



OPEN

Genotyping-in-Thousands by sequencing panel development and application for high-resolution monitoring of introgressive hybridization within sockeye salmon

Sarah L. Chang¹, Hillary G. M. Ward², Lucas D. Elliott^{1,3} & Michael A. Russello^{1✉}

Stocking programs have been widely implemented to re-establish extirpated fish species to their historical ranges; when employed in species with complex life histories, such management activities should include careful consideration of resulting hybridization dynamics with resident stocks and corresponding outcomes on recovery initiatives. Genetic monitoring can be instrumental for quantifying the extent of introgression over time, however conventional markers typically have limited power for the identification of advanced hybrid classes, especially at the intra-specific level. Here, we demonstrate a workflow for developing, evaluating and deploying a Genotyping-in-Thousands by Sequencing (GT-seq) SNP panel with the power to detect advanced hybrid classes to assess the extent and trajectory of intra-specific hybridization, using the sockeye salmon (*Oncorhynchus nerka*) stocking program in Skaha Lake, British Columbia as a case study. Previous analyses detected significant levels of hybridization between the anadromous (sockeye) and freshwater resident (kokanee) forms of *O. nerka*, but were restricted to assigning individuals to pure-stock or “hybrid”. Simulation analyses indicated our GT-seq panel had high accuracy, efficiency and power (>94.5%) of assignment to pure-stock sockeye salmon/kokanee, F₁, F₂, and B₂ backcross-sockeye/kokanee. Re-analysis of 2016/2017 spawners previously analyzed using TaqMan® assays and otolith microchemistry revealed shifts in assignment of some hybrids to adjacent pure-stock or B₂ backcross classes, while new assignment of 2019 spawners revealed hybrids comprised 31% of the population, ~74% of which were B₂ backcross or F₂. Overall, the GT-seq panel development workflow presented here could be applied to virtually any system where genetic stock identification and intra-specific hybridization are important management parameters.

Global fish populations are in decline due to agricultural development, rapidly rising riverine water temperatures, and reduced connectivity^{1–6}. In particular, 80% of salmonid populations in the Columbia River have experienced losses linked to hydroelectric development and the destruction of spawning and rearing habitat^{7,8}. In response, management strategies have included incorporating fish passage at dams to mitigate impacts to migratory fish life histories and supplementing populations in decline with hatchery fry for stock enhancement. For example, Coho salmon (*Oncorhynchus kisutch*) have been restored in the lower Columbia River using hatchery stocking, resulting in the establishment of local naturalized populations^{5,9}. Additionally, the enhancement of spawning habitat has been an effective strategy to restore fish populations, where the replacement of river substrate improved water velocities, dissolved oxygen, and usage of the site by spawning Chinook salmon (*O. tshawytscha*)¹⁰. Ultimately, the success of fish stocking programs that target species with diverse migratory and resident forms should include careful consideration of resulting hybridization dynamics and corresponding outcomes on recovery initiatives.

¹Department of Biology, University of British Columbia, Kelowna, BC, Canada. ²Lands and Natural Resource Operations and Rural Development, British Columbia Ministry of Forests, Penticton, BC, Canada. ³UiT The Arctic University of Norway, Tromsø, Norway. ✉email: michael.russello@ubc.ca

Hybridization in fish species is well documented, including inter-specific hybrids such as rainbow (*Oncorhynchus mykiss*) and cutthroat trout (*O. clarkii*)¹¹, as well as within-species hybrids between different life history forms, such as the case between anadromous sockeye salmon and freshwater resident kokanee¹². The long-term effects of hybridization are often complex, with the potential for both positive and negative outcomes¹³. Hybridization can be a powerful conservation tool harnessed to rescue populations with low genetic diversity and increase fitness through the integration of favorable traits such as larger body size, more offspring, and longer lifespans^{13–16}. On the other hand, detrimental effects have been observed such as introgression with maladapted gene complexes¹⁷, decrease in reproductive success¹⁸, and negative impacts on growth¹⁹, with hybrid fitness theorized to decrease as the divergence between parental phenotypes increases²⁰. Overall, the impacts of intra-specific hybridization in the wild are still not well understood, warranting further study, especially when observed as part of an active management program.

Oncorhynchus nerka provides an excellent system for investigating the genetic and physiological outcomes of hybridization as it exhibits tremendous life history variation, shows natal homing behavior, and represents a valuable species targeted for population restoration through restocking²¹. This species exhibits two main migratory forms, including anadromous sockeye salmon (hereafter referred to as “sockeye salmon”) and freshwater resident kokanee (hereafter referred to as “kokanee”). Kokanee are much smaller than sockeye salmon (26 cm versus > 45 cm average adult fork length) and occur sympatrically in many lakes, but tend to exhibit different spawning habitat preferences and spawning periods^{22,23}. Despite differences in spawning behavior, kokanee males are known to sneak on spawning sockeye salmon females²⁴, and size-selective mating has been observed between male sockeye salmon and female kokanee²⁵, allowing for gene flow between migratory forms. Sockeye-kokanee hybridization can lead to an increase in body size of resident hybrids that can increase angler satisfaction¹⁹ and bolster the overall genetic diversity of *O. nerka* in the system¹⁶. Conversely, hybrids can also experience negative impacts such as lower swimming capabilities than pure sockeye salmon²⁶, medial seawater adaptabilities²⁷, intermediate maturation time²³, decreased egg survival rate²⁸, and the loss of the iconic red coloration that is key in sexual selection²⁵. With a broad range of potential outcomes, accurate assessment of the extent of introgression is critical for elucidating the long-term effects of hybridization associated with supplementation and reintroduction programs.

To address large-scale declines in sockeye salmon populations in the Columbia River, an experimental reintroduction program was initiated in Skaha Lake within the Okanagan Basin of the Southern Interior of British Columbia, Canada. This program provides an excellent opportunity to investigate the extent and outcomes of intra-specific hybridization between *O. nerka* migratory and resident forms outside of a laboratory setting. Previous research examining hybridization between sockeye salmon and kokanee found *O. nerka* hybrids occupy intermediate morphologies and exhibit a largely resident life history, although this latter finding requires further investigation^{12,29}. These studies employed genetic marker sets that were effective at differentiating pure-stock from hybrid, however, they had limited power for the identification of advanced hybrid classes that has become increasingly important for genetic monitoring as the reintroduction program matures. The detection of advanced hybrid classes within a species can be challenging due to the proximity of backcrosses to pure genotype frequencies. For example, allozyme loci were only able to reveal intermediate allelic frequencies suggesting hybridization between transplanted and native sockeye³⁰, while microsatellite assays have the ability to detect introgression to the F₁ level³¹, but both marker types experience difficulty in the detection of advanced hybrid classes based on the conventional number of loci employed^{32,33}. In that regard, traditional markers may not have the statistical power to identify advanced hybrid classes³⁴, with an estimated minimum of 70 markers required to discriminate between pure parental species and advanced backcrosses³⁵. A previous simulation-based sensitivity analysis found that a panel comprised of 300 highly differentiated SNPs has the resolution needed to detect advanced hybrid classes between sockeye salmon and kokanee in Skaha Lake³⁶. With this in mind, advancements in massively parallel sequencing can be leveraged to improve the identification of hybridization in systems by pooling barcoded amplicons to increase the power of genetic panels for classification³⁷; Genotyping-in-Thousands by sequencing (GT-seq) is a particularly useful approach in cases where large sample sizes need to be cost-effectively genotyped³⁸.

Here, we demonstrate a workflow for developing, evaluating and deploying a GT-seq SNP panel with the power to detect advanced hybrid classes to assess the extent and trajectory of intra-specific hybridization, using the sockeye salmon reintroduction program in the Okanagan Basin as a case study. Using previously published restriction site associated DNA sequencing (RAD-seq) collected for this system, we first simulated two pure (kokanee, sockeye) and four hybrid classes (F₁, F₂, B₂ backcross-kokanee, B₂ backcross-sockeye) with three different SNP datasets (300 highest *F_{st}* SNPs; 600 highest *F_{st}* SNPs; 350 random SNPs selected from the 600 highest *F_{st}* SNPs) to assess information content to inform panel construction. We then evaluated the accuracy, efficiency, and power of the optimized GT-seq panel using simulation analyses. We further assessed panel performance by re-genotyping individuals sampled in Skaha Lake in 2016 and 2017 that were previously genetically assigned as pure-stock or F₁ hybrid at 32 SNPs and subjected to otolith microchemistry analysis to reconstruct migratory history²⁹. Finally, we genetically assigned *O. nerka* spawners sampled in 2019 to pure-stock or hybrid class (F₁, F₂, B₂ backcross-kokanee, B₂ backcross-sockeye) to estimate stock proportions and examine trends in this system over time.

Methods

Study system and samples. Skaha Lake is located in the Canadian portion of the Okanagan Basin, and flows south to join the Columbia River (Fig. 1). *O. nerka* spawn in the Okanagan River, upstream of Skaha Lake. Historically, sockeye salmon populations existed in the Okanagan Basin, however, the construction of a dam at McIntyre Bluff in 1921 blocked access to spawning grounds and channelization of the Okanagan River

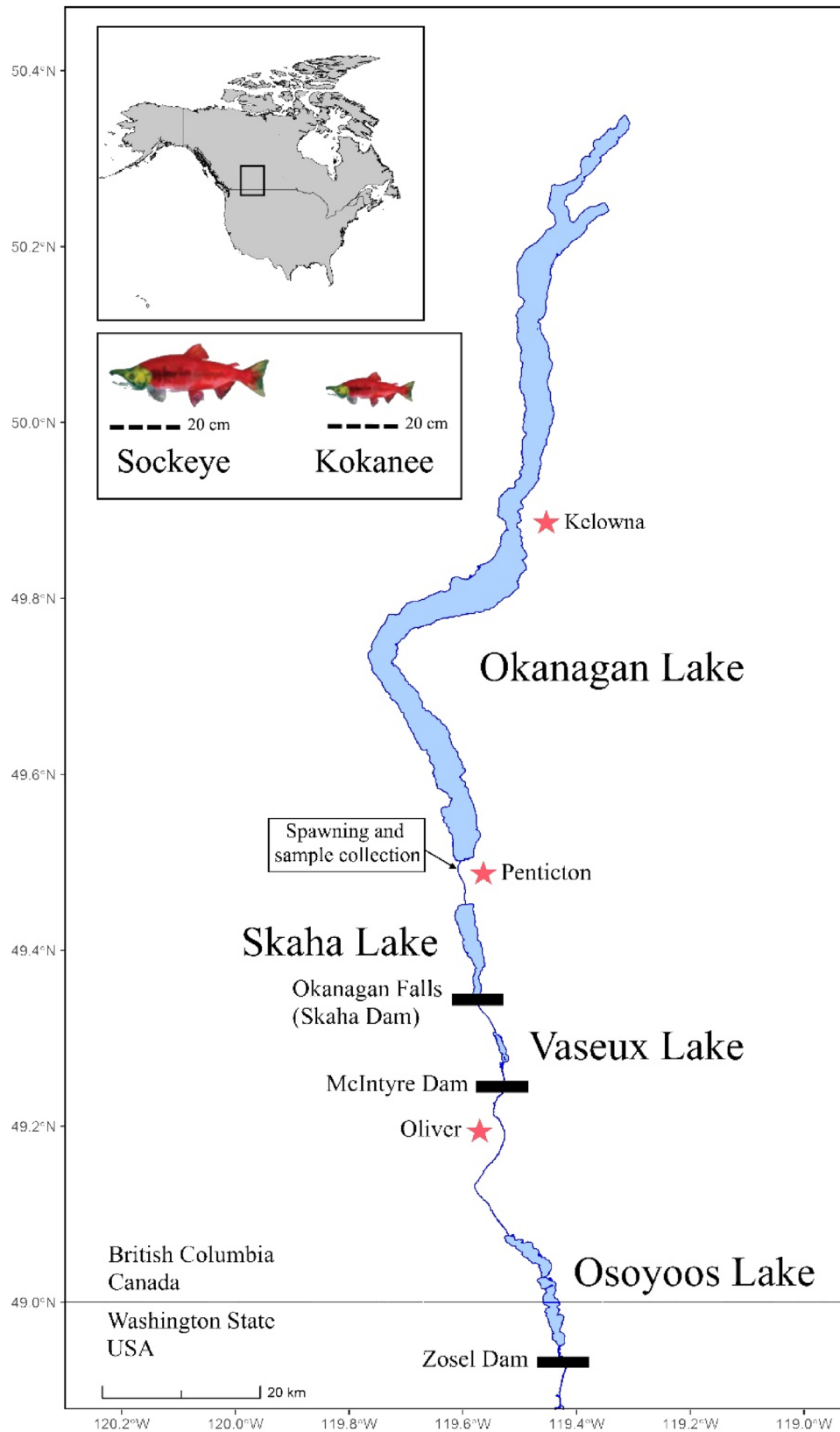


Figure 1. Map of the Columbia River displaying Skaha Lake, Okanagan River (extending southward originating at McIntyre Dam), and downstream lakes and dams. Size comparison of an average sockeye male and kokanee male included. The maps were created using a custom R script (<https://github.com/changsarah/BCmap>) with R version 4.1.1 (<https://www.R-project.org/>). Geographic data were accessed under the open government license—British Columbia: <https://catalogue.data.gov.bc.ca/dataset/freshwater-atlas-watersheds> and the Washington geospatial open data license: <https://geo.wa.gov/datasets/waacy:wa-hydrography-nhdwaterbody/about>.

further degraded spawning habitat. These activities left Skaha Lake with a sole population of stream-spawning kokanee³⁹. In attempts to restore sockeye salmon to the Okanagan Basin, a re-introduction program was implemented in 2004. Through this initiative, upstream fish passage was created by making structural improvements to migration barriers, restoring stream habitat, and stocking sockeye salmon fry annually into Skaha Lake. This system now supports a self-sustaining population of sockeye salmon^{39,40}.

We genotyped tissue samples from *O. nerka* spawners in the Okanagan River upstream of Skaha Lake that were previously sampled in 2016 (n = 45) and 2017 (n = 59)²⁹, and newly sampled in 2019 (n = 218; this study) as deadpitch (post-mortem carcass collection after spawning) by personnel from the BC Ministry of Forests, Lands, Natural Resource Operations, and Rural Development. The section of river that was sampled represents the only spawning habitat that exists for *O. nerka* populations from Skaha Lake and all forms spawn within this section of river. Samples were collected on multiple dates that spanned the duration of the spawning run. Biological data (length and sex) and tissue samples (operculum punches preserved in ethanol) were collected from all sampled fish. Moreover, the 2016 and 2017 samples were previously genetically assigned to pure-stock or F₁ hybrid using a panel of 32 TaqMan® SNP assays and subjected to otolith microchemistry analysis³⁶ to determine fish migratory history and investigate maternal migratory signatures that are passed transgenerationally to offspring⁴¹.

GT-seq SNP panel design. We used previously published genotypic data collected via RAD-seq⁴² for: (1) Skaha Lake kokanee (n = 20) collected in 2003 prior to the sockeye salmon restocking program; and (2) Okanagan River sockeye salmon (n = 35) collected in 2012 downstream of historical migration barriers. Using the *populations* module in STACKS version 2.0 beta 8⁴³, we required all loci to be present in at least 60% of individuals in both populations, with a minor allele frequency greater than 0.05. Due to the salmonid whole genome duplication event, we filtered out suspected homeologs by removing any locus with a negative F_{is} or $H_{obs} > 0.5$ that occurred in both reference populations⁴⁴. The resulting SNP dataset was then filtered for quality using VCFtools⁴⁵ to remove SNPs not in Hardy–Weinberg equilibrium and to calculate Weir and Cockerham (1984)'s θ , an unbiased estimate of F_{st} ⁴⁶, between Skaha Lake kokanee and Okanagan River sockeye salmon following³⁶. We removed loci with insufficient flanking sequence required for primer design by retaining loci with the SNP positioned between the 40th and 70th base pairs of the RAD tag sequences and selected the top 650 loci exhibiting the highest F_{st} . We then assessed all loci pairs for deviation from linkage equilibrium using GENEPOP 4.5⁴⁷, removed loci that were linked, and finalized selection of 600 candidate loci.

To examine panel informativeness of candidate loci and forecasting of panel ability after primer dropout, we simulated parental and offspring individuals (n = 1000) for six hybrid classes (kokanee, sockeye, F₁, F₂, B₂ backcross-kokanee, B₂ backcross-sockeye) with *recom-sim.py* (<https://github.com/salanova-elliott/recom-sim>). Separate simulations were conducted at: 1) 300 highest F_{st} SNPs; 2) 600 highest F_{st} SNPs; and 3) 350 SNPs randomly selected from the 600 highest F_{st} SNPs. We assigned simulated individuals to parental or hybrid class by calculating the posterior probabilities of membership as implemented in NEWHYBRIDS⁴⁸ with the reference populations flagged as known genotypes with the “z” option, and constructed confusion matrices for each panel assessment. After *in-silico* assessment, we sent the full RAD tag sequences that were associated with the pool of candidate SNPs to GTseek LLC (<https://gtseek.com/>) for custom locus-specific primer design.

GT-seq test library preparation. We constructed a GT-seq test library with the previously extracted DNA samples collected in 2016 (n = 45) and 2017 (n = 59) for which otolith microchemistry analysis was previously conducted²⁹. Extracted DNA was quantified with a Qubit 3.0 Fluorometer and the dsDNA High Sensitivity Assay Kit (Invitrogen). Library preparation followed the original protocol³⁸, with the exception that we diluted the PCR1 product to 1:10 (<https://doi.org/10.17504/protocols.io.byvppw5n>). The PCR2 product was quantified with Picogreen™ (Molecular Probes, Inc.) and each sample was normalized to a concentration of 10 ng/μL. The pooled library was purified with a MinElute PCR Purification Kit (Qiagen) and eluted into a final volume of 25 μL. Test libraries were sequenced using a Mid Output Reagent Kit (300 cycles) on an Illumina MiniSeq within the Ecological and Conservation Genomics Laboratory at the University of British Columbia Okanagan.

GT-seq genotyping and primer optimization. Demultiplexed raw sequencing files were processed with the GT-seq pipeline available on GitHub (<https://github.com/GTseq/GTseq-Pipeline>). We removed primers with non-specific *in silico* probes, candidates that were overrepresented (exhibiting > 2% of the raw read count), observed primer dimers, potential PCR artefacts, off-target amplification or *in silico* probe variation following previously published work⁴⁹. A second and third test library with the same sample composition as the first test library were prepared and tested iteratively with optimized primer pools from previous libraries using the protocols detailed above for sample preparation, sequencing, and primer dropout (Fig. 2).

To compile the final dataset based on the optimized SNP panel for downstream analyses, raw sequencing files from individuals were concatenated across sequencing runs and processed with the GT-seq pipeline available on GitHub (<https://github.com/GTseq/GTseq-Pipeline>). We filtered out individuals with > 25% missing data using PLINK⁵⁰. We decided on a cutoff of 25% to minimize the proportion of missing data that can negatively impact the detection of population structure, particularly with the high resolution needed to elucidate hybrid classes within a species⁵¹.

Panel performance. We assessed the accuracy, efficiency and power of NEWHYBRIDS assignments with simulated individuals (n = 100) using two pools (top 300 F_{st} SNPs, optimized SNP panel) with *hybridpowercomp* as implemented in the R package *hybriddetective*^{34,52}. Specifically, we evaluated assignment probabilities based on: (1) accuracy (correct assignments/total assignments per category); (2) efficiency (correct assignments/total number of individuals in a category); and (3) power (product of accuracy and efficiency). We also used *hybrid-*

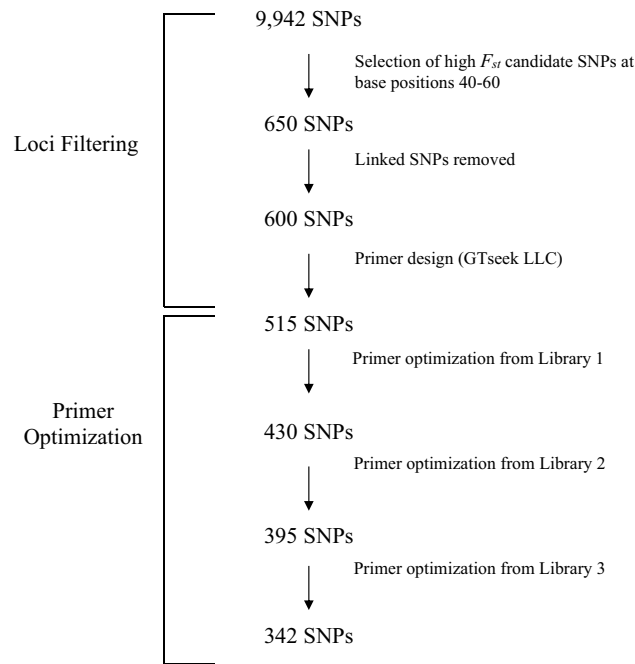


Figure 2. Workflow of *Oncorhynchus nerka* (sockeye salmon and kokanee) GT-seq hybridization panel design.

detective^{34,52} to quantify error rates associated with individual assignment (top 300 F_{st} SNPs, optimized SNP panel) as follows: (1) Type I: false positive error rate (number of known pure individuals wrongly assigned to a hybrid genotype frequency class/total number of known pure individuals); and (2) Type II: false negative error rate (number of known hybrid individuals wrongly assigned to a pure genotype frequency class/total number of known hybrid individuals).

Re-analysis of 2016–2017 samples. We genetically assigned the 2016 and 2017 sampled individuals to pure-stock or hybrid class using NEWHYBRIDS⁴⁸, the genotypic data from the optimized SNP panel, and reference baseline genetic data from 2003 Skaha Lake kokanee and 2012 Okanagan River sockeye salmon⁴². Five genotype frequency classes (kokanee, sockeye, F_1 , F_2 , B_2 backcross-kokanee, and B_2 backcross-sockeye) were specified, and the analysis initiated with a burn-in period of 10,000 followed by 50,000 iterations. We applied the s and z flags to the reference sockeye salmon and kokanee populations to designate individuals of known genotype that were sampled separately from the test individuals. We then selected the maximum probability of assignment to assign individuals to the most likely class. Lastly, we compared the NEWHYBRIDS⁴⁸ genetic assignment to pure-stock or hybrid class to the migratory history of the individuals, as previously inferred²⁹ by way of otolith microchemistry analysis.

Assignment of 2019 samples. For the new samples collected in 2019 ($n = 218$), DNA was extracted with a standard Chelex-based protocol in 96-well plates containing approximately 0.5 mm² of tissue, 5 μ l proteinase K (10 mg/ml), and 195 μ l 10% Chelex solution⁵³. Plates were incubated for 4 h at 55 °C to digest tissue, and then 95 °C for 15 min using an Applied Biosystems Veriti thermal cycler (Applied Biosystems, Foster City, CA, USA). Library preparation was conducted as above with the optimized SNP panel. The multiplexed pooled library was sequenced using a partial High Output Reagent Kit (300 cycles) on the Illumina MiniSeq within the Ecological and Conservation Genomics Laboratory at the University of British Columbia Okanagan. Samples with individual missing data > 25% were filtered out for downstream analyses. Individual assignment to hybrid class was conducted using NEWHYBRIDS⁴⁸ and the same parameters as above. Fisher's Exact Test was conducted in R to determine if proportions of hybrid class were significantly different within a sample year and between sample years.

Morphometric analyses. One-way ANOVA and post-hoc Tukey Tests were conducted in R to determine if mean fork length for the 2019 samples was significantly different between individuals assigned to pure-stock (kokanee, sockeye salmon) and the various hybrid classes (F_1 , F_2 , B_2 backcross-kokanee, and B_2 backcross-sockeye).

A: Top 300 F_{st}	Kokanee	Sockeye	F ₁	F ₂	B ₂ Kokanee	B ₂ Sockeye
Kokanee	1.000	0.000	0.000	0.000	0.000	0.000
Sockeye	0.000	1.000	0.000	0.000	0.000	0.000
F ₁	0.000	0.000	1.000	0.000	0.000	0.000
F ₂	0.000	0.000	0.020	0.980	0.000	0.000
B ₂ Kokanee	0.000	0.000	0.000	0.000	1.000	0.000
B ₂ Sockeye	0.000	0.000	0.000	0.000	0.000	1.000
B: Top 600 F_{st}	Kokanee	Sockeye	F ₁	F ₂	B ₂ Kokanee	B ₂ Sockeye
Kokanee	1.000	0.000	0.000	0.000	0.000	0.000
Sockeye	0.000	1.000	0.000	0.000	0.000	0.000
F ₁	0.000	0.000	0.960	0.040	0.000	0.000
F ₂	0.000	0.000	0.020	0.980	0.000	0.000
B ₂ Kokanee	0.000	0.000	0.000	0.000	1.000	0.000
B ₂ Sockeye	0.000	0.000	0.000	0.000	0.000	1.000
C: Random 350 F_{st}	Kokanee	Sockeye	F ₁	F ₂	B ₂ Kokanee	B ₂ Sockeye
Kokanee	1.000	0.000	0.000	0.000	0.000	0.000
Sockeye	0.000	1.000	0.000	0.000	0.000	0.000
F ₁	0.000	0.000	0.970	0.030	0.000	0.000
F ₂	0.000	0.000	0.060	0.920	0.020	0.000
B ₂ Kokanee	0.000	0.000	0.000	0.000	1.000	0.000
B ₂ Sockeye	0.000	0.000	0.000	0.010	0.000	0.990
D: 342 Final	Kokanee	Sockeye	F ₁	F ₂	B ₂ Kokanee	B ₂ Sockeye
Kokanee	1.000	0.000	0.000	0.000	0.000	0.000
Sockeye	0.000	1.000	0.000	0.000	0.000	0.000
F ₁	0.000	0.000	0.950	0.050	0.000	0.000
F ₂	0.000	0.000	0.050	0.940	0.000	0.010
B ₂ Kokanee	0.000	0.000	0.010	0.020	0.970	0.000
B ₂ Sockeye	0.000	0.000	0.000	0.020	0.000	0.980

Table 1. Confusion matrices of simulated individuals' proportional assignment with NEWHYBRIDS based on data from: (A) Top 300 F_{st} SNPs; (B) Top 600 F_{st} SNPs; (C) Random 350 SNPs from the top 600 F_{st} SNPs (Random 350 F_{st}); and (D) 342 final SNP panel. Rows represent true genetic classification, with columns as assigned hybrid class. Bold values along the diagonal are correct assignment proportions. Subsets of candidate markers were tested for informativeness to assess potential primer dropout.

Results

Initial panel ability. The simulated SNP panels provided comparable accuracy to assign individuals to hybrid class based on initial NEWHYBRIDS assignment of simulated individuals (Table 1). The top 300 F_{st} SNP panel performed best with high accuracy (>0.98), followed by the top 600 F_{st} SNP panel (>0.96) and the random 350 SNPs selected from the top 600 F_{st} SNPs (>0.92).

From the initial pool of top 600 F_{st} SNPs, primers were successfully designed for 515 SNPs after *in silico* testing. Following three rounds of multiplex amplicon sequencing and primer pool optimization, the optimized GT-seq panel consisted of 342 SNPs (Fig. 2, Supplementary Table S1). Our finalized 342 SNP panel had an accuracy of >94% across all hybrid classes (Table 1). The lowest assignment accuracies were to F₁ and F₂ across all simulated panels (but all ≥0.92; Table 1), which is consistent with previous results³⁶.

SNP panel assignment efficacy. The optimized 342 SNP GT-seq panel performed slightly worse than the top 300 F_{st} SNP pool, but still displayed high accuracy and efficiency in assigning simulated individuals to pure-stock and the different hybrid classes (Fig. 3). Pure-stocks and B₂ backcrosses were detected at >98% accuracy and >99% efficiency at a critical posterior probability threshold of 50% (Fig. 3). F₁ and F₂ classes were detected at lower, but still meaningful levels, with efficiency at >92% and accuracy at >90% at the 50% critical probability threshold (Fig. 3). The power of assignment was comparable between the 342 SNP GT-seq panel and top 300 F_{st} SNPs where pure-stocks and B₂ backcrosses remained stable over a wide range of probability thresholds: >99% at a critical probability threshold of 50%, and >98% at a threshold of 90% (Supplementary Fig. S1). However, the power of assignment for F₁ and F₂ classes declined in performance at higher critical posterior probability thresholds: >90% at a critical probability threshold of 50% and >79% at a threshold of 90% (Supplementary Fig. S1). The Type I false positive error rate and Type II false negative error rate were <0.001% for both the 342 SNP GT-seq panel and top 300 F_{st} SNPs.

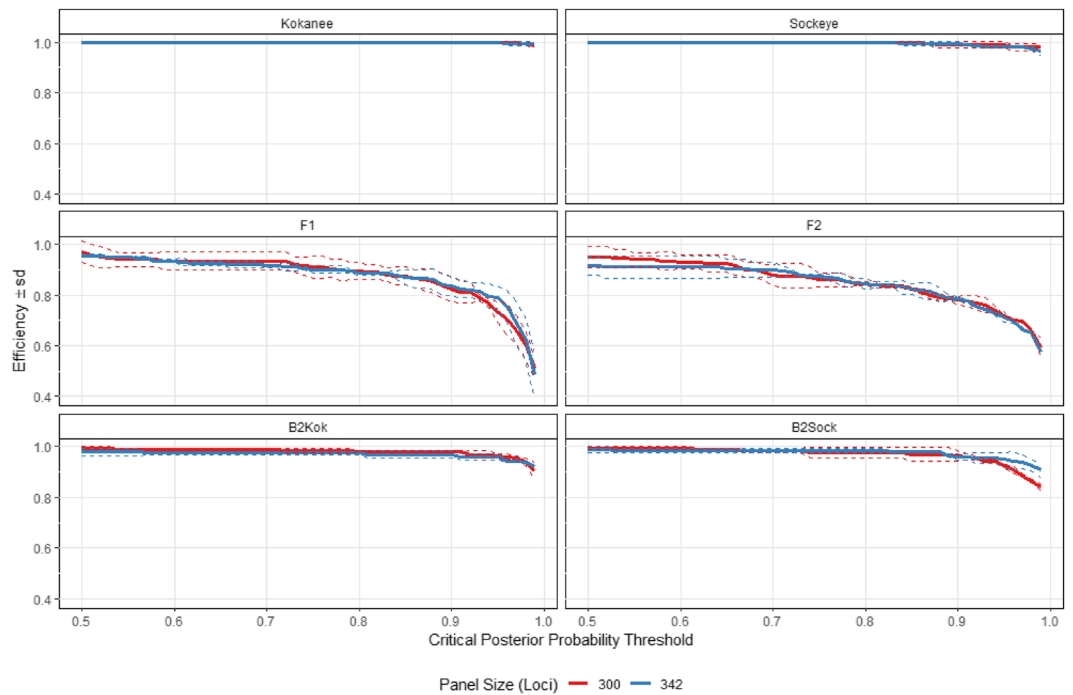


Figure 3. Accuracy and efficiency with simulated individuals of different hybrid classes for the top 300 F_{st} panel and the final optimized 342 SNP GT-seq panel.

Year	<i>n</i>	TaqMan® 32 SNP assay			342 SNP GT-seq panel					
		Kokanee	Hybrid	Sockeye	Kokanee	B ₂ Kok	F ₁	F ₂	B ₂ Sock	Sockeye
2016	36	16 (0.44)	12 (0.33)	8 (0.22)	16 (0.44)	1 (0.03)	8 (0.22)	0 (0.00)	0 (0.00)	11 (0.31)
2017	55	19 (0.35)	19 (0.35)	17 (0.31)	23 (0.42)	3 (0.05)	11 (0.20)	0 (0.00)	0 (0.00)	18 (0.33)
2019	202	–	–	–	115 (0.57)	24 (0.12)	16 (0.08)	4 (0.02)	19 (0.09)	24 (0.12)

Table 2. Stock composition of spawners by sample year assessed with the TaqMan® 32 SNP assay and 342 SNP GT-seq Panel. Stock proportions indicated in parentheses.

Re-analysis of 2016–2017 Okanagan river samples. We successfully genotyped 2016 ($n = 36$) and 2017 ($n = 55$) Okanagan River spawners with the 342 SNP GT-seq panel after filtering for 25% missing data across individuals (average read depth = 220.8; genotyping rate = 90.1%). Overall, assignments to pure-stock and hybrid class were generally similar between the different marker sets, although there was a slightly higher proportion of individuals assigned to pure-stock using the 342 SNP GT-seq panel (Table 2; Supplementary Table S2). Our 342 SNP GT-seq panel further refined assignments into multiple hybrid classes for those individuals previously classified as hybrids, both in 2016 (proportion F₁: 0.22, B₂ backcross-kokanee: 0.03) and 2017 (proportion F₁: 0.20, B₂ backcross-kokanee: 0.05) (Table 2).

A small number of assignment differences were found between the TaqMan® SNP assays²⁹ and 342 SNP GT-seq panel. In 2016, four hybrids were reclassified as sockeye salmon with three instances of microchemistry conflict, where the latter results suggested that these individuals did not migrate to the ocean (Supplementary Table S2). Additionally, one sockeye salmon was genetically reclassified to F₁ hybrid. One individual that was genetically assigned as sockeye salmon with both the 32 SNP TaqMan® assays²⁹ and 342 SNP GT-seq panel had conflicting microchemistry that suggested no history of anadromy or maternal anadromy (Supplementary Table S2). Within the 2017 cohort, five individuals that were previously genetically assigned as hybrids were refined to: sockeye salmon ($n = 1$) and kokanee ($n = 4$) (Table 2; Supplementary Table S2). All microchemistry was congruent, other than one individual that was genetically assigned as sockeye salmon by both 32 SNP TaqMan® assays²⁹ and the 342 SNP GT-seq panel; in this case, microchemistry suggested a resident maternal parent, but anadromous migratory history.

Assignment of 2019 Okanagan river samples. We genotyped 2019 Okanagan River spawners ($n = 202$) with the 342 SNP GT-seq panel after filtering for 25% missing data across individuals (average read depth = 370.9; genotyping rate = 91.2%). The composition of the 2019 Okanagan River spawners revealed a higher proportion

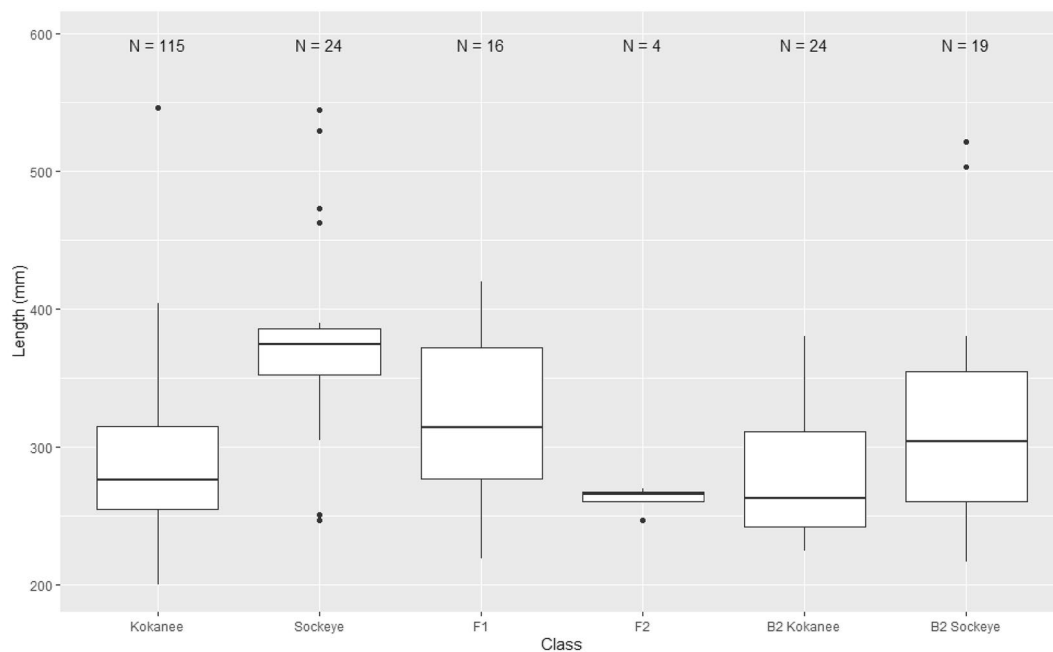


Figure 4. Body length of spawning *Oncorhynchus nerka* in Okanagan River Channel 2019.

of kokanee (0.57; Table 2; Supplementary Table S3) and lower proportion of sockeye salmon (0.12; Table 2; Supplementary Table S3) than in 2016 and 2017 (kokanee: 0.42–0.44; sockeye salmon: 0.31–0.33; Table 2; Supplementary Table S3). In addition, the 2019 spawners contained substantially more detected backcrosses, including the first B₂ backcross-sockeye individuals genetically identified (B₂ backcross-kokanee: n = 24, overall proportion = 0.12; B₂ backcross-sockeye: n = 19; overall proportion 0.09; Table 2; Supplementary Table S3). We also detected low proportions of F₂ hybrids in 2019 (n = 4, overall proportion = 0.02) when compared to other hybrid classes.

We found significant differences in proportions among hybrid classes within a year among all sample years (2016: $P < 0.001$, $\chi^2 = 45.20$, $df = 5$; 2017: $P < 0.001$, $\chi^2 = 62.68$, $df = 5$; 2019: $P < 0.001$, $\chi^2 = 292.61$, $df = 5$). Across sample years, we found that the proportions of kokanee ($P = 0.080$, $\chi^2 = 5.04$, $df = 2$), F₂ ($P = 0.779$, $\chi^2 = 1.83$, $df = 2$) and B₂ backcross-kokanee ($P = 0.140$, $\chi^2 = 4.35$, $df = 2$) were not significantly different. The proportions of sockeye salmon ($P < 0.001$, $\chi^2 = 16.98$, $df = 2$), F₁ ($P < 0.05$, $\chi^2 = 10.12$, $df = 2$), and B₂ backcross-sockeye ($P < 0.05$, $\chi^2 = 9.15$, $df = 2$) were significantly different across sample years.

In general, sockeye salmon exhibited larger mean lengths compared with kokanee, with hybrids generally occupying an intermediate range between kokanee and sockeye salmon mean lengths (Supplementary Fig. S2; Supplementary Table S2). However, it is notable that F₂ hybrids seem to occupy the same size distribution as B₂ backcross-kokanee, signifying smaller body sizes with advanced hybrid classes. The mean lengths between sockeye/kokanee, F₁/sockeye, F₂/sockeye, B₂ backcross-kokanee/sockeye, and B₂ backcross-sockeye/sockeye were significantly different (Supplementary Fig. S2; Supplementary Table S2).

Discussion

The accurate identification of advanced hybrid classes is valuable for monitoring the extent of introgression and potential fitness impacts between multiple reproductive forms of a single species or where hybridization between species is possible. SNP panels have been developed for a broad range of taxa with the ability to accurately identify hybridization between closely related species within the same genus up to the third backcross hybrid generation^{54,55}. However, to our knowledge, this is the first GT-seq SNP panel with the resolution to differentiate advanced hybrid classes up to the second backcross hybrid generation across life history forms *within-species*. Our GT-seq SNP panel represents a valuable tool that may be used to examine the ongoing progression of intra-specific hybridization and potential fitness outcomes associated with an active sockeye salmon reintroduction program in the Okanagan Basin, providing information to guide on-going management strategies and offer a roadmap to other such programs for species with complex life histories.

GT-seq panel development and evaluation. Regarding initial panel development and optimization, we effectively integrated an expected level of primer drop-out within our simulation analyses to inform locus selection and provide preliminary insights on panel performance. To that end, our simulations were consistent with previous work in this system (e.g.^{29,36}), and displayed high assignment accuracy across all hybrid classes (> 94%), as well as comparable panel accuracy, efficiency, and power between the top 300 F_{st} SNP pool, random 350 SNP subset of the top 600 F_{st} SNP pool, and ultimately, our optimized 342 SNP panel (Fig. 4, Supplementary Fig. S2). Subsequent error rates (< 0.001% at all levels) exhibited by the optimized 342 SNP panel reflected those of the

pilot analyses, effectively streamlining the GT-seq panel preparation process and helping to ensure that panel performance met management relevant benchmarks for accuracy, efficiency, and power.

When applied to 2016 and 2017 Okanagan River spawners previously analyzed using 32 SNP TaqMan® genotyping assays and otolith microchemistry, the optimized 342 SNP panel did result in a small number of shifts in assignment of hybrids to adjacent pure-stock classes or B₂ backcross classes. One previously classified pure-stock sockeye salmon had a refined assignment (> 99% probability) to an F₁ hybrid. Microchemistry revealed saltwater inhabitation for the both this individual and their maternal parent; the reclassification to F₁ hybrid suggests that that the paternal parent was likely a kokanee rather than a sockeye (Supplementary Table S2). Individuals that were refined from hybrid to pure-stock sockeye salmon (n = 2) and kokanee (n = 4) all exhibited a high probability of assignment (> 75%), consistent microchemistry, and had larger body sizes than other fish in the hybrid class (Supplementary Table S2). Reclassification of F₁ hybrids to pure-stock sockeye salmon did not display evidence of anadromy, suggesting they may be residual sockeye salmon with anadromous parentage, but did not migrate to sea⁵⁶. Moreover, all of these newly classified pure-stock sockeye salmon had larger body sizes than the average hybrid individual (Supplementary Table S2).

Management implications. Sockeye salmon have been successfully re-introduced into the Okanagan Basin and comprised 41% of the *O. nerka* population by 2014¹². The proportion of sockeye salmon in this system since that peak has significantly varied over time (12–33%), likely due to a combination of factors such as the number of spawners and fry stocked, variable ocean survival, and high water temperatures during upstream migration⁵⁷. The re-establishment of wild spawning sockeye salmon created the possibility of hybridization with resident kokanee in later years; our results demonstrate that hybrids most recently comprised 31% of the population (2019: 8% F₁, 2% F₂, 12% B₂ backcross-kokanee, 9% B₂ backcross-sockeye; Table 2), with advanced hybrid classes in this system present since at least 2016. Given these trends, further hybrid classes (ex: B₃ backcross sockeye/kokanee) may develop, while the occurrence of pure-stock sockeye salmon is likely to remain due to on-going stocking programs using broodstock collected downstream from Skaha Lake.

Understanding the fitness outcomes of hybridization is important when considering the overall productivity of the system and long-term management goals. Our results and those of a previous study²⁹ suggest that hybrid *O. nerka* overwhelmingly exhibit a resident life history, but can express a migratory life history with at least one F₁ hybrid spending time in the marine environment (Supplementary Table S1). Progressive hybrid classes may also prefer a resident life history²⁹ with similarities in size between F₂ hybrids and B₂ backcross-kokanee. However, it is still possible that backcrosses and F₂ hybrids expressing anadromy were subject to increased mortality due to smaller size at migration or genomic incompatibilities, limiting their detection within the sampled pool of Okanagan River spawners. These trends are consistent with other salmonids, where cutthroat and steelhead trout hybrids have been found to exhibit intermediate migratory behaviors when compared to parental species that may be maladaptive to their local environment⁵⁸. Such increased migration mortality in advanced hybrid classes expressing anadromy may signify a decrease of *O. nerka* productivity in the system. To examine the link between intermediate hybrid behavior and survival moving forward, a combination of identification and tracking methodologies may be appropriate, using passive integrated transponder (PIT) tags to mark juvenile movement, and subsequent comparison of recovered tags and genetic hybridization classes to evaluate survival and age at maturity⁵⁹.

The population level outcomes arising from an intermediate body size in hybrids may ultimately be detrimental to the fitness of migratory hybrids compared to the larger-bodied sockeye salmon, as body size is often directly correlated with fecundity²⁸. Conversely, the overall increase in body size of the resident population resulting from the presence of hybrids in the system may be advantageous for some management strategies, especially those that target increasing recreational angling quality or harvest opportunities for First Nations, as larger bodied fish generally have increased survival and are often more valued as a food source.

From a stock assessment perspective, our results emphasize that, though morphology and body size have been used historically to determine hybrid class in this system and others, advanced hybrid classes would be indistinguishable from pure-stock, with the B₂ backcross-kokanee, F₂ and kokanee classes having no significant mean size differences⁶⁰. Therefore, if stock assessment programs and management goals require an estimate of stock composition, genetic tools such as GT-seq panels can be an effective enumeration method for their ability to accurately detect hybrid classes and expand sample sizes.

Here, we demonstrated the effectiveness of GT-seq for identifying advanced hybrid classes and, in this case, tracking the trajectory of sockeye salmon, kokanee, and hybrid stocks as part of a re-introduction program. Given the uncertainty in the long-term composition of the population and the range of potential outcomes and impacts on management goals, continued genetic monitoring of this system is recommended. More broadly, the GT-seq panel development workflow presented here could be applied to inform other sockeye salmon restoration initiatives in the Columbia and Fraser River systems, or in other species and systems where genetic stock identification and intra-specific hybridization are important management parameters.

Data availability

All probe sequences and SNP genotypic data collected via GT-seq have been deposited in DRYAD (<https://doi.org/10.5061/dryad.z34tmgg4>).

Received: 31 December 2021; Accepted: 17 February 2022

Published online: 02 March 2022

References

- Winston, M. R. & Taylor, C. M. Upstream extirpation of four minnow species due to damming of a prairie stream. *Trans. Am. Fish. Soc.* **120**, 8 (1991).
- Graham, K. Contemporary status of the North American paddlefish, *Polyodon spathula*. *Environ. Biol. Fishes* **48**, 279–289 (1997).
- Kaushal, S. S. *et al.* Rising stream and river temperatures in the United States. *Front. Ecol. Environ.* **8**, 461–466 (2010).
- Vörösmarty, C. J. *et al.* Global threats to human water security and river biodiversity. *Nature* **467**, 555–561 (2010).
- Galbreath, P. F., Bisbee, M. A., Dompier, D. W., Kamphaus, C. M. & Newsome, T. H. Extirpation and tribal reintroduction of coho salmon to the interior Columbia river basin. *Fisheries* **39**, 77–87 (2014).
- Schmidt, B. A. *et al.* Determining habitat limitations of Maumee River walleye production to western Lake Erie fish stocks: Documenting a spawning ground barrier. *J. Great Lakes Res.* **46**, 1661–1673 (2020).
- Kendall, N. W., Marston, G. W. & Klungbein, M. M. Declining patterns of Pacific Northwest steelhead trout (*Oncorhynchus mykiss*) adult abundance and smolt survival in the ocean. *Can. J. Fish. Aquat. Sci.* **74**, 1275–1290 (2017).
- Myers, J., Bryant, G. & Lynch, J. *Factors Contributing to the Decline of Chinook Salmon: An Addendum to the 1996 West Coast Steelhead Factors for Decline Report* (Springer, 1998).
- Molony, B. W., Lenanton, R., Jackson, G. & Norriss, J. Stock enhancement as a fisheries management tool. *Rev. Fish Biol. Fish.* **13**, 409–432 (2005).
- Merz, J. E. & Setka, J. D. Evaluation of a spawning habitat enhancement site for Chinook salmon in a regulated California river. *N. Am. J. Fish. Manag.* **24**, 397–407 (2004).
- Ostberg, C. O., Chase, D. M. & Hauser, L. Hybridization between yellowstone cutthroat trout and rainbow trout alters the expression of muscle growth-related genes and their relationships with growth patterns. *PLoS ONE* **10**, e0141373 (2015).
- Veale, A. J. & Russello, M. A. Sockeye salmon repatriation leads to population re-establishment and rapid introgression with native kokanee. *Evol. Appl.* **9**, 1301–1311 (2016).
- Fraser, D. J., Cook, A. M., Eddington, J. D., Bentzen, P. & Hutchings, J. A. Mixed evidence for reduced local adaptation in wild salmon resulting from interbreeding with escaped farmed salmon: Complexities in hybrid fitness. *Evol. Appl.* **1**, 501–512 (2008).
- Stewart, G. S. *et al.* The power of evolutionary rescue is constrained by genetic load. *Evol. Appl.* **10**, 731–741 (2017).
- Weeks, A. R. *et al.* Genetic rescue increases fitness and aids rapid recovery of an endangered marsupial population. *Nat. Commun.* **8**, 1071 (2017).
- Chan, W. Y., Hoffmann, A. A. & van Oppen, M. J. H. Hybridization as a conservation management tool. *Conserv. Lett.* **12**, e12652 (2019).
- Bekkevold, D., Hansen, M. M. & Nielsen, E. E. Genetic impact of gadoid culture on wild fish populations: Predictions, lessons from salmonids, and possibilities for minimizing adverse effects. *ICES J. Mar. Sci.* **63**, 198–208 (2006).
- Muhlfeld, C. C. *et al.* Hybridization rapidly reduces fitness of a native trout in the wild. *Biol. Lett.* **5**, 328–331 (2009).
- Harvey, A. C., Glover, K. A., Taylor, M. I., Creer, S. & Carvalho, G. R. A common garden design reveals population-specific variability in potential impacts of hybridization between populations of farmed and wild Atlantic salmon, *Salmo salar* L. *Evol. Appl.* **9**, 435–449 (2016).
- Edmands, S. Does parental divergence predict reproductive compatibility?. *Trends Ecol. Evol.* **17**, 520–527 (2002).
- Johnson, B. M., Johnson, M. S. & Thorgaard, G. H. Salmon genetics and management in the Columbia river basin. *Northwest Sci.* **92**, 346–363 (2019).
- Hanson, A. J. & Smith, H. D. Mate selection in a population of sockeye salmon (*Oncorhynchus nerka*) of mixed age-groups. *J. Fish. Board Can.* **24**, 23 (1967).
- Wood, C. C. & Foote, C. J. Evidence for sympatric genetic divergence of anadromous and nonanadromous morphs of sockeye salmon (*Oncorhynchus nerka*). *Evolution* **50**, 1265–1279 (1996).
- Foote, C. J. Male mate choice dependent on male size in salmon. *Behaviour* **106**, 63–80 (1988).
- Craig, J. K., Foote, C. J. & Wood, C. C. Countergradient variation in carotenoid use between sympatric morphs of sockeye salmon (*Oncorhynchus nerka*) exposes nonanadromous hybrids in the wild by their mismatched spawning colour. *Biol. J. Linn. Soc.* **84**, 287–305 (2005).
- Taylor, E. B. & Foote, C. J. Critical swimming velocities of juvenile sockeye salmon and kokanee, the anadromous and non-anadromous forms of *Oncorhynchus nerka* (Walbaum). *J. Fish Biol.* **38**, 407–419 (1991).
- Foote, C. J., Wood, C. C., Clarke, W. C. & Blackburn, J. Circannual cycle of seawater adaptability in *Oncorhynchus nerka*: Genetic differences between sympatric sockeye salmon and kokanee. *Can. J. Fish. Aquat. Sci.* **49**, 99–109 (1992).
- Wood, C. C. & Foote, C. J. Genetic differences in the early development and growth of sympatric sockeye salmon and kokanee (*Oncorhynchus nerka*), and their hybrids. *Can. J. Fish. Aquat. Sci.* **47**, 2250–2260 (1990).
- Elliott, L. D., Ward, H. G. M. & Russello, M. A. Kokanee–sockeye salmon hybridization leads to intermediate morphology and resident life history: Implications for fisheries management. *Can. J. Fish. Aquat. Sci.* **77**, 355–364 (2020).
- Hendry, A. P., Quinn, T. P. & Utter, F. M. Genetic evidence for the persistence and divergence of native and introduced sockeye salmon (*Oncorhynchus nerka*) within Lake Washington, Washington. *Can. J. Fish. Aquat. Sci.* **53**, 823–832 (1996).
- Praebel, K. *et al.* A diagnostic tool for efficient analysis of the population structure, hybridization and conservation status of European whitefish (*Coregonus lavaretus* (L.)) and vendace (*C. albula* (L.)). *Adv. Limnol.* **64**, 247–255 (2013).
- Sanz, N., Araguas, R. M., Fernández, R., Vera, M. & García-Marin, J.-L. Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. *Conserv. Genet.* **10**, 225–236 (2009).
- McFarlane, S. & Pemberton, J. Detecting the true extent of introgression during anthropogenic hybridization. *Trends Ecol. Evol.* **34**, 315–326 (2019).
- Vähä, J.-P. & Primmer, C. R. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Mol. Ecol.* **15**, 63–72 (2006).
- Boecklen, W. J. & Howard, D. J. Genetic analysis of hybrid zones: Numbers of markers and power of resolution. *Ecology* **78**, 2611–2616 (1997).
- Elliott, L. & Russello, M. A. SNP panels for differentiating advanced-generation hybrid classes in recently diverged stocks: A sensitivity analysis to inform monitoring of sockeye salmon re-stocking programs. *Fish. Res.* **208**, 339–345 (2018).
- Twyford, A. D. & Ennos, R. A. Next-generation hybridization and introgression. *Heredity* **108**, 179–189 (2012).
- Campbell, N. R., Harmon, S. A. & Narum, S. R. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Mol. Ecol. Resour.* **15**, 855–867 (2015).
- Alexander, C. A. & Pickard, D. *Skaha Lake Experimental Sockeye Reintroduction: Synthesis of First 4 of 12 Years (2004–2007 Brood Years)* (Springer, 2009).
- McQueen, D. *et al.* *Evaluation of the Experimental Introduction of Sockeye Salmon (*Oncorhynchus nerka*) into Skaha Lake and Assessment of Sockeye Rearing in Osoyoos Lake* (Springer, 2013).
- Hegg, J. C., Kennedy, B. P. & Chittaro, P. What did you say about my mother? The complexities of maternally derived chemical signatures in otoliths. *Can. J. Fish. Aquat. Sci.* **76**, 81–94 (2019).
- Veale, A. J. & Russello, M. A. Genomic changes associated with reproductive and migratory ecotypes in sockeye salmon (*Oncorhynchus nerka*). *Genome Biol. Evol.* **9**, 2921–2939 (2017).

43. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
44. Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W. & Luikart, G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol. Ecol. Resour.* **11**, 117–122 (2011).
45. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
46. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
47. Rousset, F. genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106 (2008).
48. Anderson, E. C. & Thompson, E. A. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217–1229 (2002).
49. Schmidt, D. A., Campbell, N. R., Govindarajulu, P., Larsen, K. W. & Russello, M. A. Genotyping-in-Thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Mol. Ecol. Resour.* **20**, 114–124 (2020).
50. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
51. Reeves, P. A., Bowker, C. L., Fetting, C. E., Tembrock, L. R. & Richards, C. M. *Effect of Error and Missing Data on Population Structure Inference Using Microsatellite Data.* (2016) <https://doi.org/10.1101/080630>.
52. Wringe, B. F., Stanley, R. R. E., Jeffery, N. W., Anderson, E. C. & Bradbury, I. R. hybriddetective: A workflow and package to facilitate the detection of hybridization using genomic data in R. *Mol. Ecol. Resour.* **17**, e275–e284 (2017).
53. Walsh, P. S., Metzger, D. A. & Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506–513 (1991).
54. Russell, T. *et al.* Development of a novel mule deer genomic assembly and species-diagnostic SNP panel for assessing introgression in mule deer, white-tailed deer, and their interspecific hybrids. *Genes Genomes Genet.* **9**, 911–919 (2019).
55. Thongda, W. *et al.* Species-diagnostic SNP markers for the black basses (*Micropterus* spp.): A new tool for black bass conservation and management. *Conserv. Genet. Resour.* **12**, 319–328 (2020).
56. Ricker, W. E. 'Residual' and kokanee salmon in Cultus lake. *J. Fish. Board Can.* **27**, 192–218 (1938).
57. Crossin, G. T. *et al.* Exposure to high temperature influences the behaviour, physiology, and survival of sockeye salmon during spawning migration. *Can. J. Zool.* **86**, 127–140 (2008).
58. Moore, M. E. *et al.* Early marine migration patterns of wild coastal cutthroat trout (*Oncorhynchus clarkii clarkii*), steelhead trout (*Oncorhynchus mykiss*), and their hybrids. *PLoS ONE* **5**, e12881 (2010).
59. McCutcheon, C. S., Prentice, E. F. & Park, D. L. Passive monitoring of migrating adult steelhead with PIT tags. *N. Am. J. Fish. Manag.* **14**, 220–223 (1994).
60. Scribner, K. T., Page, K. S. & Bartron, M. L. Hybridization in freshwater fishes: A review of case studies and cytonuclear methods of biological inference. *Rev. Fish Biol. Fish.* **10**, 293–323 (2001).

Acknowledgements

We would like to acknowledge the support of the Freshwater Fisheries Society of British Columbia and BC's freshwater anglers for helping fund recreational fisheries research projects, through their license fees. We also thank the staff of the BC Ministry of Forests, Lands, Natural Resource Operations and Rural Development and their contractors for providing samples associated with this project. We are further grateful to Ty Russell (University of Alberta) for providing and assisting with custom code for *hybriddetective* sensitivity analyses. This work was funded by the Freshwater Fisheries Society of British Columbia.

Author contributions

Conception: M.A.R. Study design: M.A.R., H.G.M.W., S.L.C. Data collection: S.L.C. Data analysis: S.L.C., L.D.E. Interpretation of the data: S.L.C., M.A.R., H.G.M.W. Drafting of the article: S.L.C., M.A.R., H.G.M.W. Critical revision of the article for important intellectual content: S.L.C., M.A.R., H.G.M.W., L.D.E.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-07309-x>.

Correspondence and requests for materials should be addressed to M.A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022