Site C Clean Energy Project

Fisheries and Aquatic Habitat Monitoring and Follow-up Program

Fish Genetics Study
2020 Status Report for Bull Trout, Arctic Grayling and Rainbow Trout

Construction Year 6 (2020)

Armando Geraldes, PhD
University of British Columbia

Eric B. Taylor, PhD
University of British Columbia

May 6, 2021

EXECUTIVE SUMMARY

BC Hydro is currently constructing the Site C Clean Energy Project (the Project) near the town of Fort St. John in northeastern British Columbia which will be the third hydroelectric dam on the Peace River. BC Hydro developed the Site C Fisheries and Aquatic Habitat Monitoring and Follow-up Program (FAHMFP) in accordance with Provincial Environmental Assessment Certificate Condition No. 7 and Federal Decision Statement Condition Nos. 8.4.3 and 8.4.4 for the Project. To date, Mon-1b, Task 2c (Site C Reservoir Tributaries Fish Population Indexing Survey), Mon-2, Task 2a (Peace River Large Fish Indexing Survey) and Mon-2, Task 2b (Peace River Fish Composition and Abundance Survey) of the FAHMFP have collected DNA samples from Bull Trout (*Salvelinus confluentus*), Arctic Grayling (*Thymallus arcticus*) and Rainbow Trout (*Oncorhynchus mykiss*). The Site C Fish Genetics Study aims to: (a) determine levels and patterns of population structure for the three fish species in the Peace River and its tributaries, (b) develop genotyping assays for genetic monitoring of the system, and (c) deploy these assays in an initial number of samples available for analysis. Here we report on the progress of the Site C Fish Genetics Study from January 1, 2020 to December 31, 2020. The results and status from earlier components of this study can be found in Geraldes and Taylor (2020).

A total of 4,058 genetic samples have been collected between 2016 and 2020, from all three species. These samples were shipped to UBC where they have been stored and catalogued. For Bull Trout, 2,616 samples were received at UBC. The DNA was extracted from 1,794 Bull Trout samples, including all 2016 (N=108), 2017 (N=753) and 2018 (N=711) samples, as well as all Peace River Bull Trout samples from 2019

(N=191). All 191 Peace River Bull Trout samples collected in 2019 were genotyped at six loci previously developed and shown to be sufficient to assign samples with high confidence to two genetic groups identified with genome-wide data: one group consisted of samples that spawn upstream of the Project (UP) in the Halfway River, and the other consisted of samples that spawn downstream of the Project (DP) in the Pine River (Geraldes and Taylor 2020). Genetic analysis allowed most samples collected in 2019 to be assigned to one of the two groups with more than 95% confidence, with the vast majority being assigned to the Halfway group (N=180, 94.2% of all samples) and a small number being assigned to the Pine group (N=6, 3.1% of all samples). Five samples could not be assigned to either UP or DP with 95% confidence (2.6% of all samples). Of these latter samples, three fish were assigned to UP with 64, 93 and 95% confidence and two were assigned to DP with 61 and 89% confidence.

For Arctic Grayling, 494 samples collected between 2018 and 2020 have been received and catalogued at UBC. The DNA has been extracted and quality controlled for 320 of those samples. The vast majority of DNAs yet to be extracted were collected in 2020 (N=172). A subset of samples from Peace River tributaries where Arctic Grayling are known to spawn (Halfway, Moberly, Pine, and Beatton rivers) were selected for sequencing of small fragments throughout the species genome. The sequencing data was then used to detect genetic variants and determine levels and patterns of population structure among samples from the different tributaries. Analysis of these data revealed clear genetic differentiation between Arctic Grayling caught DP in the Beatton River watershed and samples caught UP in the Moberly River. By contrast, genetic differentiation between the Halfway River (UP) and the Pine River (DP)

watersheds was low. Eight genetic variants showing maximal differentiation between the two UP tributaries and the two DP tributaries were selected and used to develop genotyping assays to determine the tributary of origin of individual Arctic Grayling. These were supplemented with one locus showing maximal differentiation between the Halfway River watershed and all others and one showing maximal differentiation between the Pine River watershed and others, for a total of ten loci. Seven of those ten assays were used to genotype 146 Arctic Grayling samples collected in the Peace River mainstem between 2016 and 2019. This allowed 123 samples (84.2%) to be assigned to UP tributaries and 11 samples (7.5%) to be assigned to DP tributaries with more than 95% confidence. Twelve samples (8.2%) could not be assigned with at least 95% confidence to spawning UP or DP areas, but of these, five fish could be assigned to DP with between 58 and 92% confidence, five could be assigned to UP with between 55 and 93% confidence and two could not be assigned to either UP or DP with more than 50% confidence. We are currently designing additional genotyping assays to improve assignment power for Arctic Grayling.

Finally, for Rainbow Trout, 948 samples collected between 2017 and 2020 have been received and catalogued at UBC. The DNA has been extracted and quality controlled for 139 of those samples. To investigate levels and patterns of genetic differentiation among fish representative of possible provenances of fish caught in the Peace River mainstem, we selected for genome sequencing samples from a) large tributaries of the Peace River (Halfway, Moberly and Pine rivers), b) smaller tributaries of the Peace River (Farrell, Lynx and Maurice creeks), c) the Dinosaur Reservoir (created by Peace Canyon Dam located UP) and d) three strains known to be used for

restocking of fish in the area (Pennask Lake, Blackwater River, and Fraser Valley Domestic). These samples have been sequenced and mapped to the Rainbow Trout reference genome sequence. We are currently analysing those data to identify genetic variants, determine levels and patterns of genetic differentiation among those possible provenances, and design assays to genotype and assign fish to UP and DP tributaries.

ACKNOWLEDGMENTS

This work was supported by a contract from BC Hydro. We thank Nich Burnett and Brent Mossop for guidance and helpful discussions, and BC Hydro for providing the map in Figure 2. Dustin Ford of Golder Associates and Mark LeRuez of Triton Environmental Consultants are thanked for assistance with sample collection, provision and orientation. Jessica Shen and Kately Nikiforuk provided excellent assistance in the laboratory. Much of the laboratory work was supported by the Natural Sciences and Engineering Research Council of Canada grants awarded to EBT (Discovery and Equipment grant programs).

LIST OF ACRONYMS AND ABBREVIATIONS

bp	Base pairs
CVE	Cross validation error
DNA	Deoxyribonucleic Acid
DP	Downstream of the Project
FAHMFP	Fisheries and Aquatic Habitat Monitoring and Follow-up Program
GBS	Genotyping-by-sequencing
K	Number of clusters in the Admixture analysis
LAA	Local Assessment Area
PCA	Principal components analysis
PCR	Polymerase chain reaction
QC	Quality control
SNP	Single nucleotide polymorphism
SRA	Short read archive
UP	Upstream of the Project

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

of each sampling section and the bottom panel shows each section of the Peace River in detail. Map courtesy of BC Hydro. Full details can be found in Appendix III.

Supporting Figure 1. Cross Validation Error (CVE) for the Arctic Grayling Admixture analysis. Results are shown for the five replicates of the analysis with the 15,209 SNPs dataset with K varying from 1 to 7. A value of K=3, i.e., three populations, minimized the CVE and is therefore the best fit to our data. A value of K=4, i.e., four populations, showed a CVE only marginally higher.

LIST OF APPENDICES

Appendix I. A. Arctic Grayling project sample information, sequencing results, genomic coverage, frequency of missing genotypes, principal components analysis results, and Admixture analysis results (both the admixture coefficients and the standard errors of those estimates for K varying from 2 to 4). B. Details of all Arctic Grayling samples genotyped at seven loci using TaqMan assays (ag13b01, ag20b03, ag01b04, ag17a14, ag19b18, ag15b24 and ag23aha1), their genotypes at those loci and their assignment to the upstream (UP) or downstream (DP) of the project reference groups. File name: "Appendix_I_AG_GBS_TaqMan_BCH2021report.xlsx"

Appendix II. Rainbow Trout project sample information, sequencing results and genomic coverage. File name: "Appendix_II_RB_GBS_BCH2021report.xlsx"

Appendix III. Details of all Bull Trout samples genotyped at six loci using TaqMan assays (BT9_1, BT8_5, BT8_7, BT25_8, BT27_15 and BT18_16), their genotypes at those loci and their assignment to the upstream (UP) or downstream (DP) of the project reference groups. File name: "Appendix_III_BT_TaqMan_BCH2021report.xlsx"

INTRODUCTION

BC Hydro is currently constructing the Site C Clean Energy Project (the Project) near the town of Fort St. John in northeastern British Columbia (hereafter referred to as Local Assessment Area, LAA) which will be the third hydroelectric dam on the Peace River. BC Hydro and the laboratory of Eric Taylor at the University of British Columbia, Department of Zoology, entered into a three-year agreement in 2018 to apply genomic techniques to facilitate aspects of the mitigation and monitoring plan for the LAA. This work is to initially focus on three important recreational sport fishes: Bull Trout (*Salvelinus confluentus*), Arctic Grayling (*Thymallus arcticus*), and Rainbow Trout (*Oncorhynchus mykiss*) which are common in the LAA. These efforts are directly tied to the Site C Fisheries and Aquatic Habitat Monitoring and Follow-up Program (FAHMFP) that BC Hydro developed in accordance with Provincial Environmental Assessment Certificate Condition No. 7 and Federal Decision Statement Condition Nos. 8.4.3 and 8.4.4 for the Project. BC Hydro is using diverse lines of evidence to better understand the population structure, migration and movement patterns, and tributary use of these key fish species in the Peace River and its tributaries. Such evidence includes data from otolith and fin ray microchemistry, radio telemetry, fish distribution, and genetics that are being used to answer management questions and test management hypotheses posed in the FAHMFP.

Purpose and Objectives

The Site C Fish Genetics Study aims to: (a) determine levels and patterns of population structure for Bull Trout, Arctic Grayling and Rainbow Trout in the Peace River and its

tributaries in the LAA, (b) develop genotyping assays for genetic monitoring of the system, and (c) deploy these assays in an initial number of samples available for analysis. Geraldes and Taylor (2020) reported on the first year of genetic work contributing to the FAHMFP. In that report, the authors summarized genomic work focused on using genotyping-by-sequencing (GBS) across the genome to resolve differences among samples of Bull Trout collected from three tributaries of the Peace River (Halfway, Moberly, and Pine rivers) from 2016 to 2018. That work revealed pronounced differences between fish that spawn in tributaries upstream (Halfway River) and downstream (Pine River) of the Project and those genome-wide differences were used to develop a set of six TaqMan genotyping assays that differentiated samples collected from the mainstem Peace River (n=473) in terms of whether an individual fish belonged to a spawning population located upstream of the Project (UP, i.e., in the Halfway River; ~92% of all samples) or downstream of the Project (DP, i.e., in the Pine River; ~4% of all samples). About 4% of all mainstem Peace River samples of Bull Trout could not be confidently (i.e., with >95% confidence) assigned to either the Halfway or Pine river spawning groups. Full details of the rationale, methods, and results of the first year's genetic work can be found in Geraldes and Taylor (2020).

The objectives of this report are to summarize the work during the second year of the study to the end of 2020. Specifically, the report summarizes: (i) population assignment work for samples of Bull Trout collected from 2016 to 2019, (ii) results from sample processing and GBS analysis of Arctic Grayling samples from the LAA, (iii) results from an initial set of seven TaqMan genotyping assays for Arctic Grayling derived from the GBS data, and (iv) initial sample processing and GBS analyses for

Rainbow Trout collected from the LAA. The third year of the study (2021) will focus on: i) the analysis of GBS data and development of TaqMan genotyping assays for Rainbow Trout, ii) development of additional TaqMan assays for Arctic Grayling, and iii) population assignment work for all species in order to complete the assignment of all samples collected from the mainstem of the Peace River from 2016 to 2020.

MATERIALS AND METHODS

Samples

The Mon-1b, Task 2c (Site C Reservoir Tributaries Fish Population Indexing Survey), Mon-2, Task 2a (Peace River Large Fish Indexing Survey) and Mon-2, Task 2b (Peace River Fish Composition and Abundance Survey) activities of the FAHMFP collected 2,616 Bull Trout, 494 Arctic Grayling and 948 Rainbow Trout genetic samples from 2016 to 2020 and stored them in individual vials with 95% ethanol (Table 1). Golder Associates Ltd. and Triton Environmental Consultants shipped the samples to UBC for analysis. Samples were collected from both the Peace River (including Farrell, Maurice and Dry creeks) and its main tributaries: the Halfway, Moberly and Beatton rivers. DNA extraction and quality control (QC) followed Geraldes and Taylor (2020).

Finally, DNA was also extracted from historical DNA samples from our laboratory archive, both in 95% ethanol and dried in paper, and used as described below.

Genome Sequencing for Polymorphism Discovery

Genome sequencing and genetic variant discovery (single nucleotide polymorphisms, SNPs) was performed for Arctic Grayling and Rainbow Trout following protocols similar to those previously described for Bull Trout (Geraldes and Taylor 2020). Here, we detail sample selection for each species and indicate any changes to the library construction protocol and data analysis described in Geraldes and Taylor (2020).

For Arctic Grayling, we selected 88 samples from the four main spawning areas in the LAA, Halfway and Moberly rivers (i.e., those spawning upstream of the Project, UP) and Pine and Beatton rivers (i.e., those spawning downstream of the Project, DP,

Table 2). Because no samples were collected in the Pine or Halfway rivers from 2016 to 2018, historical samples from our lab archive were used (Taylor et al. 2014). A few Beatton and Moberly river samples (Taylor et al. 2014) were added to ensure that differences between watersheds were temporally stable and not artefacts of a particular sample year. Sample details can be found in Appendix I.

Table 1. Bull Trout, Arctic Grayling and Rainbow Trout samples available for genetics work from 2016 to 2020.

Species	Watershed	River/SectionID	Available[1]	DNA[2]	TaqMan[3]
Bull Trout	Halfway River	Chowade River	782	481	21
Bull Trout	Halfway River	Colt Creek	18	14	13
Bull Trout	Halfway River	Cypress Creek	653	400	13
Bull Trout	Halfway River	Fiddes Creek	249	183	12
Bull Trout	Halfway River	Halfway River	7	7	3
Bull Trout	Halfway River	Turnoff Creek	40	40	4
Bull Trout	Moberly River	Moberly River	9	5	5
Bull Trout	Peace River	Dry Creek	10	0	0
Bull Trout	Peace River	Maurice	2	0	0
Bull Trout	Peace River	Section 1	223	181	181
Bull Trout	Peace River	Section 3	290	220	220
Bull Trout	Peace River	Section 5	142	105	105
Bull Trout	Peace River	Section 6	103	90	90
Bull Trout	Peace River	Section 7	58	46	46
Bull Trout	Peace River	Section 9	30	22	22
Bull Trout	All	All	2616	1794	735
Arctic Grayling	Beatton River	Beatton River	37	37	3
Arctic Grayling	Beatton River	Bratland Creek	56	54	14
Arctic Grayling	Beatton River	La Prise Creek	39	39	13
Arctic Grayling	Beatton River	Unnamed Creek 1	1	1	1
Arctic Grayling	Halfway River	Colt Creek	1	1	1
Arctic Grayling	Moberly River	Moberly River	161	42	11
Arctic Grayling	Peace River	Section 1	4	3	3
Arctic Grayling	Peace River	Section 3	93	74	74
Arctic Grayling	Peace River	Section 5	39	23	23
Arctic Grayling	Peace River	Section 6	36	32	32
Arctic Grayling	Peace River	Section 7	21	13	13

Species	Watershed	River/SectionID	Available[1]	DNA[2]	TaqMan[3]
Arctic Grayling	Peace River	Section 9	6	1	1
Arctic Grayling	All	All	494	320	189
Rainbow Trout	Halfway River	Chowade River	14	13	0
Rainbow Trout	Halfway River	Colt Creek	106	11	0
Rainbow Trout	Halfway River	Cypress Creek	27	10	0
Rainbow Trout	Halfway River	Kobes Creek	151	8	0
Rainbow Trout	Peace River	Dry Creek	7	0	0
Rainbow Trout	Peace River	Farrell Creek	178	11	0
Rainbow Trout	Peace River	Maurice Creek	38	0	0
Rainbow Trout	Peace River	Section 1	205	86	0
Rainbow Trout	Peace River	Section 3	182	0	0
Rainbow Trout	Peace River	Section 5	23	0	0
Rainbow Trout	Peace River	Section 6	6	0	0
Rainbow Trout	Peace River	Section 7	10	0	0
Rainbow Trout	Peace River	Section 9	1	0	0
Rainbow Trout	All	All	948	139	0
All	All	All	4058	2253	924

[1]Number of samples received at UBC

[2]Number of samples for which a DNA extraction was performed

[3]Number of samples for which SNP genotyping with TaqMan assays was performed

We also included a sample of European Grayling (*Thymallus thymallus*), a closely related taxon for which a full genome sequence assembly is available, as well as five additional samples of Arctic Grayling from other regions in North America.

For Rainbow Trout, we selected 100 samples for sequencing. The majority were from the main spawning areas in the LAA (Table 3): 54 samples from the larger tributaries of the Peace River (Halfway and Moberly rivers, UP, and the Pine River, DP) and 28 samples from smaller tributaries located UP (Lynx, Maurice and Farrell creeks). We also included 12 samples from the Dinosaur Reservoir, located UP (as these may occasionally migrate downstream through Peace Canyon Dam), and two samples each

from the provenance locations of three strains (Blackwater River, Pennask Lake and Fraser Valley Domestic) that are known to have been used for restocking in the area (Freshwater Fisheries Society of BC, 2021). Historical samples collected before 2016, from our laboratory collection, were added where needed to ensure that all sources above were included and to ensure that differences between watersheds were temporally stable and not artefacts of a particular sampling year. Sample details can be found in Appendix II.

Table 2. Arctic Grayling samples from the LAA used for SNP discovery.

Watershed	Tributary	Year	N
Beatton River	NA[1]	1996	4
Beatton River	Beatton River	2018	3
Beatton River	Bratland Creek	2018	5
Beatton River	LaPrise Creek	2018	5
Beatton River	Unnamed Creek 1	2018	1
Pine River	NA[1]	2008	5
Pine River	Pine River	2010	5
Pine River	Pine River	2011	2
Pine River	Burnt River	2010	2
Pine River	Burnt River	2011	3
Pine River	Murray River	2010	3
Pine River	Murray River	2011	2
Pine River	Perry Creek	2011	2
Halfway River	Halfway River	2010	14
Halfway River	Halfway River	2011	6
Moberly River	NA[1]	2006	4
Moberly River	Moberly River	2010	9
Moberly River	Moberly River	2011	5
Moberly River	Moberly River	2018	8

[1]Tributary information not available.

To increase the chances of capturing as much genetic variability as possible, both for Arctic Grayling and for Rainbow Trout, we selected samples to maximize the

spatial coverage within each watershed. To minimize the chance of sequencing siblings, when multiple samples were included from a single sampling location we did not include fish smaller than 50 mm (likely newly-emerged fry). We also avoided the inclusion of fish larger than 300 mm to maximize the chance that the fish were born in the watershed where they were sampled (for Arctic Grayling only eight out of 80 fish for which length was available were larger than 300 mm and for Rainbow Trout, only five out of 82 samples for which length was available, Dinosaur Reservoir and restocking strains excluded, were larger than 300 mm; Appendices I and II).

Table 3. Rainbow Trout samples used for SNP discovery.

Watershed	Tributary/Strain	Year	N
Halfway River	Chowade River	2017	3
Halfway River	Chowade River	2018	2
Halfway River	Chowade River	2019	1
Halfway River	Colt Creek	2018	4
Halfway River	Colt Creek	2019	2
Halfway River	Cypress Creek	2017	3
Halfway River	Cypress Creek	2018	3
Halfway River	Halfway River	2011	6
Halfway River	Kobes Creek	2018	2
Halfway River	Kobes Creek	2019	4
Pine River	Pine River	2011	3
Pine River	Blind River	2011	5
Pine River	Burnt River	2011	3
Pine River	Willow Creek	2011	5
Moberly River	Shangweshi Creek	2011	8
Peace River	Farrell Creek	2018	4
Peace River	Farrell Creek	2019	6
Peace River	Lynx Creek	2006	8
Peace River	Maurice Creek	2011	10
Peace River	Dinosaur Reservoir	2011	12

Watershed	Tributary/Strain	Year	N
	Pennask Lake	2017	2
	Blackwater River	2014	2
	Fraser Valley	2017	2

For both Arctic Grayling and Rainbow Trout we used a reduced representation genome sequencing approach known as genotyping-by-sequencing (GBS; Elshire et al. 2011) to cost-effectively generate sequence data from a representative fraction of the genome. We used a modified GBS protocol described in detail elsewhere (Alcaide et al. 2014; Towes et al. 2016; Geraldes et al. 2019) and successfully used in this Project for SNP discovery in Bull Trout (Geraldes and Taylor 2020), to generate pooled libraries of digested and individually barcoded DNA. For Arctic Grayling, one library of 94 samples, of which 88 were from the LAA (see above), and two negative controls was prepared. For Rainbow Trout the 100 samples for the current Project (see above) were combined with samples from other projects in our lab (not discussed further) and distributed over two libraries, each having two negative controls. The Arctic Grayling library was sequenced in one lane of an Illumina HiSeq4000 and the two Rainbow Trout libraries were sequenced in one lane each of an Illumina NovaSeq6000 SP, all with 150 bp paired end reads, at the McGill University and Génome Québec Innovation Centre.

SNP Discovery Bioinformatics

For Arctic Grayling and Rainbow Trout we followed a bioinformatics pipeline for GBS read processing and mapping, and variant calling and evaluation, broadly similar to the one used for Bull Trout in this Project (Geraldes and Taylor 2020), which is available at https://doi.org/10.5061/dryad.t951d (Irwin et al. 2016). A couple modifications to the

pipeline used for Bull Trout were introduced: i) reads were demultiplexed with the function “process_radtags” from the STACKS v2.5 pipeline (Catchen et al. 2013) allowing for no more than one mismatch between the barcode sequence and the barcode detected on the sequence reads; and ii) variant identification was performed with GATK4 (Mckenna et al. 2010; instead of GATK3) and the functions “HaplotypeCaller”, “GenomicsDBImport” and “GenotypeGVCFs”.

Sequence reads from Arctic Grayling samples were aligned to the reference genome sequence (assembly ASM434828v1; Sävilammi et al. 2019) of the closely related European Grayling. This is a chromosome level genome assembly, with 94% of the sequence in 51 scaffolds assigned to the 51 chromosomes and the remaining 6% of the assembly in an additional 8,939 scaffolds.

Sequence reads from Rainbow Trout samples were aligned to the most recent and complete of three assemblies of the species genome (assembly USDA_OmykA_1.1, retrieved from https://www.ncbi.nlm.nih.gov/assembly/GCF_013265735.2/ on October 9th, 2020), a chromosome level assembly with 32 scaffolds assigned to 32 chromosomes and 710 short unplaced scaffolds.

We followed the protocols successfully employed in the Bull Trout dataset analysis (Geraldes and Taylor 2020) for SNP filtering in the Arctic Grayling dataset. We first used a custom script (Owens et al. 2016) to eliminate variants that showed an observed heterozygosity of 0.6 or higher, across all samples, as these are likely the result of mapping to paralogous regions of the genome. Then we kept only the 86 samples from the LAA and applied several filtering criteria with VCFtools v0.1.11

(Danecek et al. 2011) to arrive at a set of high-quality SNPs to form the basis of subsequent population genetic analysis. Namely, we eliminated: i) insertion/deletion polymorphisms to retain only SNPs, ii) SNPs with more than two alleles, iii) SNPs with genotype quality below 10 (these have a higher than 10% chance of being incorrect genotypes), iv) SNPs with missing genotypes in more than 30% of samples, and v) low frequency SNPs (SNPs present at a frequency below 5%). For analysis of population structure (see below), we used Plinkv1.9 (Chang et al. 2019) to remove SNPs that were in close linkage with other SNPs in the set (option "--indep-pairwise 50 10 0.2" to eliminate SNPs with r^2 greater than 0.2 in overlapping windows of 50 consecutive SNPs moving 10 SNPs at a time between windows) as they are not independent data points.

Population Genetic Analyses

We used two complementary approaches to infer patterns of population structure in Arctic Grayling. First, we ordinated samples in "genotype space" using principal components analyses (PCA) with the R package SNPrelate (Zheng et al. 2012) to summarize genetic variation into successive orthogonal principal components (PCs). Second, we used the program Admixture v1.3.0 (Alexander et al. 2009) to estimate ancestry proportions for each fish. Admixture is a clustering program that models the probability of the observed genotypes using ancestry proportions and population allele frequencies with a maximum likelihood approach to determine the most likely number of genetic groups (i.e., clusters, K). In this analysis, individual fish can be composed of more than one of these K genetic groups and the analysis provides an estimate of the proportion of each fish's genome composed of each of the K groups (i.e., its admixture

proportions). We ran five replicates of Admixture for each K from 1 to 7 and terminated each run when the difference in log-likelihood between successive iterations fell below 1 x 10^{-9}. We chose the K value that minimized the cross-validation error (CVE), i.e., that best fit the data (Alexander et al. 2009), and made one last run with the two best K values using 1,000 bootstraps to estimate the standard error of the inferred admixture proportions.

To further characterize levels of population differentiation in the study area for Arctic Grayling we used VCFtools (Danecek et al. 2011) to estimate overall weighted mean Weir and Cockerham's F_{ST} (Weir and Cockerham, 1984) across all loci between all pairs of tributaries as well as between the tributaries UP (Halfway and Moberly rivers) and DP (Pine and Beatton rivers).

SNP Genotyping Assays

For Arctic Grayling, we inspected each SNP in descending order of their F_{ST} rank between tributaries UP (Halfway and Moberly rivers) and DP (Pine and Beatton rivers) to determine their suitability for designing custom TaqMan (Applied Biosystems; Foster City, CA, USA) SNP genotyping assays. Each TaqMan assay uses coloured fluorescent "reporter" dyes (VIC and FAM) to efficiently determine the genotype of each fish at a single SNP amplified by quantitative PCR. Specifically, we only selected SNPs for assay design if they: a) had low missing data even if higher genotype filtering criteria were applied (HaplotypeCaller's genotype quality of 20 instead of 10, i.e. probability of an incorrect genotype call is 1% or lower), b) if we had sequence data for most samples for 30 bp upstream and downstream of the SNP, i.e. the flanking region, c) if there were

no other polymorphisms in the flanking region, and so that d) all selected SNPs were from different chromosomes. Eight SNPs that passed these criteria were submitted for TaqMan assay design using the ThermoFisher online design tool and ordered for testing (Table 4). In addition, because inspection of the levels of genetic differentiation among samples of Arctic Grayling caught in the different tributaries revealed that genetic differentiation between the Halfway River group (UP) and the Pine River group (DP) was relatively low (see Results), we used the same procedure above to estimate per locus F_{ST} between the Halfway River and all other tributaries, as well as between the Pine River and all other tributaries and designed and ordered TaqMan assays to genotype the highest F_{ST} SNP for each (Table 4).

Table 4. TaqMan assays ordered and tested to genotype Arctic Grayling samples. Shaded are the seven TaqMan assays that passed testing and were used for SNP genotyping.

TaqMan Assay	SNP name	LG	Rationale[1]	F_{ST}	F_{ST} Rank
ag13b01	CM014997.1:17726527	13b	UP/DP	0.71	1
ag20b03	CM015014.1:28434291	20b	UP/DP	0.64	3
ag01b04	CM015012.1:19976165	1b	UP/DP	0.62	4
ag03a10	CM015027.1:6982627	3a	UP/DP	0.55	10
ag24a12	CM015021.1:22633947	24a	UP/DP	0.55	12
ag17a14	CM015005.1:26896876	17a	UP/DP	0.54	14
ag19b18	CM015010.1:15356601	19b	UP/DP	0.51	18
ag15b24	CM015002.1:21619794	15b	UP/DP	0.48	24
ag23aha1	CM015019.1:14738697	23a	Halfway/others	1.00	1
ag10bpi1	CM014991.1:26347272	10b	Pine/others	0.77	1

[1]Describes whether a SNP was selected based on it being a top ranking F_{ST} SNP between tributaries upstream (UP) and downstream (DP) of the Project, or top ranking F_{ST} between a single tributary and all others. F_{ST} and F_{ST} rank refer to the value and rank specific to that rationale.

We selected an initial set of 31 Arctic Grayling samples for assay testing (seven from the Halfway River watershed, seven from the Moberly River watershed, ten from the Pine River watershed and seven from the Beatton River watershed). All 31 samples had been used for the GBS experiment and had amongst them representatives of all

three possible genotypes at all but one locus. Genotyping of Arctic Grayling samples followed the protocol for Bull Trout genotyping (Geraldes and Taylor 2020). Seven of the ten TaqMan assays (Table 4) passed this initial testing and were used for genotyping a total of 307 Arctic Grayling samples.

All Bull Trout samples collected in the Peace River in 2019 (N=191) as well as additional Bull Trout collected in the tributaries (N=93) were genotyped with the six TaqMan loci described in Geraldes and Taylor (2020).

Assignment Tests

We used our TaqMan generated genotype datasets for Bull Trout and Arctic Grayling to assign samples of each species to spawning tributaries UP and DP. We used the program GeneClass2 (Piry et al., 2004) to assign samples to UP and DP following the method of Rannala and Mountain (1997). Samples were considered assigned to UP or DP if they had 95% or higher probability of being from that group (i.e., $P<0.05$). To do this, for each species, we defined three sets of samples: Reference, Test and Assignment samples. Reference samples were samples from the Peace River tributaries for which we have GBS data and an inferred UP or DP genetic group (see Results) and to which we want to assign samples caught in the Peace River mainstem. Test samples were also samples collected in the Peace River tributaries, but which were not used as reference samples: they allow us to test how well the assignment tests perform. These are samples for which we do not have GBS data and hence, we assume that they belong to the genetic group of the tributary where they were caught.

Assignment samples were those caught in the Peace River mainstem that we want to assign to the UP and DP genetic groups found with the GBS data.

For Bull Trout, Geraldes and Taylor (2020) found two genetic groups: one consisted of samples caught in the Pine River watershed located DP (and henceforth called DP), and the other of samples caught UP in the Halfway and Moberly rivers watersheds (and henceforth called UP). As discussed in Geraldes and Taylor (2020), there was no genetic differentiation between fish caught in the Halfway and the Moberly Rivers and because Bull Trout do not spawn in the Moberly River, the fish caught in the Moberly River belong to the Halfway River spawning population. For Bull Trout we set as reference samples for DP, 37 samples collected in the Pine River and its tributaries for which we had GBS data and for which we inferred that more than 80% of their genome had ancestry in the DP genetic group (Geraldes and Taylor 2020). Similarly, we set as reference samples for UP, 45 samples collected in the Halfway River and its tributaries and five Moberly River samples for which we had GBS data and for which we inferred that more than 80% of their genome had ancestry in UP genetic group (Geraldes and Taylor 2020). For test samples we used 48 samples for which we do not have GBS data: 27 samples were caught in the Halfway and Moberly rivers watersheds and 21 samples were caught in the Pine River watershed. Additionally, we included in this group four samples for which we do have GBS data but for which less than 80% of their genome had ancestry in the genetic group of the tributary where they were caught (three samples from the Pine River and one from the Halfway River). For the assignment group we used 664 Peace River samples including 545 samples previously

assigned in Geraldes and Taylor (2020). We repeated the assignment of those samples because we now expanded our set of reference samples.

As above, we defined three sets of samples of Arctic Grayling. The reference samples are all those for which we have GBS data. For the UP reference set we used all 44 samples caught UP in the Halfway and Moberly rivers and for the DP reference group all 42 samples caught DP in the Pine and Beatton rivers. As test samples we used 75 samples caught in the four tributaries and for which GBS data was not available (UP test samples are 14 from the Halfway River and 18 from the Moberly River and DP test samples are 21 from the Pine River and 22 from the Beatton River). For the assignment group we used all 146 Peace River samples collected in 2018 and 2019.

RESULTS

DNA Extractions

Across all three species, DNA extraction and QC was performed for 2,253 samples out of a total of 4,058 samples collected between 2016 and 2020 (Table 1).

For Bull Trout, DNA was extracted from all samples collected between 2016 and 2018, as well as all 191 samples collected in 2019 from the Peace River and 31 samples collected in 2019 from tributaries (out of a total of 520 collected in 2019). In total 1,794 Bull Trout DNA extractions from 2016 to 2019 are available of which 222 were extracted in the current reporting year of 2020.

For Arctic Grayling, DNA extraction was performed for all samples from 2018 (N=95) and 2019 (N=227; no samples were collected in years 2016 and 2017). Overall,

322 DNA extractions were performed (out of 494 samples available, Table 1) of which 227 were performed in the current reporting year of 2020. DNA extraction from two samples from Bratland Creek in the Moberly River failed to recover usable DNA.

For Rainbow Trout, DNA extraction was performed for 139 of 948 samples available from 2017 to 2020 (Table 1) during the current reporting year of 2020.

SNP Discovery

For Arctic Grayling, sequencing of the GBS library resulted in close to 380 million pairs of sequence reads. Inspection of the barcode sequence with process_radtags resulted in 87.9% of reads being demultiplexed and assigned to a sample. This dataset will be archived and made publicly available on the Short Read Archive (SRA). An average of over 7.5 million reads were assigned to each sample (Table 5 and Appendix I) and after read QC an average of over 6.6 million reads per sample were retained (Table 5 and Appendix I). Of the 94 samples in the library, two samples, one from the Moberly River and one from the Halfway River, had much lower number of reads retained at this stage (less than 200,000) and were eliminated from further analysis (Appendix I). During library construction those same samples had PCR concentrations much lower than other samples and close to the PCR concentration of the two negative controls (Appendix I) suggesting that library construction failed for those samples. The reads for the remaining 92 samples were mapped onto the genome sequence of the European Grayling and on average over 5.2 million reads per sample (76.0% of the reads that passed QC) were used for SNP calling (Table 5 and Appendix I). Just as for Bull Trout (Geraldes and Taylor 2020) this rate of read mapping was likely not due to the fact that

we mapped the reads to the genome of a different species (79.2% of the reads from our sample of European Grayling mapped with high quality to the European Grayling reference genome sequence and were used for SNP calling; Appendix I), but because the reference genome sequence may be incomplete, some reads may be from exogenous DNA and some reads may be from repetitive or duplicated regions of the genome to which high quality mapping is challenging. On average, 2.07% of the genome of each sample had at least one read mapping to it with high quality (mapping quality of 20 or higher, i.e., with a probability of an incorrect alignment lower than 0.01; Table 5).

Table 5. Comparison of GBS sequencing output, read mapping, and genome coverage, for the three species surveyed in this project.

		Bull Trout	Arctic Grayling	Rainbow Trout
Reads per sample[a]	Average	3,788,138	7,561,469	10,013,931
	Max	11,662,316	14,432,040	16,791,758
	Min	2,186,342	158,338	557,634
QC reads per sample[b]	Average	3,788,069	6,650,546	9,457,152
	Max	11,662,135	8,904,916	15,205,423
	Min	2,186,273	664	297,606
Reads used for SNP calling (%)[c]	Average	62.1%	76.0%	58.8%
	Max	68.7%	79.2%	73.9%
	Min	54.0%	73.1%	45.9%
Genomic positions covered (%)[d]	Average	1.02%	2.07%	2.12%
	Max	1.27%	3.27%	2.58%
	Min	0.84%	1.85%	1.53%

[a]Number of reads assigned to each sample after demultiplexing.

[b]Number of reads per sample that passed quality control (QC).

[c]Proportion of total reads per sample that passed quality control that were used to identify genetic variants. Samples with low numbers of reads generated were eliminated from the dataset and not used for this calculation.

[d]Proportion of the total positions in the reference sequence genome with at least one sequence read mapping to it with high quality (mapping quality of 20 or higher) per sample. Samples with low numbers of reads generated were eliminated from the dataset and not used for this calculation.

We identified over 1.9 million variants across the 92 Arctic Grayling samples retained (86 samples from the four Peace River tributaries, five from other watersheds and one European Grayling), of those 1,042,267 variants were biallelic SNPs (i.e., insertion and deletion variants were eliminated, as well as SNPs with more than two variants segregating) with observed heterozygosity below 0.6. We then kept 272,295 of those SNPs that were present in the 86 Arctic Grayling samples from the LAA which had less than 30% missing genotypes and genotype quality of 10 or higher (i.e., genotype call accuracy of 90%). Finally, for estimation of genetic differentiation (F_{ST}), we kept 24,735 SNPs after eliminating variants segregating at low frequency in our sample (minor allele frequency below 5%). For estimation of population structure with PCA and Admixture, we further filtered our dataset to include only unlinked SNPs which resulted in 15,209 SNPs. These final datasets used for downstream analyses will be deposited on DRYAD.

For Rainbow Trout, sequencing of the two GBS libraries resulted in close to 1,034 (520 and 514) million pairs of sequence reads. Inspection of the barcode sequence with process_radtags resulted in 95.1% and 94.3% of reads per library pool being demultiplexed and assigned to a sample. This dataset will be archived and made publicly available on the SRA. Considering only the 94 samples from the LAA and the six samples from restocking strains an average of over 10 million reads were assigned to each sample (Table 5 and Appendix II) and after read quality control an average of close to 9.5 million reads were retained per sample (Table 5 and Appendix II). Of these 100 samples, four samples, two from Maurice Creek, one from the Halfway River and one from the Moberly River, had much lower number of reads retained at this stage

(less than 3.2 million reads) and were eliminated from further analysis (Appendix II). Those same samples had PCR concentration during library construction much lower than other samples and close to the PCR concentration of the two negative controls (Appendix II) suggesting that library construction failed for these samples. The reads for the remaining 96 samples were mapped onto the latest release of the Rainbow Trout reference genome sequence and on average, over 5.7 million reads per sample (or 58.8% of high-quality reads; Table 5) mapped to the genome with mapping quality of 20 or higher (reads with a probability of an incorrect alignment lower than 0.01) and were used for SNP calling. The proportion of high mapping quality reads for Rainbow Trout is lower than for Bull Trout (62.1%; Geraldes and Taylor 2020) and Arctic Grayling (76.0%; Table 5), despite Rainbow Trout being the only one of the three species with an available genome reference sequence (for Bull Trout the Dolly Varden (*S. malma*) genome was used and for Arctic Grayling the European Grayling genome was used), and may be due to a combination of a) the genome reference sequence not being a complete representation of the species genome, b) some reads being from exogenous DNA and c) some reads being from repetitive or duplicated regions of the genome to which high quality mapping is challenging. On average, 2.12% of the genome of each sample had at least one read mapping to it with high quality (mapping quality of 20 or higher, i.e., with a probability of an incorrect alignment lower than 0.01; Table 5).

Overall, both for Arctic Grayling and Rainbow Trout, we generated millions of reads per sample, resulting in over 2% of the positions in the reference genome sequence being covered by at least one sequence read mapping with high confidence. The amount of sequence generated per Bull Trout sample (Geraldes and Taylor 2020)

was approximately half of that generated for Arctic Grayling and one third that generated for Rainbow Trout (Table 5), which resulted in about 1% of the positions in the reference genome sequence being covered by at least one sequence read mapping with high confidence (Table 5).

Population Structure and Differentiation

Results from a PCA (Figure 1A; Appendix I) on the Arctic Grayling genotype dataset (15,209 SNPs) revealed that the main axis of variation (explaining 9.04% of variation) separated samples from the Beatton River watershed, located DP, from samples from all other watersheds and that the second axis of variation (explaining 6.00% of variation) separated samples from the Moberly River watershed, located UP, from all others. Samples from the Halfway River, located UP, and samples from the Pine River, located DP, separated poorly along these two axes of the PCA but separated clearly along the third axis of the PCA (explaining 3.95% of variation). The Admixture analysis indicated that the model of population structure that best fit our data was for K=3, with K=4 having only a marginally higher CVE (Supporting Figure 1). A model with two population groups (K=2; Figure 1B) indicated that the samples from the Beatton River watershed form a distinct genetic group. A model with three population groups (K=3), the best fitting model, indicated that samples from the Moberly River watershed also formed a distinct genetic group and finally, a model with four population groups (K=4) identified each of the four watersheds in the LAA as distinct genetic groups (Figure 1B). Thus, the data clearly resolved three distinct genetic groups of Arctic Grayling in the LAA, and possibly four, but the two watersheds with the least clear genetic differentiation, the Halfway

River watershed and the Pine River watershed are located UP and DP. Furthermore, for K=4, while all samples of both the Beatton and the Moberly rivers watersheds, showed little (Moberly River) or no (Beatton River) evidence of admixture (each sample has less than 10% of the genome assigned to ancestry from a different watershed; Figure 1 and Appendix I), 26% of samples from the Halfway River watershed (five out of 19 samples) and 24% from the Pine River watershed (six out of 25 samples) had more than 10% of their genome assigned to ancestry from a different watershed; Figure 1 and Appendix I).

We next calculated weighted average pairwise F_{ST} (a measure of population differentiation) between all four watersheds in the LAA with the 24,735 SNP dataset (Table 6). This analysis confirmed that genetic differentiation between the Halfway River and Pine River population groups is lowest (F_{ST}= 0.077). Additionally, we calculated the same F_{ST} statistic between all samples collected UP (Halfway and Moberly rivers) and DP (Pine and Beatton rivers) and this estimate (F_{ST}= 0.066) is even lower than that between the Halfway River and the Pine River population groups. Despite this low level of genetic differentiation there was considerable variation among loci with the highest single SNP F_{ST} estimate being 0.705 and 21 loci having F_{ST} estimates above 0.5, suggesting that despite low overall genetic differentiation some loci do carry considerable information to differentiate fish that belong to genetic groups UP and DP of the Project.

Table 6. Weighted average pairwise F_{ST} estimates among Arctic Grayling from the four watersheds surveyed.

	Beatton	**Halfway**	**Moberly**
Beatton			
Halfway	0.429		
Moberly	0.292	0.156	
Pine	0.430	0.077	0.144

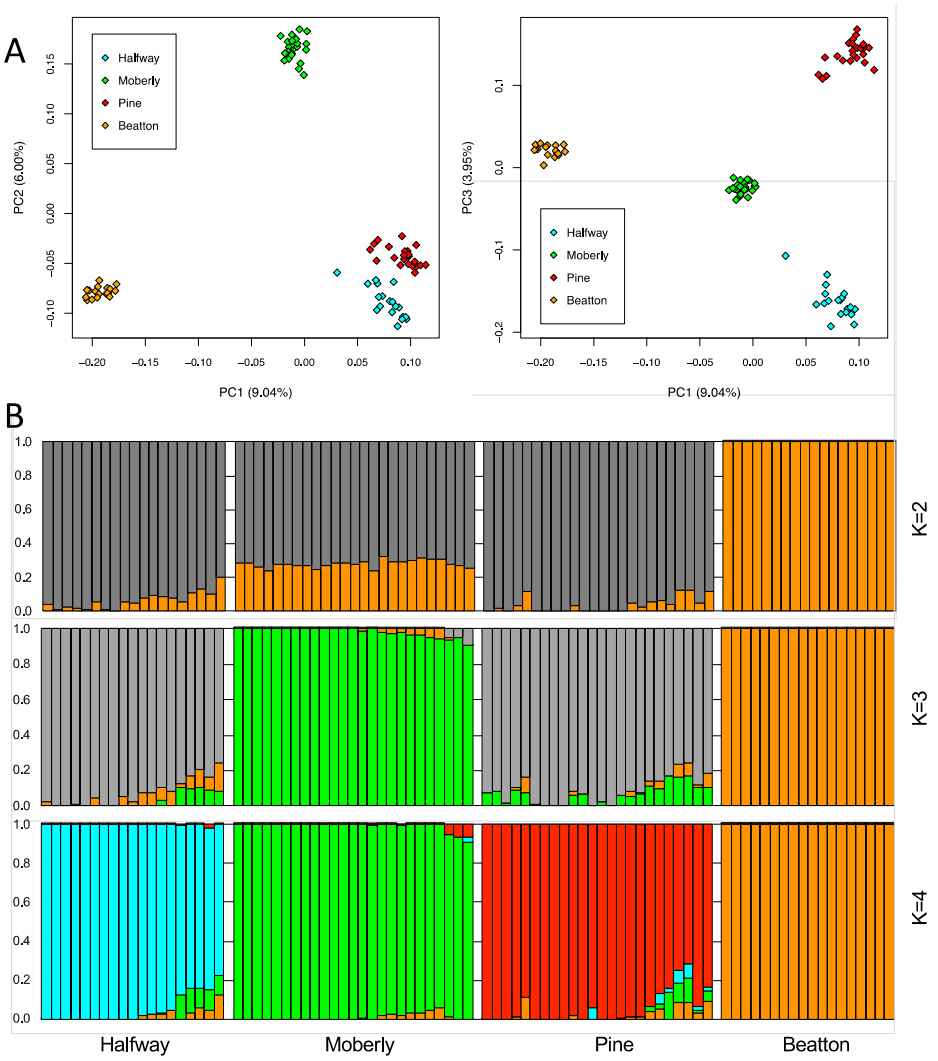


Figure 1 - Population structure of Arctic Grayling in the Peace Region. The Halfway River (n=19) and the Moberly River (n=52) are located upstream of the Project and the Pine River (n=24) and the Beatton River (n=18) are located downstream of the Project. Both the PCA (A) and Admixture (B) analysis were performed on 15,209 SNPs. (A) Each diamond represents a single Arctic Grayling sample. Halfway River samples are plotted in blue, Moberly River samples in green, Pine River samples are plotted in red, and Beatton River samples in orange. The first two principal components (PC) are plotted on the left panel and the first and third PCs are plotted on the right panel. The percentage of variation in the data explained by each PC is indicated in the axis name. (B) Admixture results are shown for K=2 (top), for K=3 (middle), and for K=4 (bottom). A K value of 3 is the number of populations that best fit the data, and K=4 had only a marginally higher CVE. Each column represents the genotype of an individual fish and the different colours represent the proportion of the genome of each fish that is assigned to each population.

SNP Genotyping

We genotyped 284 Bull Trout samples at the six ancestry informative SNPs for which TaqMan assays were developed, tested and genotyped in Geraldes and Taylor (2020).

Genotyping success was high (98.8%) with only 20 out of 1,704 genotyping reactions failing (one sample had three genotypes failing, two samples had two genotypes failing and 13 samples had one single genotype failing; Appendix III).

For Arctic Grayling we selected eight SNPs with high F_{ST} estimates between the tributaries located UP and DP for TaqMan design and testing. In addition, we selected one SNP with high F_{ST} estimates between the Halfway River population group and all others, and one with high F_{ST} between the Pine River population group and others (Table 4). Seven assays passed our initial test (Table 4) with 31 samples of known genotype, i.e., they provided both good amplification and were able to discriminate heterozygotes and homozygotes for each of the alternative alleles at each locus. These seven assays were then used to genotype 307 Arctic Grayling samples. Genotyping success was again high (97.7%) with only 50 out of 2,149 genotypes failing (one sample failed at six loci, one failed at three loci, seven failed at two loci and 27 samples failed only at one locus; Appendix I). Two samples with more than two missing genotypes were not considered further (neither fish was from the Peace River mainstem).

Population Assignment

For Bull Trout and Arctic Grayling, a sample was considered assigned to a group if the assignment score obtained in GeneClass2 (Piry et al., 2014) following the method of Rannala and Mountain (1997) was over 95% (i.e., $P<0.05$).

For Bull Trout, all samples used as reference samples (37 from the Pine River, 45 from the Halfway River and five from the Moberly River) were correctly assigned to

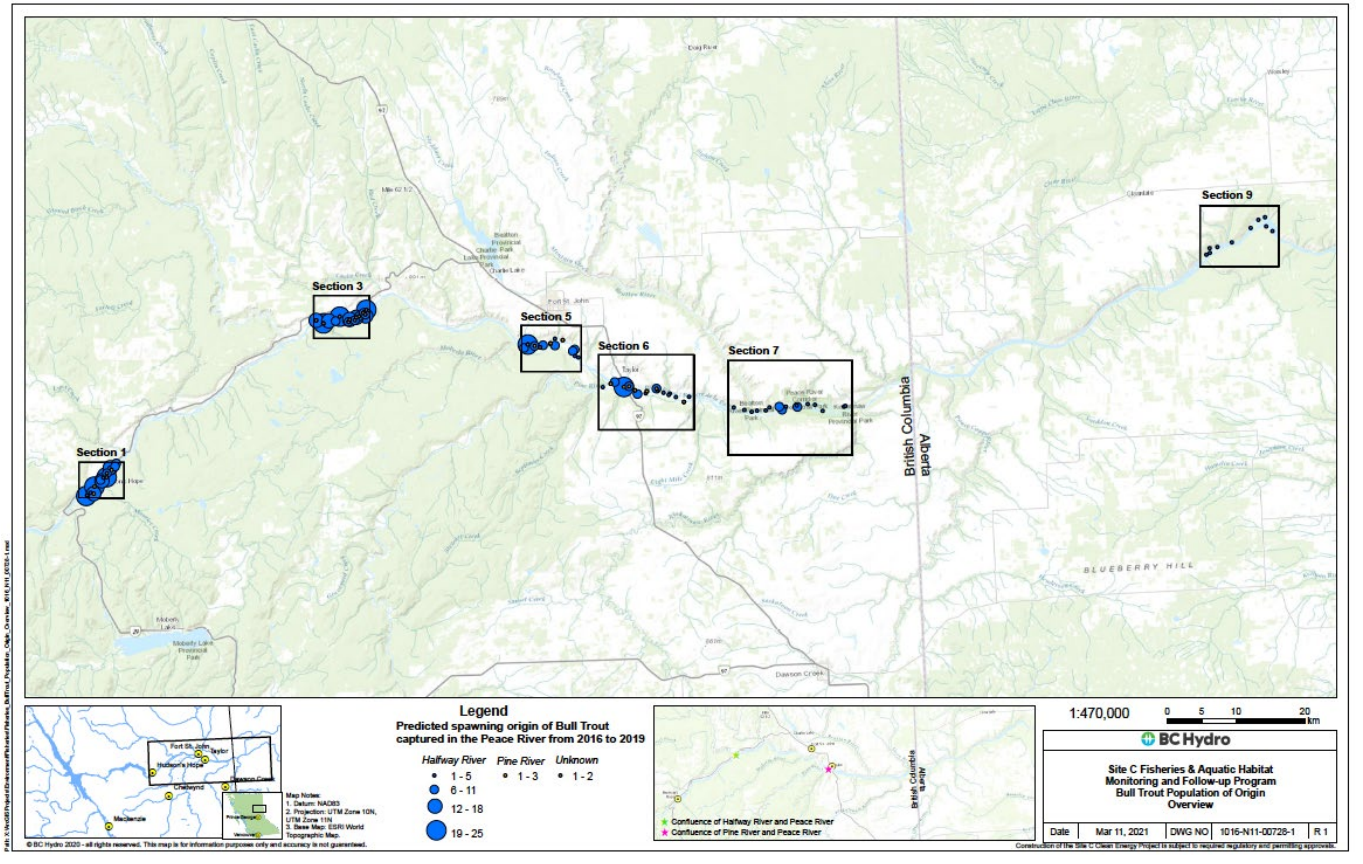
their known population groups (Appendix III). For the test sample set, all 25 samples from the Halfway River watershed and two samples from the Moberly River watershed were assigned to the UP group, and of the 21 samples from the Pine River watershed, 20 were assigned to the DP group, and one could not be assigned to either population group with more than 95% confidence (assignment score to the DP group was 64.5%). Because we have now expanded our reference samples (see Materials and Methods above for details), we performed the assignment tests again for the 473 Peace River samples collected between 2016 and 2018 that we had assigned and reported previously in Geraldes and Taylor (2020). Of the 473 samples from the Peace River assigned both in Geraldes and Taylor (2020) and here, 13 samples have discordant assignments (2.7%), all of which went from being unassigned to either population group to being assigned with high confidence. Importantly, in no case did a sample go from being assigned to one group to being assigned to a different group. Three samples unassigned in Geraldes and Taylor (2020) are now assigned to the DP group, seven samples unassigned in Geraldes and Taylor (2020) are now assigned to the UP group and three samples assigned in Geraldes and Taylor (2020) to DP are now unassigned. Across all years 93.8% of samples caught in the Peace River (623 out of 664 total samples) are assigned to the UP group, 3.8% (25 out of 664) are assigned to the DP group and 2.4% (16 out of 664 samples) cannot be assigned to either group with a score of 95% or higher (i.e., they are considered unassigned; Table 7 and Figure 2).

As part of Mon-2, Task 2a (Peace River Large Fish Indexing Survey), sampling occurred in six sections of the Peace River, with Sections 1 and 3 located UP and Sections 5, 6, 7 and 9 located DP (Figure 2, Table 7 and Appendix III). There was

considerable variation in the proportion of Bull Trout assigned to each group across sampling sections. Bull Trout assigned to the DP group were most common in Sections 5 and 6 located DP. Assignment proportions were markedly consistent among years with the exception of Sections 5 and 6 which showed differences of almost 15% between years although the clear majority of fish were always assigned to UP (Table 7).

Table 7. Number of Bull Trout samples caught in the Peace River (PR) assigned (% of total) to the UP (upstream of the Project) or DP (downstream of the Project) groups based on genotypes at six ancestry informative SNPs with more than 95% confidence.

Location	Year	Total	UP	DP	Unassigned
PR Section 1	2016-2018	118	114 (96.6%)	3 (2.5%)	1 (0.8%)
	2019	63	62 (98.4%)	0 (0.0%)	1 (1.6%)
	All years	181	176 (97.2%)	3 (1.7%)	2 (1.1%)
PR Section 3	2016-2018	161	149 (92.5%)	4 (2.5%)	8 (5.0%)
	2019	59	54 (91.5%)	2 (3.4%)	3 (5.1%)
	All years	220	203 (92.3%)	6 (2.7%)	11 (5.0%)
PR Section 5	2016-2018	81	76 (93.8%)	4 (4.9%)	1 (1.2%)
	2019	24	19 (79.2%)	4 (16.7%)	1 (4.2%)
	All years	105	95 (90.5%)	8 (7.6%)	2 (1.9%)
PR Section 6	2016-2018	68	59 (86.8%)	8 (11.9%)	1 (1.5%)
	2019	22	22 (100.0%)	0 (0.0%)	0 (0.0%)
	All years	90	81 (90.0%)	8 (8.9%)	1 (1.1%)
PR Section 7	2016-2018	27	27 (100.0%)	0 (0.0%)	0 (0.0%)
	2019	19	19 (100.0%)	0 (0.0%)	0 (0.0%)
	All years	46	46 (100.0%)	0 (0.0%)	0 (0.0%)
PR Section 9	2016-2018	18	18 (100.0%)	0 (0.0%)	0 (0.0%)
	2019	4	4 (100.0%)	0 (0.0%)	0 (0.0%)
	All years	22	22 (100.0%)	0 (0.0%)	0 (0.0%)
All PR Sections	2016-2018	473	443 (93.7%)	19 (4.0%)	11 (2.3%)
	2019	191	180 (94.2%)	6 (3.1%)	5 (2.6%)
	All years	664	623 (93.8%)	25 (3.8%)	16 (2.4%)


Section 1
Section 3
Section 5
Section 6
Section 7
Section 9
British Columbia
Alberta
BLUEBERRY HILL
Legend
Predicted spawning origin of Bull Trout captured in the Peace River from 2016 to 2019
Halfway River
Pine River
Unknown
1 - 5
6 - 11
12 - 18
19 - 25
1 - 3
1 - 2
Confluence of Halfway River and Peace River
Confluence of Pine River and Peace River
1:470,000
0 5 10 20 km
BC Hydro
Site C Fisheries & Aquatic Habitat Monitoring and Follow-up Program
Bull Trout Population of Origin Overview
Date
Mar 11, 2021
DWG NO
1016-N11-00728-1
R 1
Map Notes:
1. Datum: NAD83
2. Projection: UTM Zone 10N, UTM Zone 11N
3. Base Map: ESRI World Topographic Map.
© BC Hydro 2020 - all rights reserved. This map is for information purposes only and accuracy is not guaranteed.
Construction of the Site C Clean Energy Project is subject to required regulatory and permitting approvals.

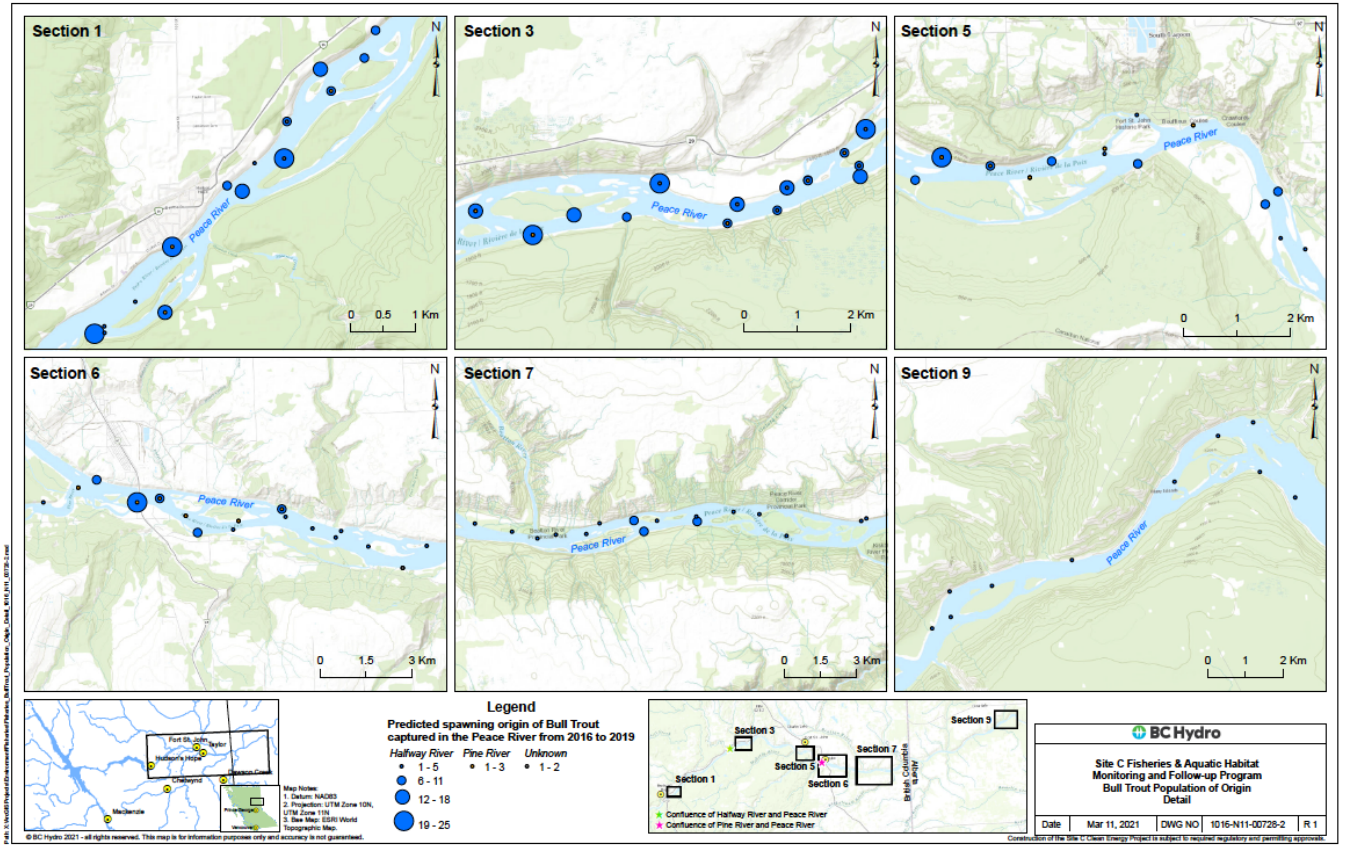


Figure 2. Predicted population of origin of subadult and adult Bull Trout captured in Sections 1, 3, 5, 6, 7, and 9 of the Peace River from 2016 to 2019 (blue: upstream of the Project-Halfway River; orange: downstream of the Project-Pine River). Circles are proportional to numbers of samples assigned. Samples that could not be assigned to either group with more than 95% confidence are shown in grey. The top panel shows the geographic location of each sampling section and the bottom panel shows each section of the Peace River in detail. Map courtesy of BC Hydro. Full details can be found in Appendix III.

For Arctic Grayling, all samples used as reference samples for watersheds located UP (19 from the Halfway River watershed and 25 from the Moberly River watershed) were correctly assigned to the UP group. For the DP group reference samples (24 from the Pine River watershed and 18 from the Beatton River watershed), all Beatton River watershed reference samples were assigned to the DP group, and 87.5% of the Pine River watershed samples were assigned to the DP (4.2% of the samples, i.e., one out of 24 samples, were assigned to UP group, and 8.3% of the samples, i.e., two out of 24 samples, could not be assigned with more than 95% confidence, but these two fish were assigned to DP with 91 and 92% confidence; Appendix I).

All 14 Halfway River watershed Arctic Grayling fish used as test samples, i.e., caught in the Halfway River watershed but for which no GBS data is available, were assigned to the UP group and all Beatton River test samples were assigned to the DP group. For the Moberly River test samples, 16 samples were assigned to the UP group and one could not be assigned to either the UP or DP groups (5.9%, i.e., one out of 17 samples, but it could be assigned to UP with 67% confidence). For the Pine River test samples, 21 were assigned to the DP group and four samples could not be assigned to either group (19.0%, i.e., four out of 21 samples were assigned to DP with 76-94% confidence; Appendix I).

Of the 146 Arctic Grayling samples caught in the Peace River mainstem, 84.2% (i.e., 123 samples) were assigned to the UP group, 7.5% (i.e., 11 samples) were assigned to the DP group and 8.2% (i.e., 12 samples) could not be assigned to either group (Table 8). All but one sample assigned to the DP group were collected in

sampling sections located DP, while samples assigned to the UP group were present in all sampling sections except Section 9 (the sampling section furthest downstream from the Project) where the only sample available could not be assigned to either the UP or DP groups. Arctic Grayling samples that could not be assigned to either group with more than 95% confidence were present in most sampling sections of the Peace River (except for Section 1, the most upstream of all sampling sections). Of the 12 samples that could not be assigned with more than 95% confidence, however, three fish could be assigned to DP with 75 to 93% confidence and three fish could be assigned to UP with 84 to 94% confidence.

Table 8. Number of Arctic Grayling samples caught in the Peace River (PR) assigned (% of total) to the UP (upstream of the Project) or DP (downstream of the Project) groups based on genotypes at seven ancestry informative SNPs with more than 95% confidence.

Location	Total	UP	DP	Unassigned
PR Section 1	3	2 (66.7%)	1 (33.3%)	0 (0.0%)
PR Section 3	74	72 (97.3%)	0 (0.0%)	2 (2.7%)
PR Section 5	23	21 (91.3%)	0 (0.0%)	2 (8.7%)
PR Section 6	32	21 (65.6%)	6 (18.8%)	5 (15.6%)
PR Section 7	13	7 (53.8%)	4 (30.8%)	2 (15.4%)
PR Section 9	1	0 (0.0%)	0 (0.0%)	1 (100.0%)
All PR Sections	146	123 (84.2%)	11 (7.5%)	12 (8.2%)

The results for Bull Trout and Arctic Grayling are concordant in assigning the vast majority of samples caught in the Peace River mainstem to spawning tributaries located UP (93.8% in Bull Trout and 84.2% in Arctic Grayling). The percentage of samples that could not be assigned to either group with more than 95% confidence was higher in Arctic Grayling (8.2% of all samples) than in Bull Trout (2.3% of all samples).

DISCUSSION

The results from this year align well with those from Geraldes and Taylor (2020) in showing that GBS is a cost-effective way of sampling large amounts of genetic variation from a moderately large number of samples and in identifying the main patterns of population differentiation in the LAA. Analysis of the GBS data generated for Arctic Grayling allowed for the selection of genetic variants highly differentiated between watersheds located UP and DP. The genotyping assays developed for seven of those genetic variants allowed for the effective genotyping of large numbers of Arctic Grayling samples collected in the Peace River and for their assignment to spawning populations located UP and DP. Analysis of the Rainbow Trout GBS data and design of genotyping assays is ongoing.

For the two species for which we have already identified SNPs and performed analysis of population structure in the LAA (Bull Trout and Arctic Grayling), our approach was very successful at identifying distinct genetic groups in the LAA. For Bull Trout (Geraldes and Taylor 2020), one genetic group corresponded to samples spawning in tributaries located UP (Halfway River) and the other corresponded to samples spawning in tributaries located DP (Pine River). These two groups are well differentiated with a weighted average F_{ST} of 0.11 between UP and DP genetic groups. This overall genomic differentiation coupled with a wide range of locus specific F_{ST} estimates and a long tail of loci showing high genetic differentiation (highest locus specific F_{ST} is 0.82 and 126 loci have F_{ST} above 0.5) resulted in the successful development of fast genotyping assays that allowed for the assignment of the vast majority (97.6%) of Bull Trout caught in the Peace River mainstem to the UP and DP

genetic groups with more than 95% confidence using just six loci (Table 7). In Arctic Grayling we were similarly successful in identifying distinct genetic groups in the LAA, but the patterns of genetic differentiation were less striking. Consequently, the assignment of fish caught in the Peace River mainstem to UP and DP genetic groups resulted in a higher number of fish whose ancestry was less certain (i.e., assignment confidence lower than 95%). The main reason for this difference in assignment success between species is that genetic differentiation between the Halfway (located UP) and Pine (located DP) rivers is fairly low (F_{ST}=0.07; Table 6) in Arctic Grayling and fewer loci showed high levels of genetic differentiation between UP and DP groups (the highest locus specific F_{ST} estimate is 0.71 and only 21 loci have F_{ST} above 0.5). These two genetic groups could, however, be distinguished both in the Admixture analysis (when four populations were assumed, even if a model with only three populations was a better for the data) and also along the third axis of a PCA (even if that axis explained a relatively small proportion of the genetic variability in the data). As a result, even with the use of genotype data from the seven genotyping assays, assignment of Arctic Grayling samples to UP and DP spawning populations was slightly less successful (91.8% of samples could be assigned with more than 95% confidence) than for Bull Trout samples (97.6% of samples could be assigned with more than 95% confidence). A larger proportion of Arctic Grayling samples remaining, for now, unassigned with more than 95% confidence, is consistent with the lower degree of population differentiation between the Halfway and Pine rivers' samples of Arctic Grayling compared to Bull Trout with microsatellites; F_{ST} between Halfway and Pine rivers' Bull Trout ranged from 0.068 to 0.081 with microsatellites, but only 0.020 to 0.067 for Arctic Grayling (Taylor et al.

2014). Bull Trout have been characterized as perhaps having amongst the most specific natal homing behaviour in salmonids which would promote divergence (McPhail 2007). We are currently developing additional genotyping assays to improve the ability to assign samples to UP and DP groups in Arctic Grayling and our preliminary results look promising.

Our results for Bull Trout and Arctic Grayling to date are highly consistent in showing that spawning populations located UP contribute most of the fish that are caught in the mainstem of the Peace River (96% of all assigned fish in Bull Trout; and 92% of all fish assigned in Arctic Grayling). In Bull Trout, where sample sizes are much larger, fish assigned to DP make up less than 10% of the fish caught in any section of the Peace River mainstem and make up more than 5% in only two sections of the Peace River mainstem, both located DP near the confluence with the Pine River, i.e., in Section 5, where 7.6% of fish are assigned to DP, and Section 6, where 8.9% of fish are assigned to DP. In Arctic Grayling, not only are a smaller proportion of fish confidently assigned to genetic groups UP and DP for now, but sample sizes are also lower. Despite these lower assignment rates and sample sizes, at this point, when considering only sections of the Peace River mainstem where more than three fish have been collected and analyzed, the same pattern emerges where fish assigned to DP are most common in sampling sections located DP than UP. In fact, from 77 fish collected in Sections 1 and 3 located UP, only one fish was assigned to the DP genetic group. This suggests that in Bull Trout and Arctic Grayling there is relatively low movement of fish that spawn DP to areas of the Peace River upstream of the Project (Sections 1 and 3), but conversely, there seems to be frequent movement of fish that spawn UP to sections

of the Peace River DP of the Project (Sections 5 to 9). Taylor et al (2014) inferred relatively low levels of movement both for Arctic Grayling and Bull Trout between DP and UP tributaries as well.

A small percentage of Bull Trout (average = 2.4%) and Arctic Grayling (average = 8.2%) were classified as 'unassigned' at a 95% confidence threshold. It is important to note, however, that at lower levels of confidence most fish can be assigned as either UP or DP. For instance, in Arctic Grayling, there were only four fish (2.7%) whose assignment could really be characterized as ambiguous because their assignments to DP or UP were within 5% of 50:50 and only one such case occurred in Bull Trout (0.5%). The very few cases of ambiguity in assignment can be due to missing data (not relevant in any individual case here) or due to mixed ancestry owing to occasional interbreeding between fish from different tributaries. Consequently, assessment of the decision to pass a fish or not in terms of their ancestry will be possible for the vast majority of individuals.

We are currently in the process of analyzing patterns of population structure for Rainbow Trout in the LAA before designing and testing genotyping assays for that species. In the coming months we will have completed DNA extraction, SNP genotyping and population assignment for all fish of the three species collected in the Peace River mainstem from 2016 to 2020.

REFERENCES

Alcaide, M., Scordato, E. S. C., Price, T. D., & Irwin, D. E. (2014). Genomic divergence in a ring species complex. Nature, 511, 83–85. https://doi. org/10.1038/nature13285

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19(9), 1655–1664. http://doi.org/10.1101/gr.094052.109

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets, GigaScience, Volume 4, Issue 1, s13742–015–0047–8, https://doi.org/10.1186/s13742-015-0047-8

J. Catchen, P. Hohenlohe, S. Bassham, A. Amores, and W. Cresko. Stacks: an analysis tool set for population genomics. Molecular Ecology. 2013.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Durbin, R. (2011). The variant call format and VCFtools. Bioinformatics (Oxford, England), 27, 2156–2158. https://doi. org/10.1093/bioinformatics/btr330

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for

high diversity species. PLoS ONE, 6, e19379. https://doi.org/10.1371/journal.pone.0019379.g006

Freshwater Fisheries Society of BC, 2021. https://www.gofishbc.com/

Geraldes, A. and Taylor, E. B. (2020). Fish Genetics Study: 2019 Status Report for Bull Trout, Arctic Grayling and Rainbow Trout. BC Hydro - Site C Clean Energy Project. 44 pages + 2 App.

Geraldes, A., Askelson, K. K., Nikelski, E., Doyle, F. I., Harrower, W. L., Winker, K., & Irwin, D. E. (2019). Population genomic analyses re-veal a highly differentiated genetic cluster of northern goshawks (Accipiter gentilis laingi) in Haida Gwaii. Evolutionary Applications, 12, 757–772. https://doi.org/10.1111/eva.12754

Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., & Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Molecular Ecology, 25, 4488–4507. https://doi.org/10.1111/mec.13792

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20, 1297–1303. https://doi. org/10.1101/gr.107524.110

McPhail, J.D. (2007). The freshwater fishes of British Columbia. University of Alberta Press, Edmonton.

Owens, G. L., Baute, G. J., & Rieseberg, L. H. (2016). Revisiting a classic case of introgression: Hybridization and gene flow in Californian sunflowers. Molecular Ecology, 25, 2630–2643. https://doi.org/10.1111/ mec.13569

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. Journal of Heredity, 95, 536-539. https://doi.org/10.1093/jhered/esh074

Rannala B and Mountain JL, 1997. Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA 94:9197–9201.

Sävilammi, T., Primmer, C. R., Varadharajan, S., Guyomard, R., Guiguen, Y., Sandve, S. R., et al. (2019). The Chromosome-Level Genome Assembly of European Grayling Reveals Aspects of a Unique Genome Evolution Process Within Salmonids. G3 (Bethesda, Md.), 9(5), 1283–1294.

Taylor, E. B., Yau, M. M., & Mattock & A. B. (2014). Population structure in three species of co-distributed salmonid fishes in the peace river and tributaries near a major

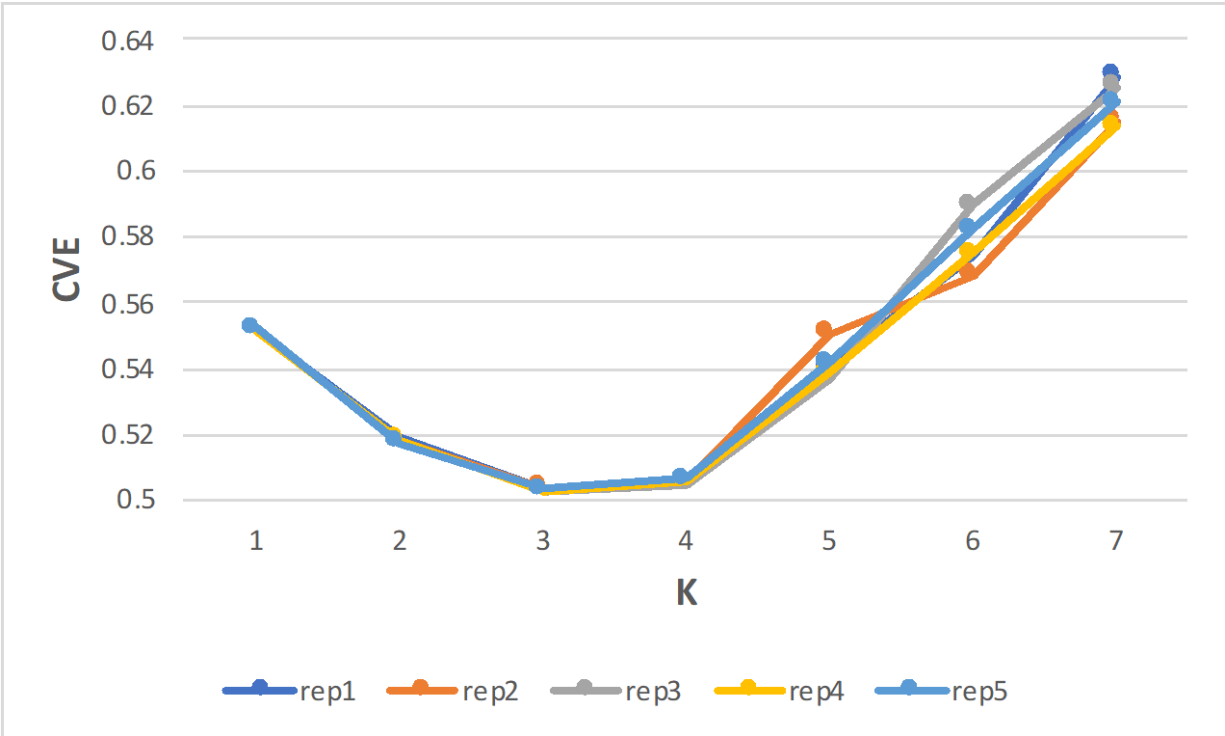
proposed hydroelectric development in northeastern British Columbia, Canada. River Research and Applications, 30, 1120-1133.

Toews, D. P. L., Brelsford, A., Grossen, C., Milá, B., & Irwin, D. E. (2016). Genomic variation across the Yellow-rumped Warbler species com- plex. Auk, 133, 698–717. https://doi.org/10.1642/AUK-16-61.1

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358–1370. https:// doi.org/10.1111/j.1558-5646.1984.tb05657.x

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., Weir, B. S. (2012). A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. Bioinformatics. DOI: 10.1093/bioinformatics/bts606.

SUPPORTING FIGURES



Supporting Figure 1- Cross Validation Error (CVE) for the Arctic Grayling Admixture analysis. Results are shown for the five replicates of the analysis with the 15,209 SNPs dataset with K varying from 1 to 7. A value of K=3, i.e., three populations, minimized the CVE and is therefore the best fit to our data. A value of K=4, i.e., four populations, showed a CVE only marginally higher.