

NPAFC

Doc. 1105

Rev. _____

Microsatellite stock identification of chum salmon on a Pacific Rim basis and a comparison with single nucleotide polymorphisms (SNPs)

Terry D. Beacham¹, John R. Candy¹, C. Wallace¹, Shigehiko Urawa²,
Shunpei Sato², Natalia V. Varnavskaya³, Khai D. Le¹, and Michael Wetklo¹

¹Fisheries and Oceans Canada, Pacific Biological Station,
3190 Hammond Bay Road, Nanaimo, B. C., Canada V9T 6N7

²National Salmon Resources Center, Fisheries Research Agency, Toyohira-ku, Sapporo 062-
0922, Japan

³Kamchatka Fishery and Oceanography Research Institute, 18 Naberezhnaya Street
Petropavlovsk-Kamchatsky 683000, Russia

Submitted to the

NORTH PACIFIC ANADROMOUS FISH COMMISSION

by

CANADA

October 2008

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:

Beacham, T.D., J.R. Candy, C.W. Wallace, S. Sato, S. Urawa, N.V. Varnavskaya, K.D. Le, and M. Wetklo. 2008. Microsatellite stock identification of chum salmon on a Pacific Rim basis and a comparison with single nucleotide polymorphisms (SNPs). NPAFC Doc. 1105. 77 pp. (Available at <http://www.npafc.org>).

Abstract

Variation at 14 microsatellite loci was analyzed for over 53,000 chum salmon (*Oncorhynchus keta*) sampled from 381 locations ranging from Korea to Washington, and the variation applied to estimate stock composition in mixed-stock fishery samples. High resolution of mixed-stock samples was possible, with reporting groups distributed as follows: Korean 1, Japan 7, Russia 8, Alaska 15, Canadian Yukon River 5, British Columbia 16, and Washington 5. The number of alleles observed at a locus was related to the power of the locus in providing accurate estimates of stock composition of single-population mixtures. Approximately 800 alleles were observed across the 14 microsatellites, providing the basis for high-resolution stock identification. Analysis of known-origin samples indicated that accurate regional estimates of stock composition were obtained. Estimated stock compositions of mixed-fishery samples from coastal Japan, the Sea of Okhotsk, the western Pacific Ocean, the Gulf of Alaska, and coastal British Columbia were quite different among samples, and clearly reflected the presence of local populations. Microsatellites have provided the ability to provide accurate estimates of stock composition from many locations in the Pacific Rim distribution of chum salmon. In comparison with single nucleotide polymorphisms (SNPs), the entire stock identification power of 30 mtSNPs was equal to or less than the power of a single microsatellite for six loci examined on a Pacific Rim basis. Accuracy and precision of stock composition estimates for western Alaska chum salmon populations provided by 61 SNPs were comparable to those provided by subsets of 1-3 microsatellites, with similar total allele counts in all marker sets. A minimum 300 and likely closer to 400 SNPs will probably be required on a Pacific Rim basis to provide accuracy and precision comparable to that currently obtained from the 14-microsatellite baseline.

Introduction

The chum salmon (*Oncorhynchus keta* Walbaum) displays one of the widest spawning distributions of Pacific salmon. In Asia, chum salmon are distributed from Korea and Japan in the south to the Arctic Ocean coast of Russia in the north, while in North America, the distribution ranges from California in the south to the Beaufort Sea coast in the north, and as far east as the Mackenzie River in North American Arctic distribution (Salo 1991). After fry emerge from the gravel nest in the spring, they generally move directly to marine residence, first to estuaries, and later in the year to nearshore and offshore waters. Cohabitation of North American and Asian chum salmon in the North Pacific and Bering Sea provides one of the more challenging regimes for stock identification, given the wide distribution from which chum salmon can originate and the complex structure that is likely in fishery samples from these locations.

Stock identification of salmon in mixed-stock fisheries is one of the key pieces of information fishery managers use to decide on the timing and area of local salmon fisheries. Various techniques have been used to provide estimates of stock composition, but in practical applications, the objective is generally to provide the greatest resolution among stocks or populations present, and to do so at the cheapest cost per fish that is practical. If stock composition information is critical to guide managers in opening fisheries in specific locations and at specific times during spawning migrations, then the technique applied must also enable rapid estimation of stock composition. Stock composition information is also vital in determining locations of ocean residence of specific stocks of immature chum salmon, and the migration routes used by immature salmon to reach seasonal rearing areas, as well as the routes used by maturing chum salmon to return to natal rivers.

Non-genetic and genetic techniques have been used to estimate stock compositions of chum salmon. An early non-genetic technique used was scale pattern variation, as variation in the scale characters used in the analyses reflect environmentally induced variation in growth rate. However, discrimination among populations was restricted to large geographic areas (Tanaka et al. 1969; Ishida et

al. 1989), although in some cases identification to local regions may be possible (Nikolayeva and Semenets 1983). Environmentally induced variation in trace element composition of otoliths has also been reported to be effective in stock identification in local applications (Sohn et al. 2005).

Early genetic techniques of stock identification centered on allozymes, whereby genetic structure of populations potentially contributing to a mixed-stock fishery was determined (Okazaki 1982a,b; Winans et al. 1994; Wilmot et al. 1994; Seeb and Crane 1999a; Efremov 2001) and then this structure was used to estimate stock composition of samples from mixed-stock fisheries (Beacham et al. 1987; Wilmot et al. 1998; Winans et al. 1998; Seeb and Crane 1999b). With the advent of direct surveys of DNA-based variation, sequence variation in mitochondrial (mt) DNA was employed to evaluate population structure (Ginatulina 1992; Sato et al. 2001). Subsequently, 30 mitochondrial (mt) single nucleotide polymorphisms (SNPs) have been used to estimate stock composition of chum salmon in the Bering Sea and North Pacific Ocean (Moriya et al. 2007), although the level of resolution achieved (Japan, Russia, North America) was less than that previously observed with allozymes or indeed scales. SNPs have also been used for stock identification of chum salmon on a local scale for populations from western Alaska (Smith and Seeb 2008). Minisatellite variation has been used to evaluate population structure in chum salmon (Taylor et al. 1996; Beacham 1996), but has not been employed in any significant way in estimation of stock composition.

Microsatellites have been demonstrated to be effective in determining population structure of chum salmon and estimating stock composition in fisheries in local areas in both Asia (Beacham et al. 2008b,c) and North America (Beacham et al. 2008d, in press). Although microsatellites have been demonstrated to be effective in estimation of stock composition in local chum salmon fisheries with local baselines, it remains to be demonstrated that microsatellites provide high resolution stock composition estimates for complex mixed-stock samples as may be encountered in North Pacific and Bering Sea sampling.

Microsatellites and single nucleotide polymorphisms (SNPs) are the two main classes of DNA markers currently used in salmon stock identification applications. In salmon population structure and

stock identification studies, the number of alleles observed at microsatellite loci ranges from fewer than 10 alleles to over 100 alleles, while most SNP loci display two alleles. For microsatellites, it is clear that a locus with greater numbers of alleles generally provides more accurate and precise estimates of stock composition than a locus with fewer numbers of alleles (Beacham et al. 2005, 2006, 2008a). As most SNPs display only two alleles, they will generally be less powerful than a microsatellite in stock identification applications. Although extensive microsatellite baselines exist for Pacific salmon (Beacham et al. 2001, 2005, 2006, 2008b,c, unpublished data), and SNP baselines are being developed for a number of salmon species (Sato et al. 2004; Smith et al. 2005), an important question that remains unanswered 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 a high-resolution microsatellite baseline.

Two studies are available to allow a comparison of stock identification accuracy and precision between the power of SNPs and the microsatellites reported in the current study. Sato et al. (2004) surveyed variation in 30 mt SNPs in 48 populations across the Pacific Rim, and this baseline was subsequently used by Moriya et al. (2007) to estimate stock compositions of chum salmon collected in the Bering Sea and North Pacific Ocean. In a second study, Smith and Seeb (2008), in a study on chum salmon populations from western Alaska, compared stock composition estimates derived from 61 SNPs with those derived from 15 microsatellites. They reported that the 61 SNPs outperformed subsets of 1-2 microsatellite markers in every single reporting region, and the 61 SNPs were described as having performed considerably better than the complete microsatellite baseline. Smith and Seeb (2008) indicated that the power of the hundreds of alleles present in existing microsatellite baselines could be replicated with far fewer SNP alleles, essentially because SNP alleles were proposed to have a higher information content. Smith and Seeb (2008) indicated that the microsatellites examined were chosen from the set of microsatellites reported in the current study or used by the U. S. Fish and Wildlife Service for chum salmon studies in western Alaska. The current study includes 14 microsatellites, of which

six loci were used by Smith and Seeb (2008) in their comparisons. Two main questions are important to evaluate: 1) Can the stock identification power of the 14-locus baseline reported in this study be replicated by the 30 mtSNPs employed by Moriya et al. (2007) or the 61-SNP baseline reported by Smith and Seeb (2008)?, 2) If not, how many SNPs are necessary to replicate the power of the existing microsatellite baseline for local applications such as in western Alaska chum salmon, and how many SNPs may be necessary for Pacific Rim applications?

In the current study, we evaluate the utility of using variation at 14 microsatellite loci for stock identification applications to region-specific identification of chum salmon over much of its natural range. This evaluation is conducted by examining the accuracy and precision of estimated stock compositions for individual microsatellites, combinations of microsatellites, and for all microsatellites combined through analysis of simulated mixtures and estimation from actual samples from fisheries in coastal Japanese waters, the Sea of Okhotsk, the western North Pacific Ocean, the eastern North Pacific Ocean, and coastal British Columbia, with the mixtures resolved using a 381-population baseline incorporating populations from Korea, Japan, Russia, the United States, and Canada. We demonstrate that sufficient population allele frequency variation exists at 14 microsatellites in chum salmon to enable accurate estimation of stock composition of mixed-stock samples to 59 reporting regions on a Pacific Rim basis. We also evaluate the accuracy and precision of estimates of stock composition derived from a 48-population, 30-mtSNP Pacific Rim baseline and an 8-population, 61-SNP western Alaska baseline with those derived from the current 14-microsatellite Pacific Rim baseline.

Methods and Materials

Collection of DNA samples and laboratory analysis

Tissue samples were collected from mature chum salmon, 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 14 microsatellite loci were conducted: Ots3 (Banks et al. 1999), Oke3 (Buchholz et al. 2001), Oki2 (Smith et al. 1998), Oki100 (Beacham et al. 2008b), Omm1070 (Rexroad et al. 2001), Omy1011 (Spies et al. 2005), One101, One102, One104, One111, and One114 (Olsen et al. 2000), Ots103 (Nelson and Beacham 1999), Ssa419 (Cairney et al. 2000), and OtsG68 (Williamson et al. 2002).

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 H₂O. 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 and were outlined by Beacham et al. (in press). PCR fragments 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 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.

Baseline populations

The baseline survey consisted of analysis of over 53,000 chum salmon from 381 populations from Korea, Japan, Russia, Alaska, Canada, and Washington (Figure 1). The

sampling sites or populations surveyed in each geographic region are outlined in Appendix Table 1. The major geographic regions and river drainages outlined in Appendix Table 1 are indicated in Figure 1. Information on regional population structure has been outlined previously for Japanese populations (Beacham et al. 2008c), Russian populations (Beacham et al. 2008b), western Alaska populations (Beacham et al. in press) and British Columbia populations (Beacham et al. 2008d). 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

Estimation of stock composition in single-population samples

Genotypic frequencies were determined for each locus in each population and the Statistical Package for the Analysis of Mixtures software program (SPAM version 3.7) (Debevec et al. 2000) was used to estimate stock composition of simulated single-population samples. The Rannala and Mountain (1997) correction to baseline allele frequencies was used in the analysis 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 loci were considered to be in Hardy-Weinberg equilibrium, 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 150 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.

The accuracy and precision of estimated stock compositions for simulated single-population samples was evaluated for each microsatellite locus individually for 40 single-

population samples spanning a Pacific Rim distribution of the populations. The mean accuracy and precision observed from these 40 simulated samples was compared with the number of alleles observed in each locus. The 40 populations were distributed as follows: Korea: Namdae (1); Japan: Chitose (2), Teshio (3), Tokachi (4), Shiriuchi (5), Yurappu (6), Tokoro (7), Shibetsu (8), Gakko (9), Sakari (10); Russia: Amur (11), Naiba (12), Ola (13), Vorovskaya (14), Ossora (15), Anadyr (16); Alaska: Kobuk (17), Pilgrim(18), Andreafsky (19), Cheena (20), Goodnews (21), Frosty (22), Volcano Bay (23), Stepovak Bay (24), DIPAC hatchery (25); Canada: Kluane River (26), Fishing Branch (27), Steel (28), Stanley (29), Kshwan (30), Nangeese (31), Wilson (32), Kitasoo (33), Nekite (34), Homathko (35), Nanaimo (36), Nitinat (37), Inch (38); Washington: Bitter (39), Hoodsport (40).

Estimation of sample sizes required for 90% correct assignment

Average baseline population sample sizes required for 90% correct assignment to population and reporting region were determined by simulating 108 single-population samples and determining the accuracy recorded on both a population and reporting region level. The 108 populations were distributed regionally as follows: Korea 1, Japan 26, Russia 34, United States 24, and Canada 23.

Estimation of stock composition in mixed-stock samples

Analysis of simulated single-population samples with SPAM provided the initial evaluation of the utility of the baseline for stock composition analysis. The next stage of the analysis was to analyze multi-population, multi-reporting region samples from a single geographic area. The key assumption in the simulations is that the baseline used will be representative of populations present when it is applied to mixed-stock fishery samples. At this stage, as fewer simulations were required, analysis of the simulated fishery samples was conducted with a Bayesian procedure (cBayes). The BAYES routine of Pella and Masuda (2001)

was modified by our laboratory to a C++-based program (cBayes), which is available from our laboratory website (Neaves et al. 2005). cBayes analyses required substantially more computer analytical time than did the SPAM software for analysis of an individual sample. Previous applications of both SPAM and cBayes to the same mixed-stock sample suggested that accuracy was improved with the cBayes application (Beacham et al. 2005; unpublished results). In the analysis, ten 20,000-iteration Monte Carlo Markov chains of estimated stock compositions were produced, with initial starting values for each chain set at 0.90 for a particular population that was different for each chain. Estimated stock compositions were considered to have converged when the shrink factor was < 1.2 for the 10 chains (Pella and Masuda 2001), and thus the starting values were considered to be irrelevant (uninformative prior). Stock composition estimates converged before 20,000 iterations, and no further improvements in the estimates were observed in excess of 20,000 iterations. Therefore, 20,000 iterations were set as the standard in the analysis. The last 1,000 iterations from each of the 10 chains were then combined, and for each fish the probability of originating from each population in the baseline was determined. These individual probabilities were summed over all fish in the sample, and divided by the number of fish sampled to provide the point estimate of stock composition. Standard deviations of estimated stock compositions were determined from the last 1,000 iterations from each of the 10 chains incorporated in the analysis.

The next stage of the analysis was to analyze multi-reporting region simulated fishery samples encompassing several geographic regions. Six complex multi-reporting region mixtures were evaluated, and the accuracy and precision of estimated stock compositions were determined on a reporting region basis. The final stage in the analysis prior to application in actual fishery samples was to estimate stock compositions of known-origin samples that were completely independent of the baseline used in the estimation. Known-origin samples were developed by randomly removing individual fish from populations in the baseline, recompiling allele

frequencies for all populations in the baseline, and then using this modified baseline to estimate stock compositions in the samples of known origin.

Eight marine fishery samples of unknown composition were analyzed to compare the performance of the baseline in estimating stock compositions. The wide geographic distribution of these samples, spanning from domestic test fisheries in Japan to those in British Columbia, suggested that divergent estimates of stock composition should be obtained when analyzed with the Pacific Rim baseline.

Comparison with SNPs

Comparison of stock composition estimates derived from the 14-microsatellite baseline with those from the 30-*mt*, 48-population SNP baseline outlined by Sato et al. (2004) was done in the following manner. The 381-population microsatellite was reduced to the equivalent SNP baseline, keeping the same populations in both baselines to the extent possible. However, some substitutions were necessary, as not all 48 populations in the survey conducted by Sato et al. (2004) were included in the microsatellite baseline. Substituted populations were from the same regional group as the original population. For Japanese populations, the Sakari River and Miomote River populations were substituted for the Otsuchi River and Jintsu River populations, respectively. Populations substituted in the Alaskan portion of the baseline included replacing Salmon River with the Noatak River in western Alaska, Belkofski River with Delta Creek on the Alaska Peninsula, Kizhuyak River with Keta Creek in central Alaska, and Long Bay with the DIPAC hatchery, Whale Bay with Gambier Creek, and Port Beauclerc with Neets Bay in southeast Alaska. In Washington, Blackjack Creek was replaced with Skagit River, and Hamilton Creek was replaced with Kennedy Creek. The 30 *mt*SNPs were treated as a single locus with 30 alleles, with the observed allele frequencies as model inputs. Estimated stock compositions were

allocated to individual populations and then summed into six regional groups: Korea, Japan, Russia, Alaska, British Columbia, and Washington.

Comparison of the accuracy and precision of the stock composition estimates derived from the 61-SNP baseline reported by Smith and Seeb (2008) was conducted by restricting the baselines to those populations with at least 80 fish surveyed. The microsatellite baseline included 8 of the 11 populations surveyed by Smith and Seeb (2008). These eight populations were distributed in Norton Sound (Kwiniuk River, Niukluk River, Nome River), the Yukon River (Andreafsky River in the lower Yukon River and Sheenjek River in the middle portion of the drainage), the Kuskokwim River (Aniak River in the lower drainage), the Nushagak River in Bristol Bay (Mulchatna River), and Frosty Creek on the Alaska Peninsula. Genotypic frequencies were determined for each locus in each population and SPAM was used to estimate stock composition of simulated mixtures. For One111, only the first 99 alleles could be incorporated in the analysis. Accuracy and precision of estimated stock compositions were derived as described previously in the section on estimation of stock composition of single-population samples.

Projections of the number of number of SNPs required for equivalency of the current microsatellite baseline were conducted by ranking SNPs according to the F_{ST} value provided by Smith and Seeb (2008). Single-population mixtures were analyzed with 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, and 61 SNPs. SNPs with the highest F_{ST} values were initially incorporated in the SPAM analyses of the single-population mixtures, with lower-valued F_{ST} SNPs sequentially added to the analyses. Average population accuracy and precision were recorded for each set of SNPs. Subsequently, SNPs with the lowest F_{ST} values were initially incorporated in the SPAM analyses, with progressively higher-valued F_{ST} 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 merging 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. Smith and Seeb (2008) indicated that not all SNPs reported were independent, and SNPs displaying linkage disequilibrium were phased into haplotypes. Frequencies of these phased haplotypes or genotypes were not available to us, so we assumed that all 61 SNPs were independent. This assumption would actually increase the power of the SNPs examined relative to the phased alleles.

Results

Population sample size

The effect of baseline population sample size on accuracy of estimated stock compositions for single-population mixtures was evaluated for 108 populations covering a wide geographic range and sample size. Substantial variation in accuracy was observed with respect to individual population sample size. The relationship between population percent correct assignment and population sample size was described by the equation:

$$\% \text{ correct accuracy} = 62.044 N^{0.0589}$$

where N is the number of fish surveyed in the population ($r^2=0.16$) (Figure 2a). An average sample size of approximately 550 fish was projected to be required for an accuracy of 90% to population. Accurate estimation of regional contributions is a less demanding objective, and the relationship between reporting region percent correct and population sample size was described by the equation:

$$\% \text{ correct accuracy} = 63.799 N^{0.0675}$$

($r^2=0.38$) (Figure 2b). A population sample size of approximately 165 fish was projected to be required for an average accuracy of 90% to a specific region. Larger sample sizes were required to obtain the same level of population-specific accuracy compared with region-specific accuracy.

Power of individual microsatellites

The number of alleles observed at the 14 loci examined in survey of microsatellite variation across the Pacific Rim ranged from 26 to 149 alleles (Table 1). The number of alleles observed at a locus was related to estimated accuracy of stock composition of the single-population mixtures of the 40 test populations. This relationship was described by the equation:

$$\% \text{ correct accuracy} = 16.55 N^{0.2829}$$

where N is the number of alleles observed at the locus ($r^2=0.32$) (Figure 3a). Mean estimated stock compositions of the single population mixtures (correct=100%) were 45.0% for single loci with < 40 alleles, 49.2% for loci with 40-49 alleles, 53.3% for loci with 50-59 alleles, and 53.7% for loci with > 59 alleles (Table 1). The number of alleles observed at a locus also had a marked effect on precision of estimated stock compositions of single population mixtures. The relationship between allele number and precision (standard deviation, SD) was:

$$SD=119.04 N^{-0.5517}$$

where N is the number of alleles observed at the locus ($r^2=0.69$, $P<0.01$) (Figure 3b). Mean standard deviations of the estimated stock compositions were 17.9% for single loci with < 40 alleles, 14.6% for loci with 40-49 alleles, 13.7% for loci with 50-59 alleles, and 11.8% for loci with > 59 alleles (Table 1). Similar results were observed for regional estimates of stock composition, with higher accuracy and lower standard deviations observed compared with average population-specific estimates. Loci that displayed more alleles during the survey of microsatellite variation generally provided more accurate and precise estimates of stock composition of the single-population mixtures than did loci with fewer observed alleles.

Number of alleles in relation to accuracy and precision of estimated stock compositions

The number of alleles employed in estimation of stock composition had a direct effect on accuracy and precision of estimates stock compositions. Starting with the locus with the highest number of alleles (One111) and sequentially adding loci with progressively fewer alleles resulted

in stock composition estimates of continually increasing accuracy (Figure 4a). This relationship was described by the equation:

$$\% \text{ correct accuracy} = 44.245 N^{0.1058}$$

where N is the number of alleles employed in the estimation ($r^2=0.98$). The rate of increase in accuracy per allele generally decreased as more alleles were used in stock composition estimation. As accuracy of estimated stock compositions was about 75% at 100 alleles, the scope for increased improvement was a maximum 25%, and thus a reduction in the effectiveness of each additional allele was not unexpected. The addition of loci and alleles for stock composition estimation always produced, on average, more accurate results. There was no indication of any decline in accuracy with increasing numbers of loci or alleles used in the estimation.

Precision of estimated stock compositions was directly influenced by the number of alleles employed in estimation. Continually adding loci (increasing allele number) resulted in stock composition estimates of continually increasing precision (Figure 4b). This relationship was described by the equation:

$$\% \text{ SD} = 79.878 N^{-0.4962}$$

where N is the number of alleles employed in the estimation ($r^2=0.99$). Relatively significant increases in precision were observed with increasing numbers of alleles, with over 50% reduction in standard deviation a seven-times increase in allele number. The addition of loci produced, on average, more precise estimates of stock composition. In summary, the number of alleles employed in the estimation of stock composition directly influenced the accuracy and precision of the estimates, with higher accuracy and precision obtained by employing the maximum number of alleles.

Analysis of simulated single-geographic region mixtures

Accurate estimates of stock composition were obtained when simulated mixtures containing chum salmon of only Asian origin were evaluated. For example, for a simulated

sample comprised solely of fish from the Namdae River in South Korea, stock composition was estimated at 100% (SD=0.5%) South Korean composition, with the 381-population baseline used in the analysis. A mixture containing only Japanese chum salmon with equal weighting of the seven defined reporting regions listed in Appendix Table 1 was estimated to contain 100% (SD=0.5%) Japanese chum salmon, with an average reporting region error of <1% (Figure 5). Similar results were observed for a simulated mixture of chum salmon from eight reporting regions in Russia. Analysis of the simulated mixtures suggested that accurate regional and national estimates of stock composition when applied to samples containing only Asian chum salmon.

For samples of North American origin, accurate geographic estimates of stock composition were generally observed. Stock composition of a simulated sample composed largely of chum salmon from western Alaska was estimated at 100% (SD=0.7%) western Alaskan origin, but the average reporting region error was 3.6% (Figure 5). The largest sources of reporting region error centered on the lower Yukon summer run and Nushagak River components. Estimated stock composition of a mixture of central and southeast Alaska origin chum salmon was estimated at 100% geographic origin composition, with the estimates for the reporting groups generally within 2% of actual values (Figure 5). Similarly, estimated stock composition of a sample containing chum salmon of only Canadian Yukon River and Mackenzie River origin was estimated at 100% geographic origin composition, with the error in estimates for the reporting groups generally < 1%. Estimated geographic stock composition of a simulated sample of chum salmon of solely northern British Columbia origin was 100% (SD=0.5%), with the estimates for reporting groups within 2% of actual values. The largest error was observed in the North Coast reporting group, with a 1.9% estimated error (Figure 5). Estimated stock composition of a sample of chum salmon of southern British Columbia origin was 99.7% accurate (SD=0.7%), with the error in the estimates for the reporting groups generally < 2%. Finally, stock composition of a sample of Washington chum salmon was estimated with a

100% geographic accuracy (SD=0.4%), with errors in the estimated stock composition of the reporting groups < 1% (Figure 5).

Analysis of simulated multi-country mixtures

Accuracy and precision of estimated stock compositions were evaluated for six multi-population, multi-region, and multi-country simulated mixture samples. The estimated stock composition of a simulated mixture of chum salmon from South Korea, Japan, and Russia was within 2% of the correct percentage for a specific population, 3% for the regional reporting group, and 1% for country of origin (Table 2, mixture 1). The stock composition of a simulated mixture of chum salmon from Japan, Russia, and western Alaska was estimated with an accuracy of within 1% of actual compositions for population, reporting region, and country of origin (Table 2, mixture 2). A simulated mixture containing primarily chum salmon from western Alaska and the Yukon River drainage (with some contribution from west Kamchatka and the northern Alaska Peninsula) was estimated with an accuracy within 4% of actual values for specific populations and reporting regions, and within 2% for country of origin (Table 2, mixture 3). The fourth mixture evaluated contained only simulated chum salmon from southeast Alaska and northern and central British Columbia. For this mixture, the estimated stock composition was within 2% for population and reporting region components, and 1% for country of origin. The last regionally-based mixture evaluated was composed of simulated chum salmon from southern British Columbia and Washington. In this case, the estimated stock composition was within 5% of actual values for population and reporting region, and 1% for country of origin (Table 2, mixture 5). With a Pacific Rim distribution of populations in a simulated mixture, the accuracy of the estimate was within 2% of actual values for population and reporting region, and 1% for country of origin (Table 2, mixture 6). Analysis of all single geographic region and multi-country simulated mixtures indicated that reliable estimates of stock composition should be provided by the 14 microsatellites evaluated in the study.

Analysis of known-origin mixtures

Analysis of simulated mixtures provided an initial evaluation of the reliability of estimated stock compositions derived from applying the 14 microsatellites surveyed in the study. This initial assessment indicated that accurate estimates of stock composition should be available on a reporting region basis, and certainly on a country-of-origin basis. A more difficult test of the effectiveness of the baseline and markers used is to estimate stock compositions of known-origin mixtures completely independent of the baseline used for stock composition. Four mixtures of known origin were generated by randomly extracting specific numbers of fish from populations with larger sample sizes and recalculating observed allele frequencies for the populations. The first sample evaluated contained chum salmon of only Asian origin. The stock composition of this sample was estimated with an accuracy within 2% of actual values for eight of 13 populations present in the sample, nine of 11 reporting regions, and with an accuracy of within 2% for country of origin (Table 3, mixture 1). The second sample evaluated contained chum salmon of Russian, Alaska, and Yukon River origin. Accuracy of estimated stock compositions were within 2% for eight of 14 populations present in the sample, as well as eight of 13 reporting regions. Accuracy of estimated country contributions were generally within 4% for the three countries represented (Table 3, mixture 2). The third sample evaluated was composed of chum salmon entirely from southeast Alaska and northern British Columbia. Accuracy of estimated stock compositions were within 3% for six of 10 populations present in the sample, within 3% for six of eight reporting regions present, and within 6% for country of origin (Table 3, mixture 3). The fourth and final sample evaluated was composed of chum salmon entirely from southern British Columbia and Washington. Accuracy of estimated stock compositions were within 3% for eight of 12 populations present in the sample, within 3% for six of eight reporting regions present, and within 5% for country of origin (Table 3, mixture 4). Analysis of these known-origin samples suggested that reasonably accurate estimates (within 5% of actual values) of stock composition

may be possible at times for specific populations, and that reasonably accurate estimates of stock composition were typically observed for regional reporting groups.

Analysis of marine samples

Analysis of simulated mixtures and known-origin samples provide some evaluation of the accuracy of estimated stock compositions that may be expected when applied to actual fishery samples of unknown origin. However, actual mixed-stock fishery samples typically contain individuals from populations not represented in the baseline used in stock composition estimation. This has the potential to result in biased estimates of stock composition. Estimation of stock composition from widely divergent samples can aid in evaluation of whether the estimated stock compositions are within expectations for the location and timing of sample collection. We tested the 14 microsatellites used in stock composition estimation by analyzing eight mixed-stock samples of chum salmon with divergent geographic and temporal origins. The first sample evaluated was obtained from maturing chum salmon sampled on their spawning migration near northern Hokkaido. Chum salmon from the Sea of Japan coast of Hokkaido were estimated to have accounted for 61% of the sample, with Chitose River the most abundant population. Chum salmon from the Sea of Okhotsk coast were estimated at 32% of the sample, with the Tokushibetsu River population the most abundant population. Chum salmon from the Sea of Japan coast of Honshu were estimated at 7% of the sample (Table 4). Chum salmon of only Japanese origin were estimated to be present in the sample. The next two samples evaluated were obtained from juvenile chum salmon in the Sea of Okhotsk. The first sample was obtained from the southwestern Sea of Okhotsk offshore from Sakhalin Island. Chum salmon from the Sea of Okhotsk coast of Hokkaido were estimated at 51% of the sample, with Sakhalin Island chum salmon estimated at 21% of the sample. Japanese-origin chum salmon were estimated at 75% of the sample, with Russian-origin chum salmon most of the remainder (Table 4). The second juvenile sample was obtained from the east central portion of the Sea of Okhotsk, off the coast of

west Kamchatka. Chum salmon from the west coast of Kamchatka were estimated at 81% of the sample, with chum salmon from the east coast Kamchatka estimated at 15% of the sample. Russian-origin chum salmon were estimated to be the dominant contributor to the sample. The fourth sample was obtained from immature and maturing chum salmon in the western Pacific Ocean off the southeast coast of Kamchatka. This sample was dominated by individuals of Japanese (67%) and Russian origin (32%), with the Sea of Okhotsk stock the major Japanese contributor and the east and west coasts of Kamchatka stocks the major Russian contributors (Table 4).

The remaining four samples evaluated were all obtained from eastern Pacific locations, with two samples from the Gulf of Alaska and two from coastal waters of British Columbia. In the Gulf of Alaska, winter samples of largely immature fish indicated that fish of North American origin were distributed in more northern locations and fish of Asian origin were distributed in more southern locations. For example, in a sample from the central east Gulf of Alaska, chum salmon of Canadian origin comprised 64% of the sample evaluated, with American-origin fish 25%. Major Canadian regional components were chum salmon of Central Coast (18%) and Fraser River origin (17%), with major American components of southeast Alaska (13%) and Prince William Sound (6%) (Table 4). Further south, Asian-origin chum salmon were the main contributors to the sample, with Japanese chum salmon estimated at 37% and Russian chum salmon at 29% (Table 4). North American contributions were estimated to be derived largely from southeast Alaska (14%) and Prince William Sound (9%). One sample was obtained from migrating maturing chum salmon in coastal waters in southern British Columbia from Johnstone Strait. Virtually all of the salmon sampled were estimated to be of southern British Columbia origin, with the sample dominated by chum salmon of Fraser River (54%), east coast Vancouver Island (28%), and southern British Columbia coastal (14%) origin (Table 4). The final sample evaluated was obtained from maturing individuals in Bute Inlet in southern British Columbia, which is the most terminal on a chum salmon migration route of the eight samples evaluated. For

this sample, virtually 100% of the sample was estimated to be of South Coast origin (Table 4), with two specific populations, Homathko River at 55% and Southgate River at 44%, contributing all of the fish in the sample.

Comparisons with SNPs

The first analysis conducted compared accuracy and precision for regional stock composition estimates derived solely for 30 mtSNPs and single microsatellites with a Pacific Rim distribution of populations. For 30 mtSNPs, estimated accuracy of single-region stock compositions estimated for six regions varied between 50-87%, with a 73% mean accuracy over all regions. Mean estimated regional accuracy for six individual microsatellite loci varied from 71-85%, with an average accuracy of 77%, 4% higher than observed with the 30 mtSNPs (Fig. 6a). In terms of accuracy, any one of six loci examined provided comparable or more accurate estimates of regional stock composition than did the 30 mtSNPs. Standard deviation of the mtSNP estimates averaged 18% over all regions. Mean estimated standard deviation for the six microsatellite loci investigated ranged from 6%-12%, with an average SD of 9% (Fig. 6b). Average SD of the regional estimated stock composition from the 30 mtSNPs was twice as large as that derived from a single microsatellite locus. On average, the entire stock identification power of the 30 mtSNPs was equal to or less than the power of a single microsatellite locus for the six loci examined.

The second analysis conducted compared population-specific accuracy and precision of stock composition estimates for 61 SNPs and sets of microsatellites for populations with only a western Alaska distribution. Two populations (Sheenjek River, Frosty Creek) were estimated with accuracies over 90%, regardless of whether SNPs or microsatellites were employed. The other populations were more difficult to identify, regardless of the marker class used. The average level of correct population assignment achieved by using 61 SNPs (61 independent alleles, 122 total alleles) could be also achieved by using only a single microsatellite (One111, 99

alleles incorporated), two three-microsatellite groups (Group 1: One114, OtsG68, One102, 155 alleles in western Alaska, Beacham et al. in press); Group 2: One101, Ots103, Omm1070, 139 alleles), and a four-microsatellite group (Omy1011, One104, Oki100, Ssa419, 127 alleles) (Figure 7a). On average, accuracy results from the 61 SNPs (122 total alleles) were replicated by a set of microsatellites incorporating 130 total alleles. A higher average correct population assignment was achieved with the 14 microsatellites compared with 61 SNPs. The average standard deviation observed by employing 61 SNPs was again replicated with the subsets of microsatellites (Figure 7b). Standard deviations declined by approximately 50% when the 14 microsatellites were used in stock composition estimation compared with 61 SNPs. Accuracy and precision of stock composition estimates provided by the 61 SNPs were comparable to those provided by a subset of 1-3 microsatellites, with similar total allele counts in all marker sets. The full 14 microsatellites provided, on average, more accurate and precise estimates of stock composition than did the 61 SNPs.

How many SNPs of quality comparable to those reported by Smith and Seeb (2008) would it require to provide results equivalent to those available from the 14 microsatellites for the eight populations evaluated? The relationship between % correct assignment and number of SNPs was described by the equation:

$$\% \text{ correct assignment} = 49.955 N^{0.1054}$$

where N is the number of SNPs used in the analysis ($r^2=0.998$) (Figure 8a). The 14 microsatellites achieved on average 84.7% correct population assignment, which is predicted to require 150 SNPs. The relationship between standard deviation and number of SNPs was described by the equation:

$$SD = 50.183 N^{-0.4847}$$

where SD is the standard deviation observed ($r^2=0.995$) (Figure 8b). The 14 microsatellites produced on average a standard deviation of 3.8, which is predicted to require 205 SNPs for comparable results.

Discussion

Population structure and sample size

A regionally-based population structure is typically required in the application of genetic markers for stock composition estimation, as an important assumption in the application is that the portion of the mixed-stock sample derived from unsampled populations is allocated to sampled populations from the same region. This assumption reduces the cost and complexity of developing a baseline for stock composition analysis. For chum salmon, regionally-based population structure has been observed in analysis of microsatellite variation. For example, populations from Japan and Russia were generally arrayed in clusters with a regional basis (Beacham et al 2008b,c). Within the Yukon River, regional stock structure has been successfully employed in implementing microsatellite-based analysis of mixed-stock samples during the fishing season (Flannery et al. 2007). In British Columbia, regional clustering of populations was observed in the large river drainages like the Taku, Skeena, and Fraser systems. Geographically-based clustering of populations in smaller river drainages was observed for populations on the Queen Charlotte Islands, the central coastal coastal area, and the east and west coasts of Vancouver Island (Beacham et al. 2008d). Chum salmon population structure thus meets the important condition that unsampled populations contributing to mixed fishery samples will likely be allocated to sampled populations in the same region.

The number of fish sampled in a population was observed to influence accuracy of estimated stock compositions. In our study, the number of fish surveyed in a population varied considerably, from fewer than 20 individuals to approximately 600 individuals in a population (Appendix Table 1). With variable population sample sizes, estimation of allele frequencies was possibly subject to sampling error for populations with smaller numbers of fish surveyed,

particularly for those loci like One111 with larger numbers of alleles. However, analyses of population structure indicated that small sample size populations were usually grouped with geographically proximate neighbors, such as the three populations from Primorye (17-49 fish sampled) (Beacham et al. 2008b). Therefore, it seems likely that sampling errors in allele frequencies have not masked existing population structure. If baseline population sample sizes of 20 to 30 fish were adequate to provide expected population structure, they were included in a baseline for stock identification applications. Clearly though, enhancement of the lower baseline population sample sizes would be required if increased accuracy and precision of estimated stock compositions were necessary.

Number of alleles

The number of alleles observed at a microsatellite locus is a key predictor in the effectiveness of a locus in determination of stock composition. Although early computer simulation studies suggested that loci with fewer than 10 alleles were the most effective for studies in population structure and stock identification (Bernatchez and Duchesne 2000), subsequent simulation studies suggested that loci with more alleles were more powerful than those with modest numbers of alleles (Kalinowski 2002; 2004). Studies with simulated loci are an initial useful stage to evaluate power in stock identification applications, but empirical applications provide more compelling evidence on which to base conclusions concerning locus stock identification power. Studies conducted in our laboratory on sockeye salmon (Beacham et al. 2005) and Chinook salmon (Beacham et al. 2006; 2008a) stock identification have consistently indicated that the number of alleles observed at a microsatellite locus was positively correlated to the power of the locus in providing accurate and precise estimates of stock composition. Similar results were observed with chum salmon, but there was certainly some variation in the relationship between allele number and accuracy. In general though, loci with larger numbers of

alleles were more effective in providing more accurate and precise estimates of stock composition than were loci with smaller numbers of alleles.

Does employing more microsatellite alleles always produce more accurate and precise estimates of stock composition? Previous studies on sockeye (Beacham et al. 2005) and Chinook salmon (Beacham et al. 2006) indicated that adding more loci consistently resulted in estimates of higher accuracy and precision, with the effect more pronounced on precision. Both studies indicated that there was rapid improvement in accuracy of estimated stock compositions until 80-100 alleles were employed in the analysis. The use of more than 100 alleles for stock composition estimation resulted in diminishing returns for accuracy per allele employed, whereas variance of the estimates continued to decline. Similar results were observed for chum salmon, as comparable accuracy of estimated stock compositions was about 75% at 100 alleles and increased more slowly to about 90% at 750 alleles. There is clearly scope to modify the set of loci used in stock identification applications to meet the level of accuracy and precision required in the application. For applications where the required resolution of stock composition estimates is less than that possible from the suite of 14 microsatellites used in the current study, then clearly fewer loci can be used in estimating stock composition, reducing the cost per fish of the analysis.

Accuracy and precision of estimated stock compositions

Initial evaluation of genetic stock identification techniques involves analysis of simulated mixtures to evaluate the accuracy and precision of estimated stock compositions. In our study, analysis of simulated single geographic region and multi-geographic region mixed-stock fishery samples indicated that reliable estimates of stock composition were obtained. The next phase of the evaluation required that analysis of known-origin samples independent of the baseline used for estimation of stock compositions. These known-origin samples generally provided reasonably accurate estimates of stock composition for reporting regions and countries. However, even if reliable estimates of stock composition have been obtained from both simulated and known-

origin samples, there is still a potential for inaccurate estimates of stock composition in real fisheries applications if a significant portion of the mixed-stock sample has been derived from populations or regions inadequately represented in the baseline. The application of the baseline to estimation of stock composition for actual mixed-stock samples is a means to evaluate whether the presence of unsampled populations in the mixture will cause bias in estimated stock compositions.

In our study, we examined eight mixed-stock samples with a Pacific Rim distribution of the samples. The sample from coastal Japanese waters was estimated to be dominated by the Chitose River population which is the most abundant of any population along the Sea of Japan coast of Hokkaido or Honshu. Given the location of the sample, as it was on the migration route for Sea of Japan populations, it was reasonable that Chitose River was estimated as a significant contributor to the sample. Similarly, the Tokushibetsu River was adjacent to the sampling location, so again a significant contribution by this population to the sample is not unexpected. Estimated stock compositions of samples of juveniles from the Sea of Okhotsk were again consistent with physical locations. A sample off Sakhalin Island displayed a significant Sakhalin Island contribution, and a sample off west Kamchatka displayed a significant component from west Kamchatka. A sample from the western Pacific Ocean was dominated by Japanese and Russian chum salmon, with North American chum salmon becoming more abundant in Gulf of Alaska samples. A sample from Johnstone Strait was estimated to be composed of populations that were regionally adjacent to known migration routes of maturing individuals. This result was accentuated in the results of the sample from Bute Inlet in southern British Columbia. There are three main rivers in the inlet in which chum salmon spawn (Orford, Homatho, and Southgate). Chum salmon return to the Orford River earlier than to the other two rivers, and have largely returned to the river by early October. As the sample from Bute Inlet was collected on 11 October, it is plausible to assume that no Orford River chum salmon should be present in the sample. With 381 populations in the baseline, the Homathko River and Southgate River were

identified as contributing 99% of the fish in the sample. In summary, microsatellites provided reliable estimates of stock composition in local fishery samples even when there was a Pacific Rim distribution of populations potentially available for allocation.

Comparisons with SNPs

When applied on a Pacific Rim basis, accuracy and precision of the estimates of stock composition produced from the 30 mtSNPs were replicated by each of six individual microsatellites evaluated. Accuracy and precision of the estimates could of course be improved by applying more than just a single microsatellite to estimate stock compositions. In this regard, as individual microsatellites evaluated provided estimates of stock composition of equal or better quality than that provided by the 30 mtSNPs, no further comparisons were considered to be necessary. Required resolution, accuracy, and precision of stock composition estimates will dictate the number of microsatellites that should be employed in the analysis.

Many of the 61 SNPs incorporated in the comparison concerning western Alaska chum salmon were specifically selected for distinguishing among western Alaskan populations (Smith and Seeb 2008). The 14 microsatellites in the current baseline did not undergo this rigorous selection procedure, as they 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 could fit on three injections in an automated DNA sequencer. Even so, the average population-specific accuracy and precision for the eight western Alaskan chum salmon populations derived from the 61 selected SNPs could be replicated by employing only 1-3 microsatellites from the microsatellite baseline. In contrast, Smith and Seeb (2008) reported that the 61 SNPs consistently produced the highest accuracy for all reporting groups when compared with a 15-microsatellite baseline. There were two significant differences in the 15-microsatellite baseline evaluated by Smith and Seeb (2008) and the 14-microsatellite baseline used in the current evaluation. Of the 15 microsatellites examined by Smith and Seeb

(2008), about half of the loci (7) displayed fewer than 10 alleles, and two loci displayed fewer than 20 alleles. Thus, there was a lower level of average polymorphism in the microsatellites evaluated by Smith and Seeb (2008). Secondly, sample sizes in the Smith and Seeb (2008) survey ranged from 78-96 fish per population, whereas in the current 14-microsatellite baseline, sample sizes ranged from 82-313 fish per population (Appendix Table 1). Population-specific estimates of stock composition derived from microsatellites depend on baseline sample size. For Chinook salmon, approximately 150 fish per population were required before 90% correct assignment was achieved in simulated single-population samples (Beacham et al. 2006). For chum salmon, population samples in excess of 500 fish were projected to be required for comparable results. If the analysis were restricted to those populations with at least 150 individuals sampled, then approximately 245 SNPs were predicted to be required to achieve comparable average accuracy as the 14 microsatellites, and 275 SNPs were predicted to be required to achieve comparable precision. 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 only the frequencies of two alleles (one independent) are estimated. 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.

If better stock ID resolution is required for western Alaska chum salmon in the existing SNP and microsatellite databases, is it more effective to add SNPs or microsatellites? The worst performing microsatellite (*Oke3*, mean assignment accuracy 55.8%) provided higher mean assignment accuracy than the best performing SNP (*Oke_RFC2_618*, 40.7%). 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. While there is little evidence that this has yet occurred for chum salmon, considerable effort would be required

for SNP discovery in order to enhance the baseline evaluated by Smith and Seeb (2008). In contrast, there are literally hundreds of microsatellites available for salmonids, and as many are easily transferrable among species, a broad range of microsatellites could be rapidly screened in judiciously chosen populations. Microsatellites specifically chosen for their stock separation of western Alaska populations could then be incorporated into existing SNP or microsatellite baselines.

The analysis conducted indicated that at least 200 SNPs of the average quality reported by Smith and Seeb (2008) would be required to equal the accuracy and precision of western Alaskan stock compositions available from the 14-microsatellite baseline. As many of the SNPs reported by Smith and Seeb (2008) were selected specifically for their population separation capability in western Alaska, it is unlikely that they would have the same relative population separation power in other areas of the species distribution, such as in Japan or British Columbia. New SNPs would likely 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 Japan (Beacham et al. 2008c), Russia (Beacham et al. 2008b), and British Columbia (Beacham et al. 2008d), as well as in western Alaska (Beacham et al. in press). As accuracy and precision of stock composition estimates provided by the 61 SNPs were comparable to those from suites of microsatellites with similar total allele counts, and at least 200 SNPs were estimated to be required to provide equivalent accuracy and precision for western Alaska alone, and the 14-microsatellite baseline incorporates approximately 800 total alleles, we suggest that at a minimum 300 and likely closer to 400 SNPs will likely be required on a Pacific Rim basis to provide accuracy and precision comparable to that currently obtained from the current 14-microsatellite baseline. This is clearly substantially greater than the 20-30 SNPs initially suggested by Seeb et al. (2004) that would be required for applications useful on a Pacific Rim basis. Stock composition estimation can clearly be

conducted with less than 300 SNPs in the baseline, but then accuracy and precision of the estimated stock compositions will be reduced relative to what can be achieved by application of a high-resolution microsatellite baseline.

Acknowledgments

A very substantial effort was undertaken to obtain samples from chum salmon sampled in this study. Starting from the south, we thank Dr. J. B. Shaklee of the Washington Department of Fish and Wildlife who provided samples from Washington State populations. In southern British Columbia, we thank various staff of the Department of Fisheries and Oceans (DFO) for baseline sample collection, as well as First Nations staff. P. Van Will of DFO coordinated fishery sampling in Johnstone Strait and Bute Inlet. In the northern British Columbia and the central coast, the Kitasoo Fisheries Program is acknowledged for some central coast populations, as well as fishery guardians and patrolmen. We would like to thank northern DFO staff who collected and supervised collections in Skeena River and central coast drainages. We also acknowledge the various agencies and organizations whose staff collected samples in northern British Columbia. These included the Gitksan Watershed Authority, the Kitselas and Kitsumkalum field staff and the Skeena Fisheries Commission in the Skeena River drainage and the Nisga'a First Nation in the Nass watershed. The Haida Fisheries Program and the crews of the Canadian Coast Guard Vessel Arrow Post assisted in sampling streams located in the Queen Charlotte Islands. In southeast Alaska, most samples were provided by Dr. R. L. Wilmot of the National Marine Fisheries Service (NMFS) Auke Bay Laboratory. Central Alaska, Alaska Peninsula, Bristol Bay, and Kuskokwim River samples were provided by Dr. L. W. Seeb of the Alaska Department of Fish and Game. S. Johnston and P. Milligan of the DFO Whitehorse office supervised collections of the Canadian portion of the Yukon River drainage. Dr. J. Wenburg of the United States Fish and Wildlife Service Conservation Genetics Laboratory in Anchorage provided samples from the

Alaskan portion of the Yukon River drainage, as well as from many Norton Sound and Kotzebue Sound populations. Dr. A. J. Gharrett of the University of Alaska Fairbanks provided some additional regional samples. For Russia, samples of populations from Primorye and Sakhalin Island, as well as some Magadan samples, were initially provided by Dr. V. V. Efremov to the NMFS Auke Bay Laboratory, where Dr. Richard Wilmot provided access to the Molecular Genetics Laboratory (MGL). Additional samples were provided by one of us (N. Varnakskaya) directly to the DFO Molecular Genetics Laboratory. Dr. G. Winans of the NMFS Montlake laboratory also provided access to some population samples from both Russia and Japan. Additional samples from Japan were collected by staffs of the National Salmon Resources Center, Fisheries Research Agency. The Korean population sample was provided by Dr. J. Park of the National Fisheries Research and Development Institute. L. Fitzpatrick drafted the map. Funding for the study was provided by Fisheries and Oceans, Canada.

References

- Banks, M. A., M. S. Blouin, B. A. Baldwin, V. K. Rashbrook, H. A. Fitzgerald, S. M. Blankenship, and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tshawytscha*). *Journal of Heredity* 90: 281-288.
- Beacham, T.D., J.R. Candy, K.J. Supernault, T. Ming, B. Deagle, A. Schultz, D. Tuck, K. Kaukinen, J.R. Irvine, K.M. Miller and R. E. Withler. 2001. Evaluation and application of microsatellite and major histocompatibility complex variation for stock identification of coho salmon in British Columbia. *Transactions of the American Fisheries Society* 130: 1116-1155.
- Beacham, T.D., J.R. Candy, B. McIntosh, C. MacConnachie, A. Tabata, K. Kaukinen, L. Deng, K. M. Miller, R. E. Withler, and N. V. Varnavskaya. 2005. Estimation of stock composition and individual identification of sockeye salmon on a Pacific Rim basis using

- microsatellite and major histocompatibility complex variation. Transactions of the American Fisheries Society 134: 1124-1146.
- Beacham, T. D., J. R. Candy, K. L. Jonsen, J. Supernault, M. Wetklo, L. Deng, K. M. Miller, and R. E. Withler. 2006. Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim using microsatellite variation. Transactions of the American Fisheries Society 135: 861-888.
- Beacham, T. D., B. Spilsted, K. D. Le, and M. Wetklo. 2008d. Population structure and stock identification of chum salmon *Oncorhynchus keta* from British Columbia determined with microsatellite DNA variation. Canadian Journal of Zoology 89: 1002-1014.
- Beacham, T. D., S. Urawa, K. D. Le, and M. Wetklo. 2008c. Population structure and stock identification of chum salmon from Japan determined with microsatellite DNA variation. Fisheries Science 74: 983-994.
- Beacham, T. D., N. V. Varnavskaya, K. D. Le, and M. Wetklo. 2008b. Determination of population structure and stock identification of chum salmon (*Oncorhynchus keta*) in Russia, determined with microsatellites. Fishery Bulletin 106: 245-256.
- Beacham, T. D., M. Wetklo, C. Wallace, J. B. Olsen, B. G. Flannery, J. K. Wenburg, W. D. Templin, A. Antonovich, and L. W. Seeb. 2008a. The application of microsatellites for stock identification of Yukon River Chinook salmon. North American Journal of Fisheries Management 28: 283-295.
- Beacham, T. D., K. D. Le, M. Wetklo, B. McIntosh, T. Ming, and K. M. Miller. In press. Population structure and stock identification of chum salmon from western Alaska determined with microsatellite and major histocompatibility complex variation. Pages XXX-XXX in C. C. Krueger, and C. E. Zimmerman, editors. Sustainability of the Arctic-Yukon-Kuskokwim salmon fisheries: what do we know about salmon ecology,

- management, and fisheries? American Fisheries Society, Symposium XX, Bethesda, Maryland.
- Beacham, T. D. 1996. The use of minisatellite DNA variation for stock identification of chum salmon, *Oncorhynchus keta*. *Fishery Bulletin* 94: 611-627.
- Beacham, T.D., A.P. Gould, R.E. Withler, C.B. Murray and L.W. Barner. 1987. Biochemical genetic survey and stock identification of chum salmon (*Oncorhynchus keta*) in British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* 44: 1702-1713.
- Beacham, T. D., A. P. Gould, R. E. Withler, C. B. Murray, and L. W. Barner. 1987. Biochemical genetic survey and stock identification of chum salmon (*Oncorhynchus keta*) in British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* 44: 1702-1713.
- Bernatchez, L. and P. Duchesne. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences* 57: 1-12.
- Buchholz, W. G., S. J. Miller, and W. J. Spearman. 2001. Isolation and characterization of chum salmon microsatellite loci and use across species. *Animal Genetics* 32: 160-165.
- Cairney, M., J. B. Taggart, and B. Hoyheim. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. *Molecular Ecology* 9: 2175-2178.
- Chen, J.-P., D.-J. Sun, C.-Z. Dong, B. Liang, W.-H. Wu, and S.-Y. Zhang. 2005. Genetic analysis of four wild chum salmon *Oncorhynchus keta* populations in China based on microsatellite markers. *Environmental Biology of Fishes* 73: 181-188.
- Debevec, E. M., R. B. Gates, M. Masuda, J. Pella, J. M. Reynolds, and L. W. Seeb. 2000. SPAM (Version 3.2): Statistics program for analyzing mixtures. *Journal of Heredity* 91: 509-510.

- Efremov, V. V. 2001. Genetic variation and differentiation of populations of chum salmon *Oncorhynchus keta* (Walbaum) from Southern Russian Far East. *Russian Journal of Genetics* 37: 283-289.
- Flannery, B. G., T. D. Beacham, R. R. Holder, Kretschmer, and J. K. Wenburg. 2007. Stock structure and mixed stock analysis of Yukon River chum salmon. U. S. Fish and Wildlife Service, Alaska Fisheries Technical Report Number 97. 53p.
- Ginatulina, L. K. 1992. Genetic differentiation among chum salmon, *Oncorhynchus keta* (Walbaum), from Primorye and Sakhalin. *Journal of Fish Biology* 40: 33-38.
- Goudet, J. 1995. FSTAT A program for IBM PC compatibles to calculate Weir and Cockerham's (1984) estimators of F-statistics (version 1.2). *Journal of Heredity* 86: 485-486.
- Ishida, Y., S. Ito, and K. Takagi. 1989. Stock identification of chum salmon *Oncorhynchus keta* from their maturity and scale characters. *Bulletin of the Japanese Society of Scientific Fisheries* 55: 651-656.
- Kalinowski, S. T. 2002. How many alleles per locus should be used to estimate genetic distances? *Heredity* 88: 62-65.
- Kalinowski, S. T. 2004. Genetic polymorphism and mixed-stock fisheries analysis. *Canadian Journal of Fisheries and Aquatic Sciences* 61: 1075-1082.
- Moriya, S., S. Sato, T. Azumaya, O. Suzuki, S. Urawa, A. Urano, and S. Abe. 2007. Genetic stock identification of chum salmon in the Bering Sea and North Pacific Ocean using mitochondrial DNA microarray. *Marine Biotechnology* 9: 179-191.
- Neaves, P. I., C. G. Wallace, J. R. Candy, and T. D. Beacham. 2005. CBayes: Computer program for mixed stock analysis of allelic data. Version v4.02. Free program distributed by the authors over the internet from http://www.pac.dfo-mpo.gc.ca/sci/mgl/Cbayes_e.htm

- Nikolayeva, Y., and N. I. Semenets. 1983. A contribution to stock differentiation of chum salmon, *Oncorhynchus keta* (Salmonidae), by scale structure in the first year of growth. *Journal of Ichthyology* 23: 18-28.
- Nelson, R. J., and T. D. Beacham. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. *Animal Genetics* 30: 228-229.
- Okazaki, T. 1982a. Genetic study on population structure in chum salmon (*Oncorhynchus keta*). *Far Seas Fisheries Research Laboratory Bulletin* 19: 25-116.
- Okazaki, T. 1982b. Geographical distribution of allelic variation of enzymes in chum salmon *Oncorhynchus keta*, river populations of Japan and the effects of transplantation. *Bulletin of the Japanese Society of Scientific Fisheries* 48: 1525-1535.
- Olsen, J. B., S. L. Wilson, E. J. Kretschmer, K. C. Jones, and J. E. Seeb. 2000. Characterization of 14 tetranucleotide microsatellite loci derived from sockeye salmon. *Molecular Ecology* 9: 2185-2187.
- Pella, J., and M. Masuda. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. *Fishery Bulletin* 99: 151-167.
- Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Science USA* 94: 9197-9201.
- Rexroad, C. E., R. L. Coleman, A. M. Martin, W. K. Hershberger, and J. Killefer. 2001. Thirty-five polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics* 32: 317-319.
- Salo, E. O. 1991. Life history of chum salmon (*Oncorhynchus keta*). In: C. Groot and L. Margolis (editors.), *Life History of Pacific Salmon*, p. 231-309. U.B.C. Press, Vancouver, B.C.

- Sato, S., J. Ando, H. Ando, S. Urawa, A. Urano, and S. Abe. 2001. Genetic variation among Japanese populations of chum salmon inferred from the nucleotide sequences of the mitochondrial DNA control region. *Zoological Science (Tokyo)* 18: 99-106.
- Sato, S., H. Kojima, J. Ando, H. Ando, R. L. Wilmot, L. W. Seeb, V. Efremov, L. LeClair, W. Buchholz, D.-H. Jin, S. Urawa, M. Kaeriyama, A. Urano, and S. Abe. 2004. Genetic population structure of chum salmon in the Pacific Rim inferred from mitochondrial DNA sequence variation. *Environmental Biology of Fishes* 69: 37-50.
- Seeb, L. W., and P. A. Crane. 1999a. High genetic heterogeneity in chum salmon in Western Alaska, the contact zone between northern and southern lineages. *Transactions of the American Fisheries Society* 128: 58-87.
- Seeb, L. W., and P. A. Crane. 1999b. Allozymes and mitochondrial DNA discriminate Asian and North American populations of chum salmon in mixed-stock fisheries along the south coast of the Alaska Peninsula. *Transactions of the American Fisheries Society* 128: 88-103.
- Seeb, L. W., C. T. Smith, W. D. Templin, R. L. Wilmot, and J. E. Seeb. 2004. Development of a Pacific Rim baseline for chum salmon based on single nucleotide polymorphism markers (SNPs). *North Pacific Anadromous Fisheries Commission Document* 824. 10p.
- Small, M. P., T. D. Beacham, R. E. Withler, and R. J. Nelson. 1998. Discriminating coho salmon (*Oncorhynchus kisutch*) populations within the Fraser River, British Columbia using microsatellite DNA markers. *Molecular Ecology* 7: 141-155.
- Smith, C. T., B. F. Koop, and R. J. Nelson. 1998. Isolation and characterization of coho salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids. *Molecular Ecology* 7: 1613-1621.
- Smith, C. T., and L. W. Seeb. 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. *Transactions of the American Fisheries Society* 137: 751-762.

- Smith, C. T., W. D. Templin, J. E. Seeb, L. W. Seeb. 2005. Single nucleotide polymorphisms (SNPs) provide rapid and accurate estimates of the proportions of U. S. and Canadian Chinook salmon caught in Yukon River fisheries. *North American Journal of Fisheries Management* 25: 944-953.
- Sohn, D., S. Kang, and S. Kim. 2005. Stock identification of chum salmon (*Oncorhynchus keta*) using trace elements in otoliths. *Journal of Oceanography* 61: 305-312.
- Spies, I. B., D. J. Brasier, P. T. L. O'Reilly, T. R. Seamons, and P. Bentzen. 2005. Development and characterization of novel tetra-, tri-, and dinucleotide microsatellite markers in rainbow trout (*Oncorhynchus mykiss*). *Molecular Ecology Notes* 5: 278-281.
- Tanaka, S., M. P. Shepard, and H. T. Bilton. 1969. Origin of chum salmon (*Oncorhynchus keta*) in offshore waters of the North Pacific in 1956-1958 as determined from scale studies. *International North Pacific Fisheries Commission Bulletin* 26: 57-155.
- Taylor, E. B., T. D. Beacham, and M. Kaeriyama. 1994. Population structure and identification of North Pacific Ocean chum salmon (*Oncorhynchus keta*) revealed by an analysis of minisatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences* 51: 1430-1442.
- Weir, B.S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Williamson, K. S., J. F. Cordes, and B. P. May. 2002. Characterization of microsatellite loci in chinook salmon (*Oncorhynchus tshawytscha*) and cross-species amplification in other salmonids. *Molecular Ecology Notes* 2: 17-19.
- Wilmot, R. L., R. J. Everett, W. J. Spearman, R. Baccus, N. V. Varnavskaya, and S. V. Putivkin. 1994. Genetic stock structure of Western Alaska chum salmon and a comparison with Russian Far East stocks. *Canadian Journal of Fisheries and Aquatic Sciences* 51 (Suppl. 1): 84-94.

- Wilmot, R. L., C. M. Kondzela, C. M. Guthrie, and M. M. Masuda. 1998. Genetic stock identification of chum salmon harvested incidentally in the 1994 and 1995 Bering Sea trawl fishery. North Pacific Anadromous Fish Commission Bulletin 1: 285-299.
- Winans, G. A., P. B. Aebersold, S. Urawa, and N. V. Varnavskaya. 1994. Determining continent of origin of chum salmon (*Oncorhynchus keta*) using genetic identification techniques: status of allozyme baseline in Asia. Canadian Journal of Fisheries and Aquatic Sciences 51(Supplement 1), 95-113.
- Winans, G. A., P. B. Aebersold, Y. Ishida, and S. Urawa. 1998. Genetic stock identification of chum salmon in highseas test fisheries in the western North Pacific Ocean and Bering Sea. North Pacific Anadromous Fish Commission Bulletin 1: 220-226.

Table 1. Number of alleles per locus, F_{ST} , and mean accuracy (%) and standard deviation (%) for estimated percentage compositions of single population mixtures (correct = 100%) for 40 representative populations of chum salmon from the Pacific Rim distribution of populations.

The number of alleles and F_{ST} were calculated from the entire baseline.

Locus	Number of alleles	F_{st}	Population mean accuracy	Population mean SD	Regional mean accuracy	Regional mean SD
1) Oke3	26	0.104 (0.005)	37.0	20.7	49.5	19.2
2) Oki100	31	0.039 (0.002)	49.5	15.0	59.1	13.3
3) Ots3	31	0.097 (0.005)	48.5	18.0	59.6	15.4
4) Oki2	42	0.062 (0.005)	41.9	16.6	50.7	15.0
5) Omy1011	44	0.027 (0.001)	52.6	14.2	59.3	12.6
6) One104	48	0.027 (0.001)	53.1	13.0	60.1	11.8
7) Ots103	54	0.019 (0.001)	59.6	10.9	66.4	9.7
8) Ssa419	54	0.028 (0.001)	41.4	16.7	47.8	16.1
9) One101	56	0.058 (0.002)	59.2	13.5	66.5	11.5
10) Omm1070	60	0.009 (0.000)	48.8	11.9	53.0	11.2
11) One114	60	0.017 (0.001)	49.0	14.0	54.8	13.1
12) One102	69	0.011 (0.001)	40.8	14.1	46.3	13.8
13) OtsG68	69	0.017 (0.001)	57.2	11.1	63.4	9.7
14) One111	149	0.036 (0.003)	72.9	7.9	78.1	6.7
Total		0.033 (0.007)	89.1	3.1	92.0	2.5

Table 2. Estimated percentage stock compositions of simulated mixtures of chum salmon as may be encountered in Pacific Rim fisheries. Standard deviation is in parentheses. Expected regional composition is obtained by adding true population components, and expected country composition is obtained by adding true regional components.

Composition	True	Estimate	Composition	True	Estimates
Population	Mixture 1		Population	Mixture 4	
Namdae	10	9.8 (2.2)	DIPAC Hatchery	15	15.6 (2.8)
Gakko	10	9.8 (2.3)	Nakut	10	9.3 (2.4)
Chitose	20	20.1 (3.0)	Clapp Basin	10	8.4 (2.1)
Tokushibetsu	20	21.7 (3.1)	Naden	10	11.0 (2.3)
Tokachi	10	9.7 (2.2)	Ensheshese	20	19.6 (3.4)
Amur	10	10.3 (2.2)	Andesite	10	9.4 (2.2)
Naiba	10	10.0 (2.2)	Markle Inlet	10	10.0 (2.1)
Hairusova	10	9.7 (2.1)	Cheenis Lake	5	4.8 (1.7)
Region			Draney	10	11.2 (3.1)
South Korea	10	10.5 (2.2)	Region		
Honshu-Japan Sea	10	8.0 (2.1)	Southeast Alaska	25	24.9 (3.4)
Hokkaido-Japan Sea	20	20.1 (3.2)	QCI- west coast	10	8.4 (2.1)
Hokkaido-Sea of Okhotsk	20	23.0 (3.2)	QCI- north coast	10	11.0 (2.3)
Hokkaido-eastern Pacific	10	9.7 (2.2)	North Coast	20	19.8 (5.0)
Amur River	10	10.3 (2.2)	Skeena River	10	9.4 (2.2)
Sakhalin Island	10	10.0 (2.2)	Grenville	10	11.2 (2.4)
			Channel		
West Kamchatka	10	9.7 (2.1)	Central Coast	5	4.9 (1.7)
Country			Rivers Inlet	10	16.1 (3.1)

South Korea	10	10.5 (2.2)	Country		
Japan	60	59.5 (3.5)	United States	25	24.9 (3.5)
Russia	30	30.0 (3.3)	Canada	75	75.1 (3.5)
Population		Mixture 2		Mixture 5	
Kawabukuro	10	10.8 (2.3)	Nimpkish	20	19.8 (3.0)
Teshio	10	9.3 (2.1)	Homathko	10	9.2 (2.4)
Abashiri	10	9.7 (2.1)	Indian	20	24.8 (3.9)
Ola	20	20.5 (2.9)	Puntledge	10	7.2 (2.4)
Anadyr	20	19.5 (3.0)	Harrison	10	9.0 (2.4)
Agiapuk	10	10.5 (2.2)	Nooksack	10	9.6 (2.6)
Pilgrim	10	9.8 (2.6)	Tulalip	10	9.0 (2.3)
Gisasa	10	9.7 (2.5)	Hoodsport	10	11.2 (2.4)
Region			Region		
Honshu- Sea of Japan	10	10.8 (2.3)	Johnstone Strait	20	19.8 (3.0)
Hokkaido- Sea of Japan	10	9.7 (2.1)	South Coast	30	34.1 (4.0)
Hokkaido- Sea of Okhotsk	10		Vancouver	10	7.2 (2.4)
		9.3 (2.1)	Island- east coast		
Magadan	20	20.5 (2.9)	Fraser River	10	9.0 (2.5)
Northeast Russia	20		North Puget	20	18.6 (3.2)
		19.5 (3.0)	Sound		
Kotzebue Sound	10	10.5 (2.2)	Hood Canal	10	11.2 (2.4)
Norton Sound	10	9.8 (2.6)	Country		
Lower Yukon- summer run	10	9.7 (2.5)	Canada	70	70.2 (3.5)
Country			United States	30	29.8 (3.5)
Japan	30	29.8 (3.3)		Mixture 6	

Russia	40	40.0 (3.6)	Teshio	10	9.9 (2.1)
United States	30	30.1 (3.3)	Naiba	10	10.0 (2.2)
Population		Mixture 3	Ossora	10	10.9 (2.3)
Imnachuk	10	10.1 (2.1)	Vorovskaya	10	8.2 (2.2)
Kobuk	10	11.8 (2.8)	Tozitna	10	9.9 (2.7)
Andreafsky	5	0.8 (1.9)	Toklat	10	9.9 (2.3)
Snake	15	13.6 (3.2)	Tatchun	10	10.2 (2.3)
Kanektok	10	8.6 (2.5)	Kanektok	10	8.6 (2.5)
Kluane River	10	10.0(2.1)	Mulchatna	10	11.9 (2.7)
Fishing Branch	10	11.1 (2.8)	Nitinat	10	10.3 (2.2)
Bolshaya	10	10.3 (2.1)	Region		
Sheenjek	10	10.2 (2.6)	Hokkaido- Sea of	10	9.9 (2.1)
			Japan		
Frosty	10	10.0 (2.1)	Sakhalin Island	10	10.0 (2.2)
Region			West Kamchatka	10	8.3 (2.2)
West Kamchatka	10	10.3 (2.1)	East Kamchatka	10	10.9 (2.3)
Kotzebue Sound	20		Lower Yukon-	10	9.9 (2.7)
		21.8 (3.3)	summer run		
Norton Sound	15	16.8 (3.6)	Tanana- fall run	10	9.9 (2.3)
Lower Yukon- summer run	5		Mainstem Yukon	10	10.2 (2.3)
		1.0 (2.0)	River		
Upper Alaska	10		Kuskowim	10	8.7 (2.5)
		10.2 (2.6)	River/Bay		
Kuskokwim River/Bay	10	8.7 (2.5)	Nushagak River	10	11.9 (2.7)
White River	10	9.9 (2.1)	Vancouver	10	10.3 (2.2)

Island- west coast					
Porcupine River	10	11.1 (2.8)	Country		
North Peninsula/ Aleutians	10	10.0 (2.1)	Japan	10	9.9 (2.1)
Country			Russia	30	29.2 (3.3)
Russia	10	10.3 (2.1)	United States	40	40.4 (3.6)
United States	70	68.6 (3.8)	Canada	20	20.5 (2.3)
Canada	20	21.1 (3.3)			

Table 3. True and estimated stock compositions (percentage, SD in parentheses) of four known-origin 200-fish samples of chum salmon. The mixture samples were constructed by randomly removing the appropriate number of individuals from existing baseline samples to generate a mixture sample of known origin, and then the baseline population allele frequencies were recalculated to reflect removal of individuals included in the mixture sample.

Composition	True	Estimate	Composition	True	Estimate
Population	Mixture 1		Population	Mixture 2	
Namdae	5	3.1 (1.3)	Ossora	5	0.2 (0.9)
Chitose	10	15.5 (2.5)	Anadyr	5	3.2 (1.9)
Tokushibetsu	10	8.8 (2.3)	Imnachuk	10	7.0 (1.8)
Shibetsu	5	0.0 (0.0)	Kobuk	5	1.7 (2.8)
Teshio	8	3.2 (1.9)	Pilgrim	10	14.5 (4.7)
Sakari	5	3.8 (1.4)	Snake	5	3.1 (3.7)
Amur	15	13.7 (2.6)	Tozitna	5	0.7 (2.2)
Naiba	5	4.0 (1.5)	Gisasa	10	0.4 (1.5)
Ryazanovka	2	0.4 (0.9)	Salcha	5	4.5 (3.7)
Ola	5	0.0 (0.2)	Fishing Branch	8	9.2 (3.1)
Vorovskaya	15	13.1 (3.4)	Kluane River	10	9.7 (2.1)
Pymta	5	1.1 (1.7)	Kanektok	5	0.1 (0.5)
Ossora	10	9.2 (2.7)	Gertrude	5	0.5 (1.2)
Region			Frosty	5	4.9 (1.5)
South Korea	5	3.1 (1.3)	Keta	2	1.0 (1.0)
Hokkaido-Sea of Japan	18	18.7 (3.3)	DIPAC hatchery	5	6.0 (1.9)
Hokkaido-Sea of Okhotsk	10	8.8 (2.3)	Region		
Hokkaido-Nemuro Strait	5	3.8 (2.1)	East Kamchatka	5	2.2 (1.3)

Honshu-Pacific Coast	5	3.8 (1.5)	Northeast Russia	5	4.2 (1.9)
Primorye	2	1.9 (1.1)	Kotzebue Sound	15	8.9 (2.8)
Amur River	15	13.7 (2.6)	Norton Sound	15	18.6 (4.7)
Sakhalin Island	5	4.2 (1.5)	Lower Yukon-summer run	15	13.4 (4.7)
Magadan	5	0.4 (1.1)	Tanana River-summer run	5	8.7 (3.7)
West Kamchatka	20	21.9 (3.8)	Porcupine River	8	9.6 (3.1)
East Kamchatka	10	13.5 (3.3)	White River	10	9.7 (2.1)
Country			Kuskokwim River/Bay	5	3.2 (2.9)
South Korea	5	3.1 (1.3)	South Bristol Bay	5	0.6 (1.2)
Japan	37	35.1 (3.6)	North Peninsula/Aleutians	5	4.9 (1.6)
Russia	58	55.6 (3.7)	Prince William Sound	2	1.0 (1.0)
			Southeast Alaska	5	6.1 (1.9)
			Country		
			Russia	10	10.7 (1.9)
			United States	72	67.8 (4.2)
			Canada	18	21.5 (3.1)
			Population		
				Mixture 4	
			Viner Sound	10	7.6 (2.0)
Population		Mixture 3	Southgate	10	7.7 (2.9)
Fish	10	5.6 (3.2)	Indian	10	8.8 (2.8)
Disappearance	10	7.0 (3.5)	Little Qualicum	10	7.3 (3.0)
Gold Harbour	10	9.5 (2.3)	Puntledge	10	0.0 (0.0)
Deena River	10	8.0 (2.2)	Nitinat	10	12.1 (2.5)
Kshwan	10	9.8 (3.2)	Inch	10	5.9 (4.0)
Andesite	10	16.1 (3.3)	Squakum	10	0.0 (0.4)

Whitebottom	10	0.1 (0.7)	Nooksack	5	1.3 (1.4)
Markle	10	9.6 (2.5)	Kennedy	5	4.9 (1.6)
Kitasoo	10	11.0 (2.6)	Bitter	5	2.9 (1.5)
Draney	10	7.6 (2.7)	Quinault	5	4.1 (1.6)
Region			Region		
Southeast Alaska	20	14.1 (4.1)	Johnstone Strait	10	7.6 (2.0)
QCI-west coast	10	9.6 (2.3)	South Coast	20	20.8 (3.8)
QCI-Skidegate Channel	10	8.1 (2.2)	Vancouver Island-east coast	20	18.6 (3.5)
North coast	10	16.3 (4.3)	Vancouver Island-west coast	10	12.2 (2.5)
Skeena River	20	17.3 (3.1)	Fraser River	20	25.1 (3.8)
Grenville Channel	10	10.1 (2.2)	North Puget Sound	5	1.4 (1.5)
Central Coast	10	12.9 (2.5)	South Puget Sound	5	5.0 (1.6)
Rivers Inlet	10	9.6 (2.5)	Coastal Washington	10	8.6 (2.1)
Country			Country		
United States	20	14.1 (4.1)	Canada	80	85.0 (3.0)
Canada	80	85.9 (4.1)	United States	20	15.0 (3.0)

Table 4. Estimated stock compositions (percentage, SD in parentheses) of eight mixed-stock samples of chum salmon . Samples were obtained from the following fisheries: 1) northern coastal Hokkaido (44°56'N, 142°35'E, 15 September 2005, N=192), 2) Southwestern Sea of Okhotsk (51°00'N, 149°00'E, 30 September 2003, N=50), 3 Central Sea of Okhotsk, (55°00'N, 155°00'E, 17 October 2003, N=48), 4) North Pacific Ocean (51°12'N, 158°58'E, 13-14 August 2005, N=49), 5) Northern Gulf of Alaska (53°01'N, 145°00'W, 16 February 2006, N=32), 6) Central Gulf of Alaska (47°54'N, 144°49'W, 18 February 2006, N=111), 7) Johnstone Strait (50°53'N, 127°35'W, 01 October 2007, N=149), and 8) Bute Inlet (50°31'N, 125°00'W, 11 October 2004, N=100). N is sample size.

Region	Mixed stock samples							
	1	2	3	4	5	6	7	8
Hokkaido Sea of Japan	61.4 (4.4)	0.1 (0.6)		17.1 (6.0)	0.1 (0.6)			
Hokkaido Sea of Okhotsk	31.8 (4.0)	50.9 (10.0)		33.7 (7.9)	1.2 (2.4)	26.3 (5.0)		
Hokkaido Nemuro Strait		2.1 (4.7)		5.3 (4.0)		0.1 (0.4)		
Hokkaido Eastern Pacific		10.0 (7.7)		9.5 (5.1)		5.4 (3.1)		
Hokkaido Western Pacific	0.1 (0.5)	3.6 (4.8)		0.7 (2.4)		5.2 (3.2)		
Honshu Pacific		7.5 (4.3)				0.1 (0.5)		
Honshu Sea of Japan	6.5 (2.5)	0.5 (2.0)		0.3 (1.3)		0.1(0.5)		
Primorye		0.1 (0.5)		6.1 (3.8)				
Sakhalin Island		21.1 (6.3)		1.0 (2.7)		5.8 (2.5)		

Magadan	0.2 (0.9)	0.1(0.7)	2.3 (3.4)		3.1 (2.1)
Northern Sea of Okhotsk			11.0 (6.4)		
Kamchatka west coast		80.6 (11.6)	12.0 (7.8)	4.8 (6.2)	11.2 (3.8)
Kamchatka east coast		14.6 (12.6)	0.1 (0.5)	1.4 (3.6)	
Northeast Russia				1.4 (3.1)	8.4 (3.1)
Norton Sound		0.1 (0.9)		0.8 (2.5)	4.3 (2.8)
Lower Yukon River summer				2.0 (3.7)	0.3 (1.1)
Upper Alaska fall					0.4 (0.9)
White River					0.1 (0.3)
Kuskokwim River/Bay				1.1 (3.0)	0.3 (1.0)
Northeast Bristol Bay				0.1 (0.7)	0.1 (0.5)
Nushagak River				0.7 (2.6)	
Southwest Peninsula			0.1 (0.5)	1.3 (3.3)	
Southeast Peninsula			0.7 (1.5)	0.1 (0.8)	
Prince William Sound				5.5 (6.6)	9.3 (3.0)
Southeast Alaska		0.1 (0.7)		13.3 (13.0)	14.2 (4.0)
Taku River		0.1 (0.7)		0.1 (0.8)	0.2 (0.9)

QCI west coast				0.9 (2.7)			
QCI east coast		0.1 (0.6)		0.1 (1.0)	0.1 (0.5)	1.3 (1.5)	
QCI north coast				0.1 (0.7)			
QCI Skidegate				0.1 (0.9)	0.1 (0.4)	0.3 (0.7)	
North Coast	0.4 (1.2)		0.1 (0.7)	1.3 (4.2)	0.3 (0.9)		
Skeena River	2.9 (3.5)	0.4 (1.2)		3.2 (5.7)	0.6 (1.2)		
Grenville Channel				4.4 (8.5)	0.2 (0.8)		
Central Coast	0.2 (1.0)	0.2 (1.2)	0.1 (0.9)	18.3 (9.6)	0.5 (1.3)	0.1 (0.4)	
Rivers Inlet				0.2 (1.1)			
Johnstone Strait	0.4 (1.3)	2.9 (5.3)		8.1 (7.7)	0.2 (0.8)	0.1 (0.6)	
South Coast				7.0 (8.7)	1.0 (1.5)	14.2 (5.5)	99.8 (1.2)
Vancouver Island east coast				1.3 (3.3)		28.4 (6.7)	
Vancouver Island west coast		0.1 (0.6)		1.3 (3.5)		0.5 (1.4)	
Fraser River		0.1 (0.9)		17.1 (10.0)	1.9 (1.7)	54.2 (6.0)	
North Puget Sound				0.2 (1.3)	0.1 (0.6)		
South Puget Sound		0.1 (0.6)		1.8 (3.6)		0.8 (1.1)	
Juan de Fuca Strait				0.6 (2.2)			

Country								
Japan	99.9 (0.6)	74.6 (6.4)	0.0 (0.7)	66.5 (6.9)	1.3 (2.6)	37.2 (4.8)	0.0 (0.2)	0.0 (0.2)
Russia	0.0 (0.2)	21.4 (6.3)	95.4 (6.0)	32.4 (6.9)	7.6 (6.9)	28.5 (5.2)	0.0 (0.2)	0.0 (0.3)
United States	0.1 (0.3)	0.0 (0.6)	0.5 (1.3)	0.8 (1.7)	27.6 (11.9)	29.1 (4.7)	0.8 (1.2)	0.1 (0.8)
Canada	0.0 (0.2)	3.9 (3.7)	4.1 (5.4)	0.2 (1.0)	63.6 (14.0)	5.3 (2.4)	99.2 (1.2)	99.9 (1.0)

Appendix Table 1. Spawning location, sample collection years, number of fish sampled per year, and total number of fish sampled for 381 populations of chum salmon in 59 reporting regions from Korea, Japan, Russia, Alaska, Canada, and Washington. Allele frequencies for all location samples surveyed in this study are available at http://www-sci.pac.dfo-mpo.gc.ca/mgl/default_e.htm

Population	Years	N
1) South Korea		
Namdae River (1)	2004	100
2) Honshu Island, Sea of Japan Coast		
Miomote River (2)	1989, 1991	95
Kawabukuro River (3)	1990, 1997	120
Hayatsuki River (4)	1990	76
Gakko River (5)	2003	160
Uono River (6)	1996	80
3) Hokkaido Island, Sea of Japan Coast		
Teshio River (7)	1991, 2001	100
Chitose River (8)	1990, 1991, 2003	280
Toshibetsu River (9)	1990	60
4) Hokkaido Island, Sea of Okhotsk Coast		
Tokoro River (10)	1988, 2005	119
Abashiri River (11)	1992, 1998	130
Horonai River (12)	1992	50
Shari River (13)	2001	80
Tokushibetsu River (14)	1997, 2004	160
5) Hokkaido Island, Nemuro Strait		

Shibetsu River (15)	1988, 1991, 2003	110
Nishibetsu River (16)	1997	80
6) Hokkaido Island, Eastern Pacific Coast		
Tokachi River (17)	1991, 2002	130
Kushiro River (18)	1998	80
7) Hokkaido Island, Western Pacific Coast		
Shiriuchi River (19)	1990, 1998	160
Shikiu River (20)	1998	80
Yurappu River (21)	1997	160
Shizunai River (22)	2002	80
8) Honshu Island, Pacific Coast		
Ohkawa River (23)	1989	19
Tsugaruishi River (24)	1999	80
Orikasa River (25)	1996	80
Sakari River (26)	1997	80
Koizumi River (27)	1996	80
9) Primorye		
Narva (28)	1994	17
Ryazanovka (29)	1994	49
Avakumovka (30)	1994	35
10) Amur River		
Amur River (31)	1994, 2001, 2004	338
11) Sakhalin Island		
Tym (32)	1995	55
Naiba (33)	1994, 1995	149

Udarnitsa (34)	1994	50
Kalininka (35)	1994	49
12) Magadan		
Tugur (36)	2004	98
Okhota (37)	2004	94
Magadan (38)	1991	79
Tauy (39)	1990	55
Ola (40)	1990, 1992	120
13) Northern Sea of Okhotsk		
Oklan (41)	1993	76
Penzhina (42)	1993	43
14) West Kamchatka		
Hairusova (43)	1990, 1993	186
Vorovskaya (44)	1991, 1993	249
Kol (45)	1991	79
Pymta (46)	1992, 1993	99
Kikchik (47)	1992, 2005	106
Utka (48)	1992	40
Bolshaya (49)	2004	96
Plotnikova (50)	2001	69
15) East Kamchatka		
Zhypanova (51)	2004	46
Kamchatka (52)	1990	76
Ivashka (53)	2005	48
Nerpichi (54)	1992	39

Karaga (55)	2005	42
Ossora (56)	1990, 1996, 2005	128
Dranka (57)	2005	44
Apuka (58)	2002	47
Olutorsky Bay (59)	2002	49
16) Northeast Russia		
Anadyr (60)	1991, 1992	94
Kanchalan (61)	1991	79
17) Mackenzie River		
Peel River (62)	1998	33
18) Kotzebue Sound		
Kelly Lake (63)	1991	92
Noatak River (64)	1991	45
Inmachuk River (65)	2005	194
Kobuk River (66)	1991, 2000	374
Agiapuk River (67)	2005	180
Koyuk River (68)	2005	44
19) Norton Sound		
Niukluk River (69)	2004, 2005	223
Pilgrim River (70)	1994, 2004, 2005	474
Kwiniuk River (71)	2004, 2005	262
Snake River (72)	2004, 2005	394
Nome River (73)	2004, 2005	204
Eldorado River (74)	2004, 2005	390
Unalakleet River (75)	2005	191

Shaktoolik River (76)	2005	196
Ungalik River (77)	2005	50
Pikmiktalik River (78)	2004, 2005	398
20) Lower Yukon River summer run (United States)		
Andreafsky River (79)	1987, 1993, 2004	313
Chulinak River (80)	1989	93
Anvik River (81)	1988, 1993	182
Nulato River (82)	1988, 2003	123
Gisasa River (83)	1988, 2003	279
Jim River (84)	2002	147
Henshaw Creek (85)	2003	193
Koyukuk (south fork) (86)	1996	92
Koyukuk (later run) (87)	1996	100
Tozitna River (88)	2002, 2003	347
Melozitna River (89)	2003, 2004	161
21) Tanana River summer run (United States)		
Cheena River (90)	1992, 1994, 1997	236
Salcha River (91)	1994, 2001	185
22) Tanana River fall run (United States)		
Toklat River (92)	1990, 1994	241
Delta River (93)	1990	80
Kantishna River (94)	2001	160
23) Upper Alaska (United States)		
Sheenjek River (95)	1987, 1988, 1989	229
Chandalar River (96)	1989, 2001	185

Black River (97)	1995, 2001	110
Big Salt (98)	2001	73
24) Porcupine River (Canada)		
Fishing Branch (99)	1987, 1989, 1992, 1994, 1997, 2007	597
Porcupine River (100)	2005, 2007	329
25) White River (Canada)		
Kluane River (101)	1987, 1992, 2001, 2007	486
Kluane Lake (102)	2002, 2003	62
Donjek River (103)	1994	72
26) Mainstem Yukon River (Canada)		
Mainstem at Big Creek (104)	1992, 1995	175
Mainstem at Minto Landing (105)	1989, 2002	145
Mainstem at Tatchun Creek (106)	1987, 1992	173
Mainstem at Pelly River (107)	1993	83
27) Teslin River (Canada)		
Teslin River (108)	1992, 2002	143
28) Upper Yukon early fall (Canada)		
Chandindu River (109)		
29) Kuskokwim River/Bay		
George River (110)	1996	82
Kasigluk River (111)	1990	68
Kwethluk River (112)	1989	76
Nunsatuk River (113)	1994	83
Aniak River (114)	1992	86
Kanektok River (115)	1989, 1994	171

30) Nushagak River		
Stuyahok River (116)	1992, 1993	74
Mulchatna River (117)	1994	82
31) North/Central Bristol Bay		
Goodnews River (118)	1991	92
Togiak River (119)	1993	75
Alagnak River (120)	1992	77
Naknek River (121)	1993	64
32) South Bristol Bay		
Egegik River (122)	1993	86
Meshik River (123)	1989	57
Gertrude Creek (124)	1987	97
Pumice Creek (125)	1993	92
33) North Peninsula/Aleutians		
Moller Bay Creek (126)	1998	93
Frosty Creek (127)	1992, 2000	179
Joshua Green River (128)	1994	95
34) Southwest Peninsula		
Coleman Creek (129)	1996	70
Volcano Bay Creek (130)	1992	104
Delta Creek (131)	1996	78
Westward Creek (132)	1993	79
35) Southeast Peninsula		
Stepovak Bay (133)	1993	94
Big River (134)	1993	87

Alogoshak River (135)	1993	91
36) Kodiak Island		
American River (136)	1992	95
Sturgeon River (137)	1992	71
Uganik River (138)	1992	100
37) Prince William Sound		
Keta Creek (139)	1992	100
Well's River (140)	1996	100
Constantine Creek (141)	1995	99
Olsen Creek (142)	1992	92
38) Southeast Alaska		
Fish Creek (143)	2006	200
DIPAC Hatchery (144)	1997, 1998	200
Gambier Bay (145)	1989, 1990	67
Well's Bridge (146)	2000	50
Green's Creek (147)	1995	50
Herman Creek (148)	1990	59
Sawmill Creek(149)	1993	36
Kennell Creek (150)	1995	50
Disappearance (151)	1986, 1988, 1998, 2007	333
Neet's Bay early (152)	1997, 2006	237
Neet's Bay late (153)	1997	150
Carroll River (154)	1986	50
Nakut (155)	2006	103
Lagoon Creek (156)	2007	79

39) Taku River

Shustnini River (157)	2004	20
Taku River (158)	2000	65
Takwahoni River (159)	2000	12
Tuskwa River (160)	2004	62
Yellow Bluff (161)	2004	13

40) Queen Charlotte Islands – west coast

Botany Creek (162)	2002	175
Clapp Basin Creek (163)	2002, 2005	393
Dawson Inlet Creek (164)	2004	203
Fairfax Inlet Creek (165)	2004, 2005	42
Gold Harbour Creek (166)	2002	203
Kano Inlet Creek (167)	2005, 2006	129
Mace Creek (168)	2002, 2003	220
Mountain Creek (169)	2004	225
Seal Inlet Creek (170)	2005, 2006	257
Security Creek (171)	2003, 2004, 2006	215
Steel Creek (172)	2005, 2006	240

41) Queen Charlotte Islands – north coast

Ain Creek (173)	2003	80
Awun Creek (174)	2003	94
Naden River (175)	2003, 2004	131
Stanley Creek (176)	2004, 2005, 2006	221

42) Queen Charlotte Islands – east coast

Bag Harbour Creek (177)	1989, 2002, 2003	138
-------------------------	------------------	-----

Dana Creek (178)	2004	202
Hutton Head Creek (179)	2005	35
Lagoon Inlet Creek (180)	1986, 2002, 2004	376
Little Goose Creek (181)	2003, 2005	220
Pacofi Creek (182)	2004, 2005	233
Pallant Creek (183)	1991, 2002	216
Salmon River (184)	2002, 2003	201
Sedgewick Creek (185)	2002	78
Surprise Creek (186)	1989	50
Thorsen Creek (187)	2001	17
43) Queen Charlotte Island – Skidegate Channel		
Buck Channel Creek (188)	2004, 2005, 2006	177
Deena River (189)	2004, 2006	214
Government Creek (190)	2002	187
Honna River (191)	2005, 2006	79
Lagins Creek (192)	2004, 2006	232
North Arm Creek (193)	2003	193
Slatechuck Creek (194)	2002	168
Tarundl Creek (195)	2004	196
44) North Coast		
Crag Creek (196)	2005, 2006	34
Dak River (197)	2003	43
Ensheshese River (198)	2004, 2005	242
Illiance River (199)	2004, 2005, 2006	158
Kateen River (200)	2003, 2004	210

Khutzeymateen River (201)	2002, 2004	27
Kincolith River (202)	2005	28
Kitsault River (203)	2004	124
Ksedin Creek (204)	2005	40
Kshwan River (205)	2002, 2003	256
Lachmach Creek (206)	2002, 2003, 2007	136
Lizard Creek (207)	2005	53
Nass River (208)	2005, 2007	137
Stagoo River (209)	2002	183
Stumaun Creek (210)	2005, 2006	38
Toon River (211)	2002	176
Tseax River (212)	2004, 2005	65
Wilauks Creek (213)	2004, 2005, 2006	162
45) Skeena River		
Andesite Creek (214)	2004, 2005, 2006, 2007	182
Date Creek (215)	2002, 2003, 2004, 2005, 2006, 2007	93
Dog-tag Creek (216)	2004, 2005, 2007	102
Ecstall River (217)	2004	104
Kitsumkalum River (218)	2004, 2005, 2006	85
Upper Kitsumkalum River (219)	2002	12
Kispiox River (220)	2004, 2005, 2006	44
Kitwanga River (221)	2002, 2003	154
Kleanza Creek (222)	2004, 2005	16
Nangeese river (223)	2002, 2003, 2004, 2005, 2006,	162

	2007	
Skeena River mainstem (224)	2004, 2005, 2006	70
Whitebottom Creek (225)	2004	139
Zymagotitz River (226)	2006	71
46) Grenville Channel		
Kumealon Creek (227)	2005, 2006, 2007	197
Kxngeal Creek (228)	2006, 2007	40
Markle Inlet Creek (229)	2002, 2003	156
Pa-aat River (230)	2004, 2005	46
Stewart Creek (231)	2005, 2006, 2007	51
Wilson Creek (232)	2002, 2003	242
47) Central Coast		
Arnoup Creek (233)	2005	181
Barnard Creek (234)	2003, 2004, 2005	156
Bish Creek (235)	2004, 2005, 2006, 2007	78
Blackrock Creek (236)	2002, 2003, 2004, 2005, 2006	189
East Arm Creek (237)	2005, 2006, 2007	82
West Arm Creek (238)	2002	201
Flux Creek (239)	2005, 2006	214
Foch Creek (240)	2002	173
Gill Creek (241)	2002, 2003, 2004, 2005	182
Gilttoyes Creek (242)	2003	177
Green River (243)	2002	170
Kemano River (244)	2002	199
Khutze River (245)	2002	187

Kiltuish River (246)	2004, 2005	197
Kitasoo Creek (247)	2002, 2005, 2007	419
Kitimat River (248)	2002	193
Nias Creek (249)	2002	180
Quaal River (250)	2002, 2003	271
Turn Creek (251)	2002, 2005, 2006	110
Tyler Creek (252)	2003	193
Turtle Creek (253)	2005, 2006, 2007	96
Bella Bella Creek (254)	2001	139
Bullock Channel Creek (255)	2004, 2006	352
Cheenis Lake (256)	2004	172
Clatse Creek (257)	2002	170
Cooper Inlet Creek (258)	2003	183
Deer Pass Creek (259)	2003, 2004	194
Duthie Creek (260)	2002	153
Kainet River (261)	2002	158
Klewnuggit Creek (262)	2007	120
Kwakusdis River (263)	2002	195
Lard Creek (264)	2003, 2004, 2006	274
McLoughin Creek (265)	2002, 2005	392
Mussel River (266)	2002	155
Neekas Creek (267)	2002	202
Quartcha Creek (268)	2002	149
Roscoe Creek (269)	2002, 2005	349
Salmon Bay Creek (270)	2004, 2005	224

Bella Coola River (271)	2002	206
Cascade River (272)	2002	194
Dean River (273)	2003	91
Elcho Creek (274)	2002	169
Frenchman Creek (275)	2003, 2004	109
Hooknose Creek (276)	2003, 2004, 2005	65
Jenny Bay Creek (277)	2002	199
Kimsquit Bay (278)	2003, 2004	59
Kimsquit River (279)	2003, 2006, 2007	151
Martin River (280)	2001, 2002, 2003	276
Necleetsconnay River (281)	2004	13
Nooseseck River (282)	2004	61
Skowquiltz River (283)	2003, 2004, 2006	194
Snootli River (284)	1992	46
48) Rivers Inlet		
Ashlulm River (285)	2003	44
Chuckwalla River (286)	2002, 2006	55
Clyak River (287)	2002, 2003, 2006	52
Draney Creek (288)	2003, 2005, 2007	153
Lockhart-Gordon Creek (289)	2002, 2003, 2005, 2006	133
McNair Creek (290)	2002, 2003, 2005	110
Milton River (291)	2003, 2006	40
Amback Creek (292)	2002	46
49) Smith Inlet		
Nekite River (293)	1989, 2002, 2004, 2005	499

Walkum Creek (294)	2003, 2004, 2005, 2006	226
50) Johnstone Strait		
Ahnuhati River (295)	2004, 2005, 2006	204
Ahta River (296)	2002, 2003	37
Glendale River (297)	2003, 2004	94
Kakweiken River (298)	2003, 2006	20
Klinaklini River (299)	1997, 2002	116
Mackenzie Sound Creek (300)	2004	18
Viner Sound Creek (301)	2002, 2003	204
Taaltz Creek (302)	2002, 2003	14
Waump Creek (303)	2002, 2003	33
Nimpkish River (304)	2002, 2004	409
Wortley Creek (305)	2002	242
Heydon Creek (306)	1998, 2001, 2003	249
51) South Coast		
Homathko River (307)	2004	203
Southgate River (308)	2003, 2004	222
Algard Creek (309)	2003	99
Orford River (310)	2003	320
Kwalate River (311)	2004	15
Mashiter Creek (312)	2004	56
Sliammon River (313)	1991	50
Squamish River (314)	2002, 2003	84
Cheakamus River (315)	1992, 2003	91
Mamquam River (316)	1991, 2002, 2004	152

Shovelnose Creek (317)	2004	147
Stawamus River (318)	2004	45
Theodosia River (319)	2002	144
Tzoonie River (320)	1991	50
Indian River (321)	2000, 2002	344

52) Vancouver Island- east coast

Campbell River (322)	2002	193
Puntledge River (323)	1991, 2007	200
Cold Creek (324)	2002	193
Big Qualicum River (325)	1992, 2007	220
Little Qualicum River (326)	1991, 2007	260
Nanaimo River (327)	1991, 1997, 2001, 2002	249
Chemainus River (328)	1992, 1997	167
Cowichan River (329)	1997, 2000	273
Goldstream River (330)	1991, 1992, 1997, 1999	285

53) Vancouver Island – west coast

Cayeghle Creek (331)	2002, 2003, 2004	137
Colonial Creek (332)	2002	221
Demamiel Creek (333)	1992	50
Goodspeed River (334)	2002	195
Hathaway Creek (335)	2002	43
Nahmint River (336)	2003	195
Nitinat River (337)	1992, 2004	243
Pegattum Creek (338)	2002	63
Smith Creek (339)	1997	88

Sugsaw Creek (340)	2004	90
54) Fraser River		
Alouette River (341)	1991	37
North Alouette River (342)	2004	57
Blaney Creek (343)	2004, 2007	55
Chehalis River (344)	1991	98
Chilliwack River (345)	1992, 2004	197
Chilqua Creek (346)	2004, 2005, 2007	218
Harrison River (347)	2002	201
Harrison River later (348)	2006	100
Hicks Creek (349)	2006, 2007	24
Hopedale Creek (350)	2005, 2007	59
Inch Creek (351)	2002, 2003	405
Kanaka Creek (352)	2004, 2005, 2006, 2007	245
Kawkawa Creek (353)	2004	65
Lower Lillooet River (354)	2002	122
Norrish Creek (355)	2004	208
Serpentine River (356)	2004	38
Silverdale Creek (357)	2000, 2004, 2005, 2007	231
Squakum Creek (358)	2000, 2004, 2005, 2007	427
Stave River (359)	1991, 2003	243
Vedder River (360)	2002, 2003	75
Wahleach Slough (361)	1991	50
Widgeon Slough (362)	2004	191
Worth Creek (363)	2005, 2007	127

55) North Puget Sound

Nooksack River (364)	1998	100
Skagit River (365)	1998	90
County Line Creek (366)	1994	97
Tulalip Creek (367)	2003	99
Grant Creek (368)	2003	50
Siberia Creek (369)	1993	75
Skykomish River (370)	2007	82

56) South Puget Sound

Kennedy Creek (371)	2003	100
Minter Creek (372)	2003	100
Green River hatchery (373)	2007	100

57) Hood Canal

Big Quilcene River (374)	1997, 2000	88
Hoodsport hatchery (375)	2003	102

58) Strait of Juan de Fuca

Salmon Creek (376)	2000	100
Elwha River (377)	1995	100

59) Coastal Washington

Ellsworth Creek (378)	2000	61
Bitter Creek (379)	2000	106
Quinault River (380)	1998	100
Satsop River (381)	1998	97

List of Figures

Figure 1. Map indicating the major geographic regions from which chum salmon were surveyed for microsatellite variation. Populations sampled in each region are outlined in Appendix Table 1.

Figure 2. Relationship between the number of fish surveyed in a specific population and the accuracy obtained for the same population (A) and regional accuracy (B) during estimation of composition of simulated single-population mixtures employing a baseline of 381 populations. Regions and populations within regions have been outlined in Appendix Table 1.

Figure 3. Relationship between the number of alleles observed at a microsatellite locus and the average percentage accuracy to population (A) and standard deviation (B) obtained for single-population mixtures using only a single locus and the 381-population Pacific Rim baseline for 40 populations of chum salmon. Loci are numbered as in Table 1.

Figure 4. Relationship between the total number of alleles used in stock composition estimation and the average percentage accuracy to population (A) and standard deviation (B) obtained for single-population mixtures using the 381-population Pacific Rim baseline for 40 populations of chum salmon.

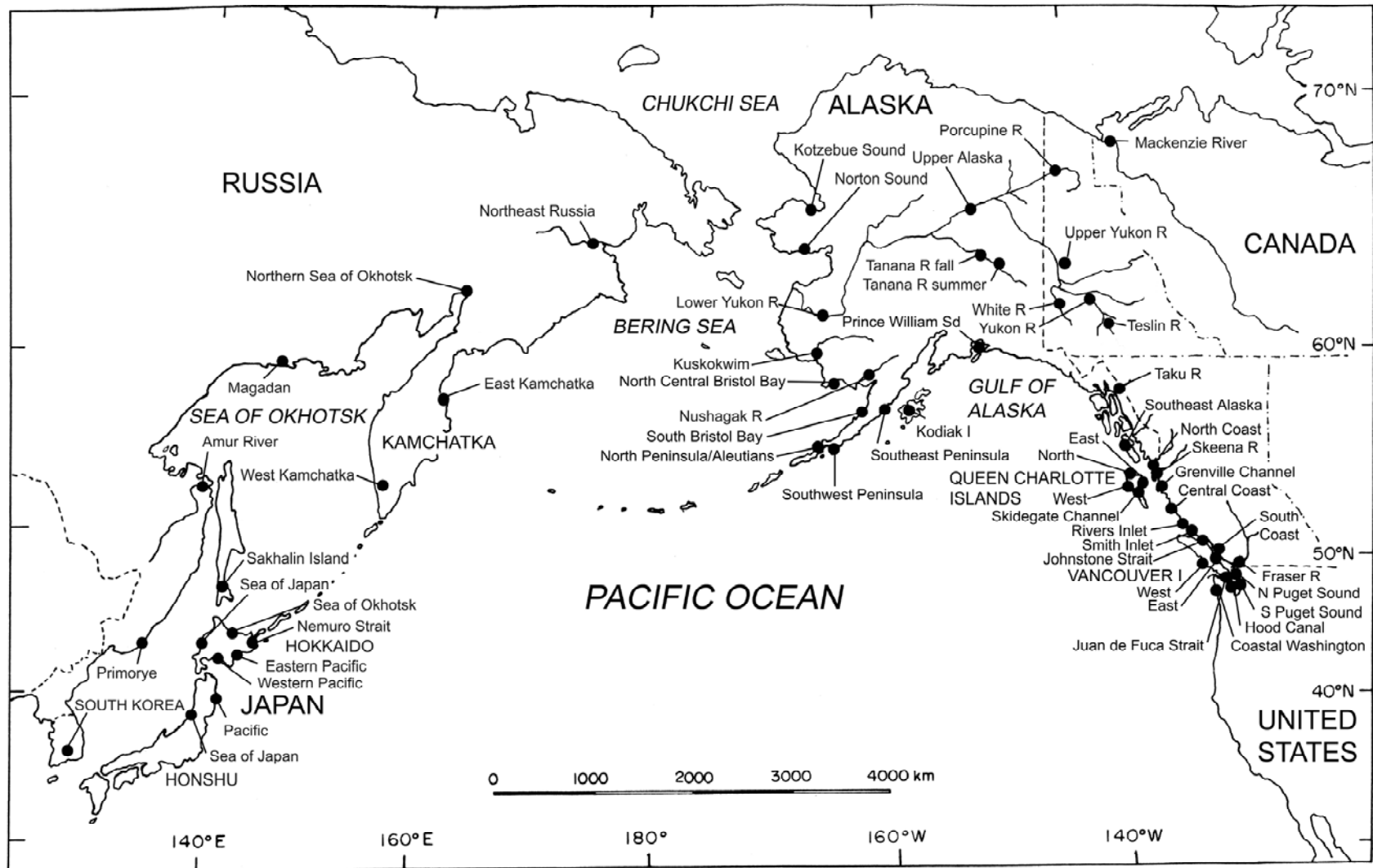
Figure 5. Accuracy and precision of estimated stock compositions by reporting region for simulated mixtures of chum salmon containing populations and regions from a single geographic area.

Figure 6. Percentage correct assignments (A) and standard deviation (B) for simulated single-region samples of chum salmon populations with a Pacific Rim distribution estimated

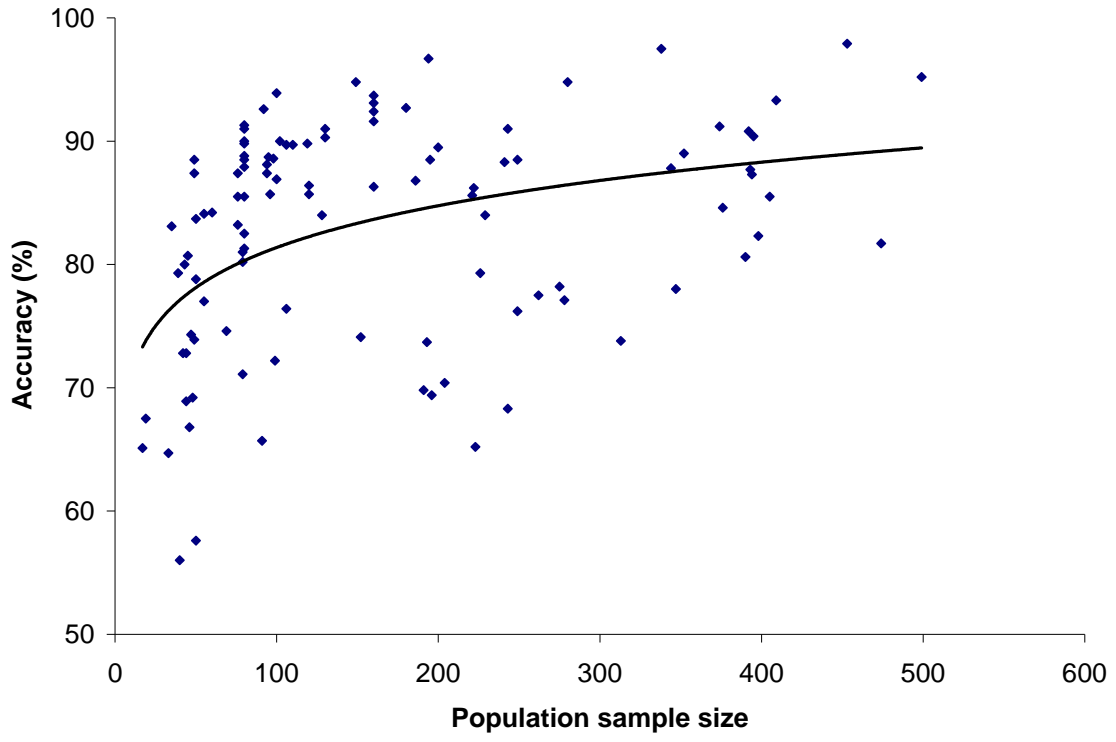
with 30 mtSNPs surveyed by Sato et al. (2004), as well as each of six microsatellites. Average accuracy and precision over all regions is also indicated.

Figure 7. Percentage correct assignments (A) and standard deviation (B) for simulated single-population samples of eight chum salmon populations from western Alaska estimated with 61 SNPs surveyed by Smith and Seeb (2008), as well as different sets of microsatellites. Average accuracy and precision over all populations is also indicated.

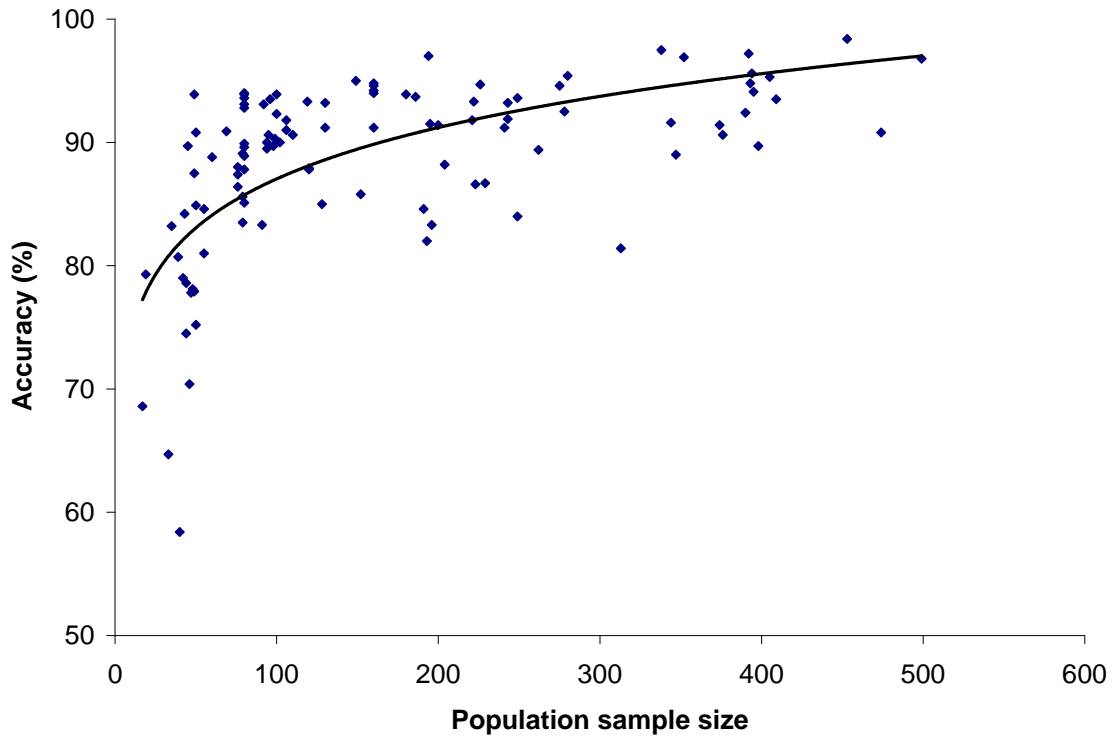
Figure 8. Relationship between percent correct assignment (A) and standard deviation (B) with number of SNPs used in stock composition estimation for western Alaska chum salmon populations.



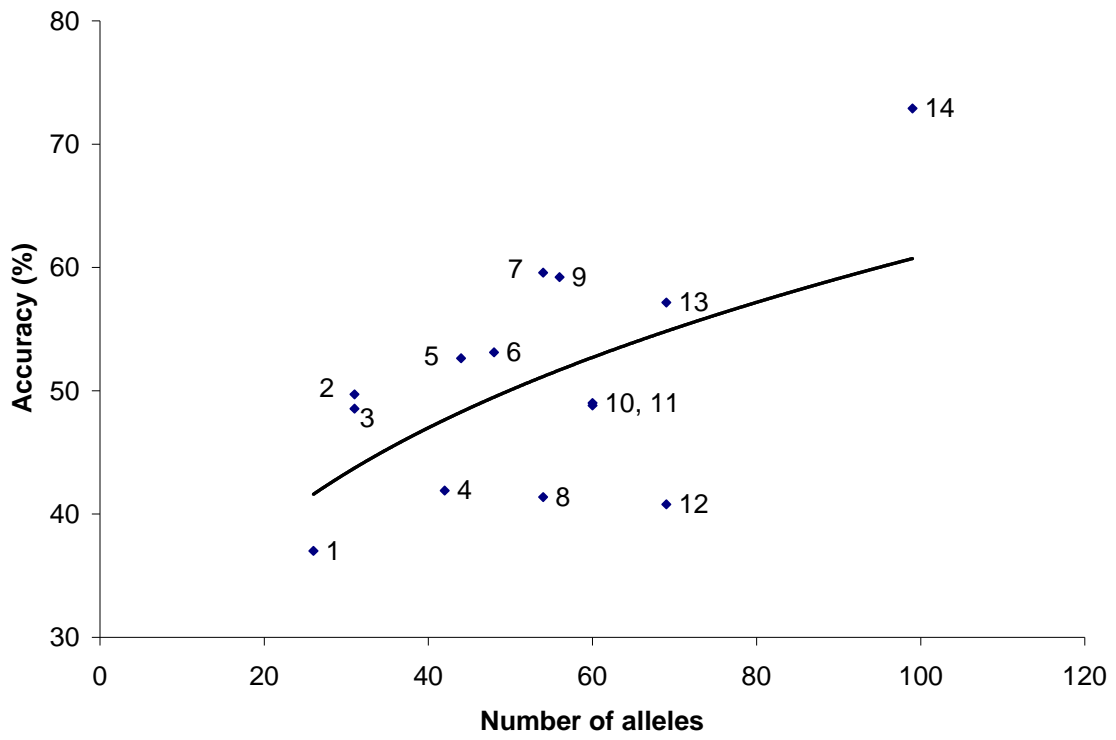
A



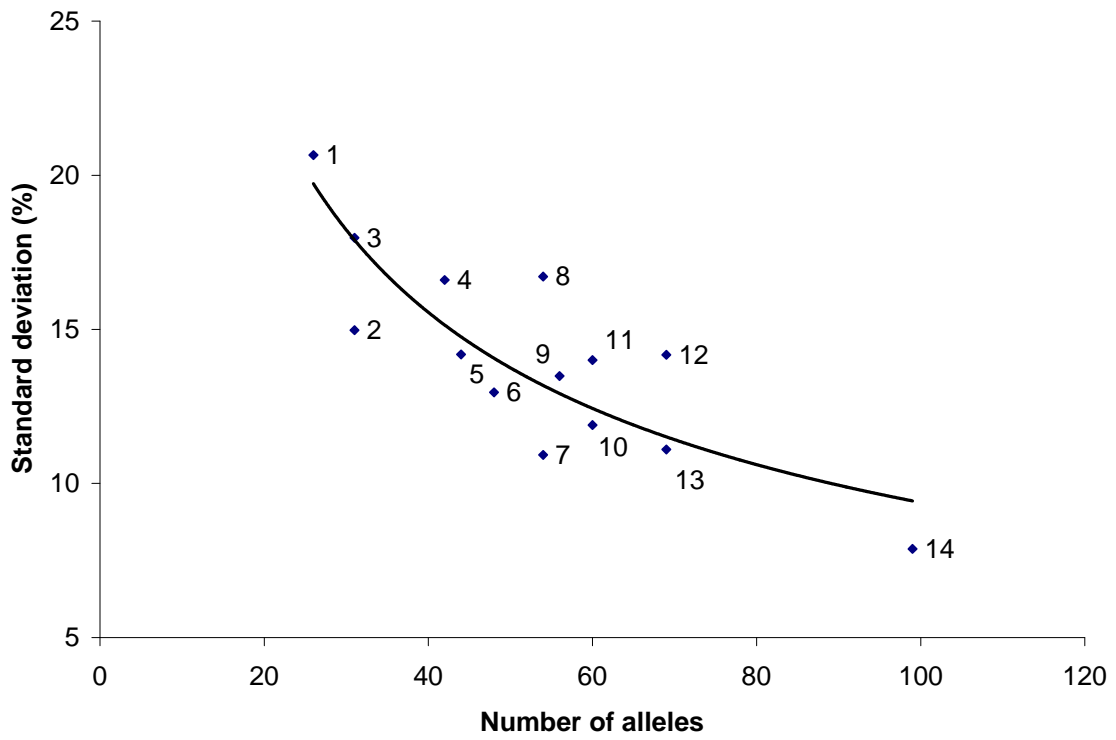
B



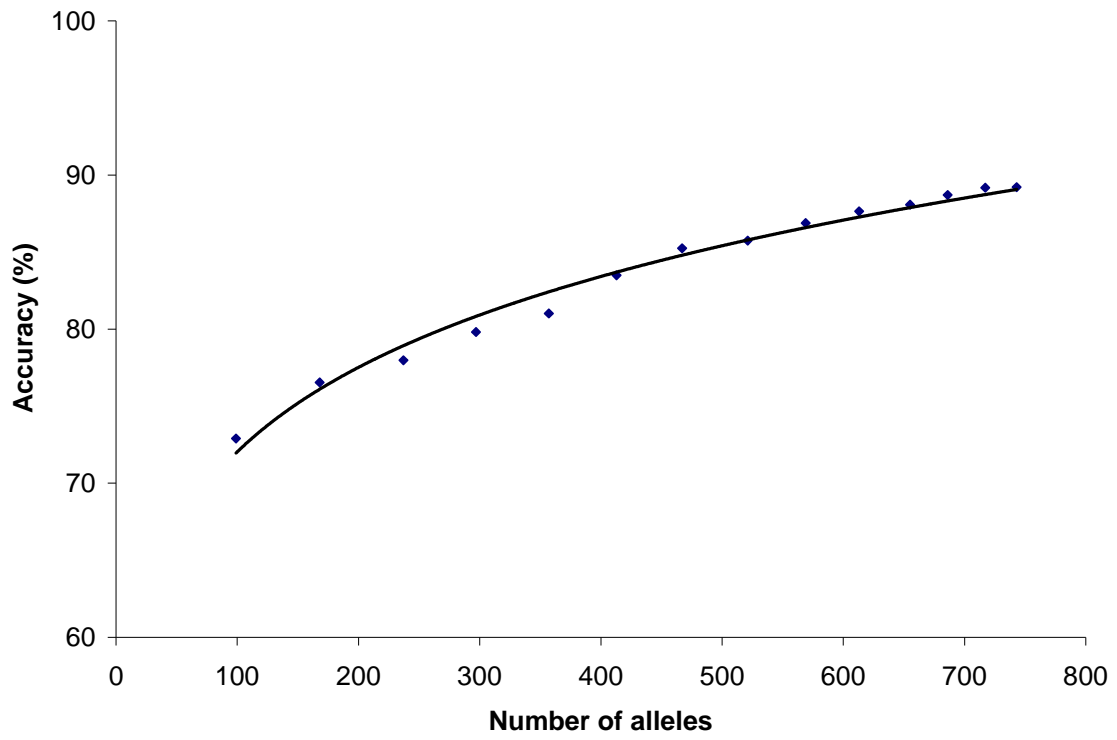
A



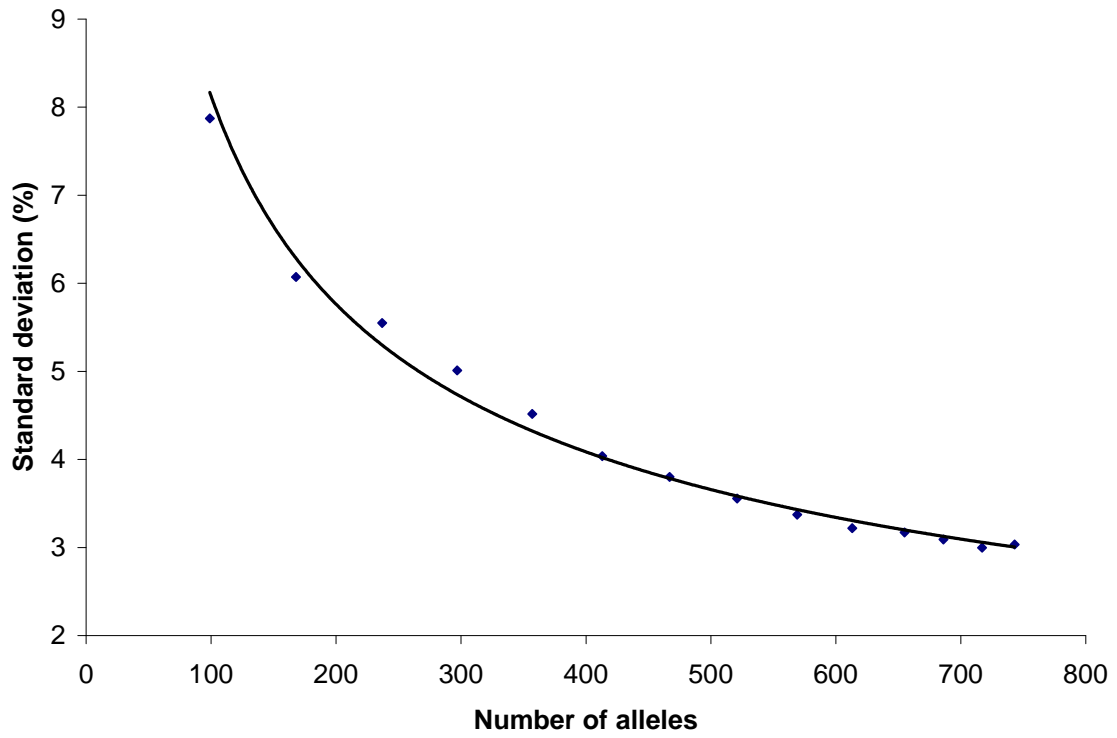
B

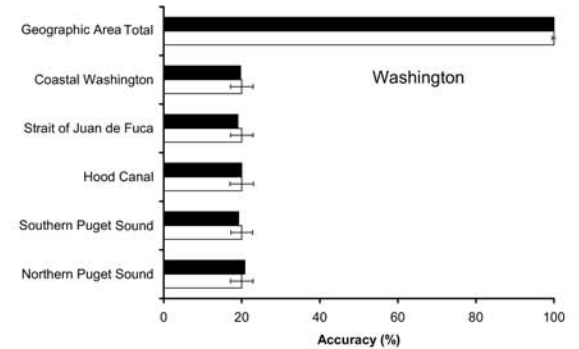
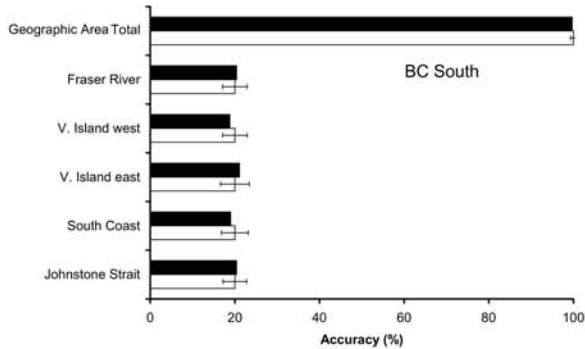
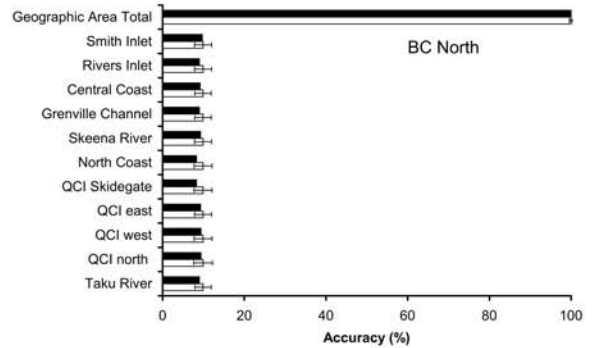
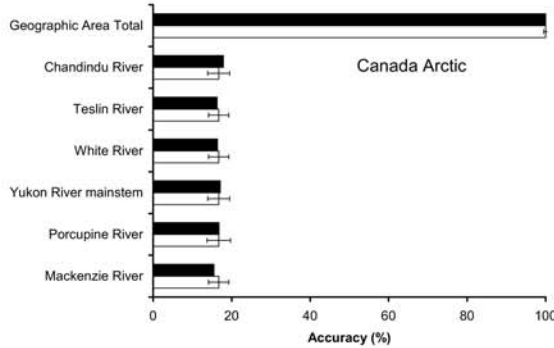
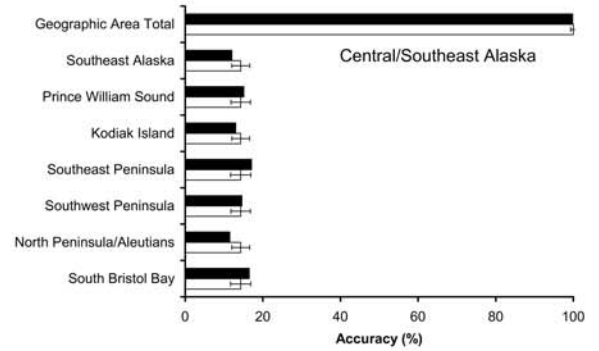
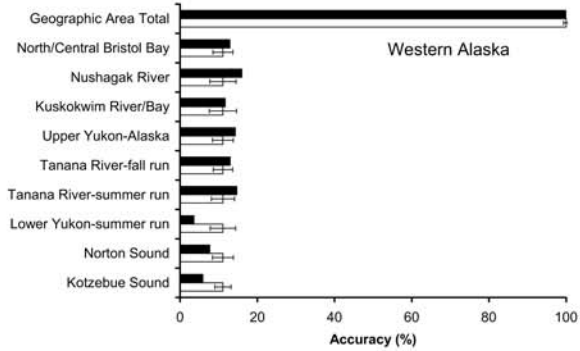
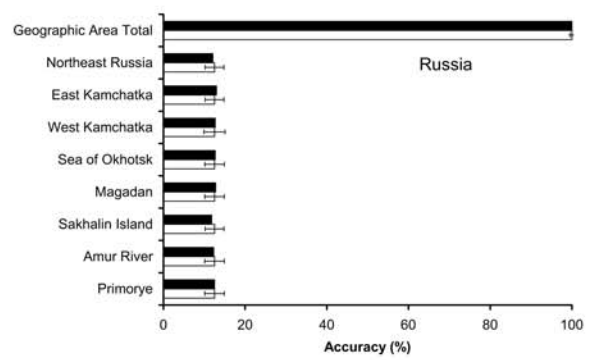
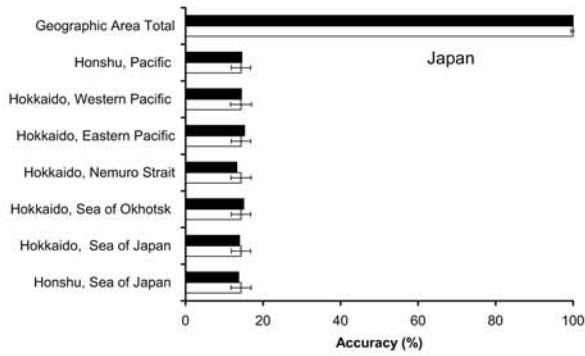


A

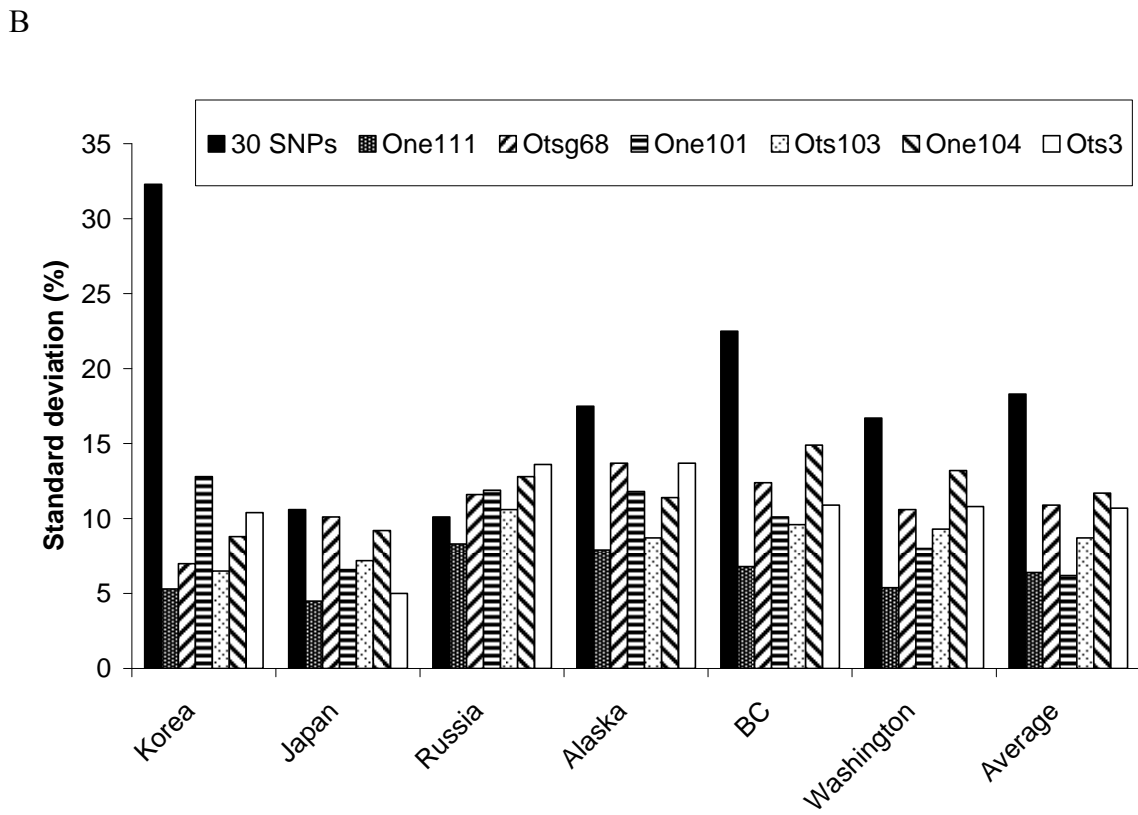
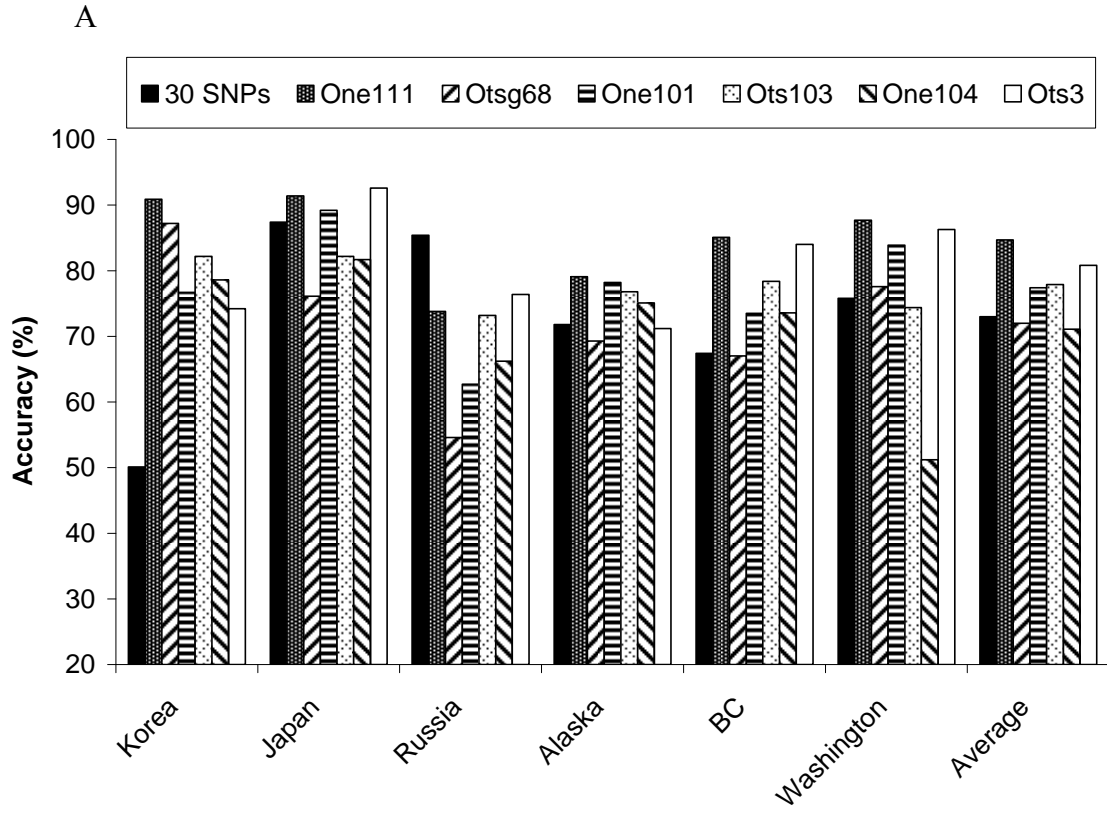


B

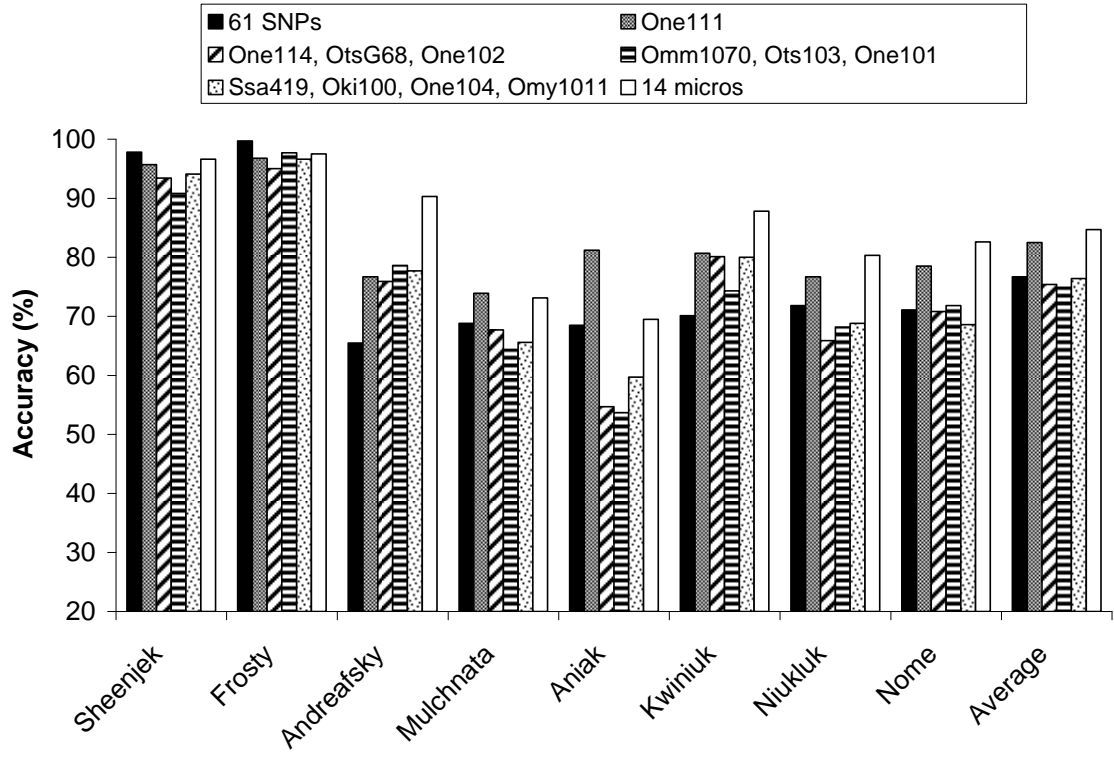




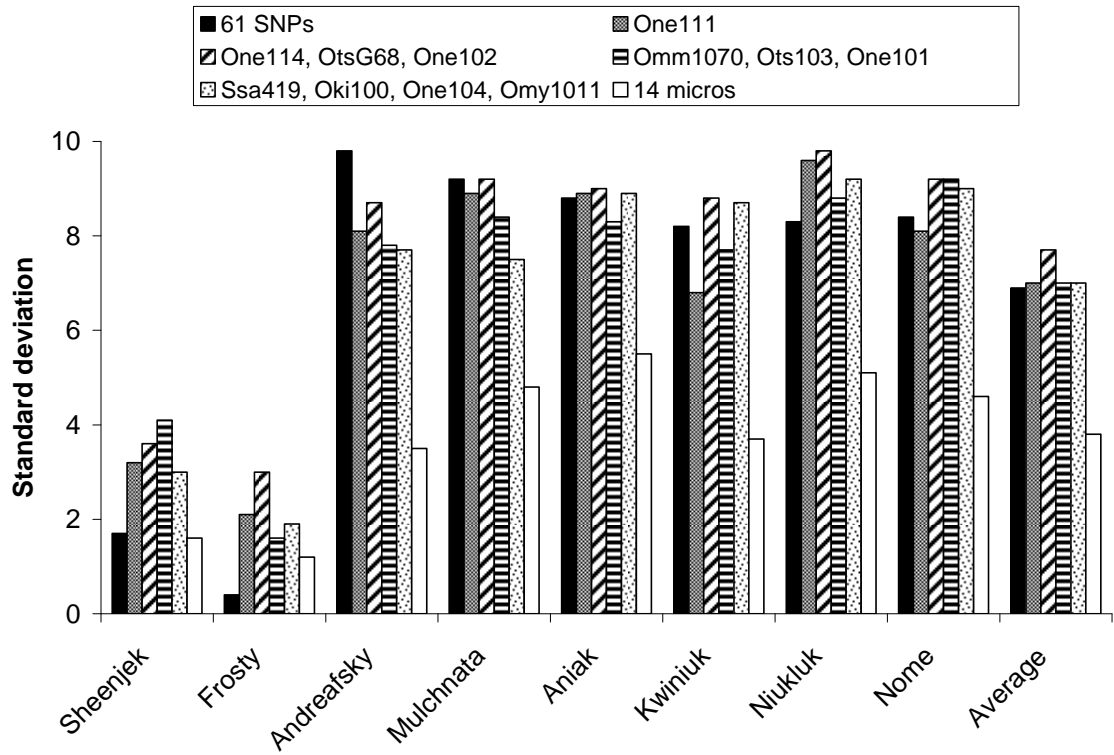
■ Actual □ Estimated



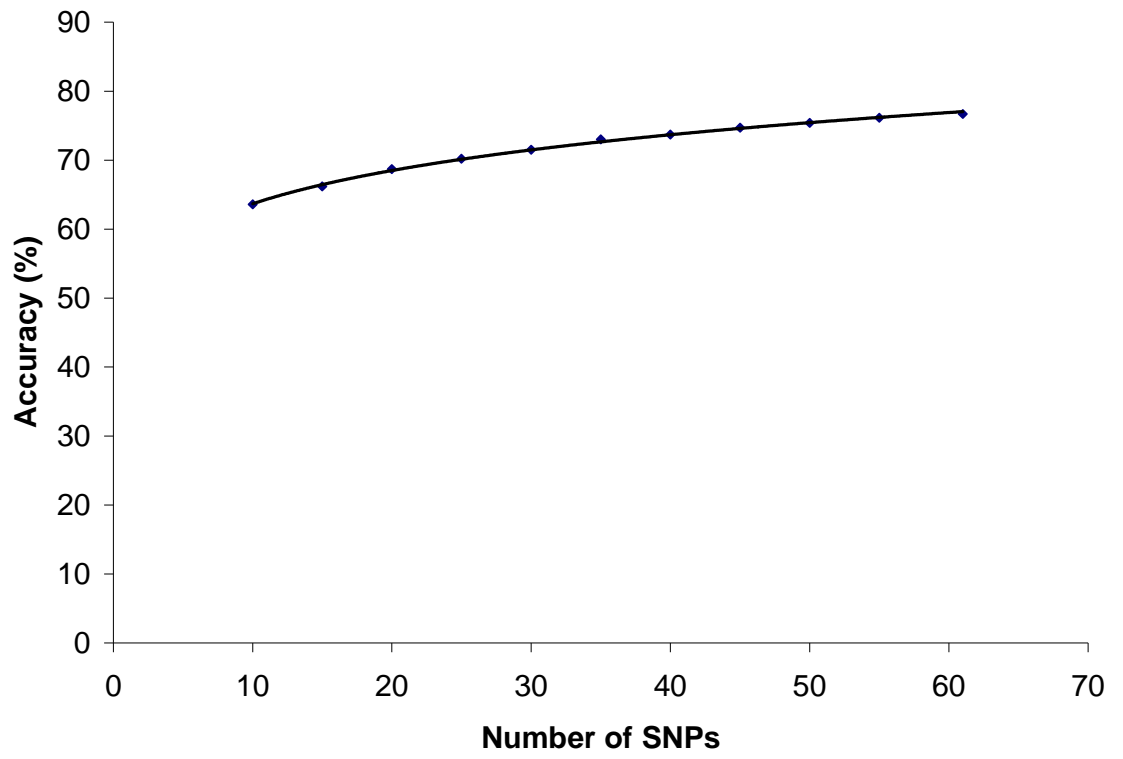
A



B



A.



B

