, the Rannala and Mountain (1997) correction to baseline allele frequencies was again implemented, with precision of the stock composition estimates calculated by bootstrapping over observed baseline population sample sizes and a mixture size of 200 fish. For both SPAM and ONCOR, allocations to individual baseline populations were summed to provide estimates of stock compositions for regional stock groups (Table 1). Additionally, ONCOR was used to provide estimates of accuracy of identification of individuals to specific populations or regional stock groups.

The effect of population sample size on average accuracy of estimated stock compositions for both population-specific and regional estimates, as well as individual assignment accuracy to population and region, was investigated for SNPs with ONCOR for population sample sizes ranging from 20, 40, 60, 80, and 95 individuals, with samples of the appropriate size created by sampling existing population samples without replacement. Mean accuracy for each population sample size was determined as the average estimate across all 44 populations.

Relative ranking of the loci

The power of individual loci for stock composition estimation was initially evaluated by incorporating only a single locus for estimation of stock composition of single-population samples. As mtSNPs were included in the analysis, only SPAM was used to provide estimates of stock composition for all 44 single population samples. Mean accuracy was determined as the average estimate across all 44 populations, with loci then ordered from the most accurate to the

least accurate (Table 2). Weir and Cockerham's (1984) F_{ST} estimates for each locus over all populations were calculated with FSTAT version 2.9.3.2 (Goudet 1995). Allele frequencies for all populations surveyed in this study are available at the Molecular Genetics Laboratory website at http://www-sci.pac.dfo-mpo.gc.ca/mgl/default_e.htm

How many SNPs for MHC equivalency?

SPAM and ONCOR were used to assess the number of SNPs that would be required to be analyzed in order to maintain the stock identification resolution provided by the MHC locus. The 14 microsatellites and the MHC locus were used to estimate stock composition of 44 single-population samples as described previously. Accuracy and standard deviation were then averaged over the 44 single-population simulations to provide the baseline case. Next, individual SNPs (starting with the locus with highest mean population accuracy) were added to the suite of 14 microsatellites, and average resolution of stock compositions determined. Additional SNPs were then added until the average population accuracy and precision of stock composition estimates provided by the suite of microsatellites and SNPs matched that provided by the microsatellites and MHC locus.

How many microsatellites for SNP equivalency?

ONCOR was used exclusively for this analysis, with the proviso that the three SNPs of mtDNA origin were eliminated from the analysis, as they provided little power for GSI applications. The microsatellite locus with the highest average population accuracy was initially incorporated into the analysis, and lower accuracy microsatellites were sequentially added to the analysis until the average population accuracy and precision of stock composition estimates provided by the suite of microsatellites matched that provided by the 46 SNPs. Single-population samples were analyzed with the best 3, 4, 5, 6, 8, 10, 12, and 14 microsatellites. Accuracy and

standard deviation for population- and region-specific estimates of stock composition were determined, and were then averaged over the 44 single-population simulations for each set of loci. Percentage of individuals correctly assigned to specific populations and regions was also determined for all 44 populations, which was then averaged over all populations for each set of microsatellites evaluated.

How many SNPs for microsatellite equivalency?

Projections of the number of number of SNPs required for equivalency of the current microsatellite baseline were conducted by ranking SNPs according to the average accuracy observed in estimation of stock composition for single-population samples over the 44 populations surveyed. The mtDNA SNPs were eliminated in the analyses. Subsequent analyses were conducted exclusively with ONCOR. Single-population samples were analyzed with 10, 15, 20, 25, 30, 35, 40, and 46 SNPs. SNPs with the highest average accuracy were initially incorporated in the analyses of the single-population mixtures, with lower-accuracy SNPs sequentially added to the analyses. Average population accuracy and precision were recorded for each set of SNPs. Additionally, SNPs with the lowest average accuracy values were initially incorporated in the analyses, with progressively higher-accuracy SNPs sequentially added to the analyses, with again average accuracy and precision recorded. Overall mean accuracy and precision for each specified number of SNPs were determined by averaging the results from both processes, and this was considered indicative of the average trend in estimating accuracy and precision when the number of SNPs employed in the analysis was increased. Average regional, population, and individual accuracy and precision over all 44 populations were recorded for each set of SNPs. A hyperbola function of the form $Y= a/X + b$ was fitted with Labfit curve fitting software (Silva and Silva 2007), with the number of SNPs incorporated in the analysis as the independent variable X, and mean observed accuracy for population and regional estimates as the dependent variable Y. Estimates of standard deviations were fitted with an inverse straight line

equation Y= 1/(aX + b), with X and Y defined as previously. Population and regional individual assignment accuracy were fitted with an inverse hyperbola equation Y= X/(a+bX), with X and Y defined as previously. Projections were then made with these regression models to estimate the number of SNPs that would be required to provide estimates of comparable resolution to that provided by the microsatellites with respect to estimated stock compositions at both the regional and population level, as well as individual identification to the regional and population level.

Results

Population sample size

Average realized population sample sizes were 97 individuals successfully genotyped at SNP loci, 254 individuals at the MHC locus, and 292 individuals at the microsatellite loci (Table 1). Sample size variation was limited at SNPs, ranging from an average of 71 to 370 individuals successfully genotyped. Sample size variation was more variable at the MHC locus, ranging from 31 to 785 individuals genotyped. The most extensive sample size variation was observed at the microsatellites, ranging from 85 to 1687 individuals genotyped at individual populations. The most extensive population survey was conducted for the Cultus Lake population, as this is the population of highest conservation concern in the Fraser River drainage.

For SNPs, 97 individuals were on average genotyped to estimate the population frequency of one independent allele, with 254 individuals genotyped to estimate population frequency of 11 independent alleles for the MHC locus, and with 292 individuals genotyped to estimate population frequency of an average of 29 independent alleles at microsatellites. Once approximately 95 individuals had been sampled for SNP variation within a population, there was virtually no increase in accuracy of estimated stock compositions or individual assignment (Fig. 2). Thus, even though the number of individuals genotyped was the smallest for SNPs,

intermediate for MHC, and greatest for microsatellites, variation in population sample size should have no marked effect on the relative performance of the different marker classes.

Relative ranking of the markers

Average population-specific accuracy for the 64 markers evaluated ranged from 81.2% to 2.3%, with a clear break in accuracy observed between microsatellites and SNPs. The top 15 markers of the 64 markers evaluated were either microsatellites or MHC, with accuracy of estimated population-specific stock compositions produced from incorporating the worst microsatellite (*Oki1b*) approximately double that of the highest ranked SNP (*One-MHC2-190*) (34.6% versus 16.0%)(Table 2). Similar results were observed for region-specific accuracy (45.4% versus 27.7%) for the two loci. F_{ST} value was a poor predictor of the relative value of each marker for stock composition analysis. Markers with higher F_{ST} values, such as the MHC β1 locus (0.245) and the SNP locus *One-MHC2-190* (0.307) did not provide as much population-specific resolution as did nine microsatellites with lower F_{ST} values (0.028-0.141) (Table 2). Two SNPs with F_{ST} values > 0.25 were not among the top five SNP loci evaluated, even though the F_{ST} values of the top five SNPs ranged from 0.21-0.31. For *One-RH2OP*-395, allele frequencies were markedly different from the norm (dominant allele frequency 0.89-1.00) for only one population (Cultus Lake, allele frequency 0.34), and for *One-STC-410*, allele frequencies were markedly different from the norm (dominant allele frequency 0.74-1.00) for three central coastal populations (Devon Lake, Kent Lake, and Tankeeah River, allele frequency range 0.38-0.47). These four populations were already well differentiated from other populations at other loci (Tables 1, 3), so no loss of stock composition resolution was incurred by not incorporating these two SNP loci in the group of the top five SNPs.

Heterozygosity was also a poor predictor of the relative power of the relative power of each marker for stock composition analysis. For microsatellites, there were only minor differences in accuracy of estimated stock compositions for loci with a heterozygosity value >

0.60 (Fig. 3A). Similarly, there were only minor differences among SNPs for loci with a heterozygosity value > 0.10 (Fig. 3B). The mtDNA SNP markers were the poorest of all markers surveyed in terms of providing accurate population-specific estimates of stock composition (Table 2).

How many SNPs for MHC equivalency?

Average stock composition of the single-population samples incorporating 14 microsatellites and the MHC locus was 92.4% when estimated with SPAM and 85.8% when estimated with ONCOR. These levels of accuracy provided the reference points when evaluating the equivalency between SNPs and MHC when used in conjunction with microsatellites to estimate sockeye salmon stock composition. Sequential addition of the five highest rated SNPs to the suite of 14 microsatellites provided the equivalent average accuracy to analyses with microsatellites and MHC (SPAM 92.3%, ONCOR 86.0%) (Table 1). Two of the top 5 SNPs were derived from MHC sequences.

Population-specific accuracy differed only modestly between SPAM and ONCOR when the populations were well defined, as populations estimated with greater than 90% accuracy in SPAM were generally estimated with greater than 90% in ONCOR. However, for those populations that were less well defined (SPAM estimate less than 90%), estimated stock compositions derived from SPAM were consistently higher than those derived from ONCOR (Table 1).

How many microsatellites for SNP equivalency?

The application of 46 SNPs to estimate stock composition of 44 single-population samples resulted in an average accuracy of 83.5% to population (Table 3). Individual population estimates ranged from 54.3% for the Takla Lake population to 99.9% for the Cultus Lake population. Starting with the microsatellite with highest observed population-specific accuracy

(*Ots100*), an average level of accuracy of 82.9% was achieved by incorporating the best six microsatellites. There was no significant difference in population-specific accuracy estimated from the 46 SNPs or six microsatellites, with accuracy of estimated stock compositions derived from these six microsatellites greater than those derived from the 46 SNPs for 20 of the populations and less than for 24 populations (sign test analysis, $P>0.10$). The average population-specific accuracy for estimates incorporating all 14 microsatellites was 85.5%. Estimates produced by incorporating 14 microsatellites were significantly more accurate than those produced incorporating 46 SNPs, with microsatellite-based estimates greater than those derived from SNPs for 32 populations and less than for 12 populations ($P<0.01$). Some populations such as Cultus Lake, Chilliwack Lake, and Gates Creek were clearly more differentiated than others with high levels of population-specific accuracy (>97%), regardless of what suite of markers was used to estimate stock composition. Similarly, estimated stock compositions for populations that displayed lower levels of accuracy, such as Lower Horsefly River, were consistently lower with both microsatellites and SNPs.

The application of 46 SNPs for estimation of standard deviation in 44 single-population samples resulted in an average standard deviation of 4.4% to population (Table 4). Individual population estimates ranged from 0.1% for the Cultus Lake population to 11.7% for the Takla Lake population. The average standard deviation of the estimated stock compositions for the top five microsatellites was 4.2%. There was no significant difference in standard deviation estimated from the 46 SNPs or five microsatellites, with standard deviations of estimated stock compositions derived from these five microsatellites less than those derived from the 46 SNPs for 24 of the populations and greater than for 20 populations (sign test analysis, $P>0.05$). The average population-specific standard deviation for estimates incorporating all 14 microsatellites was 3.2%. Estimates produced by incorporating 14 microsatellites were significantly more precise than those produced incorporating 46 SNPs, with microsatellite-based standard deviations

less than those derived from SNPs for 42 populations and greater than for two populations ($P<0.01$).

The application of 46 SNPs to assignment of individual fish resulted in an average correct assignment of 50.8% to population and 74.4% to region (Table 4). Population assignment accuracy ranged from 10.1% for the Seymour Creek population to 98.4% for the Cultus Lake population (Table 3). Regional assignment accuracy ranged from 32.9% for the North Thompson River population to 100.0% for the Kent Lake population. Comparable population-specific levels of assignment accuracy were achieved with an application of five microsatellites (52.5%), with comparable levels of regional assignment accuracy achieved also with five microsatellites (70.1%) (Table 4). In summary, the application of the best six microsatellites produced levels of population-specific and regional accuracy and precision for estimates of stock composition, as well as levels of correct individual assignment to population and region, comparable with those produced by application of the best 46 SNPs.

How many SNPs for microsatellite equivalency?

Average stock composition accuracy of the single-population samples incorporating 14 microsatellites was 85.53% to a specific population and 99.20% to geographic region (Table 5). Standard deviations of the estimates were 2.94% and 0.61%, respectively. Average accuracy of individual assignment was 67.71% to specific population and 91.48% to region. These levels of accuracy, precision, and individual assignment provided the reference points for estimation of the number of SNPs required to provide equivalent results in stock identification applications as derived with the microsatellites. The estimated number of SNPs required to provide equivalent results to those of the microsatellites depended upon the complexity of the task. The easiest task, estimation of regional stock composition, was estimated to require 53 SNPs of the average quality evaluated in the study to provide equivalent levels of accuracy as that observed with the microsatellites (Table 6; Fig. 4). However, equivalent precision in regional estimates of stock

composition was estimated to require 104 SNPs of the quality evaluated in the study (Table 6, Fig. 4). Thus, 53-104 SNPs were projected to be required to provide accuracy and precision of estimated regional stock compositions equivalent to the microsatellites.

The next most difficult task was determination of population-specific estimates of stock composition. Equivalent accuracy in population-specific estimates of stock composition was estimated to require 75 SNPs of the quality evaluated in the study, whereas equivalent precision was estimated to require 79 SNPs (Table 6, Fig. 4). Thus, 75-79 SNPs were projected to provide accuracy and precision of estimated population-specific stock compositions equivalent to the microsatellites.

The most difficult task in stock identification applications is the assignment of individuals to specific groups, with correct assignment to a region obviously easier to obtain than correct assignment to an individual population within a region. Equivalency in assignment accuracy to region was estimated to require 100 SNPs of the quality evaluated in the study, whereas equivalent accuracy in assignment to specific populations was estimated to require 124 SNPs (Table 6; Fig. 4). Clearly, the number of SNPs required to provide results of a quality equivalent to that of the existing suite of microsatellites depends upon the complexity of the application.

Discussion

Sample size

SNPs typically display two alleles at a locus, but if the frequency of one allele is determined, then the frequency of the other allele is known as well. Thus, SNPs display one independent allele per locus. The other types of markers surveyed in the study (MHC and microsatellites) display multiple independent alleles per locus, with the microsatellites displaying up to 84 independent alleles per locus. Comparison of the efficacy of SNPs and microsatellites for stock identification applications requires that adequate sample sizes be available for

populations included in the analyses for both classes of markers. Clearly, highly polymorphic microsatellites require more fish to be sampled in a population to obtain estimates of allele frequencies with similar accuracy and precision compared with SNPs, where the frequency of only one independent allele is estimated. Beacham et al. (in review) demonstrated that microsatellites required larger baseline sample samples than SNPs in order to reduce sampling variation in estimation of allele frequencies and thus increase accuracy of estimated stock compositions. Population sample size in the microsatellite baselines was required to be about two to three times larger than in the SNP baselines before equivalent levels of accuracy relative to the asymptotic value were obtained. Therefore, in comparing the effectiveness of SNPs and microsatellites for stock identification estimation for Pacific salmon, larger baseline population sample sizes will be required for microsatellites than for SNPs for all species.

Studies of salmonid population structure and stock identification incorporating nuclear SNPs typically include sample sizes of approximately 95 individuals per population (Smith et al. 2005b; Smith and Seeb 2008; Narum et al. 2008), and our study utilized the same experimental approach. Simulation modeling showed that once approximately 95 individuals had been sampled at SNP loci within a population, there was virtually no increase in accuracy of estimated stock compositions or individual assignment. Thus accuracy of estimated stock compositions and assignment of individuals derived from SNPs was not limited by population sample size in our study. Estimates of stock composition of single-population samples of salmon derived from microsatellites can be limited by population sample size, with population sample sizes of approximately 200 individuals required before there is little effect of sample size on accuracy of population-specific estimates (Beacham et al. 2006c). However, as population sample sizes for microsatellites of most populations in the study were greater than 200 fish, accuracy of estimated stock compositions and assignment of individuals derived from microsatellites was also not significantly limited by population sample size in our study (Beacham et al. in review).

Relative ranking of the markers

Several methods are available to measure the information content of genetic markers for population differentiation and individual assignment (Rosenberg et al. 2003; Hedrick 2005). The basis for much of the work on genetic variation in Pacific salmon is to develop a baseline for stock identification application. As such, our measure of the relative ranking of markers was based solely upon their observed accuracy in stock identification. Narum et al. (2008), in a study evaluating the effectiveness of 13 microsatellites and 37 SNPs in 29 Chinook salmon populations, reported that the best 10 loci for correct individual assignment were microsatellites, and that the best 15 loci included 12 microsatellites and three SNPs. The top 15 markers of the 64 markers evaluated in our study were either microsatellites or MHC, with accuracy of estimated population-specific stock compositions produced from incorporating the least informative microsatellite approximately double that of the highest ranked SNP. The number of alleles observed at each microsatellite locus ranged from 7 to 85 alleles, with an average of 30 alleles per locus. Microsatellites containing only a few alleles (<5) were not included in our suite analyzed, as we recognized early in locus selection that such loci provide limited stock resolution. This was illustrated by Beacham et al. (2005c), where the number of alleles observed at a locus was generally related to the power of the locus in providing accurate estimates of stock composition. Thus, the microsatellites included in the current suite did undergo some initial evaluation in the 1990s, but the 14 microsatellites in the current baseline were simply 14 microsatellites that initially displayed enough polymorphism to be considered for inclusion in a suite of microsatellites, were in Hardy-Weinberg equilibrium in test populations, and now fit on two injections for an automated DNA sequencer. The current suite was first used for individual assignment to three lakes in Barkley Sound on the west coast of Vancouver Island (Beacham et

al. 2002), and for in-season stock identification of Fraser River sockeye salmon in 2002 (Beacham et al. 2004a).

How many SNPs for MHC equivalency?

One of the objectives of the current study was to evaluate the number of SNPs that were required to replace the MHC locus in GSI applications for Fraser River sockeye salmon with no loss of average population resolution. Replacement of the MHC locus with SNPs allows more synchronous availability of laboratory results with the microsatellites. The best five SNPs were demonstrated to provide equivalent GSI resolution to the MHC locus, and these SNPs were chosen to replace the power provided by MHC in Fraser River sockeye salmon GSI applications. Although an all-SNP approach to Pacific salmon GSI applications was implied by PSC (2005), this change would currently appear to be of questionable merit at this time. Instead, an approach that harnesses the existing power of a combined microsatellite-SNP approach would seem to be the best current technique available for Fraser River sockeye salmon GSI applications. Inclusion of both microsatellites and SNPs in a suite of loci for Chinook salmon individual assignment was also the approach recommended by Narum et al. (2008).

Stock compositions of simulated single-population samples displayed little difference between those estimated by SPAM and ONCOR when populations were well differentiated from other populations. However, among closely related populations, estimates derived from SPAM were higher than those estimated with ONCOR. Anderson et al. (2008) reported that stock compositions derived from SPAM can overestimate the predicted accuracy of stock compositions for closely related populations relative to estimates derived from ONCOR, and the results from the current study support this conclusion.

Microsatellites or SNPs?

If better stock ID resolution is required for southern British Columbia sockeye salmon in the existing SNP and microsatellite databases, is it more effective to add SNPs or microsatellites? Morin et al. (2009) suggested that SNPs may offer significant advantages over microsatellites, including lower error genotyping rates, less effort in data standardization, and technologies for high throughput genotyping. Smith and Seeb (2008) suggested that comparable amounts of effort spent on developing SNP and microsatellite baselines will result in SNP baselines with greater information content per allele. In most applications, SNP alleles will likely have a higher information content per allele. In the current study, the top six microsatellites provided equivalent power to the 49 SNPs evaluated. The top six microsatellites displayed 245 alleles, whereas the 49 SNPs displayed 98 alleles (49 independent), illustrating that SNP alleles provided a greater information content per allele. While the relative power of individual alleles is of some interest, a more salient, practical concern is the number of loci for each class of markers that must be analyzed in order to provide comparable results. The number of loci required in each class of markers for comparable stock identification results is vital to determine, as this will largely determine the cost per individual in subsequent analysis. Vignal et al. (2002) suggested that with the progress being made in identifying and surveying SNP variation, "the effort needed to produce an equivalent amount of information as with microsatellites may some day be equivalent." The question evaluated in the current study was whether that day has arrived for sockeye salmon stock identification in British Columbia.

The results of the current study indicated that additional SNPs will be required to be surveyed in order to provide the equivalent stock identification power as the existing microsatellites. Considerable effort would be required for SNP discovery in order to enhance the

baseline evaluated in our study, as all known reported SNPs for sockeye salmon were incorporated in the survey. For microsatellites, the worst performing locus provided twice the average population-specific mean accuracy than the best performing SNP. The current suite of microsatellite loci was developed about 10 to 15 years previously, and there are now literally hundreds of microsatellites available for salmonids. Many microsatellites are easily transferrable among species, and a broad range of microsatellites could be rapidly screened in test populations. Microsatellites specifically chosen for their stock separation of southern British Columbia sockeye salmon could then be incorporated into existing SNP or microsatellite baselines. If additional stock identification resolution is required, the cost of incorporating a third injection on an automated DNA sequencer would have to be compared with the costs of surveying an unknown number of SNPs above the 124 currently estimated as required to provide equivalency with the power of the existing suite of microsatellites.

Suppose that sufficient resolution in stock identification estimates is provided for some applications by the 45 SNPs in the suite used in applications in Alaska. If so, the analysis presented herein suggests that equivalent resolution will likely be provided by applying the best six microsatellites currently used in applications in British Columbia and northern Washington. However, it is uncertain if the existing set of SNPs has same relative population separation power in other areas of the species distribution, such as in Russia. New SNPs may have to be incorporated for stock identification applications in those regions. The 14 microsatellites analyzed in the current study have already been demonstrated to be of considerable value in stock composition estimation in Russia (Beacham et al. 2006b), southeast Alaska and northern British Columbia (Beacham et al. 2005a), central British Columbia (Beacham et al. 2005b), southern British Columbia and Washington (Beacham et al. 2004a, b), as well as on a Pacific Rim basis (Beacham et al. 2005c). The current analysis indicated that up to 124 SNPs will be required to equal power of stock identification applications centering on Fraser River sockeye salmon available from the 14-microsatellite baseline. Pacific Rim stock identification applications

conducted solely with SNPs may require more than the 124 SNPs estimated to provide equivalency for Fraser River sockeye salmon applications, possibly in excess of 150 SNPs may be required in order to provide the same degree of resolution available at the individual identification level as available with the 14 microsatellite baseline.

Will discovery of a number of new SNPs in sockeye salmon over and above those evaluated in this study fundamentally change the relative power of microsatellites and SNPs for stock identification? In human genomics research, over three million SNPs have been discovered (International HapMap Consortium 2007). Advocates of a SNP-only method of stock identification in fisheries frequently claim that advancements in human genomics research will render application of SNPs the preferred approach in fisheries stock identification. This view was typified by PSC (2005), where it was noted that "the human genetics community has already undergone a transition from microsatellite to SNP markers". The implication is that advances are made first in human genetics research, which are then applied by investigators in other fields such as fisheries stock identification. Gill et al. (2004), outlining the consensus of a working group of European and American forensic scientists, assessed whether SNPs will replace STRs (microsatellites) in human forensic investigations. They noted "It is easy to be swept along in the tide of new technology, simply because it is something different. However, we must take a step back and objectively evaluate the reasons for implementing a new system. New does not necessarily mean better. We should remember that SNPs were initially developed for forensic application in the late 1980s and early 1990s, but were superseded by STRs (microsatellites) because of their marked superiority – the discussion is not new." In an evaluation of microsatellites (STRs) and SNPs for human forensic applications, Butler et al. (2007) concluded that "we do not feel that SNPs stand on the horizon as future markers (i.e., replacing STRs) for widespread use in forensic DNA testing". In spite of a suite of over three million SNPs from which to chose, human forensic investigators concluded that microsatellites will be the dominant technology employed for the foreseeable future. If indeed human applications provide the path

for future fisheries stock identification, then SNP-only applications in fisheries GSI work may be best reserved when costs of application are lower, given the number of SNPs that must be analyzed in order to provide GSI resolution comparable to that of existing suites of microsatellites.

Butler et al. (2007) noted that several significant disadvantages exist with SNP markers when considered as a possible replacement for the currently used set of 13-15 microsatellites for human forensic applications. These disadvantages are directly analogous to fisheries stock identification applications. Because SNPs are not as polymorphic as microsatellites, more SNPs are required to provide equivalent powers of discrimination, they have less ability to decipher mixtures, and the expense of screening more SNPs for each sample will be higher as compared with microsatellites (Butler et al. 2007). The microsatellites employed in human forensic studies are routinely amplified in a single multiplex amplification reaction. Cost of laboratory analysis for an individual fish is a key issue in deciding the appropriate technology to apply in a particular laboratory. In our laboratory, after DNA has been extracted from a sample, the laboratory cost of analysis incurred for sockeye salmon incorporates eight polymerase chain reaction (PCR) amplifications of the 14 microsatellites, and then sizing the amplified products on two injections on an automated DNA sequencer. For this technology to be replaced as was implied by PSC (2005), it must be demonstrated that SNPs provide either equivalent performance at a cheaper analytical cost per fish than that obtained with microsatellites, or enhanced stock identification capability at the same analytical cost per fish. To date, much like forensic applications in human genetics, a SNP-only approach to sockeye salmon GSI applications fails to meet either requirement. Technologies are replaced when one technique provides a clear advantage over the other, such as replacement of allozymes by microsatellites or SNPs as the main technique employed to survey genetic variation in Pacific salmon. The estimated number of SNPs required to provide equivalent GSI resolution and individual assignment to that of the microsatellites ranged from approximately 53-124 SNPs, dependent upon the degree of difficulty of the

application. Although there exist a number of techniques available to survey SNP variation, we are aware of no technique that will allow up to 124 SNPs to be analyzed at a cost comparable to analyzing 14 microsatellites in eight PCR amplifications coupled with two injections on an automated DNA sequencer. A combination of microsatellites and a limited number of SNPs would currently appear to be best strategy employed for sockeye salmon GSI applications in British Columbia. Indeed, this was the strategy employed in applications in southern British Columbia in 2009, with 14 microsatellites and the top 5 SNPs applied in real-time estimation of sockeye salmon stock composition from test and commercial fisheries.

Acknowledgments

We thank various field staff of the PSC and DFO for collection of the samples used in baseline development. The Alaska Department of Fish and Game provided primer and probe sequences for 45 SNPs used in the survey of SNP variation. S. Latham of the PSC provided constructive comments for improving the manuscript. Funding for the SNP portion of the study was provided by the Northern Endowment Fund of the PSC, and funding for the microsatellite and MHC portion of the study was provided by the PSC and DFO. L. Fitzpatrick drafted the map of population locations.

Literature Cited

Anderson, E. C., Waples, R. S., and Kalinowski, S. T. 2008. An improved method for predicting the accuracy of genetic stock identification. Can. J. Fish. Aquat. Sci. 65: 1475-1486.

Banks, M. A., Blouin, M. S., Baldwin, B. A., Rashbrook, V. K., Fitzgerald, H. A., Blankenship, S. M., and Hedgecock, D. 1999. Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tshawytscha*). J. Heredity 90: 281-288.

Beacham, T. D., Margolis, L., and Nelson, R. J. 1998. A comparison of methods of stock identification for sockeye salmon (*Oncorhynchus nerka*) in Barkley Sound, British Columbia. N. Pac. Anad. Fish Comm. Bull. 1: 227-239.

Beacham, T.D., McIntosh, B., and MacConnachie, C. 2002. Microsatellite identification of individual sockeye salmon *Oncorhynchus nerka*) in Barkley Sound, British Columbia. J. Fish Biol. 61: 1021-1032.

Beacham, T. D., Lapointe, M., Candy, J. R., McIntosh, B. MacConnachie, C. Tabata, A., Kaukinen, K., Deng, L., Miller, K. M., and Withler, R. E. 2004a. Stock identification of Fraser River sockeye salmon (*Oncorhynchus nerka)* using microsatellites and major histocompatibility complex variation. Trans. Am. Fish. Soc. 133: 1117-1137.

Beacham, T. D., Lapointe, M., Candy, J. R., Miller, K. M., and Withler, R. E. 2004b. DNA in action: Rapid application of DNA variation to sockeye salmon fisheries management. Cons. Genet. 5: 411-416.

Beacham, T. D., Candy, J. R., McIntosh, B., MacConnachie, C., Tabata, A., Miller, K. M., and Withler, R. E. 2005a. DNA-level variation of sockeye salmon (*Oncorhynchus nerka*) in southeast Alaska and the Nass and Skeena rivers, British Columbia, with applications to stock identification. N. Amer. J. Fish. Manage. 25: 763-776.

Beacham, T.D., McIntosh, B., and MacConnachie, C. 2005b. Population structure and stock identification of sockeye salmon (*Oncorhynchus nerka*) in coastal lakes in British Columbia, Canada. Can. J. Zool. 83: 834-844.

Beacham, T. D., Candy, J. R., McIntosh, B., MacConnachie, C., Tabata, A., Kaukinen, K., Deng, L., Miller, K. M., Withler, R. E., and Varnavskaya, N. V. 2005c. Estimation of stock composition and individual identification of sockeye salmon on a Pacific Rim basis using microsatellite and major histocompatibility complex variation. Trans. Am. Fish. Soc. 134: 1124-1146.

Beacham, T. D., McIntosh, B., MacConnachie, C., Miller, K. M., Withler, R. E., and Varnavskaya, N. V. 2006a. Pacific Rim population structure of sockeye salmon as determined from microsatellite analysis. Trans. Am. Fish. Soc. 135: 174-187.

Beacham, T. D., Varnavskaya, N. V., McIntosh, B., and MacConnachie, C. 2006b. Population structure of sockeye salmon from Russia determined with microsatellite DNA variation. Trans. Am. Fish. Soc. 135: 97-109.

Beacham, T. D., Candy, J. R., Jonsen, K. L., Supernault, J., Wetklo, M., Deng, L., Miller, K. M., and Withler, R. E. 2006c. Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim using microsatellite variation. Trans. Am. Fish. Soc. 135: 861-888.

Beacham, T. D., Wetklo, M., Wallace, C., Olsen, J. B., Flannery, B. G., Wenburg, J. K., Templin, W. D., Antonovich, A., and Seeb, L. W. 2008. The application of microsatellites for stock identification of Yukon River Chinook salmon. N. Am. J. Fish. Manage. 28: 283-295.

Butler, J. M., Coble, M. D., and Vallone, P. M. 2007. STRs vs. SNPs: thoughts on the future of forensic DNA tsting. Forensic Sci. Med. Pathol. 3: 200-205.

Cooke, S. J., Hinch, S. G., Crossin, G. T., Patterson, D. A., English, K. K., Shrimpton, J. M., Van Der Kraak, G. and Farrell, A. P. 2006. Physiology of individual late-run Fraser River sockeye salmon (*Oncorhynchus nerka*) sampled in the ocean correlates with fate during spawning migration. Can. J. Fish. Aquat. Sci. 63: 1469-1480.

Debevec, E. M., Gates, R. B., Masuda, M., Pella, J., Reynolds, J. M., and Seeb. L. W. 2000. SPAM (Version 3.2): Statistics program for analyzing mixtures. J. Hered. 91: 509-510.

Gill, P., Werrett, D. J., Budowie, B., and Guerrieri, R. 2004. An assessment of whether SNPs will replace STRs in national DNA databases – Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Methods Analysis (SWGDAM). Science & Justice 44:51-53.

Goudet, J. 1995. FSTAT A program for IBM PC compatibles to calculate Weir and Cockerham's (1984) estimators of F-statistics (version 1.2). J. Hered. 86: 485-486.

Habicht, C., Templin, W. D., Willette, T. M., Fair, L. F., Raborn, S. W., and Seeb, L. W. 2007. Post-season stock composition analysis of Upper Cook Inlet sockeye salmon harvest, 2005-2007. Alaska Dept. Fish Game Fish. Man. No. 07-07. 74p.

Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59: 1633-1638.

International Hapmap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851-862

Kalinowski, S. T., Manlove, K. R., and Taper, M. L. 2007. ONCOR a computer program for genetic stock identification. Montana State University, Bozeman. Available: www.montana.edu/kalinowski/Software/ONCOR.htm. (September, 2008)

Macdonald, R. W., V. Forsland, R. E. Withler, D. A. Patterson, and A. Demsky. 2009. The use of stable oxygen isotope ($\delta^{18}O$) composition in sockeye salmon body fluid to determine whether a fish has been caught in freshwater. North American Journal of Fisheries Management 29: 560-569.

Miller, K. M., Kaukinen, K. H., Beacham, T. D, and Withler, R. E. 2001. Geographic heterogeneity in natural selection of an MHC locus in sockeye salmon. Genetica 111: 237-257.

Miller, K. M., Ming, T. J., Schulze, A. D., and Withler, R. E. 1999. Denaturing gradient gel electrophoresis (DGGE): A rapid and sensitive technique to screen nucleotide sequence variation in populations. Biotechniques 27: 1016-1030.

Miller, K. M., Ming, T. J., Schulze, A. D., and Withler, R. E. 2000. Denaturing gradient gel electrophoresis (DGGE): a rapid and sensitive technique to screen nucleotide sequence variation in populations. P. 253-258. In: J.D. Burczak and E. Mardis (eds.). Polymorphism Detection and Analysis. Eaton Publishing. Natick, MA.

Morin, P. A., Martien, K. K., and Taylor, B. L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Mol. Ecol. Res. 9: 66-73.

Morris, D. B., Richard, K. R., and Wright, J. M. 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. Can. J. Fish. Aquat. Sci. 53: 120-126.

Narum, S. R., Banks, M., Beacham, T., Bellinger, R., Campbell, M., DeKoning, J., Elz, A., Guthrie, C., Kozfkay, C., Miller, K., Moran, P., Phillips, R. Seeb, L., Smith, C. Warheit, K., Young, S., and Garza, J. C. 2008. Differentiating populations at broad and fine geographic scales with microsatellites and SNPs. Mol. Ecol. 17: 3464-3477.

Nelson, R. J., and Beacham, T. D. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. Anim. Genet. 30: 228-229.

Pacific Salmon Commission 2005. Expert Panel on the Future of the Coded Wire Tag Program for Pacific Salmon. Pac. Sal. Comm. Tech. Rep. No. 18. 230p.

Pacific Salmon Commission 2008. Recommendations for Application of Genetic Stock Identification (GSI) methods to management of ocean salmon fisheries: special report of the Genetic Stock Identification Steering Committee and the Pacific Salmon Commission's Committee on Scientific Cooperation. Pac. Sal. Comm. Tech. Rep. No. 23. 35p.

Rannala, B., and Mountain, J. L. 1997. Detecting immigration by using multilocus genotypes. Proc. Nat. Acad. Sci USA 94: 9197-9201.

Rosenberg, N. A., Li, L. M., Ward, R., and Pritchard, J. K. 2003. Informativeness of genetic markers for inference of ancestry. Amer. J. Hum. Gen. 73: 1402-1422.

Scribner, K. T., Gust, J. R., and Fields, R. L. 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic applications. Can. J. Fish. Aquat. Sci. 53: 833-841.

Silva, W. P. da, and Silva, C. D. P. S. e. 2007. Lab Fit Curve Fitting Software. Available at http://www.labfit.net. May 2009.

Small, M.P., Withler, R. E., and Beacham, T. D. 1998. Population structure and stock identification of British Columbia coho salmon, *Oncorhynchus kisutch*, based on microsatellite DNA variation. Fishery Bulletin (US) 96: 843-858.

Smith, C. T., Koop, B. F., and Nelson, R. J. 1998. Isolation and characterization of coho salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids. Mol. Ecol. 7: 1613-1621.

Smith, C. T., and Seeb, L. W. 2008. Number of alleles as a predictor of the relative assignment accuracy of short tandem repeat (STR) and single-nucleotide-polymorphism (SNP) baselines for chum salmon. Trans. Amer. Fish. Soc. 137: 751-762.

Smith, C. T., Templin, W. D., Seeb, J. E., Seeb, L. W. 2005a. Single nucleotide polymorphisms (SNPs) provide rapid and accurate estimates of the proportions of U. S. and Canadian Chinook salmon caught in Yukon River fisheries. N. Amer. J. Fish. Manage. 25: 944-953.

Smith, C. T., Seeb, J. E., Schwenke, P., and Seeb, L. W. 2005b. Use of the 5'-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. Trans. Am. Fish. Soc. 134:207-217.

Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34: 275-305.

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

Woodey, J. C. 1987. In-season management of Fraser River sockeye salmon (*Oncorhynchus nerka*): Meeting multiple objectives. P. 367-374. In H. D. Smith, L. Margolis, and C. C. Wood (editors.) Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. Can. Spec. Publ. Fish. Aquat. Sci. 96.

Young, J. L., Hinch, S. G., Cooke, S. J., Crossin, G. T., Patterson, D. A., Farrell, A. P., Van Der Kraak, G., Lotto, A. G., Lister, A., Healy, M. C., and English, K. K. 2006. Physiological and energetic correlates of an en route mortality for abnormally early migrating sockeye salmon (*Oncorhynchus nerka*) in the Thompson River, British Columbia. Can. J. Fish. Aquat. Sci. 63: 1067-1077.

Table 1. Regions, populations surveyed within regions, realized sample sizes per locus for SNPs, MHC, and microsatellites for 44 populations of sockeye salmon from 16 regions in British Columbia, as well as stock composition for single-population samples estimated for a suite of 14 microsatellites plus MHC and a suite of 14 microsatellites plus the five highest-rated SNPs, with stock composition estimated with two software packages (SPAM and ONCOR).

		Sample size			SPAM		ONCOR	
Region	Population	SNPs	MHC	Micros	Micros+MHC	Micros+5 SNPs	Micros+MHC	Micros+5 SNPs
Upper Fraser	1) Bowron Lake	85.5	200	217.1	97.82	97.68	99.79	99.56
Early Stuart	2) Takla Lake	185.6	140	283.4	81.94	81.95	64.69	65.79
Early Stuart	3) Trembleur Lake	370.1	346	490.8	90.73	91.52	79.47	80.58
Stuart/Nechako	4) Stellako River	92.5	414	312.6	96.52	96.44	96.78	97.00
Stuart/Nechako	5) Middle River	84.7	344	282.4	95.50	95.18	93.12	93.62
Stuart/Nechako	6) Nadina River	93.5	328	329.6	97.36	97.41	98.86	98.99
Stuart/Nechako	7) Pinchi Creek	93.0	105	85.1	76.52	78.38	58.63	60.27
Chilko	8) Chilko River	89.9	597	480.9	96.42	96.08	94.57	94.63
Chilko	9) Chilko Lake south	92.9	330	334.8	94.43	94.75	90.08	89.90
Quesnel	10) Mitchell River	72.6	244	208.1	94.92	95.14	93.35	93.39
Quesnel	11) Lower Horsefly	78.5	171	192.1	77.14	76.78	51.85	50.18

Quesnel	12) Upper Horsefly	84.9	386	420.1	89.61	87.25	71.20	68.28
Quesnel	13) McKinley Creek	86.1	177	192.9	69.36	72.02	37.65	42.35
Quesnel	14) Blue Lead Creek	87.3	91	89.7	87.32	87.02	75.87	78.20
Canyon	15) Gates Creek	79.3	321	341.8	98.88	98.91	99.96	99.96
Canyon	16) Nahatlatch River	94.1	214	217.3	98.12	98.04	99.91	99.84
Harrison	17) Birkenhead River	71.3	323	261.8	98.20	98.29	99.52	99.60
Harrison	18) Weaver Creek	78.7	462	443.7	98.69	98.76	99.69	99.76
Harrison	19) Harrison River	77.3	230	240.2	97.71	97.56	99.36	99.35
Harrison	20) Big Silver Creek	92.6	155	182.4	97.45	97.53	99.68	99.60
Lower Fraser	21) Pitt Lake	90.0	326	378.8	98.91	98.91	99.71	99.74
Chilliwack	22) Cultus Lake	93.7	785	1687.1	99.84	99.76	100.00	100.00
Chilliwack	23) Chilliwack Lake	93.3	156	206.9	98.29	98.10	99.96	99.96
North Thompson	24) Fennell Creek	91.7	391	424.0	97.79	97.69	98.29	98.25
North Thompson	25) Raft River	88.2	264	269.6	94.68	94.70	93.68	93.23
North	26) North Thompson	93.5	144	177.9	92.94	93.06	90.86	90.88

Thompson								
Shuswap Early	27) Scotch Creek	90.5	245	343.3	94.72	94.63	92.08	90.96
Shuswap Early	28) Eagle River (early)	88.8	154	183.4	94.40	94.63	94.57	94.53
Shuswap Early	29) Seymour Creek	92.8	166	224.6	91.03	90.58	86.06	86.00
Shuswap Early	30) Upper Adams River	83.9	329	353.3	97.98	98.24	99.17	99.32
Shuswap Early	31) Cayenne Creek	88.7	87	87.1	89.09	89.65	94.96	94.81
Shuswap Late	32) Lower Adams River	83.6	310	251.9	92.39	91.42	85.24	84.57
Shuswap Late	33) Lower Shuswap	92.0	217	192.9	85.33	86.03	64.87	64.81
Shuswap Late	34) Middle Shuswap	87.7	200	226.9	88.66	88.41	76.23	76.30
Shuswap Late	35) Eagle River (late)	87.4	72	161.4	86.16	84.89	70.73	70.47
South Coast	36) Phillips River	90.6	163	176.1	98.30	98.35	99.91	99.89
Long Lake	37) Smokehouse Creek	86.7	138	127.6	86.60	86.01	64.17	63.94
Long Lake	38) Canoe Creek	85.9	127	129.0	88.82	88.91	63.72	63.39
Owikeno	39) Washwash River	93.2	264	316.4	87.11	87.38	71.96	74.42
Owikeno	40) Neechanz River	93.3	288	307.9	87.34	86.56	65.35	64.02
Owikeno	41) Sheemahant River	92.8	270	277.3	84.30	83.38	60.01	61.44
Central Coast	42) Devon Lake	88.4	119	279.8	98.85	98.81	99.95	99.95

Central Coast	43) Tankeeah River	94.4	301	377.1	99.25	99.22	99.97	99.97
Central Coast	44) Kent Lake	87.2	88	94.1	96.47	96.59	99.95	99.97
Mean		96.5	254.2	292.3	92.36	92.33	85.81	85.95

Table 2. Ranking of 64 markers for average estimated composition (population and regional accuracy determined with SPAM) of single-population samples over 44 populations of sockeye salmon, as well as F_{st} values. Types are: M for microsatellites, MHC for MHC, S for SNPs, and SM for mtDNA SNPs.

Marker	Type	Population %	SD %	Region %	SD%	F_{st}
1) *Ots100*	M	81.2	10.6	90.1	6.2	0.130
2) Oki16	M	80.8	12.1	89.3	7.4	0.125
3) Oki29	M	79.0	9.6	87.5	6.3	0.075
4) Ots108	M	78.3	10.4	87.6	6.5	0.061
5) Oki10	M	77.1	9.6	83.7	7.4	0.028
6) Ots103	M	76.3	10.8	83.3	8.1	0.057
7) Omy77	M	73.3	13.1	81.7	9.5	0.103
8) Ots2	M	73.3	13.3	84.0	8.5	0.108
9) Oki6	M	71.7	16.2	80.7	11.8	0.141
10) One8	M	68.7	15.7	78.4	12.3	0.082
11) Beta-1	MHC	68.4	18.7	82.0	12.4	0.245
12) Ots3	M	66.2	18.9	77.0	13.9	0.068
13) Ots107	M	54.8	22.3	68.1	19.0	0.087
14) Oki1a	M	43.6	24.0	57.9	22.7	0.085
15) Oki1b	M	34.6	23.6	45.4	22.6	0.045
16) One-MHC2-190	S	16.0	11.6	27.7	14.7	0.309
17) One-MHC2-251	S	15.4	11.1	27.1	13.2	0.330
18) One-HPAL-99	S	15.1	9.5	25.0	11.3	0.200
19) One-GPH-414	S	15.0	9.6	25.3	10.5	0.218
20) One-U503-170	S	13.4	10.6	26.1	12.2	0.207

21) One-RAG3-93	S	13.4	9.2	25.1	10.4	0.141
22) One-ZP3B-52	S	13.1	8.9	20.7	10.9	0.121
23) One-ACBP-79	S	12.6	8.4	21.7	10.0	0.091
24) One-TF-EX3-182	S	12.2	10.2	21.5	12.5	0.182
25) One-RF-112	S	12.1	7.8	20.3	10.2	0.145
26) One-STR07-190	S	11.8	9.6	22.4	12.6	0.178
27) One-HGFA-100	S	11.7	8.5	21.0	11.3	0.105
28) One-HCS71-220	S	11.6	8.1	22.2	11.8	0.129
29) One-U401-224	S	11.5	6.2	19.3	8.5	0.065
30) One-C1259	S	11.4	8.5	19.3	11.1	0.092
31) One-LEI-87	S	11.1	6.7	19.1	8.4	0.057
32) One-TF-EX10-750	S	11.0	7.02	19.5	9.6	0.074
33) One-STC-410	S	10.9	8.9	22.6	9.2	0.254
34) One-GPDH-360	S	10.4	8.8	20.4	11.2	0.111
35) One-ALDOB-135	S	10.4	7.0	20.9	9.6	0.078
36) One-SEPP	S	10.4	7.5	22.6	8.8	0.064
37) One-U404-229	S	10.3	9.0	21.3	10.5	0.114
38) One-GPDH2-211	S	10.0	9.2	19.3	11.5	0.140
39) One-PLNS-107	S	9.9	6.3	15.3	8.1	0.051
40) One-U508-533	S	9.9	8.1	21.0	11.6	0.063
41) One-OTS213-181	S	9.8	8.5	18.9	10.2	0.107
42) One-HSP47	S	9.7	8.0	18.6	10.5	0.082
43) One-U301-92	S	9.7	8.5	20.4	9.2	0.093
44) One-VIM-569	S	9.6	7.4	14.8	8.8	0.063
45) One-HPAL-436	S	9.5	6.9	18.5	9.8	0.066

46) One-PIP-1-PIP3	S	9.3	7.0	16.5	9.5	0.059
47) One-P53-576	S	9.3	7.0	16.3	8.8	0.085
48) One-RAG1-103	S	9.2	7.1	16.3	8.8	0.079
49) One-U504-141	S	9.0	9.1	21.9	11.9	0.085
50) One-PRL2	S	9.0	7.7	16.2	10.1	0.053
51) One-CTGF-301	S	8.9	7.3	18.5	11.1	0.067
52) One-ZNF-61	S	8.7	6.9	15.8	9.7	0.053
53) One-SERPIN-104	S	8.6	8.5	17.7	10.7	0.157
54) One-IL8R-362	S	8.6	6.9	15.1	8.8	0.061
55) One-RF-295	S	8.6	6.0	15.7	8.5	0.087
56) One-GHII-2461	S	8.6	7.9	14.4	9.8	0.046
57) One-RH2OP-395	S	8.3	6.1	17.6	8.6	0.271
58) One-KPNA-422	S	8.0	7.0	13.9	9.3	0.038
59) One-U502-167	S	7.6	7.1	13.4	9.3	0.063
60) One-MARCKS-241	S	7.0	6.6	15.1	10.1	0.036
61) One-E2	S	5.1	5.7	12.0	8.8	0.015
62) One-CYTB-17	SM	2.3	3.2	7.5	5.6	-
63) One-CYTB-26	SM	2.3	3.2	7.5	5.6	-
64) One-CO1	SM	2.3	3.2	7.5	5.6	-

Table 3. Mean accuracy and precision of estimated stock compositions of single-population samples, as well as percentage correct individual assignment to specific populations and region, for a suite of six microsatellites (*Ots100, Oki16, Oki29, Ots108, Oki10, Ots103*) and suite of 46 SNPs (the top 46 from Table 2), with accuracy, precision, and percentage allocation estimated from ONCOR over 44 sockeye salmon populations from British Columbia. Populations within regions were outlined in Table 1.

			Accuracy		Precision		Individual assignment	
Region	Population	Type	Pop	Region	Pop	Region	Pop	Region
Upper Fraser	Bowron Lake	M	96.53	96.76	1.77	1.77	70.1	70.1
		S	98.62	98.62	1.02	1.02	80.0	80.0
Early Stuart	Takla Lake	M	54.26	98.28	8.49	1.16	38.7	66.9
		S	62.76	97.01	11.65	2.41	32.6	60.6
Early Stuart	Trembleur Lake	M	70.88	98.12	8.04	1.49	36.2	66.6
		S	86.77	98.67	8.42	1.59	43.5	71.8
Late Stuart	Stellako River	M	93.26	98.47	2.85	1.16	48.8	73.1
		S	82.21	98.74	5.70	1.28	41.7	84.7
Late Stuart	Middle River	M	89.30	95.64	4.03	2.19	45.2	67.8
		S	73.42	92.79	6.22	3.87	44.2	69.8

Late Stuart	Nadina River	M	97.82	99.25	1.46	0.68	70.3	84.9
		S	97.67	99.41	1.88	0.85	62.7	84.0
Late Stuart	Pinchi_Creek	M	49.25	93.53	5.47	2.77	26.0	64.9
		S	70.36	97.70	7.13	2.06	25.4	81.0
Chilko	Chilko River	M	90.94	96.58	3.99	0.17	35.0	50.8
		S	83.05	96.62	4.85	1.73	38.1	66.7
Chilko	Chilko_Lake south	M	89.38	97.83	4.04	1.36	53.2	66.8
		S	86.46	98.73	5.59	1.06	48.6	70.3
Quesnel	Mitchell River	M	91.21	99.88	3.15	0.23	80.1	98.9
		S	86.56	99.38	4.89	0.52	70.4	85.2
Quesnel	Lower Horsefly	M	37.80	97.98	7.22	1.29	18.8	71.0
		S	41.27	97.71	8.18	1.74	26.0	70.0
Quesnel	Upper_Horsefly	M	70.28	98.87	7.94	1.00	26.9	78.6
		S	49.40	97.24	8.90	1.34	16.1	71.0
Quesnel	McKinley Creek	M	39.84	98.45	8.34	1.15	33.3	79.7
		S	60.28	98.56	8.49	1.07	31.1	67.2
Quesnel	Blue_Lead_Creek	M	74.17	98.85	5.13	0.79	39.5	78.9

		S	81.33	97.44	4.94	1.47	21.4	78.6
Canyon	Gates_Creek	M	99.35	99.42	0.60	0.59	93.6	94.2
		S	99.68	99.72	0.35	0.32	91.1	91.1
Canyon	Nahatlatch River	M	99.07	99.20	0.67	0.67	89.8	90.4
		S	98.88	98.89	0.80	0.80	76.4	77.8
Harrison	Birkenhead River	M	98.11	98.55	1.15	0.80	75.6	82.1
		S	99.09	99.45	0.51	0.43	73.3	86.7
Harrison	Weaver Creek	M	98.30	99.02	1.15	0.70	70.7	79.3
		S	99.02	99.46	1.06	0.55	83.3	91.7
Harrison	Harrison River	M	96.13	97.41	1.77	1.38	50.4	58.0
		S	98.53	99.18	0.90	0.67	76.5	88.2
Harrison	Big_Silver Creek	M	98.80	98.94	0.98	0.87	79.2	85.5
		S	97.08	97.73	1.59	1.29	69.4	76.4
Lower Fraser	Pitt Lake	M	98.54	98.54	0.83	0.83	73.4	73.4
		S	97.88	97.88	1.62	1.62	64.7	64.7
Chilliwack	Cultus_Lake	M	99.85	99.86	0.26	0.22	95.8	96.6
		S	99.98	99.98	0.11	0.08	98.4	98.4

Chilliwack	Chilliwack_lake	M	99.78	99.80	0.39	0.31	97.6	98.8
		S	99.92	99.93	0.17	0.15	92.3	92.3
North Thompson	Fennell Creek	M	97.02	99.39	2.18	0.59	70.3	88.3
		S	96.98	98.28	1.78	1.14	50.0	57.4
North Thompson	Raft River	M	89.97	98.81	3.49	0.93	50.8	78.5
		S	78.36	96.05	6.34	2.01	31.0	60.6
North Thompson	North Thompson	M	87.15	93.85	4.18	2.55	51.0	68.8
		S	79.37	92.38	5.63	3.32	17.1	32.9
Shuswap Early	Scotch Creek	M	88.44	99.07	4.08	0.98	58.5	79.3
		S	78.29	97.16	5.69	1.73	30.9	57.4
Shuswap Early	Eagle River (e)	M	93.98	99.22	2.51	0.69	50.3	88.7
		S	89.46	98.87	4.53	0.90	45.9	73.0
Shuswap Early	Seymour Creek	M	82.04	96.44	5.12	2.16	36.4	77.7
		S	72.90	94.75	6.13	2.72	10.1	47.8
Shuswap Early	Upper Adams River	M	97.08	98.83	1.72	0.79	62.3	88.1
		S	91.38	99.07	3.84	0.82	40.9	79.5
Shuswap Early	Cayenne Creek	M	92.80	99.99	2.45	0.04	91.7	100.0

		S	92.45	99.74	3.97	0.46	63.6	84.8
Shuswap Late	Lower Adams River	M	78.77	93.85	5.44	2.63	19.1	46.4
		S	84.91	95.33	4.39	2.37	12.2	48.8
Shuswap Late	Lower Shuswap	M	55.99	97.85	6.49	1.32	28.2	76.3
		S	68.12	96.82	6.70	2.11	25.4	54.0
Shuswap Late	Middle Shuswap	M	71.29	98.60	6.11	1.06	34.7	70.5
		S	71.05	98.49	6.78	1.05	23.1	50.0
Shuswap Late	Eagle River (l)	M	66.74	96.54	5.43	2.04	44.2	68.6
		S	84.07	96.98	5.28	1.98	29.8	61.7
South Coast	Phillips River	M	98.72	98.72	0.84	0.84	90.4	90.4
		S	99.78	99.78	0.34	0.34	95.1	95.1
Long Lake	Smokehouse Creek	M	59.43	96.27	5.96	1.70	41.0	85.9
		S	69.47	99.57	7.67	0.67	48.3	75.9
Long Lake	Canoe Creek	M	60.43	98.72	6.15	0.89	39.6	69.4
		S	71.02	98.83	7.72	1.22	45.6	78.9
Owikeno	Washwash River	M	69.94	98.93	6.47	0.74	40.7	81.9
		S	72.14	99.11	6.46	0.80	39.1	76.8

Owikeno	Neechanz River	M	64.79	98.58	6.68	0.91	32.8	77.2
		S	64.76	98.23	7.95	1.24	34.3	68.6
Owikeno	Sheemahant River	M	62.41	98.53	6.20	0.90	31.5	76.6
		S	59.89	98.03	7.41	1.26	23.0	68.9
Central Coast	Devon_Lake	M	99.05	99.26	0.84	0.81	88.6	90.9
		S	99.83	99.87	0.26	0.18	85.4	95.1
Central Coast	Tankeeah River	M	98.93	99.15	0.86	0.83	88.7	92.1
		S	99.65	99.90	0.51	0.21	90.7	97.3
Central Coast	Kent_Lake	M	99.54	99.77	0.55	0.33	91.8	94.1
		S	99.81	99.97	0.36	0.14	87.9	100.0
Mean		M	82.89	98.15	3.72	1.11	56.6	78.8
		S	83.47	98.13	4.43	1.24	50.8	74.4

Table 4. Mean estimated population and regional stock composition (% accuracy) of single-population samples determined from ONCOR from suites of markers incorporating the best 3, 4, 5, 6, 7, 8, 10, 12, and 14 microsatellites from Table 2, as well as the best 46 SNPs. Percent correct assignment of individuals to specific populations and regions is also indicated.

	Stock composition (%)		Standard deviation (%)		Individual assignment (%)	
Loci	Population	Region	Population	Region	Population	Region
3 micros	79.85	96.39	5.67	1.93	40.63	60.28
4 micros	81.05	97.50	4.72	1.48	47.96	69.55
5 micros	81.57	97.81	4.02	1.27	52.50	74.05
6 micros	82.89	98.15	3.72	1.11	56.61	78.80
7 micros	84.00	98.40	3.63	0.98	59.52	82.68
8 micros	84.37	98.71	3.35	0.86	62.05	85.73
10 micros	84.84	98.96	3.14	0.74	65.06	88.73
12 micros	85.38	99.12	2.98	0.65	66.30	90.25
14 micros	85.53	99.20	2.94	0.61	67.71	91.48
46 SNPs	83.47	98.13	4.43	1.24	50.83	74.38

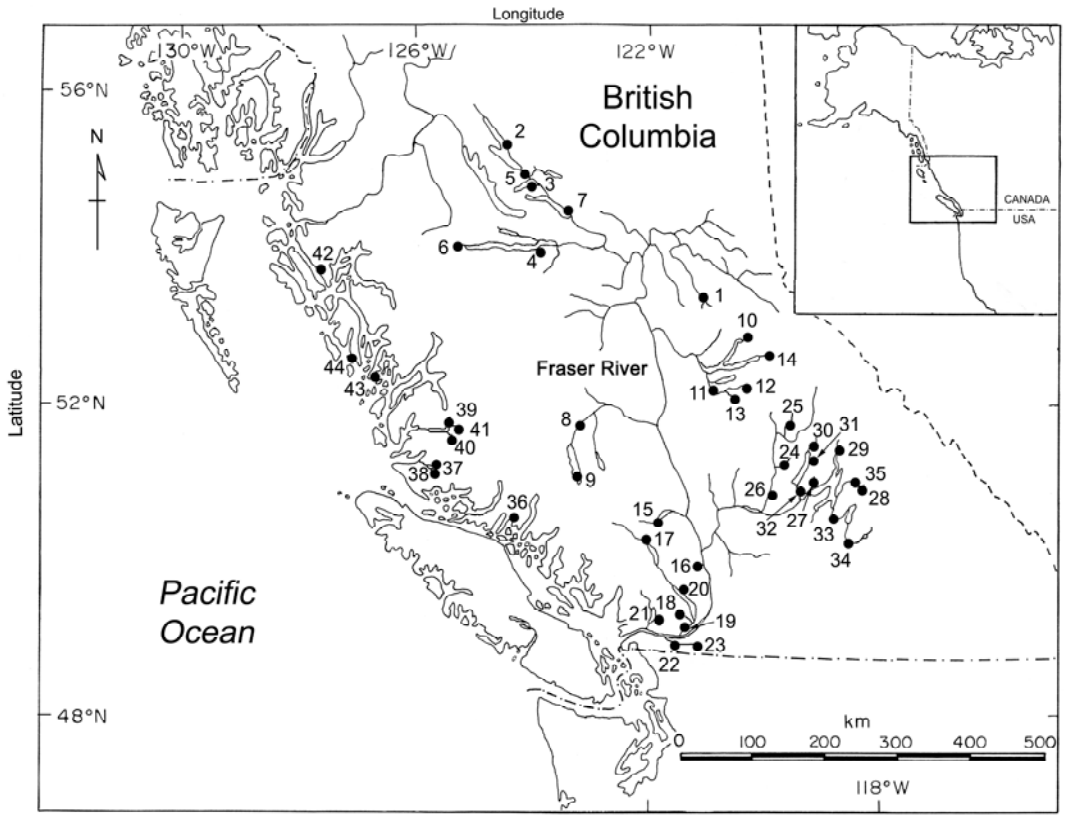
Table 5. Mean estimated population and regional stock composition (% accuracy) and precision (standard deviation, %)of single-population samples determined from ONCOR from suites of markers incorporating the best 10, 15, 20, 25, 30, 35, 40, and 46 SNPs from Table 2, as well as all 14 microsatellites. Percent correct assignment of individuals to specific populations and regions is also indicated.

	Stock composition (%)		Standard deviation (%)		Individual assignment (%)	
Loci	Population	Region	Population	Region	Population	Region
10 SNPs	68.05	90.47	8.69	4.44	28.68	44.46
15 SNPs	73.82	93.82	8.00	3.21	34.86	53.06
20 SNPs	76.61	94.98	6.62	2.62	39.36	58.78
25 SNPs	79.19	96.41	5.82	2.07	41.97	63.08
30 SNPs	80.77	97.17	5.38	1.65	45.35	67.47
35 SNPs	81.44	97.79	5.10	1.44	47.82	70.63
40 SNPs	82.11	98.01	4.71	1.30	49.48	72.60
46 SNPs	83.47	98.13	4.43	1.24	50.83	74.38
14 micros	85.53	99.20	2.94	0.61	67.71	91.48

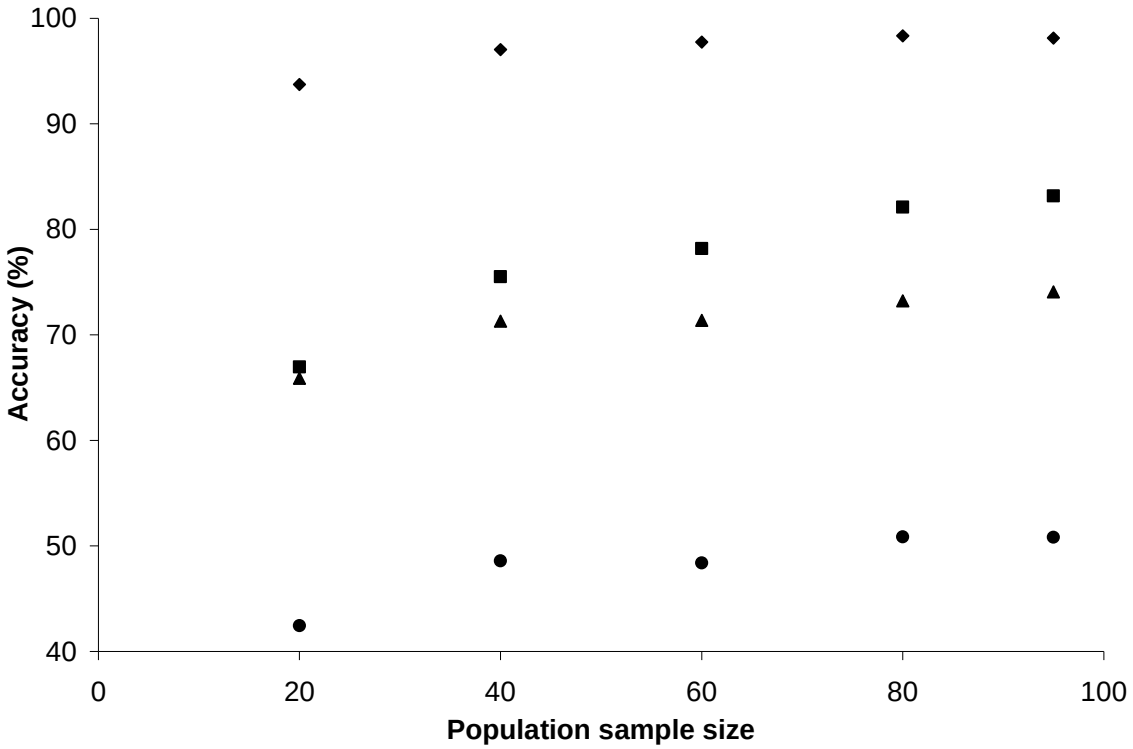
Table 6. Estimated number of SNPs required to equal performance of 14 microsatellites with respect to accuracy and precision (SD) of population and regional estimates of stock composition, as well as percentage correct individual assignment to population and region, based upon fitting functions to results outlined in Table 5. The proportion of variance accounted for by fitting each function (r^2) is also listed.

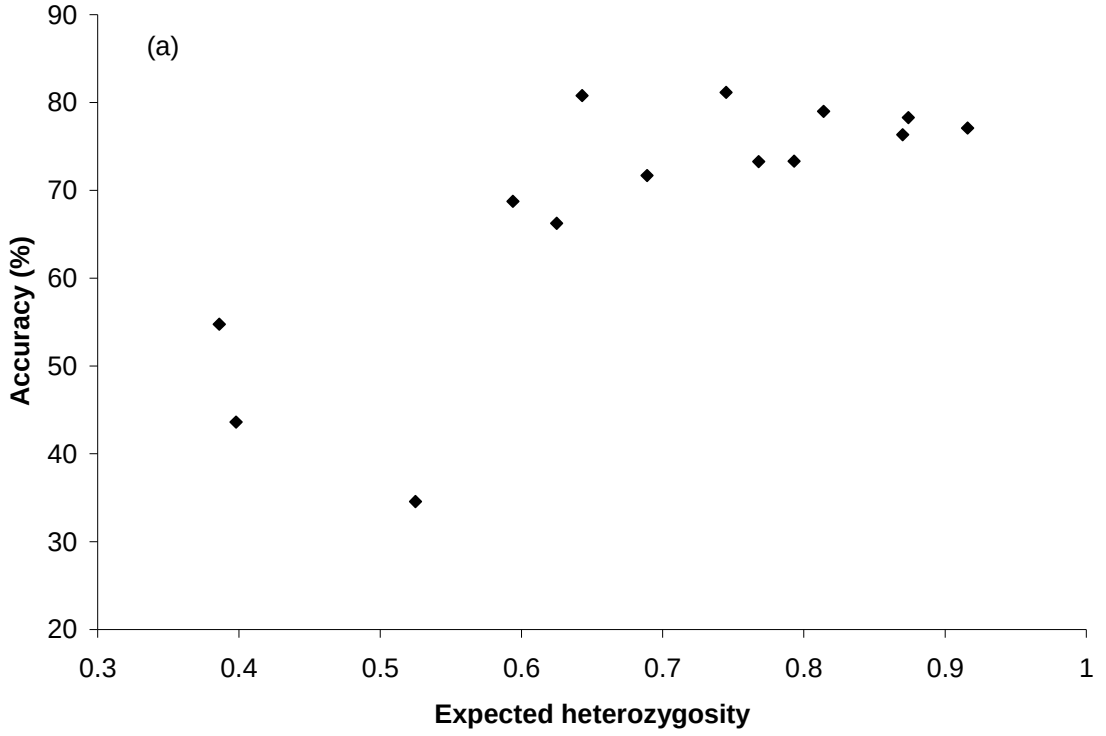
Class	Population	r^2	Region	r^2
Accuracy	75	0.9954	53	0.9964
Precision	79	0.9855	104	0.9961
Individual assignment	120	0.9931	100	0.9953

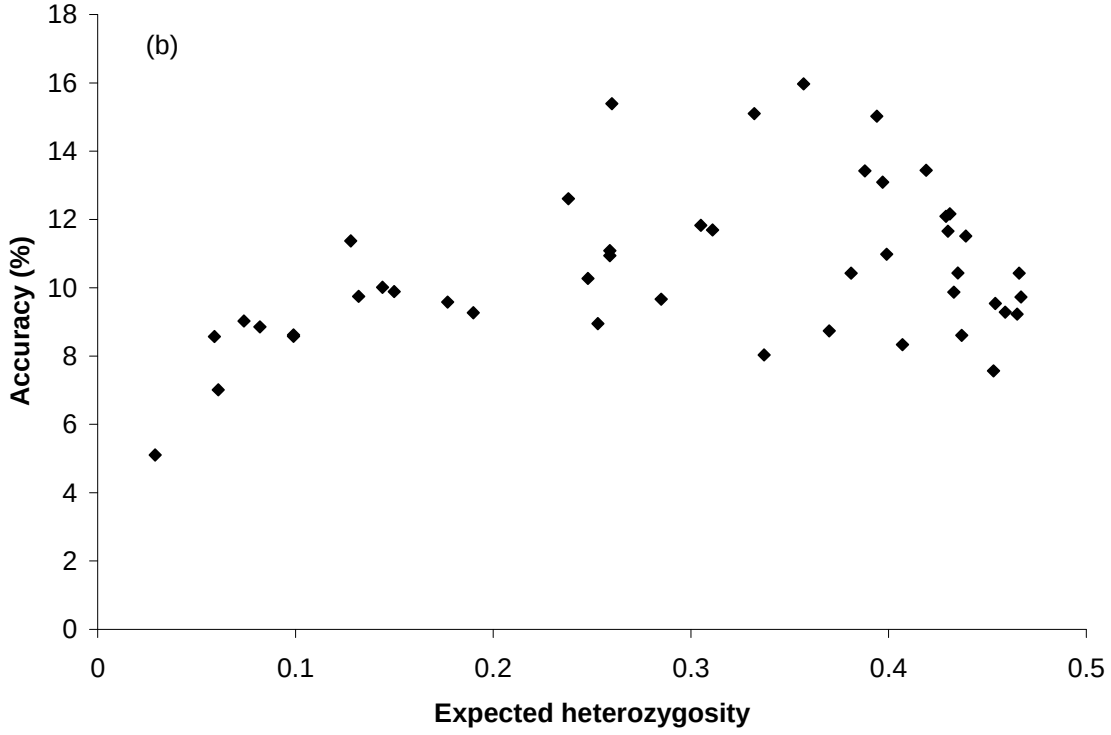
List of Figures

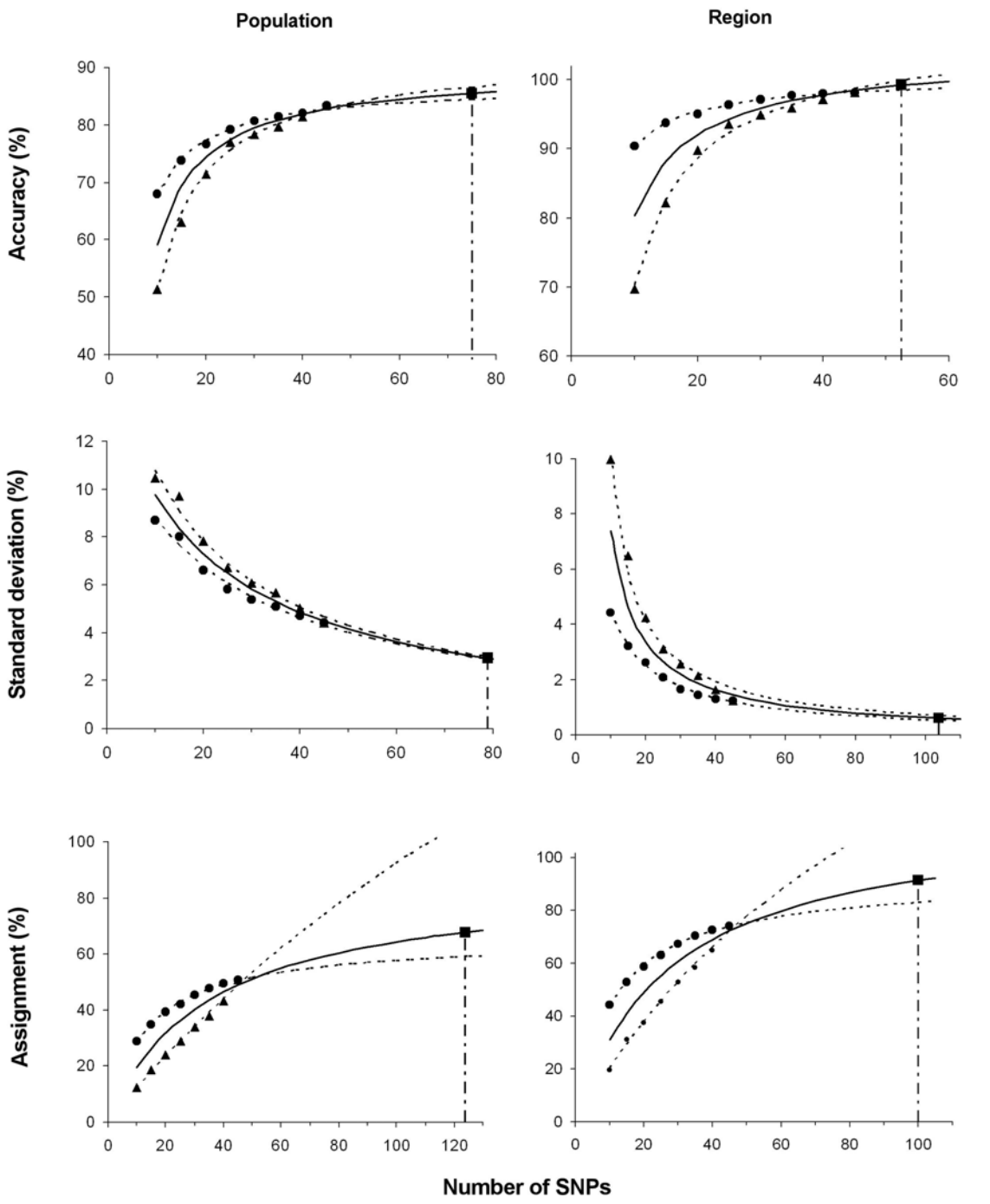

Longitude
130°W
126°W
122°W
118°W
56°N
52°N
48°N
Latitude
N
British
Columbia
Fraser River
Pacific
Ocean
CANADA
USA
km
0
100
200
300
400
500
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

50
51
52


100
90
80
70
60
50
40
Accuracy (%)
0
20
40
60
80
100
Population sample size


(a)
Accuracy (%)
90
80
70
60
50
40
30
20
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Expected heterozygosity


(b)
Accuracy (%)
18
16
14
12
10
8
6
4
2
0
0
0.1
0.2
0.3
0.4
0.5
Expected heterozygosity


Population
Region
Accuracy (%)
Standard deviation (%)
Assignment (%)
Number of SNPs

60
61

Appendix 2. A comparison of stock and individual identification for Chinook salmon (*Oncorhynchus tshawytscha*) in British Columbia provided by microsatellites (STRs) and single nucleotide polymorphisms (SNPs)

Abstract

Variation at 29 microsatellite loci (STRs) and 73 single nucleotide polymorphism (SNPs) loci was surveyed in 12 populations of Chinook salmon (*Oncorhynchus tshawytscha*) in northern and central regions in British Columbia. Microsatellites with larger observed allele numbers displayed lower F_{ST} values, but these same loci with lower F_{ST} values also tended to provide more accurate estimates of stock composition. The observed number of alleles was related to the power of the locus for providing accurate estimates of stock composition of simulated single-population samples. The number of microsatellites required to produce stock identification results equivalent to those produced from applying SNPs depended upon the specific task. Regional estimates of stock composition required at least 14 microsatellites, population-specific estimates of stock composition required six to nine microsatellites, and individual assignment to population and region required eight to 12 microsatellites. Options available for improving the accuracy and precision of stock composition estimates for 12-locus DFO microsatellite suite range include adding either four microsatellites or 25 SNPs to the existing suite. For the 13-locus GAPS microsatellites, either two microsatellites or 20-25 SNPs can be added to the existing suite. The addition of either sets of loci to both sets of microsatellites would provide generally equivalent results. When the enhanced microsatellite baselines were

evaluated, 79-88 SNPs were estimated to be required for equivalency of regional accuracy and precision for the enhanced DFO microsatellites, and 68-88 SNPs were estimated to be required for the GAPS baseline. The enhanced microsatellite baselines were projected to require 179 SNPs and 166 SNPs, respectively, for equivalency of precision of population-specific estimates. The level of regional accuracy of individual assignment available from the enhanced DFO and GAPS suites of microsatellites was projected to require 90 and 82 SNPs, respectively. The level of individual assignment to specific populations available from the enhanced DFO and GAPS suites of microsatellites was projected to require 137 and 121 SNPs, respectively. A combined microsatellite and SNP approach to estimation of stock composition in fisheries management applications may be possible for Chinook salmon stock identification applications in British Columbia.

Keywords: Chinook salmon, genetic stock identification, microsatellites, SNPs

Introduction

One key aspect in the management of Pacific salmon fisheries is to estimate stock composition in mixed-stock fishery samples with enough resolution to provide for effective management decisions, with the constraints that the estimates are timely and cost effective. For Chinook salmon (*Oncorhynchus tshawytscha*) in North America, sampling for coded-wire tags is a traditional method used to estimate stock compositions in fisheries. The use of biological markers has centered on using genetic variation for stock identification in mixed-stock Chinook salmon fisheries. Allozymes, routinely applied in some fishery samples in the 1980s, were the first method used for genetic stock identification (GSI) of mixed-stock samples of Chinook salmon (Miller et al. 1983; Milner et al. 1985; Utter et al. 1987). Allozymes were applied primarily in fisheries in the United States, but applications were limited in nature. Chinook salmon stock identification in the 1990s was largely conducted through analysis of recoveries in the coded wire tagging program.

In northern British Columbia (NBC), Chinook salmon have constituted a key component of the commercial troll fishery and sport fishery. Under the Pacific Salmon Treaty (PST) between Canada and the United States, these fisheries adjacent to the Queen Charlotte Islands are part of an aggregate abundance based management (AABM) regime implemented for NBC fisheries. Similarly, in southern British Columbia, Chinook salmon have constituted a key component of the troll fishery off the west coast of Vancouver Island, with the fishery managed under an AABM regime as well. Annual quotas for each AABM fishery are developed by prediction of Chinook salmon abundance based upon a cohort analysis model (Anonymous 2005). Poor marine survival of some Chinook salmon stocks, associated with El Niño events,

125 resulted in conservation concerns and restricted ocean exploitation in Canadian AABM fisheries
126 commencing in 1995 (Riddell et al. 2002).

127 Prior to 2002, microsatellites had been successfully applied in sockeye salmon (*O. nerka*)
128 fishery management in British Columbia where the twin management objectives were to restrict
129 exploitation on populations of conservation concern while at the same time enabling the harvest
130 of abundant populations (Beacham et al. 2004). This was precisely the dilemma confronting
131 Canadian fishery managers in the management of the Chinook salmon AABM fisheries, as only
132 a small portion of the quota available to Canadian fishermen from 1995 to 2001 was harvested
133 due to conservation concerns for specific Chinook salmon stocks. Microsatellite variation had
134 been surveyed previously for stocks likely to be present in the two Canadian AABM fisheries
135 (Beacham et al. 2003, 2006). Beginning in 2002, with managers' knowledge of timing and
136 locations of specific stocks of Chinook salmon derived from previous or in-season microsatellite-
137 based stock composition analysis, fisheries were managed with the intent of allowing Canadian
138 fishermen to harvest the quota to which they were entitled under the terms of the PST, while at
139 the same time providing protection to stocks of Canadian conservation concern (Beacham et al.
140 2008c). This change in management of Canadian AABM fisheries led to increased emphasis on
141 using microsatellites for stock identification of Chinook salmon. Subsequently, staff from
142 American agencies and Fisheries and Oceans Canada (DFO) formed a research group known as
143 the Genetic Analysis of Pacific Salmon (GAPS), and this group agreed to develop a 13-locus
144 microsatellite baseline that was to be used for stock composition estimation of Chinook salmon
145 in fisheries subject to the PST (Seeb et al. 2007). There was no formal evaluation of the GSI
146 power of the 13 microsatellites prior to their inclusion in the baseline. The set of microsatellites
147 used by Beacham et al. (2006a) for previous stock composition estimation became known as the

148 DFO loci, and four loci were in common between the GAPS suite of microsatellites and those of
149 the DFO suite. Other than the comparison outlined by Beacham et al. (2008a) for 19 populations
150 of Yukon River Chinook salmon, there has been no formal comparison of the relative power of
151 the microsatellites in the GAPS and DFO baselines, nor has the been any comparison of the
152 power of additional microsatellites currently not included in either baseline.

153 Although the GAPS microsatellite baseline was being developed, single nucleotide
154 polymorphisms (SNPs) were also developed and promoted by some fishery management
155 agencies as an alternative to microsatellites for salmon stock identification. The benefits of
156 applying SNPs relative to microsatellites were suggested to be ease of data standardization
157 among laboratories, high throughput, high among population diversity, lower genotyping errors,
158 and lower cost of analysis per individual (Smith et al. 2005a, b). However, unlike the clear
159 advantages that microsatellites displayed compared with allozymes, there was no consensus
160 among agency laboratories as to the preferred technique to apply for GSI in Chinook salmon. In
161 a report to the Pacific Salmon Commission (PSC), an "expert panel" convened to evaluate the
162 coded-wire tag program conducted by fisheries management agencies took the opportunity to
163 provide guidance on agency GSI applications. The expert panel recommended that "the Pacific
164 Salmon Commission support an immediate evaluation of the coordinated transition for all
165 salmon species from genetic stock identification (GSI) based on the use of microsatellite markers
166 to GSI based on single nucleotide polymorphism (SNP) markers" (Pacific Salmon Commission
167 2005). As a functioning GSI program for Chinook salmon in British Columbia based upon
168 microsatellites predated this report (Beacham et al. 2003, 2008c), and SNP variation had not
169 been investigated in British Columbia populations, the rationale for this recommendation was not
170 supported by DFO. Subsequently, GSI workshops sponsored by the PSC resulted in a report

171 which recommended that "the PSC should continue to support microsatellites as a demonstrated
172 coast-wide tool for GSI for Chinook salmon and continue to support the development of SNPs.
173 Both microsatellites and SNPs have demonstrated capabilities in estimation of stock
174 composition. The value of SNPs has been clearly demonstrated at a regional level, but the
175 effectiveness of a coast-wide application of SNPs remains to be explored." (Pacific Salmon
176 Commission 2008). The report noted that "Additional comparisons of current and future costs
177 and performance of microsatellites and SNPs will be required before an objective selection of a
178 single approach might be justified."

179 The key and obvious message from Pacific Salmon Commission (2008) was that
180 comparisons between microsatellites and SNPs are necessary before conclusions can be drawn
181 on the efficacy of the two techniques for Chinook salmon GSI applications. The focus of any
182 evaluation would be the resolution of stock composition and individual identification estimates
183 provided by the two techniques and the cost per individual required to obtain the observed
184 resolution. An initial comparison of stock identification resolution for Chinook salmon was
185 conducted for 19 Yukon River populations incorporating 30 microsatellites and nine SNPs
186 (Beacham et al. 2008a). In comparisons of population-specific estimation, a 9-SNP baseline was
187 approximately equivalent to a single microsatellite locus with 17-22 alleles. In a subsequent
188 study comparing the 13 GAPS microsatellites and 37 SNPs for differentiating 29 broadly
189 distributed populations, closely-related populations were better differentiated by microsatellites
190 than SNPs, but a combination of microsatellites and SNPs was indicated to be the most effective
191 suite of loci for individual assignment to population (Narum et al. 2008). More comprehensive
192 baselines, both in terms of number of microsatellites and SNPs included, as well as the number
193 of populations included in the analyses, are required to be developed before definitive

194 conclusions can be drawn with respect to the resolution of stock composition estimates derived
195 from the two techniques.

196 If a single approach to Chinook salmon GSI is to be implemented, the key question to
197 answer is how many SNPs must be used to provide stock composition and individual
198 identification estimates of equivalent quality both in terms of accuracy and precision when
199 compared with estimates produced with a high-resolution microsatellite baseline. Kalinowski
200 (2002, 2004) had previously suggested that equivalency in stock identification estimates could be
201 obtained by using a limited number of loci with many alleles, or more loci with fewer alleles.
202 Empirical evidence from microsatellites has indicated that a locus with greater numbers of alleles
203 generally provided more accurate and precise estimates of stock composition than a locus with
204 fewer numbers of alleles (Beacham et al. 2005, 2006, 2008a). SNPs generally display only two
205 alleles, and thus individual SNPs will be generally less powerful than individual microsatellites
206 in stock identification applications. The lesser power of individual SNPs relative to
207 microsatellites can be compensated by simply adding more SNPs to a GSI application so that
208 equivalency in accuracy and precision of estimated stock compositions is obtained. Once the
209 number of SNPs is determined, then evaluation of which technique produces the most cost
210 effective method of stock identification can be conducted.

211 Genetic structure of Chinook salmon is generally regionally based, with populations in
212 the same local geographic area more similar genetically to each other than to populations in more
213 distant geographic areas (Waples et al. 2004; Beacham et al. 2006b). A regional genetic
214 structure is the basis for defining reporting groups in GSI applications, with estimation of stock
215 composition by reporting group the minimum requirement for applications. In more complex
216 applications, differentiation among populations within a reporting group may be required, if

fishery management actions are directed towards specific populations (Parken et al. 2008). The final level of resolution required is the identification of individuals to specific populations in a reporting group or to specific reporting groups, and this is the most demanding aspect of GSI applications. Specific examples where this level of resolution is required are outlined by English et al. (2004) and Macdonald et al. (2009). Thus, comparison of the effectiveness of GSI techniques need to be evaluated at the level of accuracy and precision for reporting group, specific populations, and individual identification.

In the current study, 29 microsatellites were surveyed in 12 populations of Chinook salmon in northern and central British Columbia. These microsatellites included the 13 loci of the GAPS baseline, the 12 loci of the DFO baseline (4 loci in common between the two sets), as well as 8 additional microsatellites. We also evaluate the accuracy and precision of estimates of stock composition derived from a suite of 73 SNPs from the same populations. The key questions investigated were: 1) If a suite of 12-15 microsatellites were to be used in Chinook salmon GSI applications, which microsatellites should be in the suite?; 2) How many microsatellites are required to provide equivalent stock identification resolution to that of 72 SNPs?; 3) How many SNPs are required to replace the current GAPS or DFO microsatellite baselines used in GSI applications?; 4) If additional GSI power is required for either the GAPS or DFO microsatellite baselines, what is the incremental increase in power provided by adding either SNPs or microsatellites?

Methods and Materials

Collection of DNA samples and laboratory analysis

The sampling sites or populations surveyed in each geographic region are outlined in Table 1. Tissue samples were collected from mature Chinook salmon in these populations, preserved in 95% ethanol, and sent to the Molecular Genetics Laboratory at the Pacific Biological Station. DNA was extracted from the tissue samples using a variety of methods, including a chelex resin protocol outlined by Small et al. (1998), a Qiagen 96-well Dneasy® procedure, or a Promega Wizard SV96 Genomic DNA Purification system. Once extracted DNA was available, surveys of variation at 29 microsatellite loci were conducted: *Ots100*, *Ots101*, *Ots104*, *Ots107* (Nelson and Beacham 1999), *Ssa197* (O'Reilly et al. 1996), *Ogo2*, *Ogo4* (Olsen et al. 1998), *Oke4* (Buchholz et al. 2001), *Omy325* (O'Connell et al. 1997), *Oki100* (Beacham et al. 2008b), *Omm1009, Omm1037* (Rexroad et al. 2002), *Omm1080* (Rexroad et al. 2001), *Ots201b*, *Ots208b, Ots211*, *Ots212, Ots213* (Grieg 2003), *Oke4* (Buchholz et al. 2001), *Ots2*, *Ots9* (Banks et al. 1999). *Ots3M* (Grieg and Banks 1999), *Oki10* (Smith et al. 1998), *OtsG474, OtsG68 (W*illiamson et al. 2002), *OmyRGT3TUF*, *OmyRGT30TUF* (Sakamoto et al. 2000), Ssa408 (Cairney et al. 2000).

In general, PCR DNA amplifications were conducted using DNA Engine Cycler Tetrad2 (BioRad, Hercules, CA) in 6µl volumes consisting of 0.15 units of Taq polymerase, 1µl of extracted DNA, 1x PCR buffer (Qiagen, Mississauga, Ontario), 60µM each nucleotide, 0.40µM of each primer, and deionized H_2O. The thermal cycling profile involved one cycle of 15 minutes at 95°C, followed by 30 – 40 cycles of 20 seconds at 94°C, 30-60 seconds at 47 - 65°C and 30-60 seconds at 68 - 72°C (depending on the locus). Specific PCR conditions for a particular locus could vary from this general outline. PCR fragments were initially size fractionated in denaturing polyacrylamide gels using an ABI 377 automated DNA sequencer, and genotypes were scored by Genotyper

263 2.5 software (Applied Biosystems, Foster City, CA) using an internal lane sizing
264 standard. Later in the study, microsatellites were size fractionated in an ABI 3730
265 capillary DNA sequencer, and genotypes were scored by GeneMapper software 3.0
266 (Applied Biosystems, Foster City, CA) using an internal lane sizing standard. Allele
267 identification between the two sequencers were standardized by analyzing approximately
268 600 individuals on both platforms and converting the sizing in the gel-based data set to
269 match that obtained from the capillary-based set. Repeatability of genotyping was
270 evaluated by repeat PCR analysis and scoring of genotypes of individual fish.
271 Discrepancies in scoring were observed in eight genotypes of 15,360 genotypes scored
272 across all loci, for a genotyping error rate of 0.05%.

273 Variation was analyzed at 72 nuclear and 1 mitochondrial SNPs, with primer and
274 probe sequences outlined by Smith et al. (2005b, c, d), Smith et al. (2007), Campbell and
275 Narum (2008), and Miller et al. (2008). The listing of all SNPs surveyed is outlined in
276 Table 2. After PCR amplification, the plates (384-well) were read on an ABI Prism
277 7900HT Sequence Detection System by one individual using Sequence Detection
278 software from ABI. One SNP, *Ots_CI_A1*, required analysis on the automated
279 sequencer, as it was an insertion or deletion, and as such a size variant. Repeatability of
280 genotyping was evaluated by repeat PCR analysis and scoring of genotypes, with
281 discrepancies in scoring observed in three genotypes of 21,408 genotypes scored across
282 all loci evaluated, for a genotyping error rate of 0.01%.

283

284 **Estimation of stock composition in single-population samples**

285 Two software packages were utilized in estimation of stock composition of single-
286 population mixtures: Statistical Package for the Analysis of Mixtures software program
287 (SPAM version 3.7) (Debevec et al. 2000) and ONCOR (Kalinowski et al. 2007). A
288 mitochondrial DNA SNP was analyzed in the survey (*Ots_C3N3*), and as ONCOR is
289 currently unable to analyze variation incorporating mitochondrial haplotypes, SPAM was
290 used exlusively for those analyses incorporating *Ots_C3N3*. Genotypic frequencies were
291 determined for each locus in each population and were used to estimate stock
292 composition of simulated single-population samples. The Rannala and Mountain (1997)
293 correction to baseline allele frequencies was used for SPAM analyses in order to avoid
294 the occurrence of fish in the mixed sample from a specific population having an allele not
295 observed in the baseline samples from that population. All loci were considered to be in
296 Hardy-Weinberg equilibrium (Beacham et al. 2006b; Beacham et al. unpub.), and
297 expected genotypic frequencies were determined from the observed allele frequencies.
298 Reported stock compositions for simulated single-population samples are the bootstrap
299 mean estimate of each mixture of 200 fish analyzed, with mean and variance estimates
300 derived from 1000 bootstrap simulations. Each baseline population and simulated single-
301 population sample was sampled with replacement in order to simulate random variation
302 involved in the collection of the baseline and fishery samples. When ONCOR was used
303 to estimate stock compositions, the Rannala and Mountain (1997) correction to baseline
304 allele frequencies was again implemented, with precision of the stock compositions
305 calculated by bootstrapping (100 simulations) over observed baseline population sample
306 sizes and a mixture size of 200 fish. For both SPAM and ONCOR, allocations to
307 individual baseline populations were summed to provide estimates of stock compositions

for regional stock groups (Table 1). Additionally, ONCOR was used to provide estimates of accuracy of identification of individuals to specific populations or regional stock groups.

Sample sizes for the microsatellites were variable among populations and loci (Table 1). In order to control for the effect of varying sample size, construction of the microsatellite baseline for the analysis proceeded on the basis of capping population sample size at 200 individuals. Not all populations included in the analysis had at least 200 fish sampled (Table 1). A larger set of populations other than the 12 populations sampled in this study was used in the analysis of stock identification accuracy and precision. Allele frequencies for all populations surveyed in this study are available at the Molecular Genetics Laboratory website at http://www-sci.pac.dfo-mpo.gc.ca/mgl/default_e.htm

Relative ranking of the loci

The power of individual loci for stock composition estimation was initially evaluated by incorporating only a single locus for estimation of stock composition of single-population samples. As a mtSNP was included in the analysis, only SPAM was used to provide estimates of stock composition for all single population samples. Mean accuracy was determined as the average estimate across all populations, with loci then ordered from the most accurate to the least accurate (Table 2). Weir and Cockerham's (1984) F_{ST} estimates and heterozygosity for each locus over all populations were calculated with FSTAT version 2.9.3.2 (Goudet 1995). Allele frequencies for all

populations surveyed in this study are available at the Molecular Genetics Laboratory website at http://www-sci.pac.dfo-mpo.gc.ca/mgl/default_e.htm

How many microsatellites for SNP equivalency?

ONCOR was used exclusively for this analysis, with the proviso that the SNP of mtDNA origin were eliminated from the analysis. The microsatellite locus with the highest average population accuracy was initially incorporated into the analysis, and lower accuracy microsatellites were sequentially added to the analysis until the average population accuracy and precision of stock composition estimates provided by the suite of microsatellites matched that provided by the SNPs. Single-population samples were analyzed with the best 3-14 microsatellites. Accuracy and standard deviation for population- and region-specific estimates of stock composition were determined, and were then averaged over the single-population simulations for each set of loci. Percentage correct assignment of individuals to specific populations and regions was also determined for all populations, which was then averaged over all populations for each set of microsatellites evaluated.

Comparing DFO and GAPS microsatellites

The DFO suite of microsatellites is comprised of 12 loci surveyed with two injections on an automated DNA sequencer, and the GAPs suite of microsatellites is comprised of 13 loci surveyed with three injections (two dyes unutilized) on the automated DNA sequencer (Table 2). If additional accuracy of stock composition analysis is required for either set of loci, then either additional microsatellites or additional SNPs can be added to existing baselines in order to provide increased stock

identification power. If additional microsatellites are to be added to either suite and the number of injections on the automated sequencer is capped at three, then four microsatellites can be added to the DFO suite (4 dyes available for each injection), and two microsatellites can be added to the GAPS suite (2 unutilized dyes). Microsatellites added to the DFO suite would include the top-ranked non-DFO microsatellites which were determined to be *Ots201b, Ots213, Omm1080,* and *Ots212*. Similarly, microsatellites added to the GAPS suite would include *Ots107* and *Ots100*. Single-population samples were analyzed with the DFO (regular and enhanced suites) microsatellites and GAPS (regular and enhanced suites) microsatellites. Additionally, the original suites of DFO and GAPS microsatellites were enhanced with the addition of SNPs to the baselines used for stock composition analysis, with the number of SNPs added in increments of 5 SNPs, starting with the highest rated SNPs. The addition of groups of SNPs continued until the average estimated population accuracy derived from the enhanced DFO and GAPS microsatellite baselines was achieved by the addition of sets of SNPs. Accuracy and standard deviation for population- and region-specific estimates of stock composition were determined for each group of genetic markers examined, and were then averaged over the single-population simulations. Percentage correct assignment of individuals to specific populations and regions was also determined for all populations, which was then averaged over all populations for each set of microsatellites evaluated.

How many SNPs for microsatellite equivalency?

Projections of the number of number of SNPs required for equivalency of the current microsatellite baseline were conducted by ranking SNPs according to the average accuracy observed in estimation of stock composition for single-population samples over the populations surveyed. The mtDNA SNP was eliminated in the analyses. Subsequent analyses were conducted exclusively with ONCOR. Single-population samples were analyzed with 10-72 SNPs in increments of 5 SNPs. SNPs with the highest average accuracy were initially incorporated in the analyses of the single-population mixtures, with lower accuracy SNPs sequentially added to the analyses. Additionally, SNPs with the lowest average accuracy values were initially incorporated in the analyses, with progressively higher-accuracy SNPs sequentially added to the analyses, with again average accuracy and precision recorded. Overall mean accuracy and precision for each specified number of SNPs were determined by averaging the results from both processes, and this was considered indicative of the average trend in estimating accuracy and precision when the number of SNPs employed in the analysis was increased. Average regional, population, and individual accuracy and precision over all populations were recorded for each set of SNPs. A hyperbola function of the form $Y= a/X + b$ was fitted with Labfit curve fitting software (Silva and Silva 2007), with the number of SNPs incorporated in the analysis as the independent variable X, and mean observed accuracy for population and regional estimates as the dependent variable Y. Estimates of standard deviations were fitted with the power function $Y= aX^b$), with X and Y defined as previously. Population and regional individual assignment accuracy were fitted with the modified geometric function $Y= aX^{b/X}$, with X and Y defined as previously. Projections were then made with these regression models to estimate the number of SNPs that would

be required to provide estimates of comparable resolution to that provided by the microsatellites with respect to estimated stock compositions at both the regional and population level, as well as individual identification to the regional and population level.

Results

Relative ranking of the markers

The number of alleles observed at a microsatellite locus was important in determining the value of the locus for stock identification applications. The number of alleles observed at a locus varied from 7 to 62 alleles for the populations and loci surveyed in our study (Table 3). The number of alleles observed at a microsatellite locus was related to the accuracy of estimated stock composition of the simulated single-population samples (Fig. 1a). As the number of alleles increased up to about 30 alleles per locus, accuracy of population-specific estimated stock compositions increased. However, increasing numbers of alleles over 30 alleles per locus provided only a minor increase in accuracy of estimated stock composition. Mean estimated stock compositions of the single-population samples (correct=100%) were 40.2% for simulations with single loci having 10 or fewer alleles, 56.9% for loci with 11-20 alleles, 71.6% for loci with 21-30 alleles, 75.7% for loci with 31-40 alleles, 76.7% for loci with 41-50 alleles, , 78.6% for loci with 51-60 alleles, and 79.3% for a locus with more than 60 alleles (Table 2). Precision of estimated stock compositions for a locus was influenced by the number of alleles observed at a microsatellite locus, with more precise estimates derived from loci with larger numbers of alleles (Fig. 1b). Mean standard deviations of estimated stock compositions were 22.9% for loci with 10 or fewer alleles, 19.4% for loci with 11-20

alleles, 13.5% for loci with 21-30 alleles, 9.4% for loci with 31-40 alleles, 8.8% for loci with 41-50 alleles, 7.7% for loci with 51-60 alleles, and 7.1% for a locus with more than 60 alleles (Table 3). In general, microsatellites with more alleles provided more accurate and precise estimates of stock composition of the single-population samples than did loci with fewer alleles, but results could vary among specific loci.

Average population-specific accuracy for the 102 markers evaluated ranged from 2.4% to 81.1%, with a clear break in accuracy observed between SNPs and microsatellites. Average accuracy in stock identification analysis ranged from 2.4% to 13.9% for individual SNPs, and from 34.2% to 81.1% for individual microsatellites. The top 29 markers of the 102 markers evaluated were microsatellites, with accuracy of estimated population-specific stock compositions produced from incorporating the worst microsatellite (*OmyRGT30TU* with 7 alleles) 2.5 times higher that of the highest ranked SNP (*Ots_FARSLA-220*) (34.2% versus 13.9%)(Table 3). The differential between these two loci was less for region-specific accuracy (56.5% versus 39.7%).

Heterozygosity was a good predictor of the relative power of microsatellites for stock composition analysis ($r=0.85$, $P<0.01$). For microsatellites, loci with fewer alleles also tended to have lower heterozygosity (Table 3), and as loci with more alleles were more powerful for stock identification analysis, more heterozygous loci were more powerful for stock identification analysis (Fig. 2a). However, there was little increase in accuracy of estimated stock compositions for microsatellites once heterozygosity values of 0.80 were observed for microsatellites, and this corresponded approximately to loci displaying at least 30 alleles. All SNPs evaluated displayed only two alleles, and there were only minor differences in stock identification accuracy among SNPs for loci with a

444 heterozygosity value greater than 0.15 (Fig. 2b). In this case, heterozygosity at SNP loci
445 was a modest predictor of the relative power of the locus for stock identification analysis
446 ($r=0.55$, $P<0.01$).

447 F_{ST} value was a modest predictor of the relative value of each microsatellite for
448 stock composition analysis. Microsatellites with higher numbers of alleles also tended to
449 have lower F_{ST} values ($r=-0.81$, $P<0.01$). Therefore, in microsatellites, F_{ST} values were
450 negatively correlated with stock identification accuracy ($r=-0.62$, $P<0.01$) (Fig. 3a).
451 Microsatellites with lower F_{ST} values provided higher stock identification accuracy, but
452 they were also the loci that displayed more alleles. As SNPs displayed only two alleles,
453 SNPs with higher F_{ST} values should be expected to display greater power in stock
454 identification analysis. This was indeed observed in our study, with locus F_{ST} value
455 positively correlated with stock identification accuracy ($r=0.81$, $P<0.01$) (Fig. 3b). SNPs
456 that displayed higher F_{ST} values were also the most valuable for stock identification
457 analysis, the opposite of that observed for microsatellites. Obviously, when comparing a
458 mixed set of microsatellites and SNPs, it is difficult to discern the relative power of each
459 locus for stock identification analysis from a simple inspection of F_{ST} values.
460
461 **How many microsatellites for SNP equivalency?**
462 The application of 72 SNPs for estimation of stock composition of single-
463 population samples resulted in an average accuracy of 97.5% to reporting region with a
464 standard deviation of 0.8% (Table 4). These levels of accuracy and precision, along with
465 levels associated with population-level estimates of stock composition and individual
466 assignment to both region and population provided the reference points for estimation of

467 the number of microsatellites required to provide equivalent results in stock identification
468 applications as compared with the SNPs. Starting with the microsatellite with highest
469 observed population-specific accuracy (*Ots107*), at least 14 microsatellites were required
470 to be applied to achieve comparable levels of accuracy and precision in regional estimates
471 of stock composition as provided by the SNPs (Table 4).

472 The application of 72 SNPs for estimation of stock composition of single-
473 population samples resulted in an average accuracy was 83.1% to population with a
474 standard deviation of 3.7% (Table 3). Some populations such as Bulkley River were
475 clearly more differentiated than others with high levels of population-specific accuracy
476 (>97%), regardless of what suite of markers was used to estimate stock composition.
477 Similarly, estimated stock compositions for populations that displayed lower levels of
478 accuracy, such as Nahlin River, and Nakina River, were consistently lower than most
479 other populations with both microsatellites and SNPs. Standard deviations of estimated
480 stock compositions were generally higher with SNPs than with either of the DFO or
481 GAPS suites of microsatellites (Fig. 5). On average, comparable levels of accuracy and
482 precision obtained by application of the best microsatellites was achieved with nine
483 microsatellites for accuracy and six microsatellites for precision in comparison with the
484 SNPs.

485 Assignment of individuals to correct region of origin with the 72 SNPs was
486 attained with an accuracy of 87.3%, and assignment to specific population was attained
487 with an accuracy of 56.7 (Fig. 6). Comparable region-specific levels of assignment
488 accuracy were achieved with an application of 12 microsatellites, with comparable levels
489 of population-specific assignment accuracy achieved with eight microsatellites (Table 4).

In summary, the number of microsatellites required to produce stock identification results equivalent to those produced from applying 72 SNPs depended upon the specific task. Regional estimates of stock composition required at least 14 microsatellites, population-specific estimates of stock composition required six to nine microsatellites, and individual assignment to population and region required eight to 12 microsatellites.

Comparing DFO and GAPS microsatellites

Average estimates of stock composition for single-population samples were 83.0% for the DFO microsatellites and 83.1% for the GAPS microsatellites (Table 5). Both suites of loci provided essentially equal average population-specific accuracy over populations of Chinook salmon surveyed in British Columbia. The addition of the four most powerful non-DFO microsatellites (*Ots201b, Ots213, Omm1080, Ots212*) to the suite of DFO microsatellites improved estimated accuracy of the single-population samples to 85.6% (Table 5). Improvement of the level of accuracy of estimated stock compositions to this level could also be achieved by the addition of about 25 of the top-ranked SNPs to the 12 DFO microsatellites (Table 5). The addition of the two most powerful non-GAPS microsatellites (*Ots107, Ots100*) to the GAPS suite improved estimated accuracy to an average of 85.5% (Table 5), essentially equivalent to the enhanced DFO suite. Approximately 20-25 of the top-ranked SNPs would be required to be added suite of GAPS microsatellites in order to provide the same population resolution as that provided by the addition of the two most powerful non-GAPS microsatellites. In summary, at least two options are available for improving the accuracy and precision of stock composition estimates for either the DFO or GAPS suites of microsatellites if the

number of injections on the automated sequencer is capped at three injections. For the DFO microsatellites, either four microsatellites or 25 SNPs can be added to the existing suite. For the GAPS microsatellites, either two microsatellites or 20-25 SNPs can be added to the existing suite. The addition of either sets of loci to both sets of microsatellites would provide generally equivalent results.

In stock identification applications, different levels of resolution may be required among populations in different reporting regions. For example, population-specific estimates of stock composition may be required in a specific region, while region-level estimates of stock composition may be satisfactory in other regions. If an existing microsatellite baseline is to be enhanced to improve accuracy, the choice of loci to add may depend upon some degree to the specific application. Levels of accuracy of estimated stock composition at the population level clearly differed among the reporting regions (Table 6). Alternatively, accuracy of identification of Stikine River populations was most effective by enhancing the DFO loci with 25 SNPs.

In some applications, individual Chinook salmon are required to be identified to either region or population of origin. The regional origin of individuals from the Bulkley River population was identified with a high degree of accuracy, regardless of the suite of loci used in the procedure (Table 7). In northern British Columbia, identification of individuals to Stikine River, Taku River, and regional groups in the Skeena River were among the most difficult problems.

The most difficult problem encountered in stock identification applications is the correct assignment of individuals to specific populations. Lowest accuracy as typically observed for Stikine and Taku River populations (Table 8). There was no consistent

ranking within regions of the relative assignment accuracy provided by the DFO, GAPS, or SNP baselines (Table 8). For the enhanced baselines, enhancing the GAPS baseline with two microsatellites provided the least relative increase in assignment accuracy, whereas enhancing the GAPS baseline with 25 SNPs provided the greatest increase in relative assignment accuracy.

How many SNPs for microsatellite equivalency?

The average stock composition accuracy of the single-population samples incorporating the 12 DFO microsatellites was 96.9% to geographic region, and that for the 13 GAPS microsatellites was 96.7% (Table 5). In essence, the 72 SNPs employed for stock composition estimation provided at least equivalent or better accuracy in estimation of regional stock composition than the DFO or GAPS microsatellites. Average precision of regional estimates of stock composition was again essentially duplicated when compared with the DFO microsatellites, and 80 SNPs of the average quality evaluated in our study were estimated to be required in order to produce regional estimates of stock composition available from the GAPS microsatellites. When the enhanced microsatellite baselines were evaluated, 79-88 SNPs were estimated to be required for equivalency of regional accuracy and precision for the enhanced DFO microsatellites, and 68-88 SNPs were estimated to be required for the GAPS baseline (Table 10).

The average population-specific accuracy derived from the DFO microsatellites was 83.0% to specific populations, 83.1% for the GAPS microsatellites, 85.6% for the enhanced DFO microsatellites, and 85.5% for the enhanced GAPS microsatellites (Table 5). The 72 SNPs evaluated already provided equivalent population-specific accuracy to

559 that available from existing DFO and GAPS microsatellites for most populations (Fig. 4).
560 The accuracy available from enhanced DFO or GAPS suites of microsatellites was
561 projected to require between 118-122 SNPs (Table 10). Variance of population-specific
562 estimates derived from existing DFO and GAPS microsatellites was less than that
563 available from the 72 SNPs for most populations (Fig. 5). Precision of population-
564 specific estimates available from existing microsatellite baselines was projected to
565 require 122 SNPs for equivalency with the DFO microsatellites and 118 SNPs for
566 equivalency with the GAPS microsatellites. The enhanced microsatellite baselines were
567 projected to require 179 SNPs and 166 SNPs, respectively, for equivalency of precision
568 of population-specific estimates (Fig. 7, Table 10).

569 The average accuracy in assignment of individuals to specific regions was 87.0%
570 for the DFO microsatellites and 87.3% for the GAPS microsatellites, and 73-74 SNPs of
571 the average quality evaluated in the study were estimated to be required to provide
572 equivalent accuracy in assignment of individuals to region. The SNPs surveyed in the
573 study provided equivalent levels of regional assignment accuracy for individuals. The
574 level of regional accuracy of individual assignment available from the enhanced DFO and
575 GAPS suites of microsatellites was projected to require 90 and 82 SNPs, respectively
576 (Table 10). Assignment of individuals to specific populations was achieved with an
577 average accuracy of 61.5% for the DFO microsatellites and 61.8% for the GAPS
578 microsatellites. These levels of accuracy were projected to be achieved with 93-94 SNPs
579 of the average quality surveyed. The level of individual assignment available from the
580 enhanced DFO and GAPS suites of microsatellites was projected to require 137 and 121
581 SNPs, respectively (Fig. 7, Table 10).

Discussion

Sample size

Two alleles were observed at SNP loci, but up to 62 alleles were observed at the microsatellite loci. Therefore, microsatellites require more fish to be sampled in a population to obtain estimates of allele frequencies with similar accuracy and precision compared with SNPs. Beacham et al. (in review b) demonstrated that microsatellites required larger baseline sample samples than SNPs in order to reduce sampling variation in estimation of allele frequencies and thus increase accuracy of estimated stock compositions. Population sample size in the microsatellite baselines was required to be about two to three times larger than in the SNP baselines before equivalent levels of accuracy relative to the asymptotic value were obtained. As most population sample sizes were approximately 100 individuals for the survey of SNP variation in our study, microsatellite sample size was capped at a maximum of 200 individuals per population in order to estimate microsatellite allele frequencies. Estimates of stock composition of single-population samples of salmon derived from microsatellites can be limited by population sample size, with population sample sizes of approximately 200 individuals required before there is little effect of sample size on accuracy of population-specific estimates (Beacham et al. 2006a). For sockeye salmon (*O. nerka*) , Beacham et al. (in review a) showed that once approximately 95 individuals had been sampled at SNP loci within a population, there was virtually no increase in accuracy of estimated stock compositions or individual assignment. Thus accuracy of estimated stock compositions and assignment of individuals derived from SNPs was not limited by population sample size in our study. Comparison of the utility of SNPs and microsatellites for stock

identification applications requires that adequate sample sizes be available for populations included in the analyses for both classes of markers.

Relative ranking of the markers

In the current study, the survey of microsatellite variation included loci with from 7 to 62 alleles observed at a locus. The number of alleles observed at a microsatellite locus was related to the accuracy and precision of estimated stock compositions. Microsatellites with larger numbers of alleles provided more accurate and precise estimates than did loci with fewer numbers of alleles. Similar empirical results have previously been reported for Chinook salmon both in the Yukon River drainage (Beacham et al. 2008a) and on a Pacific Rim basis (Beacham et al. 2006a). Of the 10 top microsatellites surveyed in the current study, five microsatellites were also in the top 10 loci evaluated in the Yukon River, with two loci not surveyed in the previous study, and three loci of lesser value. The five loci in the top 10 of both studies were *Ots100*, *Ots107*, *Oki100*, *Omm1080*, and *Ots211*, indicative of the general power of these loci for stock identification applications, given the widely divergent geographical settings of both studies.

How does one choose which loci to include in baselines for stock identification applications? One might expect F_{ST} values should be of value in evaluating the power of loci for stock identification applications. However, for microsatellites, F_{ST} values were of little aid in predicting which loci would be of value in stock identification applications, which is not unexpected given the range in the number of observed alleles at each locus and the results of a study on marker informativeness (Rosenberg et al. 2003). F_{ST} values

628 are determined as the average value over all alleles observed at a locus, so loci with larger
629 numbers of alleles typically tend to have lower F_{ST} values as was observed in our study,
630 and thus F_{ST} values do not reflect the value of a locus for stock identification
631 applications. Similar results were reported for 13 microsatellite loci surveyed in
632 populations from a broad geographic range (Beacham et al. 2006b). Comparing F_{ST}
633 values would only be effective when comparing loci with similar numbers of alleles. In
634 those circumstances, loci with larger F_{ST} values would be expected to provide greater
635 resolution in stock composition estimates, and this was exactly the pattern observed with
636 SNPs.

637 Although other methods for evaluating the relative power of loci for population
638 differentiation and individual assignment have been applied (Rosenberg et al. 2003;
639 Hedrick 2005; Narum et al. 2008), power for stock identification formed the sole basis of
640 evaluating the relative ranking of loci surveyed in the study. In a previous study
641 evaluating the effectiveness of 13 microsatellites and 37 SNPs for assignment of
642 individuals to 29 specific populations of Chinook salmon, Narum et al. (2008), reported
643 that the best 10 loci for correct individual assignment were microsatellites, and that the
644 best 15 loci included 12 microsatellites and three SNPs. In other cases, with only 15
645 markers employed (8 SNPs, 7 microsatellites), assignment accuracy was higher than
646 using either SNPs or microsatellites alone. In our study, which centered on the accuracy
647 of population-specific estimates of stock composition as a measure of power of the locus
648 for stock identification, and which incorporated a survey of 29 microsatellites and 73
649 SNPs across populations in northern and British Columbia, the top 29 markers of the 103
650 markers evaluated in our study were all microsatellites, with accuracy of estimated

population-specific stock compositions produced from incorporating the least informative microsatellite approximately 2.5 times that of the highest ranked SNP. The 12 highest-ranked SNPs in the study of Narum et al. (2008) were incorporated in our survey, and all of these SNPs were found to be less valuable in estimation of stock composition than the least powerful microsatellite.

DFO and GAPS microsatellites

Accuracy and precision of stock composition estimates provided by either the 12-locus DFO or 13-locus GAPS sets of microsatellites were essentially equivalent, although both were lower than that provided by an optimum set of 13 microsatellites. If increased accuracy and precision of stock composition estimates is required from either set of loci, then three potential solutions exist. One potential solution is to increase the number of microsatellites incorporated in the suites of loci, another potential solution is to augment the current suite of microsatellites with higher-resolution SNPs, and a third solution is to replace entirely the microsatellites with a set of SNPs. This section of the Discussion will center on the first two potential solutions, with discussion of the third option outlined in the next section. Increasing resolution of stock composition estimates in the real world is constrained by the cost of producing the estimates for a particular sample.

On a practical basis, when microsatellites are used for stock composition estimation, constraining costs of the analysis typically center on constraining the number of injections on the automated sequencer that are required in the survey of microsatellite variation. If the number of injections is arbitrarily capped at three for analysis of Chinook salmon microsatellite variation, then four microsatellites can be added to the

DFO suite, and two microsatellites can be added to the GAPS suite. When these additional microsatellites were incorporated into the analysis, improvement of accuracy and precision of estimated stock compositions were similar with both suites of loci, with an average population-specific accuracy of 85.5% over all populations, and an average standard deviation of 2.5%. The second option for improvement of stock composition resolution was to augment the microsatellites with a suite of high-resolution SNPs. Accuracy and precision of population-specific stock composition estimates equivalent to that of the augmented suite of microsatellites was achieved with the addition of 20-25 SNPs. If the level of 85.5% average population-specific accuracy and 2.5% standard deviation observed in our survey is considered as acceptable for management applications, then the choices are clear as to how to improve accuracy of estimated stock compositions. For the GAPS microsatellites, adding two microsatellites to the suite will increase the cost of analysis by two polymerase chain reactions and the time required to analyze two addition loci. These costs would be compared to those involved in surveying an additional 20-25 SNPs, and these cost comparisons can be evaluated by individual laboratories that apply the GAPS microsatellites.

Augmentation of the DFO microsatellites required an additional injection on the automated sequencer, four additional polymerase chain reactions, and the time required to analyze these additional loci. Comparable increases in accuracy and precision of estimated stock compositions would require 25 SNPs to be analyzed. In our laboratory, we have evaluated the relative increases in cost of analysis for the DFO microsatellites augmented with both additional microsatellites and additional SNPs, and have concluded that inclusion of a third injection on the automated sequencer was more cost effective at

the present time on a cost per fish basis than inclusion of surveys of SNP variation. This change was initiated in 2009 for those applications in which the suite of DFO microsatellites was employed. However, it is still feasible to apply a limited number of SNPS in conjunction with the augmented DFO suite of microsatellites for those circumstances in which SNPs provide enhanced resolution for specific populations of management concern. A combined microsatellite and SNP approach to estimation of stock composition in fisheries management applications has already been outlined for sockeye salmon by Beacham et al. (in review a).

Microsatellites or SNPs?

In deciding what markers to use in stock identification applications, the choice typically revolves around resolution provided by the markers and the cost per fish required to screen the markers. Other factors, although of some importance, are ancillary to the practical question of determination of stock composition in mixed-stock salmon fisheries. For example, Morin et al. (2009) suggested that SNPs may offer significant advantages over microsatellites, including lower error genotyping rates, less effort in data standardization, and technologies for high throughput genotyping. Smith and Seeb (2008) suggested that comparable amounts of effort spent on developing SNP and microsatellite baselines will result in SNP baselines with greater information content per allele. All of these observations may be correct, but they are ancillary to the practical questions of cost and resolution of applying genetic markers for estimation of stock composition in mixed-stock salmon fisheries. The key question to evaluate is the number of loci required in each class of markers for comparable stock identification results, as

720 this will largely determine the cost per individual in subsequent analysis and the utility of
721 applying a single class of markers in specific laboratories. Vignal et al. (2002) suggested
722 that with the progress being made in identifying and surveying SNP variation, "the effort
723 needed to produce an equivalent amount of information as with microsatellites may some
724 day be equivalent." Progress is being made in identifying and evaluating SNPs for
725 salmon stock identification, and the current study centered around the number of SNPs
726 likely required for equivalency with microsatellites for all aspects (regional, population,
727 individual) of Chinook salmon stock identification in British Columbia.
728 Accurate and precise regional estimates of stock compositions are generally the
729 easiest to produce, followed in difficulty by population-specific estimates, with
730 assignment of individuals to specific populations the more difficult problem. Assessment
731 of the accuracy and precision of regional estimates of stock composition produced by
732 application of 72 SNPs indicated that equivalent levels of regional accuracy and precision
733 were available from the SNPs as was produced by the existing DFO or GAPS
734 microsatellites. If that level of resolution is all that is required in some applications, then
735 either existing sets of microsatellites or SNPs could be utilized. However, if the existing
736 DFO and GAPS microsatellites are enhanced by either four or two microsatellites,
737 respectively, then the SNP baseline evaluated would require some modest enhancement.
738 If population-level estimates of stock composition are required for some regions of the
739 baseline utilized, then an additional 46-50 SNPs of the average quality evaluated in the
740 study would be required to provide population-specific accuracy and precision
741 comparable to the microsatellites. An additional 94-107 SNPs were projected to be
742 required if population-specific results were required comparable to those available from

743 the enhanced microsatellite baselines. If individual assignment is part of the stock
744 identification application, then no enhancement of the of existing SNP baseline is
745 required if levels of regional levels of accuracy provided by the DFO and GAPS
746 microsatellites is acceptable, and only about an additional 20 SNPs were projected to be
747 required if the enhanced microsatellite baselines were utilized. An additional 49-65 SNPs
748 was projected to be required if population-specific assignment results were required
749 comparable to those available from the enhanced microsatellite baselines. However, if
750 more SNPs are developed that provide higher resolution than the average of the 72 SNPs
751 used in the projections, then the number of additional SNPs required to produce stock
752 composition results of equivalent quality to those as the microsatellites will be reduced
753 from the numbers projected in our study.

754 Cost of laboratory analysis for an individual fish is a key issue in deciding the
755 appropriate technology to apply in a particular laboratory. Technologies are replaced
756 when one technique provides a clear advantage over the other, such as replacement of
757 allozymes by microsatellites or SNPs as the main technique employed to survey genetic
758 variation in Pacific salmon. Although there exist a number of techniques available to
759 survey SNP variation, we are aware of no technique that will allow well over 100 SNPs
760 to be analyzed at a cost comparable to analyzing in our laboratory up to 16 microsatellites
761 surveyed with three injections on an automated DNA sequencer. For the present, the
762 combined microsatellite and SNP approach outlined by Narum et al. (2008) may be a
763 practical approach to incorporating the power of both classes of markers.

764

765 ## Acknowledgments

We thank various field staff of DFO for collection of the samples used in baseline development. Funding for the SNP portion of the study was provided by the Northern and Southern Endowment Funds of the PSC, funding for the GAPS microsatellite survey was provided by the two endowment funds as well as funds required to provide the technical work required for the U. S. section of the Pacific Salmon Treaty to implement the Chinook Salmon Agreement, and funding for the DFO microsatellite survey was provided by DFO. L. Fitzpatrick drafted the map of population locations.

Literature Cited

Anonymous, 2005. Annual Exploitation Rate Analysis and Model Calibration. Pacific Salmon Commission Joint Chinook Technical Committee Report 05-2. Vancouver, British Columbia, Canada.

Banks, M. A., Blouin, M. S., Baldwin, B. A., Rashbrook, V. K., Fitzgerald, H. A., Blankenship, S. M., and Hedgecock, D. 1999. Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tshawytscha*). J. Hered. 90: 281-288.

Beacham, T. D., Candy, J. R., Supernault, K. J., Wetklo, M., Deagle, B., Labaree, K., Irvine, J. R., Miller, K. M., Nelson, R. J., and Withler, R. E. 2003. Evaluation and application of microsatellites for population identification of Fraser River Chinook salmon (*Oncorhynchus tshawytscha*). Fish. Bull. 101: 243-259.

Beacham, T. D., Lapointe, M., Candy, J. R., McIntosh, B. MacConnachie, C. Tabata, A., Kaukinen, K., Deng, L., Miller, K. M., and Withler, R. E. 2004. Stock identification of

788 Fraser River sockeye salmon (*Oncorhynchus nerka)* using microsatellites and major
789 histocompatibility complex variation. Trans. Am. Fish. Soc. 133: 1117-1137.
790 Beacham, T. D., Candy, J. R., McIntosh, B., MacConnachie, C., Tabata, A., Kaukinen,
791 K., Deng, L., Miller, K. M., Withler, R. E., and Varnavskaya, N. V. 2005.
792 Estimation of stock composition and individual identification of sockeye salmon
793 on a Pacific Rim basis using microsatellite and major histocompatibility complex
794 variation. Trans. Am. Fish. Soc. 134: 1124-1146.
795 Beacham, T. D., Candy, J. R., Jonsen, K. L., Supernault, J., Wetklo, M., Deng, L., Miller,
796 K. M., and Withler, R. E. 2006a. Estimation of stock composition and individual
797 identification of Chinook salmon across the Pacific Rim using microsatellite
798 variation. Trans. Am. Fish. Soc. 135: 861-888.
799 Beacham, T. D., Jonsen, K. L., Supernault, J., Wetklo, M., Deng, L., and Varnavskaya,
800 N. 2006b. Pacific Rim population structure of Chinook salmon as determined
801 from microsatellite variation. Trans. Am. Fish. Soc. 135: 1604-1621.
802 Beacham, T. D., Wetklo, M., Wallace, C., Olsen, J. B., Flannery, B. G., Wenburg, J. K.,
803 Templin, W. D., Antonovich, A., and Seeb, L. W. 2008a. The application of
804 microsatellites for stock identification of Yukon River Chinook salmon. N. Am.
805 J. Fish. Manage. 28: 283-295.
806 Beacham, T. D., Urawa, S., Le, K. D., and Wetklo, M. 2008b. Population structure and
807 stock identification of chum salmon from Japan determined with microsatellite
808 DNA variation. Fish. Sci. 74: 983-994.
809 Beacham, T. D., Winther, I., Jonsen, K. L., Wetklo, M., Deng, L., and Candy, J. R.
810 2008c. The application of rapid microsatellite-based stock identification to

management of a Chinook salmon troll fishery off the Queen Charlotte Islands, British Columbia. N. Am. J. Fish. Manage. 28: 849-855.

Beacham, T. D., McIntosh, B., and Wallace, C. In review a. A comparison of stock and individual identification for sockeye salmon (*Oncorhynchus nerka*) in British Columbia provided by microsatellites (STRs) and single nucleotide polymorphisms (SNPs). Canadian Journal of Fisheries and Aquatic Sciences.

Beacham, T. D., McIntosh, B., and Wallace, C. In review b. A comparison of polymorphism of genetic markers and population sample sizes required for mixed-stock analysis of sockeye salmon (*Oncorhynchus nerka*) in British Columbia. Canadian Journal of Fisheries and Aquatic Sciences.

Buchholz, W. G, Miller, S. J., and Spearman, W. J. 2001. Isolation and characterization of chum salmon microsatellite loci and use across species. Animal Gen. 32:162-165.

Cairney, M., Taggart, J. B., and Hoyheim, B. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Mol. Ecol. 9: 2175-2178.

Campbell, N. R., and Narum, S. R. 2008. Identification of novel single-nucleotide polymorphisms in Chinook Salmon and variation among life history types. Trans. Am. Fish. Soc. 137: 96-106.

Debevec, E. M., Gates, R. B., Masuda, M., Pella, J., Reynolds, J. M., and Seeb. L. W. 2000. SPAM (Version 3.2): Statistics program for analyzing mixtures. J. Heredity 91:509-510.

833 English, K. K., Koski, W. R., Sliwinski, C., and Blakley, A. 2004. Migration timing and
834 river survival of late-run Fraser River sockeye salmon estimated using
835 radiotelemetry techniques. Trans. Am. Fish. Soc. 134: 1342-1365.
836 Goudet, J. 1995. FSTAT A program for IBM PC compatibles to calculate Weir and
837 Cockerham's (1984) estimators of F-statistics (version 1.2). Journal of Heredity
838 86:485-486.
839 Grieg, C. A., and Banks, M. A. 1999. Five multiplexed microsatellite loci for rapid
840 response run identification of California's endangered winter Chinook salmon.
841 Anim. Genet. 30: 318-320.
842 Grieg, C., Jacobson, D. P., and Banks, M. A. 2003. New tetranucleotide microsatellites
843 for fine-scale discrimination among endangered Chinook salmon (*Oncorhynchus*
844 *tshawytscha*). Molecular Ecology Notes 3: 376-379.
845 Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:
846 1633-1638.
847 Kalinowski, S. T. 2002. How many alleles per locus should be used to estimate genetic
848 distances? Heredity 88: 62-65.
849 Kalinowski, S. T. 2004. Genetic polymorphism and mixed-stock fisheries analysis.
850 Canadian Journal of Fisheries and Aquatic Sciences 61: 1075-1082.
851 Kalinowski, S. T., Manlove, K. R., and Taper, M. L. 2007. ONCOR a computer
852 program for genetic stock identification. Montana State University, Bozeman.
853 Available: www.montana.edu/kalinowski/Software/ONCOR.htm. (September,
854 2008)

Macdonald, R. W., Forsland, V., Withler, R. E., Patterson, D. A., and Demsky, A. 2009. The use of stable oxygen isotope ($\delta^{18}O$) composition in sockeye salmon body fluid to determine whether a fish has been caught in freshwater. N. Amer. J. Fish. Manage. 29: 560-569.

Miller, K. M., Beacham, T. D., and Ming, T. J. 2008. Report to the Northern Endowment Fund on project NF-2007-I4: Chinook Salmon SNP Development. Pacific Salmon Commision, Vancouver, Canada.

Miller, M., Pattillo, P., Milner, G., and Teel, D. 1983. Analysis of Chinook salmon stock composition in the May 1982 troll fishery off the Washington coast: An application of the genetic stock identification method. Wash. Depart. Fish. Tech. Rep. Number 74. 27p.

Milner, G. B., Teel, D. J., Utter, F. M., and Winans, G. A. 1985. A genetic method of stock identification in mixed populations of Pacific salmon, *Oncorhynchus* spp. Mar. Fish. Rev. 47: 1-8.

Morin, P. A., Martien, K. K., and Taylor, B. L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Mol. Ecol. Res. 9: 66-73.

Narum, S. R. , Banks, M., Beacham, T., Bellinger, R., Campbell, M., DeKoning, J., Elz, A., Guthrie, C., Kozfkay, C., Miller, K., Moran, P., Phillips, R., Seeb, L., Smith, C., Warheit, K., Young, S., and Garza, J. C. 2008. Differentiating populations at broad and fine geographic scales with microsatellites and single nucleotide polymorphisms. Mol. Ecol. 17: 3464-3477.

877 Nelson, R. J., and Beacham, T. D. 1999. Isolation and cross species amplification of
878 microsatellite loci useful for study of Pacific salmon. Anim. Genet. 30: 228-229.
879 O'Connell, M., Danzmann, R. G., Cornuet, J. M., Wright, J. M., and Ferguson, M. M.
880 1997. Differentiation of rainbow trout stocks in Lake Ontario and the evaluation
881 of the stepwise mutation and infinite allele mutation models using microsatellite
882 variability. Can. J. Fish. Aquat. Sci. 54:1391-1399.
883 Olsen, J. B., Bentzen, P., and Seeb, J. E. 1998. Characterization of seven microsatellite
884 loci derived from pink salmon. Mol. Ecol. 7: 1083-1090.
885 O'Reilly, P. T., Hamilton, L. C. , McConnell, S. K., and Wright, J. M. 1996. Rapid
886 analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR
887 multiplexing of dinucleotide and tetranucleotide microsatellites. Can. J. Fish.
888 Aquat. Sci. 53: 2292-2298.
889 Pacific Salmon Commission 2005. Expert Panel on the Future of the Coded Wire Tag
890 Program for Pacific Salmon. Pacific Salmon Commission Technical Report
891 Number 18. 230p. Vancouver, B. C.
892 Pacific Salmon Commission 2008. Recommendations for Application of Genetic Stock
893 Identification (GSI) methods to management of ocean salmon fisheries: special
894 report of the Genetic Stock Identification Steering Committee and the Pacific
895 Salmon Commission's Committee on Scientific Cooperation. Pacific Salmon
896 Commission Technical Report Number 23. 35p. Vancouver, B. C.
897 Parken, C. K., Candy, J. R., Irvine, J. R., and Beacham, T. D. 2008. Genetic and coded
898 wire tag results combine to allow more precise management of a complex
899 Chinook salmon aggregate. N. Amer. J. Fish. Manage. 28: 328-340.

Rannala, B., and Mountain, J. L. 1997. Detecting immigration by using multilocus genotypes. Proc. Nat. Acad. Sci USA 94: 9197-9201.

Rexroad, C. E., Coleman, R. L, Martin, A. M., Hershberger, W. K., and Killefer, J. 2001. Thirty-five polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). Animal Genet. 32: 283-319.

Rexroad, C. E., Coleman, R. L., Gustafson, A. L., Hershberger, W. K., and Killefer, J. 2002. Development of rainbow trout microsatellite markers from repeat enriched libraries. Mar. Biotechnol. 4: 12-16.

Riddell, B. E., Luedke, W., Till, J., Taylor, S., and Tompkins, A. 2002. Review of 2002 chinook returns to West Coast Vancouver Island, forecast of the 2002 return to the Stamp River/Robertson Creek hatchery indicator stock, and outlook for other WCVI chinook stocks. Can. Sci. Advisory Secret. Res. Doc. 2002/119. Ottawa, Ont.

Rosenberg, N. A., Li, L. M., Ward, R., and Pritchard, J. K. 2003. Informativeness of genetic markers for inference of ancestry. Amer. J. Hum. Gen. 73: 1402-1422.

Sakamoto, T., Danzmann, R. G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S. K., Woram, R. A., Okamoto, N., Ferguson, M. M., Holm, L.-E., Guyomard, R., and Hoyheim, B. 2000. A microsatellite linkage map of rainbow trout (Oncorhynchus mykiss) characterized by large sex-specific differences in recombination rates. Genetics 155: 1331-1345.

Seeb, L. W., Antonovich, A., Banks, M. A., Beacham, T. D., Bellinger, M. R., Campbell, M., Decovich, N. A., Garza, J. C., Guthrie, C. M. III, Moran, P., Narum, S. R.,

923 Stephenson, J. J., Supernault, K. J., Teel, D. J., Templin, W. D., Wenburg, J. K.,
924 Young, S. F., and Smith, C. T. 2007. Development of a standardized DNA
925 database for Chinook salmon. Fisheries 32: 540-552.
926 Silva, W. P. da, and Silva, C. D. P. S. e. 2007. Lab Fit Curve Fitting Software
927 (Nonlinear Regression and Treatment of Data Program) V 7.2.46 (1999-2009).
928 Available at http://www.labfit.net. August 2009.
929 Small M.P., Beacham, T. D., Withler, R. E., and Nelson, R. J. 1998. Discriminating coho
930 salmon within the Fraser River, British Columbia, using microsatellite DNA
931 markers. Mol. Ecol. 7:141-155.
932 Smith, C. T., Koop, B. F., and Nelson, R. J. 1998. Isolation and characterization of coho
933 salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids.
934 Mol. Ecol. 7: 1613-1621.
935 Smith, C. T., Templin, W. D., Seeb, J. E., Seeb, L. W.. 2005a. Single nucleotide
936 polymorphisms (SNPs) provide rapid and accurate estimates of the proportions of
937 U. S. and Canadian Chinook salmon caught in Yukon River fisheries. N. Amer. J.
938 Fish. Manage. 25: 944-953.
939 Smith, C. T., Seeb, J. E., Schwenke, P., and Seeb, L. W. 2005b. Use of the 5'-nuclease
940 reaction for single nucleotide polymorphism genotyping in Chinook salmon.
941 Trans. Amer. Fish. Soc. 134:207-217.
942 Smith, C. T., Baker, J., Park, L., Seeb, L. W., Elfstrom, C., Abe, S., and Seeb, J. E.
943 2005c. Characterization of 13 single nucleotide polymorphism markers for chum
944 salmon. Mol. Ecol. 14: 259-262.

945 Smith, C. T., Elfstrom, C., Seeb, L. W., and Seeb, J. E. 2005d. Use of sequence data
946 from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon.
947 Mol. Ecol. Notes 5: 4193-4203.
948 Smith C. T., Antonovich A., Templin W. D., Elfstrom C. M., Narum S. R., and Seeb, L.
949 W. 2007. Impacts of marker class bias relative to locus-specific variability on
950 population inferences in Chinook salmon: a comparison of single-nucleotide
951 polymorphisms with short tandem repeats and allozymes. Trans. Amer. Fish. Soc.
952 136: 1674-1687.
953 Smith, C. T., and Seeb, L. W. 2008. Number of alleles as a predictor of the relative
954 assignment accuracy of short tandem repeat (STR) and single-nucleotide-
955 polymorphism (SNP) baselines for chum salmon. Trans. Amer. Fish. Soc. 137:
956 751-762.
957 Utter, F., Teel, D., Milner, G., and McIssac, D. 1987. Genetic estimates of stock
958 compositions of 1983 Chinook salmon, *Oncorhynchus tshawytscha*, harvests off
959 the Washington coast and the Columbia River. Fish. Bull. 85: 13-23.
960 Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. 2002. A review on SNP and
961 other types of molecular markers and their use in animal genetics. Genet. Sel.
962 Evol. 34: 275-305.
963 Waples, R. S., Teel, D. J., Myers, J. M., and Marshall, A. R. 2004. Life-history
964 divergence in Chinook salmon: Historic contingency and parallel evolution.
965 Evolution 58: 386-403.
966 Weir, B.S., and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of
967 population structure. Evolution 38: 1358-1370.

968 Williamson K., Cordes, S., Jan, F., and May, B. 2002 Characterization of microsatellite
969 loci in Chinook salmon (*Oncorhynchus tshawytscha*) and cross-species
970 amplification in other salmonids. Mol. Ecol. Notes, 2, 17-19.
971
972

Table 1. Regions, populations within regions, and sample sizes available in the survey of SNPs and microsatellites for Chinook salmon populations in British Columbia. Region codes are: northern British Columbia mainland (NOMN), along with the Skeena, Nass, Stikine, and Taku rivers.

Region	Population	SNPs	Microsatellites
NOMN	Atnarko	95	164
NOMN	Kateen	95	123
NOMN	Kitimat	105	240
NOMN	Wannock	95	206
Skeena Upper	Sustut	95	305
Skeena Bulkley	Bulkley	95	308
Nass	Cranberry	79	122
Nass	Damdochax	95	105
Stikine	Christina	95	177
Stikine	Little Tahltan	95	262
Taku	Nahlin	95	216
Taku	Nakina	95	273
Mean		95	208

979 Table 2. Numbers of injections on the automated sequencer, primer tags employed, and
980 loci surveyed for the Fisheries and Oceans Canada (DFO) and Genetic Analysis of
981 Pacific Salmon (GAPS) surveys of microsatellite variation.

Injection	Tag	Loci
	DFO	
1	VIC	*Ssa197*
	6FAM	*Oki100, Omy325*
	NED	*Ots2, Ots9, Oke4*
	PET	*Ogo2, Ogo4*
2	VIC	*Ots101*
	6FAM	*Ots100*
	NED	*Ots104*
	PET	*Ots107*
	GAPS	
1	VIC	*Oki100*
	6FAM	*Omm1080*
	PET	*Ogo2, Ogo4*
2	VIC	*Ots213*
	6FAM	*Ots201b*
	NED	*Ots9, Ssa408*
3	VIC	*Ots208b*
	6FAM	*Ots212*
	NED	*Ots3, Ots211*

PET *OtsG474*

982

983 Table 3. Ranking of 64 markers (29 microsatellites, 73 SNPs) for average estimated composition (population and regional accuracy
984 determined with SPAM) of single-population samples over populations of Chinook salmon, as well as number of alleles observed at
985 the locus and F_{st} values. Types are: M for microsatellites, S for SNPs, and SM for mtDNA SNPs.

Marker	Type	Alleles	F_{st}	He	Population %	SD %	Region %	SD%
1) *Ots107*	M	47	0.048	0.91	81.1	7.6	91.2	4.1
2) *Ots201b*	M	52	0.048	0.90	80.8	7.7	89.6	4.8
3) *Ots213*	M	55	0.037	0.93	80.1	7.2	88.6	4.6
4) *Omm1080*	M	62	0.021	0.95	79.3	7.1	86.9	5.3
5) *Ots212*	M	34	0.049	0.87	79.3	8.7	88.0	5.9
6) *Ots100*	M	54	0.019	0.94	78.5	7.6	85.2	6.1
7) *Ots211*	M	37	0.042	0.92	77.5	8.2	86.9	5.3
8) *Ots208b*	M	54	0.029	0.93	77.5	7.8	86.4	5.5
9) *Omy325*	M	43	0.110	0.76	77.3	11.7	91.3	5.1
10) *Oki100*	M	42	0.028	0.93	77.2	8.0	85.1	5.9

11) *SSa197*	M	45	0.028	0.93	76.8	8.2	85.5	5.7
12) *Oki10*	M	57	0.032	0.94	76.3	8.0	86.3	5.3
13) *Ots101*	M	48	0.031	0.92	75.9	8.5	85.2	5.9
14) *Ssa408*	M	31	0.052	0.88	75.9	9.8	87.7	5.6
15) *Ots104*	M	41	0.028	0.93	75.6	8.5	84.5	5.9
16) *Ogo4*	M	20	0.102	0.78	75.1	12.2	88.4	6.2
17) *Otsg68*	M	49	0.021	0.93	72.7	8.8	83.0	6.3
18) *Ots2*	M	26	0.077	0.68	72.2	14.4	84.9	9.4
19) *OmyRGT3TUF*	M	22	0.074	0.79	71.6	12.1	84.3	7.4
20) *Ogo2*	M	26	0.059	0.73	70.9	14.0	83.0	9.4
21) *Omm1276*	M	38	0.049	0.87	70.2	10.8	81.7	7.2
22) *Ots3m*	M	14	0.101	0.73	69.1	15.5	84.1	9.1
23) *Otsg474*	M	16	0.141	0.51	59.6	21.0	75.6	15.2
24) *Oke4*	M	14	0.097	0.65	51.4	21.2	69.0	15.7
25) *Bhms417*	M	9	0.092	0.72	45.4	22.1	65.3	16.1

26) *Omm1009*	M	11	0.093	0.67	44.6	22.7	65.0	18.4
27) *Ots9*	M	12	0.062	0.53	41.3	23.7	58.2	20.1
28) *Omm1037*	M	10	0.091	0.57	41.1	24.5	62.3	18.1
29) *OmyRGT30TU*	M	7	0.124	0.49	34.2	22.2	56.5	17.8
30) *Ots_FARSLA-220*	S	2	0.373	0.32	13.9	11.2	39.7	13.2
31) *Ots_MHC2*	S	2	0.455	0.18	13.3	9.3	32.1	12.1
32) *Ots_RFC2-558*	S	2	0.311	0.31	12.4	10.2	34.4	11.0
33) *Ots_u211-85*	S	2	0.235	0.37	11.9	9.6	29.2	11.6
34) *Ots_hnRNPL-533*	S	2	0.247	0.38	11.8	9.4	35.6	12.0
35) *Ots_U07-25-325*	S	2	0.206	0.25	11.2	7.6	28.0	10.0
36) *Ots_NRAMP-321*	S	2	0.237	0.36	11.1	9.3	31.1	12.2
37) *Ots_PGK-54*	S	2	0.175	0.31	11.0	7.6	25.0	10.4
38) *Ots_RAG3*	S	2	0.314	0.35	10.7	9.7	32.2	14.3
39) *Ots_P450*	S	2	0.266	0.31	10.6	8.7	35.9	13.0
40) *Ots_HSP90B-100*	S	2	0.192	0.40	10.6	8.8	27.9	12.1

41) *Ots_SL*	S	2	0.167	0.41	10.5	8.4	30.5	10.3
42*) Ots_S7-1*	S	2	0.145	0.43	10.5	8.3	26.7	11.1
43) *Ots_CI-GCSH-R-R*	S	2	0.214	0.35	10.5	8.4	28.3	10.6
44) *Ots_MHC1*	S	2	0.172	0.41	10.3	8.0	29.7	11.1
45) *Ots_CI-THIO-1*	S	2	0.108	0.44	10.0	7.0	28.7	8.3
46) *Ots_P53*	S	2	0.144	0.42	10.0	8.4	28.3	10.2
47) *Ots_u202-161*	S	2	0.191	0.39	9.6	8.5	25.2	10.6
48) *Ots_NOD1*	S	2	0.190	0.31	9.6	8.6	27.2	10.3
49) *Ots_Prl2*	S	2	0.111	0.42	9.6	6.7	26.5	10.9
50) *Ots_u07-57-120*	S	2	0.175	0.35	9.6	7.9	25.8	11.7
51) *Ots_CI-PEMT-Y*	S	2	0.116	0.43	9.6	7.3	22.5	10.1
52) *Ots_IL8R-CS*	S	2	0.139	0.36	9.5	7.9	29.1	9.7
53) *Ots_IL-11*	S	2	0.157	0.42	9.4	8.5	30.9	10.9
54) *Ots_NKEF-192*	S	2	0.150	0.26	9.3	8.0	31.7	10.8
55) *Ots_ETIF1*	S	2	0.167	0.41	9.1	8.1	25.1	11.0

56) *Ots_U07-53-133*	S	2	0.094	0.37	9.0	7.3	24.6	9.5
57) *Ots_ASPAT-196*	S	2	0.193	0.14	8.9	7.5	28.0	10.9
58) *Ots_TGFb1*	S	2	0.091	0.42	8.8	7.0	25.0	9.0
59) *Ots_u4-92*	S	2	0.196	0.17	8.7	8.3	31.3	10.6
60) *Ots_CI-CIRPA-2-2*	S	2	0.127	0.43	8.7	7.3	29.0	10.4
61) *Ots_MYOD-364*	S	2	0.140	0.34	8.7	8.0	25.8	11.4
62) *Ots_TLR3*	S	2	0.079	0.43	8.7	6.6	23.1	8.3
63) *Ots_FGF6A*	S	2	0.094	0.28	8.6	6.3	25.7	9.9
64) *Ots_GTH2B-550*	S	2	0.103	0.40	8.5	7.0	28.0	12.3
65) *Ots_U07-18-378*	S	2	0.016	0.23	8.5	8.2	28.2	12.5
66) *Ots_HSP90B-385*	S	2	0.096	0.30	8.5	6.8	24.3	9.6
67) *Ots_SClkF2R2-135*	S	2	0.108	0.43	8.4	7.7	22.7	10.7
68) *Ots_TAPBP*	S	2	0.088	0.22	8.4	6.6	19.5	8.1
69) *Ots_Tnsf*	S	2	0.135	0.44	8.3	7.4	25.9	11.3
70) *Ots_GPH-318*	S	2	0.129	0.19	8.3	6.8	30.7	11.5

71) *Ots_TCL1*	S	2	0.087	0.45	8.1	6.3	20.7	9.4
72) *Ots_MYO1-384*	S	2	0.134	0.15	8.1	7.0	23.3	8.0
73) *Ots_U07-49-290*	S	2	0.094	0.33	8.1	7.0	25.2	9.2
74) *Ots_E2-275*	S	2	0.095	0.46	8.0	6.7	20.2	10.2
75) *Ots_CI-OSTM1-1*	S	2	0.086	0.45	7.9	6.9	21.4	10.5
76) *Ots_U07-07-161*	S	2	0.066	0.40	7.8	6.5	21.1	10.4
77) *Ots_CD63*	S	2	0.079	0.45	7.4	6.1	20.9	9.2
78) *Ots_cox1-241*	S	2	0.078	0.43	7.4	7.0	25.5	9.4
79) *Ots_SWS1op-182*	S	2	0.068	0.47	7.3	6.1	21.9	8.4
80) *Ots_IGF-I.1-76*	S	2	0.157	0.19	7.2	7.5	25.8	12.5
81) *Ots_GPDH-338*	S	2	0.110	0.12	7.2	5.5	17.9	8.6
82) *Ots_CI-PHOS-R*	S	2	0.102	0.25	6.9	7.3	20.7	11.1
83) *Ots_u6-75*	S	2	0.091	0.18	6.9	6.7	20.3	10.0
84) *Ots_CI-1363-W*	S	2	0.083	0.32	6.9	6.9	18.5	9.7
85) *Ots_U212-158*	S	2	0.052	0.17	6.9	5.6	17.3	8.3

86) *Ots_CI_A1*	S	2	0.053	0.45	6.8	5.8	16.7	8.6
87) *Ots_CI-1740-R*	S	2	0.050	0.47	6.8	5.4	21.0	8.6
88) *Ots_C3N3*	SM	2	-	-	6.8	1.2	20.6	2.1
89) *Ots_AsnRS-60*	S	2	0.045	0.37	6.7	6.2	19.3	9.8
90) *Ots_unkn526*	S	2	0.071	0.21	6.7	6.0	21.2	10.2
91) *Ots_U07-17-373*	S	2	0.078	0.09	6.6	6.1	22.2	7.9
92) *Ots_LWSop-638*	S	2	0.117	0.09	6.5	6.5	22.3	8.7
93) *Ots_GnRH-271*	S	2	0.073	0.14	6.4	6.4	19.6	10.5
94) *Ots_CI-1803-K*	S	2	0.056	0.33	6.1	6.1	15.3	8.7
95) *Ots_CI-ISOT-Y*	S	2	0.064	0.09	5.7	6.0	20.2	10.3
96) *Ots_OTS311-101X*	S	2	0.034	0.15	5.7	5.6	18.1	8.8
97) *Ots_ZNF330-181*	S	2	0.189	0.02	5.3	3.5	16.2	7.2
98) *Ots_arf-188*	S	2	0.057	0.02	4.8	4.5	16.1	7.9
99) *Ots_Ikaros-250*	S	2	0.078	0.01	4.3	4.2	14.1	7.1
100) *Ots_HGFA-446*	S	2	0.014	0.01	2.9	3.8	13.1	7.5

101) *Ots_E9-BAC*	S	2	0.006	0.00	2.4	3.0	11.0	5.7
102) *Ots_CYP17*	S	2	0.006	0.00	2.4	3.4	10.8	6.2

986

Table 4. Mean estimated population and regional stock composition (% accuracy) of single-population samples determined from ONCOR from suites of markers incorporating the best 3-14 microsatellites from Table 3, as well as the best 72 SNPs. Percent correct assignment of individuals to specific populations and regions is also indicated.

	Stock composition (%)		Standard deviation (%)		Individual assignment (%)	
Loci	Population	Region	Population	Region	Population	Region
3 micros	75.8	92.9	4.7	2.3	35.6	62.2
4 micros	76.3	93.2	4.2	2.1	42.0	68.0
5 micros	77.5	94.1	3.9	1.8	46.9	73.7
6 micros	80.3	94.3	3.4	1.7	52.0	76.3
7 micros	81.5	94.9	3.3	1.6	55.9	79.7
8 micros	82.2	95.3	3.1	1.4	58.2	81.9
9 micros	83.4	96.2	2.8	1.2	60.7	84.5
10 micros	83.9	96.4	2.7	1.1	62.2	85.6
11 micros	83.3	96.2	2.7	1.1	64.3	87.0
12 micros	83.8	96.5	2.6	1.0	64.5	87.5
13 micros	84.8	97.1	2.5	1.0	65.3	88.1
14 micros	84.9	97.3	2.5	1.0	67.9	89.3
72 SNPs	83.1	97.5	3.7	0.8	56.7	87.3

Table 5. Mean estimated population and regional stock composition (% accuracy) of single-population samples determined from ONCOR from suites of markers incorporating the best 13 microsatellites, 72 SNPs, the DFO microsatellites, the DFO microsatellites plus the best 20-25 SNPs, the DFO microsatellites plus 4 additional microsatellites (*Ots201b, Ots213, Omm1080*, *Ots212*), the GAPS microsatellites, the GAPS microsatellites plus the best 20-25 SNPs, and the GAPS microsatellites plus 2 additional microsatellites (*Ots107, Ots100*) over Chinook salmon populations in British Columbia. Percent correct assignment of individuals to specific populations and regions is also indicated.

	Stock composition (%)		Standard deviation (%)		Individual assignment (%)	
Loci	Population	Region	Population	Region	Population	Region
Optimum 13	84.8	97.1	2.5	1.0	65.3	88.1
72 SNPs	83.1	97.5	3.7	0.8	56.7	87.3
DFO	83.0	96.9	3.1	1.1	61.5	87.0
DFO + 20 SNPs	84.8	97.7	2.7	0.8	66.5	91.7
DFO + 4 micros	85.6	97.5	2.4	0.9	69.1	91.0
DFO + 25 SNPs	85.7	97.9	2.6	0.8	67.8	92.2
GAPS	83.1	96.7	2.8	1.0	61.8	87.3
GAPS + 20 SNPs	84.9	97.6	2.6	0.8	67.1	91.4
GAPS + 2 micros	85.5	97.1	2.5	0.9	66.9	89.2
GAPS + 25 SNPs	86.3	97.8	2.5	0.8	68.3	91.6

Table 6. Average accuracy of estimated population stock compositions by region for the Fisheries and Oceans Canada (DFO), Genetic Analysis of Pacific Salmon (GAPS), SNP, DFO plus 4 microsatellites (DFO+4m), GAPS plus 2 microsatellites (GAPS+2m), DFO plus 25 SNPs (DFO+25s), and GAPS plus 25 SNPs (GAPS+25s) baselines incorporating the regions and populations outlined in Table 1.

Region	DFO	GAPS	SNPs	DFO+4m	GAPS+2m	DFO+25s	GAPS+25s
NOML	91.7	90.2	91.5	94.7	92.8	94.1	93.1
Skeena-upper	89.8	84.5	97.1	89.5	84.6	93.9	90.2
Skeena-Bulkley	93.7	94.9	99.0	95.6	96.2	94.9	96.1
Skeena-lower	83.5	88.2	76.0	92.2	90.0	96.0	97.0
Nass	79.3	85.4	81.1	85.4	86.0	82.2	87.2
Stikine	61.1	59.2	55.3	63.2	60.2	63.8	61.8
Taku	70.0	71.7	64.2	71.3	71.4	72.1	75.1

Table 7. Average accuracy of regional assignment for individual Chinook salmon for the Fisheries and Oceans Canada (DFO), Genetic Analysis of Pacific Salmon (GAPS), SNP, DFO plus 4 microsatellites (DFO+4m), GAPS plus 2 microsatellites (GAPS+2m), DFO plus 25 SNPs (DFO+25s), and GAPS plus 25 SNPs (GAPS+25s) baselines incorporating the regions and populations outlined in Table 1.

Region	DFO	GAPS	SNPs	DFO+4m	GAPS+2m	DFO+25s	GAPS+25s
NOML	85.7	82.0	88.8	89.3	85.7	92.5	91.7
Skeena-upper	63.7	68.0	51.7	74.7	69.0	67.7	70.9
Skeena-Bulkley	77.4	85.1	81.6	84.4	85.2	81.2	88.1
Skeena-lower	49.5	54.5	66.1	68.0	55.5	75.7	77.0
Nass	84.3	88.3	92.5	86.2	89.1	93.0	95.0
Stikine	66.4	65.0	61.1	75.0	69.8	78.0	69.0
Taku	69.4	74.2	66.1	76.6	76.3	76.2	75.8

Table 8. Average accuracy of population assignment for individual Chinook salmon for the Fisheries and Oceans Canada (DFO), Genetic Analysis of Pacific Salmon (GAPS), SNP, DFO plus 4 microsatellites (DFO+4m), GAPS plus 2 microsatellites (GAPS+2m), DFO plus 25 SNPs (DFO+25s), and GAPS plus 25 SNPs (GAPS+25s) baselines incorporating the regions and populations outlined in Table 1.

Region	DFO	GAPS	SNPs	DFO+4m	GAPS+2m	DFO+25s	GAPS+25s
NOML	68.7	63.5	64.5	78.5	72.4	76.2	74.6
Skeena-upper	62.5	66.0	49.2	74.5	66.6	67.7	70.9
Skeena-Bulkley	77.4	85.1	78.8	84.4	85.2	81.2	88.1
Skeena-lower	49.5	54.5	66.1	68.0	55.5	75.7	77.0
Nass	54.3	63.7	52.4	66.9	65.4	60.7	71.0
Stikine	33.0	35.2	25.0	40.3	39.3	41.2	40.3
Taku	41.2	47.8	30.0	49.7	50.1	50.8	50.9

Table 9. Mean estimated population and regional stock composition (% accuracy) and precision (standard deviation, %) of single-population samples determined from ONCOR from suites of markers incorporating either the best or worst 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, and 72 SNPs from Table 3. Percent correct assignment of individuals to specific populations and regions is also indicated.

		Stock composition (%)		Standard deviation (%)		Individual assignment (%)	
Loci	Class	Population	Region	Population	Region	Population	Region
10 SNPs	Best	59.8	88.3	10.5	5.0	21.9	56.8
	Worst	28.9	59.5	10.2	11.7	5.6	21.9
15 SNPs	Best	68.0	91.7	8.2	3.5	27.7	64.2
	Worst	47.5	79.0	10.7	7.9	10.0	27.4
20 SNPs	Best	71.9	94.1	7.0	2.6	33.1	70.5
	Worst	63.4	85.9	8.5	5.6	16.7	36.6
25 SNPs	Best	75.5	95.2	6.2	2.2	38.0	74.2
	Worst	67.9	86.9	7.5	4.2	21.2	44.6
30 SNPs	Best	80.1	95.8	5.4	1.9	41.8	77.1
	Worst	71.2	91.3	6.9	3.5	25.5	50.1
35 SNPs	Best	79.8	96.1	5.0	1.7	44.3	79.3
	Worst	74.1	93.0	6.2	2.8	31.2	58.1
40 SNPs	Best	80.1	96.4	4.8	1.6	46.6	81.0
	Worst	75.1	93.8	5.7	2.6	34.7	63.6
45 SNPs	Best	80.6	96.6	4.5	1.5	48.9	82.7
	Worst	76.0	94.6	5.4	2.1	38.7	69.2

50 SNPs	Best	81.6	96.6	4.5	1.4	51.2	84.5
	Worst	78.6	95.6	4.9	1.9	42.6	73.5
55 SNPs	Best	81.7	96.6	4.2	1.3	52.0	84.9
	Worst	80.4	96.4	4.4	1.6	47.1	78.5
60 SNPs	Best	81.9	97.0	4.1	1.2	53.8	85.0
	Worst	80.9	96.7	4.2	1.4	50.0	81.4
65 SNPs	Best	82.7	97.0	3.9	1.2	55.7	86.8
	Worst	82.1	97.0	4.0	1.3	51.8	83.5
70 SNPs	Best	83.1	97.3	3.8	1.1	56.7	87.2
	Worst	82.6	97.1	3.9	1.2	54.6	85.7
72 SNPs	All	83.2	97.2	3.8	1.1	56.7	87.3

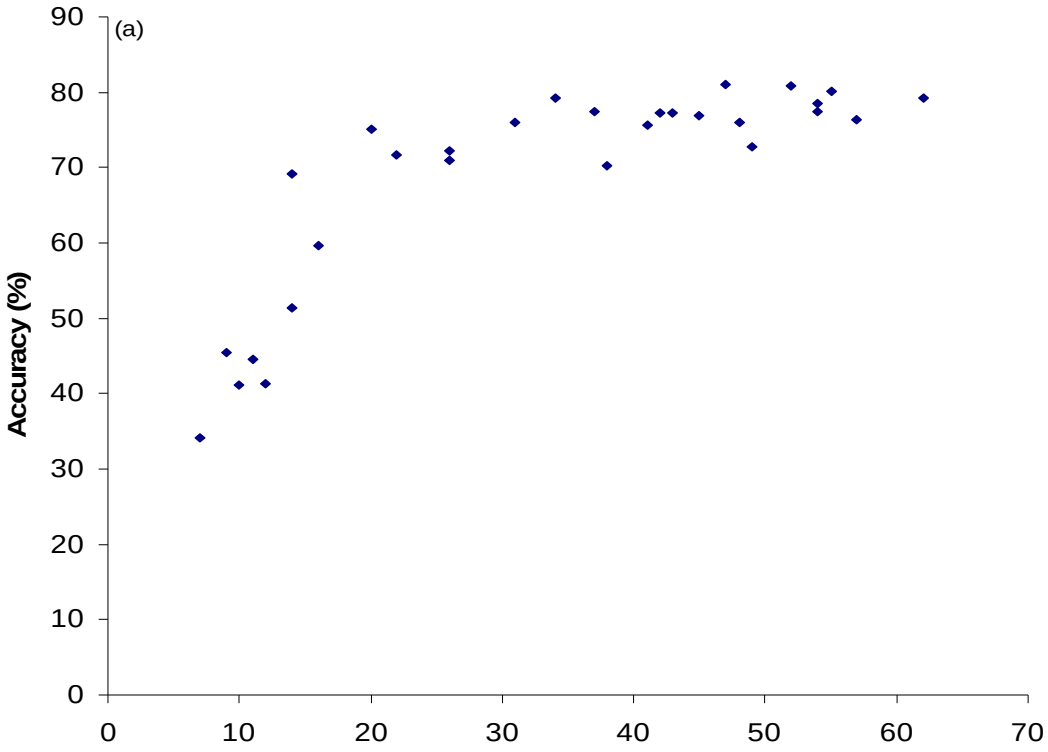
Table 10. Estimated number of SNPs required to equal performance of different suites of microsatellites with respect to accuracy and precision (SD) of population and regional estimates of stock composition, as well as percentage correct individual assignment to population and region, based upon fitting functions to results outlined in Table 9.

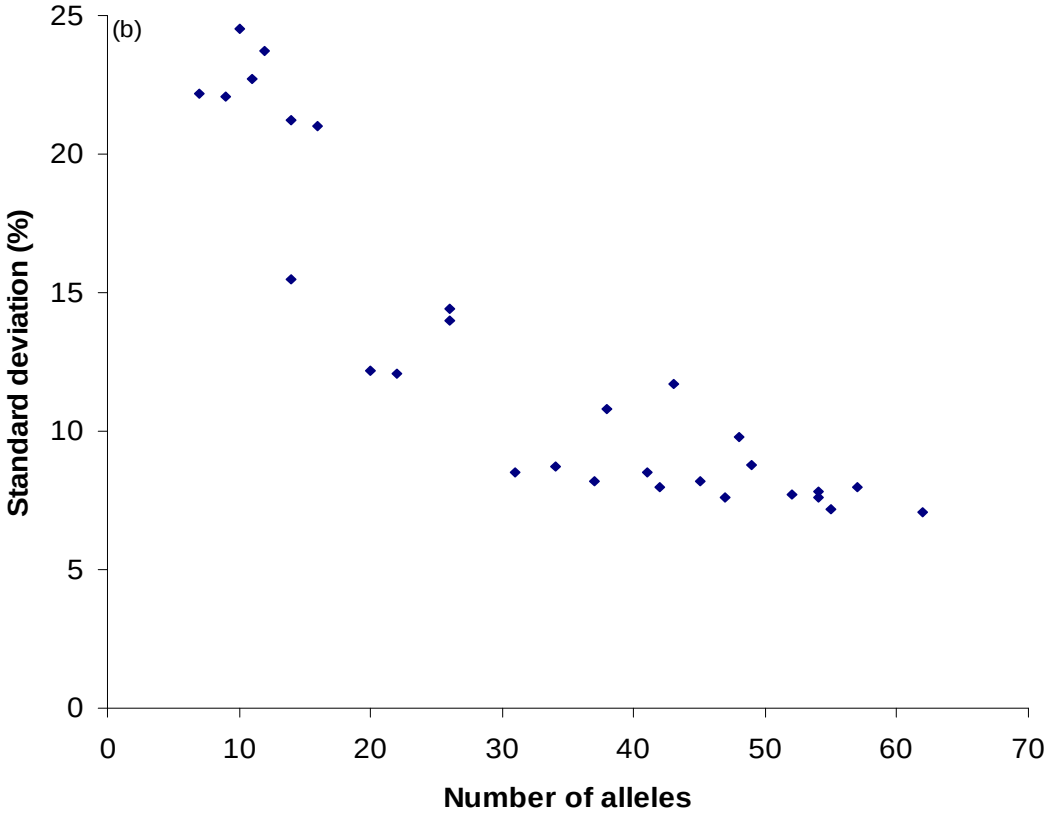
	Stock composition (%)		Standard deviation (%)		Individual assignment (%)	
Loci	Population	Region	Population	Region	Population	Region
DFO	72	64	110	73	93	73
DFO+4	122	79	179	88	137	90
GAPS	73	60	133	80	94	74
GAPS+2	118	68	166	88	121	82

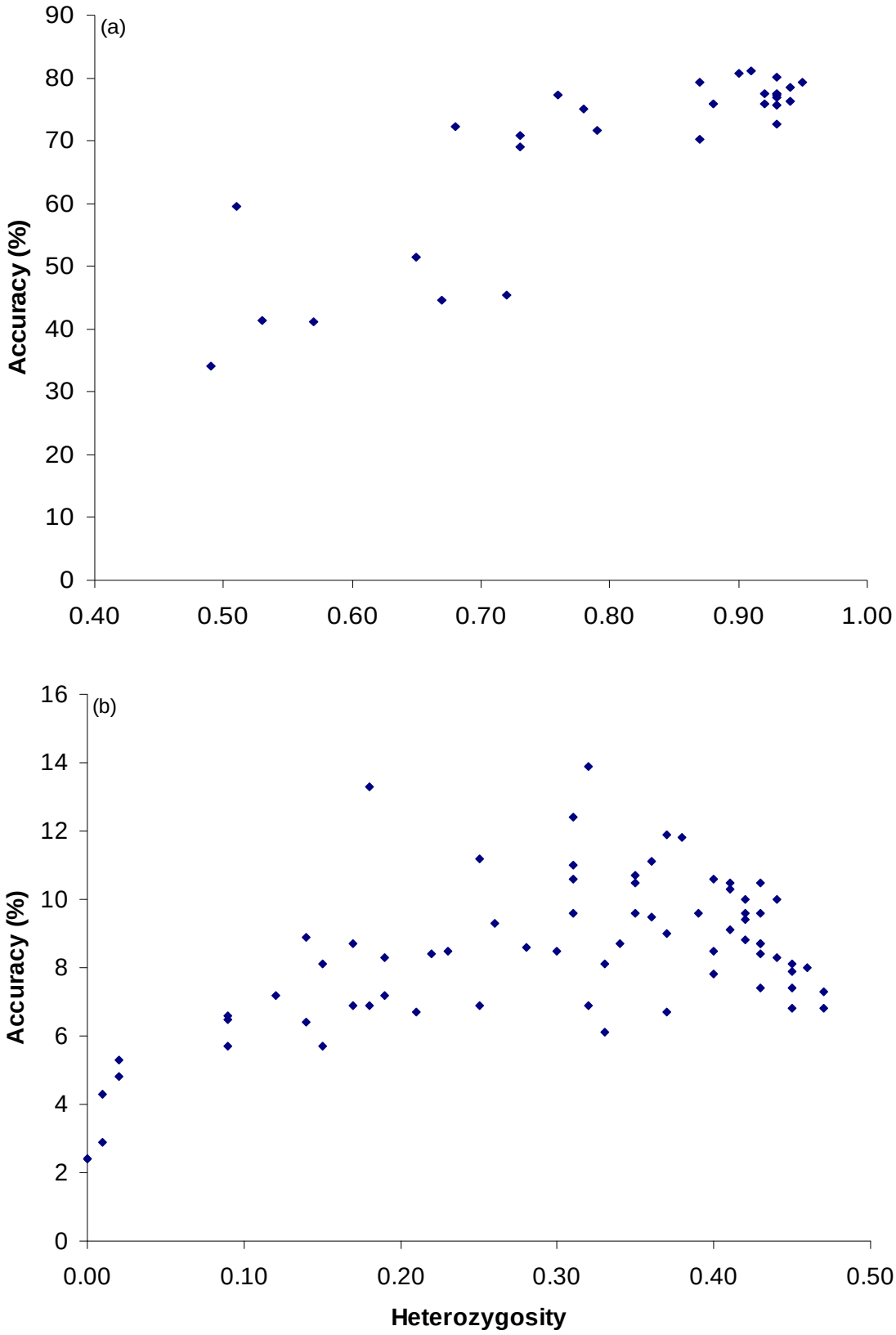
List of Figures

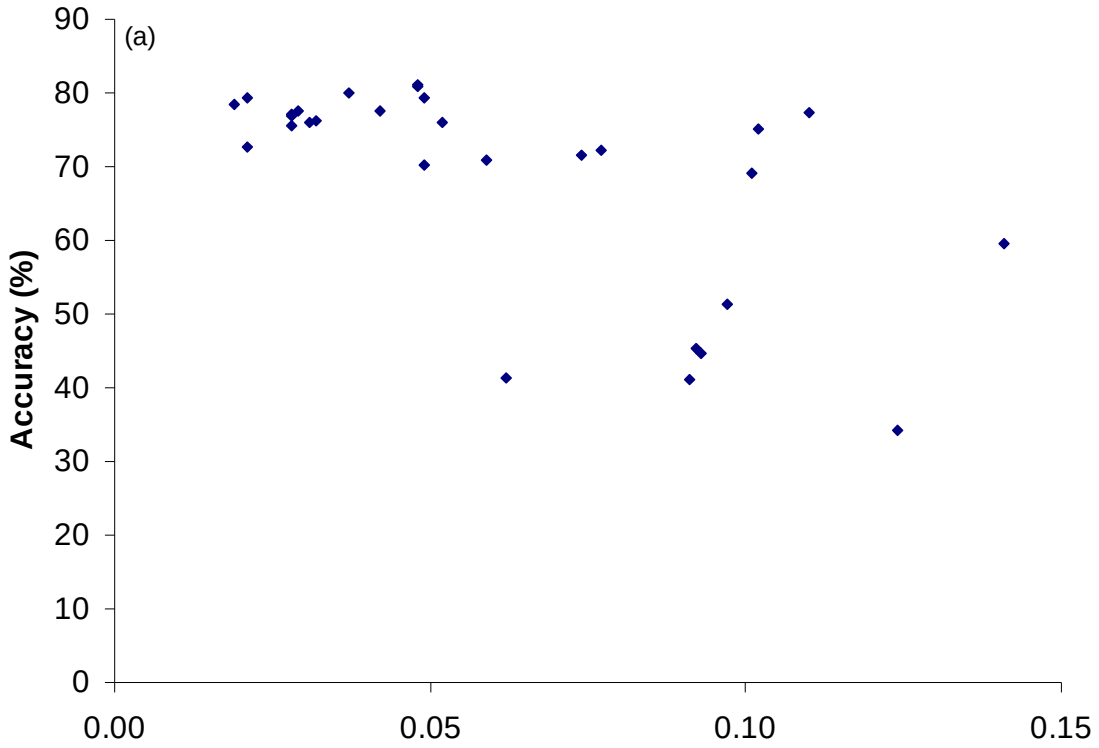
Figure 6. Mean population accuracy of individual assignment (%) by population derived from the DFO microsatellites (□), GAPS microsatellites (■), and SNPs (■) suites of loci for northern mainland, Skeena River, Nass River, Stikine River, and Taku River Chinook salmon populations for simulated single population samples.

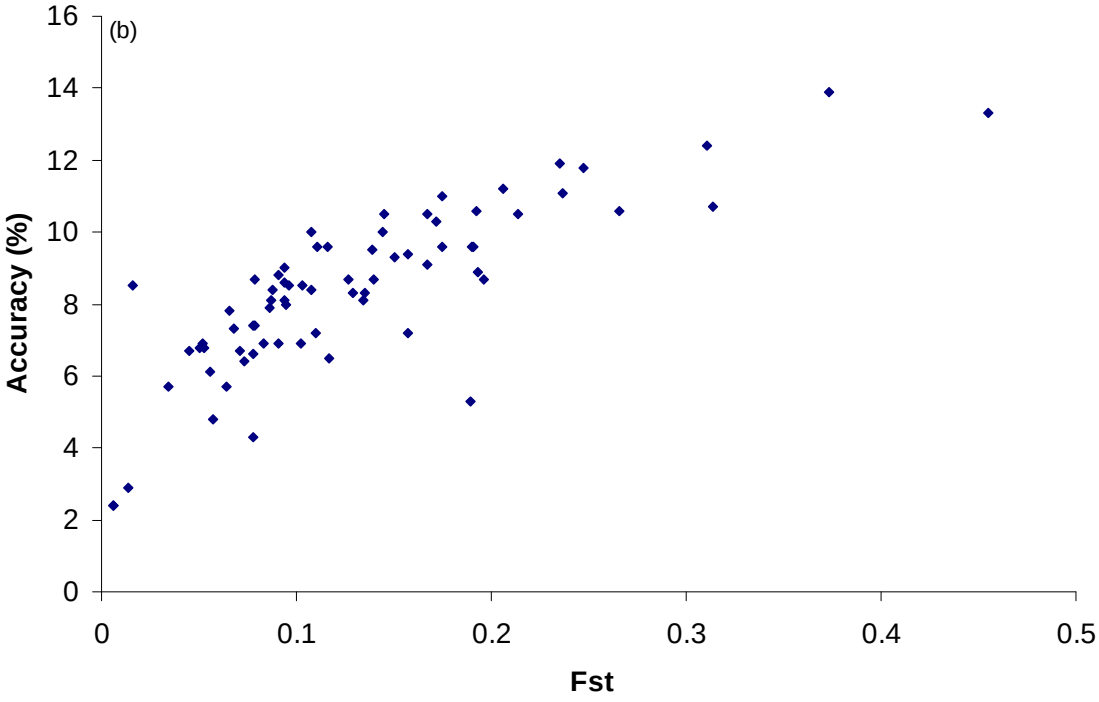
Figure 7. Accuracy and standard deviations of regional and population estimates of stock composition versus number of SNPs employed in estimation procedure, as well as accuracy of individual assignment to specific regions or populations. The number of SNPs projected to be required to provide results comparable to those of different suites of microsatellites is indicated by vertical lines. Estimates were bounded by were initially incorporating SNPs with the highest average accuracy in the analyses of the single-population mixtures, with lower-accuracy SNPs sequentially added to the analyses. Average population accuracy and precision were recorded for each set of SNPs (■). Subsequently, SNPs with the lowest average accuracy values were initially incorporated in the analyses, with progressively higher-accuracy SNPs sequentially added to the analyses, with again average accuracy and precision recorded (▲). Overall mean accuracy and precision for each specified number of SNPs were determined by averaging the results from both processes and used in projections.

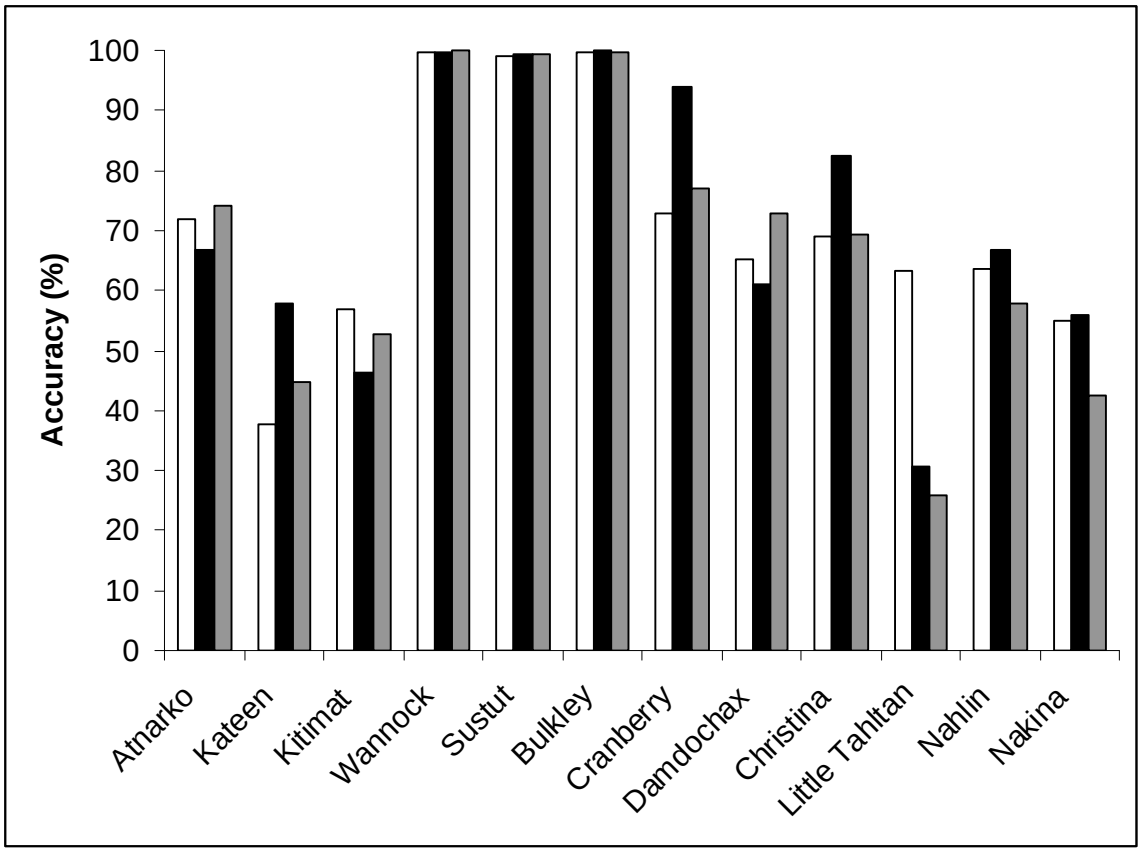

(a)
Accuracy (%)
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70

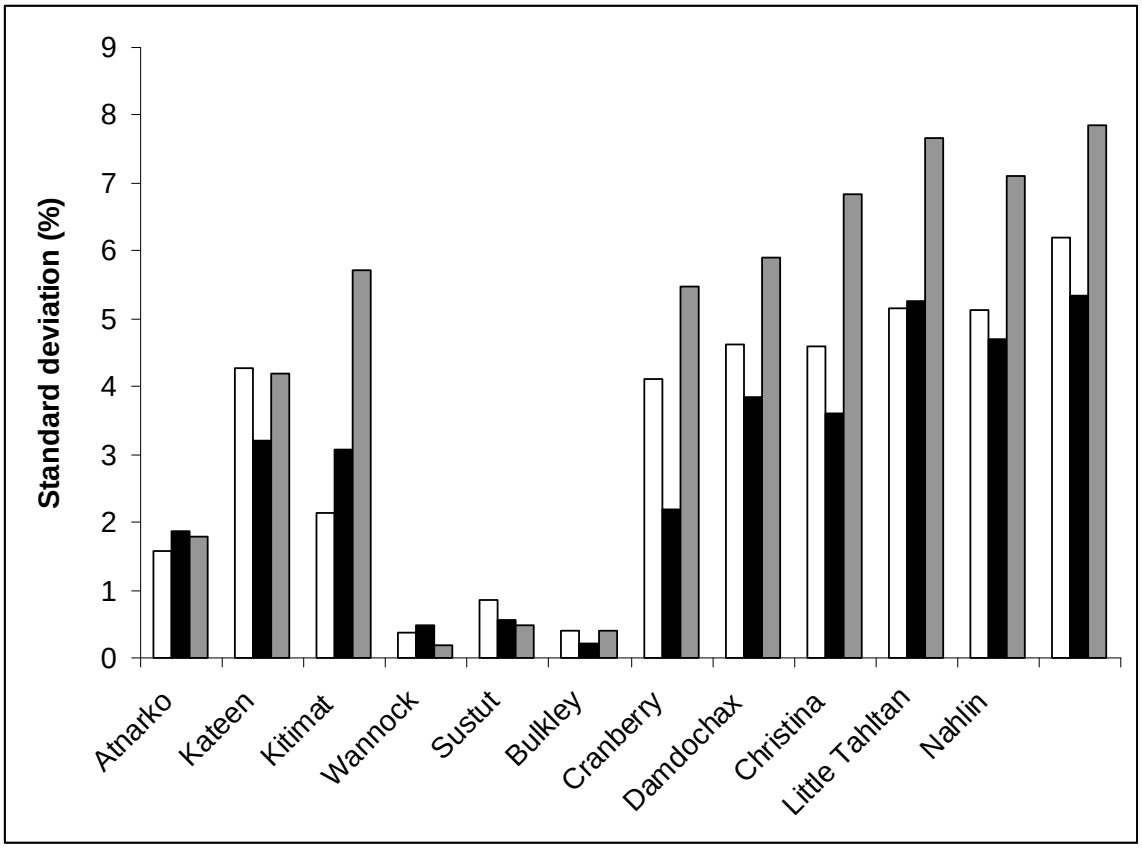

(b)
Standard deviation (%)
25
20
15
10
5
0
0
10
20
30
40
50
60
70
Number of alleles

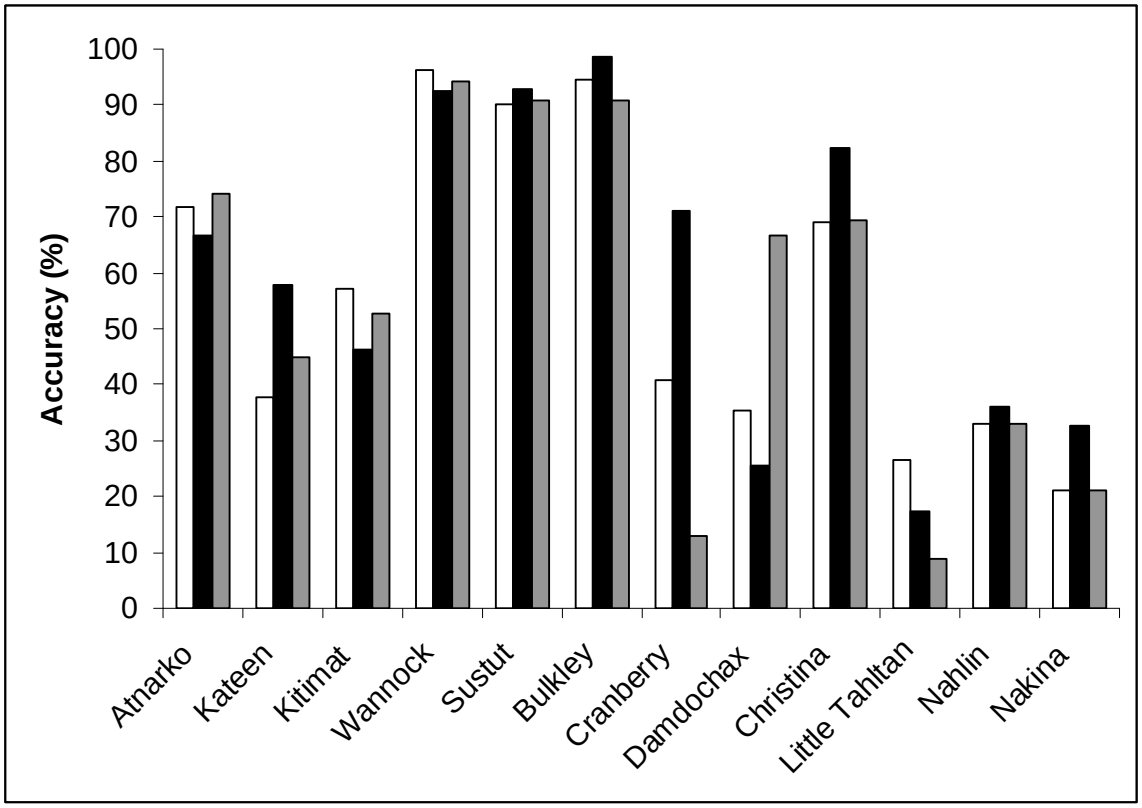

(a)
Accuracy (%)
90
80
70
60
50
40
30
20
10
0
0.40
0.50
0.60
0.70
0.80
0.90
1.00
(b)
Accuracy (%)
16
14
12
10
8
6
4
2
0
0.00
0.10
0.20
0.30
0.40
0.50
Heterozygosity

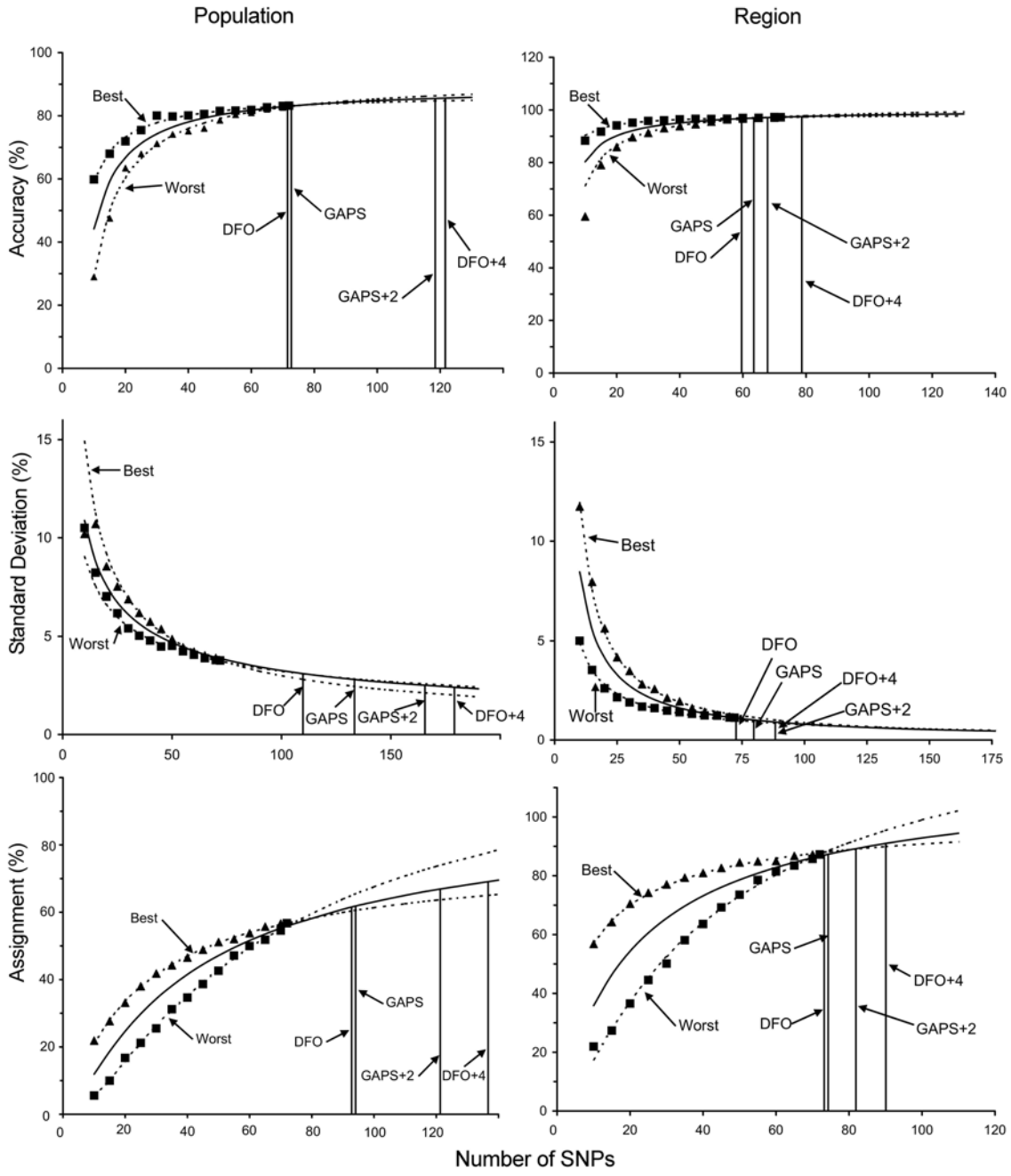

(a)
Accuracy (%)
90
80
70
60
50
40
30
20
10
0
0.00
0.05
0.10
0.15


(b)
Accuracy (%)
16
14
12
10
8
6
4
2
0
0
0.1
0.2
0.3
0.4
0.5
Fst


Accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
Atnarko
Kateen
Kitimat
Wannock
Sustut
Bulkley
Cranberry
Damdochax
Christina
Little Tahltan
Nahlin
Nakina


Standard deviation (%)
9
8
7
6
5
4
3
2
1
0
Atnarko
Kateen
Kitimat
Wannock
Sustut
Bulkley
Cranberry
Damdochax
Christina
Little Tahltan
Nahlin


Accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
Atnarko
Kateen
Kitimat
Wannock
Sustut
Bulkley
Cranberry
Damdochax
Christina
Little Tahltan
Nahlin
Nakina


Population
Region
Accuracy (%)
Standard Deviation (%)
Assignment (%)
Number of SNPs
Best
Worst
DFO
GAPS
GAPS+2
DFO+4