Development of a high-resolution SNP baseline for stock identification of Chinook salmon

Terry D. Beacham, Colin Wallace, Eric B. Rondeau

Fisheries and Oceans Canada,

Pacific Biological Station,

3190 Hammond Bay Road, Nanaimo, B. C.

Canada V9T 6N7

Final Report, May 2020

A project funded by the Northern Boundary Restoration and Enhancement Fund 2019-2020.

Abstract

Genetic stock identification (GSI) was used to assess accuracy of estimated stock compositions for accuracy of individual Chinook Salmon (*Oncorhynchus tshawytscha*) populations in the transboundary rivers of British Columbia. A regional population structure was observed for populations in transboundary rivers. A stock identification system utilizing 391 SNPs will generally provide a accurate method for estimation of stock composition by river drainage or conservation unit. A genetic system of identification will provide an alternate method of identification in the assessment and management of Canadian-origin Chinook salmon relative to the existing CWT program.

Introduction

Chinook salmon (*Oncorhynchus tshawytscha*) are produced in hatcheries in British Columbia, Canada. In some hatcheries, a portion of the juveniles is marked before release with coded-wire tags (CWTs, Jefferts et al. 1963), which are one-mm long pieces of encoded wire inserted into the nasal cartilage of juvenile salmon, and are detected in returning adults by an electronic tag detection system for individuals sampled in fisheries, hatcheries, or on the spawning grounds. Once individuals are recaptured and the CWTs recovered, the tags are decoded and the origin and age of the individual can be determined. Originally, 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 and Oregon received an adipose fin clip, but no corresponding CWT. This approach enabled mark-selective fisheries to be conducted, in which only adipose fin-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 importance of CWT recovery data for salmon management is exemplified by an annex in the Pacific Salmon Treaty (PST) between Canada and the United States that obligates both countries to maintain the CWT system in support of management of fisheries relevant to the PST. However, the increasing cost of applying CWTs has restricted the application to only a small proportion of fish released from hatcheries, with many hatcheries releasing juvenile Chinook salmon with no CWTs applied. The resulting small sample size of recovered CWTs greatly reduces confidence in estimates of stock contributions to mixed-stock fisheries because

the precision of these estimates is directly related to the number of CWTs recovered (Reisenbichler and Hartmann 1980; Hinrichsen et al. 2016).

Although CWTs are one method of stock identification in salmonids, genetic stock identification (GSI) has been a significant component of Pacific salmon fisheries management for the past 40 years. Allozymes were the first class of genetic markers used in stock identification applications (Utter et al. 1987; Shaklee et al. 1999), where the protein variation displayed by allozymes had an underlying genetic origin. The advent of newer technologies based upon DNA sequence variation provided different classes of genetic markers for application, and minisatellite DNA variation displayed substantial variation among coho salmon populations (Miller et al. 1996). While minisatellite DNA typically provided greater differentiation among populations than that observed with allozymes, the cumbersome nature of the laboratory analysis precluded widespread adoption of the technology to stock identification applications. The advent of microsatellite-based technology in the 1990s provided the first class of DNA-level variants that were effective in population differentiation and, with the development of automated DNA sequencers, practical for widespread application (Beacham et al. 2012). Single nucleotide polymorphisms (SNPs) later became a marker class of choice for many laboratories, given the ease of genotyping individuals, and the development of platforms capable of supporting genotyping individuals at large numbers of SNPs (Seeb et al. 2011).

A genetic application that provides equivalent information as CWTs, but without the limitation of small samples sizes, is parentage-based tagging (PBT; Anderson and Garza 2006). In PBT applications, the entire hatchery broodstock is genotyped annually, and subsequently progeny can be non-lethally sampled and assigned back to their parents using parentage analysis (Anderson 2012), thus identifying their hatchery of origin and brood year (i.e. age). When all

parental broodstocks are genotyped, then every offspring is genetically "tagged" which is equivalent to a 100% CWT rate, substantially higher than the current approximately 10% CWT tagging rates of juvenile Chinook salmon released from hatcheries in British Columbia. PBT has been employed in a number of applications in both natural and hatchery settings (Denson et al. 2012; Abadia-Cardosa et al. 2013; Steele et al. 2013; Ford et al. 2015, Hinrichsen et al. 2016; Hess et al. 2016), and has provided valuable information on individual reproductive success and effective population size.

In North America, GSI has replaced previous methods of salmonid stock identification, such as parasites (Margolis 1963) and scale pattern analysis (Cook and Guthrie 1987) due to improved diagnostic power of genetic methods. A key question to evaluate is whether a PBT/GSI approach can potentially replace a CWT-based approach to salmonid stock identification. Steele et al. (2013) demonstrated that for steelhead trout (*O. mykiss*) in the Snake River basin in the Columbia River drainage, stock assignments made through PBT with a panel of SNPs matched those made using CWTs, empirically confirming the equivalency of the techniques in this application.

A genetics-based approach to stock identification can incorporate both PBT and GSI. PBT allows stock of origin to be determined even when low levels of genetic differentiation among populations prohibits the use of GSI (Shaklee et al. 1999), as individuals are assigned to parents, not stocks. However, should differentiation among populations be sufficient to allow GSI to be conducted reliably, then a combined PBT/GSI approach may provide accurate estimates of stock composition. Genetically-based stock identification employs genetic markers, and for this purpose SNPs have been applied in a number of ecological, conservation, and fishery management settings (Habicht et al. 2010; Hess et al. 2011). SNP panels need to be applied

rapidly and inexpensively to genotype many individuals (Narum *et al.* 2013), incorporating sufficient numbers of SNPs to provide confidence in the accuracy of information produced. New genotyping by sequencing (GBS) techniques allow rapid SNP genotyping for many individuals at potentially hundreds of SNPs (Campbell et al. 2015), providing the capability of rapid, cost-effective genotyping of individuals through direct sequencing of amplicons.

In the present study, we evaluated whether a GSI approach can provide accurate estimation of stock composition for fisheries where transboundary river Chinook salmon populations are caught, with a transboundary baseline employed for potential population assignment. We conducted the study by applying GBS technologies to genotype Chinook salmon at 391 SNPs in 391 amplicons, and with a stock identification baseline comprising some 23 populations. We evaluated whether we could accurately estimate stock compositions of Chinook salmon through a GSI approach that incorporated a baseline of transboundary river populations that may contribute to complex marine mixed-stock fishery samples in fisheries in northern British Columbia.

The objectives of the study were:

1. Assemble 23 populations from transboundary rivers in BC for reference populations.
2. Genotype individual Chinook salmon with an amplicon panel to assay SNPs.
3. Evaluate the resolution of the SNP databases for genetic stock identification in the fisheries relevant to the Treaty.

Methods and Materials

Sample collection

The list of naturally-spawned populations sampled and their locations are listed in Table 1. Twenty-three populations were sampled. Fin tissue or operculum punches were obtained from all individuals sampled.

Library preparation and genotyping

DNA was extracted from *O. tshawystcha* tissue samples with a Biosprint robot in a 96-well format, and the DNA concentration normalized to 40 ng/µL with a Tecan LiHa robot. Four plates of extracted DNA were processed consecutively. The initial multiplex PCR amplification of 393 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 ddH_2O. Primers were developed from published sequence data for Chinook salmon (Beacham et al. 2018). Additionally, primers for SNPs originally found in coho salmon were screened for Chinook salmon (Beacham et al. 2020), and were subsequently incorporated into the panel originally outlined by Beacham et al. (2018).

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. Following this PCR, a second step employing a thermal cycler was conducted that digested the primers, whereby 1 µL of FuPa (Thermo Fisher) was added to the 10-µL cocktail, and the reaction conducted with the following conditions: 50°C – 10m; 55°C – 10m, 60°C – 20m, 10°C hold. The third and final step employing a thermal cycler to ligate the barcodes (384 individual codes) to the amplicons was initiated by adding 2µL switch solution, 1µL barcode, and 1µL ligase to the 11-µL cocktail, and was conducted with the following conditions: 22°C – 30m; 70°C – 10m, 10°C hold. Libraries were purified by addition of 22.5µL of Agencourt® AMPure® XP magnetic beads to each library, the plate was placed on a magnetic rack, supernatant

discarded, and the beads washed twice in 70% ethanol. The purified libraries were then eluted with 25µL of low TE, and 20 µL of the supernatant transferred to a fresh 96-well tray. Next, 3 µL of each of the prepared libraries were pooled into a single tube, and 20µL of the pooled libraries and 5µl of UltraPure H_2O were combined in a single tube for processing on the Ion Chef® (Thermo Fisher Scientific). Two tubes of pooled libraries were processed consecutively on the Ion Chef, and thus 768 individuals were processed on a single run of the Ion Chef. The reference genome, target regions (amplicon sequences) and hotspots (SNP sites on the amplicon) were imported to the Ion Chef. On the Ion Chef, the amplicons were attached to Ion Sphere Particles®(ISPs) via the P1 adapter, the complementary DNA sequence from the P1 adapter was extended via PCR, the libraries enriched by discarding of any ISPs with no attached DNA, the amplicons denatured with a wash of a NaOH solution, and the ISPs with attached DNA strands loaded on to a P1® chip v3. One tube of the pooled libraries was loaded on to each chip, and thus amplicons from 384 individuals were distributed on each P1 chip, with 768 individuals processed between the two chips. The chips were then loaded on to the Ion Torrent Proton sequencer, in which nucleotides flow sequentially over the Ion semiconductor P1 chip, where there is one ISP per well per sequencing reaction, and where incorporation of a specific nucleotide was detected through a change in pH adjacent to the amplicon. After the sequencing run was completed, amplicon sequences were aligned to the coho salmon (*O. kisutch*) genome (RefSeq assembly accession GCF_002021735.1), and scoring was conducted with Proton software Variant Caller®, and SNP genotypes at the sites specified by the hotspot file within target regions were called by Variant Caller. The hotspot file contained 393 SNP sites with one SNP scored at each amplicon. The barcodes attached to the amplicons identified the individual that was genotyped, and genotypes at all available SNPs for an individual were assembled to

provide a multi-locus individual genotype. In practice, our no-call rate (genotypes were not determined at all SNP sites for all individuals) was about 2.7% over all SNPs. The species identification SNP *OkiOts_120255-113* (Starks et al. 2013) and sex identification SNP *Ots_SEXY3-1* were omitted from subsequent GSI analyses, leaving 391 SNPs.

Data analysis

Individuals with more than 120 missing genotypes were eliminated from further analyses. In a test where the DNA of the same 382 individuals was genotyped on two occasions, an average genotyping error rate of 1.07% (1,220 discrepancies in 114,105 comparisons) or an allele error rate of 0.53% (1,220 discrepancies in 228,210 comparisons) was observed over the 391 SNPs scored.

For determination of population structure, F_{ST} distance was used to estimate genetic distances among all populations via hierfstat 0.04-22 (Goudet, 2004), with up to 391 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.

100% simulations

A method of individual identification is genetic stock identification (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 (RUBIAS, Moran and Anderson 2019). To test the accuracy of identifying the conservation unit and the population of origin, we performed GSI using Rubias

(Moran and Anderson 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 and river drainage to which the population belonged.

Results

Populations surveyed

One objective of the original proposal was to survey 23 populations from transboundary rivers in northern British Columbia. These populations were genotyped at up to 391 SNPs (Table 1)

Population structure

Regional population structure was observed in the 23 populations analyzed (Figure 2). Populations in the Alsek River were distinct from populations in the Taku and Stikine River drainages. In the Taku River drainage, the Nakina River population, although classified as a mid-timing return population for CU purposes, clustered genetically with populations deemed to be of early-timing return. The placement of the Nakina River population in the mid-timing CU may be open to question. The most distinctive northern population were in the Alsek River drainage. The late-returning Stikine River drainage populations tended to cluster as a group, except for the Tuya River population. This population was more similar to an early-returning

population in the drainage (Johnny Tashoots Creek), but sample size for this population (40 individuals) was smaller than for most other populations in the study. The observed presence of some regional population structuring fulfilled an important prerequisite of assignment of individuals by GSI, allowing the successful regional assignment of individuals from populations that occurred in a mixture but not in the baseline.

Single CU simulations

On average, accurate estimates of stock composition by CU were observed for simulated single CU mixture samples for transboundary river populations (Table 2, Figure 2). Average accuracy with respect to population was 57.8%, with respect to CU within drainage was 82.6%, and with respect to river drainage was 93.6%. The population with the least accuracy to CU (7.5%) was the Tuya River population, which was not surprising given the observed population structure, as it did not cluster with other late-returning populations (Figure 1). The Nakina River population also displayed poor accuracy of estimation with respect to CU (46.9%), and again this reflected the population structure, with the classified middle-returning timing population clustering with a population outside of its specified return time. Sample size was 114 individuals for this population, so the lack of clustering with other mid-timing populations was unlikely related to errors in estimation of allele frequencies due to restricted sample size. Re-evaluation of the placing of the Tuya River and Nakina River populations in their respective CUs may be in order.

Discussion

The CWT system provides three basic kinds of information: hatchery of release, year of release (and thus age of the individual), and release group within the hatchery. For wild Chinook

Salmon populations where a portion of the juveniles are marked with an adipose fin clip and a CWT, only location and year of release are typically available. The impetus for the current study originated from a critical need to evaluate whether a genetic-based stock identification system could identify individuals sampled in highly mixed-stock ocean fishery samples, in which there could be a large number of geographically-diverse contributing populations. In applications in which it can be demonstrated that estimation of an individual's origin through GSI can be obtained with a high degree of accuracy for specific populations, then GSI can be combined with PBT to provide accurate identification of individuals from those hatcheries in mixed-stock fisheries.

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). Vähä et al. (2016) noted that despite the promising SNP results, the costs for analyzing an adequate number of SNP markers was still high for GSI studies requiring the analysis of thousands or tens of thousands of individuals. Comparisons of the number of SNPs required to equal the accuracy and precision provided by existing microsatellite baselines for both sockeye salmon (Beacham et al. 2010) and Chinook salmon (Beacham et al. 2012b) provided no evidence of cost savings in our laboratory by applying SNPs in applications that were handled with microsatellites, when SNP genotyping was conducted using Taqman assays. However, the method of genotyping individuals via direct sequencing of amplicons has radically changed the cost of genotyping for individual fish in stock identification applications (Campbell et al. 2015), allowing hundreds of SNPs to be surveyed in a cost-effective manner. The number of SNPs employed in the current study was in excess of any previous stock identification application in Chinook salmon, and this evaluation has indicated that it was possible to provide

accurate estimates of stock composition of individual Chinook salmon populations from transboundary rivers with high accuracy to some specific populations (eg. Dudidontu River, Johnny Tashoots Creek, Verrett River) in some cases, to CU in most cases, and to river drainage in most cases. We expect that these individuals could be identified through GSI to smaller geographical regions of origin than is currently available from microsatellite analysis.

Acknowledgments

A very substantial effort was undertaken to obtain samples from Chinook salmon sampled in this study. We would like to thank northern DFO staff who collected and supervised collections in transboundary river drainages. We also acknowledge the various agencies and organizations whose staff collected samples in northern British Columbia.

References

Abadia-Cardosa, A., Anderson, E.C. , Pearse, D.E., and Garza, J.C. 2013. Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (*Oncorhynchus mykiss*). Mol. Ecol. 22: 4733-4746.

Anderson EC. 2018. R package: “Rubias”. https://cran.r-project.org/web/packages/rubias/rubias.pdf accessed February 10, 2018.

Anderson, E.C. 2012. Large-scale parentage inference with SNPs: an efficient algorithm for statistical confidence of parent pair allocations. Statistical Applications in Genetics and Molecular Biology Volume 11, Issue 5, ISSN (Online) 1544-6115, DOI: 10.1515/1544-6115.1833, November 2012

Anderson, E.C., and Garza, J.C. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics, 172(4):2567–2582. doi: 10.1534/genetics.105.048074.

Beacham, T. D., Jonsen, K., and Wallace, C. 2012. A comparison of stock and individual identification for Chinook salmon in British Columbia provided by microsatellites (STRs) and single nucleotide polymorphisms (SNPs). Mar. Coastal Fisheries: Dynamics, Management, and Ecosystem Sci. 4: 1-22.

Beacham, T. D., C. Wallace, C. MacConnachie, K. Jonsen, B. McIntosh, J. R. Candy, and R. E. Withler. 2018. Population and individual identification of Chinook Salmon in British Columbia through parentage-based tagging and genetic stock identification with single nucleotide polymorphisms. Canadian Journal of Fisheries and Aquatic Sciences 75: 1096-1105.

Beacham, T.D., Colin Wallace, Kim Jonsen, Brenda McIntosh, John R. Candy, Jean-Sébastien Moore, Louis Bernatchez, and Ruth E. Withler. 2020. Accurate estimation of Conservation Unit contribution to coho salmon mixed-stock fisheries in British Columbia, Canada using direct DNA sequencing for single nucleotide polymorphisms. Canadian Journal of Fisheries and Aquatic Sciences. In press

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. Res. 15:855-867.

Cook, R.C., and Guthrie, I. 1987. In-season stock identification of sockeye salmon using scale pattern recognition. P. 327-334. In H. D. Smith, L. Margolis, and C.C. Wood (eds). Sockeye salmon *(Oncorhynchus nerka)* population biology and future management. Can. Spec. Publ. Fish. Aquat. Sci. 96.

Denson, M.R., Brenkert IV, K., Jenkins, W.E., and Darden, T.L. 2012. Assessing red drum juvenile stocking in a South Carolina estuary using genetic identification. N. Amer. J. Fish. Manage. 32: 32-43.

Ford, M., Pearsons, T.N., and Murdoch, A. 2015. The spawning success of early maturing resident hatchery Chinook Salmon in a natural river system. Trans. Am. Fish. Soc. 144:539-548.

Goudet, J. 2004. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes. 5:184-186.

Habicht, C., Seeb, L.W., Myers, K.W., Farley, E.V., and Seeb, J.E. 2010. Summer–fall distribution of stocks of immature Sockeye Salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. Trans. Amer. Fish. Soc. 139:1171-1191.

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

Hess, J.E., Matala, A.P., and Narum, S.R. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Mol. Ecol. Res. 11 (Suppl.1)1-13.

Hinrichsen, R.A., Steele, C.A., Ackerman, M.W. , Campbell, M.R., Narum, S.R., Hess, M.A., Young, W.P., Shields, B.A., and Maschhoff, B.L. 2016. Maximum likelihood estimation of the proportion of hatchery-origin fish on spawning grounds using coded wire tagging and parentage-based tagging. Trans. Am. Fish. Soc. 145: 671-686.

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

Margolis, L. 1963. Parasites as indicators of the geographical origin of sockeye salmon, *Oncorhynchus nerka* (Walbaum), occurring in the North Pacific Ocean and adjacent seas. Int. N. Pac. Fish. Comm. Bull. 11: 101-156.

Miller, K.M., Withler, R.E., and Beacham, T.D. 1996. Stock identification of coho salmon (*Oncorhynchus kisutch*) using minisatellite DNA variation. Can. J. Fish. Aquat. Sci. 53: 181-195.

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

Narum, S.R., Buerkle, C.A., Davey, J.W. , Miller, M.R., and Hohenlohe, P.A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22**:**2841–2847.

Reisenbichler, R.R., and Hartmann Jr, N.A. 1980. Effect of number of marked fish and years of repetition on precision in studies of contribution to a fishery. Can. J. Fish. Aquat. Sci. 37: 576-582.

Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics, 27(4) 592-593.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L.W. 2011. Single nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol. Ecol. Res. 11 (Supplement S1): 1-8.

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

Steele, C.A., Anderson, E.C., Ackerman, M.W., Hess, M.A., Campbell, N.R., Narum, S.R., and Campbell, M.R. 2013. A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. Can. J. Fish. Aquat. Sci. 70: 1046-1054.

Vähä J-P, Erkinaro, J., Fålkegard, M., Orell, P., and Niemelä, E.E. 2016. Genetic stock identification of Atlantic salmon and its evaluation in a large population complex. Can. J. Fish. Aquat. Sci. 74: 327-338.

Utter, F.M., Teel, D.J., Milner, G.B., and McIssac, D.. 1987. Genetic estimates of stock compositions of 1983 chinook salmon harvests of the Washington coast and the Columbia River. Fish. Bull. (US) 85: 13-23.

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

Table 1. Chinook salmon spawning locations, sample collection years, and total number of fish sampled (N) for 23 populations in transboundary rivers of northern coastal British Columbia. N is the number of fish genotyped in the population.

Region/Conservation Unit	CU Number	Population	Years	N
Alsek	67	Blanchard River	2002, 2003	96
		Goat Creek	2007, 2009, 2011, 2012, 2013	100
		Klukshu River	2001	88
		Kudwat Creek	2008, 2010, 2011	69
		Takhanne River	2002, 2003, 2008, 2010, 2011	99
		Tatshenshini River	2001	24
Taku_early timing	63	Dudidontu River	2005, 2008	81
		Nahlin River	2006, 2007	97
		Tseta Creek	2008, 2010	210
Taku_mid timing	64	Little Tatsamenie River	2005, 2006, 2007	135
		Nakina River	2005, 2006	114
		Tatsmenie Lake outlet	2005	36
		Yeth River	2008, 2009, 2010	53
Taku_late timing	65	Hackett River	2008	95
		Kowatua Creek	1989, 1990, 2005	141
Stikine_early timing	60	Christina Creek	2002	86
		Johnny Tashoots Creek	2001, 2004, 2005, 2008, 2009	89
		Little Tahltan River	2010	130
		Shakes Creek	2001	90
		Tahltan River	2008, 2009, 2011	151
Stikine_late timing	61	Craig River	2002	91
		Tuya River	2008, 2009, 2011, 2012, 2013	40
		Verrett River	2007, 2009, 2010	118

Table 2. Accuracy of estimation of stock composition (%) for simulated 100% samples of single population samples with a baseline of populations of Chinook salmon populations from transboundary rivers as outlined in Table 1. CU is conservation unit.

Population	CU	River	Population	CU	River drainage
Blanchard River	Alsek	Alsek	75.9	99.5	99.5
Goat Creek	Alsek	Alsek	71.5	86.9	86.9
Klukshu River	Alsek	Alsek	59.6	99.1	99.1
Kudwat Creek	Alsek	Alsek	46.8	99.5	99.5
Takhanne River	Alsek	Alsek	73.8	99.5	99.5
Tatshenshini River	Alsek	Alsek	0.0	99.5	99.5
Christina Creek	STK-early	Stikine	67.3	91.0	99.1
Johnny Tashoots Creek	STK-early	Stikine	85.1	89.3	90.0
Little Tahltan River	STK-early	Stikine	44.7	93.1	93.1
Shakes Creek	STK-early	Stikine	22.6	85.3	85.3
Tahltan River	STK-early	Stikine	41.5	88.6	89.6
Craig River	STK-late	Stikine	57.8	93.3	96.6
Tuya River	STK-late	Stikine	7.5	7.5	73.4
Verrett River	STK-late	Stikine	83.7	97.1	99.5
Dudidontu River	TAKU-early	Taku	80.9	92.3	98.1
Nahlin River	TAKU-early	Taku	61.1	66.9	86.7
Tseta Creek	TAKU-early	Taku	87.3	87.3	96.0
Little Tatsamenie River	TAKU-mid	Taku	65.2	65.2	98.9
Nakina River	TAKU-mid	Taku	46.2	46.9	75.4
Tatsamenie Lake outlet	TAKU-mid	Taku	30.8	86.4	99.3
Yeth River	TAKU-mid	Taku	75.9	79.3	96.2
Hackett River	TAKU-late	Taku	76.5	80.4	91.9
Kowatua Creek	TAKU-late	Taku	66.7	66.7	99.0

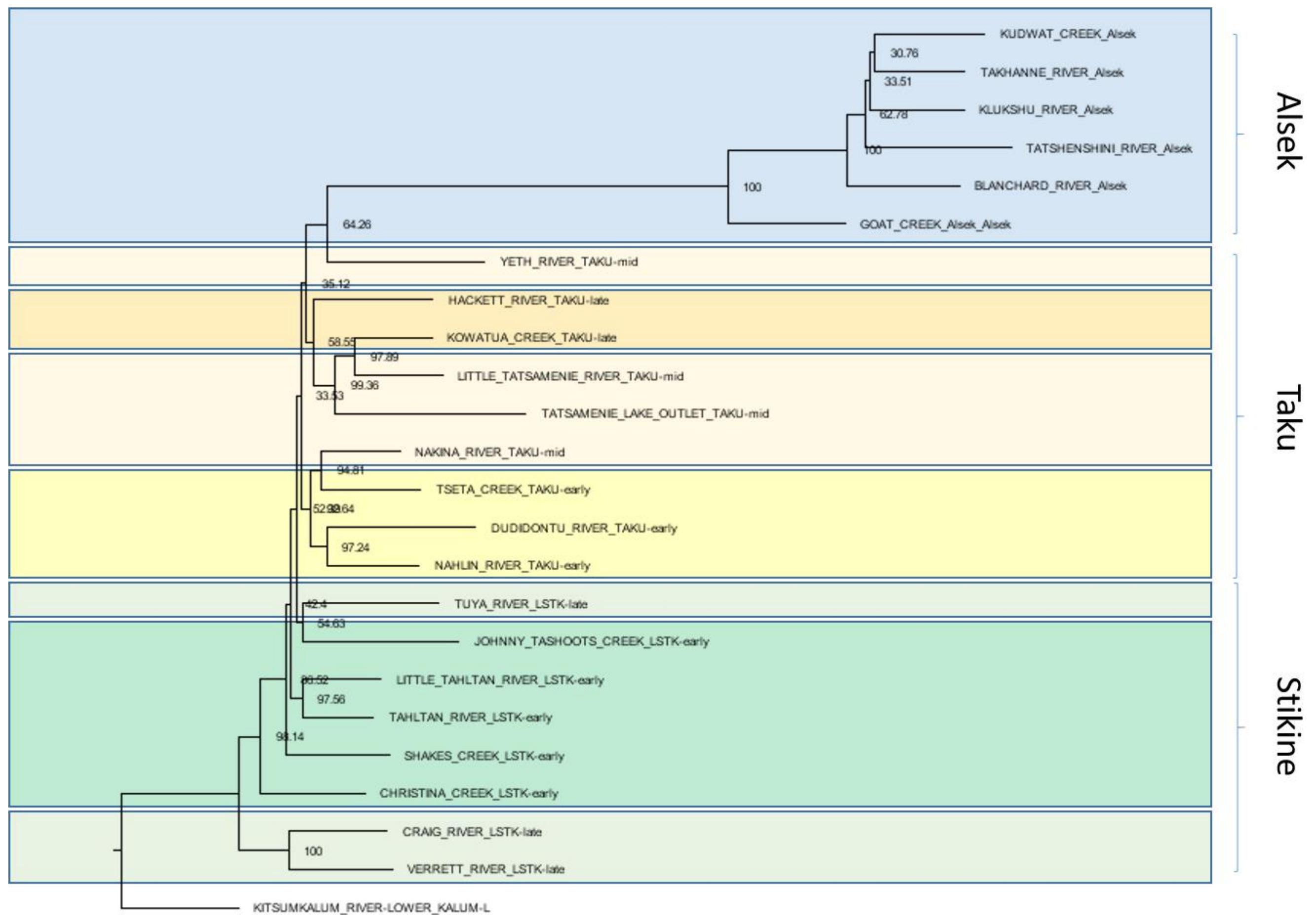

Figure 1. Neighbor-joining dendrogram of Fst distance for 23 populations of transboundary river Chinook salmon surveyed at 391 SNPs. Bootstrap values at major tree nodes indicate the percentages of 100 trees for which the populations beyond the node clustered together. The Kitsumkalum River population from the Skeena River drainage was used as an outgroup.

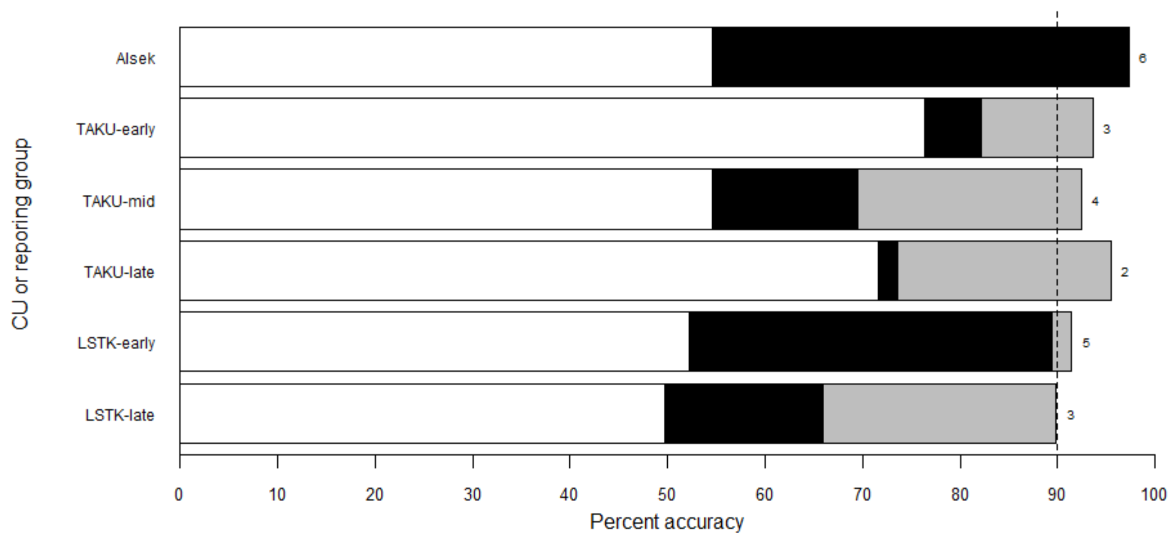

Figure 2. For simulated single-population samples, average % accuracy for estimated stock compositions for individual populations in a CU or reporting group back to population (open portion of bar), to CU (black portion of bar), and to river drainage (grey portion of bar) for all populations in a CU or reporting group, with the number of populations in the CU reported to the right of the bar. The 90% accuracy level is indicated in the figure. LSTK is the Stikine River.