FA-06 RESERVOIR NATIVE FISH GENETICS BASELINE STUDY INTERIM REPORT

SKAGIT RIVER HYDROELECTRIC PROJECT FERC NO. 553

Seattle City Light

Prepared by:
Cramer Fish Sciences

March 2022
Initial Study Report

TABLE OF CONTENTS

Section No.	Description	Page No.
1.0	**Introduction**	**1-1**
2.0	**Study Goals and Objectives**	**2-1**
3.0	**Study Area**	**3-1**
4.0	**Methods**	**4-1**
4.1	Data Requests	4-1
4.2	Genetic Analysis	4-2
4.2.1	Rainbow Trout	4-2
4.2.2	Bull Trout	4-2
4.2.3	Lineage Relationships	4-3
4.3	Availability of Existing Samples and Coordination of Sampling with Ongoing Activities	4-3
5.0	**Preliminary Results**	**5-1**
5.1	Expert Panel	5-1
5.2	Rainbow Trout (*Oncorhynchus mykiss*)	5-1
5.2.1	Collections	5-1
5.2.2	Genetic Summary Statistics	5-2
5.3	Bull Trout (*Salvelinus confluentus*)	5-10
5.3.1	Collections	5-10
5.3.2	Identification of related individuals within collections	5-11
5.3.3	Population Determination	5-13
5.3.4	Genetic summary statistics	5-15
6.0	**Summary**	**6-1**
6.1	Summary of Completed Objectives	6-1
6.2	Discussion of Rainbow Trout (*Oncorhynchus mykiss*) Genetic Data	6-2
6.3	Discussion of Bull Trout (*Salvelinus confluentus*) genetic data	6-5
6.4	Next Steps	6-6
7.0	**Variances from Proposed Study Plan and Proposed Modifications**	**7-1**
8.0	**References**	**8-1**

List of Figures

Figure No.	Description	Page No.
Figure 3.0-1.	Proposed study area and collections evaluated for Rainbow Trout.	3-2
Figure 3.0-2.	Proposed study area and collections evaluated for Bull Trout.	3-3

Figure 5.2-1. Log-likelihood ratios distribution for simulated true full-siblings versus unrelated individuals based on Roland Creek *O. mykiss* genotype data. High overlap between full-siblings and unrelated fish suggests relatively low power to detect highly related individuals. 5-3

Figure 5.2-2. Scatterplot showing $\log_{10}$-transformed relationship between sample size (n) (*x*-axis) and the R^2 estimator for pairwise linkage disequilibrium between loci (*y*-axis). Strong correlation warrants cautious interpretation of data due to possible bias. 5-5

Figure 5.2-3. Scatterplot of genetic PC-1 (2.215 percent) and PC-2 (2.044 percent) for all collections, excluding samples from the Baker River, estimated using adegenet in program R (Jombart 2011). The distribution of genetic variation appears to support some genetic structuring associated with the geographical locations of collections. River basins are provided to show the approximate geographical locations of each collection. Numbers at centroids identify the collection number listed in Table 5.2-1. Ellipses define 1.5 standard deviations of the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom left corner shows first three eigenvalues. 5-7

Figure 5.2-4. Scatterplot of genetic PC-1 (2.215 percent) and PC-3 (1.836 percent) for all collections, excluding samples from the Baker River, estimated using adegenet in program R. The distribution of genetic variation appears to support existence of genetic structuring associated with the geographical locations of collections. River basin names are provided to describe the approximate geographical locations of each collection. Numbers at centroids identify the collection number listed in Table 5.2-1. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom right corner shows first three eigenvalues. 5-8

Figure 5.2-5. Scatterplot of genetic PC-1 (3.870 percent) and PC-2 (2.028 percent) for all collections upstream of the Gorge Lake Project Boundary estimated using adegenet in program R. Numbers at centroids identify the collection number listed in Table 5.2-1. Scree plot in bottom right corner shows first three eigenvalues. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom left corner shows first three eigenvalues. 5-9

Figure 5.3-1. Log likelihood ratios distribution for simulated true full siblings versus unrelated individuals based on Skagit River *S. confluentus* genotype data. High overlap between full-siblings and unrelated fish suggests relatively low power to detect highly related individuals. 5-12

Figure 5.3-2. Visualization of k-means clustering analysis at k=5 for Bull Trout individuals in dataset for 1st and 2nd principal axes. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in upper right corner shows first three eigenvalues. Cluster 1 were Project Boundary tributary collections and contained a majority of Ross Lake samples. Cluster 2 were

individuals collected from below Gorge Dam. Clusters 3, 4 and 5 were predominantly individuals collected from Diablo and Gorge Lakes. 5-14
Figure 5.3-3. Genetic clusters visualized in Figure 5.3-2 aligned to each Bull Trout collection in dataset. Size of boxes is scaled by sample count. Genetic clusters are organized by geographic location with upper Skagit collections at the top and lower Skagit at the bottom. 5-15

List of Tables

Table No.	Description	Page No.
Table 5.2-1.	Summary statistics for samples collected from *O. mykiss* in the Skagit and Fraser river basins.	5-3
Table 5.3-1.	Bull Trout microsatellite dataset collection summary.	5-11
Table 5.3-2.	Summary statistics for samples collected from Bull Trout in the Skagit River basin.	5-16
Table 5.3-3.	Table of pairwise estimates of F_{ST} from Project Bull Trout collections	5-16

List of Attachments

Attachment A Existing Genetics Data Review Technical Memorandum

List of Acronyms and Abbreviations

AMOVA .....analysis of molecular variance

CI.....confidence interval

City Light.....Seattle City Light

CKMR.....close kin mark recapture

Expert Panel.....Salmonid Genetics Expert Panel

F_{IS}.....inbreeding coefficient

F_{ST}.....genetic divergence

FERC.....Federal Energy Regulatory Commission

FNR.....false negative rate

FPR .....false positive rate

H_O.....observed multilocus heterozygosity within populations

H_S .....expected multilocus heterozygosity within populations

HOR .....Hatchery origin

HWE .....Hardy-Weinberg equilibrium

ILP.....Integrated Licensing Process

ISR .....Initial Study Report

LD .....linkage disequilibrium

LOD .....log-of-the-odds

LP.....licensing participant

N_b .....effective number of breeders

N_e .....effective population size

NMFS.....National Marine Fisheries Service

NOR .....Natural origin

PC.....principal component

PCA.....principal component analysis

Project .....Skagit River Hydroelectric Project

RSP .....Revised Study Plan

SPAN .....Stevan Phelps Allele Nomenclature

USFWS .....U.S. Fish and Wildlife Service

USGS .....U.S. Geological Survey

USR.....Updated Study Report

WDFW.....Washington Department of Fish and Wildlife

This page intentionally left blank.

1.0 INTRODUCTION

The FA-06 Reservoir Native Fish Genetics Baseline Study (Reservoir Fish Genetics Study) is being conducted in support of the relicensing of the Skagit River Hydroelectric Project (Project), Federal Energy Regulatory Commission (FERC) No. 553, as identified in the Revised Study Plan (RSP) submitted by Seattle City Light (City Light) on April 7, 2021 (City Light 2021). On June 9, 2021, City Light filed a "Notice of Certain Agreements on Study Plans for the Skagit Relicensing" (June 9, 2021 Notice)[1] that detailed additional modifications to the RSP agreed to between City Light and supporting licensing participants (LP) (which include the Swinomish Indian Tribal Community, Upper Skagit Indian Tribe, National Marine Fisheries Service [NMFS], National Park Service, U.S. Fish and Wildlife Service [USFWS], Washington State Department of Ecology, and Washington Department of Fish and Wildlife [WDFW]). The June 9, 2021 Notice included agreed to modifications to the Reservoir Fish Genetics Study.

In its July 16, 2021 Study Plan Determination, FERC did not require implementation of the Reservoir Fish Genetics Study. Notwithstanding, City Light implemented the Reservoir Fish Genetics Study as proposed in the RSP with the agreed to modifications described in the June 9, 2021 Notice.

This interim report on the 2021 study efforts is being filed with FERC as part of City Light's Initial Study Report (ISR). City Light will perform additional work for this study in 2022 and include a report in the Updated Study Report (USR) in March 2023.

[1] Referred to by FERC in its July 16, 2021 Study Plan Determination as the "updated RSP."

2.0 STUDY GOALS AND OBJECTIVES

The goals of this study are to characterize baseline population genetic structure for three native salmonid species: Bull Trout (*Salvelinus confluentus*), Rainbow Trout (*Oncorhynchus mykiss*), and Dolly Varden (*Salvelinus malma*) (target species) in Project reservoirs and provide the basis necessary to inform the planning of long-term (i.e., over the new license term) reservoir fish management objectives. Specifically, the goals of this study are to:

- Determine the population genetic structure of within and among target species populations and assess whether management actions are necessary for genetic sustainability.
- Determine the number of fish populations, for each target species, within and among the Project reservoirs.
- Estimate the effective population size ($N_{e)}$ for each target species and reservoir.
- Identify topics and/or management objectives to be considered in the reservoir fish and aquatics management plan.

Specific objectives to meet these study goals are listed below.

Year 1

- City Light will convene an Expert Panel in consultation with LPs.
- Review, compile, and analyze target species genetics data collected by multiple researchers in the Project reservoirs.
    - Acquire and consolidate existing genetics data for Bull Trout, Rainbow Trout, and Dolly Varden.
    - Create a single, standardized data file for each species that compiles genotypes from existing studies.
- Use the standardized data files to evaluate baseline genetic metrics for Bull Trout and Rainbow Trout.
    - Calculate within- and among-population summary statistics using consistent methods for Bull Trout and Rainbow Trout.
    - Estimate relatedness for Bull Trout and Rainbow Trout and report the statistical distribution of this metric by species and reservoir.
    - Estimate the power (false detection rate) of genetic markers currently in use to identify relationships (e.g., parent-offspring pairs, full-sibling-unrelated pairs).
- Identify the availability of relevant existing genetic samples and coordinate target fish species sampling being conducted opportunistically by other relicensing studies and current license field activities.
- Expert Panel review of Year 1 study results and assistance in development of Year 2 study program.

Year 2

- Expand sample collection and/or coordinate existing samples and activities for out-of-basin and above and below dam analyses.
- Continue data collection to address heterozygosity, within- and among-population variance, and relatedness for Dolly Varden in Project reservoirs.
- Gather additional data needed to estimate N_e for each population of Bull Trout, Rainbow Trout, and Dolly Varden.
    - Gather the data needed to estimate N_e during the Integrated Licensing Process (ILP) study period.

Under the June 9, 2021 Notice, City Light and the supporting LPs agreed to four modifications to this Reservoir Fish Genetics Study:

- City Light will modify the FA-06 Reservoir Fish Genetics study plan to collect juvenile fish at spawning grounds for genetics baseline as part of field sampling program in Year 2.
- City Light will modify the study plan to expand sample collection and/or coordination of existing samples and activities and analysis out of basin and above/below dams.
- City Light will clarify the study plan to explain the role of the expert panel. The LPs and City Light agree that: (1) the expert panel will serve in an advisory role; and (2) the expert panel will include experts from fields other than genetics.
- City Light will modify the study plan to provide that City Light will seek input from LPs and advice from an expert panel on whether and how genetics information or other monitoring methods can be used to inform future evaluation of reservoir fish abundance, habitat use, and migration timing.

3.0 STUDY AREA

The study area generally includes the Project reservoirs (i.e., Gorge, Diablo and Ross lakes in the U.S.) and will include associated reservoir tributaries, as appropriate (Figure 3.0-1). Additionally, because existing data is being used and consistent with the June 9, 2021 Notice, the geographic area of the Reservoir Fish Genetics Study has expanded to include sample collection/coordination of existing samples and activities, and analysis of out of basin areas and above/below the Project dams, including below Gorge Dam (Figures 3.0-1 and 3.0-2).

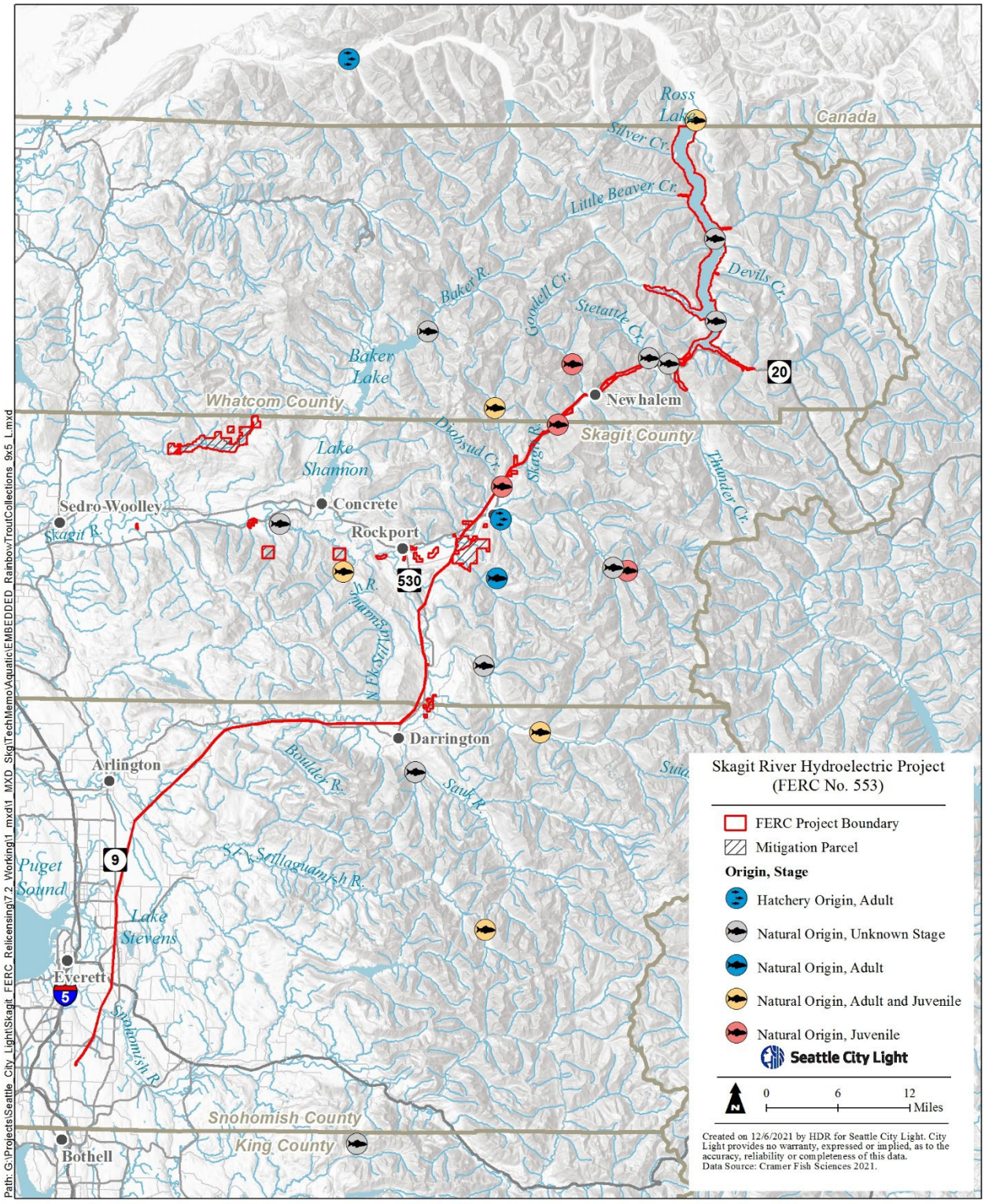

Figure 3.0-1. Proposed study area and collections evaluated for Rainbow Trout.

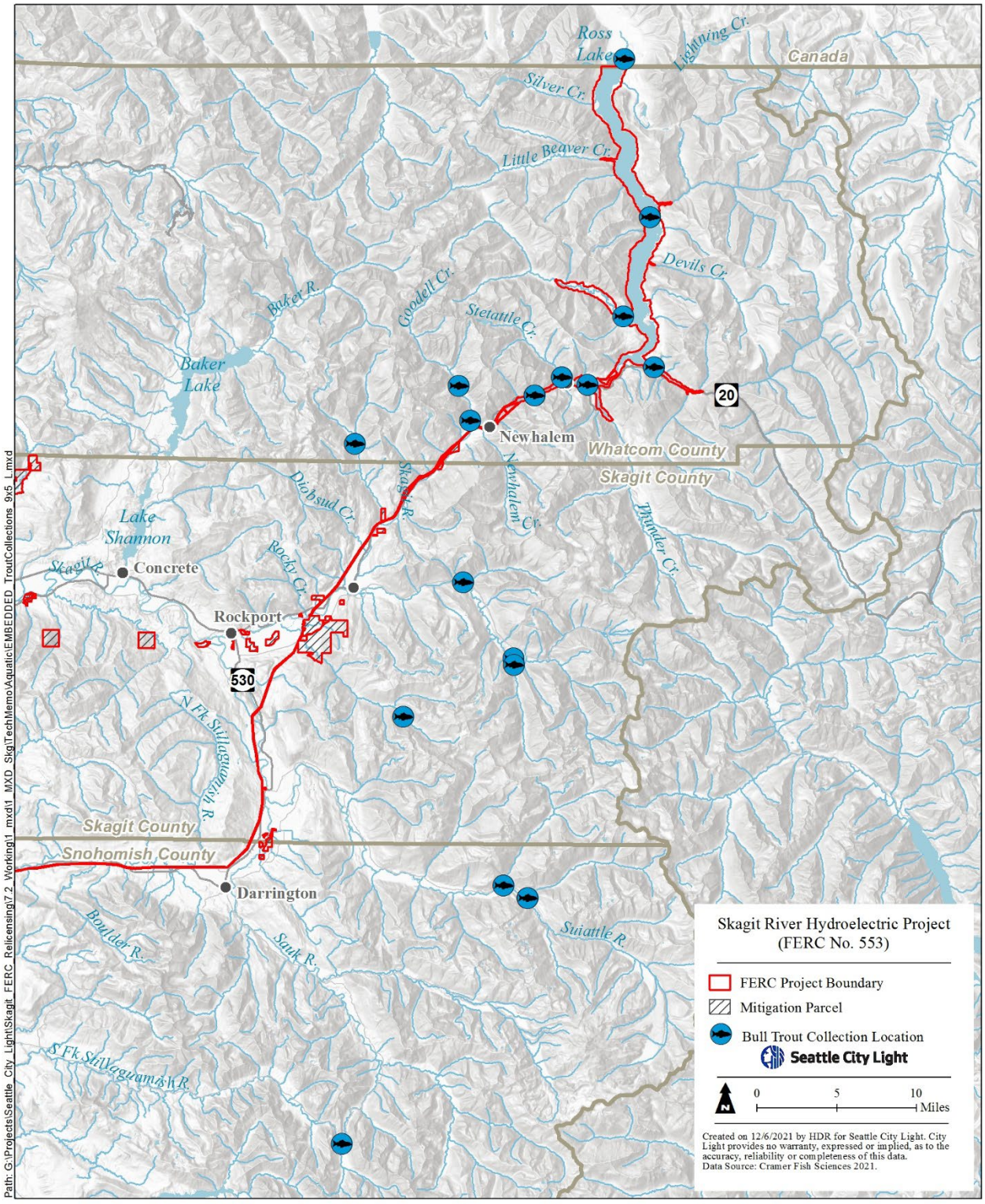

Figure 3.0-2. Proposed study area and collections evaluated for Bull Trout.

4.0 METHODS

Methods described herein apply to Year 1 objectives and are dedicated to obtaining, standardizing, and vetting existing data in consultation with the Expert Panel to evaluate the current understanding of salmonid baseline population genetics in the Project reservoirs. During Year 2, City Light will begin to fill known data gaps, as described below, and other potential gaps identified in consultation with the Expert Panel.

4.1 Data Requests

Data requests were made to obtain the previously identified and pertinent microsatellite genotypes listed in Table 2.5-1 of the RSP. The datasets contain genotypes for all three target species that were estimated in the recent past by different researchers and that may be useful for informing Year 1 objectives. To obtain those data for the purposes of this study, a request was made to the state and federal laboratories responsible for their archiving. On June 6, 2021, Cramer Fish Sciences emailed Todd Seamons, Director of the WDFW genetics laboratory, and Matt Smith, fish geneticist at the USFWS Abernathy Fish Technology Center, requesting data and metadata used in Pflug et al. (2013) and in Smith et al. (2010).

On June 13, 2021, USFWS Matt Smith provided (via email) a tab-delimited .txt file containing 563 *Salvelinus* genotypes at 16 microsatellite loci used in Smith et al. (2010): *Omm1128*, *Omm1130* (Rexroad et al. 2001); *Sco102*, *Sco105*, *Sco106*, *Sco107*, *Sco109* (WDFW unpublished); *Sco200*, *Sco202*, *Sco212*, *Sco215*, *Sco216*, *Sco218*, *Sco220* (Dehaan and Ardren 2005); *Sfo18* (Angers and Bernachez 1996); and *Smm22* (Crane et al. 2004). The dataset sent by Matt Smith included the following metadata: Individual Name, Synonym 1, Region (1), Watershed (2), Tributary (3), Capture Location (4), Age, Brood Year, Collected By, Collection Year, Comment, Date Collected, Fork Length (mm), Hatchery/Wild, HOR/NOR Assignment, Latitude, Life History, Stage, Longitude, Phenotypic Sex, PIT Tag, Population ID, Preservation Method, Project Number, Received From, Resident / Anadromous, Run Type, Spawn Date, Spawn Year, Spawned With, Species, Synonym 2, Synonym 3, Tissue Type, Total Length (mm), Used for Broodstock, and Weight (g). Only some of these metadata were relevant to this report or contained entries.

On July 28, 2021, WDFW provided (via email) an Excel spreadsheet containing 335 *Salvelinus* and 2,967 *Oncorhynchus* genotypes. The *Salvelinus* were comprised of six collections from lakes within the Project Boundary and two collections from outside the Project Boundary, with genotypes generated using the same microsatellite loci used by Smith et al. (2010). The *Oncorhynchus* genotypes were the 15 microsatellites analyzed in Pflug et al. (2013): *One-102, Ogo-4*, (Olsen et al. 1998); *Ots-100* (Nelson et al. 1998); *Oki-10*, *Oki-23* (Smith et al. 1998); *Omy-7* (K. Gharbi, unpublished, as referenced in Pflug et al. 2013*); Omy-1001*, *Omy-1011* (Spies et al. 2005); *Ots-3M*, *Ots-4* (Banks et al. 1999); *One-14* (Scribner et al. 1996); *Ssa-407, Ssa-408* (Cairney et al. 2000); *Ssa-298* (McConnell et al. 1995); and *Oke-4* (Buchholz et al. 2001). The dataset included the following metadata: Sample ID, WDFW Collection Code, Count, and Percent Missing Data. Various other metadata were available directly from the Pflug et al. (2013) and Smith (2010) reports. On September 9, 2021, Cramer Fish Sciences requested any geospatial data that could aid in identifying the specific locations that tissue samples were collected. WDFW stated that information is unavailable; if geospatial data is obtained it will be included in the USR.

4.2 Genetic Analysis

4.2.1 Rainbow Trout

The program FSTAT Version 2.9.3.1 (Goudet 1995) was used to estimate and test metrics of genetic diversity unless otherwise stated. Expected heterozygosity and allelic richness were estimated to describe genetic diversity across loci and collections. Randomization tests were performed to test the assumption of Hardy-Weinberg equilibrium (HWE) at each locus within collections. Observed (H_O) and expected (H_S) multilocus heterozygosity within populations were compared using Wright's (1951) F_{IS} to measure the magnitude of departures from HWE. To assess the assumption of random association of alleles among loci, log-likelihood ratio tests using 1,000 permutations were implemented to test for pairwise linkage disequilibrium (LD) within all collections. The Weir and Cockerham (1984) version of F_{ST} was estimated to measure genetic differentiation between all pairs of collections. A principal component analysis (PCA) of individual-based genetic distances was implemented using the R package {adegenet} (Jombart et al. 2010) to summarize the genetic diversity among the sampled individuals. The computer program POWSIM Version 4.1 (Ryman and Palm 2006) was used to estimate statistical power to detect deviation from genetic homogeneity. POWSIM is a simulation-based computer program that estimates statistical power of rejecting the null hypothesis (H_0) of genetic homogeneity for different combinations of sample sizes, number of loci, number of alleles, and allele frequencies for a hypothetical degree of true differentiation (quantified as F_{ST}). POWSIM can only accommodate 30 collections of individuals, so the first 30 collections were used to estimate power to detect low (F_{ST}=0.001) and moderate (F_{ST}=0.01) genetic differentiation by assuming allele frequencies estimated from the loci described in this report. The statistical power to observe relatives was determined using {CKMRSim} (Anderson 2019). All tests of significance were assessed at the $\alpha = 0.05$ level and applied Bonferroni corrections when conducting multiple tests.

4.2.2 Bull Trout

Exploratory analyses were conducted on Bull Trout like those described for Rainbow Trout. Partitioning of genetic variation was explored using visualization of individual-based data and genetic PCA (e.g., Jombart et al. 2010). The statistical power to observe relatives was determined using {CKMRSim} (Anderson 2019). Tests of genetic equilibrium were performed on collections. Following exploration of genetic data present in collections, summary statistics were calculated. Gene diversity (the expected frequency of heterozygotes within a population assuming HWE) was estimated following the sampling bias correction method described by Nei (1987). The observed heterozygosity (average frequency) was also estimated. A common implementation of the HWE test was used following the Guo and Thompson (1992) Markov-chain random walk extension of Fisher's (2-allele) classical exact test. Departures from HWE were also quantified using the inbreeding coefficient (F_{IS}) statistic observed from analysis of molecular variance (AMOVA) (Excoffier et al. 1992; Yang 1998), which is equivalent to Weir and Cockerham (1984) small f statistics. Collections were analyzed for evidence of LD (i.e., non-independence of alleles at different loci). Given gametic phase was unknown for previously reported data, LD between a pair of loci was tested using a likelihood-ratio test, whose empirical distribution is obtained by a permutation procedure (e.g., Excoffier and Slatkin 1998). Lastly, allelic distributions across collections were evaluated using contingency table analysis of observed allelic distributions described by Raymond and Rousset (1995).

The AMOVA framework estimates hierarchical f-statistics for any number of desired levels (e.g., within individuals, within populations, among populations). This allows for population differentiation (allele frequency variance) to be quantified. In other words, the degree that individuals within a population (collection) are more similar to each other than are individuals from different populations (collection). There are many formulations of the population differentiation variance component measure, although a common implementation is a form of the fixation index (e.g., genetic divergence [F_{ST}]). Estimates of F_{ST} were estimated pairwise following Weir and Goudet (2017) and used as a measure genetic divergence, with statistical significance calculated following likelihood-ratio tests (Goudet et al. 1996).

4.2.3 Lineage Relationships

While correlations among alleles estimated within and among populations (e.g., f statistics) attempt to account for relatedness and population genetic structure, the underlying pedigrees for sampled fish are unknown. Directly documenting relatedness among individuals is a useful measure to evaluate the genetic structure and integrity of a population over time. Parentage can determine whether fish move between reservoirs and subsequent survival, as well as gauge reproductive success within reservoirs. There are many formulations for estimating relatedness. For Bull Trout and Rainbow Trout populations, the statistically unbiased Queller and Goodnight (1989) Rxy estimator was used. The power (false detection rate) of genetic markers used to identify relationships among individual Bull Trout and Rainbow Trout (e.g., parent-offspring pairs, full-sibling-unrelated pairs) was also estimated.

4.3 Availability of Existing Samples and Coordination of Sampling with Ongoing Activities

City Light is identifying the availability of existing genetic samples from past studies (e.g., unanalyzed samples from past studies, archived samples from fieldwork in Project reservoirs, samples used in previous analyses for which a partial sample may still be available for additional analyses, etc.). City Light is also coordinating potential opportunistic sampling conducted by other relicensing studies and ongoing licensing-related field activities. These additional sampling opportunities may include, but are not limited to, the U.S. Geological Survey (USGS) Food Web Study, the Acoustic Telemetry Monitoring Program, and the FA-03 Reservoir Fish Stranding and Trapping Risk Assessment (City Light 2022). A summary of these available samples will help to identify information data gaps and inform the scope of additional data collection activities in Year 2 of the study.

5.0 PRELIMINARY RESULTS

5.1 Expert Panel

The Reservoir Fish Genetics Study is being conducted in consultation with an advisory Expert Panel. In accordance with the RSP and the June 9, 2021 Notice, City Light convened an Expert Panel composed of resource agency specialists and experts from academia with backgrounds in genetics and/or ecology. The purpose of the Expert Panel is to provide input and recommendations to inform City Light's study approach and decisions at specific milestones. Three meetings will be held with the Expert Panel throughout the study process. The first meeting was a "meet and greet" virtual gathering held in October 2021 for the Expert Panel to be introduced to the LPs.

The members of the Expert Panel are provided below:

- Hope Draheim (USFWS)
- Jason Dunham (USGS)
- Alex Fraik (NMFS Affiliate)
- Jim Meyers (NMFS)
- Meryl Mims (Virginia Tech)
- Krista Nichols (NMFS)
- Carl Ostberg (USGS)
- George Pess (NMFS)
- Todd Seamons (WDFW)
- Matt Smith (USFWS)
- Adrian Spidel (NW Indian Fisheries Commission)
- Rick Taylor (University of British Columbia)

The second Expert Panel meeting (per the RSP) occurred in January 2022 to discuss the Year 1 characterization of existing data and its sufficiency to address fish resource questions articulated by City Light and the LPs. The final Expert Panel meeting will be held in fall 2022 to discuss the results of the two-year study and recommend potential topics to be addressed in a long-term reservoir fish and aquatics management plan.

5.2 Rainbow Trout (*Oncorhynchus mykiss*)

5.2.1 Collections

In the data provided by WDFW in July 2021 (described in Section 4.1 of this study report), the microsatellites appeared to be a subset of the standardized Stevan Phelps Allele Nomenclature (SPAN) markers described in Stephenson et al. (2009) that were developed to ensure data quality (repeatable allele scoring) across laboratories. The data were provided in a Microsoft Excel spreadsheet that included Sample ID, WDFW Collection Code, Count, and Percent Missing Data. Exact sampling locations were not provided but collections appeared to be from the sites in the Skagit and Fraser river basins that are described in Pflug et al. (2013), which included tributaries,

mainstem rivers, hatcheries, and Project reservoirs. Some sites appeared to have been sampled across multiple years. The collections ranged in size from 1 in the Suiattle River in 2009 to 106 in Diablo Lake in 2005. No metadata were provided regarding sampling field methods (e.g., electrofishing), whether samples were collected randomly, or targeted life stages, life histories, morphologies, taxa, etc.

Of the 2,697 samples provided by WDFW, 536 were removed due to missing genotypes at two or more loci (e.g., Reeves et al. 2016), and 20 were removed because of duplicated genotypes. Pooling of samples from the same locations across years reduced the number of analyzed collections from 76 to 25; however, City Light retained only four of the pooled collections due to decreased deviation from HWE: Bacon Creek (2007 to 2010), Clear Creek (2009 and 2010), Blackwater River (2009 and 2010), and the Suiattle River (1981 and 2009 to 2011). Putative siblings from the *O. mykiss* dataset were not omitted because multiple age classes appeared to have been sampled and doing so could reduce precision of analyses as cautioned by Waples and Anderson (2017) (but see analyses of *Salvelinus* spp.). Collections with fewer than 25 individuals were removed to avoid biased estimates of allele frequencies within sub-populations (Hale et al. 2012). Data were not sufficient to describe hybridization with *O. clarkii* because the submitted spreadsheet from WDFW did not contain known nonhybridized *O. clarkii* genotypes to use as positive controls for estimating taxon-diagnostic allele size distributions. The final dataset contained 1,900 individuals from 40 collection events but only 38 were analyzed due to possible hybridization with *O. clarkii* that was not apparent until most analyses were completed. The genotypes are available upon request in GENEPOP format (Rousset 2020).

5.2.2 Genetic Summary Statistics

Comparison of observed (H_o=0.729) and expected (H_S=0.747) heterozygosity across all collections and loci suggested a relatively small but overall deficit of heterozygotes (F_{IS}=0.025 95 percent confidence interval [CI]; 0.01, 0.03). Eighty-six of 600 (14 percent) randomization tests for HWE (15 markers x 40 collections) using FSTAT (Goudet 1995) were significant at the α=0.05 level with 68 (79 percent) of the tests showing a deficit of heterozygotes. No tests for HWE were significant at the adjusted level of α=0.00008. The locus *One-14* deviated from HWE in 17 of 40 (42.50 percent) total collections with all tests showing a deficit of heterozygotes. By contrast, most other markers (11 of 15) produced various combinations of heterozygote excess and deficiency. Therefore, the locus *One-14* was omitted from further analysis due to the possibility of genotyping problems. This adjustment decreased mean F_{IS} to 0.017, though the difference was not statistically significant (95 percent CI: 0.03, 0.01). The remaining 14 microsatellite loci had a total of 312 alleles, ranging from 11 at *Ots-4* to 32 at *Omy-1001*. Across all 14 loci and 40 collections, the estimated false detection rate of a parent-offspring pair was 0.00000811, 0.00000033 of full siblings, and 7.277×10^{-21} of unrelated individuals. However, within any single collection, power is expected to be substantially lower. For example, the false positive rate (FPR) for related individuals in Roland Creek, a tributary within the Project Boundary, is 0.0000161 and the false negative rate (FNR) is 0.392 (Figure 5.2-1). Gene diversity (H_S) within each collection ranged from 0.36 in the collection from North Fork Cascade River in 2010 to 0.83 in the Baker River in 2010 (Table 5.2-1). Average gene diversity in collections from upstream of the Project Boundary at Gorge Lake (H_S=0.74) was similar diversity in all other collections (H_S=0.74).

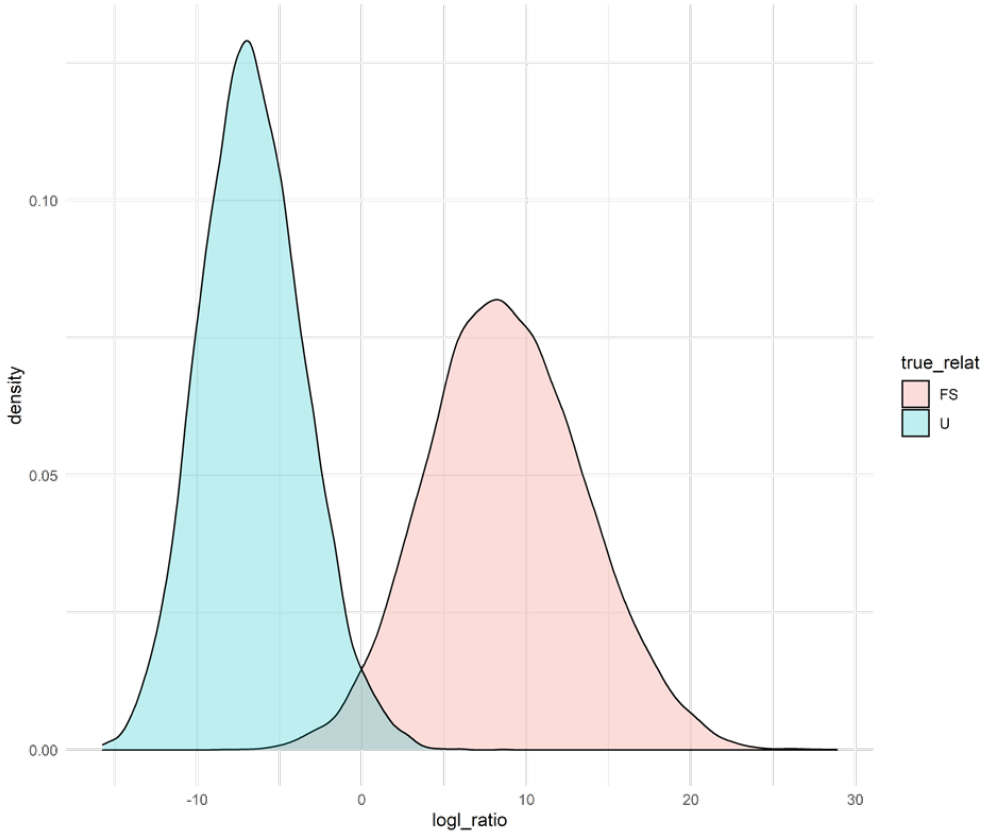

Figure 5.2-1. **Log-likelihood ratios distribution for simulated true full-siblings versus unrelated individuals based on Roland Creek *O. mykiss* genotype data. High overlap between full-siblings and unrelated fish suggests relatively low power to detect highly related individuals.**

Table 5.2-1. **Summary statistics for samples collected from *O. mykiss* in the Skagit and Fraser river basins.**

Collection number [1]	Collection size	WDFW Code [2]	Location	Origin[3]	Upper Skagit[4]	Stage	Phenotype[5]	F_{IS} [6]	H_S [7]	A_R [8]	R^2 [9]
1	57	07MS, 08MI, 10BA	Bacon Creek	NOR	No	Juvenile, adult		0.01	0.79	9.45	0.02
X	57	09EL	Baker River 09	NOR	No		Trout	0.09	0.82	10.44	0.03
X	42	10AU	Baker River 10	NOR	No		Trout	0.11	0.84	11.13	0.04
2	51	09EU	Big Creek 09	NOR	No		Trout	0.04	0.66	5.63	0.02
3	48	10BG	Big Creek 10	NOR	No		Trout	0.06	0.67	5.16	0.03
4	52	09JB, 10BJ	Blackwater River	NOR	No	Juvenile	Trout	0.11	0.74	7.61	0.02
5	66	10MZ	Chilliwack Hatchery	HOR	No	Adult		0.00	0.76	8.02	0.02
6	94	09ET, 10BE	Clear Creek	NOR	No		Trout	0.06	0.68	8.56	0.01

Collection number [1]	Collection size	WDFW Code [2]	Location	Origin[3]	Upper Skagit[4]	Stage	Phenotype[5]	F_{IS} [6]	H_S [7]	A_R [8]	R^2 [9]
7	38	10BB	County Line Ponds	NOR	NO	Juvenile		0.01	0.80	9.10	0.04
8	26	05NG	Diablo	NOR	Yes		Trout	0.06	0.75	8.50	0.04
9	41	10BK	Diobsud	NOR	No	Juvenile		0.02	0.79	9.77	0.03
10	43	03OA	Dry Creek	NOR	Yes		Trout	0.02	0.71	7.50	0.03
11	47	09EH	Finney Creek	NOR	No	Juvenile		0.00	0.78	9.27	0.03
12	47	10AT	Finney Creek	NOR	No	Juvenile		0.01	0.80	9.59	0.02
13	30	11BK	Finney Creek	NOR	No	Adult		-0.02	0.80	10.40	0.04
14	38	09IZ	Goodell Creek	NOR	No	Juvenile		0.01	0.77	8.76	0.03
15	41	10BC	Goodell Creek	NOR	No	Juvenile		0.00	0.79	9.12	0.03
16	47	09EE	lower Cascade	NOR	No	Juvenile		-0.05	0.77	8.23	0.03
17	44	10AV	lower Cascade	NOR	No	Juvenile		0.03	0.79	9.26	0.03
18	48	10AY	lower Skagit	NOR	No	Juvenile		0.02	0.79	9.51	0.03
19	28	08LF	lower Skagit	NOR	No	Adult		0.01	0.78	9.26	0.04
20	59	09CF	Marblemount	HOR	No	Adult		0.01	0.82	9.68	0.02
21	44	10AN	Marblemount	HOR	No	Adult		0.03	0.79	8.89	0.03
22	39	09BM	mid Skagit	NOR	No	Adult		0.01	0.80	10.49	0.03
23	31	10AS	mid Skagit	NOR	No	Adult		0.04	0.80	10.14	0.03
24	47	09ES	NF Cascade	NOR	No		Trout	0.11	0.41	4.30	0.02
25	45	10BF	NF Cascade	NOR	No		Trout	-0.08	0.36	3.98	0.02
26	79	02FB	Roland Creek	NOR	Yes		Trout	0.01	0.71	7.68	0.01
27	30	06AF	Ross	NOR	Yes		Trout	0.03	0.73	8.20	0.04
28	44	09MA	Ross	NOR	Yes		Trout	-0.01	0.69	6.65	0.04
29	47	10BH	Ross	NOR	Yes		Trout	-0.03	0.70	6.40	0.04
30	45	10AX	Sauk	NOR	No	Juvenile		0.04	0.80	9.66	0.03
31	29	83AAA	Sauk	NOR	No	Adult		0.06	0.80	10.29	0.04
32	32	09JA	Stetattle	NOR	Yes		Trout	0.03	0.76	8.66	0.04
33	41	10BI	Stetattle	NOR	Yes		Trout	0.03	0.77	8.79	0.03
34	115	09DT, 09EF, 10AQ, 10AW, 11BM	Suiattle	NOR	No	Juvenile, adult		0.01	0.79	10.05	0.01
35	51	09EV	upper Finney	NOR	No		Trout	0.03	0.74	6.52	0.02
36	49	10BD	upper Finney	NOR	No		Trout	0.04	0.72	6.77	0.02
37	56	10AZ	upper Skagit	NOR	No	Juvenile		0.01	0.79	9.56	0.02
38	32	11BI	upper Skagit	NOR	No	Adult		0.00	0.81	10.43	0.03

1 Collection number: corresponds to Figures 5.1-3 through 5.1-5.
2 WDFW code: WDFW collection identification with apparent sample year as the prefix.
3 Origin: hatchery (HOR) or natural (NOR) origin.
4 Upper Skagit: collections from upstream of the Project Boundary in the Skagit River and from B.C.

5 Phenotype: identifies whether collections were from apparent trout as determined by WDFW.
6 F_{IS}: estimated deviation from HWE.
7 H_S: estimated expected heterozygosity within sub-populations (i.e., gene diversity).
8 A_R: estimated allelic richness.
9 R^2: is the estimated pairwise correlation of alleles among loci.

Six-hundred-forty of 3,640 (17.5 percent) log-likelihood (G) tests for pairwise LD using FSTAT were significant at the α=0.05 level. However, only 15 (<1 percent) tests were significant at the adjusted table-wide level of α=0.00007. The greatest disequilibrium was observed in the collection from Diablo Lake in 2005 (R^2=0.04) and the least in Suiattle River (R^2=0.01) (Table 5-2.1). Notably, there was a consistent, negative relationship between sample size and R^2 (R^2=0.796), which is the estimator for pairwise LD (Figure 5-2.2), suggesting cautious interpretation of LD analysis is warranted.

Fisher's exact tests using POWSIM (Ryman and Palm 2006) which were based on sample sizes and estimated allele frequencies of the dataset, suggested power to detect deviation from genetic homogeneity was 0.32 for F_{ST}=0.001 and was 1.00 for F_{ST}=0.01. The overall estimated proportion of genetic variance explained by population structure (F_{ST}) was 0.094. Log-likelihood (G) tests for population differentiation were significant for each locus and across all loci (P<0.001). Estimates of pairwise F_{ST} ranged from -0.004 between collections from Stetattle Creek in 2009 and 2010 to 0.39 between collections from Ross Lake and North Fork Cascade.

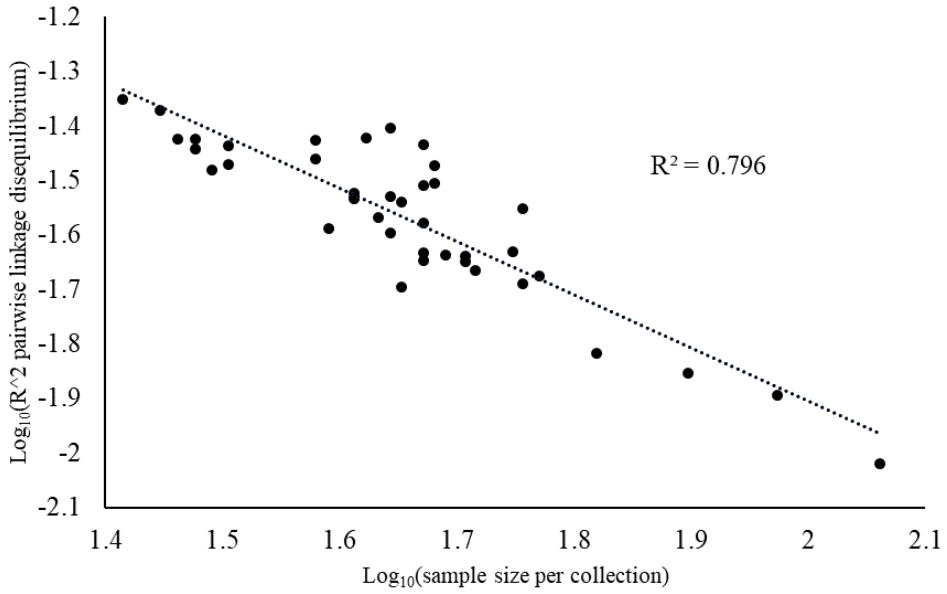

Figure 5.2-2. Scatterplot showing log_{10}-transformed relationship between sample size (n) (x-axis) and the R^2 estimator for pairwise linkage disequilibrium between loci (y-axis). Strong correlation warrants cautious interpretation of data due to possible bias.

PCA of individual-based genetic distances using {adegenet} (Jombart 2011) accounted for a relatively small amount of projected inertia—a metric of the magnitude of the explained genetic variance among individuals (cumulative inertia explained by PC-1 through 3=5.924 percent).

Genetic population structuring was apparent in scatterplots of the first three principal components (PC). However, several samples from the Baker River collections appeared to be outliers along axes 1 and 2. Notes provided by WDFW suggested the samples could be hybrids with *O. clarkii*. Reanalysis without the Baker River collections only slightly improved projected inertia of the first three PCs (6.095 percent); however, it did improve visualization of genetic population structure (Figure 5.2-3). Specifically, PC-1 (2.215 percent) clearly distinguished the North Fork Cascade River (Collections 24 and 25) from all other collections. PC-2 (2.044 percent) highlighted additional population structuring with collections from upstream of the Project boundary tending to display positive inertia, collections from the Sauk River basin tending to display negative inertia, and remaining collections falling in between. PC-3 (1.836 percent) nearly distinguished Big Creek (Collections 2 and 3) from all other collections (Figure 5.2-4).

Limiting PCA to collections from upstream of the Project Boundary at Gorge Lake identified three samples that might be hybrids between *O. mykiss* and *O. clarkii* based on notes from WDFW; they were subsequently removed from the analysis (09JA0030, 05NG0056, and 10BI0047). Reanalysis without the potential hybrid samples indicated that the first three PCs explained 5.898 percent of the total inertia (Figure 5.2-5) and appeared to support some genetic structuring associated with location but statistical support for individual genetic groups was low.

Effective population size (N_e) of *O. mykiss* was not estimated in the Project reservoirs because of sampling considerations. Firstly, hybridization with *O. clarkii* could bias estimates of N_e by creating genetic disequilibria that is not associated with genetic drift. Secondly, estimating N_e in an iteroparous species with overlapping generations requires extensive sampling effort and significant data on life stage specific survival and reproduction. Though it is common to estimate effective number of breeders (N_b), unbiased estimates typically call for sampling of individuals of the same cohort or across multiple generations, and such data was not available.

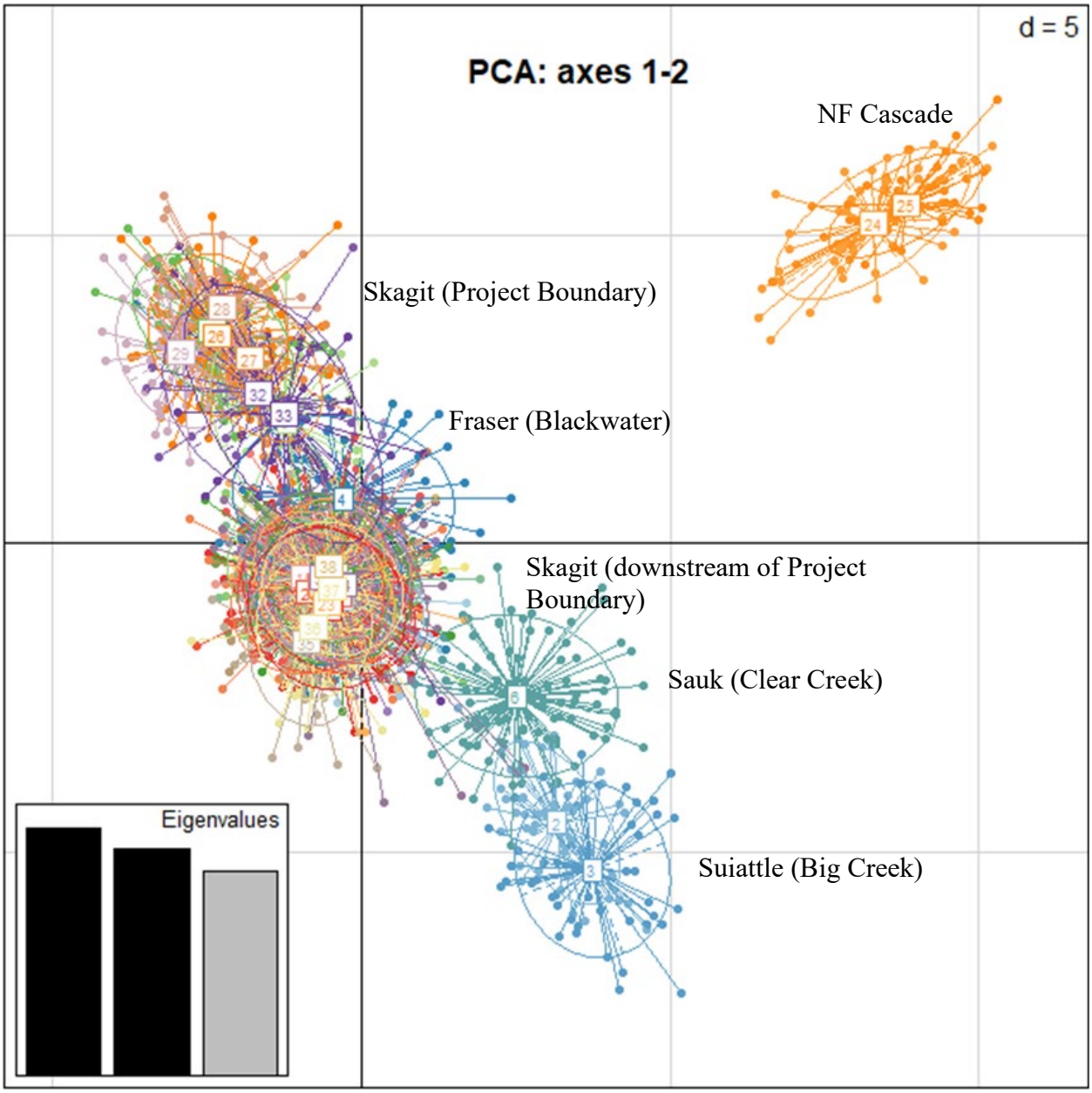

Figure 5.2-3. **Scatterplot of genetic PC-1 (2.215 percent) and PC-2 (2.044 percent) for all collections, excluding samples from the Baker River, estimated using adegenet in program R (Jombart 2011). The distribution of genetic variation appears to support some genetic structuring associated with the geographical locations of collections. River basins are provided to show the approximate geographical locations of each collection. Numbers at centroids identify the collection number listed in Table 5.2-1. Ellipses define 1.5 standard deviations of the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom left corner shows first three eigenvalues.**

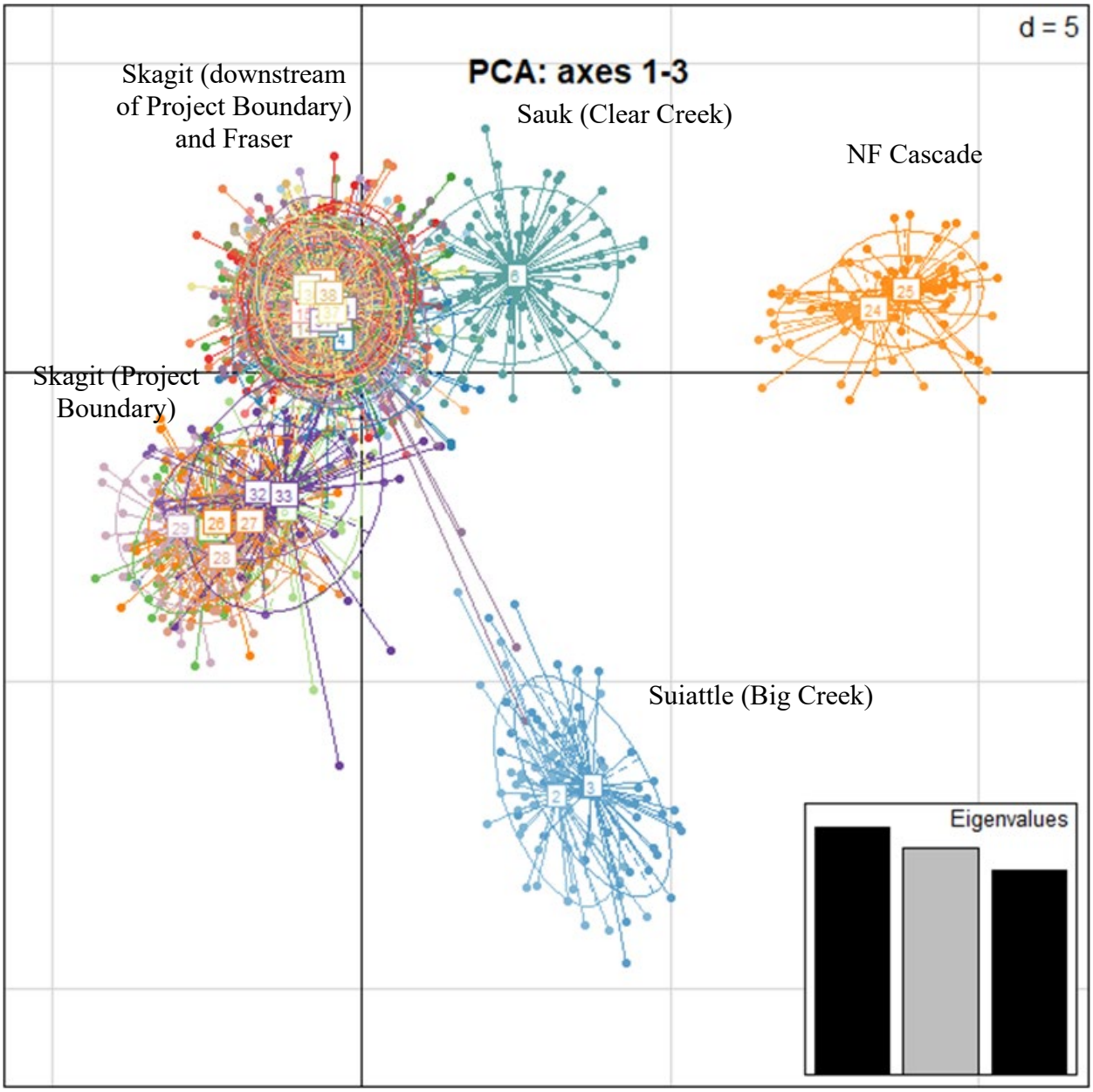

Figure 5.2-4. **Scatterplot of genetic PC-1 (2.215 percent) and PC-3 (1.836 percent) for all collections, excluding samples from the Baker River, estimated using adegenet in program R. The distribution of genetic variation appears to support existence of genetic structuring associated with the geographical locations of collections. River basin names are provided to describe the approximate geographical locations of each collection. Numbers at centroids identify the collection number listed in Table 5.2-1. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom right corner shows first three eigenvalues.**

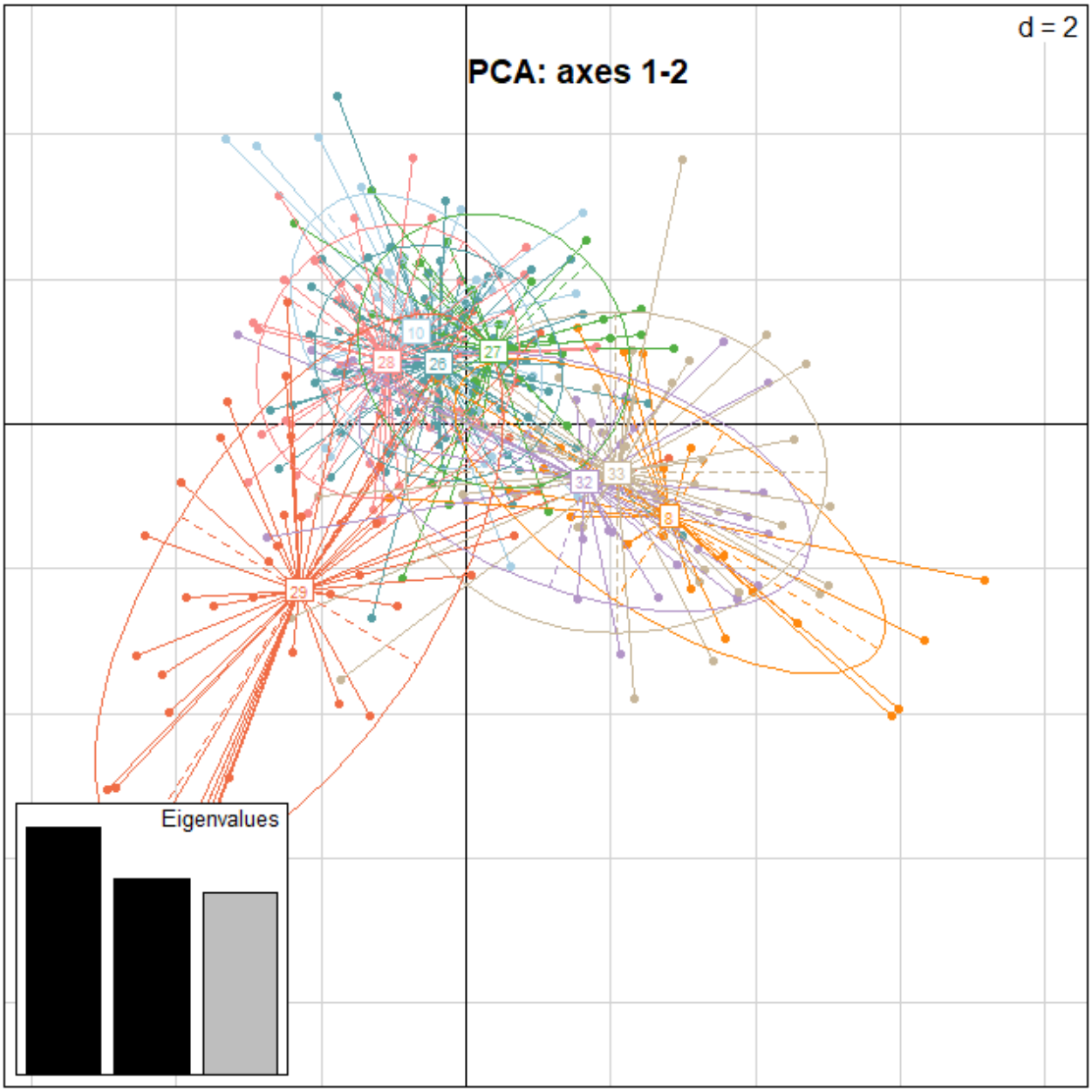

Figure 5.2-5. **Scatterplot of genetic PC-1 (3.870 percent) and PC-2 (2.028 percent) for all collections upstream of the Gorge Lake Project Boundary estimated using adegenet in program R. Numbers at centroids identify the collection number listed in Table 5.2-1. Scree plot in bottom right corner shows first three eigenvalues. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom left corner shows first three eigenvalues.**

5.3 Bull Trout (*Salvelinus confluentus*)

5.3.1 Collections

Eight hundred and ninety-eight *Salvelinus* spp. genotypes at 16 microsatellite loci were provided by USFWS and WDFW following a request for existing Bull Trout data within the Skagit River basin (Table 5.3-1). USFWS provided 563 of the genotypes and WDFW provided 335. The standardized markers included *Omm1128*, *Omm1130* (Rexroad et al. 2001); *Sco102*, *Sco105*, *Sco106*, *Sco107*, *Sco109* (WDFW unpublished); *Sco200*, *Sco202*, *Sco212*, *Sco215*, *Sco216*, *Sco218*, *Sco220* (Dehaan and Ardren 2005); *Sfo18* (Angers and Bernachez 1996); and *Smm22* (Crane et al. 2004). The collections were from four Project vicinity tributaries (upper Skagit, Big Beaver, Ruby, and Stetattle creeks) and all three reservoirs (Ross, Diablo, and Gorge lakes). It was unclear which *Salvelinus* spp. taxa or their hybrids were included in the dataset. It was also unclear to what extent collections comprised highly related individuals, which is a common concern in genetic studies of Bull Trout (DeHaan et al. 2014). Furthermore, USFWS communicated that the juvenile collections likely contained related individuals (M. Smith, personal communication).

Sampling location metadata were not provided for USFWS samples, so sampling locations were assumed to be the same as reported in Smith (2010). The stated purpose of the collections from Smith (2010) was to assess genetic variability within and between Bull Trout populations, with sampling methods including a combination of electrofishing, snorkeling, and angling.

No metadata were provided by WDFW other than collection code. Location data were not provided, so samples obtained from within the Project Boundary were considered "at-large" from reservoirs. The stated purpose of WDFW collections was to characterize the genetic variation of Bull Trout, Dolly Varden, and Brook Trout in the Skagit reservoirs, but no collection methodology was described. The degree to which samples were collected randomly across *Salvelinus* spp. taxa was unknown, including whether any special effort was made to target Bull Trout, Dolly Varden, Brook Trout, or whether potential hybrids were targeted or avoided. Sampling considerations are a key concern because targeted collections (i.e., based on morphology) can bias inference into studies of genetic variation.

City Light conducted quality assurance/quality control procedures to obtain a final dataset in which basic population genetic analyses could be reasonably implemented. Duplicate genotypes were observed for sample IDs 12FG008 and 12FG0009, and so sample 12FG0009 was omitted from dataset. City Light removed all individuals with missing genotypes at three or more loci, which is more than the 14 loci chosen for *O. mykiss* (see above). This was necessary, however, because the Bull Trout data production appeared to have been conducted in four by four-locus panels (i.e., multiplexes), with many samples missing a single four locus block. Following data quality assurance/quality control, 589 samples were retained for analysis (Table 5.3-1).

Table 5.3-1. Bull Trout microsatellite dataset collection summary.

Collection Location	River	Life Stage	WDFW Code	Number Collected	Number Evaluated	Number Analyzed
Upper Skagit River	Skagit	adult		16	14	14
Big Beaver Creek	Skagit	adult		21	21	21
Ruby Creek	Skagit	adult		43	41	41
Stetattle Creek	Skagit	juvenile		59	41	41
Lower Goodell Creek	Goodell	juvenile		60	46	46
Upper Goodell Creek	Goodell	juvenile		19	8	8
Bacon Creek	Bacon	juvenile		61	24	24
Cascade River	Cascade	juvenile		39	33	33
Marble Creek	Cascade	juvenile		28	18	18
Kindy Creek	Cascade	juvenile		30	17	17
Illabot Creek	Illabot	juvenile		70	60	60
South Fork Sauk River	Sauk	juvenile		59	54	54
Downey Creek	Sauk	juvenile		58	44	44
Ross Lake	Skagit	unk	12FG	54	47	42
Ross Lake	Skagit	unk	15OW	28	22	20
Diablo Lake	Skagit	unk	13PS	40	29	8
Gorge Lake	Skagit	unk	14ST	27	5	3
Gorge Lake	Skagit	unk	19NL	109	22	0
Sulfur	Skagit	unk	050F	4	4	4
Sulfur	Skagit	unk	06JQ	28	23	23
Diablo, Gorge Lake	Skagit	unk	11LX	45	16	9
			Total	**898**	**589**	**530**

5.3.2 Identification of related individuals within collections

Statistical power was estimated to correctly classify related individuals. This was completed to evaluate the possible effects of violations of sampling assumptions common to the analysis of Bull Trout microsatellite data; specifically, that highly related individuals (i.e., full siblings) are common in samples of Bull Trout (particularly samples of juveniles), which can result in pseudo-replication of genotypes and thus biased estimates of allele frequencies (DeHaan et al. 2014). Statistical power of pedigree analysis to identify parent-offspring and full-sibling pairs was conducted using the close kin mark recapture (CKMR) R package CKMRSim version 0.1 (Anderson 2019; formerly NMFS Southwest Fisheries Science Center). During pedigree analysis, all samples are examined for relatedness in pairwise comparisons, and so the FPR increases exponentially with sample size. It is recommended to choose a FPR threshold approximately 10 times smaller than the reciprocal number of pairwise comparisons. In this case, $1.4\ e^{-5}$ was the target FPR used to evaluate the power to detect relatives (i.e., $0.10 * (100 \times 100)^{-1} = 0.000014$). To simulate the related and unrelated individuals needed to estimate power of pedigree analysis, all collections from the Skagit River dataset were used. The distribution of log-of-the-odds (LOD) values are shown in Figure 5.3-1 for full-sibling pairs. The expected distributions overlap between full-sibling and unrelated individuals, which means that choosing a FPR that provides reasonable

assurance no unrelated pairs will be falsely called full-siblings will result in an undesirably high FNR. For Skagit River Bull Trout, a LOD value = 8.0 (corresponding to FPR = 1.4 e^{-5}) results in a FNR = 0.15, meaning approximately 15 percent of true full-sibling comparisons would be misclassified as unrelated with an α=0.05 as the typical standard.

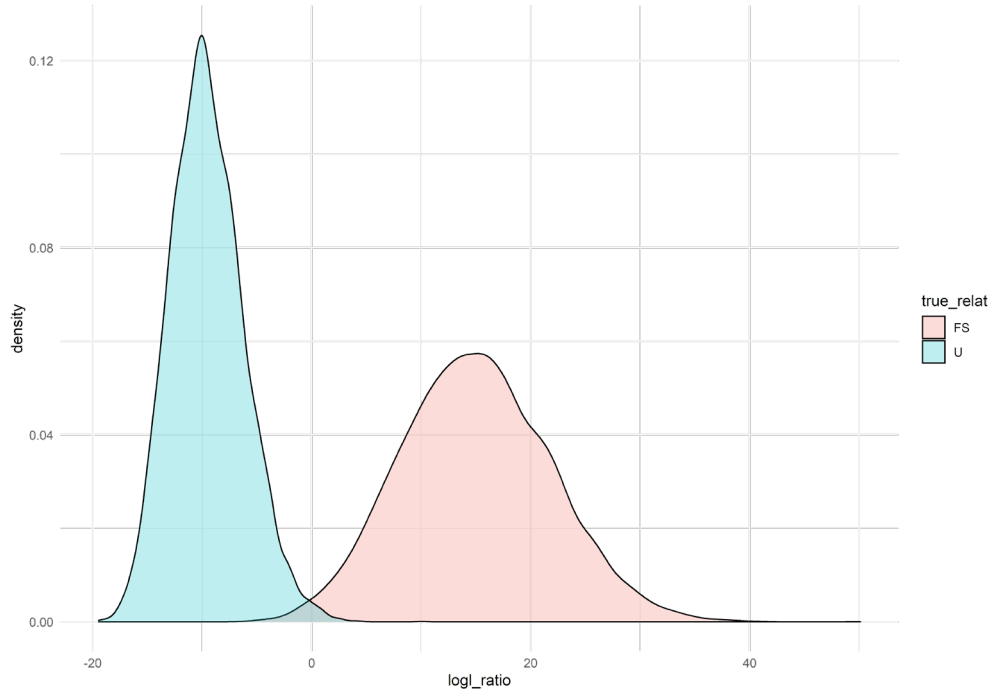

Figure 5.3-1. Log likelihood ratios distribution for simulated true full siblings versus unrelated individuals based on Skagit River *S. confluentus* genotype data. High overlap between full-siblings and unrelated fish suggests relatively low power to detect highly related individuals.

Note that these estimated rates were based on all the available collections (n=530), which would likely overestimate power for studies of "real-world" populations. A more realistic evaluation would consider collections from a single Project Boundary tributary, as opposed to considering potential comparisons between unrelated individuals across the entire Skagit River basin. City Light therefore repeated the analysis, using only collections from Big Beaver, Ruby, and Stetattle creeks in the Project Boundary. The FNR estimated for Big Beaver, Ruby, and Stetattle collections were 0.857, 0.868, and 0.95, respectively, meaning pedigree analysis is expected to result in more false relationship assignments than true assignments.

Understanding power to detect related individuals helped identify individual samples that might need to be removed from analysis to reduce violation of sampling assumptions. COLONY (Jones and Wang 2010) was used to screen collections for full sibling families, and based on power estimates above, applied probability of inclusion = 1.0 and a probability of exclusion = 0.99 to accept family classifications. Inclusion probability gives the probability that all individuals (in that family) are indeed full siblings from the same family. Exclusion probability is the probability those individuals are full siblings, and no other individuals are full siblings with this family. There is no accepted convention or criterion for identifying and removing related individuals from a dataset,

although the criteria used here are more stringent than those referenced in literature pertaining to this Bull Trout dataset (Smith 2010). All full siblings but one[2] were omitted from identified families within the collection.

5.3.3 Population Determination

Like for Rainbow Trout, PCA of allele frequencies (adegenet package) was used to examine genetic variation among collections. Data modeling suggested retention of approximately 15 PCs and 5 discriminant functions (*k*) would result in reliable partitioning of genetic variation among group clusters. With the number of genetic group clusters fixed at two (i.e., *k*=2), samples partitioned into genetic groupings associated with Diablo/Gorge Lakes and all other samples. With an additional cluster allowed (*k*=3), individuals partitioned into (1) Project Boundary tributaries and some reservoir samples; (2) Project Boundary reservoir samples; and (3) samples from below Gorge Dam. With the allowance of fourth and fifth genetic clusters (*k*=4 and *k*=5), Project Boundary reservoir samples became split among the newly allowed clusters. No further refinement of Project Boundary samples was observed at higher numbers of clusters. A visualization of the k-means clustering at k=5 is shown on Figure 5.3-2. Clusters 3, 4 and 5 were predominantly individuals collected from Diablo and Gorge lakes. Cluster 1 were Project Boundary tributary collections, constituting of a majority of Ross Lake samples. Cluster 2 were individuals collected from below Gorge Dam.

As mentioned, collections submitted by WDFW were a part of evaluations intended to assess hybridization among Bull Trout, Dolly Varden and Brook Trout. Reports pertaining to data noted that hybrids were observed within these collections (e.g., Small et al. 2013; Small et al. 2016). City Light was unable to directly ascribe clusters 3, 4, and 5 to hybridization among individuals or genetic introgression because (1) taxon-diagnostic alleles among taxa were unknown; (2) sample IDs for individuals WDFW considered hybrids were not provided; (3) the methods by which WDFW determined individuals to be hybrids was not provided; and (4) the selection strategy (if any) of field personnel collecting individuals "at large" from reservoirs was also not provided.

[2] The presence of multiple representatives from the same family skews allele frequencies from true population proportions, creating a bias. Removing all but one sibling removes this bias.

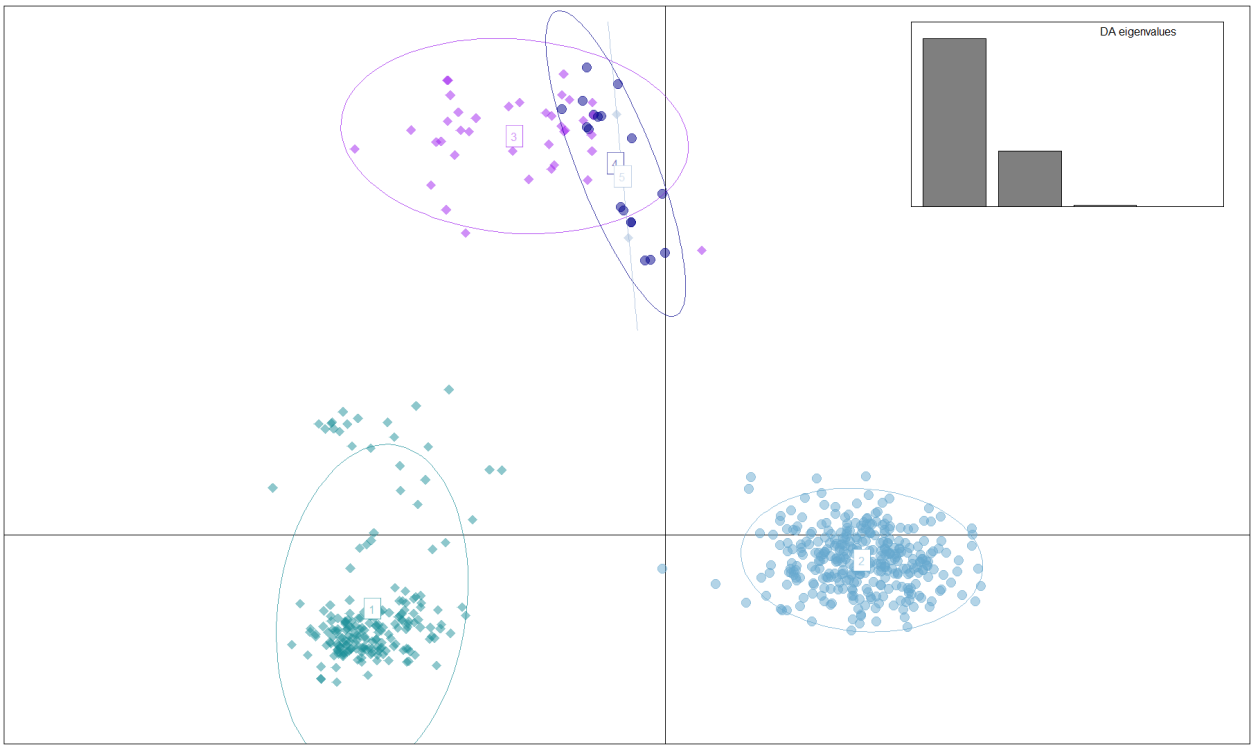

Figure 5.3-2. **Visualization of k-means clustering analysis at k=5 for Bull Trout individuals in dataset for 1st and 2nd principal axes. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in upper right corner shows first three eigenvalues. Cluster 1 were Project Boundary tributary collections and contained a majority of Ross Lake samples. Cluster 2 were individuals collected from below Gorge Dam. Clusters 3, 4 and 5 were predominantly individuals collected from Diablo and Gorge Lakes.**

Small sample sizes of *Salvelinus* spp. (median=26) relative to *O. mykiss* (median=45) highlighted limitations associated with balancing precision and bias. For instance, collections with fewer than 25 individuals are typically not recommended for analyses using microsatellite data, however, adopting this criterion for the *Salvelinus* spp. dataset would have resulted in exclusion of about 50 percent of Project Bull Trout collections from an already sparse dataset. The genetic groupings shown in Figure 5.3-3 also underscore the challenges associated with choosing which fish to retain in any given collection due to genetic admixture. All individuals in clusters 3, 4, and 5 were considered potentially admixed and omitted from the dataset prior to estimating genetic summary statistics for each collection. The current sample size threshold pertaining to Bull Trout collections may be modified based upon future discussions of hypotheses and research questions with the Expert Panel and LPs. The resulting final dataset comprised n=530 samples (Table 5.3-1). The genotypes are saved in GENEPOP format and are available upon request.

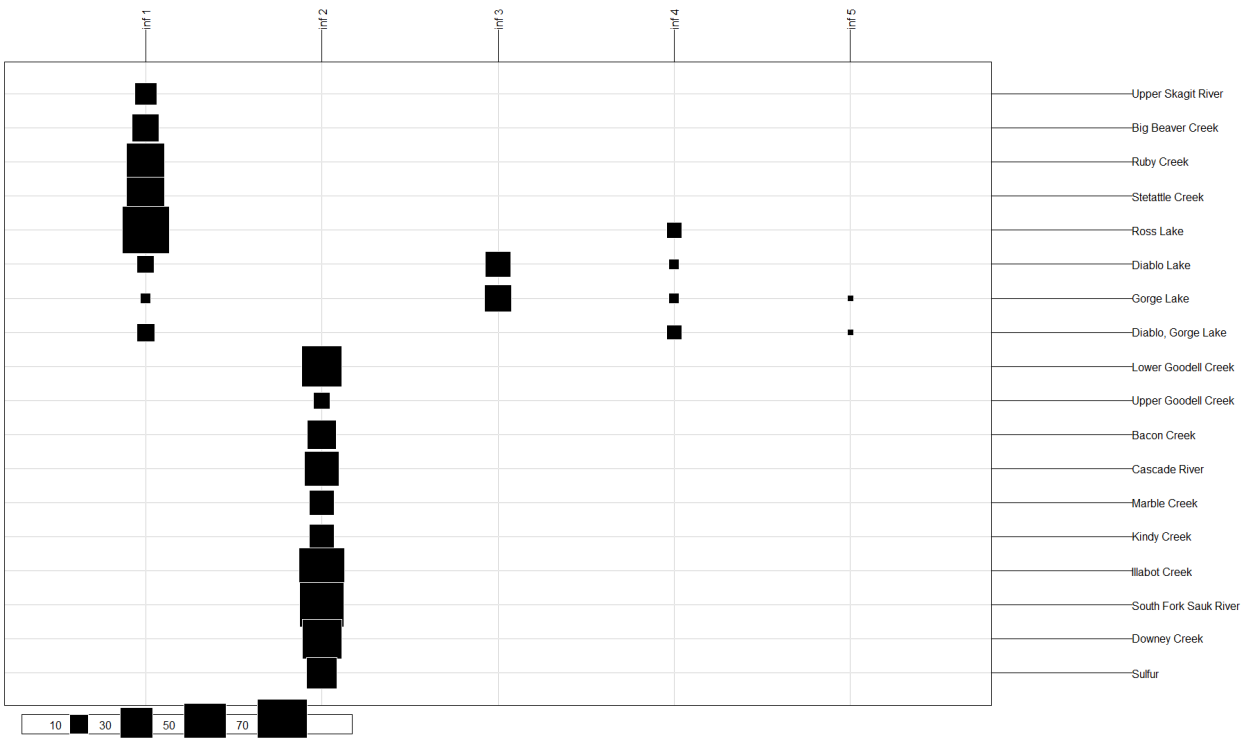

Figure 5.3-3. Genetic clusters visualized in Figure 5.3-2 aligned to each Bull Trout collection in dataset. Size of boxes is scaled by sample count. Genetic clusters are organized by geographic location with upper Skagit collections at the top and lower Skagit at the bottom.

5.3.4 Genetic summary statistics

Heterozygosity in the Bull Trout collections ranged from 0.337 to 0.467 within collections from Project Boundary tributaries (above Gorge Dam) and was 0.473 in the Ross Lake collection (Table 5.3-2). The collections from within the Project Boundary (above Gorge Dam) had lower heterozygosity than the collections from below Gorge Dam (Chi-square p-value = 0.0027). Our attempt to reduce violation of HWE appeared successful, as mean F_{IS} across all collections was not statistically different from 0.00 (F_{IS}=0.008, 95 percent CI: -0.024-0.051). Each Project vicinity tributary collection (upper Skagit, Big Beaver, Ruby, Stetattle) did not deviate significantly from expectations. The Ross Lake collection was not in HWE, along with potentially several collections from below Gorge Dam, particularly Bacon Creek and Illabot Creek. Potential cause(s) of observed HWE deviations (e.g., data quality, inbreeding, population mixing) have not yet been determined. We measured LD using log-likelihood (G) tests for all pairwise locus comparisons. Of the 1,680 comparisons (overall collections), 271 were significant at the α=0.05 level. No Project Boundary tributary collections (above Gorge Dam) had statistically significant LD tests using the adjusted table wide significance level α=0.0003. The Ross Lake collection had 11 significant LD tests out of 120. The greatest number of significant log-likelihood tests was observed for the Illabot Creek collection (16).

The estimated proportion of genetic variance explained by population structure (F_{ST}) across all Bull Trout collections was 0.188, and 0.03 among Project Boundary tributary collections, only. Pairwise log-likelihood (G) tests for population differentiation were not statistically significant

between the upper Skagit, Big Beaver, and Ruby Creek collections (adjusted nominal level 5 percent). The Upper Skagit River collection was not differentiated from the Ross Lake collection, but the Ross Lake collection was differentiated from both the Big Beaver and Ruby Creek collections. The Stetattle Creek collection was differentiated from all other Project Boundary collections. Note that the Marble Creek collection was not differentiated from any collection in the dataset except South Fork Sauk River. This seemed anomalous, so results that follow exclude consideration of Marble Creek collection. All Project Boundary collections (above Gorge Dam) were differentiated from below Gorge Dam collections. Recall, F_{ST} is the proportion of genetic variation that is attributable to population subdivision with F_{ST}=0.00 reflecting no differences and F_{ST}=1.00 reflecting complete differentiation (i.e., all genetic diversity is partitioned among subpopulations). The F_{ST} estimated (pairwise) between the Project Boundary collections are shown in Table 5.3-3). For context, F_{ST} estimated from comparisons between the Project Boundary collections with those from below Gorge Dam ranged from a low of 0.207 to a high of 0.397.

Table 5.3-2. Summary statistics for samples collected from Bull Trout in the Skagit River basin.

Collection	Sample Size	F_{IS}[1]	H_s[2]	MNA[3]
Upper Skagit River	14	0.080	0.467	5.00
Big Beaver Creek	21	0.042	0.410	4.44
Ruby Creek	41	-0.021	0.384	4.75
Ross Lake	62	0.105	0.473	7.16
Stetattle Creek	41	-0.078	0.337	2.94
Goodell Creek	54	0.046	0.647	6.97
Bacon Creek	24	0.038	0.678	7.56
Illabot Creek	60	-0.050	0.634	7.44
Cascade River	33	0.033	0.662	8.19
Marble Creek	18	-0.080	0.679	6.94
Kindy Creek	17	0.016	0.689	7.19
S.F. Sauk River	54	-0.032	0.656	8.31
Downey Creek	44	0.010	0.709	9.88
Sulfur	27	0.035	0.607	6.13

1 $F_{IS:}$ estimated deviation from Hardy-Weinberg proportions.
2 $H_{S:}$ estimated expected heterozygosity within sub-populations (i.e., gene diversity).
3 MNA: is the mean number of alleles observed over all loci.

Table 5.3-3. Table of pairwise estimates of F_{ST} from Project Bull Trout collections.

	Upper Skagit River	Big Beaver Creek	Ruby Creek	Ross Lake
Big Beaver Creek	0.001			
Ruby Creek	0.028	0.014		
Ross Lake	0.023	0.043	0.061	
Stetattle Creek	0.068	0.030	0.034	0.105

6.0 SUMMARY

6.1 Summary of Completed Objectives

This section provides a summary of completed objectives for Year 1.

- City Light convened an Expert Panel. Panel members are identified in Section 5.1 above.
- City Light reviewed, compiled, and summarized genetics data collected in the Project reservoirs by multiple researchers. Specifically, City Light consultants (i.e., Cramer Fish Sciences) contacted the WDFW fish genetics laboratory and the USFWS Abernathy Fish Technology Center via email to request all genetic data and metadata. WDFW provided 2,697 genotypes for 15 microsatellite loci that appeared to have also been analyzed by Pflug et al. (2013). Ambiguity exists because individual identification for each genotype was not provided in the Pflug et al. (2013) report. WDFW and USFWS provided 898 genotypes for 16 microsatellites that appeared to have been analyzed by Smith (2010). Summaries of the review, compilation, and analysis for each taxon are provided in Sections 6.2 and 6.3 of this study report.
- City Light acquired and consolidated existing genetics data for Bull Trout, Rainbow Trout, and Dolly Varden. No information was provided on how the samples were collected or what hypotheses were being tested by the data. Due to this ambiguity, City Light's consolidation efforts focused on reducing violation of statistical assumptions that are common to the analysis of microsatellite data in general. Specifically, efforts attempted to increase biologically meaningful signals within the data by reducing noise associated with (1) possible hybridization with *O. clarkii*; (2) small sample sizes; (3) missing and erroneous data; and (4) violation of HWE and linkage equilibrium. Discussion of the baseline genetic metrics for the Project taxa are presented in Sections 6.2 and 6.3 of this study report.
- City Light created a single, standardized data file for each species that compiles genotypes from existing studies. The genotypes were compiled into GENEPOP files (Raymond and Rousset 1995) that are available upon request.
- City Light used the standardized GENEPOP files to evaluate baseline genetic metrics for the three Project taxa. Summaries of the baseline genetic metrics for the Project taxa are presented in Sections 6.2 and 6.3 of this study report.
- City Light calculated within- and among-population summary statistics using consistent methods for Bull Trout and Rainbow Trout. Within sub-population genetic diversity was estimated as expected heterozygosity (H_S). Mean H_S was equal to 0.74 for Rainbow Trout and 0.57 for Bull Trout (Tables 5.2-1 and 5.3-2). Among-population genetic diversity was estimated using F_{ST} and summarized using PCA (see Figures 5.1-3, 5.1-4, 5.1-5, and 5.2-2). Overall F_{ST} was equal to 0.09 for Rainbow Trout and 0.19 for Bull Trout. Discussion of within- and among-population genetic variation of the Project taxa are presented in Sections 6.2 and 6.3 of this report.
- City Light estimated the power (false detection rate) of genetic markers currently in use to identify relationships (e.g., parent-offspring pairs, full-sibling-unrelated pairs). For *O. mykiss* sampled in Roland Creek, a tributary within the Project Boundary, the FNR for identifying related individuals was 0.392. For Bull Trout sampled in Big Beaver, Ruby, and Stetattle

Creeks, the FNR estimated for collections were 0.857, 0.868, and 0.95, respectively, meaning pedigree analysis is expected to result in more false relationship assignments than true assignments.

- Results of this data review are a key step towards identifying how existing genetic information can support coordination of fish species sampling being conducted opportunistically by other relicensing studies and ongoing relicensing field activities.
- Expert Panel review of Year 1 study results and assistance in development of Year 2 study guidance.

6.2 Discussion of Rainbow Trout (*Oncorhynchus mykiss*) Genetic Data

Review of 2,697 microsatellite genotypes provided by WDFW highlighted potential data gaps and opportunities for characterizing baseline genetic structure of *O. mykiss* in the Skagit River basin. Most of the samples appear to have been previously analyzed by Pflug et al. (2013); however, the samples analyzed in common between this report and the Pflug et al. (2013) report were uncertain because sample identification was not included in their reporting. Although the general objective of describing genetic structure is in the Pflug et al. (2013) report, it is unclear in the submitted WDFW datasheet if sampling was implemented to test specific *a priori* hypotheses. Consequently, the purpose of the WDFW data is somewhat ambiguous. Specifically, the dataset contained no information on how samples were collected, assumptions, statistical power of intended analytical approaches, or scope of inference. Therefore, the genotypes analyzed in this report were compiled in a way that attempted to reduce biases that are common to microsatellite datasets to support basic inferences about genetic population structure of *O. mykiss* in the Skagit River basin.

Standard sampling guidance for population genetic studies of salmonids is to collect enough samples in each subpopulation to accurately characterize allele frequencies (Landguth et al. 2010). For iteroparous species like *O. mykiss* that display overlapping generations, effective sampling should at least be representative of the age groups (cohorts comprising the generation[s]) of interest. The assumption is that sampling occurs within the discreet subpopulations and that genetic differentiation (changes in allele frequencies among populations) occurs along a reproductive continuum ranging from panmixia (random mating) to complete isolation (mating is restricted to within subpopulations) (Waples and Gaggiotti 2006). Along this spectrum, nonrandom mating may support the evolution of genetic population structure. Genetic structure is assumed to be a function of genetic drift (random changes in allele frequencies in a population between generations due to sampling individuals that become parents and binomial sampling of alleles during meiosis) and gene flow (exchange of genetic information between subpopulations). This is because microsatellite genetic markers are presumed to have no effect on reproductive success (microsatellites are selectively neutral) and any mutations (changes in the DNA sequence or chromosome in the transmission of genetic information from parent to progeny) have negligible effects on reproductive success (relative proportion of offspring contributed to the next generation).

Assuming further that there is sufficient statistical power (sufficient sampling and genetic markers), and absence of genotyping and sampling error, estimates of diversity (allele frequency distributions at genetic markers) and divergence (differences in observed frequencies) are expected to reflect the true genetic structure of the population. In practice, all these assumptions are likely

never met, because population boundaries are unknown, genotyping and sampling are imperfect with respect to populations, and the statistical power to reliably observe differences in allele frequencies varies. Hence, genetic analyses are accompanied by descriptions of how individuals are sampled in space and time, what decisions occurred regarding how samples were genotyped, how genotypes were summarized into alleles frequencies, and how the observed variation in allele frequencies were compared and analyzed.

The methods for presenting a compiled dataset are described below. The approach attempted to increase the biologically meaningful signal by reducing noise associated with (1) hybridization with *O. clarkii*; (2) small sample sizes; (3) missing and erroneous data; and (4) violation of Hardy-Weinberg and LD. City Light concludes with a discussion of inference on population genetic structure.

Absence of hybridization with *O. clarkii* is a common assumption of analysis of *O. mykiss* genetic structure because many classical analyses assume that genetic variation is a function of effective population size (N_e) and migration (m) within a single taxon (mutation is assumed negligible). For example, the equation $F_{ST} \approx 1/(4N_em + 1)$ used to describe the strength of gene flow on genetic divergence assumes the subpopulations contributing migrants comprise only *O. mykiss*. Although the dataset analyzed contained genotypes at genetic markers apparently diagnostic for *O. clarkii*, their diagnostic properties were unknown because positive control genotypes for nonhybridized *O. clarkii* were not provided. In practice, this limits the ability to estimate evolutionary relationships among subpopulations, which are typically assumed to be a function of genetic drift and gene flow within *O. mykiss*, as opposed to ongoing genetic introgression of alleles from *O. clarkii*.

Small sample sizes can result in imprecise estimates of allele frequencies and thus weak biological inference. There is no accepted threshold or rule for sample sizes because sampling needs vary by hypotheses, research questions, and marker types (Landguth et al. 2010). For the *O. mykiss* dataset, the recommendation of Hale et al (2016) was adopted—that 25 individuals are typically enough to accurately estimate allele frequencies using microsatellites. Nevertheless, others have cautioned that when allelic diversity per population is high, as is the case with microsatellites, sampling effort may need to surpass 80 – 100 individuals to have a high probability of detecting low frequency alleles (Ott 1992: Seeb et al. 2007). Yet, other studies have reported that for isolated populations (n=8,000 individuals), 20 individuals genotyped at 6 microsatellites could produce an accurate allele frequency distribution (Siniscalco et al. 1999). For the present dataset, excluding collections of n<25 provided high power (1.00, $P<0.05$) to detect moderate differentiation ($F_{ST}=0.01$), but low power (0.32, $P<0.05$) to detect low differentiation ($F_{ST}=0.001$).

Like questions of sample size, there is no accepted threshold or rule for treating missing and erroneous microsatellite data. Using computer simulations, Reeves et al. (2016) estimated that for every 1 percent of missing genotypic data, 2 to 4 percent fewer correct population assignments can be expected. They recommended limiting the percentage of missing data to approximately 2 percent, unless a greater amount can be justified. City Light therefore removed all individuals with missing data at two or more loci (approximately 6 percent), which was the most missing data that could be accommodated in a dataset of 15 microsatellites without allowing only individuals with complete genotypes to be included. Regarding genotyping errors, 1 to 2 percent fewer correct population assignments are expected for every percentage increase in genotyping error (Reeves et

al. 2016). Although there are a variety of computer programs available to estimate the frequency of genotyping errors in a dataset, most techniques are based on conformance of genotypes to Hardy-Weinberg proportions.

Hardy-Weinberg and linkage equilibrium are common assumptions of population genetic analyses for a variety of reasons. Metrics of HWE, for example, can provide insight into mating systems of populations (i.e., inbreeding) or data quality problems like genotyping issues, overrepresentation of families, etc. The data compilation method for *O. mykiss* attempted to reduce violations of HWE that might result from data quality problems with the goal to increase chances that metrics reflect the actual underlying mating system. The compilation method of removing markers with consistent deviation from HWE (i.e., *One-14*) and combining collections from the same tributaries but in different years that produced fewer deviations from HWE resulted in a dataset with a lower overall F_{IS} than the original dataset, though the decrease was not statistically significant. For clarity, lower F_{IS} suggests the compilation method succeeded in reducing deviations from HWE.

Regarding LD, City Light observed a consistent negative relationship between sample size (n) and the estimator R^2, which could indicate that, on average, sampling was not sufficient to obtain unbiased estimates of LD. The potential bias presents a challenge to data interpretation. For example, the collection from Diablo Lake contained the highest LD (R^2=0.04) but also one of the smallest sample sizes in the entire dataset (n=26). This presents a data interpretation challenge because the collection from Diablo Lake also contained apparent hybridization with *O. clarkii*, as noted by WDFW, which is expected to cause an increased LD associated with genetic admixture between genetically dissimilar populations. It is therefore uncertain whether high LD in Diablo Lake is associated with something biologically meaningful, like hybridization, or is simply an artifact of bias associated with small sample size.

The type of inference that can be drawn from analysis of the existing data could be limited to basic descriptions of genetic diversity and population structure. Genetic structure was apparent in the analyzed collections. The overall estimated proportion of genetic variance explained by population structure (F_{ST}) was 0.094, and the PCA appeared to provide some evidence that geography affects structure. Nevertheless, specific hypotheses about how geography affects structure were not tested (e.g., isolation-by-distance versus possible gene flow from historical hydrogeological connectivity with the Fraser River). Cautious interpretation of the observed patterns is warranted because the proportion of variation that can be explained by hybridization with *O. clarkii* and/or hatchery introgression was not directly examined due to power limitations (e.g., Vaha and Primmer 2006), which is important because observed patterns of diversity may not reflect natural genetic drift and gene flow.

The existing Rainbow Trout dataset provides limited inference about effective population size (N_e), individual and evolutionary relationships, hybridization, hatchery introgression, and quantitative genetics (e.g., genes and phenotypes under selection). Effective population size (N_e), for example, is arguably the most important metric in conservation biology because it determines how a population evolves. N_e was not estimated, however, because of sampling limitations. Firstly, hybridization with *O. clarkii* or hatchery introgression could bias estimates of N_e by creating genetic disequilibria that is not associated with genetic drift. Secondly, estimating N_e in an iteroparous species with overlapping generations requires extensive sampling effort and significant data on life-stage specific survival and reproduction. Though it is common to estimate effective

number of breeders (N_b), unbiased estimates typically call for sampling of large numbers of individuals (i.e., on the same order of magnitude as true N_b) from the same cohort or across multiple generations, and such data was not available. Likewise, effects of hybridization with either *O. clarkii* or hatchery-origin fish were not directly examined because statistical power of the preexisting dataset was relatively low with just 14 microsatellites, positive control genotypes for *O. clarkii* were not available, and Pflug et al. (2013) have already addressed several questions about hatchery introgression using this dataset.

6.3 Discussion of Bull Trout (*Salvelinus confluentus*) genetic data

Review of 898 microsatellite genotypes provided by USFWS and WDFW highlighted potential data gaps and opportunities for characterizing baseline Bull Trout genetic structure in the Skagit River basin. Most of the samples evaluated appear to have been previously analyzed by Smith (2010), Small et al. (2013), and Small et al. (2016). Yet, it was uncertain which samples were evaluated in common because sample identification was not included in the reporting. The purpose for tissue sampling varied by collection. Smith (2010) stated that the study's collection purpose was to assess genetic variability within and between Bull Trout populations. The stated purpose of WDFW collections was to characterize the genetic variation of Bull Trout, Dolly Varden, and Brook Trout in the Skagit reservoirs. Given the sampling objectives differed for the collections, methodology was not well described, and multiple *Salvelinus* taxa (or hybrids) may have been incorporated into the collections, it was challenging to compile a Bull Trout dataset. As with *O. mykiss*, this approach focused on compiling the dataset that reduced violations of basic assumptions common to the analysis of microsatellites and, in general, to support basic inferences about genetic population structure of *S. confluentus* in the Skagit River basin.

Like descriptions above for *O. mykiss*, evaluation of *Salvelinus* spp. data had to contend with inclusion of potential hybridized individuals, small collection sample sizes, missing and erroneous data, and violation to genetic equilibria. Data quality recommendations as noted above for Rainbow Trout were also applied to Bull Trout. The sample size threshold was reduced to retain collections from within the Project Boundary. Additionally, the missing genotype data threshold was increased to 25 percent and *Salvelinus* spp. samples that appeared ambiguous were omitted from summary statistic estimations. Lastly, while there was limited power to identify related individuals within collections, full-sibling families that were inferred using established methods were reduced in size. These steps resulted in a dataset that largely conformed to genetic equilibrium expectations, which was an improvement in data quality.

If genotypes at each genetic marker location (locus) occur at a frequency expected by random associations of alleles (a function of the allele frequency), genotypes are said to be in HWE, or alleles within loci are uncorrelated (statistically independent). Many phenomena may cause deviations for HWE expectations (e.g., null alleles, inbreeding, population mixing), with the deviation quantifiable using an analysis of variance approach. F statistics partition the reduction (or excess) in heterozygotes relative to HWE. One component, F_{IS}, is the individual relative to the subpopulation (collection). Globally across all collections, the mean F_{IS} observed was low (F_{IS} =0.008) and the 95 percent confidence interval overlapped zero. Further, all Project Boundary tributary collections were statistically consistent with HWE. LD quantifies the correlation of alleles between loci. LD is a useful quantity to measure, as the pattern of LD in the genome is influenced by population history, the breeding system, the pattern of geographic subdivision,

natural selection, gene conversion, and mutation (Slatkin 2008). No Project Boundary tributary collections had statistically significant LD tests. The Ross Lake collection was not in HWE and had 11 statistically significant LD tests.

From a genotype frequency perspective, population structure results in an inbreeding-like effect (a reduction in heterozygotes expected relative to HWE) due to nonrandom mating among all individuals analyzed. As such, measuring the deviation from HWE expectations due to population structure acts as a measure of genetic distance between two populations (or collections in this case). The degree of deviation is quantified using another F statistic, F_{ST}, the component of genetic variance within subpopulation (collection) relative to total population (paired collections being tested). The estimated proportion of genetic variance explained by population structure (F_{ST}) across all Bull Trout collections was 0.19. Project Boundary tributary collections were more similar to each other than any were to collections from below Gorge Dam, with collections from Ross Lake tributaries not statistically different. The Stetattle Creek collection was genetically differentiated from Ross Lake tributary collections. The "at large" Ross Lake collection was genetically differentiated from all Project Boundary collections except upper Skagit River. Pairwise F_{ST} estimates comparing Project Boundary collections with those from below Gorge Dam were substantial in magnitude, with the minimum estimate observed being 0.207.

6.4 Next Steps

Inferences made in this report underscore several next steps towards achieving objectives of the Reservoir Fish Genetics Study. Specifically, this report provides estimates of genetic structure within and among target species, including estimated statistical power to detect different levels of genetic structure and perspectives on quality of existing data, possible data gaps, and opportunities for existing data to meet the objectives of the two-year study. The Expert Panel is expected to provide advisory input into the findings and perspectives of this report as they pertain to the listed objectives of the Reservoir Fish Genetics Study, including:

(1) The genetic structure of salmonid populations in the Project Boundary and whether it is sufficient to inform actions necessary for genetic sustainability;

(2) Determining the number of fish populations, for each target species, within and among the Project reservoirs;

(3) Perspectives on adequacy (or lack thereof) of existing data to estimate the effective size (N_e) for each target species and reservoir; and

(4) Identification of topics and/or management objectives to be considered in the reservoir fish and aquatics management plan.

The Expert Panel was provided a technical memo summarizing the initial data summary created to support this study and met with City Light and LPs on January 18, 2022. An overview of the information contained in the technical memo was provided, and the Expert Panel and LPs were given the opportunity to ask questions. The next steps for the Reservoir Fish Genetics Study are:

- Development of a sampling plan for Year 2 which supports the goals and objectives as described in the RSP and June 9, 2021 Notice, which includes:

  - Collecting field data to support genetic baseline analysis of juveniles at reservoir tributary spawning grounds;
  - Expanding sample collection and/or coordinate existing samples and activities for out-of-basin and above and below dam analyses;
  - Estimating heterozygosity, within- and among-population variance, and relatedness for Dolly Varden in Project reservoirs;
  - Collecting additional data to estimate N_e for each population of Bull Trout, Rainbow Trout, and Dolly Varden; and
  - Logistics of gathering metadata needed to estimate N_e during the ILP study period.

- Continued engagement with the Expert Panel and LPs to support the development of LP research questions into focused research questions with testable hypotheses that could be considered for incorporation into the Year 2 sampling plan or as topics for future long-term management planning.

7.0 VARIANCES FROM PROPOSED STUDY PLAN AND PROPOSED MODIFICATIONS

There are no variances from or proposed modifications to the study plan for the Reservoir Fish Genetics Study.

8.0 REFERENCES

Anderson, S. 2019. CKMRsim package, efficient sampling for close-kin mark recapture power estimation. [Online] URL: https://github.com/eriqande/CKMRsim. September 2021.

Angers, B. and L. Bernachez. 1996. Usefulness of heterologous microsatellites obtained from brook charr, *Salvelinus fontinalis* Mitchill, in other *Salvelinus* species. Molecular Ecology 5:317–319.

Banks, M. A., M. S. Blouin, V. K. Baldwin, H. A. Rashbrook, H. A. Fitzgerald, S. M. Blankenship, and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in Chinook Salmon (*Oncorhynchus tschawytscha*). The Journal of Heredity 2.

Buchholz, W. G., S. J. Miller, and W. J. Spearman. 2001. Isolation and characterization of chum salmon microsatellite loci and use across species. Animal Genetics 32.

Cairney, M., J. B. Taggart, and B. Hoyheim. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Molecular Ecology 9:2155-2234.

Crane, P. A., C. J. Lewis, E. J. Kretschmer, S. J. Miller, W. J. Spearman, A. L. DeCicco, M. J. Lisac, and J. K. Wenburg. 2004. Characterization and inheritance of seven microsatellite loci from Dolly Varden, *Salvelinus malma*, and cross-species amplification in Arctic char, S. alpinus. Conservation Genetics 5:737-741.

DeHaan P. W. and W. R. Ardren. 2005. Characterization of 20 highly variable tetranucleotide microsatellite loci for bull trout (*Salvelinus confluentus*) and cross amplification in other Salvelinus species. Mol. Ecol. Notes 5: 582-585.

DeHaan, P., B. J. Adams, and M. C. Nelson. 2014. Fine-scale population structure analysis and genetic population assignments of Wenatchee River sub-basin bull trout. USFWS, editor, Longview, WA.

Excoffier, L., and M. Slatkin. 1998. Incorporating genotypes of relatives into a test of linkage disequilibrium. Am. J. Hum. Genet. 171-180.

Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485-486.

Goudet, J., M. Raymond, T. de Meeüs, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144: 1933-1940.

Guo, S. and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372.

Hale, M. L., T. M. Burg, and T. E. Steeves. 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population Is enough to accurately estimate allele frequencies. PLoS ONE 7:1-10.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal 732 components: a new method for the analysis of genetically structured populations. BMC 733 Genet. 11:94. BioMed Central Ltd.

Jombart, T. 2011. *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10:551-555.

Landguth, E. L., S. A. Cushman, M. K. Schwartz, K. S. McKelvey, M. Murphy, and G. Luikart. 2010. Quantifying the lag time to detect barriers in landscape genetics. Molecular Ecology 19:4179–4191.

McConnell, S., L. Hamilton, D. Morris, D. Cook, D. Paquet, P. Bentzen, and J. Wright. 1995. Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian east coast stocks of Atlantic salmon. Aquaculture 137:19-30.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA.

Nelson, R. J., T. D. Beacham, and M. P. Small. 1998. Microsatellite analysis of the population structure of a Vancouver Island sockeye salmon (*Oncorhynchus nerka*) stock complex using nondenaturing gel electrophoresis. Molecular Marine Biology and Biotechnology 7:312-319.

Olsen, J. B., P. Bentzen, and J. E. Seeb. 1998. Characterization of seven microsatellite loci derived from pink salmon. Molecular Ecology 7:1087-1089.

Ott, J. 1992. Strategies for characterizing highly polymorphic markers in human gene mapping. American Journal of Human Genetics 51:283-290.

Pflug, D. E., E. J. Connor, B. Hayman, T. W. Kassler, K. Warheit, B. McMillan, and E. Beamer. 2013. Ecological, genetic, and productivity consequences of interactions between hatchery- and natural-origin steelhead of the Skagit watershed. Skagit River System Cooperative.

Queller, D.C. and K.F. Goodnight. 1989. Estimating relatedness using genetic markers. Evolution 43(2):258–275.

Raymond M. and F. Rousset. 1995. An exact test for population differentiation. Evolution 49:1280-1283

Reeves, P. A., C. L. Bowker, C. E. Fettig, L. R. Tembrock, and C. M. Richards. 2016. Effect of error and missing data on population structure inference using microsatellite data. BioRxiv. 080630.

Rexroad, C. E. III, R. L. Coleman, A. M. Martin, W. K. Hershberger, and J. Killefer. 2001. Thirty-five polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). Animal Genetics 32:317-319.

Rousset, F. 2020. Genepop Manual. Version 4.7.5.

Ryman, N., and S. Palm. 2006. POWSIM: A computer program for assessing statistical power when testing for genetic differentiation. Molecular Ecology Notes 6:600-602.

Scribner, K. T., J. R. Gust, and R. L. Fields. 1996. Isolation and characterization of novel microsatellite loci: cross species amplification and population genetic applications. Canadian Journal of Fisheries & Aquatic Sciences 53.

Seattle City Light (City Light). 2021. Revised Study Plan (RSP) for the Skagit River Hydroelectric Project, FERC Project No. 553. April 2021.

_____. 2022. FA-03 Reservoir Fish Stranding and Trapping Risk Assessment, Interim Report for the Skagit River Hydroelectric Project, FERC Project No. 553. Prepared by Four Peaks Environmental Science & Data Solutions and Blue Coast Engineering. March 2022.

Seeb, L. W., A. Antonovich, M. A. Banks, T. D. Beacham, M. R. Bellinger, S. M. Blankenship, M. R. Campbell, N. A. Decovich, J. C. Garza, C. M.Guthrie III, T. A. Lundrigan, P. Moran, S. R. Narum, J. J. Stephenson, K. J. Supernault, D. J. Teel, W. D. Templin, J. K. Wenburg, S. F. Young, and C.T. Smith. 2007. Development of a standardized DNA database for Chinook Salmon. Fisheries 32:540–552.

Siniscalo, M., R. Robledo, P. K. Bender, C. Carcassi, L. Contu, and J. C. Beck. 1999. Population genomics in Sardinia: a novel approach to hunt for genomic combinations underlying complex traits and diseases. Cytogenetic and Genome Research 86:148-152.

Slatkin, M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 9: 477–485.

Small, M.P., E. Martinez, and S. Peterson. 2013. Genetic analysis of native char collected in Diablo Lake, Ross Lake, Gorge Reservoir, Skagit mainstem, and Skagit Bay. WDFW Molecular Genetics Lab, Olympia WA.

Small, M.P., S. Bell, and C. Bowman. 2016. Genetic analysis of native char collected in Diablo Lake, Gorge Reservoir, and Ross Lake in the Skagit River basin. Washington Department of Fish and Wildlife Molecular Genetics Lab, Olympia, Washington.

Smith, C. T., B. F. Koop, and R. J. Nelson. 1998. Isolation and characterization of coho salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids. Molecular Ecology 7:1614-1616.

Smith, M. J. 2010. Population structure and genetic assignment of bull trout (*Salvelinus confluentus*) in the Skagit River Basin. School of Aquatic and Fishery Sciences University of Washington.

Smith, M. 2021. Personal communication between Scott Blankenship, Science Director Cramer Fish Sciences and Matt Smith, Fish Geneticist USFWS. June 14, 2021.

Spies, I. B., D. J. Brasier, P. T. L. O'Reilly, T. R. Seamons, and P. Bentzen. 2005. Development and characterization of novel tetra-, tri-, and dinucleotide microsatellite markers in rainbow trout (*Oncorhynchus mykiss*). Molecular Ecology Notes 5:278-281.

Stephenson, J. J., M. R. Campbell, J. E. Hess, C. C. Kozfkay, A. Matala, M. V. McPhee, P. Moran, S. Narum, M. M. Paquin, O. Schlei, M. P. Small, D. M. Van Doornik, and J. Wenburg. 2009. A centralized model for creating shared, standardized, microsatellite data that simplifies inter-laboratory collaboration. Conservation Genetics 10:1145-1149.

Vaha, J., C. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci.

Molecular Ecology 15:63–72. doi:10.1111/j.1365-294X.2005.02773.x.

Waples, R., and E. C. Anderson. 2017. Purging putative siblings from population genetic data sets: a cautionary view. Molecular Ecology 26:1211-1224.

Waples, R., and O. E. Gaggiotti. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15:1419–1439.

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

Weir, B.S. and J. Goudet. 2017. A Unified Characterization of Population Structure and Relatedness. Genetics 206(4):2085–2103.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics. 15:323–354.

Yang, R.-C. 1998. Estimating Hierarchical F-Statistics. Evolution 52(4):950–956.

This page intentionally left blank.

RESERVOIR NATIVE FISH GENETICS BASELINE STUDY
INTERIM REPORT

ATTACHMENT A

EXISTING GENETICS DATA REVIEW TECHNICAL MEMORANDUM

Technical Memorandum

Date:	Tuesday, January 04, 2022
Project:	Skagit River Hydroelectric Project
To:	Seattle City Light
From:	Scott Blankenship and Dan Bingham, Cramer Fish Sciences
Subject:	FA-06 Reservoir Native Fish Genetics Baseline Study – Existing Genetics Data Review

1.0 INTRODUCTION

1.1 Introduction

The Skagit River Hydroelectric Project is owned and operated by Seattle City Light (City Light) and is undergoing Federal Energy Regulatory Commission (FERC) relicensing with its current operating license expiring in 2025. The FA-06 Reservoir Native Fish Genetics Baseline Study (Reservoir Fish Genetics Study) was not required by FERC in its July 16, 2021 Study Plan Determination; however, City Light is implementing the Reservoir Fish Genetics Study as proposed in the Revised Study Plan (RSP; City Light 2021) with the agreed upon modifications described in the "Notice of Certain Agreements on Study Plans for the Skagit Relicensing" (June 9, 2021 Notice)[1]. The Year 1 objectives outlined in the Reservoir Fish Genetics Study aim to use genetic data produced by previous studies to better understand the types of inferences that can be drawn about the genetic diversity and population structure of native fish and to identify possible data gaps that might prevent satisfactory answers to questions submitted by licensing participants (LP) to City Light. Year 2 will use inferences gleaned from the Year 1 efforts (i.e., information provided in this memo) to fill any data gaps and provide answers to any outstanding questions pursued.

The microsatellite genotypes analyzed in this memo were previously identified in the FA-06 Reservoir Fish Genetics Study Plan as potentially providing useful baseline genetic data for native salmonids (City Light 2021). Since the early 2000s, microsatellite data were collected by multiple researchers from native salmonids sampled in the Project reservoirs, their tributaries, and from outside the Project vicinity but within the Skagit and nearby basins. City Light worked with primary researchers that produced these datasets (U.S. Fish and Wildlife Service [USFWS] and Washington Department of Fish and Wildlife [WDFW]) and with CFS to obtain and compile those existing native salmonid genetics data and to unify them for a common analysis of the baseline genetics metrics identified in subsequent sections of the RSP. This memo therefore reflects a *post hoc* analysis of existing microsatellite genotypes from Rainbow Trout (*Oncorhynchus mykiss*) and Bull Trout (*Salvelinus confluentus*) collected from inside and outside of the Project reservoirs

[1] Referred to by FERC in its July 16, 2021 Study Plan Determination as the "updated RSP."

(Figures 1 and 2) and is intended to establish a basis of existing information to support the Reservoir Fish Genetics Study.

Data requests were made by City Light to obtain the previously identified and pertinent microsatellite genotypes listed in Table 2.5-1 of the RSP. The genotypes for Rainbow Trout were originally produced and analyzed by Pflug et al (2013). Most genotypes for Bull Trout were produced and analyzed by Smith (2010), but genotypes from Small et al. (2013), (2016) and (2020) were also included. Methods described in this memo were not chosen to test hypotheses of the original studies, to provide a peer-review, nor to implement meta-analysis of their results. Rather, methods were chosen to describe the existing data in the context of the Reservoir Fish Genetics Study objectives. Due to inherent risks associated with evaluating existing genetic information for *post hoc* scientific investigations in general (i.e., inference based on analyses outside of the intended scope of the initial study design), caution is warranted during interpretation of results presented in this memo.

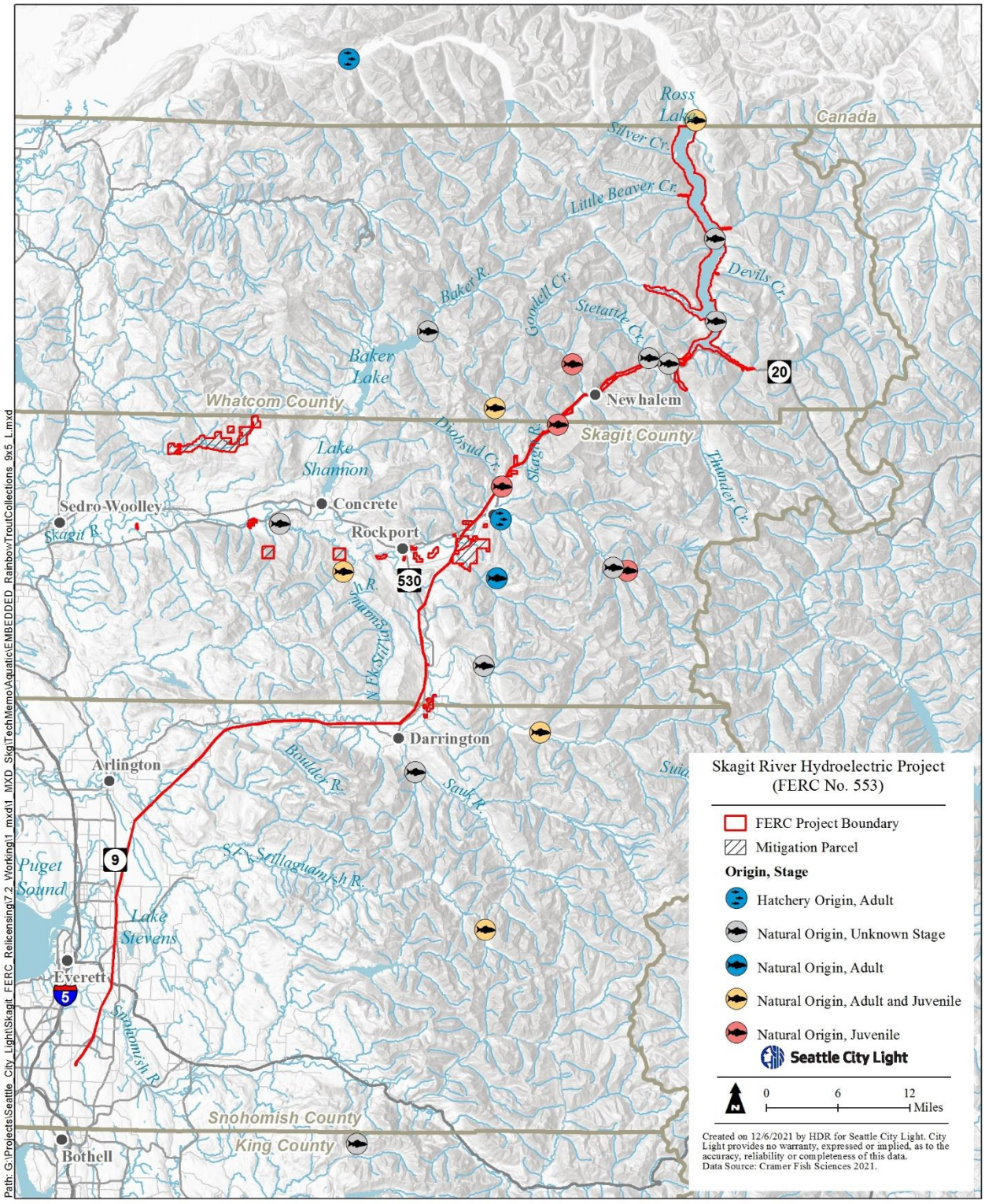

Figure 1. Proposed study area and collections evaluated for Rainbow Trout.

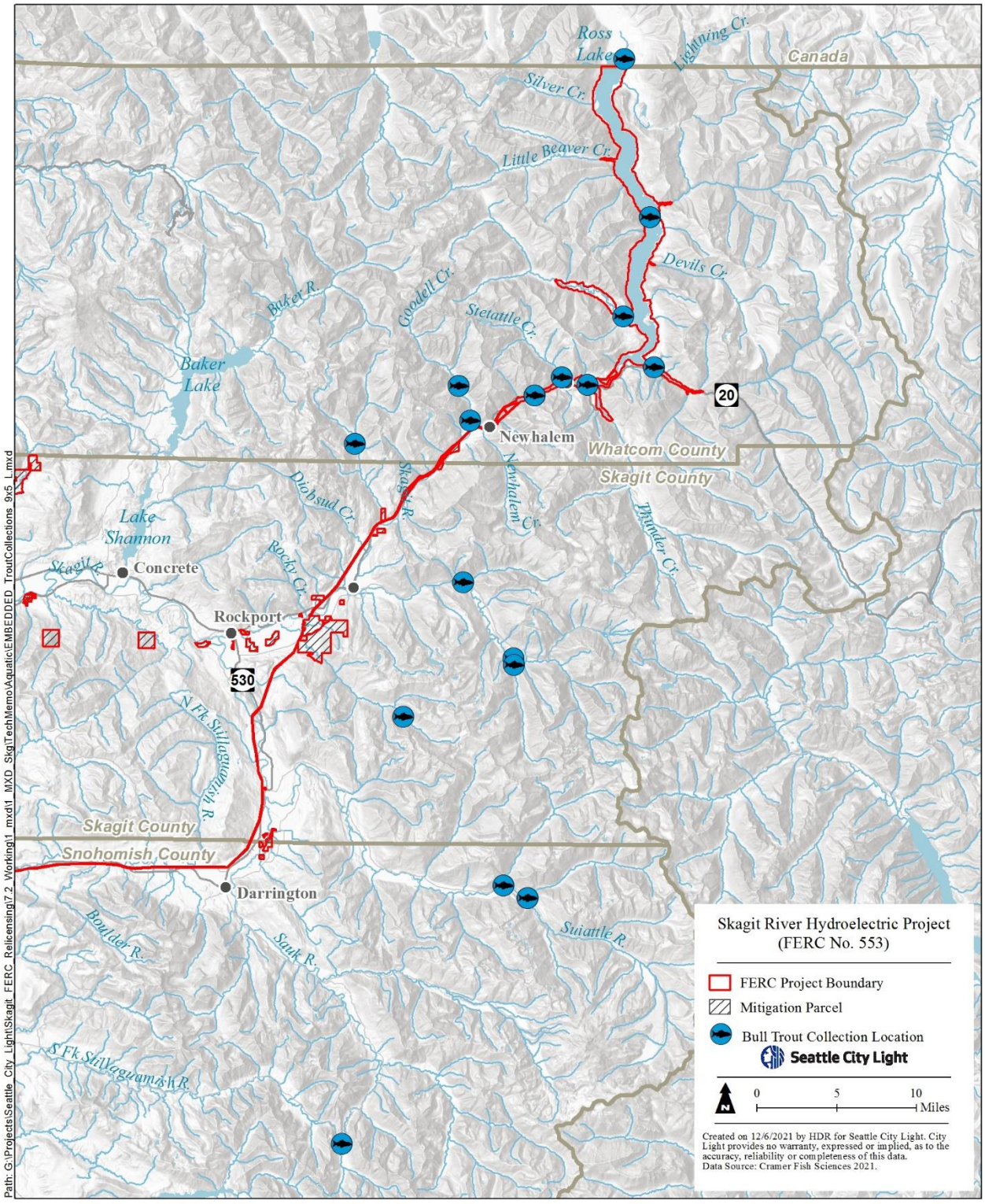

Figure 2. **Proposed study area and collections evaluated for Bull Trout.**

2.0 METHODS

2.1 What preexisting genetic data are available?

On June 6, 2021, Cramer Fish Sciences (CFS) emailed Todd Seamons of the WDFW genetics laboratory (Olympia, WA) and Matt Smith of the USFWS Abernathy Fish Technology Center (Longview, WA), requesting the microsatellite genotypes and metadata analyzed in Pflug et al. (2013), Smith et al. (2010), Small et al. (2013), (2016) and (2020). On September 9, 2021, CFS sent an additional request for any geospatial data that could aid in identifying the specific locations that tissue samples were collected. Todd Seamons and Matt Smith each forwarded the requested genotypes (See Section 3 of this memo for details) but indicated geospatial data were unavailable, and thus, precise reaches of rivers, streams, or positions in a lake are not known beyond descriptions in the original reports.

2.2 What types of inference can be drawn?

Understanding the designs of the original studies is important because any new conclusions are naturally limited by the initial scopes of inference (e.g., sampling designs, genetic marker choices, etc.). Briefly, the study design of Pflug et al. (2013) was observational in nature (i.e., not experimental). Sampling occurred between 2008 and 2010 and appeared to be loosely stratified by life history (i.e., anadromous versus resident), life stage (juvenile versus adult), degree of isolation (upstream versus downstream of a migration barrier), and by origin (hatchery- versus natural-origin), but allocation of sampling effort within strata was not defined *a priori*. The extent to which sampling was random, opportunistic, or targeted was unclear, but the report stated that for each collection, attempts were made to obtain 100 adult steelhead, 100 juvenile steelhead from anadromous zones, and 100 resident rainbow trout from above barriers. The geographic extent of sampling was broad across the Skagit River basin, including multiple collections from above and below the Project Boundary and within the Project reservoirs (Figures 1 and 2). Thus, some collections may characterize subpopulations while others may be mixtures. Samples from natural populations in the Sauk and Fraser rivers were also examined and so were samples from hatchery populations commonly used to supplement natural populations of Rainbow Trout in the Skagit River. Fifteen microsatellite markers were analyzed to *"…provide information about basic genetic characteristics of natural and hatchery origin steelhead populations and resident [O.] mykiss populations.*" Nine additional objectives that can be broadly categorized as descriptions of genetic population structure were also listed but no testable hypotheses were defined. The statistical populations of inference to which descriptions of genetic diversity likely apply (i.e., the extent over which inferences applied) might therefore be loosely defined as: (1) naturally reproducing subpopulations of resident and anadromous *O. mykiss* affected by hydropower management on the Skagit River (e.g., hatchery supplementation, isolation, and hybridization with *O. clarkii*); and (2) artificially reproducing hatchery populations that are commonly used to supplement *O. mykiss* in the Skagit River basin. Due to the relatively short, two-year period over which sampling occurred, inferences might only reflect a "snapshot" of genetic diversity and could be limited to the one or two generations sampled between 2008 to 2010.

Because the Bull Trout samples analyzed in this memo were collected by multiple researchers, the scope of inference is harder to define than for Rainbow Trout. The study designs of Bull Trout described in Smith (2010), Small et al. (2013), and Small et al. (2016) were observational in nature. The most comprehensive of the three studies was completed by Smith (2010) in which 16

microsatellite markers were analyzed to "*…complete an assessment of the genetic variability within and among bull trout populations of the Skagit River Basin and subbasins in the vicinity of Seattle City Light's (SCL) Skagit Hydroelectric Project.*" Five objectives were listed, which can broadly be categorized as descriptions of genetic diversity and assignments of individuals from potentially mixed fishery collections to natal subpopulations of origin. Five-hundred-ninety-five juvenile and adult Bull Trout were sampled from fourteen localities due to their proximity to City Light's hydroelectric facilities and to represent populations' "baseline" localities likely to contribute to the genetic diversity of the adult Bull Trout found in the mainstem Skagit River. Samples were collected using a combination of electrofishing, snorkeling, and angling. No data were available regarding the extent to which sampling was opportunistic, targeted, or random with respect to age, phenotype, and life history. During 2006-2008, 435 fluvial adult and sub-adult Bull Trout were collected from the mainstem Skagit River from the Gorge Powerhouse to the confluence of the Sauk River. These samples represent a potential mixture from several spawning populations located throughout the Skagit basin. Samples were collected primarily by angling. The scope of inference for Bull Trout considered by this memo might therefore be defined as naturally reproducing subpopulations of Bull Trout in the Project vicinity that were affected by management of the Skagit Hydroelectric Project (e.g., isolation and hybridization) from 2005 to 2015.

2.2.1 Limitations of *Post Hoc* Studies

Due to the *post hoc* nature of analyses presented in this memo, statistical methods were chosen based on their ability to accommodate assumptions common to population genetic studies of **microsatellite** genotypes in general (words in bold appear in Glossary Section 6.0 of this memo). Typical guidance is to collect enough genetic samples in each **subpopulation** to accurately characterize **allele frequencies** (Landguth et al. 2010). For **iteroparous** species like Rainbow Trout and Bull Trout that display **overlapping generations**, sampling should at least be representative of the cohorts comprising the generation(s) of interest (Allendorf and Phelps 1980). A common assumption is that sampling occurs within predefined and discreet subpopulations and that **genetic population structure** occurs along a reproductive continuum ranging from panmixia (random mating) to complete isolation (mating is restricted to within subpopulations) (Waples and Gaggiotti 2006). Genetic diversity is often assumed to be a function of **genetic drift** and **gene flow** because microsatellites are assumed to be **selectively neutral** and the rate at which new genetic diversity enters the population through mutation is negligible (Waples and Gaggiotti 2006). Assuming further that there is sufficient **statistical power**, and absence of genotyping and sampling errors, estimates of genetic diversity are expected to reflect the true genetic structure of the population. In nature, all these assumptions are never met.

2.2.2 Genetic Analysis of Rainbow Trout

Due to the limitations described above, collections of genotypes that appeared to have been sampled from the same localities were pooled and treated them as random samples from individual "subpopulations" of Rainbow Trout unless statistical evidence that suggested they should be separated was observed (i.e., temporally spaced collections from the same locations were pooled unless the null hypothesis of genetic homogeneity was rejected). The computer program POWSIM Version 4.1 (Ryman and Palm 2006) was used to estimate statistical power to detect deviation from genetic homogeneity. POWSIM is a simulation-based computer program that estimates statistical power of rejecting the null hypothesis (H_0) of genetic homogeneity for different combinations of sample sizes, number of loci, number of alleles, and allele frequencies for a

hypothetical degree of true differentiation (quantified as F_{ST}). POWSIM can only accommodate 30 collections of individuals, so the first 30 collections (Table 1) were used to estimate power to detect low (F_{ST}=0.001) and moderate (F_{ST}=0.01) genetic differentiation by assuming allele frequencies estimated in Pflug et al. (2013). The program FSTAT Version 2.9.3.1 (Goudet 1995) was used to estimate and test metrics of genetic diversity unless otherwise stated. Expected heterozygosity (H_S, i.e., gene diversity assuming Hardy-Weinberg equilibrium [HWE]) and allelic richness (A_R) were estimated to describe genetic diversity across loci and collections. One thousand randomizations of alleles at each locus were performed to test the assumption of HWE at within collections. Observed (H_O) and expected multilocus heterozygosity within subpopulations were compared using Wright's (1951) F_{IS} to measure the magnitude of departures from HWE. To assess the assumption of random association of alleles among loci, log-likelihood ratio tests using 1,000 permutations were implemented to test for pairwise linkage disequilibrium (LD) within all collections. The Weir and Cockerham (1984) version of F_{ST} was estimated to measure genetic differentiation between all pairs of collections. To summarize genetic diversity among the sampled individuals, a principal component analysis (PCA) of individual-based genetic distances based on allele frequencies was implemented using the R package {adegenet} (Jombart et al. 2010). Ordination in "allelic space" along the first three PC axes was visualized using ggplot in program R. The statistical power to observe relatives was determined using {CKMRSim} (Anderson 2019). All tests of significance at the $\alpha = 0.05$ level were assessed and applied Bonferroni corrections when conducting multiple tests.

Table 1. Summary statistics for samples collected from *O. mykiss* in the Skagit and Fraser River basins.

Collection number [1]	Collection size	WDFW Code [2]	Location	Origin[3]	Upper Skagit[4]	Stage	Phenotype[5]	F_{IS} [6]	H_S [7]	A_R [8]	R^2 [9]
1	57	07MS, 08MI, 10BA	Bacon Creek	NOR	No	Juvenile, adult		0.01	0.79	9.45	0.02
X*	57	09EL	Baker River 09	NOR	No		Trout	0.09	0.82	10.44	0.03
X*	42	10AU	Baker River 10	NOR	No		Trout	0.11	0.84	11.13	0.04
2	51	09EU	Big Creek 09	NOR	No		Trout	0.04	0.66	5.63	0.02
3	48	10BG	Big Creek 10	NOR	No		Trout	0.06	0.67	5.16	0.03
4	52	09JB, 10BJ	Blackwater River	NOR	No	Juvenile	Trout	0.11	0.74	7.61	0.02
5	66	10MZ	Chilliwack Hatchery	HOR	No	Adult		0.00	0.76	8.02	0.02
6	94	09ET, 10BE	Clear Creek	NOR	No		Trout	0.06	0.68	8.56	0.01
7	38	10BB	County Line Ponds	NOR	No	Juvenile		0.01	0.80	9.10	0.04
8	26	05NG	Diablo	NOR	Yes		Trout	0.06	0.75	8.50	0.04
9	41	10BK	Diobsud	NOR	No	Juvenile		0.02	0.79	9.77	0.03
10	43	03OA	Dry Creek	NOR	Yes		Trout	0.02	0.71	7.50	0.03
11	47	09EH	Finney Creek	NOR	No	Juvenile		0.00	0.78	9.27	0.03
12	47	10AT	Finney Creek	NOR	No	Juvenile		0.01	0.80	9.59	0.02

Collection number [1]	Collection size	WDFW Code [2]	Location	Origin[3]	Upper Skagit[4]	Stage	Phenotype[5]	F_{IS} [6]	H_S [7]	A_R [8]	R^2 [9]
13	30	11BK	Finney Creek	NOR	No	Adult		-0.02	0.80	10.40	0.04
14	38	09IZ	Goodell Creek	NOR	No	Juvenile		0.01	0.77	8.76	0.03
15	41	10BC	Goodell Creek	NOR	No	Juvenile		0.00	0.79	9.12	0.03
16	47	09EE	lower Cascade	NOR	No	Juvenile		-0.05	0.77	8.23	0.03
17	44	10AV	lower Cascade	NOR	No	Juvenile		0.03	0.79	9.26	0.03
18	48	10AY	lower Skagit	NOR	No	Juvenile		0.02	0.79	9.51	0.03
19	28	08LF	lower Skagit	NOR	No	Adult		0.01	0.78	9.26	0.04
20	59	09CF	Marblemount	HOR	No	Adult		0.01	0.82	9.68	0.02
21	44	10AN	Marblemount	HOR	No	Adult		0.03	0.79	8.89	0.03
22	39	09BM	mid Skagit	NOR	No	Adult		0.01	0.80	10.49	0.03
23	31	10AS	mid Skagit	NOR	No	Adult		0.04	0.80	10.14	0.03
24	47	09ES	NF Cascade	NOR	No		Trout	0.11	0.41	4.30	0.02
25	45	10BF	NF Cascade	NOR	No		Trout	-0.08	0.36	3.98	0.02
26	79	02FB	Roland Creek	NOR	Yes		Trout	0.01	0.71	7.68	0.01
27	30	06AF	Ross	NOR	Yes		Trout	0.03	0.73	8.20	0.04
28	44	09MA	Ross	NOR	Yes		Trout	-0.01	0.69	6.65	0.04
29	47	10BH	Ross	NOR	Yes		Trout	-0.03	0.70	6.40	0.04
30	45	10AX	Sauk	NOR	No	Juvenile		0.04	0.80	9.66	0.03
31	29	83AAA	Sauk	NOR	No	Adult		0.06	0.80	10.29	0.04
32	32	09JA	Stetattle	NOR	Yes		Trout	0.03	0.76	8.66	0.04
33	41	10BI	Stetattle	NOR	Yes		Trout	0.03	0.77	8.79	0.03
34	115	09DT, 09EF, 10AQ, 10AW, 11BM	Suiattle	NOR	No	Juvenile, adult		0.01	0.79	10.05	0.01
35	51	09EV	upper Finney	NOR	No		Trout	0.03	0.74	6.52	0.02
36	49	10BD	upper Finney	NOR	No		Trout	0.04	0.72	6.77	0.02
37	56	10AZ	upper Skagit	NOR	No	Juvenile		0.01	0.79	9.56	0.02
38	32	11BI	upper Skagit	NOR	No	Adult		0.00	0.81	10.43	0.03

* Collections removed from PCA due to indirect evidence of hybridization with *O. clarkii*.

1 Collection number: corresponds to Figures 5 through 7.

2 WDFW code: WDFW collection identification with apparent sample year as the prefix

3 Origin: hatchery (HOR) or natural (NOR) origin.

4 Upper Skagit: collections from upstream of the Project Boundary in the Skagit River and from B.C.

5 Phenotype: identifies whether collections were from apparent trout as determined by WDFW.

6 F_{IS}: estimated deviation from HWE.

7 H_S: estimated expected heterozygosity within sub-populations (i.e., gene diversity).

8 A_R: estimated allelic richness.

9 R^2: is the estimated pairwise correlation of alleles among loci.

2.2.3 Genetic Analysis of Bull Trout

Exploratory analyses were conducted on Bull Trout similar to those described for Rainbow Trout. Partitioning of genetic variation was explored using visualization of individual-based data and genetic principal component analysis (e.g., Jombart et al. 2010). The statistical power to observe relatives was determined using {CKMRSim} (Anderson 2019). H_S was estimated following the sampling bias correction method described be Nei (1987). A common implementation of HWE test was used following Guo and Thompson (1992) Markov-chain random walk extension of Fisher's (2-allele) classical exact test. Departures from HWE were also quantified using the inbreeding coefficient (F_{IS}) statistic observed from analysis of molecular variance (AMOVA) (Excoffier et al. 1992; Yang 1998), which is equivalent to Weir and Cockerham (1984) small *f* statistics. Collections were analyzed for evidence of LD (i.e., non-independence of alleles at different loci). Given gametic phase was unknown for previously reported data, LD between a pair of loci was tested using a likelihood-ratio test, whose empirical distribution is obtained by a permutation procedure (e.g., Excoffier and Slatkin 1998). Lastly, allelic distributions across collections were evaluated using contingency table analysis of observed allelic distributions described by Raymond and Rousset (1995).

The AMOVA framework used to describe genetic structure of Bull Trout estimates hierarchical f-statistics for any number of desired levels (e.g., within individuals, within subpopulations, among subpopulations). This allows for subpopulation differentiation (allele frequency variance) to be quantified. In other words, the degree that individuals within a subpopulation (collection) are more similar to each other than are individuals from different subpopulations (collection). There are many formulations of the population differentiation variance component measure, although a common implementation is a form of the fixation index (e.g., genetic divergence [F_{ST}]). F_{ST} metrics were estimated pairwise following Weir and Goudet (2017) and used as a measure genetic divergence, with statistical significance calculated following likelihood-ratio tests (Goudet et al. 1996).

3.0 RESULTS

3.1 What genetic data are available?

3.1.1 Rainbow Trout

On 28 July 2021, Todd Seamons (WDFW) provided an Excel spreadsheet containing 2,967 genotypes for the 15 microsatellites analyzed in Pflug et al. (2013): *One-102*, *Ogo-4* (Olsen et al. 1998), *Ots-100* (Nelson et al. 1998), *Oki-10*, *Oki-23*(Smith et al. 1998), *Omy-7* (K. Gharbi, unpublished, as referenced in Pflug et al. 2013), *Omy-1001*, *Omy-1011*(Spies et al. 2005), *Ots-3M*, *Ots-4* (Banks et al. 1999), *One-14* (Scribner et al. 1996), *Ssa-407*, *Ssa-408* (Cairney et al. 2000), *Ssa-298* (McConnell et al. 1995), and *Oke-4* (Buchholz et al. 2001). The dataset included the following metadata: 'Sample Name', 'WDFW code', 'Count', 'Percent Missing Data'. Various other metadata were available directly from the Pflug et al. (2013) report (Table 1). The microsatellites appeared to be a subset of the standardized Stevan Phelps Allele Nomenclature (SPAN) markers described in Stephenson et al. (2009) that were developed to ensure data quality (repeatable allele scoring) across laboratories. Exact sampling locations (i.e., GPS coordinates) but based on the Pflug et al. (2013) report, appeared to include tributaries, mainstem rivers, hatcheries, and Project reservoirs. Some sites appeared to have been sampled across multiple years. The collections ranged in size from n=1 in the Suiattle River in 2009 to n=106 in Diablo Lake in 2005. No metadata were provided regarding sampling field methods (e.g., electrofishing), whether samples were collected randomly, or targeted life stages, life histories, morphologies, taxa, etc.

3.1.2 Bull Trout

On 13 June 2021, Matt Smith (USFWS) provided a tab delimited .txt file containing 563 genotypes at 16 microsatellite loci previously analyzed in Smith et al. (2010): *Omm1128*, *Omm1130* (Rexroad et al. 2001), *Sco102*, *Sco105*, *Sco106*, *Sco107*, *Sco109* (WDFW unpublished), *Sco200*, *Sco202*, *Sco212*, *Sco215*, *Sco216*, *Sco218*, *Sco220* (Dehaan and Ardren 2005), *Sfo18* (Angers and Bernachez 1996), and *Smm22* (Crane et al. 2004). The dataset included the following metadata: 'Individual Name', 'Synonym 1', 'Region (1)', 'Watershed (2)', 'Tributary (3)', 'Capture Location (4)', 'Age', 'Brood Year', 'Collected By', 'Collection Year', 'Comment', 'Date Collected', 'Fork Length (mm)', 'Hatchery/Wild', 'HOR/NOR assignment', 'Latitude', 'Life History', 'Stage', 'Longitude', 'Phenotypic Sex', 'PIT Tag', 'Population ID', 'Preservation Method', 'Project number', 'Received From', 'Resident / Anadromous', 'Run Type', 'Spawn Date', 'Spawn Year', 'Spawned With', 'Species', 'Synonym 2', 'Synonym 3', 'Tissue Type', 'Total Length (mm)', 'Used for Broodstock', 'Weight (g)'. Only some of these metadata were relevant to this report or contained entries. The same email sent from WDFW on 28 July 2021 contained 335 genotypes from six collections from the Project lakes and two collections from outside the Project Boundary.

3.2 What types of inference can be drawn from preexisting data?

3.2.1 Genetic analysis of Rainbow Trout

Of the 2,697 Rainbow Trout genotypes provided by WDFW, 536 were removed prior to analysis due to missing genotypes at two or more loci (e.g., as recommended by Reeves et al. 2016), and 20 were removed because their genotype was duplicated elsewhere in the dataset (i.e., they were removed due to possible pseudo replication). Pooling of samples from the same locations and across years reduced the number of analyzed collections from 76 to 25, however, only four of the

pooled collections contained fewer deviations from HWE and were thus retained: Bacon Creek (2007 to 2010), Clear Creek (2009 and 2010), Blackwater River (2009 and 2010), and the Suiattle River (1981 and 2009 to 2011). Collections with fewer than 25 individuals were removed to avoid biased estimates of allele frequencies within sub-populations as recommended by Hale et al. 2012. The provided table of genotypes did not contain any information about which markers were diagnostic for *O. clarkii* and so hybridization was not directly assessed. However, certain genotypes were removed from analyses if hybridization was indirectly apparent (i.e., if genotypes appeared to be statistical outliers).

Comparison of observed (H_O=0.729) and expected (H_S=0.747) heterozygosity across all collections and loci suggested a relatively small but overall deficit of heterozygotes (F_{IS}=0.025 95 percent CI [0.03, 0.01]). Eighty-six of 600 (14 percent) randomization tests for HWE (15 markers x 40 collections) were significant at the α=0.05 level with 68 (79 percent) of the tests showing a deficit of heterozygotes. No tests for HWE were significant at the adjusted level of α=0.00008. The locus *One-14* deviated from HWE in 17 of 40 (42.50 percent) total collections with all tests showing a deficit of heterozygotes. By contrast, most other markers (11 of 15) produced various combinations of heterozygote excess and deficiency. Therefore, the locus *One-14* was omitted from further analysis. This adjustment decreased mean F_{IS} to 0.017, though the difference was not statistically significant (95 percent CI: 0.03, 0.01).

The final dataset contained 1,900 individuals from 40 collections[2] genotyped at 14 microsatellites. However, in some instances, fewer than 40 collections and 1,900 individuals were analyzed (i.e., PCA) due to indirect evidence of hybridization with *O. clarkii*. The 14 microsatellite loci had a total of 312 alleles, ranging from 11 at *Ots-4* to 32 at *Omy-1001*. Across all 14 loci and 40 collections, the estimated false detection rate of a parent-offspring pair was 0.00000811, 0.00000033 of full siblings, and 7.277×10^{-21} of unrelated individuals. However, within any single collection, power is expected to be substantially lower. For example, the false positive rate (FPR) for related individuals in Roland Creek, a tributary within the Project Boundary, is 0.0000161 and the false negative rate is 0.392 (Figure 3). Gene diversity (H_S) within each collection ranged from 0.36 in the collection from North Fork Cascade River in 2010 to 0.83 in the Baker River in 2010 (Table 1). Average gene diversity in collections from upstream of the Project Boundary at Gorge Reservoir (H_S=0.74) was similar diversity in all other collections (H_S=0.74).

[2] While 40 collections were included in these study analyses, collections from the Baker River were removed from the PCA due to hybridization with *O. clarkii*.

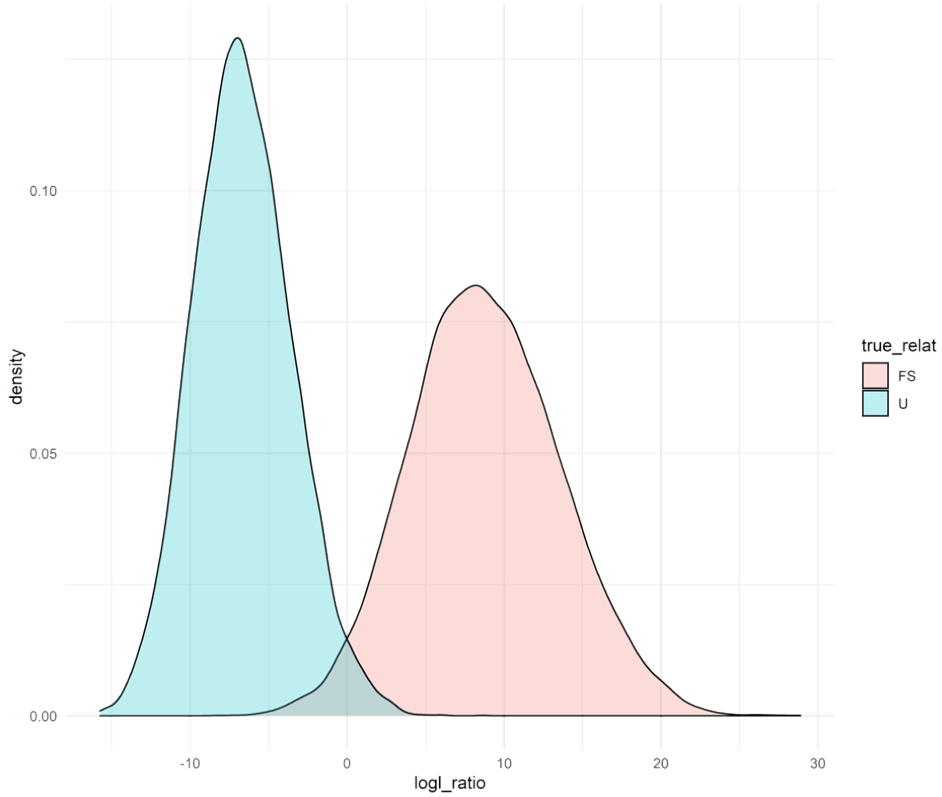

Figure 3. **Log-likelihood ratios distribution for simulated true full-siblings versus unrelated individuals based on Roland Creek *O. mykiss* genotype data. High overlap between full-siblings and unrelated fish, suggests relatively low power to detect highly related individuals.**

Six-hundred-forty of 3,640 (17.5 percent) log-likelihood (*G*) tests for pairwise LD using FSTAT were significant at the α=0.05 level. However, only 15 (<1 percent) tests were significant at the adjusted table-wide level of α=0.00007. The greatest disequilibrium was observed in the collection from Diablo Lake in 2005 (R^2=0.04) and the least in Suiattle River (R^2=0.01) (Table 1). Notably, there was a consistent, negative relationship between sample size and the estimator for pairwise LD, R^2 (Figure 4).

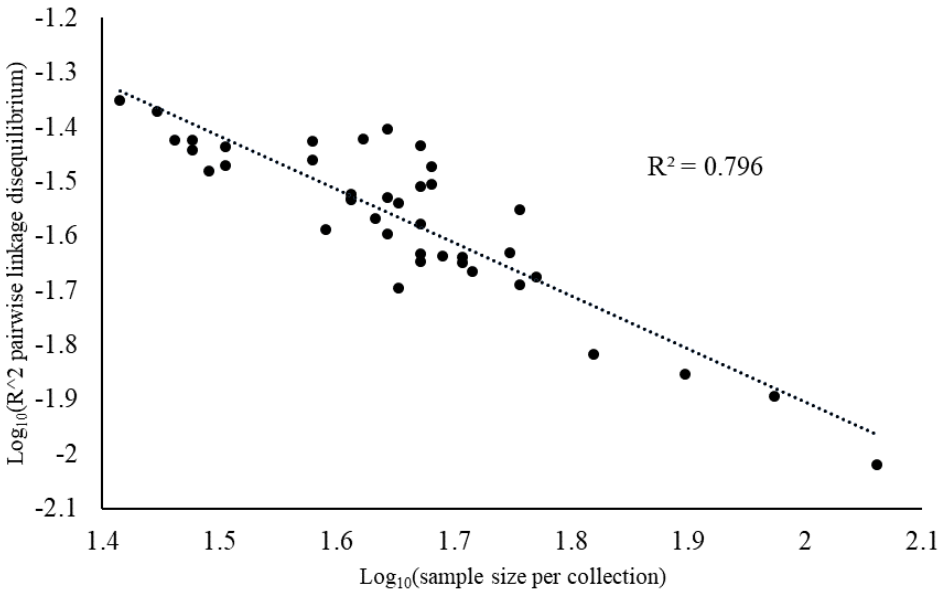

Figure 4. **Scatterplot showing $\log_{10}$-transformed relationship between sample size (n) (*x*-axis) and the R^2 estimator for pairwise linkage disequilibrium between loci (*y*-axis). Strong correlation warrants cautious interpretation of data due to possible bias.**

Fisher's exact tests using POWSIM (Ryman and Palm 2006) which were based on sample sizes and estimated allele frequencies of the dataset, suggested power to detect deviation from genetic homogeneity was 0.32 for F_{ST}=0.001 and was 1.00 for F_{ST}=0.01. The overall estimated proportion of genetic variance explained by population structure (F_{ST}) was 0.094. Log-likelihood (G) tests for population differentiation were significant for each locus and across all loci (P<0.001). Estimates of pairwise F_{ST} ranged from -0.004 between collections from Stetattle Creek in 2009 and 2010 to 0.39 between collections from Ross Reservoir and North Fork Cascade.

Principal components analysis (PCA) of allele frequencies using {adegenet} (Jombart 2011) accounted for a relatively small amount of projected inertia ─ a metric of the magnitude of the explained variance (Cumulative inertia explained by PC-1 through 3=5.924 percent). Genetic population structuring was apparent in scatterplots of the first three principal components (PC). However, several samples from the Baker River collections appeared to be outliers along axes 1 and 2. Notes provided by WDFW suggested the samples could be hybrids with *O. clarkii*. Reanalysis without the Baker River collections only slightly improved projected inertia of the first three PCs (6.095 percent); however, it did improve visualization of genetic population structure (Figures 5 and 6). Specifically, PC-1 (2.215 percent) clearly distinguished the North Fork Cascade River (Collections 24 and 25) from all other collections. PC-2 (2.044 percent) highlighted additional population structuring with collections from upstream of the project boundary at Gorge Reservoir tending to display positive inertia, collections from the Sauk River basin tending to display negative inertia, and remaining collections falling in between. PC-3 (1.836 percent) nearly distinguished Big Creek (Collections 2 and 3) from all other collections.

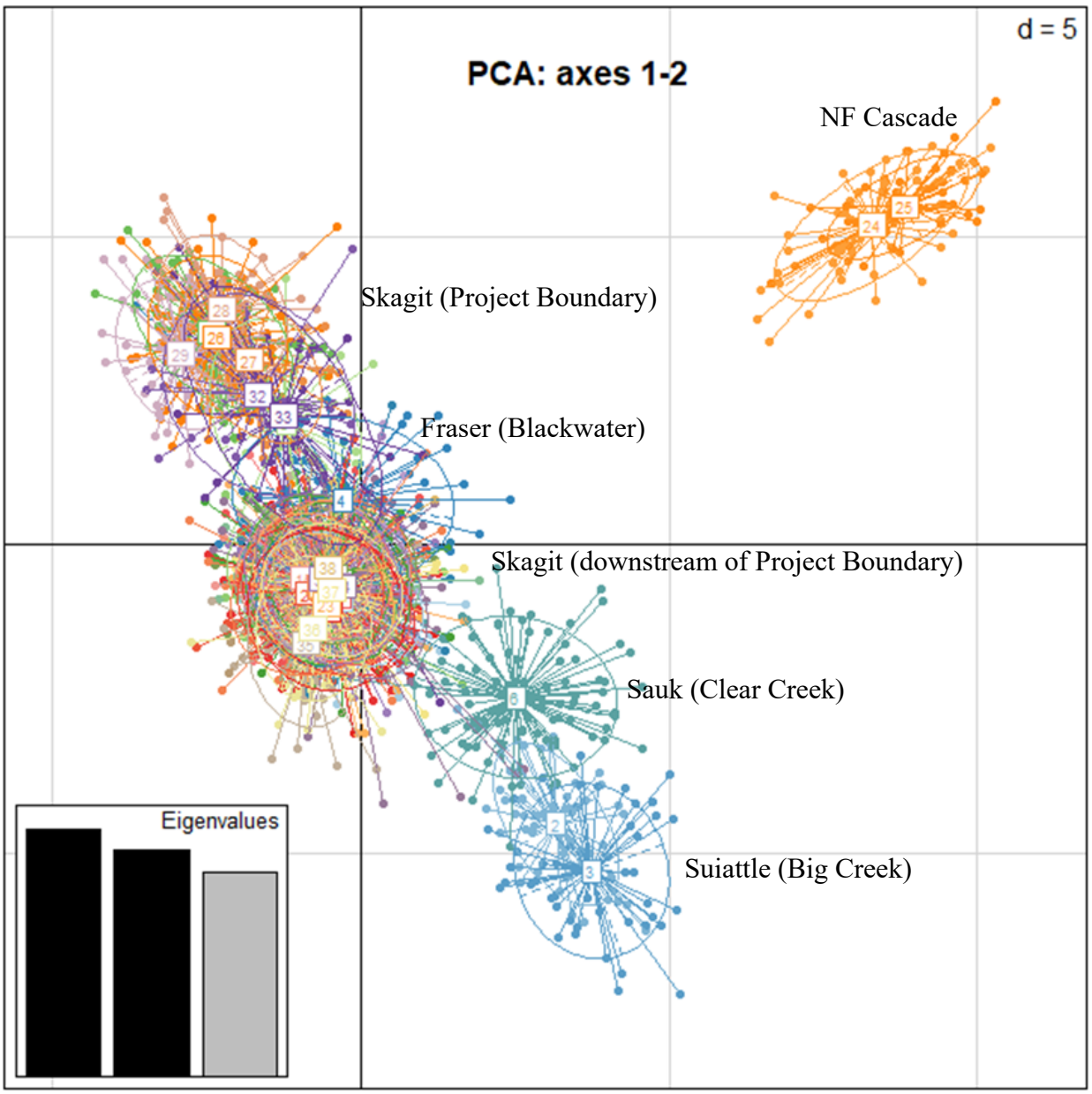

Figure 5. **Scatterplot of genetic principal components 1 (2.215 percent) and 2 (2.044 percent) for all Rainbow Trout collections, excluding samples from the Baker River, estimated using adegenet in program R (Jombart 2011). River basin names are provided to describe the approximate geographical locations of each collection. Numbers at centroids identify the collection number listed in Table 1. Ellipses define 1.5 standard deviations of the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom left corner shows first three eigenvalues.**

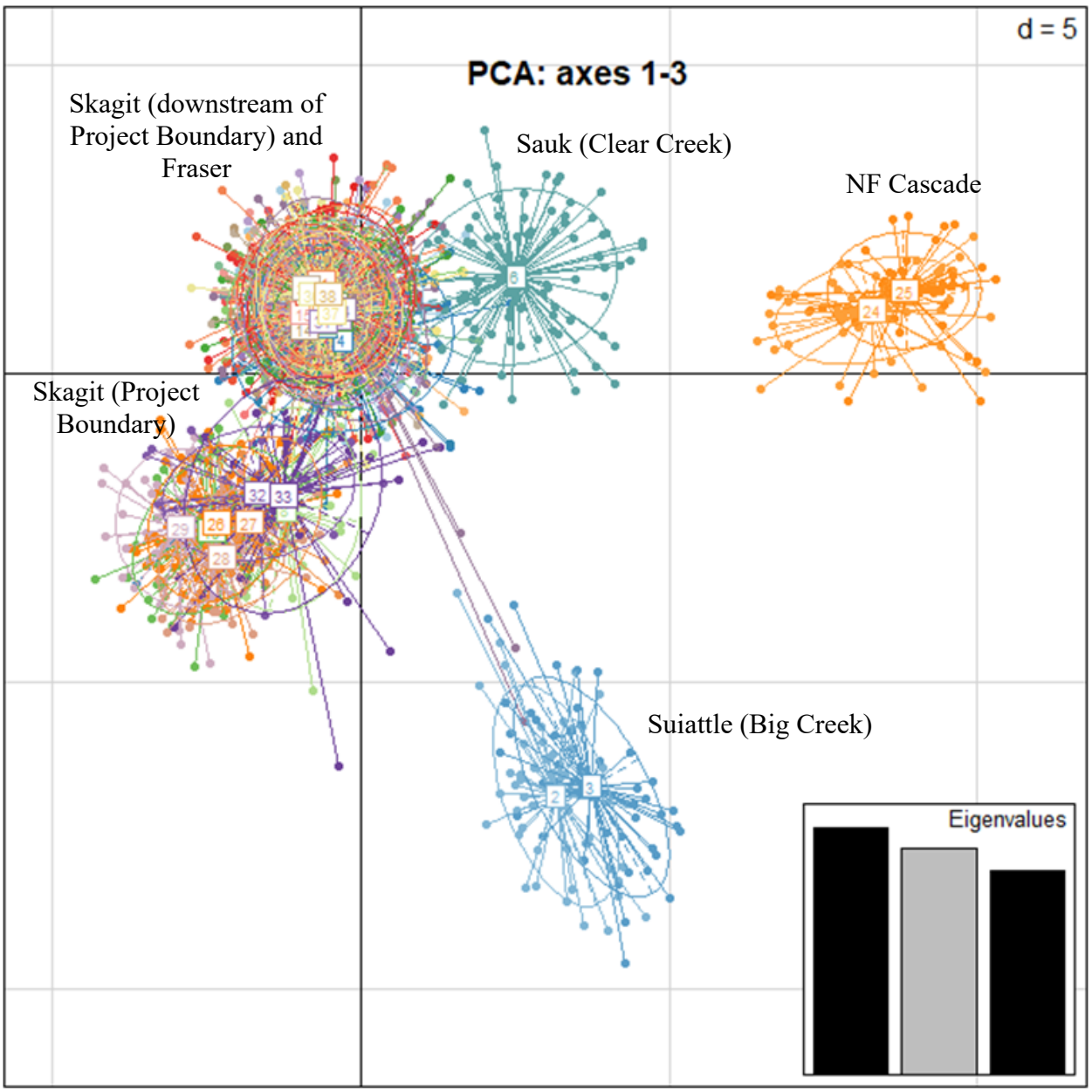

Figure 6. **Scatterplot of genetic principal components 1 (2.215 percent) and 3 (1.836 percent) for all Rainbow Trout collections, excluding samples from the Baker River, estimated using adegenet in program R. River basin names are provided to describe the approximate geographical locations of each collection. Numbers at centroids identify the collection number listed in Table 1. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom right corner shows first three eigenvalues.**

Limiting PCA to collections from upstream of the Project Boundary at Gorge Reservoir identified three samples that might be hybrids between *O. mykiss* and *O. clarkii* based on notes from WDFW; they were subsequently removed from the analysis (09JA0030, 05NG0056, and 10BI0047). Reanalysis without the potential hybrid samples indicated that the first three PCs explained 5.898

percent of the total inertia (Figure 7) and appeared to support some genetic structuring associated with location but statistical support for individual genetic groups was low.

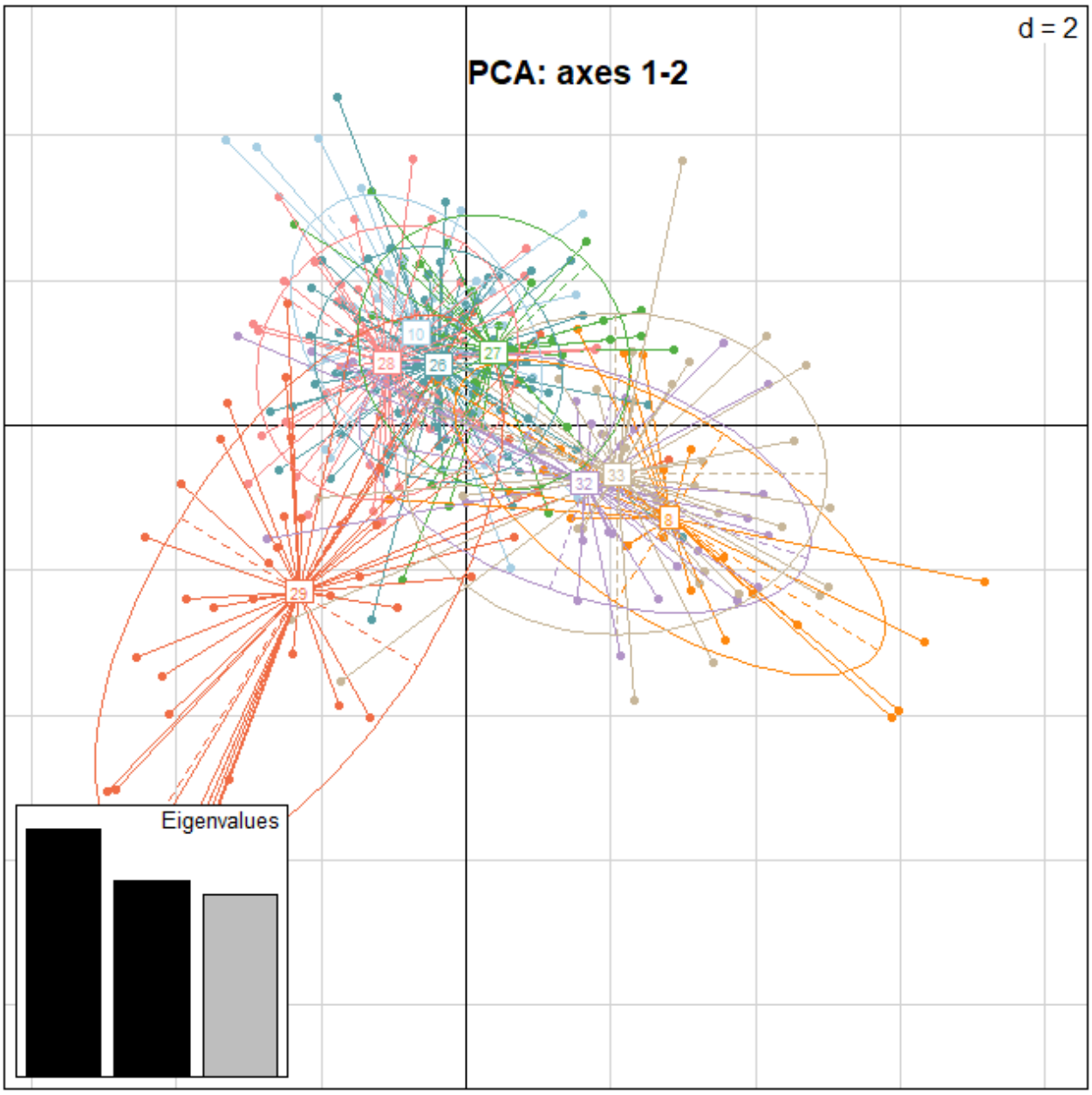

Figure 7. **Scatterplot of genetic principal components 1 (3.870 percent) and 2 (2.028 percent) for all Rainbow Trout collections upstream of the Gorge Lake Project Boundary estimated using adegenet in program R. Numbers at centroids identify the collection number listed in Table 1. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in bottom left corner shows first three eigenvalues.**

3.2.2 Genetic analysis of Bull Trout

Eight hundred and ninety-eight *Salvelinus* spp. genotypes at 16 microsatellite loci were provided by USFWS and WDFW following a request for existing Bull Trout data within Skagit River basin (Table 2). USFWS provided 563 of the genotypes and WDFW provided 335. The standardized markers included *Omm1128*, *Omm1130* (Rexroad et al. 2001), *Sco102*, *Sco105*, *Sco106*, *Sco107*, *Sco109* (WDFW unpublished), *Sco200*, *Sco202*, *Sco212*, *Sco215*, *Sco216*, *Sco218*, *Sco220* (Dehaan and Ardren 2005), *Sfo18* (Angers and Bernachez 1996), and *Smm22* (Crane et al. 2004). The collections were from four Project vicinity tributaries (upper Skagit, Big Beaver, Ruby, and Stetattle creeks) and all three reservoirs (Ross, Diablo, and Gorge lakes). It was unclear which *Salvelinus* taxa or their hybrids were included in the dataset. It was also unclear to what extent collections comprised highly related individuals, which is a common concern in genetic studies of Bull Trout (DeHaan et al. 2014). Furthermore, USFWS communicated that the juvenile collections likely contained related individuals (Smith 2021a).

Sampling location metadata were not provided for USFWS samples, so sampling locations were assumed to be the same as reported in Smith (2010). The stated purpose of the collections from Smith (2010) was to assess genetic variability within and between Bull Trout populations, with sampling methods including a combination of electrofishing, snorkeling, and angling.

No metadata were provided by WDFW other than collection code. Location data were not provided, so samples obtained from within the Project boundary were considered "at large" from reservoirs. The stated purpose of WDFW collections was to characterize the genetic variation of Bull Trout, Dolly Varden, and Brook Trout in the Skagit reservoirs, but no collection methodology was described. The degree to which samples were collected randomly across *Salvelinus* taxa was unknown, including whether any special effort was made to target Bull Trout, Dolly Varden, Brook Trout or whether potential hybrids were targeted or avoided. Sampling considerations are a key concern because targeted collections (i.e., based on morphology) can bias inference into studies of genetic variation.

Quality assurance/quality control (QA/QC) procedures were conducted to obtain a final dataset in which basic population genetic analyses could be reasonably implemented. Duplicate genotypes were observed for sample IDs 12FG008 and 12FG0009, and so sample 12FG0009 was omitted from dataset. All individuals with missing genotypes at three or more loci, which is more than the 14 loci chosen for *O. mykiss* (see above), were removed. This was necessary, however, because the Bull Trout data production appeared to have been conducted in four by four-locus panels (i.e., multiplexes), with many samples missing a single four locus block. Following data QA/QC, 589 samples were retained for analysis (Table 2).

Table 2. Bull Trout microsatellite dataset collection summary.

Collection Location	River	Life Stage	WDFW Code	Number Collected	Number Evaluated	Number Analyzed
Upper Skagit River	Skagit	adult		16	14	14
Big Beaver Creek	Skagit	adult		21	21	21
Ruby Creek	Skagit	adult		43	41	41
Stetattle Creek	Skagit	juvenile		59	41	41
Lower Goodell Creek	Goodell	juvenile		60	46	46
Upper Goodell Creek	Goodell	juvenile		19	8	8
Bacon Creek	Bacon	juvenile		61	24	24
Cascade River	Cascade	juvenile		39	33	33
Marble Creek	Cascade	juvenile		28	18	18
Kindy Creek	Cascade	juvenile		30	17	17
Illabot Creek	Illabot	juvenile		70	60	60
South Fork Sauk River	Sauk	juvenile		59	54	54
Downey Creek	Sauk	juvenile		58	44	44
Ross Lake	Skagit	unk	12FG	54	47	42
Ross Lake	Skagit	unk	15OW	28	22	20
Diablo Lake	Skagit	unk	13PS	40	29	8
Gorge Lake	Skagit	unk	14ST	27	5	3
Gorge Lake	Skagit	unk	19NL	109	22	0
Sulfur	Skagit	unk	050F	4	4	4
Sulfur	Skagit	unk	06JQ	28	23	23
Diablo, Gorge Lake	Skagit	unk	11LX	45	16	9
			Total	**898**	**589**	**530**

Statistical power was estimated to correctly classify related individuals. This was completed to evaluate the possible effects of violations of sampling assumptions common to the analysis of Bull Trout microsatellite data; specifically, that highly related individuals (i.e., full siblings) are common in samples of Bull Trout (particularly samples of juveniles), which can result in pseudo-replication of genotypes and thus biased estimates of allele frequencies (DeHaan et al. 2014). Statistical power of pedigree analysis to identify parent-offspring and full-sibling pairs was conducted using the close kin mark recapture R package CKMRSim version 0.1 (Anderson 2019; Formerly NMFS Southwest Fisheries Science Center). During pedigree analysis, all samples are examined for relatedness in pairwise comparisons, and so the false positive rate (FPR) increases exponentially with sample size. It is recommended to choose a FPR threshold approximately 10 times smaller than the reciprocal number of pairwise comparisons. In this case, $1.4\ e^{-5}$ was the target FPR used to evaluate the power to detect relatives (i.e., $0.10 * (100 \times 100)^{-1} = 0.000014$). To simulate the related and unrelated individuals needed to estimate power of pedigree analysis, all collections from the Skagit River dataset were used. The distribution of log-of-the-odds (LOD) values are shown in Figure 8 for full-sibling pairs. The expected distributions overlap between full-sibling and unrelated individuals, which means that choosing a FPR that provides reasonable assurance no unrelated pairs will be falsely called full-siblings will result in an undesirably high false negative rate (FNR). For Skagit River Bull Trout, a LOD value = 8.0 (corresponding to FPR

= 1.4 e^{-5}) results in a FNR = 0.15, meaning approximately 15 percent of true full-sibling comparisons would be misclassified as unrelated with an α=0.05 as the typical standard.

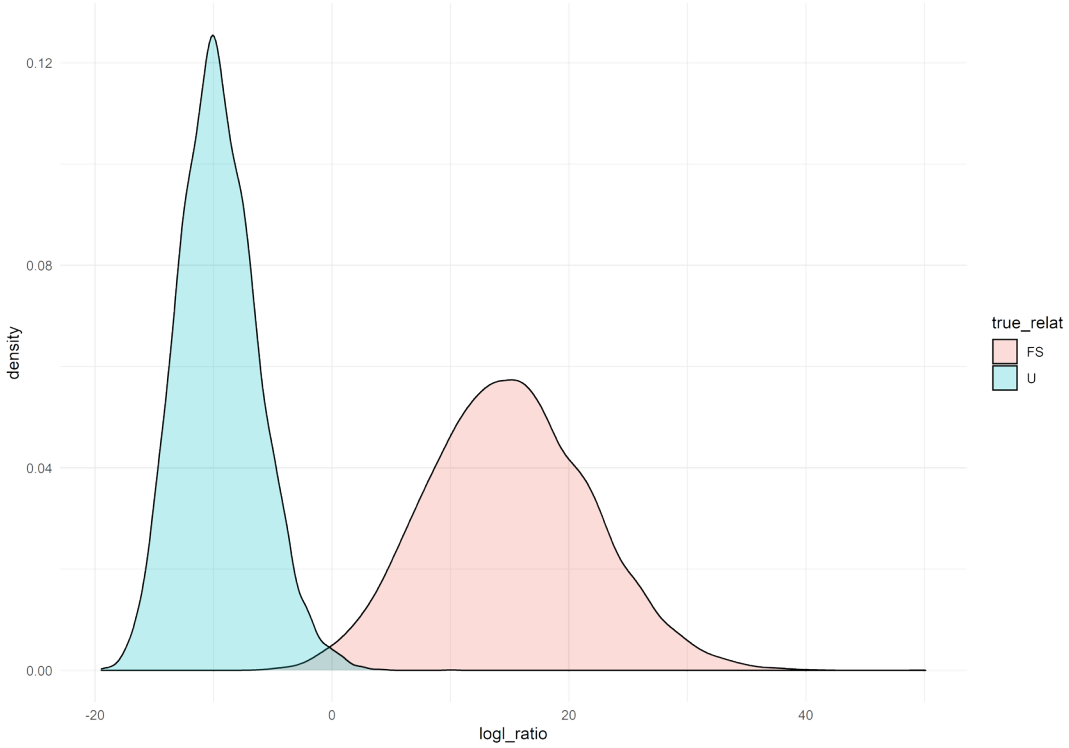

Figure 8. **Log likelihood ratios distribution for simulated true full siblings versus unrelated individuals based on Skagit River S. confluentus genotype data. High overlap between full-siblings and unrelated fish, suggests relatively low power to detect highly related individuals.**

Note that these estimated rates were based on all the available collections (n=530), which would likely overestimate power for studies of "real-world" populations. A more realistic evaluation would consider collections from a single Project Boundary tributary, as opposed to considering potential comparisons between unrelated individuals across the entire Skagit River basin. Therefore, the analysis was repeated, using only collections from Big Beaver, Ruby, and Stetattle creeks in the Project Boundary. The FNR estimated for Big Beaver, Ruby, and Stetattle collections were 0.857, 0.868, and 0.95, respectively, meaning pedigree analysis is expected to result in more false relationship assignments than true assignments.

Understanding power to detect related individuals helped identify individual samples that might need to be removed from analysis to reduce violation of sampling assumptions. COLONY (Jones and Wang 2010) was used to screen collections for full sibling families, and based on power estimates above, applied probability of inclusion = 1.0 and a probability of exclusion = 0.99 to accept family classifications. Inclusion probability gives the probability that all individuals (in that family) are indeed full siblings from the same family. Exclusion probability is the probability those individuals are full siblings, and no other individuals are full siblings with this family. There is no accepted convention or criterion for identifying and removing related individuals from a dataset, although the criteria used here are more stringent than those referenced in literature pertaining to

this Bull Trout dataset (Smith 2010). All full siblings but one[3] were omitted from identified families within collection.

Like for Rainbow Trout, PCA of allele frequencies (adegenet package) was used to examine genetic variation among collections. Data modeling suggested retention of approximately 15 PCs and 5 discriminant functions (*k*) would result in reliable partitioning of genetic variation among group clusters. With the number of genetic group clusters fixed at two (i.e., *k*=2), samples partitioned into genetic groupings associated with Diablo/Gorge lakes and all other samples. With an additional cluster allowed (*k*=3), individuals partitioned into (1) Project Boundary tributaries and some reservoir samples; (2) Project Boundary reservoir samples; and (3) samples from below Gorge Dam. With the allowance of fourth and fifth genetic clusters (*k*=4 and *k*=5), Project Boundary reservoir samples became split among the newly allowed clusters. No further refinement of Project Boundary samples was observed at higher numbers of clusters. A visualization of the k-means clustering at k=5 is shown on Figure 9. Clusters 3, 4 and 5 were predominantly individuals collected from Diablo and Gorge lakes. Cluster 1 were Project Boundary tributary collections and contained a majority of Ross Lake samples. Cluster 2 were individuals collected from below Gorge Dam.

As mentioned, collections submitted by WDFW were a part of evaluations intended to assess hybridization among Bull Trout, Dolly Varden and Brook Trout. Reports pertaining to data noted that hybrids were observed within these collections (e.g., Small et al. 2013; Small et al. 2016). Clusters 3, 4, and 5 were unable to be directly ascribed to hybridization among individuals or genetic introgression because: (1) taxon-diagnostic alleles among taxa were unknown; (2) sample IDs for individuals WDFW considered hybrids were not provided; (3) the methods by which WDFW determined individuals to be hybrids was not provided; and (4) the selection strategy (if any) of field personal collecting individuals "at large" from reservoirs was also not provided.

[3] The presence of multiple representatives from the same family skews allele frequencies from true population proportions, creating a bias. Removing all but one sibling removes this bias.

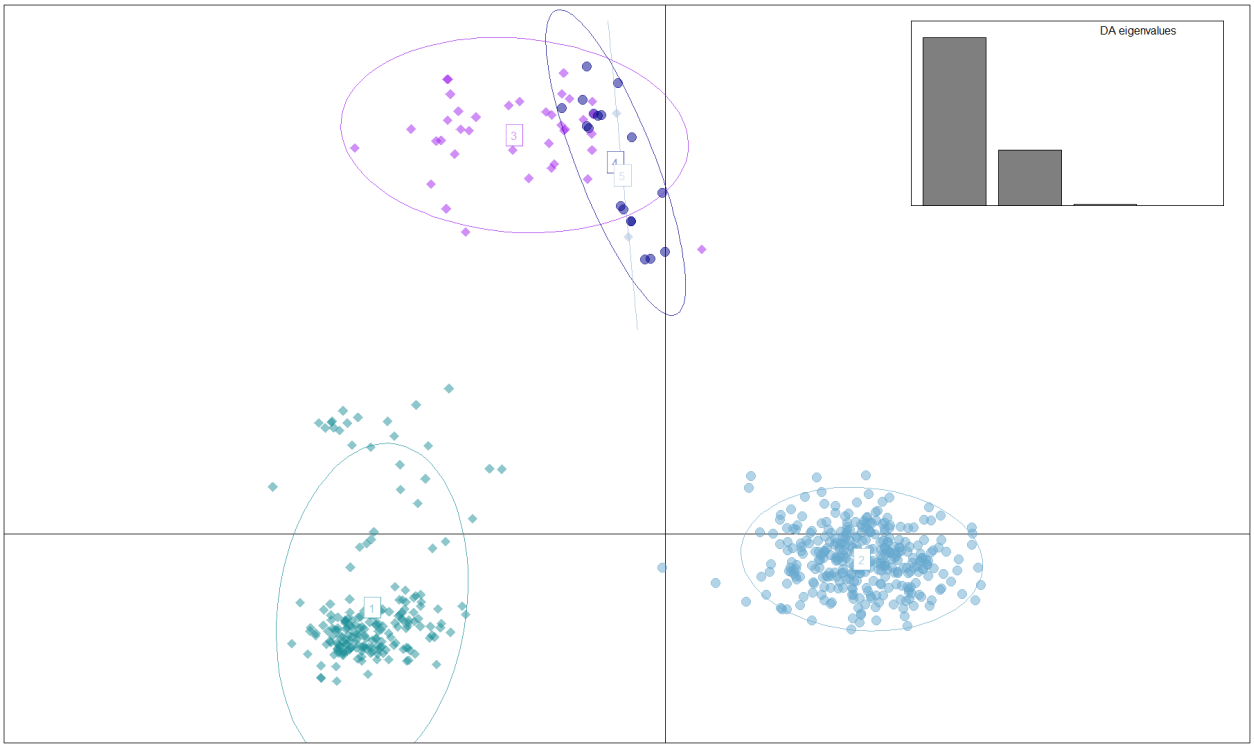

Figure 9. **Visualization of k-means clustering analysis at k=5 for Bull Trout individuals in dataset for 1st and 2nd principal axes. Ellipses define 1.5 standard deviations for the inertia (variance) around each centroid, where ellipses that overlap more are less distinct. Scree plot in upper right corner shows first three eigenvalues. Cluster 1 were Project Boundary tributary collections and contained a majority of Ross Lake samples. Cluster 2 were individuals collected from below Gorge Dam. Clusters 3, 4 and 5 were predominantly individuals collected from Diablo and Gorge Lakes.**

Small sample sizes of *Salvelinus* spp. (median=26) relative to *O. mykiss* (median=45) highlighted limitations associated with balancing precision and bias. For instance, collections with fewer than 25 individuals are typically not recommended for analyses using microsatellite data, however, adopting this criterion for the *Salvelinus* spp. dataset would have resulted in exclusion of about 50 percent of Project Boundary Bull Trout collections from an already sparse dataset. The genetic groupings shown in Figure 10 also underscore the challenges associated with choosing which fish to retain in any given collection due to genetic admixture. All individuals in clusters 3, 4, and 5 were considered potentially admixed and omitted from the dataset prior to estimating genetic summary statistics for each collection. The current sample size threshold pertaining to Bull Trout collections may be modified based upon future discussions of hypotheses and research questions with the Expert Panel and LPs. The resulting final dataset comprised n=530 samples (Table 2). The genotypes are saved in GENEPOP format and are available upon request.

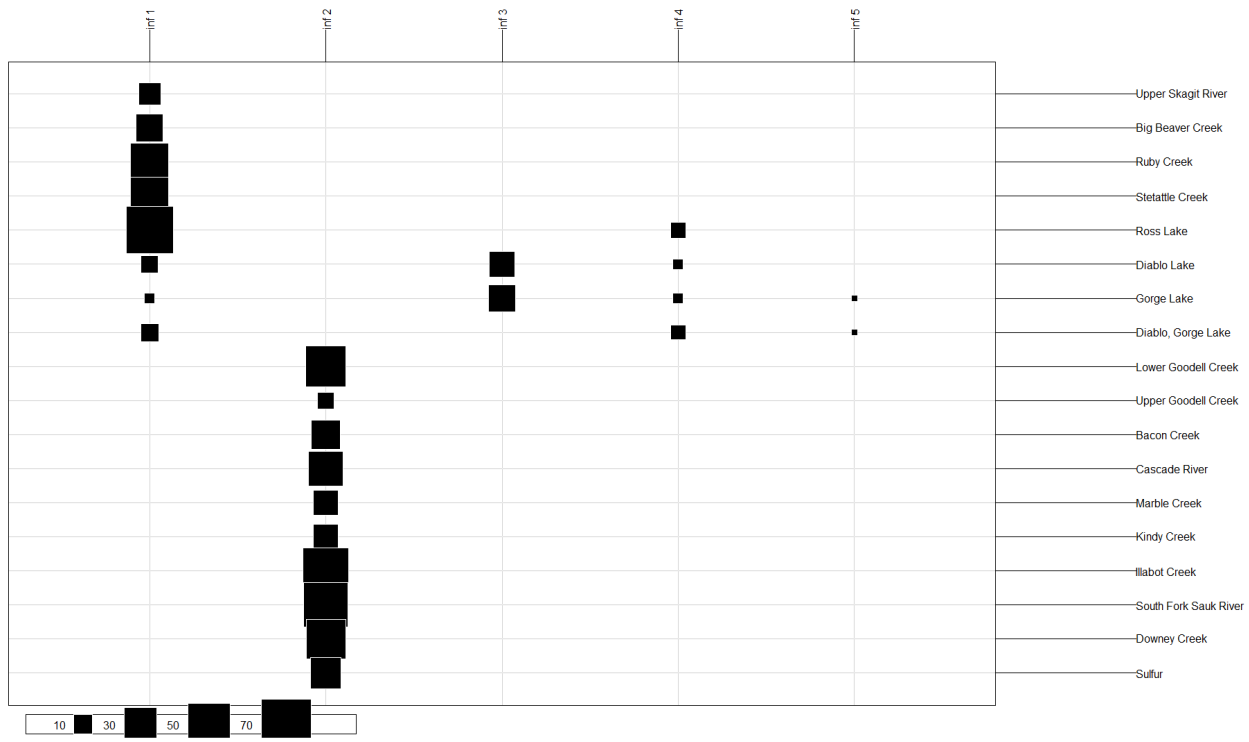

Figure 10. **Genetic clusters visualized in Figure 9 aligned to each Bull Trout collection in dataset. Size of boxes is scaled by sample count. Genetic clusters are organized by geographic location with upper Skagit collections at the top and lower Skagit at the bottom.**

Heterozygosity (gene diversity) in the Bull Trout collections ranged from 0.337 to 0.467 within collections from Project Boundary tributaries (above Gorge Dam) and was 0.473 in the Ross Lake collection (Table 3). The collections from within the Project Boundary (above Gorge Dam) had lower heterozygosity than the collections from below Gorge Dam (Chi-square p-value = 0.0027). Our attempt to reduce violation of HWE appeared successful, as mean F_{IS} across all collections was not statistically different from 0.00 (F_{IS}=0.008, 95 percent CI: -0.024-0.051). Each Project vicinity tributary collection (upper Skagit, Big Beaver, Ruby, Stetattle) did not deviate significantly from expectations. The Ross Lake collection was not in HWE, along with potentially several collections from below Gorge Dam, particularly Bacon Creek and Illabot Creek. LD was measured using log-likelihood (G) tests for all pairwise locus comparisons. Of the 1,680 comparisons (overall collections), 271 were significant at the α=0.05 level. No Project Boundary tributary collections (above Gorge Dam) had statistically significant LD tests using the adjusted table wide significance level α=0.0003. The Ross Lake collection had 11 significant LD test out of 120. The greatest number of significant log-likelihood tests was observed for the Illabot Creek collection (16).

The estimated proportion of genetic variance explained by population structure (F_{ST}) across all Bull Trout collections was 0.188, and 0.03 among Project Boundary tributary collections, only. Pairwise log-likelihood (G) tests for population differentiation were not statistically significant between upper Skagit, Big Beaver, and Ruby Creek collections (adjusted nominal level 5 percent). Upper Skagit River collection was not differentiated from Ross Lake collection, but Ross Lake collection was differentiated from both Big Beaver and Ruby Creek collections. The Stetattle

Creek collection was differentiated from all other Project Boundary collections. Note that the Marble Creek collection was not differentiated from any collection in the dataset except South Fork Sauk River. This seemed anomalous, so results that follow exclude consideration of Marble Creek collection. All Project Boundary collections (above Gorge Dam) were differentiated from below Gorge Dam collections. Recall, F_{ST} is the proportion of genetic variation that is attributable population subdivision with F_{ST}=0.00 reflecting no differences and F_{ST}=1.00 reflecting complete differentiation (i.e., all genetic diversity is partitioned among subpopulations). The F_{ST} estimated (pairwise) between Project Boundary collections are shown in Table 4). For context, F_{ST} estimated from comparisons between Project Boundary collections with those from below Gorge Dam ranged from a low of 0.207 to a high of 0.397.

Table 3. Summary statistics for samples collected from Bull Trout in the Skagit River basin.

Collection	Sample Size	F_{IS}[1]	H_s[2]	MNA[3]
Upper Skagit River	14	0.080	0.467	5.00
Big Beaver Creek	21	0.042	0.410	4.44
Ruby Creek	41	-0.021	0.384	4.75
Ross Lake	62	0.105	0.473	7.16
Stetattle Creek	41	-0.078	0.337	2.94
Goodell Creek	54	0.046	0.647	6.97
Bacon Creek	24	0.038	0.678	7.56
Illabot Creek	60	-0.050	0.634	7.44
Cascade River	33	0.033	0.662	8.19
Marble Creek	18	-0.080	0.679	6.94
Kindy Creek	17	0.016	0.689	7.19
S.F. Sauk River	54	-0.032	0.656	8.31
Downey Creek	44	0.010	0.709	9.88
Sulfur	27	0.035	0.607	6.13

1 $F_{IS:}$ estimated deviation from HWE proportions.
2 $H_{S:}$ estimated expected heterozygosity within sub-populations (i.e., gene diversity).
3 MNA: is the mean number of alleles observed over all loci.

Table 4. Table of pairwise estimates of F_{ST} from Project Boundary Bull Trout collections.

	Upper Skagit River	Big Beaver Creek	Ruby Creek	Ross Lake
Big Beaver Creek	0.001			
Ruby Creek	0.028	0.014		
Ross Lake	0.023	0.043	0.061	
Stetattle Creek	0.068	0.030	0.034	0.105

4.0 SUMMARY

4.1 Rainbow Trout

4.1.1 What genetic data are available?

Review of 2,697 preexisting microsatellite genotypes provided by WDFW highlighted opportunities and gaps to drawing new inference about population genetic characteristics of Rainbow Trout affected by FERC relicensing. Any new inference gleaned from these existing data is naturally limited by the design of the original research that estimated the genotypes. In general, inferences drawn from tens of presumably neutral microsatellites ─ in this case 14 ─ are naturally limited to basic descriptions of genetic diversity and population structure because of required analytical assumptions (see Section 2.1 of this memo). Likewise, new inference is also limited by the spatial, temporal, and ecological scope of the original sampling. The statistical populations to which any new inference applies should therefore be defined and agreed upon by the study team prior to determination of how this existing data addresses outstanding conservation questions or how new sampling might be most effective.

4.1.2 What types of inference can be drawn from the existing data?

The type of inference that can be drawn from analysis of the existing data could be limited to basic descriptions of genetic diversity and population structure. Genetic structure was apparent in the analyzed collections. The overall estimated proportion of genetic variance explained by population structure (F_{ST}) was 0.094, and the PCA appeared to provide some evidence that geography affects structure. Nevertheless, specific hypotheses about how current or historical geography affects structure were not tested (e.g., isolation-by-distance versus historical hydrogeological connectivity with the Fraser River). The proportion of variation that can be explained by hybridization with either *O. clarkii* or HOR fish were not directly addressed. Firstly, notwithstanding completely diagnostic makers, the set of microsatellites has limited power due to the number and diversity of markers (Vaha and Primmer 2006). Secondly, the question of hatchery introgression was addressed by Pflug et al (2013). Pflug et al (2013) used a liberal hybrid cut off threshold of 20 percent introgression from HOR fish and stated, "the juvenile collections showed the presence of presumptive [HOR] hybrids in all collection areas". All these factors are important to consider because observed patterns of diversity may not reflect natural genetic drift and gene flow within and among natural-origin *O. mykiss*.

4.1.3 Analytical considerations

Rainbow Trout genotypes analyzed in this memo were compiled in a way that attempted to reduce biases common to microsatellite datasets (See Section 2.1 of this memo). The approach attempted to increase any biologically meaningful signal by reducing noise associated with (1) hybridization with *O. clarkii*; (2) small sample sizes; (3) missing and erroneous data; and (4) violation of HWE and LD.

Absence of hybridization with *O. clarkii* is a common assumption of *O. mykiss* genetic structure analysis because many classical analyses assume that genetic variation is a function of effective population size (N_e) and migration (m) within a single taxon (mutation is assumed negligible). For example, the equation $F_{ST} \approx 1/(4N_em + 1)$ used to describe the strength of gene flow on genetic divergence assumes the subpopulations contributing migrants comprise only *O. mykiss*. Although

the dataset analyzed contained genotypes at genetic markers apparently diagnostic for *O. clarkii*, their diagnostic properties were unknown because positive control genotypes for nonhybridized *O. clarkii* were not provided. In practice, this limits the ability to estimate evolutionary relationships among subpopulations, which are typically assumed to be a function of genetic drift and gene flow within *O. mykiss*, as opposed to ongoing genetic introgression of alleles from *O. clarkii*.

Small sample sizes can result in imprecise estimates of allele frequencies and thus weak biological inference. There is no accepted threshold or rule for sample sizes because sampling needs vary by hypotheses, research questions, and marker types (Landguth et al. 2010). For the *O. mykiss* dataset, the recommendation of Hale et al. (2012) was adopted – those 25 individuals are typically enough to accurately estimate allele frequencies using microsatellites. Nevertheless, others have cautioned that when allelic diversity per population is high, as is the case with microsatellites, sampling effort may need to surpass 80–100 individuals to have a high probability of detecting low frequency alleles (Ott 1992: Seeb et al. 2007). Yet, other studies have reported that for isolated populations (n=8,000 individuals), 20 individuals genotyped at 6 microsatellites could produce an accurate allele frequency distribution (Siniscalco et al. 1999). For the present dataset, excluding collections of n<25 provided high power (1.00, $P<0.05$) to detect moderate differentiation ($F_{ST}=0.01$), but low power (0.32, $P<0.05$) to detect low differentiation ($F_{ST}=0.001$).

Like questions of sample size, there is no accepted threshold or rule for treating missing and erroneous microsatellite data. Using computer simulations, Reeves et al. (2016) estimated that for every 1 percent of missing genotypic data, 2 to 4 percent fewer correct population assignments can be expected. They recommended limiting the percentage of missing data to approximately 2 percent, unless a greater amount can be justified. Therefore, all individuals with missing data at two or more loci (approximately 6 percent), which was the most missing data that could be accommodated in a dataset of 15 microsatellites without allowing only individuals with complete genotypes to be included, were removed. Regarding genotyping errors, 1 to 2 percent fewer correct population assignments are expected for every percentage increase in genotyping error (Reeves et al. 2016). Although there are a variety of computer programs available to estimate the frequency of genotyping errors in a dataset, most techniques are based on conformance of genotypes to HWE proportions.

HWE and absence of LD are common assumptions of population genetic analyses for a variety of reasons. Metrics of HWE, for example, can provide insight to mating systems of populations (i.e., inbreeding) or to data quality problems like genotyping issues, overrepresentation of families, etc. The data compilation method for *O. mykiss* attempted to reduce violations of HWE that might result from data quality problems with the goal being to increase chances that metrics reflect the actual underlying mating system. The compilation method of removing markers with consistent deviation from HWE (i.e., *One-14*) and combining collections from the same tributaries but in different years that produced fewer deviations from HWE resulted in a dataset with a lower overall F_{IS} than the original dataset, though the decrease was not statistically significant. For clarity, lower F_{IS} suggests the compilation method succeeded in reducing deviations from HWE.

Regarding LD, a consistent negative relationship between sample size (n) and the estimator R^2, which could indicate that, on average, sampling was not sufficient to obtain unbiased estimates of LD, was observed. The potential bias presents a challenge to data interpretation. For example, the

collection from Diablo Lake contained the highest LD (R^2=0.04) but also one of the smallest sample sizes in the entire dataset (n=26). This presents a data interpretation challenge because the collection from Diablo Lake also contained apparent hybridization with *O. clarkii*, as noted by WDFW, which is expected to cause an increase LD associated with genetic admixture between genetically dissimilar populations. It is therefore uncertain whether high LD in Diablo Lake is associated with something biologically meaningful, like hybridization, or is simply an artifact of bias associated with small sample size.

4.1 Bull Trout

4.1.1 What genetic data are available?

Review of 898 microsatellite genotypes provided by USFWS and WDFW highlighted potential data gaps and opportunities to drawing new inference about population genetic characteristics of Bull Trout affected by FERC relicensing. Like Rainbow Trout, there are limitations to any new inference gleaned from these existing data. Most of the samples evaluated appear to have been previously analyzed by Smith (2010), Small et al. (2013), and Small et al. (2016) and so the collective scope of inference is somewhat ambiguous and should be discussed by the study group prior to making decisions on what questions can be answered by the existing data and what new samples need to be collected. Yet, it was uncertain which samples were evaluated in common among all three studies because sample identification were not included in the original reporting. The purpose for tissue sampling varied by collection. Smith (2010) stated that study's collection purpose was to assess genetic variability within and between Bull Trout populations of the Skagit River Basin and subbasins in the vicinity of City Light's Skagit Hydroelectric Project. The stated purpose of WDFW collections was to characterize the genetic variation of Bull Trout, Dolly Varden, and Brook Trout in the Skagit reservoirs. Given the sampling objectives differed for collections and multiple *Salvelinus* taxa (or hybrids) may have been incorporated into collections, it was challenging to compile a Bull Trout dataset. As with the *O. mykiss*, this approach focused on compiling the dataset that reduced violations of basic assumptions common to the analysis of microsatellites and, in general, to support basic inferences about genetic population structure of Bull Trout in the Skagit River basin. Nevertheless, as a working definition, the scope of inference of the analyses in this memo might apply to the few generations of naturally reproducing subpopulations of Bull Trout living within and downstream of the Project reservoirs and that might have been affected by key factors such as hydropower management (e.g., isolation) and hybridization from 2005 to 2015.

4.1.2 What types of inference can be drawn from the existing data?

Similar to Rainbow Trout, the types of inference that can be made about Bull Trout are likely limited to the few generations and subpopulations sampled by the original studies. The estimated proportion of genetic variance explained by population structure (F_{ST}) across all Bull Trout collections was 0.19. Project Boundary tributary collections were more similar to each other than any were to collections from below Gorge Dam, with collections from Ross Lake tributaries not statistically different. The Stetattle Creek collection was genetically differentiated from Ross Lake tributary collections. The "at large" Ross Lake collection was genetically differentiated from all Project Boundary collections except upper Skagit River. Pairwise F_{ST} estimates comparing Project Boundary collections with those from below Gorge Dam were relatively large in magnitude, with the minimum estimate observed being 0.207.

4.1.3 Analytical considerations

Like descriptions above for *O. mykiss*, evaluation of *Salvelinus* spp. data had to contend with inclusion of potential hybridized individuals, small collection sample sizes, missing and erroneous data, and violation to genetic equilibria. Data quality recommendations as noted above for Rainbow Trout were also applied to Bull Trout. The sample size threshold was reduced to retain collections from within the Project Boundary. Additionally, the missing genotype data threshold was increased to 25 percent and *Salvelinus* spp. samples that appeared ambiguous were omitted from summary statistic estimations. Lastly, while there was limited power to identify related individuals within collections, full-sibling families that were inferred using established methods were reduced in size. These steps resulted in a dataset that largely conformed to genetic equilibrium expectations, which was an improvement in data quality.

If genotypes at each genetic marker location (locus) occur at a frequency expected by random associations of alleles (a function of the allele frequency), genotypes are said to be in HWE, or alleles within loci are uncorrelated (statistically independent). Many phenomena may cause deviations for HWE expectations (e.g., null alleles, inbreeding, population mixing), with the deviation quantifiable using an analysis of variance approach. F statistics partition the reduction (or excess) in heterozygotes relative to HWE. One component, F_{IS}, is the individual relative to the subpopulation (collection). Globally across all collections, the mean F_{IS} observed was low (F_{IS} =0.008) and the 95 percent confidence interval overlapped zero. Further, all Project Boundary tributary collections were statistically consistent with HWE. LD quantifies the correlation of alleles between loci. LD is a useful quantity to measure, as the pattern of LD in the genome is influenced by population history, the breeding system, the pattern of geographic subdivision, natural selection, gene conversion, and mutation (Slatkin 2008). No Project Boundary tributary collections had statistically significant LD tests. The Ross Lake collection was not in HWE and had 11 statistically significant LD tests.

From a genotype frequency perspective, population structure results in an inbreeding like effect (a reduction in heterozygotes expected relative to HWE) due to nonrandom mating among all individuals analyzed. As such, measuring the deviation from HWE expectations due to population structure acts as a measure of genetic distance between two populations (or collections in this case). The degree of deviation is quantified using another F statistic, F_{ST}, the component of genetic variance within subpopulation (collection) relative to total population (paired collections being tested). For example, as mentioned above, Project Boundary tributary collections were more similar to each other than any were to collections from below Gorge Dam.

5.0 CONCLUSION

Existing microsatellite genotypes and metadata were available for Rainbow Trout (Pflug et al. 2013) and Bull Trout (Smith 2010; Small et al. 2013, 2016 and 2020) sampled within the Project Boundary and in the Project vicinity at different times and locations over the last twenty years. Genetic structure was apparent in both taxa and geography appeared to be an important factor in how the variation was distributed. Hypotheses about what created the structure were not tested. How structure was affected by hatchery introgression, hybridization, historical hydrogeological connectivity, genetic drift, etc. may be included in Expert Panel discussions and FA-06 Reservoir Fish Genetics Study reporting. Further, whether these data are sufficient to inform topics of interest communicated by LPs will be considered as part of Expert Panel discussions when topics are transformed to specific scientific question that can be applied to these data. Additionally, there are other genetic datasets that have been collected and analyzed by researchers that were not considered in this memo. As stated above, the data analyzed in this tech memo were recognized as potentially helpful for describing the baseline genetics of native fish in the RSP. Yet, unconsidered data could be brought to bear during formulation of research questions or designs for additional field sampling. Datasets not included here include genotypes for different genetic markers, including mitochondrial DNA haplotypes for Bull Trout inside of the Project Boundary (Smith 2021b) and for outside of the Project Boundary (Taylor and May-McNally 2015). There are also single nucleotide polymorphism (SNP) genotypes for 30 Rainbow Trout sampled from the drawdown zone of Gorge Lake (Small et al. 2020).

6.0 GLOSSARY

Allele frequency: a measure of the relative frequency of an allele at a genetic locus in a population.

genetic drift: random changes in allele frequencies in a population between generations due to sampling individuals that become parents and binomial sampling of alleles during meiosis.

Gene flow: exchange of genetic information between subpopulations.

Genetic population structure: systematic difference in allele frequencies between subpopulations in a population resulting from non-random mating between individuals.

Iteroparous: a reproductive strategy characterized by multiple reproductive cycles over the course its lifetime.

Microsatellite: tandemly repeated DNA consisting of short sequences of 1 to 6 nucleotides repeated approximately 5 to 100 times.

Overlapping generations: a breeding system where sexual maturity does not occur at a specific age, or where individuals breed more than once, causing individuals of different ages to interbreed in a given year.

Selectively neutral: an allele that is not under selection because it has no effect on fitness.

Statistical power: probability of obtaining a statistically significant result given a true effect occurs in a population.

Subpopulations: groups of individuals within a population delineated by reduced gene flow with other groups.

7.0 REFERENCES

Allendorf, F. W., and S. R. Phelps. 1981. Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences 38:1507-1514.

Anderson, S. 2019. CKMRsim package, efficient sampling for close-kin mark recapture power estimation. Available at https://github.com/eriqande/CKMRsim

Angers, B. and L. Bernachez. 1996. Usefulness of heterologous microsatellites obtained from brook charr, *Salvelinus fontinalis* Mitchill, in other *Salvelinus* species. Molecular Ecology 5:317–319.

Banks, M. A., M. S. Blouin, V. K. Baldwin, H. A. Rashbrook, H. A. Fitzgerald, S. M. Blankenship, and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in Chinook Salmon (*Oncorhynchus tschawytscha*). The Journal of Heredity 2.

Buchholz, W. G., S. J. Miller, and W. J. Spearman. 2001. Isolation and characterization of chum salmon microsatellite loci and use across species. Animal Genetics 32.

Cairney, M., J. B. Taggart, and B. Hoyheim. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Molecular Ecology 9:2155-2234.

Crane, P. A., C. J. Lewis, E. J. Kretschmer, S. J. Miller, W. J. Spearman, A. L. DeCicco, M. J. Lisac, and J. K. Wenburg. 2004. Characterization and inheritance of seven microsatellite loci from Dolly Varden, *Salvelinus malma*, and cross-species amplification in Arctic char, S. alpinus. Conservation Genetics 5:737-741

DeHaan P. W. and W. R. Ardren. 2005. Characterization of 20 highly variable tetranucleotide microsatellite loci for bull trout (*Salvelinus confluentus*) and cross amplification in other Salvelinus species. Mol. Ecol. Notes 5: 582-585.

DeHaan, P., B. J. Adams, and M. C. Nelson. 2014. Fine-Scale Population Structure Analysis and Genetic Population Assignments of Wenatchee River Sub-Basin Bull Trout.in D. o. t. I. U. S. F. W. Service, editor., Longview, WA.

Excoffier, L., and M. Slatkin. 1998. Incorporating genotypes of relatives into a test of linkage disequilibrium. Am. J. Hum. Genet. 171-180.

Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485-486.

Goudet, J., M. Raymond, T. de Meeüs, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144: 1933-1940.

Guo, S. and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372.

Hale, M. L., T. M. Burg, and T. E. Steeves. 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population Is enough to accurately estimate allele

frequencies. PLoS ONE 7:1-10.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal 732 components: a new method for the analysis of genetically structured populations. BMC 733 Genet. 11:94. BioMed Central Ltd.

Jombart, T. 2011. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10:551-555.

Landguth, E. L., S. A. Cushman, M. K. Schwartz, K. S. McKelvey, M. Murphy, and G. Luikart. 2010. Quantifying the lag time to detect barriers in landscape genetics. Molecular Ecology 19:4179–4191.

McConnell, S., L. Hamilton, D. Morris, D. Cook, D. Paquet, P. Bentzen, and J. Wright. 1995. Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian east coast stocks of Atlantic salmon. Aquaculture 137:19-30.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA.

Nelson, R. J., T. D. Beacham, and M. P. Small. 1998. Microsatellite analysis of the population structure of a vancouver island sockeye salmon (*Oncorhynchus nerka*) stock complex using nondenaturing gel electrophoresis. Molecular Marine Biology and Biotechnology 7:312-319.

Olsen, J. B., P. Bentzen, and J. E. Seeb. 1998. Characterization of seven microsatellite loci derived from pink salmon. Molecular Ecology 7:1087-1089.

Ott, J. 1992. Strategies for characterizing highly polymorphic markers in human gene mapping. American Journal of Human Genetics 51:283-290.

Pflug, D. E., E. J. Connor, B. Hayman, T. W. Kassler, K. Warheit, B. McMillan, and E. Beamer. 2013. Ecological, genetic, and productivity consequences of interactions between hatchery- and natural-orgin steelhead of the Skagit watershed. Skagit River System Cooperative.

Queller, D.C. and K.F. Goodnight. 1989. Estimating Relatedness Using Genetic Markers. Evolution 43(2):258–275.

Raymond M. and F. Rousset. 1995. An exact test for population differentiation. Evolution 49:1280-1283

Reeves, P. A., C. L. Bowker, C. E. Fettig, L. R. Tembrock, and C. M. Richards. 2016. Effect of error and missing data on population structure inference using microsatellite data. BioRxiv. 080630. Available online at https://www.biorxiv.org/content/ early/2016/10/13/080630.

Rexroad, C. E. III, R. L. Coleman, A. M. Martin, W. K. Hershberger, and J. Killefer. 2001. Thirty-five polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). Animal Genetics 32:317-319.

Rousset, F. 2020. Genepop Manual. Version 4.7.5.

Ryman, N., and S. Palm. 2006. POWSIM: a computer program for assessing statistical power

when testing for genetic differentiation. Molecular Ecology Notes 6:600-602.

Scribner, K. T., J. R. Gust, and R. L. Fields. 1996. Isolation and characterization of novel microsatellite loci: cross species amplification and population genetic applications. Canadian Journal of Fisheries & Aquatic Sciences 53.

Seattle City Light (City Light). 2021. Revised Study Plan (RSP) for the Skagit River Hydroelectric Project, FERC Project No. 553. April 2021.

Seeb, L. W., A. Antonovich, M. A. Banks, T. D. Beacham, M. R. Bellinger,S. M. Blankenship, M. R. Campbell, N. A. Decovich, J. C. Garza, C. M.Guthrie III, T. A. Lundrigan, P. Moran, S. R. Narum, J. J. Stephenson, K. J. Supernault, D. J. Teel, W. D. Templin, J. K. Wenburg, S. F. Young, and C.T. Smith. 2007. Development of a standardized DNA database for Chinook Salmon. Fisheries 32:540–552.

Siniscalo, M., R. Robledo, P. K. Bender, C. Carcassi, L. Contu, and J. C. Beck. 1999. Population genomics in Sardinia: a novel approach to hunt for genomic combinations underlying complex traits and diseases. Cytogenetic and Genome Research 86:148-152.

Slatkin, M. 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 9: 477–485.

Small, M.P., E. Martinez, and S. Peterson. 2013. Genetic analysis of native char collected in Diablo Lake, Ross Lake, Gorge Reservoir, Skagit mainstem, and Skagit Bay. WDFW Molecular Genetics Lab, Olympia WA.

Small, M.P., S. Bell, and C. Bowman. 2016. Genetic analysis of native char collected in Diablo Lake, Gorge Reservoir, and Ross Lake in the Skagit River basin. Washington Department of Fish and Wildlife Molecular Genetics Lab, Olympia, Washington.

Small, M., E. Lowery, and M. Kissler. 2020. Genetic analysis of Char from Gorge Lake, Upper Skagit River, Washington. Final Report.

Small, M., E. Lowery, and M. Kissler. 2020. Genetic analysis of Char from Gorge Lake, Upper Skagit River, Washington. Final Report.Smith, C. T., B. F. Koop, and R. J. Nelson. 1998. Isolation and characterization of coho salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids. Molecular Ecology 7:1614-1616.

Smith, M. J. 2010. Population structure and genetic assignment of bull trout (*Salvelinus confluentus*) in the Skagit River Basin. School of Aquatic and Fishery Sciences University of Washington.

_____. 2021a. Personal communication between Scott Blankenship, Science Director Cramer Fish Sciences and Matt Smith, Fish Geneticist USFWS. June 14, 2021.

_____. 2021b. Personal communication between Dan Bingham, Senior Geneticist Cramer Fish Sciences and Matt Smith, Fish Geneticist USFWS. December 28, 2021.

Spies, I. B., D. J. Brasier, and P. T. L. O'Reilly, T. R. Seamons, and P. Bentzen. 2005. Development and characterization of novel tetra-, tri-, and dinucleotide microsatellite markers in rainbow trout (*Oncorhynchus mykiss*). Molecular Ecology Notes 5:278-281.

Stephenson, J. J., M. R. Campbell, J. E. Hess, C. C. Kozfkay, A. Matala, M. V. McPhee, P. Moran, S. Narum, M. M. Paquin, O. Schlei, M. P. Small, D. M. Van Doornik, and J. Wenburg.

2009. A centralized model for creating shared, standardized, microsatellite data that simplifies inter-laboratory collaboration. Conservation Genetics 10:1145-1149.

Taylor, E. and May-McNally, S. L. 2015. Genetic analysis of Dolly Varden (*Salvelinus malma*) across its North American range: evidence for a contact zone in southcentral Alaska. Canadian Journal of Fisheries and Aquatic Sciences. 72:1-10.

Vaha, J. and C. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology 15:63–72. doi:10.1111/j.1365-294X.2005.02773.x.

Waples, R., and O. E. Gaggiotti. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15:1419–1439.

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

Weir, B.S. and J. Goudet. 2017. A Unified Characterization of Population Structure and Relatedness. Genetics 206(4):2085–2103.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics. 15:323–354.

Yang, R.C. 1998. Estimating Hierarchical F-Statistics. Evolution 52(4):950–956.