**Population and individual identification of Chinook Salmon
 in British Columbia through parentage-based tagging and
 genetic stock identification with single nucleotide
 polymorphisms**

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2017-0168.R1
Manuscript Type:	Article
Date Submitted by the Author:	31-Jul-2017
Complete List of Authors:	Beacham, Terry; Department of Fisheries and Oceans, Wallace, Colin; Department of Fisheries and Oceans, MacConnachie, Cathy; Fisheries and Oceans Canada - Pacific Biological Station, Jonsen, Kim; Fisheries and Oceans Canada - Pacific Biological Station, McIntosh, Brenda; Fisheries and Oceans Canada - Pacific Biological Station, Molecular Biology Candy, John; Department of Fisheries and Oceans, Withler, Ruth; Pacific Biological Station, Fisheries and Oceans Canada
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	Chinook Salmon, parentage-based tagging, genetic stock identification, single nucleotide polymorphisms

SCHOLARONE™
 Manuscripts

Population and individual identification of Chinook Salmon in British Columbia through
parentage-based tagging and genetic stock identification with single nucleotide polymorphisms
Terry D. Beacham^{1,2}, Colin Wallace¹, Cathy MacConnachie¹, Kim Jonsen¹, Brenda McIntosh¹, John R.
Candy¹, and Ruth E. Withler¹

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

²Corresponding author

Abstract

A study was undertaken to evaluate whether a parentage-based tagging (PBT) and genetic stock identification (GSI) program has the potential to emulate the results from an existing coded-wire tag (CWT) assessment program in British Columbia. A PBT/GSI approach was used to identify Chinook Salmon (*Oncorhynchus tshawytscha*) to specific populations and broodyears where 36,241 individuals from 45 populations were genotyped at 321 single nucleotide polymorphisms (SNPs). Known-origin and known-age age 1 juveniles from seven test populations were assigned via PBT (two-parental genotypes required, 538 of 656 juveniles assigned; one parental genotype required, 636 of 656 juveniles assigned) with a minimum accuracy of 99.9%. Assignment accuracy via PBT of 1,026 age 1, 2 or 3 Chinook Salmon returning to nine populations in 2015 or 2016 (two-parental genotypes required, 556 of 1,026 individuals assigned; one parental genotype required, 898 of 1,026 individuals assigned) was a minimum of 99.8%. A PBT/GSI or PBT system of identification may provide an alternate cost-effective method of identification in the assessment and conservation of Canadian-origin Chinook Salmon relative to the existing CWT program, thereby providing very high resolution of mixed-stock fishery samples containing both hatchery-origin (adipose fin clipped) and wild-origin (unclipped) populations.

Keywords: Chinook Salmon, parentage-based tagging, genetic stock identification, single nucleotide polymorphisms

Introduction

Assignment of individuals to specific populations can be an important component of analysis of mixed-stock salmon fisheries and other applications, with identification of individuals conducted either through physical tags (coded-wire tags (CWTs); Jefferts et al. 1963) or via genetic methods (Manel et al. 2005). Each method has had a long history in salmon fisheries management, with CWTs routinely applied to some juveniles released from hatcheries in British Columbia beginning in the 1970s, and with genetic methods applied to salmon fisheries management beginning in the 1980s (Utter et al. 1987). The basic CWT has remained relatively unchanged since its initial application, whereas genetic technology applied to salmon fisheries management has changed substantially during this interval (Shaklee et al. 1999; Seeb et al. 2011; Campbell et al. 2015).

In some hatcheries in British Columbia, a portion of the juvenile Chinook Salmon (*Oncorhynchus tshawytscha*) production is marked prior to release with CWTs, and which are later detected in returning adults by an electronic tag detection system applied to individual salmon in fisheries, hatcheries, or on the spawning grounds. Once the CWT is recovered, the tag is decoded and the origin and age of the individual can be determined. Initially, salmon marked with CWTs upon hatchery release also received an adipose fin clip, with this externally-visible mark allowing CWT-marked fish to be identified visually and sampled in a fishery. However, since the 1990s, in order to facilitate fisheries that exploited Chinook Salmon produced only in hatcheries, most Chinook Salmon released from many hatcheries in Washington, Oregon, and the Columbia River drainage received an adipose fin clip, but no corresponding CWT. This approach enabled mark-selective fisheries to be conducted, in which only adipose-clipped hatchery fish were harvested, but it impaired the recovery of CWTs, as now many adipose fin-

clipped individuals did not carry a CWT. The necessity of maintaining a viable CWT system for salmon assessment was recognized under the Pacific Salmon Treaty between Canada and the United States through a Memorandum of Understanding. In 2004, given the impairment of CWT recovery, the Pacific Salmon Commission (PSC) convened an expert panel to examine limitations of the CWT program and to evaluate the capacity of alternative technologies to provide data to improve assessment of Chinook Salmon. One finding of the panel was that a parentage-based tagging (PBT) technique would provide the equivalent of CWT recovery data, that is, hatchery of release and age of the sampled individual, and could be easily integrated with a genetic stock identification (GSI) system to provide stock of origin for all fish from PBT hatcheries (PSC 2005). However, the panel noted that an empirical demonstration was needed to validate theoretical PBT results that suggested broad feasibility.

PBT (Anderson and Garza 2006) provides equivalent information as CWTs. In PBT applications, the entire hatchery broodstock is genotyped annually, and subsequently progeny can be assigned back to their parents using parentage analysis (Anderson 2012; Wang 2016), thus identifying their hatchery of origin and brood year (i.e. age). When genotypes are available from all individuals in the broodstock, then every offspring is genetically “tagged” which is equivalent to a 100% CWT mark rate, higher than the current approximately 10% CWT tagging rates of juvenile Chinook Salmon released from selected hatcheries in British Columbia.

With the advent of genotyping by sequencing (GBS) techniques, rapid single nucleotide polymorphism (SNP) genotyping for many individuals at potentially hundreds of SNPs provides the capability of rapid, cost-effective genotyping of individuals through direct sequencing of amplicons (Campbell et al. 2015). Genotyping of hundreds of SNPs in a cost-effective manner

81 potentially increases the power of PBT for assignment of individuals in complex marine mixed-
82 stock fishery samples.

83 An initial evaluation of the equivalency of CWT and PBT recovery information was
84 reported by Steele et al. (2013) who demonstrated that for steelhead trout (*O. mykiss*) in the
85 Snake River basin in the Columbia River drainage, stock assignments made through PBT with a
86 panel of SNPs matched those made using CWTs, empirically confirming the equivalency of the
87 techniques in this application. This approach was expanded by Hess et al. (2016), who used both
88 PBT and GSI to investigate run timing of Steelhead Trout in the upper Columbia River drainage.
89 These applications confirmed the practicality of using PBT and GSI for stock identification and
90 the equivalency of CWT recovery data, but they were limited in geographic scale and focused on
91 Steelhead Trout, a species with limited CWT application in comparison with Chinook and Coho
92 Salmon. Beacham et al. (2017) demonstrated, on a coastal scale, that PBT and GSI could be
93 applied to Coho Salmon in British Columbia to provide the equivalent of CWT recovery data for
94 populations in British Columbia where CWTs are currently applied, as well as for additional
95 hatchery populations where no CWTs are applied.

96 Chinook Salmon is the most important Pacific salmon species in terms of CWT
97 application, and was the species of most concern to the expert CWT panel (PSC 2005). The
98 challenge would be to demonstrate empirically, on a coastal scale, that a PBT or a PBT/GSI
99 approach could provide the equivalent of CWT recovery data. All juvenile Chinook Salmon that
100 are adipose fin clipped upon release from major production hatcheries in British Columbia are
101 also marked with a CWT, as there is currently no mass marking (adipose fin clip) of hatchery
102 production in British Columbia. If adipose fin clipped individuals are sampled in fisheries in
103 British Columbia, will it be possible to emulate the results available from the current CWT

program for British Columbia production hatcheries? Evaluation of this question provided the impetus for the present study.

In the present study, we evaluated whether a combined PBT/GSI approach can emulate and perhaps improve upon the results available from the current CWT program for Chinook Salmon in British Columbia production hatcheries, with a British Columbia baseline employed for potential population assignment. In essence, we evaluated whether we could correctly identify the origin and age of one-year old, two-year old, and three-year old Chinook Salmon either rearing in or returning to hatcheries in British Columbia where a portion of the production is marked with CWTs upon release from the hatchery. Age at maturity in Chinook Salmon in British Columbia (ages 1 to 6 years) is more variable than that of Coho Salmon (2 to 4 years), and thus would present the largest challenge for Pacific Salmon in correct assignment of individuals via PBT. However, as initiation of genotyping of hatchery broodstock was initiated in 2013 for the current study, only returning one-year old and two-year old individuals could be assigned via PBT in 2015, and one-year old, two-year old, and three-year old individuals assigned via PBT in 2016. We conducted the study by applying GBS technologies to genotype Chinook Salmon at 321 SNPs in 321 amplicons, and with a stock identification baseline comprising up to 45 populations (hatchery and wild populations) and 36,241 individuals genotyped. We evaluated whether we could correctly identify the origin and age of individual Chinook Salmon through a combined PBT/GSI approach that incorporated a baseline of populations where CWTs are or may be applied to juveniles upon hatchery release in British Columbia and that these individuals may contribute to marine mixed-stock samples in fisheries. A case for application of the genetic approach rests upon demonstration of the potential ability to identify individuals to the correct population concurrent with the correct age assignment of

individuals, and that task was the focus of the current study. We provide evidence that at least for one-year-old, two-year-old, and three-year-old individuals, this task was accomplished for populations in British Columbia.

Methods and Materials

Sample collection

The hatchery populations were sampled in 2013, 2014 and 2015, where the objective was complete broodstock sampling at hatcheries in British Columbia where a portion of the juveniles was adipose fin clipped and coded-wire tagged upon release. Additionally, samples of adipose fins of juveniles that were clipped as part of CWT marking were obtained in 2016 from six hatcheries (seven populations), as these individuals were the offspring of the hatchery broodstock sampled in 2015. These juveniles constituted samples of known origin and age, and were subsequently used in evaluation of population and broodyear assignment. In 2015, samples were also collected from individuals returning to hatcheries visually identified as putative “jacks” (age 2) or “jimmies” (age 1) based upon their body size. These individuals could have been adipose fin clipped (hatchery origin) or unclipped (hatchery or natural spawning origin). In 2016, samples were collected from adipose fin clipped individuals returning to hatcheries that were visually identified as putative age 3, jacks, or jimmies based upon their body size. In sampling of hatchery escapements, we targeted the smaller individuals, as only jacks and jimmies could be assigned via PBT in 2015 (parents in 2013 and 2014 broodstocks), and age 3, jacks, and jimmies in 2016 (parents in 2013, 2014, and 2015 broodstocks). We also assumed that most individuals would return to their natal hatcheries and river drainages. However, if straying did occur, we evaluated whether it was possible to validate straying estimated via PBT with recovered CWTs. The juveniles sampled at the hatcheries in 2016, along with the smaller

individuals sampled in the hatchery escapements in 2015 and 2016, provided the basis for evaluation of the potential ability to identify individuals to the correct population concurrent with the correct age assignment of individuals. Additionally, samples were collected opportunistically from spawning populations on the west coast of Vancouver Island when PBT was not the objective of sampling. The list of hatchery broodstocks or naturally-spawned populations sampled and their locations are listed in Supplementary Table S1 and Figure 1. Fin tissue was obtained from all individuals sampled.

Library preparation and genotyping

The detailed procedure for library preparation and genotyping was outlined by Beacham et al. (2017). Summarized briefly, DNA was extracted from *O. tshawytscha* tissue samples and the DNA concentration normalized to 40 ng/μL with a Tecan LiHa robot. The initial multiplex PCR amplification of 321 target amplicons was conducted with a cocktail of 2μL of normalized DNA extract, 5μL of 2X Ion Agriseq primer pool, 2μL of Ion Agriseq HiFi mix, and 1μL of ddH₂O. Thermal cycling was conducted in 96-well PCR plates (one individual per well) with the following conditions for PCR: 99°C – 2m; 17 cycles [99°C – 15s, 60°C – 4m]; 10°C hold. Primer sequences for each amplicon are outlined in Supplementary Table S2. Primers were developed from published sequence data for Chinook Salmon (Campbell and Narum 2008; Clemento et al. 2011; Warheit et al. 2013; Hecht et al. 2015). Additionally, a panel of primers developed for Coho Salmon (*O. kisutch*) from published DNA sequence sources (Smith et al. 2006; Campbell and Narum 2011; Starks et al. 2016; N. Campbell, unpub.) was applied to Chinook Salmon in our laboratory in order to detect additional SNPs. Primers for scorable SNPs from the Coho Salmon panel were subsequently incorporated into the panel that was applied to Chinook Salmon.

173 Following the initial PCR, a second step employing a thermal cycler was conducted that
174 partially digested the primers on the amplicons, and the reaction conducted with the following
175 conditions: 50°C – 10m; 55°C – 10m, 60°C – 20m, 10°C hold. A third and final step employing
176 a thermal cycler was initiated to ligate the barcodes (384 individual codes) to the amplicons, and
177 was conducted with the following conditions: 22°C – 30m; 70°C – 10m, 10°C hold. Libraries
178 were purified by addition of 22.5µL of Agencourt® AMPure® XP magnetic beads to each
179 library, the plate was placed on a magnetic rack, supernatant discarded, and the beads washed
180 twice in 70% ethanol. The purified libraries were then eluted with 25µL of low TE, and 20 µL of
181 the supernatant transferred to a fresh 96-well tray. Next, each of the 384 prepared libraries was
182 pooled into a single tube for processing on the Ion Chef® (Thermo Fisher Scientific). Two
183 tubes of pooled libraries were processed consecutively on the Ion Chef, and thus 768 individuals
184 were processed on a single run of the Ion Chef. One tube of the pooled libraries was loaded on
185 to each P1® chip v3, and thus amplicons from 384 individuals were distributed on each P1 chip,
186 with 768 individuals processed between two chips. The chips were then loaded on to the Ion
187 Torrent Proton sequencer. After the sequencing run was completed, comparisons with the
188 reference genome of the Rainbow Trout (*O. mykiss*) (Berthelot et al. 2014), supplemented with
189 the sequences containing the observed Chinook Salmon SNPs, were conducted with Proton
190 software Variant Caller®, and SNP genotypes at the sites specified by the hotspot file within
191 target regions were called by Variant Caller. The hotspot file contained 321 SNP sites (254
192 amplicons from Chinook Salmon-origin primers, 67 amplicons from Coho Salmon-origin
193 primers), with one SNP scored at each amplicon. Genotypes at all available SNPs for an
194 individual were assembled to provide a multi-locus individual genotype. The species

identification SNP *OkiOts_120255-113* (Starks et al. 2013) and sex identification SNP *Ots_SEXY3-1* were omitted from subsequent parentage and GSI analyses, leaving 319 SNPs.

Data analysis

F_{st} was estimated via the R software package *pegas* (Paradis 2010) and was used to estimate genetic differentiation among all populations, with 319 SNPs incorporated in the analyses. Expected heterozygosity by locus was estimated via the R software package *ade4* (Jombart 2008; Jombart and Ahmed 2011). An unrooted neighbor-joining tree based upon F_{st} differentiation was generated using *TreeFit* (Kalinowski 2009). Bootstrap support for the major nodes in the tree was evaluated based upon 100 replicate trees.

Assignments of individuals to specific populations were conducted with two methods. The first method is PBT where the genotypes of individuals to be identified are matched to the genotypes of prospective parents (SNPPIT, Anderson 2012; COLONY, Jones and Wang 2010, Wang 2016). If all individuals in a hatchery broodstock are sampled and subsequently genotyped, then all offspring from the broodstock are genetically marked. The genotypes of individuals of unknown origin are statistically compared with the genotypes of potential parents, and if a match is made, the offspring are assigned to the parents, and thus the origin and age of the fish are determined. Genotypes of both parents must be available for assignments to be made with SNPPIT. SNPPIT allows different populations to be input as broodstock, requires the year of collection of potential parents to be included in the input files, and allows the sex of some individuals to be undefined. In order to evaluate the power of SNPPIT to deliver correct assignments for a geographically diverse collection of potential broodstock populations, individuals in all collection years within a population were included as potential parents for all individuals assigned with SNPPIT. All 36,241 individuals sampled in baseline populations in

the study were available for potential parental assignment. Similar to the technique outlined by Hess et al. (2016), a threshold value of ≥ 11 was chosen for likelihood of difference (LOD) to minimize simultaneously the false positive and false negative assignments based upon the known-origin and known-age samples available from populations in British Columbia. Individuals with more than 120 missing genotypes were eliminated from further analyses. 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. An estimated genotyping error rate of 1% or a per allele rate 0.5% was used for SNPPIT assignments. Higher genotyping error rates were not possible to incorporate with our dataset, as SNPPIT analyses would not complete successfully with higher rates.

Additional parentage assignment software was utilized (COLONY version 2.0.6.2, Jones and Wang 2010; Wang 2016) to assign offspring to parents, as COLONY can produce assignments when the genotype of one of the parents is missing, either due to a missing parental sample, or failure to produce a parental genotype from an existing sample. COLONY was run with all potential broodstock input as a single unit, with no differentiation among populations or broodyears within populations. Two-parent assignments were accepted only when both assigned parents originated from the same population and same sampling year, otherwise the individual was passed to genetic stock identification (GSI) for potential assignment. Two-parent and single-parent assignments were accepted only when the probability of correct assignment was ≥ 0.90 , otherwise the individual was passed to potential assignment by GSI. Individuals for which no prospective parents were identified in the broodstock were passed to GSI for potential assignment. Polygamous mating was assumed for the COLONY analysis. Simple pairwise

comparisons between offspring and potential parents were conducted. Prior to the 2013 broodstock sampling, the sex of the individual sampled was usually not available. However, COLONY requires that the sex of the prospective parental individuals be known. Therefore, for individuals where the sex was unknown, we assumed that all individuals were female in order to allow them to be available for potential single-parent assignments. In fact, no age 1, 2, or 3 individual assignments were ever made to parents sampled prior to the initiation of hatchery broodstock genotyping in 2013, so this conservative assumption had no effect on results, other than allowing the individuals sampled prior to 2013 to be available for potential parentage assignment. The baseline for juveniles sampled in 2016 included all broodstocks sampled in 2015 and previous years. Analysis of putative jacks and jimmies sampled in 2015 included all broodstocks sampled in 2014 and previous years, and analysis of putative age 3, jacks, and jimmies sampled in 2016 included all broodstocks sampled in 2015 and previous years. An estimated genotyping error rate of 1% was used for COLONY assignments.

The second method of individual identification is GSI, in which the genetic profiles of whole populations potentially contributing to a mixed-stock sample are used to estimate the stock composition of the sample, and in some instances estimate the origin of each individual in the sample (GSI-sim, Anderson et al. 2008). For each sample, individuals not assigned by SNPPIT or COLONY were then potentially assigned with GSI-sim, with the origin of individuals assigned with a probability of < 0.85 (assignment to population) classified as undetermined.

Results

Population structure

A total of 36,241 individuals from 45 populations was genotyped at 319 SNPs used in assessment of population structure and individual assignment. The average coverage rate over all SNPs was 259 X (range 28 X to 700 X over 319 SNPs), with a median coverage per SNP of 212 X. In practice, our no-call rate (genotypes were not determined at all SNP sites for all individuals) was about 1.4% over all 319 SNPs (one SNP per amplicon, range 1.2% to 14.0% over 319 SNPs). Average heterozygosity over all SNPs was 0.30 (range 0.00 to 0.50), and average F_{st} over all SNPs was 0.07 (range 0.01-0.30) (Supplementary Table S2).

Accuracy of assignment via GSI hinges upon a regional population structure, particularly if individuals in mixed-stock samples are derived from unsampled populations. Accordingly, population structure was evaluated, and regional structure was observed in the 45 populations analyzed (Figure 2). Populations in southern British Columbia, Vancouver Island, and the Fraser River clustered in regional groups. Populations on the east coast of Vancouver Island were well differentiated from those on the west coast of Vancouver Island and the Fraser River. Within the east coast of Vancouver Island region, the most northern populations sampled (Quinsam River, Campbell River) were differentiated from those further south, with 100% bootstrap support on the dendrogram. In the Fraser River drainage, populations in the South Thompson River tributary (lower Shuswap River, middle Shuswap River; 100% bootstrap support) and lower Thompson River tributary (Nicola River, Coldwater River, and Spius Creek) were differentiated from other populations sampled in the drainage.

Assignment by PBT and GSI

Genotypes were available from 656 one-year-old Chinook Salmon of known origin from seven populations. A total of 82.0% (538/656) of the individuals was subsequently assigned via

SNPPIT. All assignments via SNPPIT were 100% accurate for both population of origin and age, with a baseline of 45 populations and 36,241 individuals with multiple age classes within populations available for assignment. In total, 94.1% (617/656) of the individuals were correctly assigned via a combination of SNPPIT and GSI (Table 1). The unassigned individuals failed to meet the probability of correct assignment level of ≥ 0.85 . Juveniles sampled at the Chehalis River hatchery originated from two broodstocks (Harrison River and Chilliwack River), with assignments to either population considered correct.

Assignments of known-origin and known-age individuals via COLONY resulted in assignments of 636 of 656 individuals, with an accuracy of 99.8% (635/636) to population and 100.0% to age (Table 1). The single error observed in assignment was a Cowichan River population juvenile assigned via single parentage analysis to a 2015 male in the Big Qualicum River population broodstock. All individuals assigned via two parents were assigned with 100% accuracy. In total, 99.2% (651/656) of the individuals were correctly assigned via a combination of COLONY and GSI.

2015 jack assignments

Putative jacks were targeted in the 2015 hatchery broodstock sampling, and it was expected that assignment via PBT should assign these individuals to parents sampled from the hatchery broodstock during 2013. Assignments other than as expected were checked for validation with CWTs (location and age) or scales (age), with validations outlined in Table 2. Overall, of the 466 putative jacks sampled, 48.7% (227/466) of the individuals were assigned via SNPPIT analysis. With COLONY, 83.9% (391/466) of the individuals were assigned, with 16 individuals identified as jimmies. For the SNPPIT age 1 assignments, one individual from the

Quinsam River was confirmed to be age 1 through CWT recovery, with the assignment of a second unclipped individual (Quinsam River) assumed correct (Table 2). For the COLONY age 1 assignments, seven individuals were confirmed age 1 through CWT recovery, with an additional nine unclipped individuals identified as jimmies. Of these nine putative jimmies, six individuals were observed in the Quinsam River population. Of these six individuals, three individuals were identified as Quinsam River in origin via both COLONY ($P=1.000$) and GSI-sim ($P=1.000$) and were considered assigned correctly. Two unclipped individuals were assigned to the Lower Shuswap River population via both COLONY ($P=1.000$) and GSI ($P=1.000$), but no judgment was made as to accuracy of assignment (Table 2). The final unclipped jimmie at Quinsam River was identified as Chilliwack River in origin via single-parent assignment in COLONY ($P=1.000$), but Lower Shuswap River in origin via GSI-sim ($P=1.000$), and no judgment was made as to accuracy of assignment (Table 2). In the Big Qualicum River population, two unclipped individuals (potentially jimmies) were identified as Big Qualicum River in origin via single-parent analysis in COLONY ($P=1.000$), but possibly Puntledge River in origin via GSI-sim ($0.51 < P < 0.90$), and no judgment was made as to accuracy of assignment. In the Sarita River population, one unclipped individual (potential jimmie) was identified as Robertson Creek in origin via single parent analysis in COLONY ($P=0.9558$), but Thornton Creek in origin via GSI-sim ($P=1.000$), and no judgment was made as to assignment accuracy.

Straying by jacks between river drainages was observed. One Puntledge River origin jack was confirmed to have strayed to the Big Qualicum River through CWT recovery, and this was also indicated to have occurred through parentage analysis via both SNPPIT and COLONY. Additional straying of jacks was also estimated to have occurred at the Quinsam and Big

Qualicum River hatcheries as assignments were made through parentage analysis, but these assignments could not be confirmed by CWTs as the individuals were not adipose fin clipped. One assignment made by COLONY considered to be in error was an adipose fin-clipped jack sampled at the Big Qualicum hatchery and assigned via single-parent analysis to the 2013 Chilliwack River broodstock, yet no Chilliwack-origin CWT was associated with the jack returns at the hatchery.

At the Puntledge River hatchery, individuals are putatively classified as either “summer” or “fall” based upon the timing of their return. Of the 13 summer jacks sampled in 2015, 84.6% (11/13) were assigned to summer parents sampled during 2013 via SNPPIT and COLONY, and two individuals were assigned to the summer population via GSI (Table 2). Of the 124 fall jacks sampled, 61.3% (76/124) were assigned to fall parents and 21.0% (26/124) to summer parents via SNPPIT. With COLONY, 68.5% (85/124) of the fall jacks and 100% of the jimmies (3/124; age confirmed by CWT recovery) were assigned to Puntledge River fall parents, with an additional 21.8% (27/124) assigned to summer parents, and with 0.8% (1/124) of the returning jacks unassigned either through COLONY or GSI.

2016 jack returns

Smaller adipose fin-clipped individuals were targeted in the 2016 hatchery broodstock sampling, and it was expected that assignment via PBT should assign these individuals to parents sampled from the hatchery broodstock during 2013 (age 3), 2014 (jacks), or 2015 (jimmies). Similar to the 2015 returns, assignments through PBT other than as expected were checked for validation with CWTs (location and age) or scales (age), with validations outlined in Table 3. Returning sampled individuals at the Big Qualicum River included both age 2 and age 3

352 individuals, with the majority age 2 (Table 3). Four jacks from the Cowichan River were also
353 observed, with the assignments made through both SNPPIT and COLONY, and with the
354 assignments confirmed with CWT recoveries. COLONY assigned via single-parent analysis one
355 individual as a Puntledge River jimmie, but as this assignment was unconfirmed by a CWT
356 recovery, the assignment was considered incorrect. All assignments of Chilliwack River “jacks”
357 were considered to be correct.

358 CWTs were not applied to juveniles from the Nitinat River and Sarita River populations,
359 but sampling of jack or jimmie returns still provided information on assignment of individuals.
360 The broodstock was not completely genotyped for the Nitinat River population, but when jacks
361 and age 3 individuals were assigned via PBT, all individuals (except one jimmie) were assigned
362 to the Nitinat River population (Table 3). The single individual not assigned to the Nitinat River
363 population was an individual identified as a Sarita River jimmie through both SNPPIT and
364 COLONY assignments, and was considered to be correctly assigned. When all returning jacks
365 sampled at the Sarita River population were assigned via PBT, all individuals were assigned to
366 parents in the 2015 broodstock by both SNPPIT and COLONY, identifying them as jimmies
367 which was confirmed by scale analysis.

368 Smaller individuals returning to the Quinsam River were primarily identified as age 3 via
369 COLONY assignments, but age 1 and age 2 individuals were also observed, with the ages
370 confirmed by recovery of CWTs. At the Puntledge River, for summer jacks, 91.2% (52/57) of
371 the individuals were assigned to summer parents sampled during 2013 or 2014 via SNPPIT, and
372 one individual was assigned to the fall population (Table 3). Similar results were observed with
373 the COLONY assignments, with 96.5% (55/57) assigned to summer parents, and one individual
374 assigned to fall parents. Of the 85 fall jacks sampled, 48.2% (41/85) were assigned to fall

parents and 24.7% (21/85) to summer parents via SNPPIT. With COLONY, 63.5% (54/85) of the fall jacks were assigned to fall parents, and 24.7% (21/85) to summer parents. Both SNPPIT and COLONY assigned four fall jacks to Cowichan River parents, and these assignments were verified by recoveries of CWTs. Overall, of the 560 putative jacks sampled, 58.8% (329/560) of the individuals were assigned via SNPPIT analysis with apparent 100% accuracy with respect to population and age. For COLONY, 90.5% (507/560) of the individuals were assigned with an accuracy of 99.8% with respect to population and age.

Discussion

Determination of accurate origin and age of sampled individuals in mixed-stock fishery samples through genetic analysis is of paramount importance if a genetic method can complement or replace the current CWT method of assessment. Hess et al. (2016) suggested that when PBT is combined with GSI, there is the potential to achieve the most complete and accurate information for both hatchery-origin and wild-origin individuals when sampled in mixed-stock fisheries. They noted that nearly 100% accuracy for PBT assignments is achieved when applied (e.g. Steele 2013; Beacham et al. 2017). This level of accuracy can have many applications other than analysis of mixed-stock fishery samples, ranging from PBT being increasingly applied in both natural and hatchery settings (Sekino et al. 2005; Denson et al. 2012; Ashton et al. 2016), and has provided valuable information on individual reproductive success, effective population size, and proportion of hatchery-origin individuals in salmon escapements (Abadia-Cardosa et al. 2013; Ford et al. 2015, Hinrichsen et al. 2016).

Evaluation of the genetic method of assignment in the current study hinged upon being able to demonstrate that when assignments of individuals were made, they were accurate with

respect to population of origin and age. As DNA profiles are invariant over an individual's life, the demonstrated ability to identify juveniles, jimmies, jacks, and age 3 individuals to the correct hatchery and broodyear should translate directly to the ability to identify individuals in subsequent mixed-stock fishery samples. The juveniles sampled in 2016 prior to release from the hatchery were of known origin and age. If all parents were sampled and genotyped in the broodstock, and genotyping errors of both parents and juveniles were within the 1.00% per SNP limits used for parentage analysis, then 100% of the juveniles should have been assigned to the correct hatchery and the correct broodyear. In the current study, the high accuracy obtained from the assignment of the juveniles indicated that, when they were assigned via parentage analysis, it was possible to assign individuals both with respect to population and broodyear within population with confidence.

Juveniles sampled at the Chehalis River hatchery in 2016 were assigned to parents from both the Harrison River and Chilliwack River populations. Broodstock for the hatchery is normally obtained from the Harrison River, and thus assignments to the Harrison River population would be expected. Prior to the sampling of juveniles during 2016, 100,000 fry from the Chilliwack River hatchery were transferred to the Chehalis River hatchery on April 18, 2016. Hatchery records indicated that these 100,000 fry originated from parents spawned during the October 20, 2015 egg take at the Chilliwack River hatchery. Assignments to the Chilliwack River population via parentage analysis corresponded to broodstock spawned on October 20, 2015, and thus assignments to the Chilliwack River population were considered to be correct.

In 2015, returning putative jacks were sampled at hatcheries under the expectation that age 2 year returning individuals should be available to be assigned to parents in the 2013 broodyear, the initial year of large-scale genotyping of hatchery broodstock. In practice, there is

an overlap in size between age 2 and age 3 Chinook Salmon, so sampling based upon the visual size of the individual may have included some age 3 individuals that could not be assigned via PBT. Jimmies were also observed in the putative jacks sampled at Quinsam River and Puntledge River hatcheries, with one adipose fin-clipped jimmie confirmed with the recovery of one CWT age 1 year individual at the Quinsam River hatchery, and with three clipped individuals (CWTs also recovered) at the Puntledge River hatchery.

Straying of adipose fin-clipped jacks at some of the east coast of Vancouver Island hatcheries was observed via recovery of CWTs, a result that was also observed via assignments through parentage analysis. In addition to the straying confirmed by the recovery of adipose fin clipped CWT individuals, additional straying was observed by assignment via parentage analysis of unclipped individuals to populations other than those in which they were recovered. For example, jacks of Chilliwack River and Big Qualicum River origin were observed in the Quinsam River, and a jack of Little Qualicum River origin was observed in the Big Qualicum River. Analysis of CWT recoveries had previously indicated that Chinook Salmon can stray from the lower Fraser River into east coast Vancouver Island (ECVI) rivers, as well as stray among ECVI rivers (Candy and Beacham 2000), so assignments of the unclipped individuals to non-natal populations were possible.

Chinook Salmon returning to the Puntledge River have been defined as summer-run for those individuals returning prior to August 1st, and fall-run for individuals returning later (L. Terry, DFO, pers. comm.). The fall-run population in the drainage was supplemented with a series of transplants, primarily from the Little Qualicum River (1985-2000), Big Qualicum River (1985-2000), and Quinsam River (1984-1992) populations (C. Lynch, DFO Salmonid Enhancement Program, pers. comm.). If juveniles were assigned via parentage analysis, then all

known-origin juveniles were assigned to the appropriate timing group. Parentage analysis also assigned summer-run jacks to mainly (one case of assignment to fall-run parents) summer-run parents, but jacks defined as fall-run based upon timing were assigned to both summer-run and fall-run parents. For those individuals unassigned by parentage analysis but subsequently assigned via GSI, all summer-origin jacks were correctly assigned to the summer population. Fall-run jacks were assigned via GSI to the fall-run population, the Big Qualicum River population, and the Little Qualicum River population, but with 54% unassigned to population. With COLONY, only 14 jacks remained unassigned following parentage analysis, and of these 14 jacks, nine individuals were assigned to the Puntledge River fall population, one individual to the Big Qualicum River population, and five individuals remained unassigned via GSI. The timing of the return of summer-run jacks indicated that 42% of the jacks returned during the period defined as the fall-run return, so it may be that summer-run jacks return later than the rest of the summer-run population.

We undertook the current study to evaluate whether a genetic-based stock identification system could likely identify individuals that may be sampled in highly mixed-stock ocean fisheries, in which there could be a large number of geographically-diverse contributing populations. We undertook the evaluation by assigning smaller individuals sampled in hatchery escapements, as well as juveniles that were adipose fin clipped as part of being marked with CWTs, and assessing the reliability of the assignments made both in terms of hatchery of release and age of the individual. In British Columbia, adipose-clipped Chinook Salmon can mature either at ages 1-6 years. Although a combined PBT/GSI approach has been outlined in our study, higher accuracy of assignment was observed through PBT than with GSI. GSI has the advantage of assigning individuals unassigned by PBT but, as there is more uncertainty

associated with GSI-based assignments than with PBT-based assignments, one could choose to accept only PBT-based assignments in an application.

Beacham et al. (2017) outlined a number of reasons that a PBT/GSI or PBT system of identification could provide an alternate method of identification in the assessment and management of Canadian-origin Coho Salmon relative to the existing CWT program. These included higher “marking” rates using genetic tags, ability to identify unclipped individuals, non-lethal sampling requirements, ability to identify the sex and origin of both marked and unmarked individuals, and ability to construct pedigrees. A similar rationale exists with respect to Chinook Salmon with respect to implementation of a PBT/GSI system of identification, coupled with cheaper operating costs compared with CWTs (Supplementary Table S3). The ability to identify Chinook Salmon to specific Canadian hatcheries provides an opportunity for a genetic-based system to enhance or replace the current CWT system for Coho Salmon in British Columbia. Satterthwaite et al. (2015) explored various scenarios under which PBT could be expanded to a coast-wide application for both Chinook Salmon (*O. tshawytscha*) and Coho Salmon.

If stock composition of adipose-clipped Chinook Salmon in a mixed-stock fishery sample is required, we have the ability to identify the contributions of specific hatcheries in southern British Columbia with a high degree of accuracy. Given the strong regional population structure observed in Chinook Salmon with the SNP baseline, standard GSI techniques could be applied to include unclipped individuals as well, and we expect that estimated stock compositions through GSI with the SNP panel could be available from smaller geographical regions of origin or reporting groups than is currently available from microsatellite analysis (Beacham et al. 2006) or other SNP baselines (Larson et al. 2012; Clemento et al. 2014).

488 Acknowledgments

489 A very substantial effort was undertaken to obtain samples from Chinook Salmon
490 sampled in this study. In southern British Columbia, we thank various staff of the Salmon
491 Enhancement Program of Fisheries and Oceans Canada (DFO), including hatchery managers and
492 staff, C. Lynch, and D. Willis for sample collection and coordination. Dr. S. R. Narum and N. R.
493 Campbell of the Columbia River Inter-Tribal Fish Commission provided valuable information on
494 DNA sequence information for many of the SNPs surveyed in the study. Critical assistance was
495 provided by Thermo Fisher Scientific staff in project execution, including Dr. L. Chen for
496 analytical support, Dr. C. Scafe for bioinformatics, Dr. M. Andersen who suggested trying Coho
497 Salmon-origin primers on Chinook Salmon and provided laboratory procedure advice, C. Adams
498 who provided early access to the barcodes, and D. Bong for project supplies. Dr. E. C. Anderson
499 of the National Marine Fisheries Service Southwest Fisheries Center graciously modified the
500 code for both SNPPIT and GSI-sim to allow the analysis to proceed. Partial funding support was
501 provided by the Genomics Research and Development Initiative of Fisheries and Oceans
502 Canada.

503

504 References

505 Abadia-Cardosa, A., Anderson, E.C., Pearse, D.E., and Garza, J.C. 2013. Large-scale parentage
506 analysis reveals reproductive patterns and heritability of spawn timing in a hatchery
507 population of steelhead (*Oncorhynchus mykiss*). *Mol. Ecol.* 22: 4733-4746.
508 Anderson, E.C. 2012. Large-scale parentage inference with SNPs: an efficient algorithm for
509 statistical confidence of parent pair allocations. *Statistical Applications in Genetics and*

- 510 Molecular Biology Volume 11, Issue 5, ISSN (Online) 1544-6115, DOI: [10.1515/1544-](https://doi.org/10.1515/1544-6115.1833)
511 [6115.1833](https://doi.org/10.1515/1544-6115.1833), November 2012
- 512 Anderson, E.C., and Garza, J.C. 2006. The power of single-nucleotide polymorphisms for
513 large-scale parentage inference. *Genetics*, 172(4):2567–2582. doi:
514 10.1534/genetics.105.048074.
- 515 Anderson, E.C., Waples, R.S. and Kalinowski, S.T. 2008. An improved method for predicting
516 the accuracy of genetic stock identification. *Can. J. Fish. Aquat. Sci.* 65:1475-1486.
- 517 Ashton, N.K., Campbell, M.R., Anders, P.J., Powell, M.S., and Cain, K.D. 2016. Evaluating
518 microsatellite markers for parentage-based tagging of hatchery burbot. *Northwest Sci.*
519 90: 249-259.
- 520 Beacham, T.D., Candy, J.R., Jonsen, K.L., Supernault, J., Wetklo, M., Deng, L., Miller, K.M.,
521 and Withler, R.E. 2006. Estimation of stock composition and individual identification of
522 Chinook salmon across the Pacific Rim using microsatellite variation. *Trans. Am. Fish.*
523 *Soc.* 135: 861-888.
- 524 Beacham, T.D., Wallace, C., MacConnachie, C., Jonsen, K., McIntosh, B., Candy, J.R., Devlin,
525 R., and Withler, R.E. In press. Population and individual identification of Coho Salmon
526 in British Columbia through parentage-based tagging and genetic stock identification: an
527 alternative to coded-wire tags. *Can. J. Fish. Aquat. Sci.*
- 528 Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva,
529 C., Labadie, K., Alberti, A., Aury, J.M., Louis, A., Dehais, P., Bardou, P., Montfort, J.,
530 Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G.H., Boussaha, M., Quillet, E., Guyomard,
531 R., Galiana, D., Bobe, J., Volff, J.N., Genêt, C., Wincker, P., Jaillon, O., Roest Crollius,

- 532 H., and Guiguen, Y. 2014. The rainbow trout genome provides novel insights into
533 evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5: 3657. doi:
534 10.1038/ncomms4657
- 535 Campbell, N.R., Harmon, S., and Narum, S.R. 2015. Genotyping-in-Thousands by sequencing
536 (GT-seq): A cost effective SNP genotyping method based on custom amplicon
537 sequencing. *Mol. Ecol. Res.* 15:855-867.
- 538 Campbell, N.R., and Narum, S.R. 2008. Identification of novel single-nucleotide
539 polymorphisms in Chinook salmon and variation among life history types. *Trans. Amer.*
540 *Fish. Soc.* 137:96–106.
- 541 Campbell, N.R., and Narum, S.R. 2011. Development of 54 novel single-nucleotide
542 polymorphism (SNP) assays for sockeye and coho salmon and assessment of available
543 SNPs to differentiate stocks within the Columbia River. *Mol. Ecol. Res.* 11(Suppl 1), 20–
544 30.
- 545 Candy, J.R., and Beacham, T.D. 2000. Patterns of homing and straying in southern British
546 Columbia coded-wire tagged chinook salmon (*Oncorhynchus tshawytscha*) populations.
547 *Fish. Res.* 47: 41-56.
- 548 Clemento, A.J., Abadía-Cardoso, A., Starks, H.A., and Garza, J.C. 2011. Discovery and
549 characterization of single nucleotide polymorphisms in Chinook salmon, *Oncorhynchus*
550 *tshawytscha*. *Mol. Ecol. Res.* 11(Suppl. 1): 50–66.
- 551 Clemento, A.J., Crandall, E.D., Garza, J.C., and Anderson, E.C. 2014. Evaluation of a single
552 nucleotide polymorphism baseline for genetic stock identification of Chinook Salmon
553 (*Oncorhynchus tshawytscha*) in the California Current large marine ecosystem. *Fish.*
554 *Bull.* 112: 112-131.

- 555 Denson, M.R., Brenkert IV, K., Jenkins, W.E., and Darden, T.L. 2012. Assessing red drum
556 juvenile stocking in a South Carolina estuary using genetic identification. *N. Am. J. Fish.*
557 *Manage.* 32: 32-43.
- 558 Ford, M., Pearsons, T.N., and Murdoch, A. 2015. The spawning success of early maturing
559 resident hatchery Chinook Salmon in a natural river system. *Trans. Am. Fish. Soc.*
560 144:539-548.
- 561 Hecht, B.C., Matala, A.P., Hess, J.E., and Narum, S.R. 2015. Environmental adaptation in
562 Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range.
563 *Mol. Ecol. Res.* 24: 5573–5595.
- 564 Hess, J.E., Ackerman, M.W., Fryer, J.K., Hasselman, D.J., Steele, C.A., Stephenson, J.J.,
565 Wheteaker, J.M., and Narum, S.R. 2016. Differential adult migration-timing and stock-
566 specific abundance of steelhead in mixed stock assemblages. *ICES J. Mar. Sci.*
567 doi:10.1093/icesjms/fsw138.
- 568 Hess, J.E., Matala, A.P., and Narum, S.R. 2011. Comparison of SNPs and microsatellites for
569 fine-scale application of genetic stock identification of Chinook salmon in the Columbia
570 River Basin. *Mol. Ecol. Res.* 11 (Suppl.1) 1-13.
- 571 Hinrichsen, R.A., Steele, C.A., Ackerman, M.W., Campbell, M.R., Narum, S.R., Hess, M.A.,
572 Young, W.P., Shields, B.A., and Maschhoff, B.L. 2016. Maximum likelihood estimation
573 of the proportion of hatchery-origin fish on spawning grounds using coded wire tagging
574 and parentage-based tagging. *Trans. Am. Fish. Soc.* 145: 671-686.
- 575 Jefferts, K.B., Bergmann, P.K., and Fiscus, H.F. 1963. A coded wire tag identification system
576 for macro-organisms. *Nature (London)* 198:460-462.

- 577 Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
578 Bioinformatics 24:1403-1405.
- 579 Jombart T. and Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP
580 data. Bioinformatics 27:3070-3071.
- 581 Jones, O.W., and Wang, J. 2010. COLONY: a program for parentage and sibship inference
582 from multilocus genotype data. Mol. Ecol. Res. 10:551-555.
- 583 Larson, W.A., Seeb, J.E., Pascal, C.E., Templin, W.D., and Seeb, L.W. 2014. Single-nucleotide
584 polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic
585 stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western
586 Alaska. Can. J. Fish. Aquat. Sci. 71: 698-708.
- 587 Manel, S., Gaggiotti, O.E., and Waples, R.S. 2005. Assignment methods: matching biological
588 questions with appropriate techniques. Trends Ecol. Evol. 20:136-142.
- 589 Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R., and Hohenlohe, P.A. 2013. Genotyping-
590 by-sequencing in ecological and conservation genomics. Mol. Ecol. 22:2841-2847.
- 591 Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular
592 approach. Bioinformatics 26: 419-420.
- 593 PSC. 2005. Expert Panel on the Future of the Coded Wire Tag Program for Pacific Salmon.
594 Pacific Salmon Comm. Tech. Rep. No. 18: 230 p.
595 <http://www.psc.org/publications/workshop-reports/coded-wire-tag-program-review/>
- 596 Satterthwaite, W., Anderson, E., Campbell, M., Garza, J.C., Mohr, M., Narum, S., and Speir, C.
597 2015. Multidisciplinary evaluation of the feasibility of parentage-based genetic tagging

- 598 (PBT) for management of Pacific salmon. Report to the Pacific Salmon Commission,
599 Vancouver.
- 600 Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L.W. 2011. Single
601 nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in
602 nonmodel organisms. *Mol. Ecol. Res.* 11 (Supplement S1): 1-8.
- 603 Sekino, M., Saitoh, K., Yamada, T., Hara, M., and Yamashita, Y. 2005. Genetic tagging of
604 released Japanese flounder (*Paralichthys olivaceus*) based on polymorphic DNA
605 markers. *Aquaculture* 244: 49-61.
- 606 Shaklee, J.B., Beacham, T.D., Seeb, L., and White, B.A. 1999. Managing fisheries using genetic
607 data: Case studies from four species of Pacific salmon. *Fish. Res.* 43: 45-78.
- 608 Smith C.T, Park, L., VanDoornik, D., Seeb, L.W., and Seeb, J.E. 2006. Characterization of 19
609 single nucleotide polymorphism markers for coho salmon. *Mol. Ecol. Notes*, 6: 715–
610 720.
- 611 Starks, H.A., Clemento, A.J., and Garza, J.C. 2016. Discovery and characterization of single
612 nucleotide polymorphisms in coho salmon, *Oncorhynchus kisutch*. *Mol. Ecol. Res.* 16:
613 277-287.
- 614 Steele, C.A., Anderson, E.C., Ackerman, M.W., Hess, M.A., Campbell, N.R., Narum, S.R., and
615 Campbell, M.R. 2013. A validation of parentage-based tagging using hatchery steelhead
616 in the Snake River basin. *Can. J. Fish. Aquat. Sci.* 70: 1046-1054.
- 617 Utter, F.M., Teel, D.J., Milner, G.B., and McIssac, D.. 1987. Genetic estimates of stock
618 compositions of 1983 chinook salmon harvests of the Washington coast and the
619 Columbia River. *Fish. Bull. (US)* 85: 13-23.

620 Wang, J. 2016. Individual identification from genetic marker data: developments and accuracy
621 comparisons of methods. *Mol. Ecol. Res.* 16: 163-175.

622 Warheit, K.I., Seeb, L., Templin, W.D., and Seeb, J. 2013. Moving GSI into the Next Decade:
623 SNP Coordination for Pacific Salmon Treaty Fisheries. Pacific Salmon Commission
624 Report for Chinook Technical Committee (US Section) for Funding under the Letter of
625 Agreement (LOA). 47p.

Draft

Table 1. Accuracy of population assignment and age determination (%) for known-origin samples of Chinook Salmon sampled from seven populations in southern British Columbia during 2016 with individuals assigned through either parentage-based tagging (PBT, primary method) or genetic stock identification (GSI, used for individuals unassigned by PBT), and with a baseline of 36,241 potential parents in 45 populations. N is the number of individuals in each sample assigned through either PBT or GSI. Values in parenthesis in the table footnote are the number of individuals assigned via GSI to the listed population. Juveniles sampled at the Chehalis River hatchery in 2016 originated from parents in the Harrison River and Chilliwack River populations.

		SNPPIT					COLONY				
		PBT			GSI		PBT			GSI	
		% assignment			% assign		% assignment			% assign	
Test population	Age	N	Pop	Age	N	Pop	N	Pop	Age	N	Pop
Big Qualicum	1	64	100.0	100.0	32	50.0 ¹	93	100.0	100.0	3	66.7 ⁸
Quinsam	1	83	100.0	100.0	13	100.0	96	100.0	100.0	0	-
Cowichan	1	76	100.0	100.0	20	70.0 ²	93	98.9	100.0	3	66.7 ⁹
Puntledge summer	1	81	100.0	100.0	9	100.0	87	100.0	100.0	3	100.0
Puntledge fall	1	80	100.0	100.0	10	30.0 ⁴	88	100.0	100.0	2	0.0 ¹⁰
Chilliwack	1	74	100.0	100.0	19	68.4 ⁵	86	100.0	100.0	7	100.0
Chehalis	1										

Harrison	42	100.0	100.0	7	71.4 ⁶	49	100.0	100.0	0	-
Chilliwack	38	100.0	100.0	8	75.0 ⁷	44	100.0	100.0	2	100.0
Summary	538	100.0	100.0	118	66.9	636	99.9	100.0	20	80.0

632 Alternate GSI assignments: ¹ Puntledge (1), unassigned (15); ² unassigned (6); ³ Big Qualicum (1), unassigned (26) ⁴ unassigned (7); ⁵
633 unassigned (6); ⁶ unassigned (2); ⁷ unassigned (2); ⁸ unassigned (1); ⁹ unassigned (1); ¹⁰ unassigned (2)

634

635

636

637

638

639

Table 2. Accuracy of population assignment and age determination (%) for escapement samples of assumed jack Chinook Salmon sampled from seven populations in southern British Columbia during 2015 with individuals assigned through either parentage-based tagging (PBT, primary method) or genetic stock identification (GSI, used for individuals unassigned by PBT), and with a baseline of 36,241 potential parents in 45 populations sampled in multiple years available for potential assignment through both PBT and GSI. N is the number of individuals in each sample assigned through either PBT or GSI. For PBT, assignments with SNPPIT were accepted when the likelihood of difference (LOD) was ≥ 11 . Unassigned individuals were then considered for population assignment via GSI. Assignments with COLONY were accepted when the probability associated with an estimated single parent or parent pair was ≥ 0.90 . For GSI, individuals with an assignment probability of < 0.85 were considered as unassigned in origin. Values in parenthesis in the table footnote are the number of individuals assigned via GSI to the listed population.

Test population	Age	SNPPIT					COLONY				
		PBT			GSI		PBT			GSI	
		% assignment			% assign		% assignment			% assign	
		N	Pop	Age	N	Pop	N	Pop	Age	N	Pop
Chilliwack	2	28	100.0	100.0	72	100.0	67	100.0	100.0	33	100.0
Lower Atnarko											
Lower Atnarko	2	1	100.0	100.0	5	100.0	4	100.0	100.0	1	100.0
Upper Atnarko	2	1	100.0	100.0 ^b			2	100.0	100.0 ^b	-	

Quinsam											
Quinsam	2	13	100.0	100.0	75	90.7 ¹	68	100.0	100.0	15	80.0 ⁶
Quinsam	1	1	100.0	100.0 ^a			1	100.0	100.0 ^a		
Quinsam	1	1	100.0	100.0 ^b			2	100.0	100.0 ^b		
Chilliwack	2	1	100.0	100.0 ^b			1	100.0	100.0 ^b		
Chilliwack	1	0					1	-	- ^e		
Lower Shuswap	1	0					2	-	- ^e		
Big Qualicum	2	1	100.0	100.0 ^b			2	100.0	100.0 ^b		
Big Qualicum											
Big Qualicum	2	48	100.0	100.0	53	49.1 ²	76	100.0	100.0	15	6.7 ⁷
Big Qualicum	1	0	-	-			3	100.0	100.0 ^a		
Big Qualicum	1	0	-	-			2	-	- ^e		
Little Qualicum	2	1	100.0	100.0 ^b			1	100.0	100.0 ^b		
Puntledge	2	1	100.0	100.0 ^a			1	100.0	100.0 ^a		
Cowichan	2	0	-	-			3	100.0	100.0 ^d		
Cowichan	2	0					1	100.0	100.0 ^b		
Chilliwack	2	0	-	-			1	0.0 ^f	100.0		
Sarita											
Sarita	2	17	100.0	100.0	10	50.0 ⁴	25	100.0	100.0	0	-
Robertson	2	0	-	-			1	-	- ^e		

Robertson	1	0	-	-			1	-	- ^e		
Puntledge summer	2	11	100.0	100.0	2	100.0	11	100.0	100.0	2	100.0
Puntledge fall											
Puntledge fall	2	76	100.0	100.0	22	40.9 ⁵	85	100.0	100.0	9	88.9 ⁸
Puntledge fall	1	0	-	-			3	100.0	100.0 ^a		
Puntledge summer	2	26	100.0	100.0			27	100.0	100.0		
Summary		227	100.0	100.0	239	81.7	391 ⁹	99.8	100.0	75	84.3

^a CWT also recovered; ^b Individuals were unclipped, assignment assumed correct, $\text{LOD} \geq 11$ or $P > 0.90$; ^c Age confirmed through scale

analysis, ^d Two Cowichan CWTs recovered, ^e single-parent assignment, unclipped individual, unable to evaluate accuracy of

assignment, ^f clipped individual, no Chilliwack CWT recovered.

Alternate GSI assignments: ¹ Lower Shuswap (4), unassigned (3); ² Cowichan (3), Chilliwack (1), unassigned (23); ³ Sooke (1), Sarita

(1), unassigned (8); ⁴ Nitinat (2), unassigned (3); ⁵ Big Qualicum (3), unassigned (10); ⁶ Lower Shuswap (1), unassigned (2); ⁷

unassigned (13); ⁸ unassigned (1); ⁹ Actual sample size is 384 individuals, as seven individuals have been excluded as they were

unclipped and there were discrepancies between PBT and GSI assignments, with GSI assigning to a different population or individuals

were unassigned via GSI.

658 Table 3. Accuracy of population assignment and age determination (%) for escapement samples of assumed jack Chinook Salmon
 659 sampled from six populations in southern British Columbia during 2016 with individuals assigned through either parentage-based
 660 tagging (PBT, primary method) or genetic stock identification (GSI, used for individuals unassigned by PBT), and with a baseline of
 661 36,241 potential parents in 45 populations sampled in multiple years available for potential assignment through both PBT and GSI. N
 662 is the number of individuals in each sample assigned through either PBT or GSI. For PBT, assignments with SNPPIT were accepted
 663 when the likelihood of difference (LOD) was ≥ 11 . Unassigned individuals were then considered for population assignment via GSI.
 664 Assignments with COLONY were accepted when the probability associated with an estimated single parent or parent pair was ≥ 0.90 .
 665 For GSI, individuals with an assignment probability of < 0.85 were considered as unassigned in origin. Values in parenthesis in the
 666 table footnote are the number of individuals assigned via GSI to the listed population.

		SNPPIT					COLONY				
		PBT			GSI		PBT			GSI	
		% assignment			% assign		% assignment			% assign	
Test population	Age	N	Pop	Age	N	Pop	N	Pop	Age	N	Pop
Big Qualicum					23	69.6 ¹				9	100.0
Big Qualicum	2	32	100.0	100.0			44	100.0	100.0		
Big Qualicum	3	2	100.0	100.0			3	100.0	100.0		
Cowichan	2	4	100.0	100.0 ^a			4	100.0	100.0 ^a		

Puntledge	1	0	-	-		1	0.0	0.0		
Chilliwack					39	100.0			22	100.0
Chilliwack	2	54	100.0	100.0		69	100.0	100.0		
Chilliwack	3	0	-	-		2	100.0	100.0 ^a		
Nitinat					64	29.7 ³			0	-
Nitinat	1	2	100.0	100.0 ^b		66	100.0	100.0 ^b		
Nitinat	2	3	100.0	100.0 ^b		3	100.0	100.0 ^b		
Sarita	1	1	100.0	100.0 ^b		1	100.0	100.0 ^b		
Sarita					3	66.7 ⁴			0	-
Sarita	1	92	100.0	100.0 ^b		95	100.0	100.0 ^b		
Quinsam					79	89.9 ⁵			15	60.0 ⁶
Quinsam	1	1	100.0	100.0 ^a		2	100.0	100.0 ^a		
Quinsam	2	10	100.0	100.0		15	100.0	100.0		
Quinsam	3	9	100.0	100.0		67	100.0	100.0		
Puntledge summer					4	100.0			1	100.0
Puntledge summer	2	44	100.0	100.0		47	100.0	100.0		
Puntledge summer	3	8	100.0	100.0		8	100.0	100.0		
Puntledge fall	2	1	100.0	100.0		1	100.0	100.0		

Puntledge fall					19	21.1 ⁸				6	16.7 ⁷
Puntledge fall	2	41	100.0	100.0			54	100.0	100.0		
Puntledge summer	2	21	100.0	100.0			21	100.0	100.0		
Cowichan	2	4	100.0	100.0 ^a			4	100.0	100.0 ^a		
Summary		329	100.0	100.0	231	67.1	507	99.8	99.8	53	79.2

^a CWT also recovered; ^b Age confirmed through scale analysis

Alternate GSI assignments: ¹ Puntledge (1), Little Qualicum (1), unassigned (5); ² unassigned (7); ³ Sarita (6), Sooke (6), Chapman (1), unassigned (32); ⁴ unassigned (1); ⁵ Big Qualicum (5), unassigned (3); ⁶ Big Qualicum (5), unassigned (1); ⁷ Big Qualicum (1), unassigned (4); ⁸ Big Qualicum (2), Little Qualicum (1), unassigned (12).

List of Figures

Figure 1. Map indicating sampling locations for 45 populations of Chinook Salmon in British Columbia. The specific populations, sampling years, and the number of individuals genotyped are outlined in Supplementary Table S1.

Figure 2. Neighbor-joining dendrogram of F_{st} distance for 45 populations of Chinook Salmon surveyed at 319 SNPs. Bootstrap values at major tree nodes indicate the percentages of 100 trees for which the populations beyond the node clustered together.

Draft

Draft