# Site C Clean Energy Project

## Fisheries and Aquatic Habitat Monitoring and Follow-up Program

***Fish Genetics Study***
***2019 Status Report for Bull Trout, Arctic Grayling and Rainbow Trout***

**Construction Year 5 (2019)**

**Armando Geraldes, PhD**
**University of British Columbia**

**Eric B. Taylor, PhD**
**University of British Columbia**

**June 15, 2020**

## EXECUTIVE SUMMARY

BC Hydro is currently constructing the Site C Clean Energy Project (the Project), which will be the third hydroelectric dam on the Peace River near the town of Fort St. John in northeastern British Columbia. 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) and Mon-2, Task 2a (Peace River Large Fish Indexing Survey) of the FAHMFP have examined 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 during the first implementation year of the Site C Fish Genetics Study from September 1, 2018 to December 31, 2019.

Work during the first year focused on Bull Trout. Extraction and quality control of DNA was completed for 1,572 Bull Trout samples from the Peace River and its tributaries collected from 2016 to 2018. A subset of samples from select tributaries of the Peace River (Halfway, Pine and Moberly rivers) was used to develop genetic markers and determine levels and patterns of population structure among samples from the different tributaries. Analysis of these data revealed clear genetic differentiation between Bull Trout caught in the Pine River, downstream of the Project, and samples caught in the Moberly and Halfway rivers upstream of the Project. Henceforth we refer to these as the Halfway

and Pine genetic groups because a) Bull Trout are not known to spawn in the Moberly River, and b) no genetic differentiation was detected between Bull Trout from the Halfway and Moberly rivers. From a pool of 7,564 quality filtered genetic markers, 11 markers that showed maximal differentiation among samples from upstream (Halfway) and downstream (Pine) of the Project were selected and used to develop genotyping assays to determine the tributary of origin of individual Bull Trout. Six of those 11 assays were selected to genotype 517 Bull Trout samples, including all available samples from the Peace River (n=473), as well as a few samples from the Pine (n=28), Moberly (n=2) and Halfway (n=14) rivers. Analysis of the genotype data allowed most samples (499 out of 517) to be assigned to each of the two genetic groups with a high degree of confidence. Eighteen samples from the Peace River (4%) could not be assigned unambiguously to either group. With this method, we estimated that 92% of Bull Trout sampled in the Peace River likely originated from the Halfway River, 4% likely originated from the Pine River, and 4% were undetermined. Bull Trout assigned to the Pine River were most common in Section 6 of the Peace River, near the confluence of the Peace and Pine rivers, where they made up 7.4% of all Bull Trout captured. Bull Trout assigned to the Pine River were absent from Sections 7 and 9, and made up less than 5% of fish sampled in Sections 1, 3 and 5. Bull Trout assigned to the Halfway River were widely distributed throughout the Peace River (more than 90% of fish sampled in any section of the Peace River likely originated from the Halfway River).

## ACKNOWLEDGMENTS

This work was supported by a contract from BC Hydro. We thank Nich Burnett and Brent Mossop for guidance and helpful discussions. Dustin Ford of Golder Associates is thanked for assistance with sample collection, provision and orientation. Jessica Shen 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 |
| CV | Coefficient of variation |
| DNA | Deoxyribonucleic Acid |
| 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 |
| LG | Linkage group |
| MYA | Million years ago |
| NCBI | National Center or Biotechnology Information |
| PCA | Principal components analysis |
| PCR | Polymerase chain reaction |
| Q | Proportion of ancestry in a given genetic cluster in the Admixture analysis |
| SNP | Single nucleotide polymorphism |
| Rn | Reporter dye signal of each allele (FAM or VIC) normalized by the fluorescence signal of the ROX dye in the TaqMan assays |
| SRA | Short read archive |

## TABLE OF CONTENTS


Executive Summary ..... 2
Acknowledgments ..... 4
List of Acronyms and Abbreviations ..... 5
Table of Contents ..... 6
List of Tables ..... 7
List of Figures ..... 8
List of Appendices ..... 9
Introduction ..... 10
  Purpose and Objectives ..... 13
Materials and Methods ..... 13
  Samples ..... 13
  DNA Extractions and Quality Control ..... 14
  Bull Trout Sequencing ..... 16
  SNP Discovery Bioinformatics ..... 19
  Population Genetic Analyses ..... 21
  SNP Genotyping Assays ..... 21
  Assignment Tests ..... 23
Results ..... 24
  SNP Discovery ..... 24
  Population Structure and Differentiation ..... 26
  SNP Genotyping ..... 29
  Population Assignment ..... 30
Discussion ..... 34
References ..... 37
Supporting Figures ..... 42

## LIST OF TABLES

**Table 1.** Samples available for genetics work from 2016 to 2018.

**Table 2.** Samples used for Bull Trout SNP discovery.

**Table 3.** TaqMan assays ordered and tested to genotype Bull Trout samples.

**Table 4.** Average, maximum and minimum number of reads generated, mapped and used for SNP calling per Bull Trout sample used in the GBS library sequencing (across the 92 samples in total).

**Table 5.** Weighted average pairwise $F_{ST}$ estimates among Bull Trout from the three watersheds surveyed.

**Table 6.** Number of Bull Trout assayed and assigned (% of total) to the Halfway or Pine groups after being genotyped at six ancestry informative SNPs with TaqMan assays.

## LIST OF FIGURES

**Figure 1.** Population structure of Bull Trout in the Peace Region.

**Figure 2.** Levels of population differentiation between Bull Trout samples from the Halfway (upstream of the Project) and Pine (downstream of the Project) groups for each SNP in the dataset as estimated by $F_{ST}$.

**Figure 3.** Predicted population of origin (blue: Halfway, Pine: orange) of subadult and adult Bull Trout captured in Sections 1, 3, 5, 6, 7, and 9 of the Peace River from 2016 to 2018.

**Supporting Figure 1.** Coefficient of variation for the Bull Trout Admixture analysis.

**Supporting Figure 2.** Additional Bull Trout population structure in the Pine River watershed.

**Supporting Figure 3.** PCA analyses of genetic variability among Bull Trout from the Halfway and Moberly rivers.

## LIST OF APPENDICES

**Appendix I.** Bull Trout project sample information, sequencing results (reads generated, mapped to the reference genome, and used for SNP calling), genome coverage (the percentage of the genome with at least one and at least five high quality mapping reads), frequency of missing genotypes (out of the unfiltered 663,244 variants identified with GenotypeGVCFs in GATK3), 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 5). File Appendix_I_BT_GBS_Results.xlsx

**Appendix II.** 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 one of two reference populations as well as the probability of exclusion for each population (i.e., the probability of observing the six locus genotype within 100,000 simulated genotypes from each putative source population). File AppendixII_TaqMan_Results.xlsx

## INTRODUCTION

BC Hydro is currently constructing the Site C Clean Energy Project (the Project), which will be the third hydroelectric dam on the Peace River near the town of Fort St. John in northeastern British Columbia (hereafter referred to as the Local Assessment Area, LAA). The BC portion of the Peace River has 41 native freshwater fish species and among them are important recreational sport fishes such as Bull Trout (*Salvelinus confluentus*), Arctic Grayling (*Thymallus arcticus*), and Rainbow Trout (*Oncorhynchus mykiss*, McPhail 2007) which are common in the LAA. Furthermore, there are numerous forage species that support, to varying degrees, these sport fishes (e.g., Mountain Whitefish, *Prosopium williamsoni*). All of these species, as well as many others, regularly migrate through areas that are upstream and downstream of the Project and occur in major tributaries upstream and downstream of the Project (e.g., Farrell Creek, Halfway, Moberly, Pine, and Beatton rivers). 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 developed the Site C Fisheries and Aquatic Habitat Monitoring and Follow-up Program (FAHMFP). As part of the FAHMFP, BC Hydro will use various 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. For instance, information from otolith and fin ray microchemistry, radio telemetry, fish distribution, and genetics will be used to answer management questions and test management hypotheses posed in the FAHMFP.

There is a rich, and at least 50-year-old history of the development and application of genetic tools in many different aspects of fisheries management and conservation (see reviews by Carvalho and Pitcher 2012; Valenzuela-Quiñonez 2016; Allendorf 2017). In the case of the Project, Taylor et al. (2014) studied the degree of population distinction among samples of Bull Trout, Arctic Grayling and Mountain Whitefish in the LAA using microsatellite DNA variation. This study clearly indicated substantial differences between fish sampled from tributaries upstream and downstream of the Project in Bull Trout and Arctic Grayling, but rather less distinction among samples of Mountain Whitefish. The microsatellite DNA-based work of Taylor et al. (2014) also inferred movements of fish among tributaries (via the mainstem Peace River) that were roughly concordant with results observed with radio telemetry. Consequently, this work clearly indicated the potential for genetic analyses to address various fisheries management questions in the LAA.

Since the work of Taylor et al. (2014), however, developments in the kinds of molecular analyses that can provide high resolution, diverse and efficient genetic assays at reasonable cost have progressed rapidly. The most salient feature of this progress has been the development of techniques, both “at the bench” and in terms of bioinformatics and associated population genetic analyses, that sequence the entire genome of a study species, or at least a modest percentage of the entire genome (see various chapters in Rajora 2019). In salmonid fishes, with a haploid genome size of about 3 billion base pairs (bp), even assaying a modest percentage of the genome results in many (i.e., millions) genetic characters to examine. By contrast, “traditional” techniques (and by “traditional” is meant some techniques that were, and still are in some cases, widely applied only a decade or so ago) typically assay only a few dozen

or so genetic traits (or loci) in a typical fisheries application. For instance, Taylor et al. (2014) examined only 9-10 genetic loci in each of the three study species. Consequently, these recent developments have pushed fisheries “genetics” solidly into the realm of fisheries “genomics” where literally millions of genetic sites are examined across the genome and thus provide un-precedented power to address applications in fisheries including questions that have been difficult to address in the past (e.g., concurrently studying adaptive and neutral traits and the molecular basis of adaptation, e.g., Hand et al. 2016). Consequently, genomic analysis can inform the biology and management of Bull Trout, Arctic Grayling and Rainbow Trout, which are of key interest to Indigenous groups and regulatory agencies. Further, the Province of BC may use the genomic data generated in its management plans and BC Hydro can use the findings of the current study to help inform key uncertainties in the management of fishes in the LAA. For instance, there is no provision for volitional movement of migratory fishes upstream of the Project, and BC Hydro will implement a trap and haul program to maintain connectivity in the Peace River. Here, fish that pass a weir-orifice fishway at times of the year when upstream migration could be anticipated (e.g., spawning migrations from mainstem habitats to tributaries) are captured, tagged, transported and released upstream of the Project. Some individuals that appear at the fishway, however, spawn in tributaries downstream of the Project. In such instances, moving fish upstream may reduce their spawning success. Consequently, there is an important need for a way to differentiate fish spawning in tributaries downstream of the Project from those spawning in upstream tributaries. Analyses that exploit the occurrence of genetic differences among populations spawning in different tributaries, for which there are

myriad examples across the globe and in diverse taxa, provide a powerful toolbox to fill such a need.

### 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. The initial phase of the study focused on developing assays for Bull Trout to reliably distinguish fish sampled in the Peace River that originated from spawning tributaries either upstream or downstream of the Project. The initial phase of the study also began the process of similar assay design for Arctic Grayling. Subsequent phases of the study will complete assay design for Arctic Grayling and Rainbow Trout. The report focuses on the completed analysis for Bull Trout and the progress on Arctic Grayling to the end of 2019.

## MATERIALS AND METHODS

### Samples

Mon-1b, Task 2c (Site C Reservoir Tributaries Fish Population Indexing Survey) and Mon-2, Task 2a (Peace River Large Fish Indexing Survey) of the FAHMFP collected 1,572 Bull Trout, 95 Arctic Grayling and 296 Rainbow Trout genetic samples from 2016 to 2018 and stored them in individual vials with 95% ethanol (Table 1). Golder Associates Ltd. shipped the samples to UBC for analysis. Samples were collected from both the Peace River (including Farrell Creek) and its tributaries: the Halfway, Moberly

and Beatton rivers. Additional historical DNA samples from the Pine River, a tributary of the Peace River downstream of the Project, and from the Halfway and Moberly rivers from our laboratory archive were also used as described below (Table 1).

## DNA Extractions and Quality Control

Tissue samples were digested and total genomic DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen Inc., Valencia, CA, USA) following the manufacturer's instructions. The DNA was eluted in 200 µL of AE buffer provided with the kit. The DNA concentration of each extract was measured with the Qubit dsDNA broad range kit (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions in a Qubit4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). In cases where the DNA amount was below the detection limit of the assay, quantification was repeated with the Qubit dsDNA high sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA). The purity level of all DNA extracts was determined by spectrophotometry using a Nanodrop 8000 (Thermo Fisher Scientific, Waltham, MA, USA). Finally, DNA integrity – i.e., whether high molecular weight was extracted – was checked by running 2 µL of each DNA extraction in 2% agarose with 1% SYBR safe DNA gel stain (Thermo Fisher Scientific, Waltham, MA, USA).

**Table 1.** Samples available for genetics work from 2016 to 2018. Available refers to the overall total, DNA refers to DNA samples extracted, GBS refers to the number of DNAs used in SNP discovery, TaqMan refers to the number of DNAs used in the TaqMan assays.

| Species | Watershed | River | Available | DNA | GBS[1] | TaqMan |
|---|---|---|---|---|---|---|
| Bull Trout | Halfway River | Chowade River | 473 | 473 | 8 | 2 |
| Bull Trout | Halfway River | Colt Creek | 9 | 9 | 8 | 3 |
| Bull Trout | Halfway River | Cypress Creek | 392 | 392 | 8 | 4 |
| Bull Trout | Halfway River | Fiddes Creek | 175 | 175 | 8 | 0 |
| Bull Trout | Halfway River | Halfway River | 7 | 7 | 5 | 3 |
| Bull Trout | Halfway River | Turnoff Creek | 40 | 40 | 4 | 1 |
| Bull Trout | Moberly River | Moberly River | 3 | 3 | 3 | 2 |
| Bull Trout | Peace River | Peace River - Section 1 | 118 | 118 | 0 | 118 |
| Bull Trout | Peace River | Peace River - Section 3 | 161 | 161 | 0 | 161 |
| Bull Trout | Peace River | Peace River - Section 5 | 81 | 81 | 0 | 81 |
| Bull Trout | Peace River | Peace River - Section 6 | 68 | 68 | 0 | 68 |
| Bull Trout | Peace River | Peace River - Section 7 | 27 | 27 | 0 | 27 |
| Bull Trout | Peace River | Peace River - Section 9 | 18 | 18 | 0 | 18 |
| | | | | | | |
| Arctic Grayling | Beatton River | Beatton River | 3 | 3 | 3 | 0 |
| Arctic Grayling | Beatton River | Bratland Creek | 15 | 15 | 5 | 0 |
| Arctic Grayling | Beatton River | LaPrise Creek | 13 | 13 | 5 | 0 |
| Arctic Grayling | Beatton River | Unnamed Creek 1 | 1 | 1 | 1 | 0 |
| Arctic Grayling | Moberly River | Moberly River | 8 | 8 | 8 | 0 |
| Arctic Grayling | Peace River | Peace River - Section 1 | 3 | 3 | 0 | 0 |
| Arctic Grayling | Peace River | Peace River - Section 3 | 36 | 36 | 0 | 0 |
| Arctic Grayling | Peace River | Peace River - Section 5 | 10 | 10 | 0 | 0 |
| Arctic Grayling | Peace River | Peace River - Section 6 | 5 | 5 | 0 | 0 |
| Arctic Grayling | Peace River | Peace River - Section 7 | 1 | 1 | 0 | 0 |
| | | | | | | |
| Rainbow Trout | Halfway River | Chowade River | 11 | 0 | 0 | 0 |
| Rainbow Trout | Halfway River | Colt Creek | 32 | 0 | 0 | 0 |
| Rainbow Trout | Halfway River | Cypress Creek | 21 | 0 | 0 | 0 |
| Rainbow Trout | Halfway River | Kobes Creek | 42 | 0 | 0 | 0 |
| Rainbow Trout | Peace River | Farrell Creek | 45 | 0 | 0 | 0 |
| Rainbow Trout | Peace River | Peace River - Section 1 | 62 | 0 | 0 | 0 |
| Rainbow Trout | Peace River | Peace River - Section 3 | 66 | 0 | 0 | 0 |
| Rainbow Trout | Peace River | Peace River - Section 5 | 11 | 0 | 0 | 0 |
| Rainbow Trout | Peace River | Peace River - Section 6 | 2 | 0 | 0 | 0 |
| Rainbow Trout | Peace River | Peace River - Section 7 | 4 | 0 | 0 | 0 |

[1] Additional samples from previous study years were included in order to have representative samples from potential spawning areas upstream and downstream of the Project (Table 2).

## Bull Trout Sequencing

Samples for sequencing and genetic variant discovery (single nucleotide polymorphisms, SNPs) were selected from the two main spawning areas in the LAA, Halfway (upstream of the Project) and Pine rivers (downstream of the Project, Table 2). Because no samples were collected in the Pine River from 2016 to 2018, historical samples from 2011 were used (Taylor et al. 2014). A few 2011 Halfway 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. To increase the chances of capturing as much genetic variability as possible, we selected samples to maximize the spatial coverage within each watershed. We did not include fish smaller than 50 mm (likely newly-emerged fry) from a single sampling location to minimize the chance of sequencing close relatives. 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 (only 15 fish were larger than 300 mm). We also included five samples from the Moberly River (Table 2). Bull Trout are not known to spawn in the Moberly River, and Bull Trout found there are likely using it as a foraging site. Additionally, one Arctic Char (*Salvelinus alpinus*) and one Dolly Varden (*Salvelinus malma*) were included (Table 2).

In order to be able to cost-effectively generate sequence data from a representative fraction of the genome of these 94 samples, we used a reduced representation genome sequencing approach known as genotyping-by-sequencing (GBS; Elshire et al. 2011). We used a modified GBS protocol described in detail elsewhere (Alcaide et al. 2014; Towes et al. 2016; Geraldes et al. 2019) to generate a pooled library of digested and individually barcoded DNA.

**Table 2.** Samples used for Bull Trout SNP discovery.

| Species | Watershed | Tributary | Year | N |
|---|---|---|---|---|
| Bull Trout | Halfway River | Chowade River | 2017 | 4 |
| Bull Trout | Halfway River | Chowade River | 2018 | 4 |
| Bull Trout | Halfway River | Colt Creek | 2017 | 3 |
| Bull Trout | Halfway River | Colt Creek | 2018 | 5 |
| Bull Trout | Halfway River | Cypress Creek | 2017 | 4 |
| Bull Trout | Halfway River | Cypress Creek | 2018 | 4 |
| Bull Trout | Halfway River | Fiddes Creek | 2017 | 4 |
| Bull Trout | Halfway River | Fiddes Creek | 2018 | 4 |
| Bull Trout | Halfway River | Halfway River | 2011 | 5 |
| Bull Trout | Halfway River | Halfway River | 2016 | 5 |
| Bull Trout | Halfway River | Turnoff Creek | 2018 | 4 |
| Bull Trout | Moberly River | Moberly River | 2006 | 1 |
| Bull Trout | Moberly River | Moberly River | 2016 | 2 |
| Bull Trout | Moberly River | Moberly River | 2018 | 2 |
| Bull Trout | Pine River | Blind Creek | 2011 | 3 |
| Bull Trout | Pine River | Burnt River | 2011 | 7 |
| Bull Trout | Pine River | Callazon Creek | 2011 | 7 |
| Bull Trout | Pine River | Fellers Creek | 2011 | 7 |
| Bull Trout | Pine River | North Burnt River | 2011 | 2 |
| Bull Trout | Pine River | Pine River | 2011 | 4 |
| Bull Trout | Pine River | Willow Creek | 2011 | 4 |
| Bull Trout | Pine River | Wolverine River | 2011 | 7 |
| Dolly Varden[1] | | | | 1 |
| Arctic Char[1] | | | | 1 |

[1] Arctic Char and Dolly Varden are sister species closely related to Bull Trout and the reference genome sequence of *Salvelinus* used to map the sequencing reads was from an Arctic Char

Specifically, for each sample, we digested 100 ng of genomic DNA with the enzyme *Pst*I (New England Biolabs, Ipswhich, MA, USA) at 37°C for 3 hours in the presence of barcoded and common adaptors. Next, to attach the unique barcode (4 to 8 bp long) and common adaptors required for PCR amplification to the digested DNA fragments, all three components were ligated with T4 DNA ligase (New England Biolabs, Ipswhich, MA, USA) following the manufacturer's instructions for 1 hour at 22°C

followed by enzyme inactivation at 65°C for 10 minutes. The resulting reactions were then cleaned with AMPure XP beads (Beckman-Coulter, Brea, CA, USA) with a 30:20 beads:ligated DNA solution to remove DNA fragments smaller than 100 bp (including non-ligated barcodes and common adaptors) as well as other reagents that might inhibit the subsequent PCR reaction. Purified DNA was eluted in 40 µL of AE buffer (Qiagen Inc., Valencia, CA, USA) of which 5 µL were used for PCR amplification with Phusion High-Fidelity DNA polymerase (New England Biolabs, Ipswhich, MA, USA) at a final reaction volume of 20 µL. Primer, barcode and common adaptor sequences were taken from Alcaide et al. (2014) and the PCR mix followed the manufacturer's instructions. The PCR program consisted of 18 amplification cycles of 98°C for 10 s, 65°C for 30 s and 72°C for 30 s, preceded by an initial DNA denaturation for 30 s at 98°C and followed by a final DNA extension for 5 m at 72°C. We ran 2 µL of each PCR amplified DNA to check for a DNA smear indicating that there was no preferential amplification of some fragment sizes but rather that a large range of product sizes were amplified.

The amplified DNA was quantified with Qubit dsDNA high sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA) and 150 ng of each sample's barcoded DNA was added to a common pool. Amplified DNA concentrations varied between 23 and 50 ng/µL except for two negative controls: one had water instead of the barcoded adaptor (PCR concentration: 5.0 ng/µL) and the other water instead of DNA (PCR concentration: 3.9 ng/µL). The 380 µL DNA pool was then concentrated in an Eppendorf Vacufuge (Hamburg, Germany) at 30°C under vacuum for 1 hour to a final volume of 80 µL. We ran the concentrated pooled library over six lanes of a 2% agarose gel stained with 1% SYBR safe DNA gel stain (Thermo Fisher Scientific, Waltham, MA, USA) at 75V for 2 hours and then excised the 600-700 bp gel section from each lane. The DNA was

extracted and purified from the agarose gel with the QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA, USA) and the size distribution of the fragments in the library was checked in an Agilent High Sensitivity DNA chip ran on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The average fragment size of the resulting DNA library was 650 bp (coefficient of variation = 15.3%). The pooled library was sequenced in an Illumina HiSeq 4000 with 150 bp paired end reads at the McGill University and Génome Québec Innovation Centre.

SNP Discovery Bioinformatics

Analysis of the sequence data followed a bioinformatics pipeline available at https://doi.org/10.5061/dryad.t951d (Irwin et al. 2016) for GBS read processing and mapping. We used a perl script from Baute et al. (2016) to: demultiplex the raw sequencing reads according to the barcode sequence of each sample, remove the barcode sequence, and remove sequences shorter than 30 bp. The resulting reads for each sample were then trimmed with Trimmomatic-0.39 (Bolger et al. 2014) with options TRAILING:3, SLIDINGWINDOW:4:10, MINLEN:30. We used BWA-MEM (Li & Durbin, 2009) with default settings to align the trimmed reads from each sample to the Arctic Char (*Salvelinus alpinus*) reference genome sequence (assembly ASM291031v2; Christensen et al. 2018). This reference genome contains 39 scaffolds assigned to the different linkage groups of Arctic Char. For the most part, each linkage group/scaffold corresponds to a chromosome, except for the putative sex chromosome (chromosome 4) which is broken up into three different linkage groups and chromosome 6 which is broken up into two scaffolds. In addition to these scaffolds, the reference genome sequence contains a scaffold that corresponds to the mitochondrial sequence and

15,215 small scaffolds that have not been mapped to any chromosomes or linkage groups (i.e. unplaced scaffolds). We then used Picard (http://broadinstitute.github.io/picard/) and SAMtools (Li et al. 2009) to generate a BAM file containing the alignment information for each sample. We used HaplotypeCaller from GATK3 (Mckenna et al. 2010) to call genotypes one sample at a time. We then generated a VCF file with the genotypes for all 94 samples with the function GenotypeGVCFs in GATK3 with default settings, except for option "-hets" where the default value was changed from 0.001 to 0.01.

We applied several filtering criteria with VCFtools v0.1.11 (Danecek et al. 2011) and a custom script (Owens et al. 2016) to arrive at a set of high-quality SNPs to form the basis of subsequent population genetic analysis. Namely, we eliminated insertion/deletion polymorphisms to retain only SNPs that had only two alleles, and SNPs that were only present in the outgroups (Dolly Varden and Arctic Char). We also eliminated SNPs that showed an observed heterozygosity of 0.6 or higher as these are likely the result of mapping to paralogous regions of the genome, SNPs with genotype quality below 10 (these have a 10% chance of being incorrect genotypes), SNPs mapping to the mitochondria or the three sex-linked scaffolds, SNPs with missing genotypes in more than 30% of Bull Trout samples, and low frequency SNPs (SNPs present at a frequency below 5% in our Bull Trout sample). 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 among Bull Trout samples using our dataset that was trimmed to eliminate SNPs in high linkage disequilibrium. First, to generate a general picture of differences among samples, 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. 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 in an effort 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, Q). We ran five replicates of Admixture for each K from 1 to 9 and terminated each run when the difference in log-likelihood between successive iterations fell below $1 \times 10^{-9}$. We chose the K value that minimized the cross-validation error, i.e., that best fit the data (Alexander et al. 2009). After determining that K=2 was the best fit to our data, we made one last run with 1,000 bootstraps to estimate the standard error of the inferred admixture proportions.

## SNP Genotyping Assays

We used VCFtools (Danecek et al. 2011) to estimate Weir and Cockerham’s $F_{ST}$ (Weir and Cockerham, 1984) for each SNP in our autosomal dataset (prior to removal of

SNPs in close linkage). We then inspected each SNP in descending order of their $F_{ST}$ rank 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 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), 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 either from different chromosomes or if from the same chromosome, we required they be at least 30 Mbp from another selected SNP. Eleven SNPs that passed these criteria were submitted for TaqMan assay design using the ThermoFisher online design tool and ordered for testing (Table 3).

**Table 3.** TaqMan assays ordered and tested to genotype Bull Trout samples. Shaded are the six TaqMan assays that passed testing and were used for SNP genotyping.

| TaqMan Assay | SNP name | LG | $F_{ST}$ | $F_{ST}$ Rank | Amplification | Discrimination |
|---|---|---|---|---|---|---|
| BTU_2 | NW_019943400.1:117942 | unplaced | 0.823 | 1 | yes | no |
| BT9_1 | NC_036849.1:30074307 | LG9 | 0.811 | 2 | yes | yes |
| BT8_5 | NC_036848.1:43995570 | LG8 | 0.764 | 7 | yes | yes |
| BT8_7 | NC_036848.1:13985783 | LG8 | 0.747 | 9 | yes | yes |
| BT25_8 | NC_036865.1:22128237 | LG25 | 0.707 | 21 | yes | yes |
| BT28_6 | NC_036868.1:10445677 | LG28 | 0.696 | 23 | no | no |
| BT35_12 | NC_036874.1:13256446 | LG35 | 0.687 | 29 | no | no |
| BT27_15 | NC_036867.1:10122966 | LG27 | 0.680 | 33 | yes | yes |
| BT5_45 | NC_036844.1:14802962 | LG5 | 0.660 | 36 | yes | no |
| BT18_16 | NC_036858.1:29856596 | LG18 | 0.660 | 37 | yes | yes |
| BT14_20 | NC_036854.1:25021369 | LG14 | 0.627 | 52 | yes | no |

We selected an initial set of 94 samples for assay testing: 86 were samples of unknown genotype collected from the Peace River, and 8 had been used for GBS and

had known genotypes for most of the 11 loci (5 samples were from the Pine River and 3 were from the Halfway River). We ensured that amongst the 8 samples all possible genotypes at each of the 11 loci were included (i.e., for each locus, genotypes homozygous for allele 1, homozygous for allele 2 and heterozygous were represented). Samples were genotyped in 384-well plates following the manufacturer's instructions: 2.5 µL TaqMan genotyping master mix 2X, 0.25 µL TaqMan assay 20X, 1 µL DNA (concentration between 2 and 5 ng/µL), and 1.25 µL water. Genotyping was performed in a Viia7 Real-Time PCR system (Applied Biosystems, Foster City, CA, USA) with the following cycling conditions: 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. We called the genotypes for each sample at each locus by visual inspection of the plots of the ΔRn values of each allele. The Rn value is the reporter dye (VIC or FAM) signal of each allele normalized by the fluorescence signal of the ROX dye, and ΔRn is Rn minus the baseline. Genotyping was successful at 6 of the 11 assays (i.e. good amplification and allelic discrimination was achieved, Table 3). The remaining samples from the Peace River and additional samples from the tributaries of known genotypes were genotyped for the six successful TaqMan assays with the conditions detailed above.

Assignment Tests

We used our TaqMan generated dataset (six SNP genotypes for 517 samples) to assign Bull Trout samples to the genetic groups identified. We used the program GeneClass2 (Piry et al., 2004) to compute the probability that the multilocus genotype of each individual belonged in one of two genetic groups (Halfway or Pine river; see the Results section for details). To do this we first defined the two reference samples

(Halfway: 14 Halfway and 2 Moberly individuals, and Pine: 26 Pine individuals; all fish used for the reference samples had low admixture coefficients in the other genetic group in the Admixture analysis below). We tested our ability to assign these fish to their known population of origin using the standard assignment test employing maximum likelihood using GeneClass2. Next, we simulated a dataset of 100,000 multilocus genotypes based on the allele frequencies in those reference samples following the Monte-Carlo resampling procedure of Paetkau et al. (2004). Fish in the Halfway and pine reference samples were then subject to an 'exclusion' test. Here, a fish was excluded as belonging to one *or either* of the reference samples if its probability of belonging to one or either reference sample was less than 0.01. Any individual was considered unassigned if the probability of belonging to each of the two reference groups was lower than 0.01. Finally, fish of unknown origin from the Peace River mainstem were subject to the exclusion tests after the veracity of our assays to correctly identify fish of known origin was established (see Results).

## RESULTS

### SNP Discovery

From a total of 382 million paired end reads obtained through the sequencing of our GBS pooled library, an average of 3.8 million reads were assigned to each sample after demultiplexing (Table 4; Appendix I). These data will be deposited at the NCBI SRA. Of those reads, an average of 98.9% mapped to the Arctic Char reference sequence which suggests that the reference sequence is close to compete. On average, 61.2% of the reads per sample were used for variant identification with the HaplotypeCaller function in GATK3. The main reason for this is low read mapping quality, not because the

reference is from a closely related species and not from Bull Trout itself (63.3% of the Arctic Char reads mapped to the Arctic Char genome, a fraction similar to the average 61.2%), but most likely because of the salmonid specific whole genome duplication ~95 MYA (Mcqueen and Johnston 2014), which may result in a large fraction of reads not mapping uniquely in the genome. On average, 0.45% of the genome of each sample had 5 or more reads mapping to it with high quality (BWA-MEM mapping quality of 20 or higher, i.e., reads with a probability of an incorrect alignment lower than 0.01).

**Table 4.** Average, maximum and minimum number of reads generated, mapped and used for SNP calling per Bull Trout sample used in the GBS library sequencing (across the 92 samples in total).

| | Mean (SE) | Maximum | Minimum |
|---|---|---|---|
| Demultiplexed reads[1] | 3,788,138 (1,218,964) | 11,662,316 | 2,186,342 |
| Mapped reads[2] | 3,745,522 (1,194,180) | 11,546,014 | 2,168,762 |
| SNP calling reads[3] | 2,339,906 (714,949) | 7,274,051 | 1,404,927 |

[1] Number of reads assigned to each sample with the demultiplexing script of Baute et al. (2016)
[2] Total reads mapped with BWA-MEM
[3] Number of reads used by HaplotypeCaller to call variants

Over all 94 samples, Arctic Char and Dolly Varden included, GenotypeGVCFs in GATK3, identified 663,244 putative genetic variants. We eliminated insertion and deletion variants from this file and retained 532,689 SNPs, of which 524,075 SNPs passed the observed heterozygosity filter (heterozygosity<0.6). We next eliminated SNPs mapping to the mitochondrial genome and the three sex-linked contigs and retained 485,022 SNPs. Of those, we then kept only 99,120 SNPs present in the 92 Bull Trout samples, with 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 7,564 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 and used 3,582 SNPs. These final datasets used for downstream analyses will be deposited on DRYAD.

Population Structure and Differentiation

Results from a PCA (Figure 1A; Appendix I) revealed that the main axis of variation (explaining 8.8% of variation) in our dataset separated samples from the Pine River watershed from samples from the Halfway River and Moberly River watersheds. The Admixture analysis indicated that the model of population structure that best fit our data was for K=2 (Supporting Figure 1). In other words, the data resolved two distinct genetic groups of Bull Trout in the LAA. The average admixture coefficients expressed in terms of group 1 ($Q_1$) were 0.01 for the Halfway River watershed, 0.01 for the Moberly River watershed and 0.95 for the Pine River watershed. Overall there was little evidence of admixture between the two genetic groups (Halfway/Moberly vs Pine watersheds, Figure 1B; Appendix I). The exceptions were: one sample in the Halfway River watershed (Colt Creek) which had $Q_1$=0.22 (indicating that it had a substantial component of genetic group 1) and three samples from the Pine River watershed with $Q_1$<0.8 (one from the Pine River itself, $Q_1$=0.55; one from the Callazon Creek, $Q_1$=0.72; and one from the Blind Creek, $Q_1$=0.79).

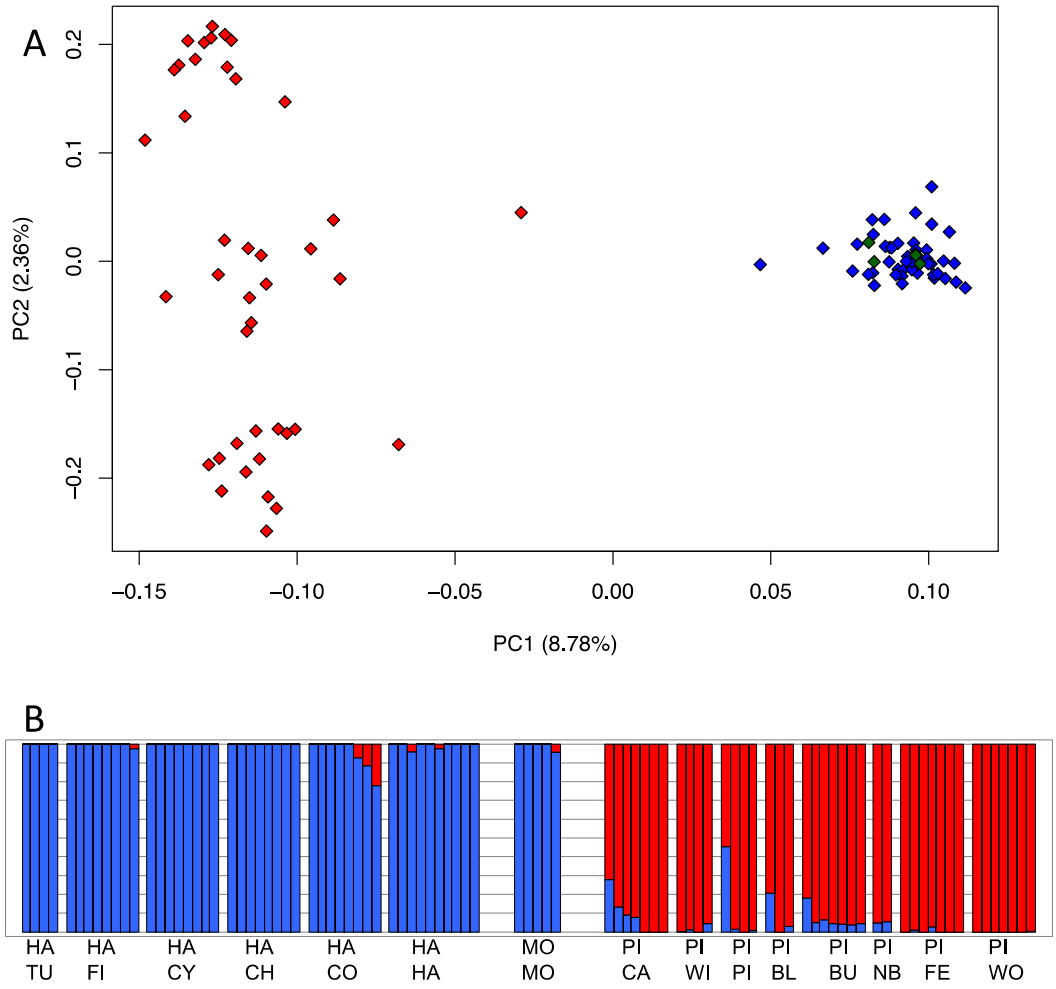



**Figure 1. Population structure of Bull Trout in the Peace Region**. The Halfway River (and its tributaries; n=46) and the Moberly River (n=5) are located upstream of the Project and the Pine River (and its tributaries; n=41) is located downstream of the Project. Both the PCA (A) and Admixture (B) analysis were performed on 3,582 SNPs. (A) Each diamond represents a single Bull Trout sample. Pine samples are plotted in red, Moberly samples in green and Halfway samples in blue. The first two principal components (PC) are plotted and the percentage of variation in the data explained by each PC are indicated in the axis name. (B) Admixture results shown are for K=2, the number of populations that best fit the data. Each column represents the genotype of an individual fish and the two colours in each column represent the proportion of the genome of each fish that is assigned to each population: blue, Halfway and Moberly rivers and red, the Pine River. Samples from different tributaries in each watershed are separated by an empty column and different watersheds are separated by two empty columns. Top letters indicate the watersheds: HA (Halfway River), MO (Moberly River), and PI (Pine River). Bottom letters indicate the tributary within the watershed: TU (Turnoff Creek), FI (Fiddes Creek), CY (Cypress Creek), CH (Chowade River), CO (Colt Creek), HA (Halfway River), MO (Moberly River), CA (Callazon Creek), WI (Willow Creek), PI (Pine River), BL (Blind Creek), BU (Burnt River), NB (North Burnt River), FE (Fellers Creek), and WO (Wolverine River).

The second PCA axis (Figure 1A; Appendix I) explained much less variation than the first one (2.4 vs 8.8%) and revealed some additional genetic structuring within the

Pine River watershed (Figure 1A and Supporting Figure 2). No genetic differentiation was detected between the Moberly River watershed and Halfway River watershed in further PCA axes (Supporting Figure 3). The PCA and Admixture analyses that included more polymorphisms (i.e., adding SNPs that were below our threshold of a minimum allele frequency of 0.05) also failed to detect genetic differentiation between these watersheds (results not shown).

We next calculated weighted average pairwise $F_{ST}$ (a measure of population differentiation) between all three watersheds, with the 7,564 SNP dataset (Table 5). This analysis confirmed the inference above of no genetic differentiation between the Halfway River and Moberly River watersheds ($F_{ST}$=0.002) and their level of genetic differentiation is close to two orders of magnitude lower than between either and the Pine ($F_{ST}$=0.105 between Halfway and Pine and $F_{ST}$=0.091 between Moberly and Pine). Henceforth, we refer to the two genetic groups identified as the Halfway and the Pine to reflect the fact that samples caught in the Moberly likely originated in the Halfway River watershed (see discussion). Weighted average pairwise $F_{ST}$ between the Halfway and Pine groups was 0.105, but there was considerable variation among SNPs, with the vast majority close to zero, and the 99$^{th}$ percentile at 0.59, i.e., 1% of SNPs (approximately 75 SNPs) showed an $F_{ST}$ between these two groups of 0.59 or higher (Figure 2).

**Table 5.** Weighted average pairwise $F_{ST}$ estimates among Bull Trout from the three watersheds surveyed.

| | Halfway River | Moberly River | Pine River |
|---|---|---|---|
| Halfway River | | | |
| Moberly River | 0.002 | | |
| Pine River | 0.105 | 0.091 | |

SNP Genotyping

We selected 11 SNPs with high $F_{ST}$ estimates between the Halfway and Pine groups to design TaqMan assays to genotype samples of Bull Trout from the Peace River to allow their assignment to the Halfway or Pine groups (Table 3). Six assays passed our initial test, because they provided both good amplification and were able to discriminate homozygotes for alternative alleles and heterozygotes at each locus (Table 3). Two were located in the same Arctic Char linkage group (LG8) but were more than 30 Mbp away from each other and genetically unlinked in our sample. The remaining four were located in linkage groups LG9, LG18, LG25 and LG27. Genotyping success was high: only 10 out of 3,102 genotypes (517 samples over 6 loci) failed (Appendix II).

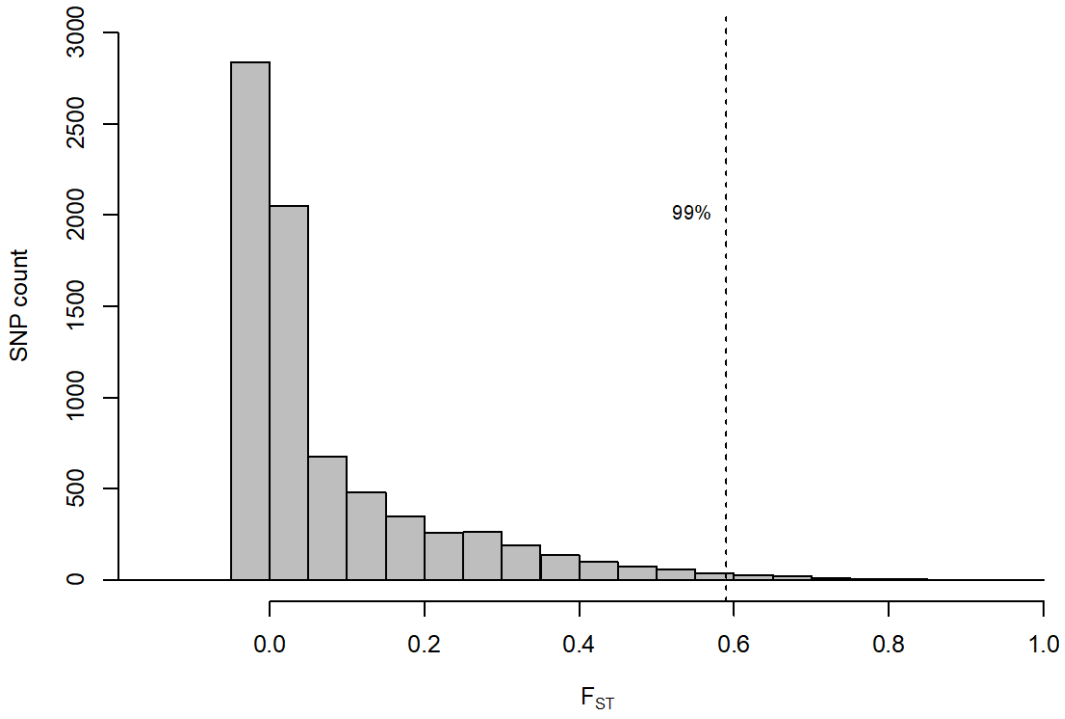



**Figure 2. Levels of population differentiation between Bull Trout samples from the Halfway (upstream of the Project) and Pine (downstream of the Project) groups for each SNP in the dataset as estimated by $F_{ST}$.** The dashed vertical line is at 0.59, the 99$^{th}$ percentile of $F_{ST}$ estimates.

Population Assignment

All samples used as reference samples (26 from the Pine, 14 from the Halfway and 2 from the Moberly) were correctly assigned to their known population groups with a high degree of confidence (i.e., their assignment to one population or the other was greater than 99.5% and their observed likelihood of population membership was typically represented by > 1% of those generated from the simulated genotypes of one population and 1% or fewer from the simulated genotypes of the other population; Appendix II). Two samples from the Pine River watershed that had coefficients of Admixture ($Q_1$) too high to be included in the reference samples ($Q_1$=0.545 and $Q_1$=0.723) were also assigned to the Pine group. These tests (Table 6) revealed that 92% (436 samples) of Bull Trout caught in the mainstem Peace River belong to the Halfway group while only 4% (19 samples) belong to the Pine group.

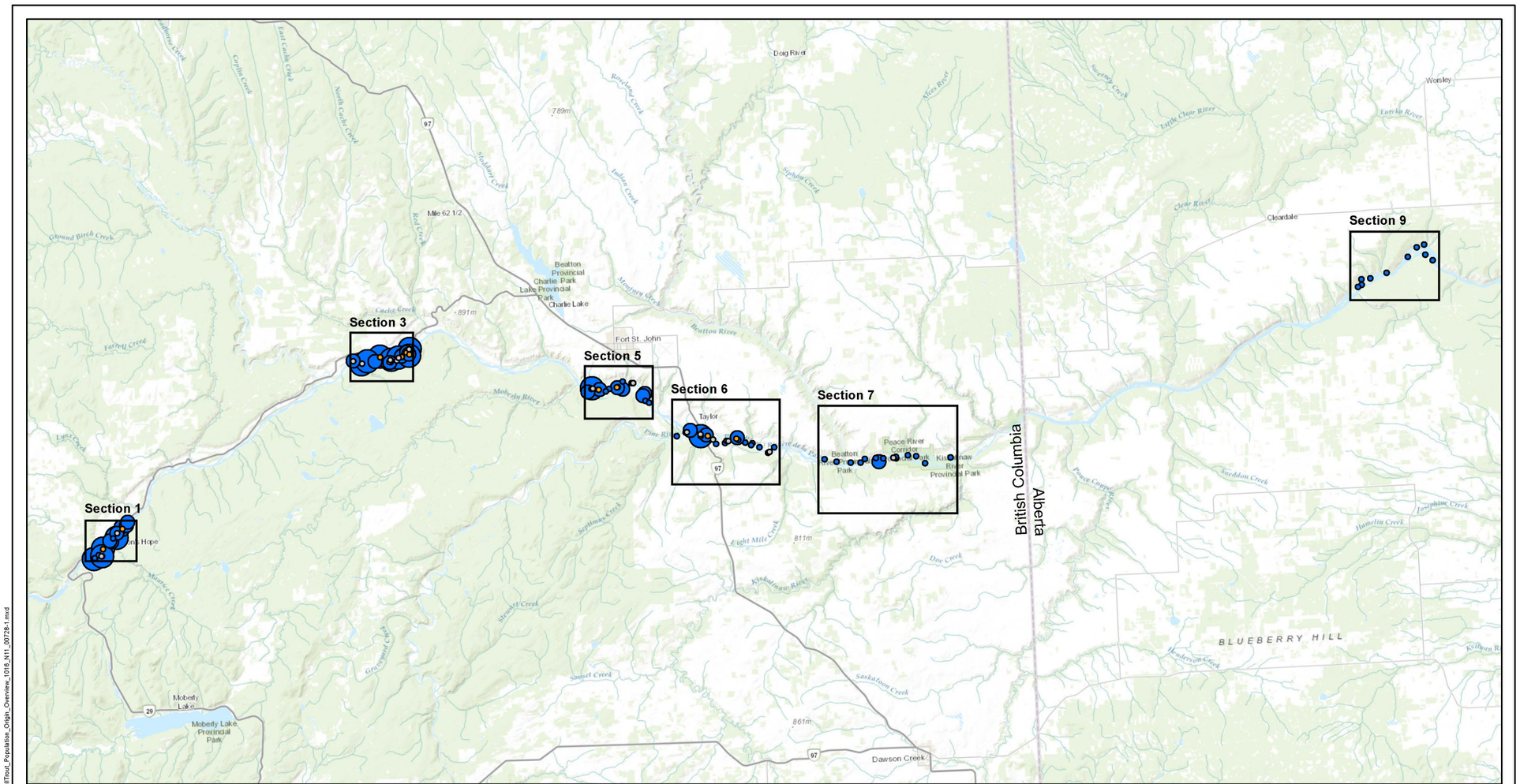
We could not assign 4% of the samples (18 fish) to either group because they had a probability of belonging to either group of less than 0.01 (Appendix II). There was little variability in the proportion of fish in each of these two groups between sampling years 2017 (n=222) and 2018 (n=210). Results from 2016 were less similar to the results of 2017 and 2018 however the sample size in 2016 was considerably lower (n=41). Finally, 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 upstream of the Project and Sections 5, 6, 7 and 9 located downstream of the Project (Figure 3, Table 6 and Appendix II). There was considerable variation in the proportion of fish assigned to each group across sampling sections. Fish assigned to the Pine group and fish that we could not assign (see above) were most common in Section 6

(located at the confluence of the Peace and Pine rivers; Figure 3, Table 6 and Appendix II).

**Table 6.** Number of Bull Trout assayed and assigned (% of total) to the Halfway or Pine groups after being genotyped at six ancestry informative SNPs with TaqMan assays.

| Location | Year | Total | Halfway | Pine | Unassigned |
|---|---|---|---|---|---|
| Peace River Section 1 | All Years | 118 | 113 (95.8%) | 3 (2.5%) | 2 (1.7%) |
| | 2016 | 8 | 8 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| | 2017 | 54 | 51 (94.4%) | 1 (1.9%) | 2 (3.7%) |
| | 2018 | 56 | 54 (96.4%) | 2 (3.6%) | 0 (0.0%) |
| Peace River Section 3 | All Years | 161 | 146 (90.7%) | 7 (4.3%) | 8 (5.0%) |
| | 2016 | 8 | 8 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| | 2017 | 79 | 72 (91.1%) | 3 (3.8%) | 4 (5.1%) |
| | 2018 | 74 | 66 (89.2%) | 4 (5.4%) | 4 (5.4%) |
| Peace River Section 5 | All Years | 81 | 75 (92.6%) | 4 (4.9%) | 2 (2.5%) |
| | 2016 | 8 | 7 (7.5%) | 0 (0.0%) | 1 (12.5%) |
| | 2017 | 38 | 35 (92.1%) | 3 (7.9%) | 0 (0.0%) |
| | 2018 | 35 | 33 (94.3%) | 1 (2.9%) | 1 (2.9%) |
| Peace River Section 6 | All Years | 68 | 58 (85.3%) | 5 (7.4%) | 5 (7.4%) |
| | 2016 | 9 | 8 (88.9%) | 0 (0.0%) | 1 (11.1%) |
| | 2017 | 33 | 27 (81.8%) | 4 (12.1%) | 2 (6.1%) |
| | 2018 | 26 | 23 (88.5%) | 1 (3.8%) | 2 (7.7%) |
| Peace River Section 7 | All Years | 27 | 26 (96.3%) | 0 (0.0%) | 1 (3.7%) |
| | 2016 | 3 | 3 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| | 2017 | 10 | 10 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| | 2018 | 14 | 13 (92.9%) | 0 (0.0%) | 1 (7.1%) |
| Peace River Section 9 | All Years | 18 | 18 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| | 2016 | 5 | 5 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| | 2017 | 8 | 8 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| | 2018 | 5 | 5 (100.0%) | 0 (0.0%) | 0 (0.0%) |
| All Peace River Sections | All Years | 473 | 436 (92.2%) | 19 (4.0%) | 18 (3.8%) |
| | 2016 | 41 | 39 (95.1%) | 0 (0.0%) | 2 (4.9%) |
| | 2017 | 222 | 203 (91.4%) | 11 (5.0%) | 8 (3.6%) |
| | 2018 | 210 | 194 (92.4%) | 8 (3.8%) | 8 (3.8%) |

**A**




**Predicted spawning origin of Bull Trout captured in the Peace River from 2016 to 2018**

| Halfway River | Pine River | Unknown |
|---|---|---|
| 1 - 4 | 1 - 2 | 1 - 2 |
| 5 - 9 | | |
| 10 - 17 | | |

Map Notes:
1. Datum: NAD83
2. Projection: UTM Zone 10N, UTM Zone 11N
3. Base Data: ESRI National Geographic World Topographic

★ Confluence of Halfway River and Peace River
★ Confluence of Pine River and Peace River

1:468,000 0 5 10 20 km

BC Hydro

**Site C Fisheries & Aquatic Habitat Monitoring and Follow-up Program**
**Bull Trout Population of Origin**
**Overview**

| Date | Jan 7, 2020 | DWG NO | 1016-N11-00728-1 | R 0 |
|---|---|---|---|---|

Path: X:\ArcGIS\Projects\Environment\Fisheries_BullTrout_Population_Origin_Overview_1016_N11_00728-1.mxd



Construction of the Site C Clean Energy Project is subject to required regulatory and permitting approvals.

**B**

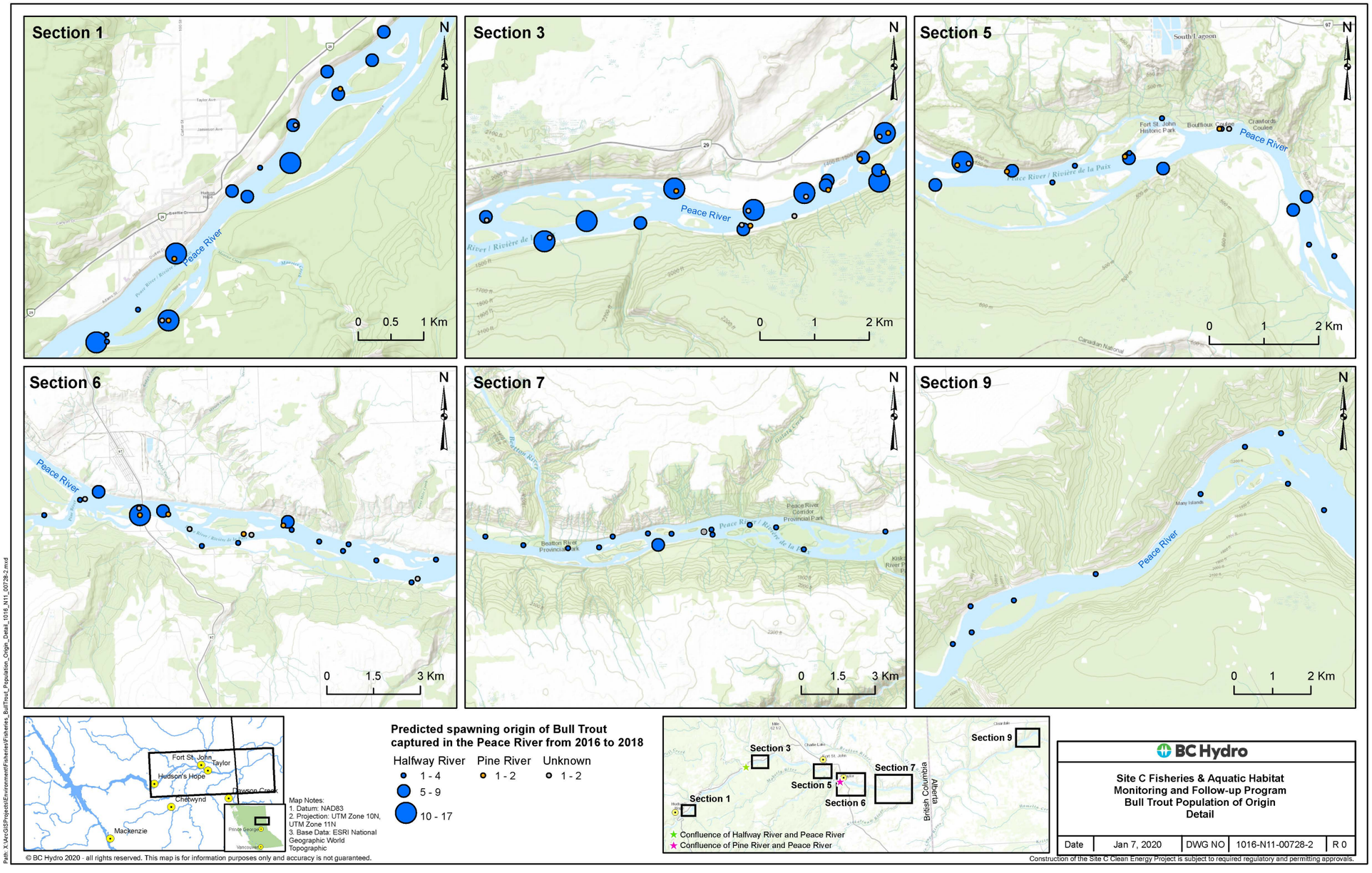



**Figure 3. Predicted population of origin (blue: Halfway, Pine: orange) of subadult and adult Bull Trout captured in Sections 1, 3, 5, 6, 7, and 9 of the Peace River from 2016 to 2018.** Circles are proportional to frequency. Samples that could not be assigned to either group are shown in gray. Panel A shows the geographic location of each sampling section and Panel B shows each section of the Peace River in detail. Map courtesy of BC Hydro. Full details can be found in Appendix II.

## DISCUSSION

The GBS approach that we adopted to assay genetic variation in Bull Trout has provided an efficient mechanism by which to guide fisheries management in the Peace Region and fish passage management at Site C. From the millions of base pairs of the genome interrogated, we developed a small panel of SNPs to identify the spawning population of origin of fish captured in the Peace River. We found a major difference between fish sampled in the Halfway and Moberly rivers from those sampled in the Pine River. The lack of differentiation between Bull Trout sampled in the Halfway and Moberly rivers is consistent with there being no known spawning population of Bull Trout in the Moberly River (Mainstream Aquatics Ltd. 2012) and with the idea that Bull Trout use the latter as an occasional foraging area. Consequently, our discussion focuses on the identified genetic groups as the Halfway and Pine. Our assays showed that the vast majority of Bull Trout captured in the Peace River likely originated from the Halfway River watershed, which is consistent with the findings from otolith and fin ray microchemistry and fish sampling (Mainstream Aquatics Ltd. 2012; Earth Tone Environmental R&D & Mainstream Aquatics Ltd. 2013; TrichAnalytics 2020). Future SNP panel assays in this context can likely be achieved with relatively short turn-around time, i.e., two weeks or perhaps less from tissue receipt for a sample of approximately 300 fish. The overall productivity of our SNP discovery was consistent with recent work on Lake Trout (*Salvelinus namaycush*) across a similar geographic expanse in Québec where Bernatchez et al. (2016) recovered an average of about 3.2 million sequence reads per individual (n=320) and a final quality-filtered dataset of between 3,295 and 4,968 SNPs.

Our data were also highly consistent with the microsatellite DNA work of Taylor et al. (2014) which also demonstrated strong allele frequency differences between the Halfway and Pine rivers, and $F_{ST}$ values were comparable between the two studies although slightly higher in the current study (0.07-0.08 in Taylor et al. (2014) and 0.11 in the current study). Levels of divergence between the Moberly and Halfway rivers were not assessed by Taylor et al. (2014) owing to a lack of Bull Trout samples from the Moberly River. Furthermore, the current analysis corroborates the distinction among various tributaries of the Pine River watershed that was demonstrated by Taylor et al. (2014).

A small number (18 individuals or 4%) of the Peace River samples were confidently excluded as originating from the Halfway and Pine groups (i.e., the observed multiple locus genotypes of these fish had less than a 1% chance of occurring in either group). This suggests that these fish likely originated in tributaries that were not represented in our original SNP discovery library and that differ substantially from the genetic character of Bull Trout from the Halfway and Pine rivers. Such “ghost populations” (Beerli 2004; Slatkin 2005) have long been known to be a factor potentially influencing estimates of migration rates between populations and related tasks such as assignment analyses. Given that Bull Trout may travel well over 100 kilometres within large rivers systems (e.g., Baxter 1997; Starcevich et al. 2012; Taylor et al., 2020, in review), including the Peace River (AMEC Earth & Environmental and LGL Ltd. 2008, 2010a,b), it is entirely likely that the occasional migrant from well upstream or downstream of the Halfway or Pine rivers may occur within the vicinity of the Project and lead to an ambiguous assignment (see also Mainstream Aquatics Ltd. 2012). Because, however, the “unassigned” Bull Trout represented such a small proportion of

the large sample assayed it is perhaps not a top priority to sample more distant putative source populations.

Sex-specific markers have been developed for *Salvelinus* spp. (Yano et al. 2013) and preliminary work in our lab confirmed a PCR protocol for fast and accurate sex identification in Bull Trout. This may be useful to assess if there are any sex-related biases in movement-related behaviour (e.g., timing of arrival at the Project) or in survival and productivity after trap-and-haul.

Finally, we have begun the bioinformatic analysis of a successfully-prepared and sequenced Arctic Grayling library to design assays similar to those employed for Bull Trout. We have also assembled the sample panel for Rainbow Trout which will form the basis of the GBS library used in SNP discovery. This work will be proceeding concurrently with the ongoing bioinformatic analysis of the Arctic Grayling sequence data.

## 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

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology, 26*, 420-430. https://doi.org/10.1111/mec.13948

AMEC Earth & Environmental and LGL Ltd. 2008. Movements of radio-tagged Bull Trout and Arctic Grayling in the upper Peace River system, 1996 to 1999. Report prepared for BC Hydro Vancouver, BC.

AMEC Earth & Environmental and LGL Ltd. 2010a. Peace River Fisheries Investigation, Peace River and Pine River Radio Telemetry Study 2009. Report prepared for BC Hydro Vancouver, BC.

AMEC Earth & Environmental and LGL Ltd. 2010b. Analysis and Assessment of the Ministry of Environment's Peace River Bull Trout and Arctic Grayling Radio Telemetry Database 1996 to 1999. Report prepared for BC Hydro Vancouver, BC.

Baute, G. J., Owens, G. L., Bock, D. G., & Rieseberg, L. H. (2016). Genome- wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. *American Journal of Botany*, *103*, 2170–2177. https://doi. org/10.3732/ajb.1600295

Baxter, J.S. (1997). Aspects of the reproductive ecology of bull trout (Salvelinus confluentus) in the Chowade River, British Columbia. M.Sc. Thesis, University of British Columbia, Vancouver, British Columbia, 97 p.

Beerli, P. (2004). Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, *13*, 827-836. https://doi.org/10.1111/j.1365-294X.2004.02101.x

Bernatchez, S., Laporte, M., Perrier, C., Sirois, P., & Bernatchez, L. (2016). Investigating genomic and phenotypic parallelism between piscivorous and planktivorous lake trout (*Salvelinus namaycush*) ecotypes by means of RADseq and morphometric analyses. *Molecular Ecology, 25,* 4773.4792. https://doi.org/10.1111/mec.13795

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Carvalho, G. R., & Pitcher, T.J. (2012). Molecular genetics in fisheries. Springer Science & Business Media.

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

Christensen, K. A., Rondeau, E. B., Minkley, D. R., Leong, J. S., Nugent, C. M., Danzmann, R. G., Ferguson, M. M., Stadnik, A., Devlin, R. H., Muzzerall, R., Edwards, M., Davidson, W. S., & Koop, B. F. (2018). The Arctic charr (Salvelinus alpinus) genome and transcriptome assembly. *PloS one*, *13*(9), e0204076. https://doi.org/10.1371/journal.pone.0204076

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

Earth Tone Environmental R&D & Mainstream Aquatics Ltd. (2013). Site C Fisheries Studies 2012 Elemental Signature Study - Final Report. Prepared for B.C. Hydro Site C Project, Corporate Affairs Report No. 12007F: 163 p. + appendices.

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

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

Hand, B. K., Muhlfeld, C. C., Wade, A. A., Kovach, R. P., Whited, D. C., Narum, S. R., Matala, A. P., Ackerman, M. W., Garner, B. A., Kimball, J. S., & Stanford, J.A. (2016). Climate variables explain neutral and adaptive variation within salmonid metapopulations: the importance of replication in landscape genetics. *Molecular Ecology*, *25*(3), pp.689-705. https://doi.org/10.1111/mec.13517

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

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England), 25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England), 25,* 2078–2079. https:// doi.org/10.1093/bioinformatics/btp352

Macqueen, D. J., & Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. R. Soc. B, 281,* 20132881. http://doi.org/10.1098/rspb.2013.2881

Mainstream Aquatics Ltd. (2012). Site C Clean Energy Project – Fish and Fish Habitat Technical Data Report. Prepared for BC Hydro Site C Project, Corporate Affairs Report No. 12002F: 239 p.

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

Paetkau, D., Slade, R., Burden, M., & Estoup, A. (2004). Direct, real-time estimation of migration rate using assignment methods: a simulation-based exploration of accuracy and power. *Molecular Ecology, 13*, 55-65. https://doi.org/10.1046/j.1365-294X.2004.02008.x

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

Rajora, O.P. ed. (2019). Population Genomics: Concepts, Approaches and Applications. Springer.

Slatkin, M. (2005). Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology, 14,* 67-73. https://doi.org/10.1111/j.1365-294X.2004.02393.x

Starcevich, S. J., Howell, P. J., Jacobs, S. E., & Sankovich, P. M. (2012). Seasonal movement and distribution of fluvial adult bull trout in selected watersheds in the mid-Columbia River and Snake River basins. *PLoS ONE, 7.* e37257 https://doi.org/10.1371/journal.pone.0037257

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.

Taylor, E. B., Chudnow, R., Spendlow, I., Pillapow, R., & van Poorten, B. (*in review*). Microsatellite DNA analysis of overwintering bull trout (*Salvelinus confluentus*) and its implications for harvest regulation and habitat management. Submitted to *Fisheries Management and Ecology*, Jan. 2020.

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

TrichAnalytics. 2020. Fish Otolith and Fin Ray Microchemistry Study. Report prepared for BC Hydro – Site C Clean Energy Project – Vancouver, BC. 102 pages + 3 appendices.

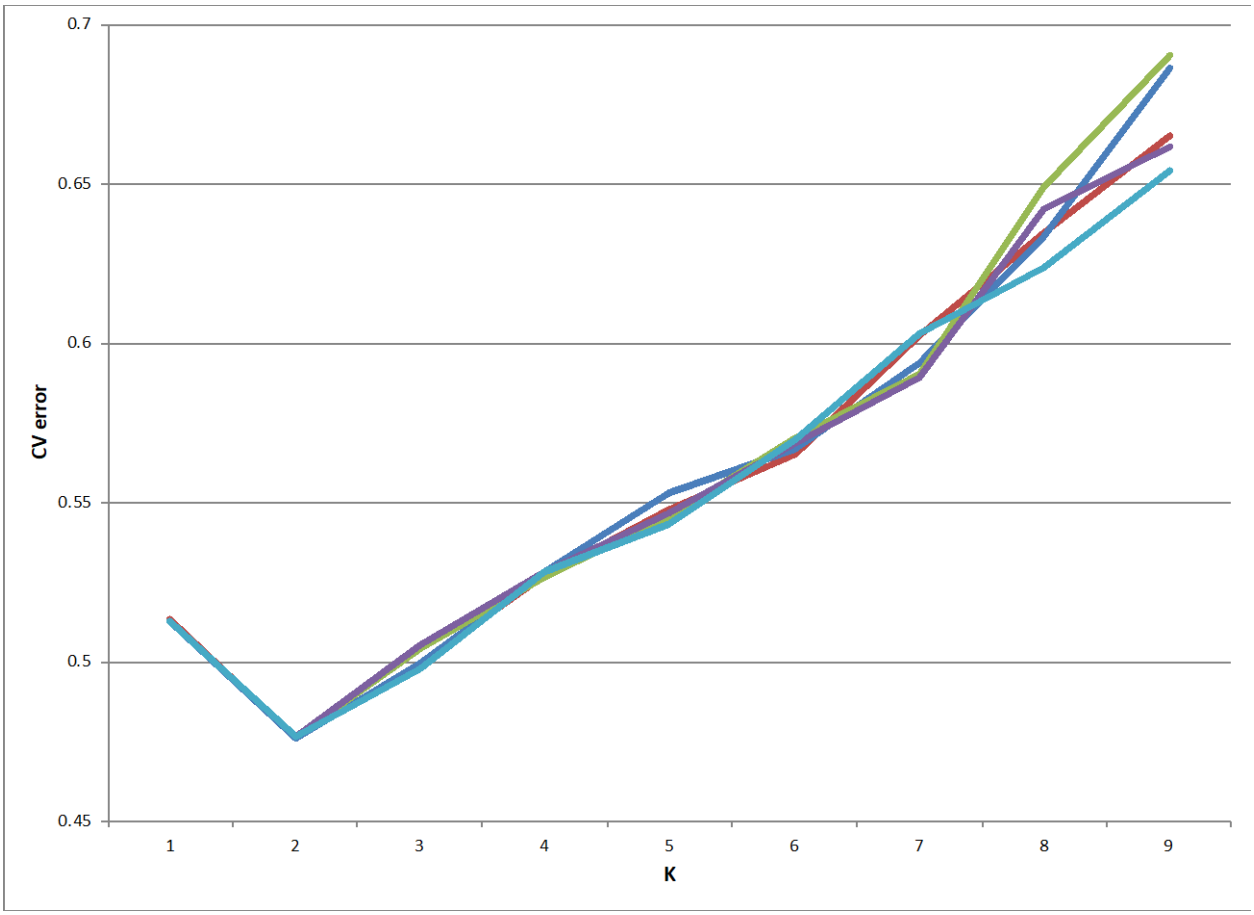
Valenzuela-Quiñonez, F. (2016). How fisheries management can benefit from genomics? *Briefings in functional genomics*, *15*, 352-357. https://doi.org/10.1093/bfgp/elw006

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

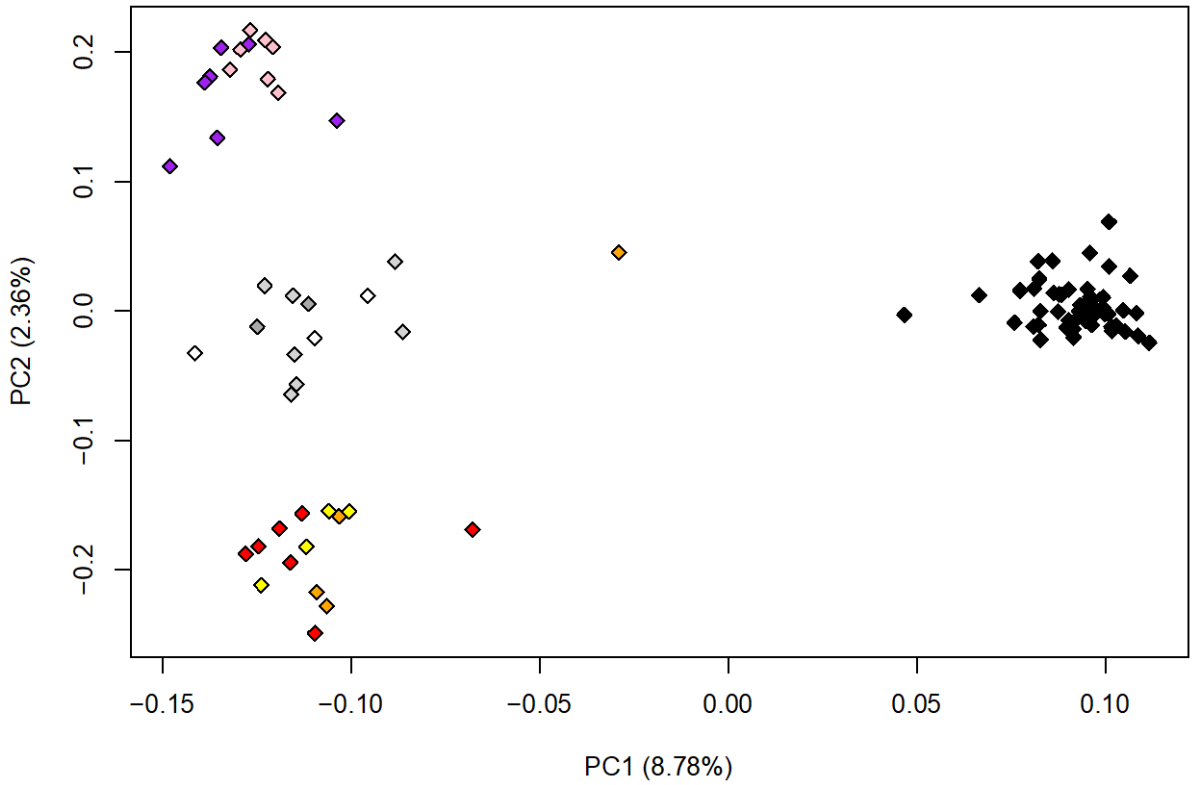
Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R. and Guiguen, Y., 2013. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications*, *6(3),* 486-496. https://doi.org/10.1111/eva.12032

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.

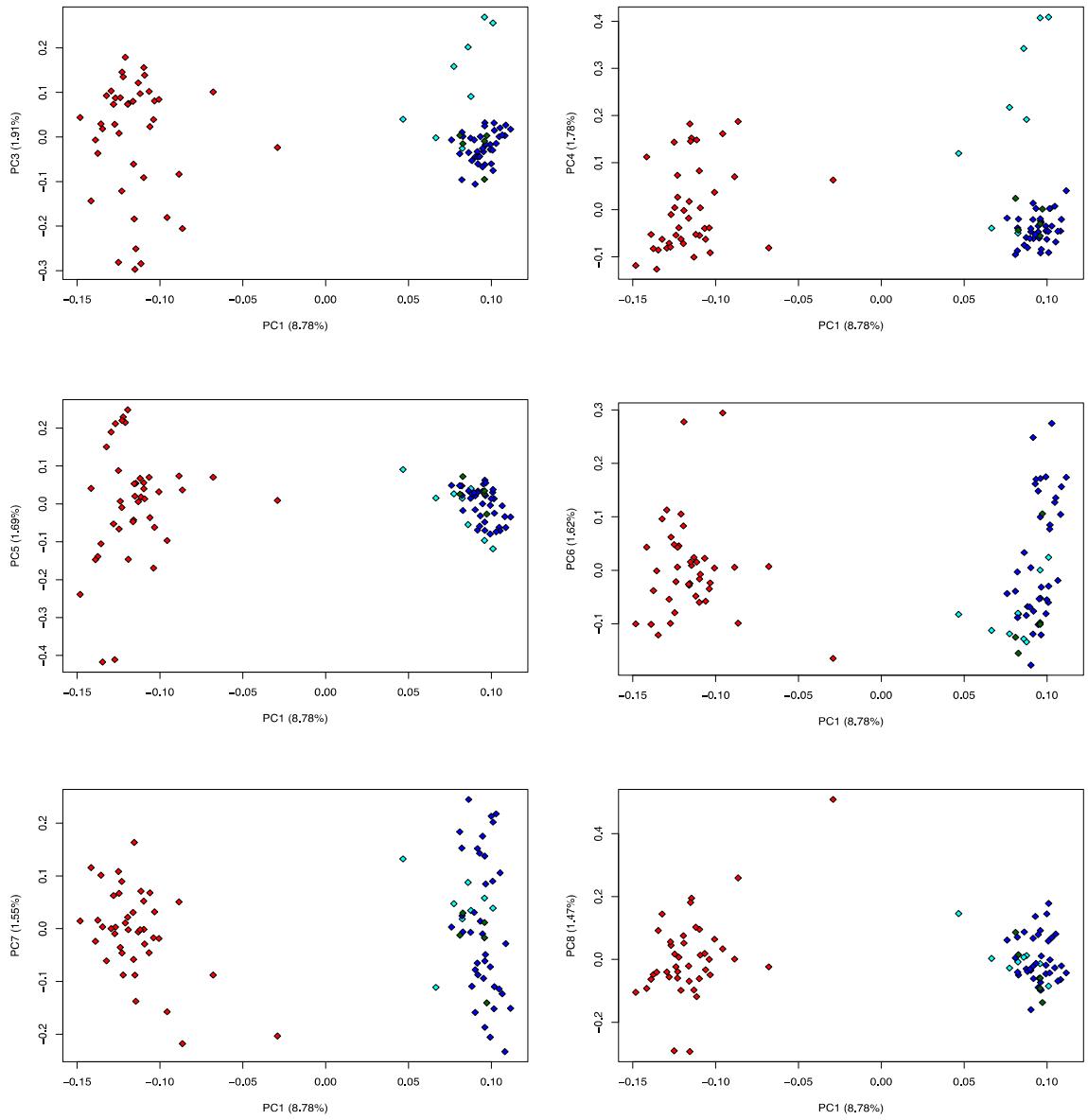
## SUPPORTING FIGURES




**Supporting Figure 1. Coefficient of variation for the Bull Trout Admixture analysis.** Results are shown for the five replicates of the analysis with the 3,582 SNPs dataset with K varying from 1 to 9. K=2, i.e. two populations, minimized the coefficient of variation and is therefore the best fit to our data.




**Supporting Figure 2. Additional Bull Trout population structure in the Pine River watershed.** The results presented are the same as in Figure 1A, but the colour scheme used is different to highlight the patterns in the Pine River watershed along PC2. Northwestern most tributaries are shown in red (Callazon Creek), orange (Pine River) and yellow (Willow Creek), southeastern most tributaries are shown in pink (Fellers Creek) and purple (Wolverine River) and the central ones in white (Blind Creek), light gray (Burnt River) and dark gray (North Burnt River). All Moberly River and Halfway River watershed samples are shown in black.




**Supporting Figure 3. PCA analyses of genetic variability among Bull Trout from the Halfway and Moberly rivers.** Depicted are the plots of PCs 3-8 on the y-axis vs PC1 on the x-axis. Samples from the Pine River are shown in red, Moberly River in green and Halfway River in blue and cyan (Colt Creek). Note that PCs 3 and 4 reveal some differentiation of most Colt Creek samples, but no PC axis reveals differentiation between the Halfway and Moberly rivers.