Estimation of Conservation Unit and population contribution to Chinook salmon mixed-stock fisheries in British Columbia, Canada using direct DNA sequencing for single nucleotide polymorphisms

Terry D. Beacham[1], Colin Wallace, Kim Jonsen, Ben J. G. Sutherland, Carrie Gummer, and Eric B. Rondeau[1]

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

[1] Corresponding author

ABSTRACT

Determination of population structure and stock identification is a general problem in fisheries assessment and management. Pacific salmon fishery management regimes are evolving to require higher resolution of stock composition on increasingly smaller reporting units. For Chinook salmon (*Oncorhynchus tshawytscha*), a stock identification baseline comprised of some 125,198 individuals from 369 populations ranging from Russia to California was employed for genetic stock identification (GSI). GSI analysis based upon variation at up to 547 single nucleotide polymorphisms (SNPs) was demonstrated to provide accurate estimates of stock composition for 68 Conservation Units (CUs) in British Columbia, 23 reporting groups in the United States, and one reporting group in Russia. In many instances, accurate population-specific estimates of stock composition within a CU were possible in fishery samples, as well as identifying individuals to some specific populations. A genetics-based assessment system provides an opportunity for conservation-based management of Canadian Chinook salmon.

Keywords: conservation, Chinook salmon, genetic stock identification, fisheries management

Introduction

In Canada, the Policy for Conservation of Wild Pacific Salmon (WSP) was established with the goal of maintaining and restoring healthy and diverse Pacific salmon populations, making conservation of wild salmon and their habitats the highest priority for resource management decision-making (Fisheries and Oceans Canada 2005). Under the WSP, wild salmon populations are identified and maintained in Conservation Units (CUs) that are identified based on genetic traits, biogeographic distribution, life-history characteristics, and local knowledge where available. For Chinook salmon (*Oncorhynchus tshawytscha)*, 84 CUs have been defined for Canadian populations. Under the WSP, the objective is to manage fisheries and hatchery production to ensure that wild populations are safeguarded and harvest benefits are sustainable. Price et al. (2014; 2017) suggested that any suitable assessment technique must provide resolution for all individual CUs to meet the conservation requirements of Canada's WSP. Pacific salmon fishery management regimes are evolving to require higher resolution of stock composition on increasingly smaller reporting units. The question becomes how can fishery management regimes adapt to meet the requirement of evaluating fishery impacts on increasingly smaller assessment units.

In British Columbia (BC), Chinook salmon fisheries are currently assessed and managed via coded-wire tags (CWT, Jefferts et al. 1963) applied to juveniles prior to hatchery release in fewer than 20 populations ("indicator" populations) where abundance is augmented via hatchery production, and by subsequent recovery of the CWTs in fishery and escapement sampling. The CWT evaluation program provides direct information on only the indicator populations, and is unable to estimate stock composition or fishery impacts by CU. It is neither technically nor financially feasible to expand the CWT marking program to have an indicator population for

every Chinook salmon CU in BC. If successful implementation of the WSP requires an assessment of fishery impacts by individual CU, the obvious question arises as to how this can be achieved.

Genetic stock identification (GSI) has a long history of being associated with management and assessment of Pacific salmon fisheries (Fournier et al. 1984; Milner et al. 1985; Shaklee et al. 1999). GSI has undergone a series of technological improvements over the last 40 years, with genetic markers ranging from allozymes, minisatellites, microsatellites, and single nucleotide polymorphisms (Utter et al. 1987; Beacham et al. 1996; Moran et al. 2018; Larson 2014). GSI of Pacific salmon in mixed-stock fisheries is important to undertake to allow fishery managers to decide on the timing and area of local salmon fisheries, as well as assess the impact of the fisheries on stocks, particularly those of conservation concern (Hess et al. 2016). Beacham et al. (2021) concluded that a combined GSI and parentage-based tagging (PBT) approach can provide information to improve Chinook salmon assessment and conservation in British Columbia (BC). PBT uses molecular-based approaches to conduct large-scale parentage assignments and has resulted in the unprecedented ability to identify genetically millions of hatchery-origin salmonids (Steele et al. 2019). GSI and PBT applied in combination can provide high-resolution estimates of stock composition, as assignment of individual Chinook salmon via PBT is virtually 100% accurate with respect to hatchery of origin and age of the individual (Beacham et al. 2018; 2021). However, PBT has not been applied to wild Chinook salmon populations in BC, and thus accuracy of estimated stock compositions of fisheries where hatchery-produced individuals comprise only a small portion of the fishery will be dependent mainly upon the accuracy of estimates generated via GSI. Thus, it is essential that GSI alone

provides accurate estimates of stock composition to CU in order to evaluate fishery impacts on both wild and hatchery-enhanced populations.

Direct DNA sequencing, coupled with automated scoring of the genotypes, results in cost-effective genotyping and unprecedented ability to provide accurate estimates of stock composition or individual identification to very discrete geographic regions or CUs. As noted by Beacham et al. (2020), it is a new era in the application of genetic variation to resource management. They reported that accurate identification of Canadian coho salmon (*O. kisutch*) sampled from mixed-stock fisheries was possible to the CU level, enabling assessment of fishery impacts that was sufficiently informative for conservation-based management as envisaged in the WSP. However, in BC and the Yukon River, approximately twice as many CUs have been defined for Chinook salmon (84) as compared with coho salmon (44). Previously, based upon a microsatellite baseline, the contributions of 22 Canadian reporting groups to mixed-stock fisheries were routinely estimated (Beacham et al. 2006a). Resolution of mixed-stock Chinook salmon fishery stock compositions in Canada to the CU level would be unprecedented, as accurate estimation of fishery contributions for up to 84 reporting groups would be required, and GSI has never been applied to provide such fine-scale resolution of stock composition in Canada for Chinook salmon. Evaluation of the ability to assess accurately the stock composition of mixed-stock Chinook salmon fishery samples by CUs and population within CU was the key objective of the current study because if possible, evaluation of fishery impacts on both hatchery and natural-origin salmon would be possible with a single technique, particularly if all hatchery-origin individuals were marked with an adipose fin clip

In January 2021, the Canadian Prime Minister's Office directed the Minister of Fisheries and Oceans and the Canadian Coast Guard to develop a Pacific Salmon Strategy that is intended

to conserve and protect wild Pacific salmon and their habitats and ecosystems. At this time it is uncertain whether a GSI-PBT system of assessment will be included for Chinook salmon fisheries as part of the new strategy, but if it is, Canada will apply GSI and PBT for Canadian fisheries assessment for Chinook salmon, and the GSI component outlined in the current manuscript is a critical part of the application.

The current study is an evaluation of the application of the GSI methodology initially outlined by Beacham et al. (2018) to determine whether GSI can be used to provide accurate information on fishery contributions by CU for Chinook salmon CUs currently identified under the WSP. Hundreds of SNPs were amplified in a single PCR and, combined with direct DNA sequencing of the resultant amplicons and automated scoring of the SNP genotypes, resulted in rapid and cost-effective genotyping. Although the current study was directed towards Chinook salmon, Beacham et al. (2020) demonstrated accurate estimation of stock composition by CU for coho salmon, and it seems likely that similar procedures could be used for other salmonid and non-salmonid species. In the current study, a stock identification baseline comprised of some 125,198 individuals from 369 populations ranging from Russia to California was employed for Chinook salmon GSI evaluation. Genotypes from each population in a CU or geographic region were simulated and estimated stock composition of the single-population and multi-population simulated mixtures determined via GSI referencing the 369-population baseline.

Methods and Materials

General Methods

Evaluation of stock identification capability initially proceeded by development of a baseline of populations likely to contribute to mixed-stock fishery samples. Once the baseline

was available, a series of tests was conducted in order to evaluate the effectiveness of the baseline in producing reliable estimates of stock composition in mixed-stock fishery samples. The initial step included arranging Canadian populations into CUs and American and Russian populations into geographically-based reporting units. CUs in southern BC are outlined in Supplementary Figure S1, CUs in northern BC in Supplementary Figure S2, and CUs in the Yukon River drainage in Supplementary Figure S3. With potential reporting units and populations within reporting units determined, an analysis was undertaken whereby simulated single-population samples were analyzed, and the baseline was used to estimate stock composition of the simulated samples for both population and reporting group. An estimated stock composition value of 90% to the reporting group for the 100% single-population sample is generally considered as satisfactory for fishery management applications (Seeb and Crane 1999; Seeb et al. 2000; Beacham et al. 2012a). The next step was an evaluation of individual self-assignment accuracy to both population and reporting group. For a specific population, an 80% self-assignment accuracy has been considered as sufficient for maintaining the population as a reporting unit (Gilbey et al. 2016). We next evaluated the baseline by simulating two mixed-stock fishery samples comprised of 10 populations each, with the simulated samples focusing on populations potentially present in northern BC and southern BC mixed-stock fisheries. The analysis continued with evaluation of samples of known origin as initially represented by 1,309 individuals recovered from 2017 hatchery broodstocks or escapements that had been marked with CWTs. The next known-origin sample was obtained from the 2018 escapement in the Campbell River drainage on the east coast of Vancouver Island. Known-origin samples were also developed by grouping of individuals that had been identified via PBT for 2018 and 2019 fishery sampling and 2019 hatchery broodstocks. Even if estimated stock compositions of

simulated mixed-stock fishery samples are accurate, there is a potential for biased estimates of stock composition from actual mixed-stock fisheries if a substantial portion of the fishery sample is derived from reporting groups inadequately represented in the baseline. The final step in the analysis was estimation of stock composition of actual fishery samples from geographically-dispersed fisheries in BC as a means to evaluate whether the presence of unsampled populations in the mixed-stock sample will cause bias in estimated stock compositions which would be indicated by unexpected stock compositions in the mixed-stock fisheries.

Baseline sample collection

The initial baseline was outlined by Beacham et al. (2018), and consisted of 36,241 individuals from 45 populations genotyped at up to 321 SNPs, with the distribution of populations largely in southern BC. With the original SNP panel modified, Beacham et al. (2021) reported that the baseline was subsequently expanded to include 105,722 individuals genotyped at up to 391 SNPs, ranging from Russia, the Yukon River drainage, southeast Alaska, BC, the Pacific Northwest, and California. The SNP panel was again modified in 2019, and the baseline in the current study was expanded again to include 125,198 individuals from 369 populations genotyped at up to 549 SNPs (Figure 1). Populations included in the baseline are outlined in Supplementary Table S1, with the populations from BC arranged by CU, and with Russian and United States of America (US) populations arranged by geographic (reporting) region. The SNPs genotyped in the expanded panel are outlined in Supplementary Table S2, along with primer sequences for the amplicons and F_{ST} and heterozygosity estimates for the SNPs. Some SNPs were found to have duplicate positions when aligned to the Chinook reference genome; these are most likely a result of genome assembly artifacts and not true

duplications (K. Christensen, Univ. of Victoria, pers. comm.). For markers with "Y" in the "Multiple" column in Supplementary Table 2, the position represented reflects one of the possible locations. Prior to 2013, most samples were collected opportunistically to provide a baseline for the previous microsatellite analyses concerning population structure and stock identification (Beacham et al. 20006a,b). From 2013 onwards, PBT was the objective of population sampling for selected hatcheries in British Columbia, and samples were collected to allow complete genotyping of the broodstock in a particular year. Any new population samples collected outside of the PBT populations were also genotyped with the panel in use at the time of analysis. Fin tissue or operculum punches were obtained from all individuals sampled.

Fishery sample collection

General fishery sampling procedures were described by Beacham et al. (2021). Summarized briefly, samples in 2018 and 2019 were collected from marine commercial, recreational, and First Nations fisheries throughout British Columbia. In the Strait of Georgia and the Juan de Fuca Strait recreational fishery off Victoria, BC, samples from the recreational fishery were obtained from a DFO creel survey program supplemented by samples provided by the Avid Anglers. The Avid Anglers are a “citizen science” group of volunteers who, if given the opportunity, fish year round, and collect biological information and tissue samples.

Library preparation and genotyping

The detailed procedure for library preparation and genotyping was outlined by Beacham et al. (2018), and a summarized version provided by Beacham et al. (2021). The process

involved loading amplified DNA from 768 individuals (up to 549 amplicons per individual) on a P1 chip v3 (chip used with the Ion Torrent Proton sequencer) with an Ion Chef (laboratory instrument used to robotically load DNA libraries on to a sequencing chip). Two chips were loaded consecutively with one run of the Ion Chef, and both chips were then subsequently loaded on to an Ion Torrent Proton sequencer. After the sequencing run was completed, amplicon sequences were aligned using BWA mem v0.7.17-r1198 (Li 2013) against the coho salmon (*O. kisutch*) genome (RefSeq assembly accession GCF_002021735.1) that also contained sequences used for Chinook-specific amplicon design which couldn't be placed in the coho genome; any potential conflicts based on the appended sequences were masked out. Genotype determination was conducted with Proton software Variant Caller®, and SNP genotypes at the sites specified by the hotspot file within target regions were called by Variant Caller. Genotypes at all available SNPs for each individual were collected to provide multi-locus genotypes. Genotypes had to be available for at least 150 SNPs for an individual to be retained in the baseline. In a test where the DNA of the same 768 individuals was genotyped on two occasions, an average genotyping error rate of 1.14% (1,839 discrepancies in 161,280 single-locus genotype comparisons) was observed over the 319 SNPs scored (Beacham et al. 2018). The species identification SNP *OkiOts_120255-113* (Starks et al. 2016) and sex identification SNP *Ots_SEXY3-1* were omitted from subsequent parentage and GSI analyses, leaving 389 SNPs for subsequent general analysis, and 547 SNPs for a group of 58 populations that had been genotyped with the most recent version of the SNP panel.

Data analysis

Expected and observed heterozygosities by locus were determined with adegenet (Jombart and Ahmed 2011). Estimation of F_{ST} by locus was conducted with ape (Paradis and Schliep 2018). BWA mem 0.7.17-r1188 (Li, 2013) was used to align amplified sequences to the reference, and a modified version of SNP-placer (https://github.com/CNuge/snp-placer) written in R (https://github.com/erondeau/snp-placer commit: eabfc78) was used to place the markers onto the Chinook reference genome assembly GCF_002872995.1 (Christensen et al, 2018). Bedtools getfasta v2.26.0 (Quinlan and Hall, 2010) was used to extract flanking sequences before and after alignment, and manually reviewed to ensure correct position was identified. Multiple best-alignments (MapQ = 0 in sam file) were flagged in Supplementary Table S2; the majority were determined most likely to be a result of genome assembly artifacts and not true duplications, and alternate mapping locations were removed from the genome using bedtools maskfasta v2.26.0 (Quinlan and Hall, 2010) for subsequent alignments. Positional information was used to plot F_{ST} by marker along the genome using R (R core team, 2020).

Genetic stock identification

The same techniques as outlined by Beacham et al. (2018) for GSI analysis were used in the current study. Summarized briefly, the genetic profiles of populations potentially contributing to a mixed-stock sample were used in conditional maximum likelihood mixture modeling as implemented in Rubias (Moran and Anderson 2019). For each sample, individuals were assigned with Rubias, with the population posterior means file as the basic file used for subsequent analyses. This file contained the probability of origin of the individual to each of the 369 populations in the baseline. After an initial burn-in of 25,000 iterations, the last 5,000 iterations from the Monte Carlo Markov Chain from Rubias were used to estimate the origin of

individuals and stock composition, with the mean allocation to each population in the baseline determined. Standard deviations of estimated stock compositions were also determined from the last 5,000 iterations from the Monte Carlo Markov Chain. Stock composition by CU or reporting group was determined by summation of allocations to all populations in the baseline that belonged to the CU or reporting group under consideration.

Simulated mixtures

To test the accuracy of identifying the CU and the population of origin, we performed GSI using Rubias where each individual in the mixture was probabilistically assigned to the closest genetic match from the set of populations in the baseline. To conduct 100% single-population simulations, we simulated mixture genotypes from each population sequentially and determined the allocation to the specific population simulated, as well as the allocation to the CU (Canada) or reporting unit (Russia, United States) to which the population belonged.

Assessment of accuracy of self assignment of individuals was conducted with Rubias where each individual in the baseline was evaluated for self-assignment accuracy both to individual population and to CU or reporting region. Leave-one-out cross validation analysis provided about 125,198 independent tests of known origin as they were collected from known specific Chinook salmon populations, with the assumption of limited straying among populations. The GSI and self-assignment of individuals analyses were conducted with the conditional maximum likelihood mixture modeling as implemented in Rubias.

The next stage of the evaluation incorporated analyses of two multi-population simulated mixed-stock fishery samples (200 individuals in each sample) for simulated mixtures as may be encountered in fishery sampling in northern and southern BC. Ten populations were

incorporated into each simulated mixture at set limits ranging from 1%-19% of the sample. Stock compositions of the resultant mixtures were estimated, and means and standard deviations determined for population and CU or reporting region estimates for 100 simulations of each mixture. Stock composition by CU or reporting group was determined by summation of allocations to all populations in the baseline that belonged to the CU or reporting group under consideration.

For estimation of stock composition in the fishery samples, after an initial burn-in of 25,000 iterations, the last 5,000 iterations from the Monte Carlo Markov Chain from Rubias were used to estimate the origin of individuals and stock composition, with the mean allocation to each population in the baseline determined. As with the simulated fishery samples, stock composition by CU or reporting group was determined by summation of allocations to all populations in the baseline that belonged to the CU or reporting group under consideration.

Known-origin samples

The first known-origin sample was constructed from individuals that contained a CWT and were genotyped from five populations in broodstock and escapement sampling in 2017. A total of 1,309 individuals was available for analysis, but the geographic range of these populations was restricted to Vancouver Island. The second known-origin sample was derived from sampling the 2018 escapement of the Campbell River population, also on the east coast of Vancouver Island, and was represented in the baseline by the Campbell River and Quinsam River populations. In order to increase the geographic range and the number of CUs and populations tested, we developed three additional test samples via PBT. PBT assignments were made to parents of known origin, and with that information, it was possible to determine the

origin and age of individuals sampled in fisheries or escapements. Two-parent PBT assignments have been demonstrated to be virtually 100% accurate with respect to population of origin for BC Chinook salmon (Beacham et al. 2018; Beacham et al. 2021). Accordingly, all two-parent PBT assignments observed in 2018 fishery sampling were considered as a known-origin sample (849 individuals), as were 2019 fishery identifications (2,295 individuals) and hatchery broodstock identifications (4,199 individuals). The PBT analysis here was independent of the subsequent GSI analysis, although the multi-locus genotype for an individual is the same basic input for both types of analyses. Estimated stock compositions for these samples were determined as outlined previously.

Results

Analysis of simulated single-population samples

The ability to obtain accurate estimates of stock composition to reporting unit is a key measure to evaluate in any assessment of the efficacy of estimating stock composition of mixed-stock fishery samples. Accurate estimates of stock composition were observed for most populations in the baseline at the CU or reporting group level (Figure 2). For Canadian-origin populations, mean accuracy of estimated stock composition was 93.7% to CU for 68 CUs encompassing 252 populations. For American-origin populations, overall accuracy of estimated stock composition was 95.4% to reporting unit for 23 geographic regions or return-time groups encompassing 104 populations. For Russian-origin populations, overall accuracy of estimated stock composition was 99.4% to the single geographic region encompassing 13 populations. An overall accuracy of 93.8% was observed across 92 CUs or geographic regions for all 369

298 populations evaluated, fulfilling the initial requirement of accurate estimation of stock
299 composition by reporting group.

300 In addition to generally accurate estimates of stock composition by reporting group,
301 accurate allocations to many individual populations were observed (Supplementary Figure S6).
302 For example, in the West Vancouver Island south age 0.x CU (31), estimated population stock
303 compositions for single-population samples were 99.1% for the Robertson Creek and 98.0% for
304 the Nitinat River populations. These high levels of accuracy were observed even though there
305 were 14 populations in the CU (Supplementary Figure S6). Similarly, in the East Vancouver
306 Island north fall age 0.x CU (29) with five populations, the estimated stock composition of the
307 Quinsam River sample was 98.9%. In the Lower Thompson River spring age 1.2 CU (17) with
308 eight populations, highly accurate estimates of stock composition were observed for the Nicola
309 River (98.0%) and Lewis Creek (99.5%) populations (Supplementary Figure S6). In essence, for
310 a number of populations in the baseline, accurate population-specific estimates of stock
311 composition should be available when applied to mixed-stock fishery samples.
312
313 Assignment of individuals
314 Estimation of stock composition of single-population samples displayed high accuracy
315 across CUs and many populations, and assignment of individuals, the most difficult of all stock
316 identification applications, mirrored the results observed in the single-population samples. In
317 general, accurate assignments (>80% accuracy) of individuals were observed across a range of
318 CUs or reporting regions, corresponding to those CUs or regions that displayed high accuracy in
319 estimation of single-population samples (Figure 3). For example, individuals from CUs in the

central coastal BC and the east and west coasts of Vancouver Island generally displayed a high level of assignment accuracy to CU, as did individuals from CUs in the Fraser River drainage.

High levels of accuracy were also observed for specific populations in the CUs or reporting units. Accurate self assignment of individuals was observed for a number of populations in the Yukon River drainage, southeast Alaska, central coastal BC, Vancouver Island, the Fraser River drainage, Oregon, and California (Supplementary Figure S7). For example, observed self-assignment accuracy was 99.7% for the Wannock River population (n=387) and 100% for the Docee River population (n=86) in central coastal BC, 98.7% for the Quinsam River population (n=9,556) on eastern Vancouver Island, and 100% for the Birkenhead River population (n=86) in the lower Fraser River drainage. Assignment of individuals could be accurate to both CU and potentially population within CU, dependent upon the population under evaluation.

Analysis of simulated multi-population fishery mixtures

Although accurate estimates of CU stock compositions and in some cases populations were observed in the single-population simulated fishery samples, the question remains as to whether accurate estimates can be obtained when individuals from multiple populations are present in a mixed-stock fishery sample. To evaluate model performance, we simulated two multi-population fishery samples as may be expected in northern and southern fisheries in BC and estimated for both population and CU/regional components. The average error of estimated stock compositions of a simulated mixed-stock fishery sample from northern BC containing individuals from 10 populations and nine CUs or regions was 0.8% for population and 0.6% for CU or region, with actual population contributions to the sample ranging from 2%-19% and CU

contributions ranging from 2%-30% (Table 1). For the southern BC simulated mixture, the average error was 0.3% for population and 0.2% for CU or region, with actual population and CU contributions to the sample ranging from 1%-16%. In particular, accurate estimates of stock composition were obtained for the Nicola River population, comprising only 1% of the sample. Accurate estimates of stock composition were obtained for simulated mixed-stock fishery samples if all populations present in the mixed-stock sample were present in the baseline, which indicated a successful completion in this step of the evaluation of the baseline for mixed-stock fishery analysis.

Effect of SNP addition to the panel

The enhanced panel was applied to genotyping 58 populations in the baseline, with the enhanced panel incorporating variation at 547 SNPs compared with the previous 389 SNPs for other populations in the baseline. For these 58 populations which were geographically spread throughout BC, the average increase in accuracy in the simulated single-population samples with respect to population estimation was modest (79.4% versus 79.1%), as was the increase in accuracy with respect to CU (96.0% versus 95.8%). Accurate identification of individuals to population was slightly higher with the enhanced panel (71.7%) than with the previous panel (69.5%), and accurate identification to CU was similarly modestly increased (94.4% versus 93.8%). Overall, there was a very limited response in estimation accuracy for both stock composition and individual identification in moving from a 389-SNP to a 547-SNP panel.

Analysis of known-origin mixtures

Estimated stock compositions of simulated single-population and multi-population samples suggested that accurate estimates of stock composition by CU and potentially by some individual populations should be possible when applied to mixed-stock fishery samples of unknown origin. Accuracy of estimated stock compositions was evaluated by analysis of five mixtures of known origin. The first mixture consisted of 1,309 individuals from five Vancouver Island populations from four CUs where individuals were identified via CWTs in 2017 hatchery broodstocks or escapements. Average error in estimation to specific population was 1.5% and that to CU 0.3% (Table 2). The second mixture was a 2018 escapement sample of 90 individuals from the Campbell River on Vancouver Island, with the Quinsam River a tributary in the drainage, and evaluated under the assumption of no straying individuals in the sample. No error was observed in the estimated stock composition, as the sample was estimated to have been comprised of 100% Quinsam River individuals.

PBT assignments allowed a wider range of populations and CUs to be included for known-origin analysis than available from CWTs and escapement sampling. Three known-origin samples were generated via two-parent PBT assignments in fishery or hatchery broodstock assignments, under the assumption of 100% accuracy to population of such assignments. The average error of GSI-determined stock composition of 849 individuals sampled in 2018 fisheries in BC was 0.6% per population across 19 populations and 0.2% per CU across 10 CUs. The average error of GSI-determined stock composition of 2,295 individuals sampled in 2019 fisheries in BC was 0.3% per population across 22 populations and 0.1% per CU across 13 CUs. For 2019 hatchery broodstocks comprising 4,199 individual identifications, the average error of GSI-determined stock composition was 0.6% per population across 18 populations and 0.1% per CU across 11 CUs. The results from estimation of stock composition of the simulated single-

population fishery samples suggested that accurate estimates of contributions of specific CUs and populations were possible. This capability has been verified for those CUs and populations in BC where a high accuracy of identification in the simulated single-population samples was observed.

Application to mixed-stock fishery sampling

Based upon the geographic locations and fishery timing, inferences can be drawn as to the reliability of the estimated stock compositions in actual practice. Estimated stock compositions from three terminal commercial fisheries (Areas 8, 23, and 25) reflected the terminal nature and geographic location of the fishery. The commercial fishery on the approach to the Bella Coola River exploited primarily the Bella Coola-Bentinck CU (90.5%) with some contribution from the adjacent Dean River CU (8.5%) (Table 3). On the west coast of Vancouver Island, terminal fisheries in Area 23 (Barkley Sound) exploited the West Vancouver Island-South fall age 0.x CU (100.0%), while that in Area 25 (Nootka Sound) exploited the West Vancouver Island-Nootka and Kyuquot fall age 0.x CU (99.8%).

Three sport fisheries adjacent to Vancouver Island were also evaluated for August stock compositions. In the northern Strait of Georgia fishery, the East Vancouver Island-Qualicum and Puntledge fall age 0.x (24.6%) and Lower Fraser River fall age 0.3 (31.3%) CUs were important contributors to the fishery, whereas in the Juan de Fuca Strait fishery, the South Thompson Summer age 0.3 was the dominant contributor (64.9%) (Table 3). In the southern west coast Vancouver Island fishery, the West Vancouver Island-South fall age 0.x CU was the dominant contributor (67.3%). In summary, estimated stock compositions of the fishery samples corresponded very well to those expected based upon fishery location and timing. In addition,

Z-scores generated from Rubias for the fishery mixtures displayed very few outlier individuals, indicative of a suitable baseline for analysis of the mixtures.

Discussion

Implementation of the WSP in Canada, along with the associated CUs, required estimation of stock composition to a resolution never before available in assessment of Canadian fisheries. The current study has provided accurate identification of Canadian Chinook salmon sampled from mixed-stock fisheries to the CU level, enabling assessment of fishery impacts for conservation-based management as envisaged in the WSP. Chinook salmon caught in Canadian commercial and recreational fisheries were identified to Canadian CUs and American geographic regions, confirming the utility of a GSI approach for conservation-based assessment of mixed-stock harvest on a wide geographic scale.

Effect of increasing SNP number in panel

There are three general levels of estimation in stock identification applications, namely reporting group, specific population, and identification of individuals, with the difficulty of the problem increasing from the former to the latter application. Increasing the number of SNPs employed in stock identification applications should intuitively increase the accuracy of estimation across all three levels of assignment, as well as increase the precision of the estimate. In the current study, moving from 389 SNPs to 547 SNPs provided very little increase in average accuracy to estimation of CU composition (0.2%), which may be expected given the high initial accuracy (95.8%) provided by the 389-SNP panel. The differential in population-specific accuracy increased between the panels, but only marginally (0.3%). As might be expected, the largest differential between the panels was observed in individual identifications (2.2%). With average accuracies of individual assignment to population ranging between 69.5% and 71.7% for

the two panels, these accuracies were considerably higher than the 56.7% accuracy reported by Beacham et al. (2012b) for a 72-SNP panel applied in a 60-population Chinook salmon baseline. Increasing the number of SNPs employed in estimation has the greatest effect on individual identification, although the effect of this appears to plateau as more SNPs are added.

If the objective of stock identification applications is to increase accuracy of population-specific estimates, one could opt for increasing either the number of SNPs employed or increasing population sample size in the baseline. Morin et al. (2009) reported that increasing sample size up to 100 individuals for detection of population structure was beneficial. Beacham et al. (2012b, 2020) reported that less genetically distinct populations required larger population sample sizes to achieve a given level of accuracy in estimated stock compositions. When increased accuracy of estimated stock compositions is required for a particular population in the baseline, if a sufficient number of markers is included in the panel already, the most direct route to follow is to increase sample size for the target population, rather than increasing the number of SNPs in the application.

Simulated fishery samples

A geographically-based population structure is usually required in the application of GSI in Pacific salmon, as an important assumption in the application is that the portion of the mixed-stock sample derived from populations not in the baseline is allocated to sampled populations from the same CU or region. GSI works well when this assumption is met, as the cost and complexity of developing a baseline for stock composition analysis is reduced when not all populations potentially contributing to a mixed-stock sample are included in the baseline. Population samples were available from 68 of the 84 CUs defined for Chinook salmon in

Canada, with the unrepresented CUs generally restricted to where the remoteness of the locations has precluded sample collection to date. A review of Z-scores from fishery mixtures did not demonstrate significant numbers of outliers. This suggests that missing CUs were either 1) not sufficiently distinct to be picked up as an outlier in the test, or more likely 2) represented a small number of individuals in the fishery. Planning is underway for collection of samples in some of these CUs, and we expect that population structure within these unsampled CUs will reflect a CU basis. If not, re-evaluation of defined CUs may be required.

Initial evaluation of genetic stock identification techniques includes analysis of simulated mixtures to evaluate the accuracy and precision of estimated stock compositions. In our study, analysis of simulated single geographic region or CU and multi-geographic region or CU mixed-stock fishery samples indicated that estimates with ample power for fishery assessment and management were obtained. If analysis of simulated mixtures produces satisfactory resolution in estimates of stock composition, then the next phase of the evaluation requires analysis of known-origin samples independent of the baseline used for estimation of stock compositions.

Analysis of known-origin mixtures

Accurate estimation of stock composition via GSI relies on a baseline that includes all major populations potentially contributing to a mixed-stock fishery sample. Analyses of the known-origin CWT, escapement, and PBT-derived samples confirmed the ability of the baseline to provide reliable estimates of stock composition for CUs when applied to analysis of known-origin mixed-stock samples, a result which had been suggested by single-population simulations. Furthermore, reliability of estimates of contributions for specific populations was generally confirmed, provided that high accuracy had been observed in the single-population sample

480 simulations. For those populations where underestimation of actual contributions was observed
481 and would be expected as illustrated by the single-population simulations, PBT can substantially
482 improve individual assignment and mixture allocation for specific populations (Beacham et al.
483 2018, 2020). Even if reliable estimates of stock composition have been obtained from known-
484 origin samples derived entirely from populations in the baseline, there is still a potential for
485 inaccurate estimates of stock composition in real fisheries applications if a significant portion of
486 the mixed-stock sample has been derived from populations or regions inadequately represented
487 in the baseline. The application of the baseline to estimation of stock composition for actual
488 mixed-stock samples is a means to evaluate whether individuals from unsampled populations
489 may be present in the mixture as evidenced by Z-scores or estimated stock compositions that are
490 unlikely to be accurate given the geographic location or timing of the fishery.

491

492 Mixed-stock fisheries

493 The GSI baseline can be applied to fisheries where both hatchery-origin and wild-origin
494 Chinook salmon are caught. In BC, there are many integrated hatcheries resulting in very similar
495 or undifferentiated hatchery and wild populations (Le Luyer et al. 2017). As long as the
496 population is represented in the baseline, no difference in accuracy or precision of estimated
497 stock compositions is expected between wild- and hatchery-origin individuals in a mixed-stock
498 fishery sample. However, if out-of-region population transfers have occurred with some
499 hatchery broodstocks, it may be difficult to ascribe region of origin of individuals from the
500 transferred population with the use of only genetic variation. As Canada does not yet mass mark
501 (remove adipose fin) all Chinook salmon hatchery production and thus there are no mark-
502 selective fisheries, exploitation rates between the hatchery and wild components of a population

503 should be similar. Stock composition of six fishery samples was estimated in order to evaluate
504 reliability of the estimates with the existing baseline. Given the geographic locations and timing
505 of the fisheries, it should be possible to evaluate estimated stock compositions of actual fishery
506 samples against expectations to measure performance of the baseline for stock identification
507 applications. In general, stock composition estimates were broadly concordant with estimates
508 from CWTs (historical or contemporary) and expectations based on time and area of the fishery
509 and proximity of putative source populations.

510 The GSI approach to fishery assessment enables catch by CU to be determined for any
511 Canadian fishery, and provides for managing a combination of mixed-stock ocean fisheries and
512 potential in-river fisheries that exploit only healthy CUs. There is no other method of fishery
513 assessment that can provide this level of resolution for mixed-stock analysis. A joint GSI-PBT
514 application marked the first time that Chinook salmon fisheries impacts in Canada were
515 evaluated by CU, and enabled an assessment that was sufficiently informative for conservation-
516 based management as envisaged in the WSP (Beacham et al. 2021). COSEWIC (2018) reported
517 that of 16 Designatable Units (DUs) of southern BC Chinook salmon with little or no hatchery
518 enhancement, 13 were identified as endangered, threatened, or of special concern. DUs
519 correspond reasonably well with CUs, and with all DUs/CUs represented in the baseline, genetic
520 analysis of samples from the 2018 and 2019 fisheries indicated areas and timing of fishery
521 impacts on DUs/CUs that were endangered, threatened, or of special concern (Beacham et al.
522 2021). The GSI baseline should prove informative with unprecedented accuracy of exploitation
523 of constituent populations and DUs/CUs within the COSEWIC assessment process.
524
525 Summary

526 The ability to provide reliable estimates of stock composition by CU was facilitated by
527 the switch from a microsatellite baseline (Beacham et al. 2006a,b) to a SNP baseline with the
528 SNPs genotyped via direct DNA sequencing of amplicons. Approximately 180 SNPs had been
529 projected to be required in order for equivalency in population and individual accuracy estimates
530 between the existing microsatellite and new SNP baselines (Beacham et al. 2012b). Hundreds of
531 SNPs were amplified in a single PCR in the current study, and direct DNA sequencing of the
532 resultant amplicons, coupled with automated scoring of the genotypes, resulted in cost-effective
533 genotyping. The increased number of SNPs available, coupled with higher minor allele
534 frequencies in the SNPs as compared with the microsatellites, resulted in an unprecedented
535 ability to provide accurate estimates of stock composition to very discrete geographic regions or
536 CUs.

537 The current study has demonstrated that it was possible to identify Chinook salmon
538 mixed-stock fishery contributions by CU, and also in many instances to specific populations
539 within a CU. A genetics-based assessment regime can accompany the mass marking of
540 hatchery-produced salmon, thereby facilitating the evaluation of hatchery contributions to
541 harvest. We have demonstrated that a genetics-based assessment system provides an opportunity
542 for conservation-based management of Canadian Chinook salmon (Beacham et al. 2021).
543 Similar results were obtained in coho salmon applications (Beacham et al. 2020) and can be
544 expected when applied to other salmonid species. A new era in salmonid stock identification is at
545 hand, powered by genotyping via direct DNA sequencing, which results in unprecedented high-
546 resolution stock identification estimation.

547

548 Acknowledgments

549 A very substantial effort was undertaken to obtain samples from Chinook salmon
550 sampled in this study. We thank various staff of the Salmon Enhancement Program of Fisheries
551 and Oceans Canada (DFO), including hatchery managers and staff for broodstock sample
552 collection and coordination. Baseline samples from the United States were obtained from the
553 National Marine Fisheries Service (NMFS), the Washington Department of Fish and Wildlife
554 (WDFW), the Alaska Department of Fish and Game (ADF&G), U.S. Fish and Wildlife Service,
555 Institute of Marine Biology, Russian Academy of Sciences, Far East Division staff, and the
556 Federal Research Institute of Fisheries and Oceanography (KamchatNIRO). Canadian samples
557 were obtained from DFO field staff and First Nations staff including the Kitasoo Fisheries
558 Program and the Gitxsan Watershed Authority. LGL Ltd. Environmental Research Associates
559 also provided samples. K. Flynn, and A. Duguid provided mixed-stock fishery DNA sample
560 extractions. Financial support was provided by DFO's "Results Fund" in support of project
561 RF121.

562

563 References

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

568 Beacham, T.D., Jonsen, K.L., Supernault, J., Wetklo, M., Deng, L., and Varnavskaya. N. 2006b.
569 Pacific Rim population structure of Chinook salmon as determined from microsatellite
570 variation. Transactions of the American Fisheries Society 135: 1604-1621.

571 Beacham, T.D., Jonsen, K,. and Wallace, C. 2012b. A comparison of stock and individual
572 identification for Chinook salmon in British Columbia provided by microsatellites (STRs)
573 and single nucleotide polymorphisms (SNPs). Mar. Coast. Fish.: Dynam, Manage, Eco.
574 Sci. 4: 1-22.

575 Beacham, T.D., Wallace, C.G., Le, K.D., and Beere, M. 2012a. Population structure and run
576 timing of steelhead trout in the Skeena River, British Columbia. N. Amer. J. Fish.
577 Manage. 32: 262-275.

578 Beacham, T.D., Wallace, C., MacConnachie, C., Jonsen, K., McIntosh, B., Candy, J.R., and
579 Withler, R.E. 2018. Population and individual identification of Chinook salmon in
580 British Columbia through parentage-based tagging and genetic stock identification with
581 single nucleotide polymorphisms. Can. J. Fish. Aquat. Sci. *75*(7): 1096-1105.
582 doi:10.1139/cjfas-2017-0168

583 Beacham, T.D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J.R., Willis, D., Lynch, C.,
584 Moore, J.-S., Bernatchez, L., amd Withler, R.E. 2020. Accurate estimation of
585 Conservation Unit contribution to coho salmon mixed-stock fisheries in British
586 Columbia, Canada using direct DNA sequencing for single nucleotide polymorphisms.
587 Can. J. Fish. Aquat. Sci. 77: 1302-1315.

588 Beacham, T.D., Wallace, C.G., Jonsen, K., McIntosh, B., Candy, J,R,, Horst, K. Lynch, C.
589 Luedke, W., Keary, L., and Rondeau, E.B. 2021. Parentage-based tagging and genetic
590 stock identification is a cost-effective and viable replacement for coded-wire tagging in
591 large-scale assessments of marine Chinook salmon fisheries in British Columbia, Canada.
592 Evol. Appl. In press

593 Beacham, T.D., Withler, R.E., and Stevens, T.A. 1996. Stock identification of chinook salmon
594 (*Oncorhynchus tshawytscha*) using minisatellite DNA variation. Can. J. Fish.Aquat. Sci.
595 53: 380-394.

596 Christensen, K.A., Leong, J.S., Sakhrani, D., Biagi, C.A., Minkley, D.R., Withler, R.E.,
597 Rondeau, E,B., Koop, B.F., and Devlin, R.H. 2018. Chinook salmon (Oncorhynchus
598 tshawytscha) genome and transcriptome. PLOS ONE 13(4): e0195461.
599 https://doi.org/10.1371/journal.pone.0195461

600 COSEWIC, 2018. COSEWIC assessment and status report on the Chinook Salmon
601 Oncorhynchus tshawytscha, Designatable Units in Southern British Columbia (Part One
602 – Designatable Units with no or low levels of artificial releases in the last 12 years), in
603 Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. xxxi + 283
604 pp. (http://www.registrelepsararegistry.gc.ca/default.asp?lang=en&n=24F7211B-1).

605 Fisheries and Oceans Canada (DFO). 2005. Canada's policy for conservation of wild Pacific
606 salmon. 57p. Fisheries and Oceans Canada, 401 Burrard Street, Vancouver, BC, V6C
607 3S4.

608 Fournier, D.A., Beacham, T.D., Riddell, B.E., and Busack, C.A. 1984. Estimating stock
609 composition in mixed stock fisheries using morphometric, meristic, and electrophoretic
610 characteristics. Can. J. Fish. Aquat. Sci. **41**: 400–408.

611 Gilbey, J., Cauwelier, E., Coulson, M.W., Stradmeyer, L. Sampayo, J.N., Armstrong, A.,
612 Verspoor, E., Corrigan, L., Shelley, J., and Middlemas, S. 2016. Accuracy of
613 assignment of Atlantic Salmon (*Salmo salar* L.) to rivers and regions in Scotland and
614 northeast England based on single nucleotide polymorphism (SNP) markers. PLoS One.
615 2016; 11(10): e0164327

Hess, J.E., Ackerman, M.W., Fryer, J.K., Hasselman, D.J., Steele, C.A. , Stephenson, J.J. , Whiteaker, J.M., and Narum, S.R. 2016. Differential adult migration-timing and stock-specific abundance of steelhead in mixed stock assemblages. ICES Journal of Marine Science 73, 2606-2615. doi:10.1093/icesjms/fsw138

Jefferts, K.B., Bergmann, P.K., and Fiscus, H.F. 1963. A coded wire tag identification system for macro-organisms. Nature (London), 198, 460-462.

Jombart T. and Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 1 Nov 1;27(21):3070-3071. doi: 10.1093/bioinformatics/btr521

Kahle, D. and Wickham, H. 2013. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Larson, W.A., Seeb, J.E., Pascal, C.E., Templin, W.D., and Seeb, L.W. 2014. Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. Can. J. Fish. Aquat. Sci, 71: 698-708.

Le Luyer J., Laporte, M., Beacham, T.D., Kaukinen, K.H., Withler, R.E., Koop, B., and Bernatchez, L. 2017. Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. Proc. Nat. Acad. Sci. (USA), 114: 12964-12969. DOI10.1073/pnas.1711229114

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN]

637 Milner, G.B., Teel, D.J., Utter, F.M., and Winans, G,A. 1985. A genetic method of stock
638 identification in mixed populations of Pacific salmon, Oncorhynchus spp. Mar. Fish. Rev.
639 47:1–8.

640 Moran, B.M, and Anderson, E.C. 2019. Bayesian inference from the conditional genetic stock
641 identification model. Can. J. Fish. Aquat. Sci. 76: 551-560.

642 Moran, P., Dazey, J., LaVoy, L., and Young, S. 2018. Genetic mixture analysis supports
643 recalibration of the fishery regulation assessment model. Fisheries, 43, 83-97.

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

646 Paradis E. and Schliep K. 2018. ape 5.0: an environment for modern phylogenetics
647 and evolutionary analyses in R. Bioinformatics 35: 526-528.

648 Price, M.H.H, Rosenberger,A.G.J., Taylor, G.G., and Stanford, J.A. 2014. Comment on
649 "Population Structure and Run Timing of Sockeye Salmon in the Skeena River, BC". N.
650 Am. J. Fish. Manage., 34: 1167-1170.

651 Price, M.H.H, English, K.K., Rosenberger, A.G., MacDuffee, M., and Reynolds, J.D. 2017.
652 Canada's Wild Salmon Policy: an assessment of conservation progress in British
653 Columbia. Can. J. Fish. Aquat. Sci., 74: 1507-1518.

654 Quinlan A.R. and Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic
655 features. Bioinformatics. 26(6): 841–842.

656 R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for
657 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Seeb, L.W., and Crane, P.A. 1999. High genetic heterogeneity in chum salmon in Western Alaska, the contact zone between northern and southern lineages. Trans. Am. Fish. Soc. 128: 58-87.

Seeb, L.W., Habicht, C., Templin, W.D., Tarbox, K.E., Davis, R.Z., Brannian, L.K., and Seeb, J.E. 2000. Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the *Exxon Valdez* oil spill. Trans. Am. Fish. Soc. 129: 1223-1249.

Shaklee, J.B., Beacham, T.D., Seeb, L., and White, B.A. 1999. Managing fisheries using genetic data: Case studies from four species of Pacific salmon. Fish. Res. 43: 45-78.

Starks, H.A., Clemento, A.J., Garza, J.C. 2016 Discovery and characterization of single nucleotide polymorphisms in coho salmon, *Oncorhynchus kisutch*. Mol Ecol Res 16: 277-287.

Steele, C.A., Hess, M., Narum, S., Campbell, M. 2019. Parentage-based tagging: Reviewing the implementation of a new tool for an old problem. Fisheries 44: 412-422. https://doi.org/10.1002/fsh.10260

Utter, F.M., Teel, D.J., Milner, G.B., and McIsaac, D. 1987. Genetic estimates of stock compositions of 1983 chinook salmon harvests off the Washington coast and the Columbia River. Fish. Bull. **85**: 13–23.

Table 1. Estimated percentage stock composition of simulated mixed-stock samples of Chinook salmon (n=200) as may be encountered in northern and southern British Columbia (BC). The expected CU/Regional compositions were obtained by adding the true population components for each CU/Region. Standard deviation is in parentheses.

CU/Region		Population	True	Population	CU/Region
		Northern BC			
Southeast Alaska		Situk River	15	15.0 (2.6)	15.0 (2.6)
	61	Verrett River	5	4.7 (1.8)	4.9 (1.6)
	43	Yakoun River	2	2.2 (1.0)	2.2 (1.0)
	58	Meziadin River	19	16.9 (2.3)	17.9 (2.4)
	57	Kincolith River	3	2.8 (1.3)	2.9 (1.3)
	54	Suskwa River	16	14.6 (2.6)	
					27.5 (3.7)
	54	Bulkley River	14	11.8 (2.4)	
	50	Kitsumkalum River	7	7.5 (2.1)	7.5 (2.1)
	39	Atnarko River	14	13.0 (3.5)	13.6 (2.6)
	40	Dean River	5	4.6 (1.6)	4.6 (1.6)
		Southern BC			
	32	Conuma River	14	13.5 (2.7)	13.7 (2.7)

31	Robertson Creek	12	11.9 (2.3)	12.0 (2.3)
29	Quinsam River	16	15.9 (2.3)	15.9 (2.3)
27	Puntledge River (fall)	14	12.9 (2.5)	13.2 (2.5)
10	Cariboo River	3	2.9 (1.3)	3.0 (1.3)
17	Nicola River	1	1.0 (0.7)	1.1 (0.7)
North Puget Sound	Skagit River	15	14.6 (2.5)	15.1 (2.5)
Coastal Washington	Hoh River	9	8.8 (2.0)	8.9 (2.0)
Snake River spring/summer	Marsh Creek	6	5.7 (1.7)	5.8 (1.7)
Northern and central coastal Oregon	Siuslaw River	10	10.0 (2.1)	10.1 (2.1)

Table 2. Accuracy of estimated population-specific and Conservation Unit (CU)-specific stock composition (percent) for known-origin individuals sampled in 2017-2019 fisheries and escapements in BC and estimated with only genetic stock identification (GSI). Known-origin was derived from coded-wire tags (CWT), escapement location, or parentage-based tagging (PBT) identifications. Standard deviation is in parentheses.

Population			CU		
Name	Actual	Estimated	Name (CU number)	Actual	Estimated
2017 CWT n=1,309					
Robertson	20.6	20.6 (1.1)	SWVI (31)	20.6	20.6 (1.1)
Quinsam	38.8	38.8 (1.2)	NEVI (29)	38.9	38.8 (1.2)
Puntledge summer	6.2	6.7 (0.6)	EVIGStr-sum (83)	6.2	6.7 (0.7)
Puntledge fall	17.2	13.4 (1.2)	QP fall (27)	34.3	33.8 (1.3)
Big Qualicum	17.2	20.3 (1.3)			
2018 Campbell River escapement n=90					
Quinsam	100.0	100.0 (0.9)	NEVI (29)	100.0	100.0 (0.9)
2018 fishery PBT identifications n=849					
Kitsumkalum	0.6	0.6 (0.3)	Kalum (50)	0.6	0.6 (0.3)
Atnarko	0.1	0.1 (0.1)	BC-Bent (39)	0.1	0.1 (0.1)
Shuswap	2.7	2.7 (0.4)	STh-SHU (15)	2.7	2.7 (0.7)
Chilliwack	27.8	25.8 (1.3)	LFR-fall (3)	29.0	29.0 (1.4)
Harrison	1.2	3.2 (0.9)			
Quinsam	5.8	5.9 (0.8)	NEVI (29)	5.8	5.9 (0.8)
Puntledge summer	0.8	0.5 (0.3)	EVIGStr-sum (83)	0.8	0.5 (0.3)

Puntledge fall	7.5	6.5 (1.0)	QP fall (27)	13.7	13.2 (1.4)
Big Qualicum	6.1	6.6 (0.9)			
Cowichan	4.2	4.8 (0.7)	CWCH-KOK (22)	4.2	4.8 (0.7)
Burman	0.4	0.0 (0.0)	NoKy (32)	1.2	1.2 (0.4)
Conuma	0.6	1.2 (0.4)			
Leiner	0.1	0.0 (0.0)			
Tlupana	0.1	0.0 (0.0)			
Nahmint	0.4	0.6 (0.3)	SWVI (31)	41.9	42.0 (1.6)
Nitinat	0.8	0.8 (0.4)			
Robertson	38.5	36.0 (1.4)			
Sarita	1.1	1.1 (0.4)			
Thornton	1.2	1.4 (0.4)			
2019 fishery PBT identifications n=2,295					
Kitsumkalum	0.7	0.7 (0.2)	Kalum (50)	0.7	0.7 (0.2)
Kitimat	0.2	0.2 (0.1)	NCC-stream (42)	0.2	0.2 (0.1)
Atnarko	2.9	2.9 (0.4)	BC-Bent (39)	2.9	2.9 (0.4)
Phillips	0.3	0.3 (0.1)	SMn-SFj (28)	0.3	0.3 (0.1)
Squamish	0.5	0.5 (0.1)	SMn-GStr (20)	0.5	0.5 (0.1)
Shuswap	0.9	0.9 (0.2)	STh-SHU (15)	0.9	0.9 (0.2)
Chilliwack	18.3	18.0 (0.9)	LFR-fall (3)	19.3	19.3 (0.9)
Harrison	1.0	1.3 (0.3)			
Quinsam	5.5	5.5 (0.5)	NEVI (29)	5.5	5.5 (0.5)

Puntledge summer	0.3	0.6 (0.2)	EVIGStr-sum (83)	0.3	0.6 (0.2)
Puntledge fall	11.6	9.8 (0.7)	QP fall (27)	15.9	15.3 (0.8)
Big Qualicum	4.3	2.8 (0.6)			
Cowichan	1.6	1.6 (0.3)	CWCH-KOK (22)	1.6	1.6 (0.3)
Burman	3.1	2.6 (0.4)	NoKy (32)	3.6	3.6 (0.4)
Conuma	0.3	1.0 (0.3)			
Leiner	0.2	0.0 (0.0)			
Bedwell	0.1	0.0 (0.0)	SWVI (31)	48.4	48.4 (1.1)
Nahmint	0.4	0.5 (0.2)			
Nitinat	0.9	0.9 (0.2)			
Robertson	44.0	43.5 (1.0)			
Sarita	2.1	1.9 (0.4)			
Thornton	1.0	1.2 (0.2)			
2019 broodstock PBT identifications, n=4,199					
Kitimat	2.9	2.9 (0.2)	NCC-stream (42)	2.9	2.9 (0.2)
Atnarko	6.0	6.0 (0.5)	BC-Bent (39)	6.0	6.0 (0.3)
Phillips	0.3	0.3 (0.1)	SMn-SFj (28)	0.3	0.3 (0.1)
Shuswap	0.4	0.4 (0.1)	STh-SHU (15)	0.4	0.4 (0.1)
Coldwater	0.3	0.3 (0.1)	Lower Thompson (17)	1.7	1.7 (0.2)
Nicola	0.9	1.1 (0.2)			
Spius	0.6	0.3 (0.1)			
Chilliwack	12.9	12.9 (0.5)	LFR-fall (3)	12.9	12.9 (0.5)

Quinsam	18.4	18.4 (0.6)	NEVI (29)	18.4	18.4 (0.6)
Puntledge summer	2.7	2.9 (0.3)	EVIGStr-sum (83)	2.7	2.9 (0.3)
Puntledge fall	15.6	12.1 (0.6)	QP fall (27)	27.0	26.2 (0.7)
Big Qualicum	11.4	8.8 (0.8)			
Cowichan	0.5	0.6 (0.1)	CWCH-KOK (22)	0.5	0.6 (0.1)
Nahmint	0.5	0.6 (0.1)			
Nitinat	1.0	1.3 (0.2)			
Robertson	19.7	20.0 (0.7)	SWVI (31)	27.2	27.4 (0.7)
Sarita	5.0	2.6 (0.3)			
Thornton	1.0	1.0 (0.2)			

Table 3. Percentage stock composition by geographic region or CU of 2019 terminal commercial fisheries (Area 8, 23, and 25) and Strait of Georgia (north), Juan de Fuca Strait, and west coast Vancouver Island sport fisheries. Standard deviation is in parentheses.

Conservation Unit	Area 8 Gillnet	Area 23 Net	Area 25 Net	SoG north sport	JDF sport	WCVI south sport
	July-August	August	August	August	August	August
Sample size	180	122	197	690	263	913
Upper Nass	1.1 (0.7)	0.0 (0.0)	0.0 (0.2)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Bella Coola-Bentinck	90.5 (2.2)	0.0 (0.1)	0.0 (0.0)	0.1 (0.1)	0.0 (0.1)	0.0 (0.0)
Dean River	8.3 (2.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Southern Mainland-Southern Fjords_FA_0.x	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.2 (0.2)	0.0 (0.1)	0.0 (0.0)
Southern Mainland-Georgia Strait_FA_0.x	0.0 (0.0)	0.0 (0.1)	0.0 (0.1)	0.7 (0.3)	0.1 (0.3)	0.0 (0.0)
Upper Fraser River_SP_1.3	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.1 (0.2)	0.0 (0.1)	0.0 (0.0)
Middle Fraser River_SU_1.3	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.1 (0.1)
Middle Fraser River-Portage_FA_1.3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.4 (0.3)	0.0 (0.0)
North Thompson_SP_1.3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.3)	0.0 (0.0)
North Thompson_SU_1.3	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.7 (0.5)	0.1 (0.1)
Shuswap River_SU_0.3	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	2.2 (0.6)	0.8 (0.6)	0.9 (0.3)
South Thompson_SU_0.3	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	17.9 (1.9)	64.9 (3.0)	16.8 (1.2)
Lower Fraser River_FA_0.3	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	31.3 (1.8)	2.3 (1.0)	1.1 (0.4)
East Vancouver Island-North_FA_0.x	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	2.2 (0.7)	0.0 (0.0)	0.2 (0.2)

East Vancouver Island-Qualicum and Puntledge_FA_0.x	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	24.6 (1.9)	0.2 (0.3)	0.6 (0.4)
East Vancouver Island-Georgia Strait_SU_0.3	0.0 (0.3)	0.0 (0.0)	0.0 (0.0)	0.3 (0.2)	0.0 (0.0)	0.0 (0.0)
East Vancouver Island-Cowichan and Koksilah_FA_0.x	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	13.3 (1.2)	0.1 (0.2)	0.8 (0.3)
West Vancouver Island-Nootka and Kyuquot_FA_0.x	0.0 (0.1)	0.0 (0.1)	99.8 (0.6)	0.0 (0.0)	0.0 (0.1)	3.4 (0.6)
West Vancouver Island-South_FA_0.x	0.0 (0.1)	100.0 (0.8)	0.2 (0.5)	0.3 (0.2)	7.6 (1.6)	67.3 (1.7)
Coastal Washington	0.0 (0.2)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.4 (0.3)	0.1 (0.1)
North Puget Sound	0.0 (0.0)	0.0 (0.4)	0.0 (0.0)	1.6 (0.5)	4.6 (1.2)	1.8 (0.5)
South Puget Sound	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	4.2 (0.7)	16.3 (2.3)	2.9 (0.6)
Lower Columbia River	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.1 (0.2)	0.4 (0.4)	1.2 (0.3)
Mid Columbia River_SP	0.0 (0.0)	0.0 (0.2)	0.0 (0.1)	0.5 (0.4)	0.0 (0.0)	0.0 (0.0)
Upper Columbia River_SU_FA	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	1.1 (0.7)	0.8 (0.3)
Snake River_FA	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.8 (0.3)
North & Central Oregon	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.1)
South Oregon coastal	0.0 (0.0)	0.0 (0.0)	0.0 (0.2)	0.0 (0.0)	0.0 (0.0)	0.3 (0.2)
California Klamath Trinity	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.1)
California Central Valley_Fall	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.2 (0.1)

18

19

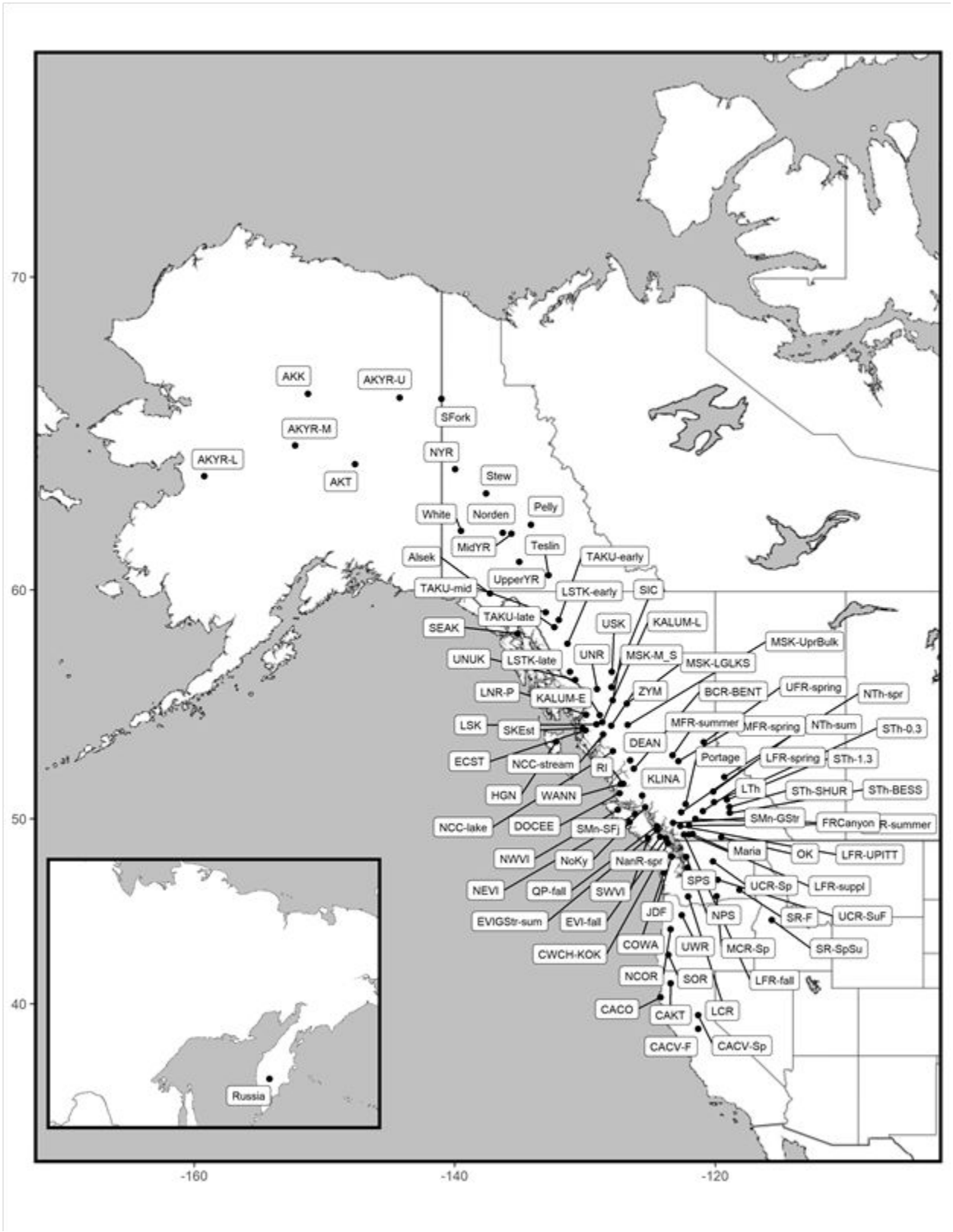
List of Figures

Figure 1. Distribution of reporting units represented in the Chinook baseline. Position displayed is the mean latitude and longitude of all collections within the reporting unit in this study. Base map and data from OpenStreetMap and OpenStreetMap Foundation, and plotted in R using the ggmap package (Kahle and Wickam, 2013). The specific populations sampled in each region or CU are outlined in Supplementary Table S1.

Figure 2. For simulated single-population samples (sample comprised 100% of a single population), average percent accuracy for estimated stock compositions of the sample to population (open portion of bar) and to CU (shaded portion of bar) for all populations in a CU or reporting unit for which the single population was member, with the number of populations in the CU reported to the right of the bar. The shaded portion of the bar is the sum of the population in question plus any other allocation to any population in the same CU or reporting unit. For example, the average assignment success for the 13 Russian populations was 49.2% to population and 99.4% to reporting unit. The 90% accuracy level is indicated in the figure.

Figure 3. Average percent accuracy for self assignment of individual Chinook salmon ("leave one out") to population (open portion of bar) and to CU (shaded portion of bar) for all populations in a CU or reporting unit, with the number of populations in the CU reported to the right of the bar. The shaded portion of the bar is the sum of the population in question plus any other allocation to any population in the same CU or reporting unit. For example, the average assignment success for the 13 Russian populations was 41.0% to population and 99.8% to reporting unit. The 80% accuracy level is indicated in the figure.

23 Figure 1



24

Figure 2

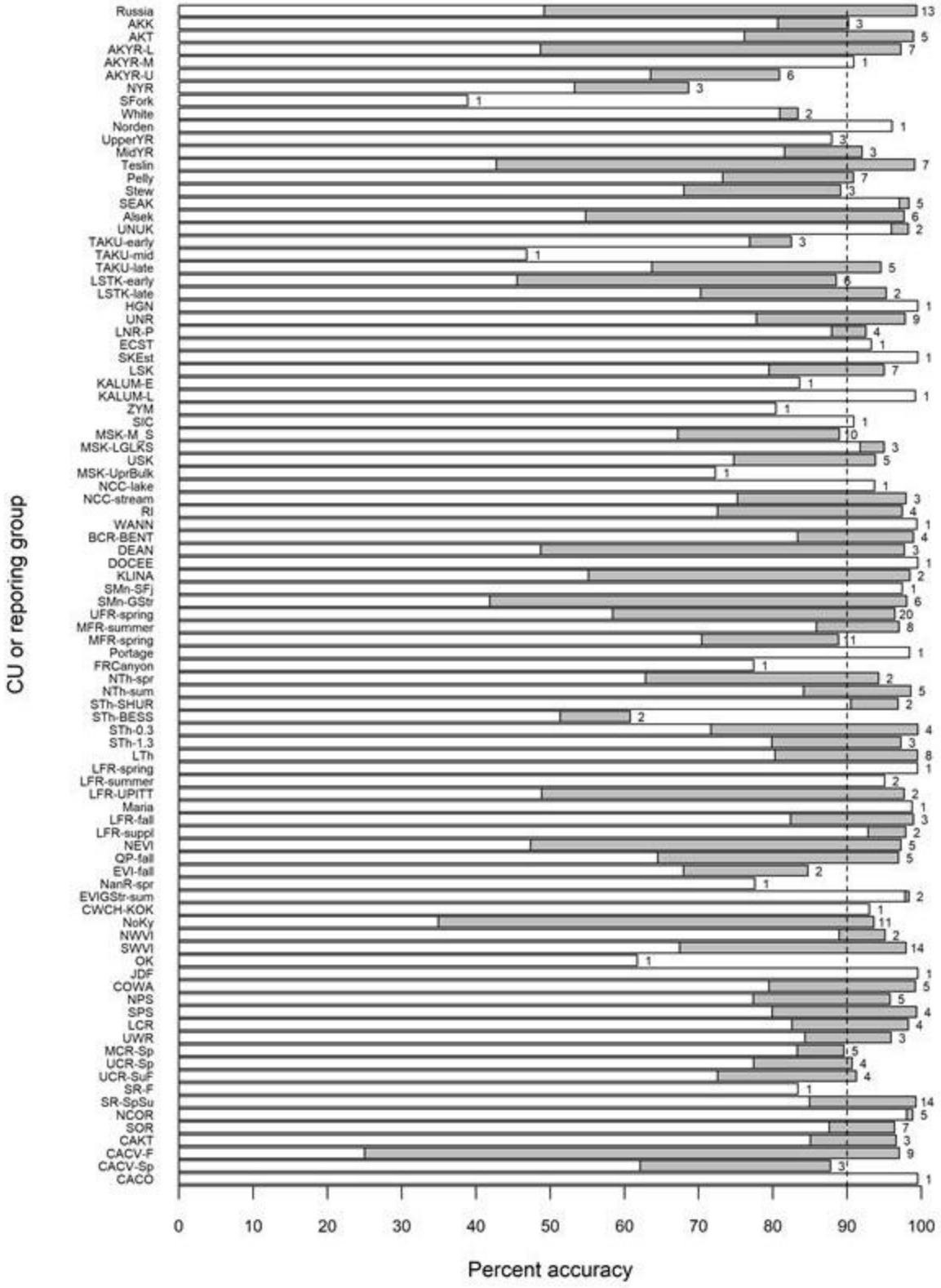


Figure 3