NRC Research Press

ARTICLE

Accurate estimation of conservation unit contribution to coho salmon mixed-stock fisheries in British Columbia, Canada, using direct DNA sequencing for single nucleotide polymorphisms

Terry D. Beacham, Colin Wallace, Kim Jonsen, Brenda McIntosh, John R. Candy, Eric B. Rondeau, Jean-Sébastien Moore, Louis Bernatchez, and Ruth E. Withler

Abstract: Determination of population structure and stock identification is a ubiquitous 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 coho salmon (*Oncorhynchus kisutch*), a stock identification baseline composed of some 57 982 individuals from 332 populations ranging from southeast Russia to California was employed for genetic stock identification (GSI). GSI analysis based upon variation at up to 480 single nucleotide polymorphisms (SNPs) was demonstrated to provide accurate estimates of stock composition for 37 conservation units (CU) in British Columbia, 13 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 coho salmon.

Résumé : La détermination de la structure des populations et l'identification au stock sont des problèmes généralisés en évaluation et en gestion des pêches. Les régimes de gestion des pêches aux saumons du pacifique évoluent pour exiger une plus haute résolution de la composition des stocks dans des unités de déclaration de plus en plus petites. Pour le saumon coho (*Oncorhynchus kisutch*), une base de référence pour l'identification au stock comprenant quelque 57 982 individus provenant de 332 populations allant du sud-est de la Russie à la Californie a été utilisée pour l'identification génétique au stock (IGS). Il est démontré que l'analyse IGS basée sur les variations de jusqu'à 480 polymorphismes mononucléotidiques (SNP) fournit des estimations exactes de la composition des stocks pour 37 unités de conservation (UC) en Colombie-Britannique, 13 groupes de déclaration aux États-Unis et un groupe de déclaration en Russie. Dans de nombreux cas, des estimations propres à la population exactes de la composition des stocks au sein d'une UC sont possibles pour des échantillons tirés de la pêche, ainsi que l'affectation des individus à certaines populations précises. Un système d'évaluation reposant sur la génétique offre la possibilité d'une gestion basée sur la conservation des saumons cohos canadiens. [Traduit par la Rédaction]

Introduction

The general objective of genetic stock identification (GSI) for Pacific salmon is to provide the optimal resolution among stocks or populations present when applied to mixed-stock fishery and forensic samples at an affordable cost (Beacham et al. 2009). Stock identification of Pacific salmon in mixed-stock fisheries is important to enable 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). Stock composition information is also important in determining locations of ocean residence of specific stocks of immature salmon (Farley et al. 2011) and the migration routes used by immature salmon to reach seasonal rearing areas (Beacham et al. 2014), as well as the routes used by maturing salmon to return to natal rivers.

GSI based on DNA variation has been widely applied in the assessment of mixed-stock Pacific salmon fisheries. For coho salmon (*Oncorhynchus kisutch*), microsatellites have been used for many years to evaluate population structure or estimate stock composition in mixed-stock fisheries (Small et al. 1998*a*, 1998*b*; Beacham et al. 2001, 2012*a*; Smith et al. 2001; Ford et al. 2004). Population structure is important to evaluate for stock identification applications. If there is a regional basis to population structure, individuals in the mixture sample from populations not in the baseline used for stock composition estimation will generally be assigned to other sampled populations from the same region (Beacham et al. 2012*a*). Microsatellites also provide the basis for estimation of stock composition of juvenile coho salmon sampled off the coasts of Washington and Oregon (Van Doornik et al. 2007), as well as off coastal British Columbia (BC) (Beacham et al. 2016). Recent studies in salmonids have indicated that incorporation of several hundred or thousands of SNPs in GSI applications can improve stock assignment accuracy (Larson et al. 2014; Moore et al. 2014).

Direct DNA sequencing is powering a revolution in the application of genetics to fisheries management and assessment, providing cost-effective genotyping at single nucleotide polymorphism (SNP) loci (Campbell et al. 2015) or microsatellites (Bradbury et al.

Received 3 October 2019. Accepted 22 March 2020.

T.D. Beacham, C. Wallace, K. Jonsen, B. McIntosh, J.R. Candy, E.B. Rondeau, and R.E. Withler. Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada.
J.-S. Moore and L. Bernatchez. Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC G1V OA6, Canada.

Corresponding author: Terry D. Beacham (email: terry.beacham@dfo-mpo.gc.ca).

2018). Incorporating this new technology, Beacham et al. (2019*a*) provided a brief summary of current assessment techniques for coho salmon in BC based upon coded-wire tags (CWTs) and concluded that a combined GSI and parentage-based tagging (PBT) approach can provide critical information to improve coho salmon assessment and conservation. 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). Assignments are made to parents of known origin, and with that information, it is possible to determine the origin and age of individuals sampled in fisheries. Application of a PBT–GSI system of identification of coho salmon in fisheries and escapements in BC provided high-resolution estimates of stock composition, catch, and exploitation rate by conservation unit (CU) or population, providing an alternate and more effective method in the assessment and management of Canadian-origin coho salmon relative to CWTs (Beacham et al. 2019*a*). In addition, PBT can determine population- and family-specific distributions among fisheries, origins of hatchery brood stocks and associated stray rates among populations, and productivity of specific components of some hatchery brood stocks (Beacham et al. 2019*b*).

GSI and PBT applied in combination can provide high-resolution estimates of stock composition, as assignment of individual coho salmon via PBT is virtually 100% accurate with respect to hatchery of origin and age of the individual (Beacham et al. 2017, 2019*a*). However, PBT has not been applied to wild coho salmon populations to date 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. Although initial analyses suggested that accurate fine geographic scale estimates of stock composition were available through GSI for some coho salmon populations in southern BC (Beacham et al. 2019*a*), demonstration of the general applicability of the accuracy of GSI-derived stock composition estimates for wild and hatchery populations throughout BC would be desirable.

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). Fisheries and hatchery production were to be managed to ensure that wild populations were safeguarded and harvest benefits sustainable. As a cornerstone of the WSP, wild salmon populations are identified and maintained in CUs that reflect their geographic, ecological, and genetic diversity. The 43 coho salmon CUs originally identified by Holtby and Ciruna (2007) have been modified to a current number of 44 CUs (Fisheries and Oceans Canada 2019). Price et al. (2014, 2017) suggested that any suitable assessment technique must provide individual resolution for all CUs to meet the conservation requirements of Canada's WSP. Accordingly, then, this implies that GSI should provide accurate estimates of stock composition by CU where the majority of the individuals captured in fisheries are wild in origin. There will be limited opportunity to enhance GSI-derived estimates of wild-origin coho salmon via PBT analysis because of the difficulty and expense of complete adult sampling in the wild. Resolution of mixed-stock fishery stock compositions in BC to the CU level would be unprecedented, as GSI has never been applied to provide such fine-scale resolution of stock composition in BC. Identification of individual populations within CUs adds to the difficulty of the task, as some mixed-stock fishery applications may depend on accurate population identification, as well as escapement surveys to estimate fishery-specific population exploitation rates (Beacham et al. 2019*a*). Evaluation of the ability to assess accurately the stock composition of mixed-stock coho salmon fishery samples by CUs and population within CU was the key objective of the current study.

The current study is an evaluation of the application of the GSI methodology initially outlined by Beacham et al. (2017) to determine whether GSI can be used to provide information on fishery contributions by CU for coho salmon CUs currently identified under the WSP. Ampliseq (Thermo Fisher Scientific) was used to amplify hundreds of SNPs in 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 coho salmon, similar procedures could be used for other salmonid and nonsalmonid species. A stock identification baseline composed of some 57 982 individuals from 332 populations ranging from southeast Russia to California was employed for GSI. Population structure of all populations in the baseline was evaluated, and genotypes from each population in a CU or geographic region were simulated and estimated stock composition of the single-population and multipopulation simulated mixtures determined via GSI referencing the 332-population baseline. The baseline was subsequently used to estimate stock composition in a putatively known-origin sample and series of 2018 fisheries in BC.

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 to evaluate the effectiveness of the baseline in producing reliable estimates of stock composition in mixed-stock fishery samples. The initial step included determination of whether population structure was geographically or regionally based, aiding in definition of reporting units. 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. 2012*b*). 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). Although not mandatory for evaluation of the reliability of stock composition estimates, we also evaluated the effect of baseline population sample size for populations with varying levels of genetic distinction as measured by mean pairwise population F_{ST}. We next evaluated the baseline by simulating two mixed-stock fishery samples composed of eight 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 a sample of known origin as represented by 884 individuals recovered from fishery sampling that had been marked with CWTs. 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.

Baseline sample collection

The initial baseline was outlined by Beacham et al. (2017) and consisted of 20 242 individuals from 117 populations, with the distribution of populations ranging from southeast Alaska to Puget Sound in Washington State. Beacham et al. (2019*a*) reported that the baseline was subsequently expanded to include 40 774 individuals from 267 populations, ranging from southeast Alaska to Oregon. The primary expansion of the baseline included a survey of additional populations in southern BC, coastal Washington, the Columbia River drainage, and Oregon. The baseline in the current study was expanded again to include 57 982 individuals from 332 populations, ranging from Russia to California. Prior to 2014, most samples were collected opportunistically to provide a baseline for the previous microsatellite analyses concerning population structure and stock identification (Beacham et al. 2011*b*, 2012*a*). From 2014 onwards, PBT was the objective of population sampling for selected hatcheries in BC, and samples were collected to allow complete genotyping of the brood stock in a particular year. Fin tissue or operculum punches were obtained from all individuals sampled.

Fishery sample collection

General fishery sampling procedures were described by Beacham et al. (2019*a*). Summarized briefly, in the northern BC troll fishery, samplers electronically checked adipose fin-clipped individuals for the presence of a CWT, and heads from those individuals containing a CWT were subsequently sent to the central laboratory for CWT recovery and tissue sampling for subsequent genotyping. Tissue samples from clipped individuals with no CWTs were directly provided for genotyping, as were samples from unclipped individuals. The origin of the samples from the recreational fishery in BC included voluntary head recoveries of adipose fin-clipped coho salmon from recreational fisheries in southern BC and direct creel sampling in some northern recreational fisheries. Samples from the recreational fishery in southern BC were derived from clipped individuals, but they may not have been marked with a CWT when delivered to the CWT head recovery laboratory.

We tested accuracy of estimated stock composition for mixed-stock fishery samples by first estimating stock composition of a known-origin sample. If a CWT (Jefferts et al. 1963) was detected in 2018 sampling of commercial fisheries, the head was sent to a central CWT head recovery laboratory in Vancouver, BC, where the DNA sample was subsequently taken. A CWT provided a putative known origin of the individual, as the number on the tag is recorded prior to release of the individual from a hatchery or during smolt migration for a wild population. Heads of adipose fin-clipped individuals from 2018 recreational fisheries were also sampled for CWTs. If a CWT was successfully recovered and decoded from an individual head from a fishery, and a genotype was successfully obtained for the individual sample, then the genotypes of all of these individuals were pooled into a single mixed-stock sample of known origin to evaluate accuracy of stock compositions by CU and population. The known-origin sample was obtained by genotyping 884 coho salmon from 2018 fisheries that were marked with CWTs.

We tested the baseline used in stock composition estimation by analyzing eight 2018 mixed-stock samples of coho salmon with divergent geographic origins. Actual fishery samples are valuable to investigate as performance of the baseline in estimation of stock composition can be evaluated relative to expected stock compositions in the fishery. The geographic distribution of these samples, spanning sites from the northern BC to a freshwater fishery in the Fraser River, suggested that divergent estimates of stock composition should be obtained when analyzed with the baseline evaluated in the study. Geographic locations for most of the fisheries were outlined by Beacham et al. (2019*a*).

Genotyping

The detailed procedures for DNA extraction, library preparation, and genotyping were described by Beacham et al. (2017), and a summarized version was provided by Beacham et al. (2019*a*). The process involved loading amplified DNA from 768 individuals (up to 482 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 a single run of the Ion Chef, both chips were then subsequently loaded on to an Ion Torrent Proton sequencer, and the genotype of each individual was recorded with automated scoring of the genotype via Proton software Variant Caller at one SNP site in each amplicon. Other than the target SNP, additional sequence variation within the same amplicon was not incorporated in the analysis. Genotypes at all available SNPs for an individual were assembled to provide a single multilocus individual genotype for each individual fish. This multilocus genotype was the basic input for subsequent analyses. The species identification SNP *OkiOts_120255-113* and sex identification SNP *Ots_SEXY3-1* were omitted from subsequent GSI analyses, leaving up to 480 SNPs included. The species identification SNP was omitted as it was monomorphic if only coho salmon were sampled and thus of no value for GSI analyses. It was, however, of considerable value in eliminating any non-coho salmon that may have been included in mixed-stock fishery samples. The sex identification SNP was eliminated because the genotype was either monomorphic (male) or "no call" (female) and of no value for GSI analyses. The 480 remaining SNPs were derived from a previous SNP panel outlined by Beacham et al. (2017), supplemented with 209 SNPs derived from the Genome Canada project Enhancing Production in Coho: Culture, Community, Catch (EPIC4).

The baseline

A listing of populations in the baseline that has been genotyped, along with the corresponding CU (Canadian populations) or geographic region (Russian and American populations) is outlined in the online Supplementary Table 1[1]. Summarized briefly, the baseline survey consisted of genotyping coho salmon in populations from Russia, Alaska, BC, Washington, Oregon, and California, with locations of Canadian CUs outlined in Fig. 1. The locations for some of the populations listed in Supplementary Table 1[1] in each geographic region were illustrated by Beacham et al. (2017), with most of the samples collected subsequent to 1990.

Data analysis

Genotypes had to be obtained for at least 150 SNPs for the individual to be retained in the baseline. In a test where the DNA of the same 382 individuals was genotyped on two occasions, a mean genotyping error rate of 1.07% (1220 discrepancies in 114 105 comparisons) or an allele error rate of 0.53% (1220 discrepancies in 228 210 comparisons) was observed over the 302 SNPs scored (Beacham et al. 2017). Expected and observed heterozygosities by locus were determined with adegenet (Jombart and Ahmed 2011). Population heterozygosity was determined as the average heterozygosity over all loci. Estimation of F_{ST} by locus was conducted with ape (Paradis and Schliep 2018). Amplicon sequences were aligned to the genome (RefSeq assembly accession GCF_002021735.1) with BWA mem version 0.7.17-r1198 (Li 2013) and filtered using filter_sam_file.py in snp-placer (commit 8bd5e72; https://github.com/CNuge/snp-placer). F_{ST} by SNP was plotted in R version 3.6.0 (R Core Team 2019) by position along the chromosomes, except for SNPs aligned to unplaced scaf-

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2019-0339.

Fig. 1. Boundaries of conservation units defined for coho salmon in British Columbia. The map was prepared with ArcGIS with the shape files located at Open Government at https://search.open.canada.ca/en/od/?sort=last_modified_tdt%20desc&page=1&od-search-portal=Open%20Data&search_text=NuSEDS. [Colour online.]

folds, which were plotted based on the average marker spacing observed across the genome. For determination of population structure, F_{ST} distance was used to estimate genetic distances among all populations via StAMPP (Pembleton et al. 2013), with up to 480 SNPs incorporated in the analyses. An unrooted neighbor-joining tree based upon F_{ST} distance was generated using phangorn 2.5.5 (Schliep 2011). Bootstrap support for the major nodes in the tree was evaluated based upon 1000 replicate trees.

To test the accuracy of identifying the CU and the population of origin, we performed GSI using Rubias (Moran and Anderson

Fig. 2. Distribution of observed heterozygosity for 480 single nucleotide polymorphisms (SNPs) surveyed in coho salmon populations ranging from Russia to California.

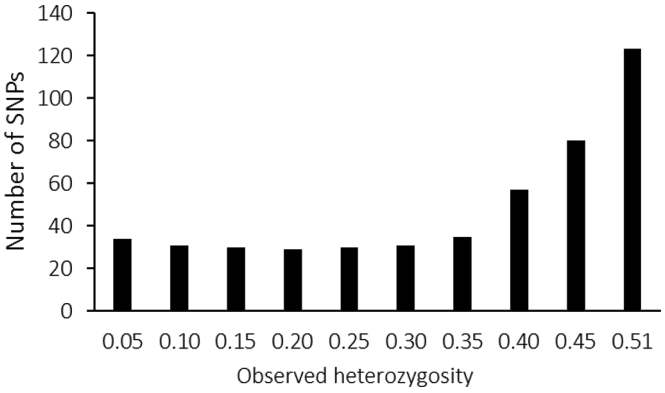
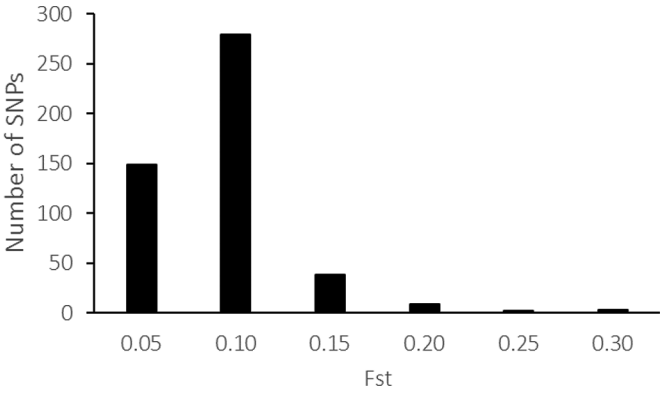


Fig. 3. Distribution of F_{ST} for 480 SNPs surveyed in coho salmon populations ranging from Russia to California.



2019), which employs Bayesian inference from a conditional genetic stock identification model. In general, the algorithm estimates the conditional probability distribution for each individual in the mixture, so it is probabilistically assigned to the closest genetic match from the set of populations in the baseline. To conduct 100% single-population simulations via Rubias, 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 region (Russia, United States) to which the population belonged.

Assessment of accuracy of self-assignment of individuals was conducted via 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 58 000 independent tests of known origin, as they were collected from known specific coho salmon populations, with the assumption of limited straying among populations.

The effect of baseline population sample size on accuracy of estimated stock composition of single-population samples was evaluated for six populations: one with high average F_{ST} (Noyo River, F_{ST} = 0.157), one with intermediate F_{ST} (Robertson Creek, F_{ST} = 0.097), and four with lower F_{ST} (Stave River, F_{ST} = 0.060; Chilliwack River, F_{ST} = 0.064; Chehalis River, F_{ST} = 0.060; and Qualicum River, F_{ST} = 0.056). Population sample sizes of 20, 40, 60, 80, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, and 900 individuals were consecutively evaluated for each population when available. Baseline sample sizes were constructed by randomly selecting the desired sample size from previously genotyped individuals in the population and incorporating these individuals into the baseline used for stock composition analysis. As only 97 individuals had been genotyped for the Noyo River (Supplementary Table 1[1]), the analysis was terminated at a baseline sample size of 80 individuals for this population.

The next stage of the evaluation incorporated analyses of two multipopulation 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. Eight populations were incorporated into each simulated mixture at set limits ranging from 5% to 20% of the sample. Rubias was used to estimate stock composition of the resultant mixture, and means and standard deviations were 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 1000 iterations from the Markov chain Monte Carlo 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 1000 iterations from the Markov chain Monte Carlo. 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.

Results

Heterozygosity and F_{ST}

Expected heterozygosity ranged from 0.00 to 0.50 across the 480 SNPs surveyed (Supplementary Table 2[1], excluding species ID and sex ID markers), and observed heterozygosity ranged from 0.00 to 0.51. Observed heterozygosities were >0.40 for 42% of the SNPs surveyed (Fig. 2). Global F_{ST} across SNPs ranged from 0.00 to 0.28, and 58% of the SNPs displayed an F_{ST} value between 0.05 and 0.10 (Fig. 3). SNPs included in the panel included many from a previous version of the panel (Beacham et al. 2017) and with the new EPIC4 SNPs (Supplementary Table 2[1]). The EPIC4 SNPs were chosen for their initial stock identification capability separating BC populations, so heterozygosities and F_{ST} values may not be representative of SNPs present in the genome.

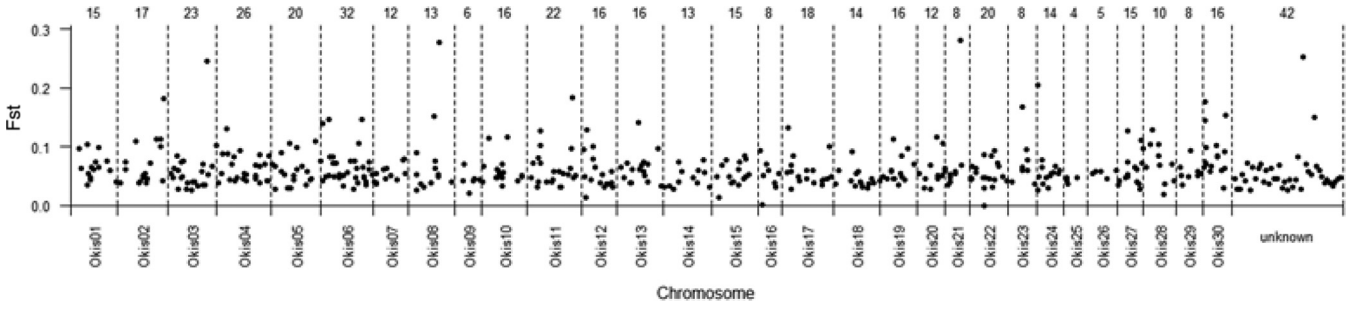
Genomic distribution of SNPs

The 480 SNPs surveyed were broadly distributed over the 30 chromosomes present in the coho salmon genome, ranging from a minimum of four SNPs present on chromosome Okis25 to 32 SNPs present on chromosome Okis6 (Fig. 4). Mean marker spacing across the chromosomes was 3.85 Mb (per chromosome mean 2.40–8.7 Mb). There were 42 SNPs present in scaffolds unassigned to specific chromosomes. SNPs with higher F_{ST} values were also widely distributed across chromosomes, so there was no clustering of these SNP sites on specific chromosomes.

Population structure

Significant genetic differentiation was observed among coho salmon populations sampled in the different CUs and geographic regions surveyed. The most distinctive stocks in the survey included the following: Russia (mean population F_{ST} = 0.130), California (mean F_{ST} = 0.150), interior Fraser River (CO-47, mean F_{ST} = 0.113), lower Thompson River (CO-7, mean F_{ST} = 0.125), North Thompson River (CO-9, mean F_{ST} = 0.114), and South Thompson River (CO-8, mean F_{ST} = 0.123) (Supplementary Fig. 1[1]). Populations on the islands of Haida Gwaii were also quite distinct, with those in the Haida Gwaii – Graham Island Lowlands CU (CO-25, mean F_{ST} = 0.103), Haida Gwaii – East CU (CO-23, mean F_{ST} = 0.094), and Haida Gwaii – West CU (CO-24, mean F_{ST} = 0.097) were generally distinct from those on mainland BC. The greatest average differentiation observed was between populations separated by the

Fig. 4. Distribution of F_{ST} across 30 coho salmon chromosomes and unassigned scaffolds for the 480 SNPs. The number of SNPs originating from each of the chromosomes is listed at the top of the figure. Position of the points along the *x* axis reflects the position of the SNP within each chromosome. For unassigned scaffolds, the points are separated by even spacing. The plot does not include Ots_SEXY3_1 or OkiOts_120255_113 used for sex ID and species ID, respectively.



greatest geographic distance. Russian populations were very distinct when compared with California populations, with an mean population pairwise F_{ST} of 0.200 (SD = 0.026).

Genetic differentiation was also observed at finer geographic scales. Coho salmon spawning populations generally clustered together in CUs, river drainages, and local geographic areas throughout the geographic range surveyed. For example, there was substantial clustering of populations in the Skeena River and Fraser River drainages (Supplementary Fig. 1[1]). Regional clustering was observed in populations in southeast Russia, Alaska, Washington, Columbia River, Oregon, and California.

Analysis of simulated single-population samples

The analysis of population variation indicated that there was a structure based on CUs for Canadian populations and a geographically based regional structure for Russian and American populations. This structure formed the basis to conduct estimation of stock composition for simulated single-population fishery samples for all populations in the baseline at both the population and reporting group (CU or region) level. In general, accurate estimates of stock composition were possible for most populations in the baseline at the CU or reporting group level (Fig. 5) and for many individual populations (Supplementary Fig. 2[1]). For example, overall accuracy for a population estimate to the correct CU was 93.4% for 258 Canadian populations, 94.8% to the correct geographic region for 65 American populations, 98.8% to the correct geographic region for nine Russian populations, and an overall accuracy of 93.8% to the correct CU or geographic region for all 332 populations.

Accurate allocations to many individual populations were observed. For example, in the East Vancouver Island – Georgia Strait CU (CO-13), estimated population stock compositions for single-population samples resolved with a 332-population baseline were 95.7% for Qualicum River, 98.6% for Puntledge River, and 99.3% for Quinsam River. These high levels of accuracy were observed even though there were 17 populations in the CU (Supplementary Fig. 2[1]). Similarly for the Lower Fraser River CU (CO-47), 24 populations were present in the baseline, yet estimates of stock composition for single-population samples were 98.5% for Chehalis River, 99.4% for Chilliwack River, 90.9% for Stave River, 96.6% for Norrish Creek, and 87.3% for Inch Creek (Supplementary Fig. 2[1]).

Assignment of individuals

Estimation of stock composition of single-population samples displayed high accuracy across CUs and populations, and further analyses investigated the accuracy of assignment of individuals, the most difficult of all stock identification applications. In general, accurate assignments (>80% accuracy) of individuals were observed across a range of CUs or reporting regions, corresponding to those CUs or regions that displayed the most genetic distinctiveness (Fig. 6). For example, individuals from North Thompson River populations (CO-9), South Thompson River populations (CO-8), lower Thompson River populations (CO-7), and interior Fraser River populations (CO-48) displayed a high level of assignment accuracy to CU, as did individuals from the Haida Gwaii CUs (CO-23, CO-24, CO-25). Accurate self-assignment of individuals was also observed for a number of populations on Haida Gwaii, Hood Canal, California, and southeast Alaska (Supplementary Fig. 3[1]). Within CUs, high assignment accuracy of individuals to the correct population (>90%) was also observed for specific populations, such as Conuma River, Maggie River, and Robertson Creek in the west coast Vancouver Island (WCVI) CU (CO-17) and Sangan River, Tell River, and Yakoun River in the northeastern Haida Gwaii CU (CO-25). Assignment of individuals could be accurate to both CU and population within CU, dependent upon the population under evaluation.

Population sample size in relation to accuracy of estimated stock compositions

Accuracy of estimated stock compositions varied among populations, and we investigated the effect of population sample size on estimation of stock composition accuracy of single-population samples for populations of varying levels of genetic distinctiveness. The Noyo River population was considered quite distinct, and only about 40 individuals were required to be genotyped in order for highly accurate estimates of stock composition to be obtained for the single-population samples (>99% estimate for a 100% simulated single-population sample; Fig. 7). Highly accurate estimates of stock composition for the Robertson Creek population were achieved at a baseline population size of 60 individuals. For less differentiated populations like Chilliwack River and Chehalis River, accurate estimates of stock composition were observed at a baseline population sample size of 150 individuals. Accuracy continued to increase for less differentiated populations of Stave River and Qualicum River up to a baseline sample size of 400 individuals, and marginal increases in accuracy were observed up to a baseline sample size of 900 individuals. The target sample size for a population to be included in a SNP GSI baseline varies with the genetic distinctiveness of the population.

Analysis of simulated multipopulation fishery mixtures

The accuracy and precision of two multipopulation simulated fishery samples were estimated for both population and CU–regional components. The mean error of estimated stock compositions of a simulated mixed-stock fishery sample from northern BC containing individuals from eight populations was 1.0% for population and 0.4% for CU or region (Table 1). For the southern BC simulated mixture, the mean error was 0.5% for population and 0.2% for CU or region. 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

Fig. 5. For simulated single-population samples, average percent accuracy for estimated stock compositions for individual populations in a conservation unit (CU) or reporting group back to population (open portion of bar) and to CU (black portion of bar) for all populations in a CU or reporting group, with the number of populations in the CU reported to the right of the bar. The 90% accuracy level is indicated in the figure.

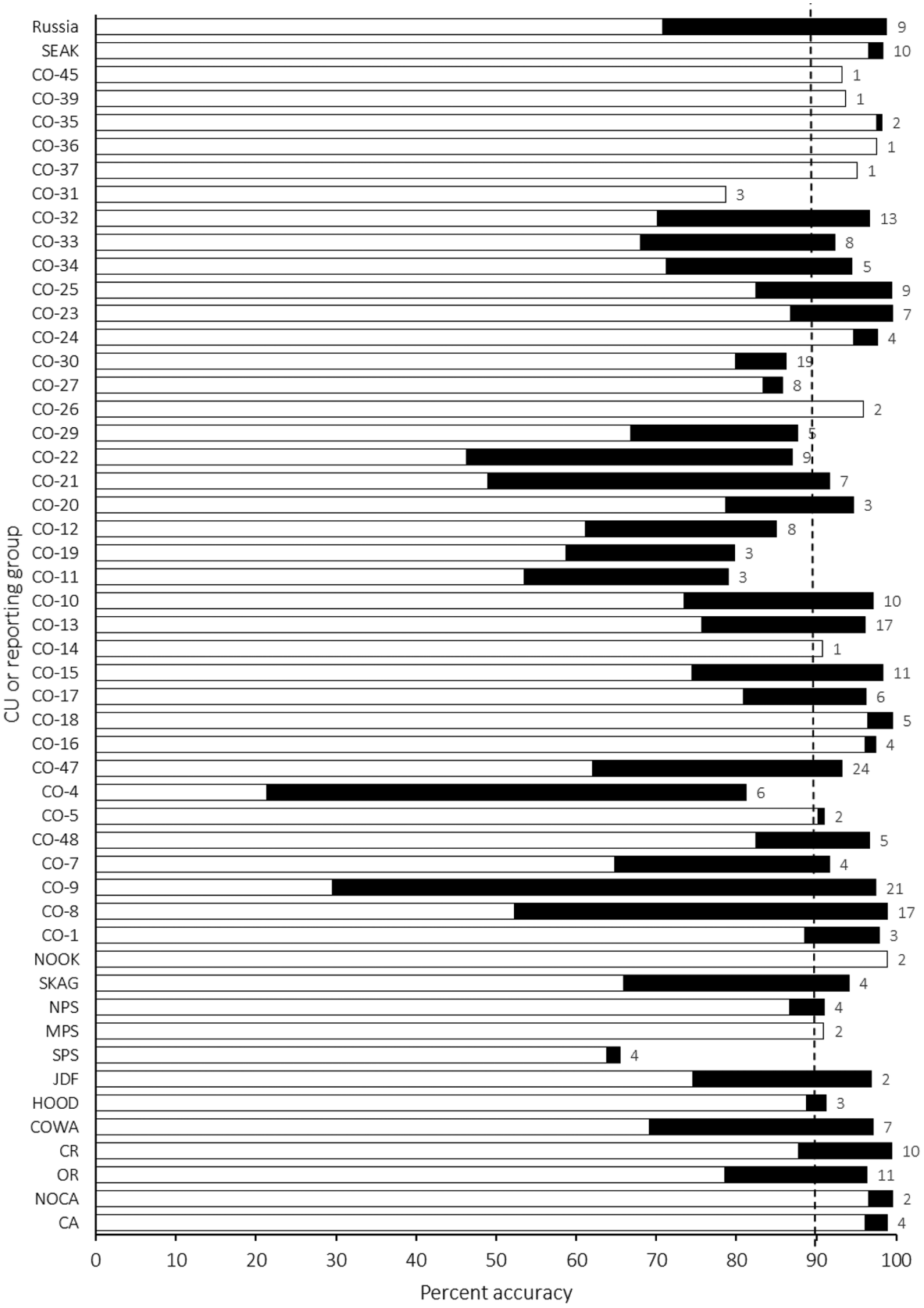
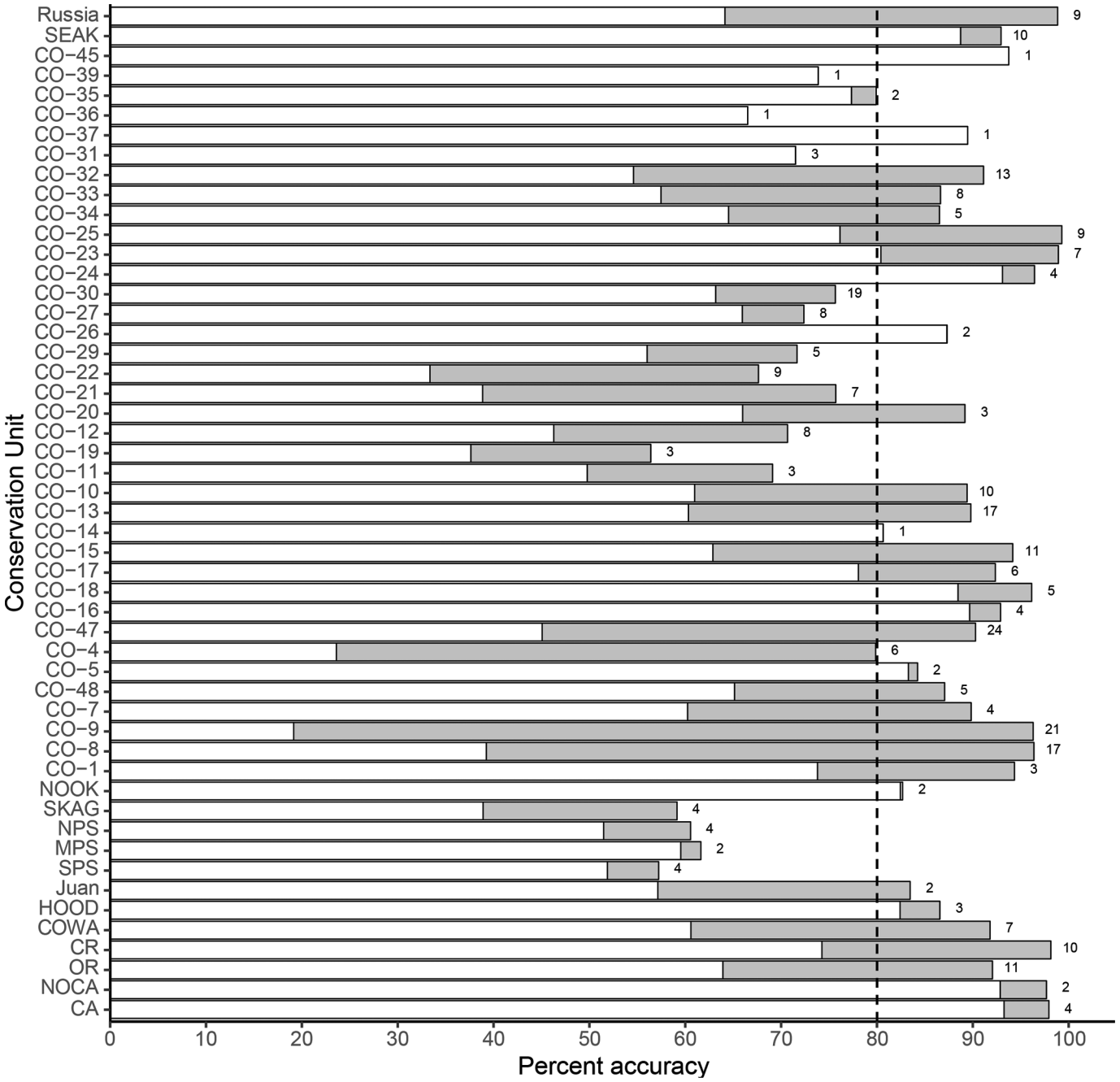


Fig. 6. Average percent accuracy for self-assignment for individual populations in a CU or reporting group back to population (open portion of bar) and to CU (grey portion of bar) for all populations in a CU or reporting group, with the number of populations in the CU reported to the right of the bar. The 80% accuracy level is indicated in the figure.



step of the evaluation of the baseline for mixed-stock fishery analysis.

Analysis of known-origin mixture

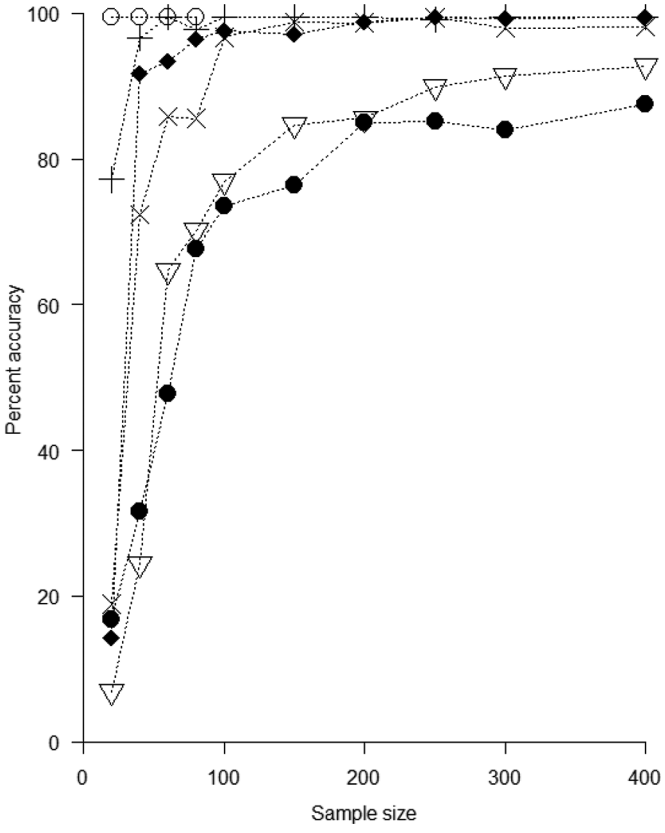
Estimations of stock compositions of simulated single-population and multipopulation samples suggested that accurate estimates of stock composition by reporting group and in some cases by population should be possible when applied to mixed-stock fishery samples of unknown origin. Assessment of this potential capability was tested by estimation of stock composition of a known-origin sample of 884 genotyped individuals from both Canadian and American populations that had been previously marked with CWTs. There were 525 tags from Canadian populations across 13 CUs. With only Canadian-origin tags considered, the mean error of estimation was 0.4% across the 13 CUs. With the addition of 359 American-origin tags, the mean error of estimation was 0.5% across the 13 CUs (Fig. 8). The increase in error of stock composition was largely a result of the estimate for the lower Stikine River CU, with an error of 0.2% (actual 2.1%, estimated 2.3%) when only Canadian-origin tags were considered, but 3.0% (actual 1.4%, estimated 4.4%) when both Canadian-origin and American-origin tags were considered. Only one population (Scud River) was genotyped in the CU, and self-assignment of individuals to the population was relatively accurate (93.6%; Supplemen-

Table 1. Estimated percent stock composition of simulated mixed-stock samples of coho salmon (n = 200) as may be encountered in northern and southern British Columbia (BC).

CU/Region	Population	True	Population	CU–region
Northern BC				
Southeast Alaska	Ford Arm Lake	15.0	14.9 (2.7)	14.9 (2.7)
CO-25	Yakoun River	5.0	5.1 (1.7)	5.2 (1.7)
CO-23	Pallant Creek	10.0	9.7 (1.9)	9.7 (1.9)
CO-36	Meziadin River	20.0	19.0 (2.3)	19.0 (2.3)
CO-34	Motase River	10.0	10.2 (2.3)	10.3 (2.3)
CO-27	Tyler Creek	15.0	13.4 (2.7)	13.8 (2.8)
CO-29	Kitimat River	10.0	12.9 (2.7)	25.9 (3.3)
	Gilttoyees Creek	15.0	13.0 (2.4)	
Southern BC				
CO-17	Conuma River	15.0	15.0 (2.7)	25.1 (3.4)
	Robertson Creek	10.0	10.0 (2.0)	
CO-13	Cowichan River	10.0	9.1 (2.2)	20.0 (3.1)
	Quinsam River	10.0	10.8 (2.1)	
CO-10	Capilano River	15.0	14.5 (2.8)	14.8 (2.8)
CO-47	Chilliwack River	10.0	9.9 (2.1)	10.0 (2.1)
Northern Puget Sound	Snohomish River	15.0	14.0 (2.1)	14.0 (2.1)
Columbia River	Clackamus River	15.0	14.3 (2.6)	15.2 (2.7)

Note: The expected conservation unit (CU)–regional compositions were obtained by adding the true population components for each CU–region. Standard deviation is in parentheses.

Fig. 7. Average percent accuracy for estimated stock compositions for simulated single-population samples of 200 individuals for Noyo River (open circles), Robertson Creek (+ symbols), Stave River (solid circles), Chilliwack River (diamonds), Chehalis River (× symbols), and Qualicum River (triangles) populations for actual baseline sample sizes varying from 20 to 400 individuals where available.



tary Fig. 3). The Klawock Lake population, not in the current baseline, contributed 35.7% (46 of 129 tags) of Alaskan-origin tags, and there was likely some misallocation of these individuals to the lower Stikine River CU.

American-origin tags originated from populations in 10 reporting groups. The mean error of estimation was 1.8% across the 10 reporting groups, but this result was heavily influenced by the southeast Alaska reporting group error (actual 14.6%, estimated 6.8%; Fig. 8). The mean error of estimation across the nine US west coast reporting groups was 1.0%. CWTs originating from populations in Washington and the Columbia River constituted 26.0% of the tags in the sample, and total estimated stock contribution of the nine US west coast reporting groups was 25.6%, indicating that errors in estimation were largely distributed among US west coast reporting groups, rather than between Canadian CUs and US west coast reporting groups.

The results from estimation of stock composition of the simulated single-population fishery samples suggested that accurate estimates of contributions of specific populations may be possible. The Canadian-origin tags previously noted were recovered from 16 populations. The mean error of estimation across the 16 populations was 0.4% (Fig. 9). The capability of estimating accurate stock composition of specific populations of coho salmon has been verified for those populations where a high accuracy of identification in the simulated single-population samples was observed.

Application to mixed-stock fishery sampling

Eight actual 2018 mixed-stock fishery samples of unknown origin were evaluated for the study (Table 2). Based upon the geographic locations and fishery timing, inferences can be drawn as to the reliability of the estimated stock compositions in actual practice. Further information is provided in the online Supplementary Materials[1].

Discussion

In Canada, the adoption of CUs under the WSP presented a fisheries management and assessment challenge. There are currently 44 CUs identified for coho salmon in British Columbia, and successful implementation of the WSP will likely require identification of fishery impacts by CU. At the time of the adoption of the WSP, there were no known methods available to enable resolution of fishery impacts across all CUs. Some fishery managers were doubtful that such resolution of mixed-stock fishery stock compositions in BC to the CU level would be possible, as GSI had never been applied to provide such fine-scale resolution of stock composition over wide geographic areas. The current study has

Table 2. Percent stock composition (standard deviation in parentheses) by geographic region or CU of various 2018 fisheries.

	Stikine River, commercial and test	North Coast, troll	Johnstone Strait, sport	WCVI, sport	Barkley Sound, sport	Juan de Fuca Strait, sport	Northern Strait of Georgia, sport	Nicomen Slough, sport
	September	August	July	August	September	September	July	October
Sample size	12	189	172	368	120	242	105	96
Region or CU								
Southeast Alaska	0.0 (0.8)	20.5 (3.6)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.1)
Alsek River CO-45	0.0 (1.4)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)
Lower Stikine CO-39	100.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Lower Nass CO-35	0.0 (6.5)	2.9 (1.3)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Upper Nass CO-36	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Portland Sound–Observatory Inlet–Portland Canal CO-37	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Skeena Estuary CO-31	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)
Lower Skeena CO-32	0.0 (0.4)	1.1 (1.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.1)
Middle Skeena CO-33	0.0 (0.1)	0.3 (0.6)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Upper Skeena CO-34	0.0 (0.1)	2.2 (1.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)
Haida Gwaii–Graham Island Lowlands CO-25	0.0 (0.1)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.1)
Haida Gwaii–East CO-23	0.0 (0.2)	14.8 (2.6)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.3)
Haida Gwaii–West CO-24	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Northern Coastal Streams CO-30	0.0 (0.0)	0.2 (0.3)	0.0 (0.2)	0.0 (0.1)	0.0 (0.2)	0.0 (0.1)	0.0 (0.0)	0.0 (0.2)
Hecate Strait Mainland CO-27	0.0 (0.6)	3.6 (1.8)	0.0 (0.1)	0.3 (0.4)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.4)
Mussel–Kynoch CO-26	0.0 (0.6)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Douglas Channel–Kitimat Arm CO-29	0.0 (0.1)	0.1 (0.5)	0.0 (0.0)	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Bella Coola–Dean Rivers CO-22	0.0 (0.0)	0.4 (0.5)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)
Rivers Inlet CO-21	0.0 (2.2)	0.3 (1.0)	0.0 (0.2)	0.0 (0.1)	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)
Smith Inlet CO-20	0.0 (1.8)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Southern Coastal Streams–Queen Charlotte Strait–Johnstone Strait–Southern Fjords CO-12	0.0 (0.1)	0.0 (0.1)	0.2 (0.4)	0.1 (0.2)	0.0 (0.2)	0.0 (0.4)	0.0 (0.0)	0.0 (0.0)
Homathko–Klinaklini Rivers CO-19	0.0 (1.3)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Georgia Strait Mainland CO-11	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Howe Sound–Burrard Inlet CO-10	0.0 (0.0)	0.0 (0.2)	9.3 (2.2)	2.4 (0.7)	0.8 (0.7)	12.3 (1.9)	22.6 (3.7)	0.0 (0.1)
East Vancouver Island–Georgia Strait CO-13	0.0 (1.5)	10.1 (1.9)	39.8 (3.7)	2.9 (1.3)	0.0 (0.1)	7.6 (1.6)	8.8 (2.9)	0.0 (0.2)
East Vancouver Island–Johnstone Strait–Southern Fjords CO-14	0.0 (0.7)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Nahwitti Lowland CO-15	0.0 (0.0)	2.3 (1.2)	2.3 (1.2)	1.4 (0.6)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.3)
West Vancouver Island CO-17	0.0 (1.1)	4.8 (1.3)	3.6 (1.5)	18.6 (2.2)	99.1 (0.8)	0.0 (0.4)	0.0 (0.0)	1.0 (0.9)
Clayoquot CO-18	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.2)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Juan de Fuca–Pachena CO-16	0.0 (0.4)	0.0 (0.4)	1.6 (0.9)	2.5 (0.8)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Lower Fraser CO-47	0.0 (0.2)	2.2 (0.9)	32.0 (3.4)	15.0 (1.7)	0.0 (0.1)	26.9 (3.4)	43.1 (4.0)	99.0 (1.2)
Lillooet CO-4	0.0 (1.6)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)
Fraser Canyon CO-5	0.0 (0.8)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Interior Fraser CO-48	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Lower Thompson CO-7	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	1.2 (0.7)	0.0 (0.0)	0.0 (0.0)
North Thompson CO-9	0.0 (1.3)	0.0 (0.1)	0.0 (0.4)	0.0 (0.1)	0.0 (0.3)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)
South Thompson CO-8	0.0 (1.8)	0.0 (0.1)	0.6 (0.6)	0.0 (0.1)	0.0 (0.1)	0.4 (0.4)	0.0 (0.1)	0.0 (0.0)
Boundary Bay CO-1	0.0 (1.7)	0.0 (0.1)	0.0 (0.1)	0.9 (0.7)	0.0 (0.0)	2.8 (1.1)	1.0 (1.4)	0.0 (0.0)
Nooksack River	0.0 (0.2)	0.0 (0.0)	0.3 (0.8)	10.0 (2.0)	0.0 (0.0)	0.9 (1.0)	7.6 (3.0)	0.0 (0.0)
Skagit River	0.0 (0.1)	5.5 (1.8)	0.3 (0.3)	10.0 (2.2)	0.0 (0.0)	19.0 (3.1)	5.2 (2.4)	0.0 (0.0)
Northern Puget Sound	0.0 (0.2)	8.6 (1.8)	2.4 (1.4)	15.1 (2.1)	0.0 (0.0)	16.7 (3.4)	5.7 (2.5)	0.0 (0.0)
Mid-Puget Sound	0.0 (0.0)	0.7 (0.7)	0.5 (0.6)	8.8 (1.9)	0.0 (0.1)	9.1 (2.1)	3.8 (2.3)	0.0 (0.0)
Southern Puget Sound	0.0 (0.3)	0.3 (0.8)	0.0 (0.0)	0.2 (0.5)	0.0 (0.0)	0.4 (0.6)	2.2 (2.0)	0.0 (0.1)
Juan de Fuca Strait	0.0 (1.0)	0.9 (0.7)	0.0 (0.0)	0.9 (0.5)	0.0 (0.0)	1.4 (0.7)	0.0 (0.0)	0.0 (0.0)
Hood Canal	0.0 (0.0)	12.5 (2.5)	0.0 (0.0)	9.3 (1.7)	0.0 (0.1)	0.1 (0.1)	0.0 (0.2)	0.0 (0.0)
Coastal Washington	0.0 (0.3)	5.8 (1.7)	1.2 (0.8)	1.6 (0.5)	0.0 (0.1)	0.1 (0.3)	0.0 (0.0)	0.0 (0.0)
Columbia River	0.0 (0.3)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.1)	0.0 (0.0)
Oregon	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)
Northern California	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

provided accurate identification of Canadian coho salmon sampled from mixed-stock fisheries to the CU level and has enabled assessment of fishery impacts that is sufficiently informative for conservation-based management as envisaged in WSP. Coho salmon harvested in Canadian commercial and recreational fisheries were identified to Canadian CUs and American geographic regions from southeast Alaska to California, confirming the utility of a GSI approach for conservation-based assessment of mixed-stock harvest on a wide geographic scale.

The adoption of the WSP required that populations of all five main species of Pacific salmon in BC be classified into CUs for assessment purposes, even though at the time of adoption there was no method to identify fishery impacts by CU. Direct DNA sequencing, coupled with Ampliseq technology that allowed

Fig. 8. Accuracy of regional (United States) and conservation unit (Canada) estimated stock composition (%) for a mixed-stock sample of 884 coded-wire tagged individuals sampled in 2018 fisheries in BC estimated with genetic stock identification (GSI). Actual percentage is the black bar, estimated percentage is the open bar. Standard deviation is indicated for the estimate.

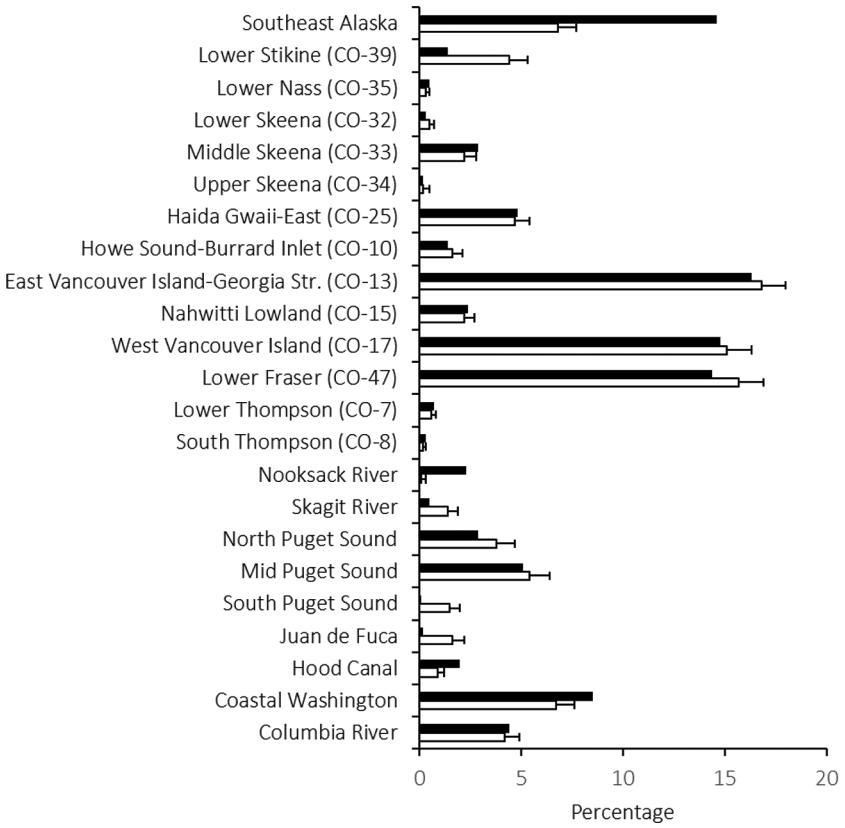
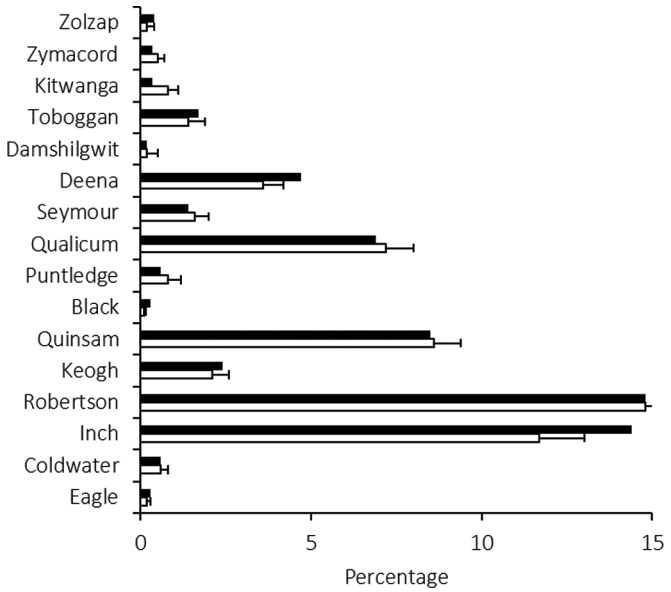


Fig. 9. Accuracy of estimated Canadian population-specific stock composition (%) for a mixed-stock sample of 884 coded-wire tagged (CWT) individuals sampled in 2018 fisheries in BC estimated with only GSI. Actual percentage is the black bar, estimated percentage is the open bar. Only Canadian populations for which CWTs were recovered and GSI analysis was conducted are illustrated, along with standard deviation of the estimate.



genotyping at hundreds of SNPs through a single PCR, provided the advance in applied genetics that allowed identification of fishery impacts by CU in coho salmon. This paradigm-shifting advance can be applied to other species, and we are in the process of applying it to Chinook salmon (*Oncorhynchus tshawytscha*), chum salmon (*Oncorhynchus keta*), and sockeye salmon (*Oncorhynchus nerka*) and coupling it with PBT technology for Chinook salmon (Beacham et al. 2018). High-resolution stock identification, which may combine both GSI and PBT, can be achieved and will provide increased accuracy in estimation of stock composition for those populations of potential conservation concern, typically comprising <5% of a fishery sample.

Population structure

A regionally based population structure is generally 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 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. Thus, a study of population structure can yield valuable insights as to how GSI will perform in mixed-stock fishery application. Previous microsatellite-based population structure studies indicated that a regional, geographically based structure was apparent (Olsen et al. 2003; Bucklin et al. 2007; Johnson and Banks 2008; Beacham et al. 2011*b*). Population structure of coho salmon populations surveyed in the current study displayed a pattern of CU or regionally based population structure. Therefore, coho salmon population structure meets the important condition that unsampled populations

contributing to mixed fishery samples will likely be allocated to sampled populations in the same region. In applications where errors in population estimation are considered to be too large for satisfactory use, then it is necessary to increase the sample size of key existing baseline populations or sample additional baseline populations in the CU or region to enhance the reliability of regional estimates of stock composition.

Population samples were available from 37 of the 44 CUs defined for BC coho salmon, with the unrepresented CUs restricted to northern BC, where the remoteness of the locations has precluded sample collection to date. 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.

Population sample size in relation to accuracy of estimated stock compositions

The target baseline population sample size is dependent upon population genetic differentiation. Through simulation of SNP variation, Morin et al. (2009) reported that increasing sample size up to 100 individuals for detection of population structure was beneficial. However, for populations with $F_{ST} = 0.10$, little increase in detection power was observed above 60 sampled individuals, similar to the results for the Robertson Creek population in the current study. Beacham et al. (2011*a*) reported that less genetically distinct populations required larger population sample sizes to achieve a given level of accuracy in estimated stock compositions, similar to the results observed in the current study. When increased accuracy of estimated stock compositions is required for a particular population in the baseline, the most direct route to follow is to increase sample size for the target population. For SNP-based baseline development, 100 genotyped individuals per population is a reasonable initial target. However, for populations more difficult to discriminate, sample sizes of up to 500 individuals may be required. In the current baseline, 25.9% of the populations contained at least 100 genotyped individuals, 50.9% of the populations contained 40–99 genotyped individuals, 15.7% of the populations contained 20–39 individuals, and 7.5% of the populations contained 11–19 individuals. Many of the poorly identified populations were associated with population sample sizes of <40 genotyped individuals.

Analysis of known-origin mixture

Accurate estimation of stock composition via GSI relies on a baseline that includes all major populations potentially contributing to a mixed-stock fishery sample. Our study provided an illustration of the misallocation that may occur when this assumption is violated. The Klawock Lake population comprised 35.7% of the Alaskan-origin CWTs and 5.2% of the total CWT sample, but the population was not in the baseline. The resultant misallocation was observed across northern CUs; the most notable was the lower Stikine River CU, where the contribution of the CU to the known-origin CWT sample was overestimated by 3.0%, some 58% of the Klawock Lake contribution to the total sample. The Stikine River originates in northern BC and flows across southeast Alaska to enter the Pacific Ocean proximal to other southeast Alaska populations, with geographic proximity accounting for the majority of the misallocation of the Klawock Lake population component. The addition of the Klawock Lake population to the baseline would likely remediate the observed misallocation.

In general, analysis of the known-origin CWT sample confirmed the ability of the baseline to provide reliable estimates of stock composition for CUs when applied to analysis of mixed-stock fishery samples, a result that 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 simulations. For those populations such as Inch Creek where underestimation of actual contributions was observed and would be expected as illustrated by the single-population simulations, PBT may provide a method to enhance estimates of stock composition for specific populations (Beacham et al. 2019*a*). For example, the Inch Creek contribution to the total CWT sample was underestimated by 2.7% by GSI alone (actual 14.4%, estimated 11.7%), whereas when both GSI and PBT were applied to estimate the Inch Creek contribution to the CWT sample, the estimate was 14.7%, an error of 0.3%.

The CWT sample consisted of tags recovered from both wild- and hatchery-origin individuals. The GSI baseline can be applied to fisheries where both hatchery-origin and wild-origin coho are caught. In BC, there are many integrated hatcheries resulting in very similar or undifferentiated hatchery and wild populations (Le Luyer et al. 2017). As long as the population is represented in the baseline, no difference in accuracy or precision of estimated stock compositions is expected between wild- and hatchery-origin individuals in a mixed-stock fishery sample.

Mixed-stock fisheries

In our study, we estimated stock composition of eight mixed-stock samples of unknown origin of coho salmon with a 332-population baseline arranged by CU for Canadian populations and geographic reporting group for American populations (Supplementary Results[1]). The rationale was that given the geographic locations and timing of the fisheries, it should be possible to evaluate estimated stock compositions of actual fishery samples against expectations to measure performance of the baseline for stock identification applications. In general, stock composition results met expectations based upon the geographic location of the fishery. The highly-mixed stock troll fishery off Haida Gwaii in northern BC displayed contributions from southeast Alaska, Haida Gwaii, northern and central BC, southern BC, Washington, and the Columbia River, as would be expected (Beacham et al. 2012*a*). In southern BC fisheries, all individuals sampled were adipose fin-clipped and indicative of their hatchery origin. Little if any contributions from the northern or central coasts (CUs north of CO-12) would be expected, and these were the exact results observed in the fisheries. For example, the August WCVI recreational fishery was expected to be composed primarily of individuals from the WCVI and lower Fraser River CUs, along with Washington reporting regions, and these were indeed the observed results. The September sample from Barkley Sound and Alberni Inlet is from a much more inshore fishery than the WCVI fishery, of which it is a part. The later timing, coupled with a restricted geographic location, led to virtually all of the individuals originating from the WCVI CU, again meeting expectations. There are only two populations (Robertson Creek, Conuma River) in the CU for which adipose fin clipping is conducted, and given the location of the fishery, virtually all adipose fin-clipped individuals should be of Robertson Creek origin. Virtually all WCVI CU individuals ($n = 119$) were identified as originating from Robertson Creek ($n = 118$), with one individual identified from Conuma River, confirming the reliability of the stock composition estimates to a specific population. The freshwater fishery investigated was in Nicomen Slough, and estimated stock composition of the fishery was composed almost entirely of individuals of lower Fraser River origin. The one non-Fraser River-origin individual in the October 2018 fishery sample was identified as Robertson Creek origin with a 100% probability level, presumably indicative of a stray. Inch Creek individuals were estimated to comprise about 60% of the sample and Norrish Creek about 34% of the sample. Norrish Creek is a later-returning population than Inch Creek, and this was illustrated by Norrish Creek comprising 91.5% of the November fishery sample ($n = 59$) and Inch Creek 8.5%. In summary, estimated stock compositions of the fishery samples corresponded

very well to those expected based upon fishery location and timing.

Summary

Current and historical assessment of coho salmon fisheries impacts in BC has been conducted with the application of CWTs, but CWTs are not applied to releases from some of the largest hatcheries in southern BC due to funding limitations, and thus their specific contributions to mixed-stock ocean fisheries are unknown in a CWT-based assessment system. The current study has not only demonstrated that it was possible to identify coho salmon mixed-stock fishery contributions by CU, but it is also possible in many instances to identify specific populations within a CU in the fishery samples, as well as identifying individuals to some specific populations. Coho salmon are mass-marked (juvenile adipose fin-clipped) upon hatchery release in many hatcheries in BC, and distinguishing between hatchery-origin and natural-origin individuals can be done visually. A genetics-based assessment regime benefits from the mass marking of hatchery-produced salmon, thereby facilitating improved monitoring of wild-enhanced fish interactions and the evaluation of hatchery contributions to harvest. We have demonstrated that a genetics-based assessment system can overcome the deficiencies present in the current CWT-based assessment regime (Beacham et al. 2019*a*) and also provides an opportunity for conservation-based management of Canadian coho salmon.

The ability to provide reliable estimates of stock composition by CU was facilitated by the switch from a microsatellite-based baseline to an SNP-based baseline with the SNPs genotyped via direct DNA sequencing of amplicons. Ampliseq allowed hundreds of SNPs to be amplified in single PCR, and direct DNA sequencing of the resultant amplicons, coupled with automated scoring of the genotypes, resulted in cost-effective genotyping and unprecedented ability to provide accurate estimates of stock composition to very discrete geographic regions or CUs. Similar results can be expected when applied to other nonsalmonid species, and a new era in salmonid stock identification is dawning.

Acknowledgements

A very substantial effort was undertaken to obtain baseline samples from coho salmon sampled in this study. Starting from the south, we thank J.C. Garza and D.J. Teel of the National Marine Fisheries Service for baseline samples from California, Oregon, and the Columbia River. J.B. Shaklee of the Washington Department of Fish and Wildlife provided samples from Washington State populations, including the Columbia River. In southern BC, we thank various staff of the Fisheries and Oceans Canada (DFO) for baseline sample collection, as well as First Nations staff. We 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 BC. These included the Gitxsan Watershed Authority and the Skeena Fisheries Commission in the Skeena River drainage and the Nisga'a First Nation in the Nass watershed. The Haida Fisheries Program assisted in sampling streams located in Haida Gwaii. In southeast Alaska, samples were provided by R.L. Wilmot of the National Marine Fisheries Service (NMFS) Auke Bay Laboratory and S. Miller of the United States Fish and Wildlife Service Anchorage Laboratory. KamchatNIRO staff provided the samples from Russian populations. The J.O. Thomas CWT recovery laboratory provided tissue samples from many individuals processed from 2018 commercial, recreational, and First Nations fisheries in BC, as well as samples of adipose fin-clipped individuals from the northern BC troll fishery. B. Baxter (DFO) prepared the map depicting the coho salmon conservation units in BC. This research was carried out in conjunction with EPIC4 (Enhancing Production in Coho: Culture, Community, Catch), a project supported in part by DFO and the government of Canada through Genome Canada, Genome British Columbia, and Genome Quebec. Funding support was also previously provided by the Northern Endowment Fund of the Pacific Salmon Commission for baseline development. Three anonymous reviewers provided comments and edits on the manuscript.

References

Beacham, T.D., Candy, J.R., Supernault, K.J., Ming, T., Deagle, B., Schulze, A., et al. 2001. Evaluation and application of microsatellite and major histocompatibility complex variation for stock identification of coho salmon in British Columbia. Trans. Am. Fish. Soc. **130**: 1116–1149. doi:10.1577/1548-8659(2001)130<1116:EAAOMA>2.0.CO;2.

Beacham, T.D., Candy, J.R., Wallace, C., Urawa, S., Sato, S., Varnavskaya, N.V., et al. 2009. Microsatellite stock identification of chum salmon on a Pacific Rim basis. N. Am. J. Fish. Manage. **29**: 1757–1776. doi:10.1577/M08-188.1.

Beacham, T.D., McIntosh, B., and Wallace, C. 2011*a*. A comparison of polymorphism of genetic markers and population sample sizes required for mixed-stock analysis of sockeye salmon (*Oncorhynchus nerka*) in British Columbia. Can. J. Fish. Aquat. Sci. **68**(3): 550–562. doi:10.1139/F10-167.

Beacham, T.D., Wetklo, M., Deng, L., and MacConnachie, C. 2011*b*. Coho salmon population structure in North America determined from microsatellites. Trans. Am. Fish. Soc. **140**: 253–270. doi:10.1080/00028487.2011.558782.

Beacham, T.D., Candy, J.R., Wallace, C., Wetklo, M., Deng, L., and MacConnachie, C. 2012*a*. Microsatellite mixed-stock identification of coho salmon in British Columbia. Mar. Coastal Fish. **4**: 85–100. doi:10.1080/19425120.2012.661393.

Beacham, T.D., Wallace, C.G., Le, K.D., and Beere, M. 2012*b*. Population structure and run timing of steelhead in the Skeena River, British Columbia. N. Am. J. Fish. Manage. **32**: 262–275. doi:10.1080/02755947.2012.675953.

Beacham, T.D., Beamish, R.J., Candy, J.R., Tucker, S., Moss, J.H., and Trudel, M. 2014. Stock-specific migration pathways of juvenile sockeye salmon in British Columbia waters and in the Gulf of Alaska. Trans. Am. Fish. Soc. **143**: 1386–1403. doi:10.1080/00028487.2014.935476.

Beacham, T.D., Beamish, R.J., Neville, C.M., Candy, J.R., Wallace, W., Tucker, S., and Trudel, M. 2016. Stock-specific size and migration of juvenile Coho Salmon in British Columbia and southeast Alaska waters. Mar. Coastal Fish. **8**: 292–314. doi:10.1080/19425120.2016.1161683.

Beacham, T.D., Wallace, C., MacConnachie, C., Jonsen, K., McIntosh, B., Candy, J.R., et al. 2017. Population and individual identification of Coho salmon in British Columbia through parentage-based tagging and genetic stock identification: an alternative to coded-wire tags. Can. J. Fish. Aquat. Sci. **74**(9): 1391–1410. doi:10.1139/cjfas-2016-0452.

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

Beacham, T.D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J.R., Willis, D., et al. 2019*a*. Comparison of coded-wire tagging with parentage-based tagging and genetic stock identification in a large-scale coho salmon fisheries application in British Columbia, Canada. Evol. Appl. **12**: 230–254. doi:10.1111/eva.12711. PMID:30697336.

Beacham, T.D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J.R., Willis, D., et al. 2019*b*. Variation in migration pattern, broodstock origin, and family productivity of coho salmon hatchery populations in British Columbia, Canada, derived from parentage-based tagging. Ecol. Evol. **9**: 9891–9906. doi:10.1002/ece3.5530. PMID:31534702.

Bradbury, I.R., Wringe, B.F., Watson, B., Paterson, I., Horne, J., Beiko, R., et al. 2018. Genotyping-by-sequencing of genome-wide microsatellite loci reveals fine-scale harvest composition in a coastal Atlantic salmon fishery. Evol. Appl. **11**: 918–930. doi:10.1111/eva.12606. PMID:29928300.

Bucklin, K.A., Banks, M.A., and Hedgecock, D. 2007. Assessing genetic diversity of protected coho salmon (*Oncorhynchus kisutch*) populations in California. Can. J. Fish. Aquat. Sci. **64**(1): 30–42. doi:10.1139/f06-171.

Campbell, N.R., Harmon, S., and Narum, S.R. 2015. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. **15**: 855–867. doi:10.1111/1755-0998.12357. PMID:25476721.

Farley, E.V., Starovoytov, A., Naydenko, S., Heintz, R., Trudel, M., Guthrie, C., et al. 2011. Implications of a warming eastern Bering Sea for Bristol Bay sockeye salmon. ICES J. Mar. Sci. **68**: 1138–1146. doi:10.1093/icesjms/fsr021.

Fisheries and Oceans Canada. 2005. Canada's policy for conservation of wild Pacific salmon. Fisheries and Oceans Canada, 401 Burrard Street, Vancouver, BC, V6C 3S4, Canada.

Fisheries and Oceans Canada. 2019. Coho salmon (*Oncorhynchus kisutch*) conservation units, sites, and status [online]. Available from https://open.canada.ca/data/en/dataset/bfeb3ecd-b99a-4d1a-b8c1-87373f31f4bd.

Ford, M.J., Teel, D., Van Doornik, D.M., Kuligowski, D., and Lawson, P.W. 2004. Genetic population structure of central Oregon Coast coho salmon (*Oncorhynchus kisutch*). Conserv. Genet. **5**: 797–812. doi:10.1007/s10592-004-1983-5.

Gilbey, J., Cauwelier, E., Coulson, M.W., Stradmeyer, L., Sampayo, J.N., Armstrong, A., et al. 2016. Accuracy of assignment of Atlantic Salmon (*Salmo salar* L.) to rivers and regions in Scotland and northeast England based on single nucleotide polymorphism (SNP) markers. PLoS ONE, **11**(10): e0164327. doi:10.1371/journal.pone.0164327. PMID:27723810.

Hess, J.E., Ackerman, M.W., Fryer, J.K., Hasselman, D.J., Steele, C.A., Stephenson, J.J., et al. 2016. Differential adult migration-timing and stock-specific abundance of steelhead in mixed stock assemblages. ICES J. Mar. Sci. **73**: 2606–2615. doi:10.1093/icesjms/fsw138.

Holtby, B.L., and Ciruna, K.A. 2007. Conservation Units for Pacific salmon under the Wild Salmon Policy. Can. Sci. Adv. Secret. Sci. Adv. Report 2007/070.

Jefferts, K.B., Bergmann, P.K., and Fiscus, H.F. 1963. A coded wire identification system for macro-organisms. Nature, **198**: 460–462. doi:10.1038/198460a0.

Johnson, M.A., and Banks, M.A. 2008. Genetic structure, migration, and patterns of allelic richness among coho salmon (*Oncorhynchus kisutch*) populations of the Oregon coast. Can. J. Fish. Aquat. Sci. **65**(7): 1274–1285. doi:10.1139/F08-044.

Jombart, T., and Ahmed, I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, **27**(21): 3070–3071. doi:10.1093/bioinformatics/btr521. PMID:1926124.

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**(5): 698–708. doi:10.1139/cjfas-2013-0502.

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. Natl. Acad. Sci. U.S.A. **114**: 12964–12969. doi:10.1073/pnas.1711229114. PMID:29162695.

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

Moore, J.-S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., et al. 2014. Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. Mol. Ecol. **23**: 5680–5697. doi:10.1111/mec.12972. PMID: 25327895.

Moran, B.M., and Anderson, E.C. 2019. Bayesian inference from the conditional genetic stock identification model. Can. J. Fish. Aquat. Sci. **76**(4): 551–560. doi:10.1139/cjfas-2018-0016.

Morin, P.A., Martien, K.K., and Taylor, B.L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Mol. Ecol. Resour. **9**: 66–73. doi:10.1111/j.1755-0998.2008.02392.x. PMID:21564568.

Olsen, J.B., Miller, S.J., Spearman, W.J., and Wenburg, J.K. 2003. Patterns of intra- and inter-population genetic diversity in Alaskan coho salmon: implications for conservation. Conserv. Genet. **4**: 557–569. doi:10.1023/A:1025684104113.

Paradis, E., and Schliep, K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, **35**: 526–528. doi:10.1093/bioinformatics/bty633.

Pembleton, L.W., Cogan, N.O.I., and Forster, J.W. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. **13**(5): 946–952. doi:10.1111/1755-0998.12129. PMID:23738873.

Price, M.H.H., Rosenberger, A.G.J., Taylor, G.G., and Stanford, J.A. 2014. Comment: Population structure and run timing of sockeye salmon in the Skeena river, British Columbia. N. Am. J. Fish. Manage. **34**: 1167–1170. doi:10.1080/02755947.2014.956162.

Price, M.H.H., English, K.K., Rosenberger, A.G., MacDuffee, M., and Reynolds, J.D. 2017. Canada's Wild Salmon Policy: an assessment of conservation progress in British Columbia. Can. J. Fish. Aquat. Sci. **74**(10): 1507–1518. doi:10.1139/cjfas-2017-0127.

R Core Team. 2019. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Schliep, K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics, **27**(4): 592–593. doi:10.1093/bioinformatics/btq706. PMID:21169378.

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. doi:10.1577/1548-8659(1999)128<0058:HGHICS>2.0.CO;2.

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. doi:10.1577/1548-8659(2000)129<1223:GDOSSO>2.0.CO;2.

Small, M.P., Beacham, T.D., Withler, R.E., and Nelson, R.J. 1998*a*. Discriminating coho salmon (*Oncorhynchus kisutch*) populations within the Fraser River, British Columbia, using microsatellite DNA markers. Mol. Ecol. **7**: 141–155. doi:10.1046/j.1365-294x.1998.00324.x.

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

Smith, C.T., Nelson, R.J., Wood, C.C., and Koop, B.F. 2001. Glacial biogeography of North American coho salmon (*Oncorhynchus kisutch*). Mol. Ecol. **10**: 2775–2785. doi:10.1046/j.1365-294X.2001.t01-1-01405.x. PMID:11903891.

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

Van Doornik, D.M., Teel, D.J., Kuligowski, D.R., Morgan, C.A., and Casillas, E. 2007. Genetic analyses provide insight into the early ocean stock distribution and survival of juvenile coho salmon off the coasts of Washington and Oregon. N. Am. J. Fish. Manage. **27**: 220–237. doi:10.1577/M06-130.1.