FISHERY RESEARCH

CHINOOK SALMON AND STEELHEAD GENOTYPING FOR GENETIC STOCK IDENTIFICATION AT LOWER GRANITE DAM

ANNUAL PROGRESS REPORT
January 1, 2017 — December 31, 2017

Prepared by:

John H. Powell, Fishery Biologist 2
Ninh Vu, Fishery Biologist 2
Jesse McCane, Data Management Specialist 2
Micah Davison, Supervisory Fishery Biologist
Matthew R. Campbell, Fisheries Genetics Program Coordinator

IDFG Report Number 18-03
January 2018

Chinook Salmon and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam

Project Progress Report

2017 Annual Report

By

John H. Powell
Ninh Vu
Jesse McCane
Micah Davison
Matthew R. Campbell

Idaho Department of Fish and Game
600 South Walnut Street
P.O. Box 25
Boise, ID 83707

and

Daniel J. Hasselman
Shawn R. Narum

Columbia River Inter-Tribal Fish Commission
Hagerman Fish Culture Experiment Station
3059-F National Fish Hatchery Road
Hagerman, ID 83332

Project Number 2010-026-00
Contract Number 73153

IDFG Report Number 18-03
January 2018

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ABSTRACT

This report summarizes progress in the development and implementation of genetic stock identification (GSI) in the Snake River basin for natural-origin steelhead and spring/summer (spring/summer) Chinook Salmon for the 01/01/2016 to 12/31/2017 reporting period. Four objectives for the GSI project are addressed in this report: 1) the maintenance and evaluation of single nucleotide polymorphism (SNP) panels for high-throughput genotyping of steelhead and Chinook Salmon in the Snake and Columbia river basins; 2) the updating, maintenance, and testing of SNP baselines to describe genetic variation and for use as a reference in conducting GSI for both species; 3) the implementation of GSI to estimate genetic stock composition and life history diversity of steelhead and spring/summer Chinook Salmon passing Lower Granite Dam (LGR); and 4) the summarization of life history and genetic diversity information for steelhead and spring/summer Chinook Salmon detected at PIT-tag detection systems. For both species, panels of up to 191 SNPs have been in use for GSI and parentage-based tagging (PBT) at both Idaho Department of Fish and Game's Eagle Fish Genetics Lab, and its collaborating laboratory, the Columbia River Inter-Tribal Fish Commission's Hagerman Genetics Lab. Steelhead SNP baseline version v3.1 consists of 66 collections and 6,150 individuals. Chinook Salmon SNP baseline v3.1 consists of 46 collections and 4,604 individuals. SNP baselines are used to describe genetic diversity and structure of natural-origin populations throughout the Snake River. Based on population structure we have defined 10 genetic stocks for steelhead and 7 genetic stocks for Chinook Salmon for GSI analysis at LGR. We summarize GSI results for returning adults and emigrating juveniles during 2016-2017 at LGR using v3.1 baselines as reference. Finally, we describe the life history variation and genetic diversity of steelhead and Chinook Salmon detected at IPTDS. The information presented in this report provides critical data for viable salmonid population (VSP) monitoring of the Snake River steelhead DPS and the Snake River spring/summer Chinook Salmon ESU.

Authors:

John H. Powell, Fishery Biologist 2

Ninh Vu, Fishery Biologist 2

Jesse McCane, Data Management Specialist 2

Micah Davison, Supervisory Fishery Biologist

Matthew R. Campbell, Fisheries Genetics Program Coordinator

Suggested citation:

Powell, J.H., N. Vu, J. McCane, M. Davison, M. R. Campbell, D.J. Hasselman, and S. R. Narum. 2018. Chinook and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam. Idaho Department of Fish and Game Report 18-03. Annual Report, BPA Project 2010-026-00.

INTRODUCTION

Abundance (i.e. number of adults on spawning grounds) is a primary metric needed for monitoring the status of steelhead and salmon populations in the Columbia River basin (McElhany et al. 2000). Estimates of abundance combined with age and sex information over time allows estimation of productivity (e.g. recruits-per-female). Both abundance and productivity metrics provide indicators of the resiliency of populations and allow assessments of extinction risk. Estimates of these metrics at the population or major population group (MPG) scale is information that fisheries managers can use to achieve sustainable harvest of larger populations, while protecting weaker stocks and the biodiversity present within them.

Population-level assessments of abundance and productivity for ESA-threatened Snake River steelhead and Chinook Salmon can be particularly difficult due to the wide distribution and location of spawning areas (many populations are present in remote or wilderness areas). Additionally, environmental conditions at the time of spawning, especially for Snake River steelhead, often prevent the use of traditional counting methodologies (weirs, rotary screw traps, and redd count surveys). This is less of a problem for spring/summer Chinook Salmon, although turbid water conditions resulting from storms and forest fires have at times impacted the ability to estimate adult abundance using redd-based surveys in the Middle Fork and South Fork Salmon rivers (Thurow 2000). Snake River steelhead monitoring is further hampered due to high turbidity and changing flow conditions during the time of spawning (Thurow 1985). As a result, escapement estimates (and other demographic information) have not been available for most Snake River populations (Busby et al. 1996; Good et al. 2005) until recently.

In lieu of more detailed basin-level and population-specific information, steelhead in the Columbia River basin have traditionally been assigned to two groups (A-run and B-run), based on life history characteristics and bimodal passage timing at Bonneville Dam in the mid-Columbia River (Busby et al. 1996). By definition, A-run steelhead pass Bonneville Dam before August 25 and tend to return after one year in the ocean. B-run steelhead pass Bonneville Dam after August 25, tend to return after two years in the ocean. B-run steelhead are thought to be larger at age than A-run steelhead. Upstream migrating steelhead adults at Lower Granite Dam (LGR) do not exhibit a bimodal passage distribution. A-run and B-run adults at LGR are enumerated based on length (A-run, ≤78 cm; B-run, >78 cm) as a proxy for ocean age. A-run and B-run steelhead also exhibit differences in spawning distribution. A-run steelhead spawn throughout the Columbia basin, whereas the majority of B-run steelhead originate primarily from the Clearwater, Middle Fork Salmon, and South Fork Salmon rivers in Idaho. The putative differences in migration timing, morphology, and life history characteristics have been used as a surrogate for biodiversity in conservation planning for Snake River steelhead. However, the relationship between the morphological and life history characteristics to time of passage at Bonneville Dam is uncertain (Good et al. 2005). Further, the bimodal passage distribution at Bonneville Dam has become unimodal in recent years (Robards and Quinn 2002).

Two management concerns regarding Snake River steelhead have arisen in the last several years. First, populations classified as “B-run” do not appear to be self-sustaining (NMFS 2007) and their presence in the basin has affected operation of the Columbia River hydrosystem and fisheries management in the lower Columbia River. In particular, harvest of fall Chinook Salmon is constrained in order to limit impacts to B-run steelhead concurrently present in the Columbia River fishery. Secondly, there are substantial data needs to refine population delineations and conservation assessments (ICTRT 2003), but data have been lacking. Although Snake River “B-run” steelhead are currently identified as a biologically significant and distinct component of the Snake River DPS, their management is confounded by the lack of a

clear and detailed understanding of their actual spawning distribution and population structure. Nielsen et al. (2009) found that steelhead in Idaho Snake River tributaries exhibit a complicated pattern of genetic structure with populations clustering according to drainage locality, not simply by “A-run” or “B-run” designations.

The above issues and similar conservation and management questions relating to Snake River steelhead and spring/summer Chinook Salmon may be addressed through genetic stock identification (GSI). GSI uses multilocus genotype data from reference populations (representing the contributing stocks) as a baseline and complimentary genotype data from mixtures of fish of unknown origin to estimate stock proportions within the mixture (Shaklee et al. 1999; Anderson et al. 2008). GSI has been used extensively to understand and manage mixed stock fisheries for a variety of Pacific salmonids including Chinook Salmon (Smith et al. 2005), Sockeye Salmon (Habicht et al. 2010), Coho Salmon (Beacham et al. 2001), and steelhead (Beacham et al. 2000). In the Snake River basin, studies have indicated that both steelhead and Chinook Salmon exhibit significant genetic structuring at the watershed (or subbasin) level (Moran 2003; Narum et al. 2007; Nielsen et al. 2009; Matala et al. 2014). Previously, researchers have made use of this genetic structure to identify the genetic stock origin of kelt steelhead at LGR (Narum et al. 2008) and to estimate the stock composition of natural-origin and hatchery Chinook Salmon (Smith 2007) and natural-origin steelhead and Chinook Salmon (Ackerman et al. 2012; Campbell et al. 2012; Camacho et al. 2017; Camacho et al. 2018a; Camacho et al. 2018b) at LGR.

The results of the studies summarized above demonstrate the utility of GSI to obtain genetic stock abundance estimates for steelhead and Chinook Salmon in the Snake River basin. Continuation of GSI at LGR will allow us to 1) monitor genetic structure throughout the basin over time, and 2) estimate abundance, productivity, and life history diversity for genetic stocks throughout the Snake River. Sustained development and evaluation of GSI has been strongly recommended by regional RME workgroups. Similar work initiated at Bonneville Dam and in the lower Columbia River has been supported by the Independent Scientific Review Panel (http://www.nwcouncil.org/library/isrp/isrp2008-15.pdf).

REPORT STRUCTURE

This report contains four sections, one for each of the study objectives. Section 1 addresses the development of GT-Seq (Genotyping in Thousands by Sequencing) for more efficient and cost effective high-throughput genotyping for GSI in the Snake River basin. Section 2 summarizes efforts to update, maintain, and test SNP baselines for Snake River steelhead and spring/summer Chinook Salmon. These baselines are used to monitor genetic diversity and structure of natural-origin populations and are the reference for GSI at LGR. Section 3 addresses the use of GSI to estimate genetic stock proportions and life history diversity for natural-origin stocks at LGR. Section 4 summarizes life history and genetic diversity of steelhead and spring/summer Chinook Salmon detected at PIT-tag detection systems.

In this report, we refer to adult steelhead and Chinook Salmon migrating past LGR using spawn years (SY). For steelhead, a spawn year refers to adults that migrate upstream past LGR during the fall of the previous calendar year and the spring of the current calendar year (e.g., SY2017 steelhead are adults that migrated past LGR between 7/1/16 – 6/30/17 and spawned in spring of 2017). For spring/summer Chinook Salmon, a spawn year refers to adults that migrate past the dam prior to August 17 and spawn that same fall. We refer to juveniles of both species migrating past LGR using migratory years (MY). A migratory year refers to juveniles migrating downstream past LGR from the end of March to the end of July that year.

SECTION 1: EVALUATE AND MAINTAIN SNP MARKER PANELS

INTRODUCTION

The conclusion of calendar year 2017 marks our second full year of genotyping using the GT-seq platform for GSI and PBT applications. Our initial GT-seq marker panels comprised all markers used previously on the Fluidigm platform. As discussed in past reports, we chose to use the same markers on both genotyping platforms for continuity and because current PBT projects still rely on the original panel that had been produced on the Fluidigm platform. To review, the original Chinook Salmon GT-seq panel consists of 95 PBT loci, 96 GSI loci, and 1 sex marker. Similarly, the original Steelhead GT-seq panel consists of 95 PBT loci, 96 GSI loci, and 1 sex marker. Starting this year, in collaboration with the CRITFC Hagerman Genetics Lab, we chose to expand both the Chinook Salmon and steelhead GSI/PBT GT-seq panels. In total, 107 and 77 new markers have been added to the Chinook Salmon and steelhead GSI/PBT GT-seq panels, respectively. All GSI and PBT work started in 2017, as well as current projects, have been genotyped using these larger panels.

Data for existing and the newer panels are available on FishGen (https://www.fishgen.net/WebPages/CustomMarkerSet/MarkerExport.aspx). Metadata for each marker include synonym of species, Vic probe, Vic allele, Fam probe, Fam allele, forward primer, and reverse primer. The new larger Chinook panel known as 'CRITFC/IDFG Chinook GTseq v3.0 299' consists of 95 PBT loci, 96 GSI loci , 1 sex marker, and 107 new SNP markers. The new larger Steelhead panel known as 'CRITFC/IDFG Steelhead GTseq v1.0 269' consists of 95 PBT loci, 96 GSI loci, 1 sex marker, and 77 new SNP markers.

DISCUSSION

Marker panels have continued to expand since their inception, and this trend may slow in the future. It is important to keep in mind that the expansion of these panels does not affect past and future GSI/PBT work from the perspective of the original panels. That means that genotypes of markers from the smaller panels are contained within the larger panels. Current and future panels must adhere to one criteria: they must be compatible with existing panels. We have begun the process of evaluating these additional makers for both GSI and PBT applications. We intend to perform a thorough evaluation over the next performance period. For example, work is currently underway to re-genotype the Snake River steelhead GSI baseline. As more fish are genotyped using these larger panels, we will have the opportunity to evaluate the utility of these markers for PBT work as well. However, it is worth noting that we will continue to use the original 95 PBT marker panels for PBT work for the next several years, as only fish from the past year have been genotyped on the larger panels.

SECTION 2: UPDATE, MAINTAIN, AND TEST SNP BASELINES FOR STEELHEAD AND CHINOOK SALMON IN THE SNAKE RIVER

INTRODUCTION

The Snake River SNP baselines for steelhead and Chinook Salmon serve two primary purposes: 1) to monitor genetic structure and diversity of natural-origin Snake River populations both spatially and temporally, and 2) to serve as a reference for GSI work at LGR.

First, the monitoring of genetic structure over time and space provides insight regarding gene flow, both historic and contemporary, from natural (successful straying) and manmade (i.e. out-of-basin hatchery stocking) causes. Monitoring genetic diversity of populations provides information about gain or loss in genetic diversity over time and provides insight into the adaptive potential of populations. In this section, we provide genetic structure and diversity information for 23 extant steelhead TRT populations and 28 extant Chinook Salmon TRT populations throughout the Snake River basin to aid in viable salmonid population (VSP; McElhany et al. 2000) monitoring of the Snake River steelhead DPS and spring/summer Chinook ESU.

Second, the Snake River SNP baselines serve as a reference for GSI conducted at LGR to estimate genetic stock composition of out-migrating smolts (e.g. Stark et al. 2016, Camacho et al. 2018b) and returning adults (e.g. Camacho et al. 2017, Camacho et al. 2018b). Genetic stock composition estimates of adults and juveniles at LGR, combined with sex and age data, will allow us to estimate abundance, productivity, and life history diversity of genetic stocks over time for VSP monitoring. For GSI, our objective is to periodically update and maintain the SNP baselines to accurately estimate contemporary allele frequencies (genetic structure) of natural-origin populations throughout the Snake River contributing to production at LGR.

In this report, we tested the accuracy and bias of the current Snake River basin GSI baselines for steelhead and Chinook Salmon. We investigated the effects of the marker panel and population composition on the assignment of individuals to reporting units. We also tested the effects of assignment method on the estimated stock composition of the run at LGR. Finally, we tested the accuracy of the Chinook Salmon Snake River SNP baseline v3.1 in classifying recent collections of Chinook Salmon sampled from the Chamberlain Creek watershed.

METHODS

We assessed whether the mixing proportions estimated for each reporting unit are biased due to heterogeneity in the number of populations included in each reporting unit using the R package *rubias* (E. Anderson unpublished, available at https://github.com/eriqande/rubias) as described in Hasselman et al. (2016). For this test, we resampled 4,000 individuals from the steelhead or Chinook Salmon Snake River GSI baseline version 3.1 1,000 times. In each iteration the proportion of samples from a reporting unit was drawn from a Dirichlet distribution with all parameters equal to 100 times the average mixture proportions observed at Lower Granite Dam from SY2012–2015. The proportion of samples from a population within each reporting unit was then drawn from a Dirichlet distribution with all parameters equal to 0.5.

We assessed whether increasing the number of genotyped loci, or individuals, would lead to an increase in the assignment probability of individuals, and a decrease in the bias and variance of estimated mixing proportions at Lower Granite Dam. These tests were implemented

by randomly removing 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 70%, 60%, or 50% of the loci or individuals in the current baseline 1,000 times. Therefore, a total of 22,000 baselines of various size were constructed. For each of these reduced size baselines 4,000 individuals were resampled from the steelhead or Chinook Salmon Snake River GSI baseline version 3.1. The reporting unit proportions for these tests were set equal to the average mixture proportions observed at Lower Granite Dam from SY2012–2015. The proportion of samples from a population within each reporting unit was then drawn from a Dirichlet distribution with all parameters equal to 0.5.

When removing individuals from the baseline for these tests we made an effort to equalize sample sizes in each of the populations. Therefore, we started removing individuals from the largest populations in the baseline first working backwards to equate sample sizes. That is, if 60 individuals were to be removed from the baseline, and the largest population was more than 60 individuals larger than the next largest population, all individuals were removed from the largest population. If for example, the largest population was not 60 individuals larger than the next largest population, the largest population was equaled in size to the next largest population and then the remaining individuals to cut were removed from each of the two largest populations equally.

For each iteration we calculated the mean posterior individual assignment probability for each population to a reporting unit, as well as root mean squared error for the estimated mixture proportions in the baseline.

In addition to these bias analyses we completed two additional tests using our GSI baseline. First, we examined the effect of using different assignment methods available in the program *gsi_sim* on the assignment rate of baseline samples. We detail the results of this study in Appendix A. We also tested the genetic divergence between, and assignment rates of, recent juvenile and adult Chinook Salmon samples collected in the Chamberlain Creek watershed and our Snake River SNP baseline v3.1. We describe the results of this study in Appendix B.

RESULTS

We observed low overall bias in the estimated mixture proportions due to unequal numbers of populations in both the Snake River steelhead and Chinook Salmon GSI baselines version 3.1 (Figures 1 and 2). However, in general the Snake River steelhead GSI baseline version 3.1 does overestimate the mixture proportions from the lower Clearwater River and lower Snake River reporting units, and underestimate the proportions from the Imnaha River and lower Salmon River reporting units (Figure 1).

In general reducing the number of loci or fish from the baseline led to increased root mean squared error in the estimated mixture proportions for steelhead reporting units in GSI baseline version 3.1 (Figure 3). Overall, it appeared that increasing the number of loci in the analysis led to greater reductions in root mean squared error than increasing the number of genotyped individuals (Figure 3). We observed larger increases in the average individual mean posterior probability of assignment to populations in the Snake River steelhead GSI baseline version 3.1 when increasing genotyped loci (Figure 4). Reducing the number of genotyped loci and fish resulted in similar patterns of increasing root mean squared error in Chinook Salmon (Figure 5). As with steelhead, the larger increases in the average individual mean posterior probability of assignment to populations in the Snake River Chinook Salmon GSI baseline version 3.1 were also observed when increasing genotyped loci (Figure 6).

DISCUSSION

We have continued to modify and update the GSI baselines used to assign returning Snake River basin steelhead and Chinook Salmon to their reporting unit of origin over the course of this project. Ackerman et al. (2011b) used GSI baseline version 1.0 (Table 1) to assign migrating fish in 2009 and 2010 to their reporting group of origin using individual based assignments to their "best-estimate" stock in the program ONCOR (Kalinowski et al. 2007). GSI baseline 2.0 was developed in 2011 (Table 1). Ackerman et al. (2012) used Bayesian mixture modeling in a newly developed software program called *gsi_sim* (Anderson et al. 2008, Anderson 2010) to determine the proportion of fish crossing Lower Granite Dam that assigned to each reporting unit. Ackerman et al. (2014) assigned individuals to a slightly modified GSI baseline version 2.0 in 2012 (Table 1). Vu et al. (2015) updated the GSI baseline to version 3.0 (Table 1) by incorporating adult fish PIT-tagged at Lower Granite Dam in SY2010–SY2012 and detected at PIT-tag detection arrays. Fish migrating in 2013 were assigned to reporting units using GSI baseline 3.0. Ackerman et al. (2016) created GSI baseline version 3.1 by removing individuals detected at IPTDS locations from the GSI baseline. This change was made to allow a re-analysis of all adults sampled at LGR (SY2009 through SY2015) using the same baseline. Fish migrating in 2014 and 2015 were assigned to reporting unit using GSI baseline version 3.1 (Table 1). We assigned fish migrating in 2016 and 2017 to GSI baseline version 3.1 using 179 loci for steelhead and 173 loci for Chinook Salmon (Table 1). The reduction in the size of the marker panels used for GSI analysis was due to a shift to genotyping individuals using the "Genotyping-in-Thousands by sequencing" methodology (Campbell et al. 2015). All individuals migrating past Lower Granite Dam since 2012 have been assigned to a reporting unit by summing the scaled likelihood of the genotype arising in a given population using in the program *gsi_sim* (Anderson et al. 2008; Anderson 2010).

We chose to keep the same overall baselines from 2015. However, we reduced the steelhead panel by 6 loci (Omy_crb-106, Omy_rapd-167, OMS00087, OMS00169, OMS00176, Omy_u09-52.284) to align with the CRITFC Columbia River steelhead baseline (Hess et al. 2013). We also removed 7 loci from the Chinook Salmon panel (Ots_ARNT, Ots_110495-380, Ots_C3N3, Ots_104569-86, Ots_unk8200-45, Ots_il13Ra2B-37, Ots_il1racp-166) because they did not make the initial GT-seq marker panel (Hess et al. 2016). These smaller marker panels led to a reduction in the number of samples included in each baseline due to the requirement that individuals included in the GSI baselines are successfully genotyped at a minimum of 90% of the SNP loci included in the baseline. Steelhead baseline v3.1 now consists of 5,967 samples genotyped at 179 loci. These collections represent all 23 TRT populations and all 6 major population groups. These collections were pooled into 45 GSI populations encompassing 10 genetic stocks. Chinook Salmon v3.1 now consists of 4,356 samples genotyped at 173 loci. These collections represent 31 of 41 TRT populations and all 5 major population groups. These collections were pooled into 30 GSI populations encompassing seven genetic stocks. Methods, results, and analyses of both baselines can be found in Vu et al. (2015).

We observed overall low bias in the estimated mixture proportions in both the Snake River steelhead and Chinook Salmon GSI baselines version 3.1 (Figures 1 and 2). This indicates that while individual assignment probabilities may be low in some populations in our baselines (Vu et al. 2015), the overall abundance estimates at Lower Granite Dam derived from estimated stock composition using GSI baselines v3.1 are unbiased.

We observed a pattern of increasing root mean squared error and decreasing mean individual posterior probability of assignment when reducing the number of genotyped loci in both the steelhead and Chinook Salmon GSI baselines (Figures 3-6). These reductions were greater than what we observed when reducing the number of genotyped individuals (Figures 3–6). Therefore, there appears to be a greater benefit to increasing the number of genotyped loci for both baselines than adding additional samples to the baselines. This coming year we will focus on expanding the marker panel in our current Snake River steelhead and Chinook Salmon GSI baselines.

Based on the results of our self-assignment tests we will use maximum likelihood estimates of the reporting unit of origin for individual steelhead and Chinook Salmon sampled at Lower Granite Dam (Appendix A). We did not observe substantial genetic divergence between recent collections of Chinook Salmon from Chamberlain Creek and our GSI baseline version 3.1 (Appendix B). Therefore, we did not update the Chamberlain Creek reporting unit in GSI baseline version 3.1 from what Vu et al. (2015) describe.

SECTION 3. IMPLEMENT GSI METHODS TO ESTIMATE PROPORTIONS AND BIOLOGICAL PARAMETERS OF NATURAL-ORIGIN STOCKS AT LOWER GRANITE DAM

INTRODUCTION

The Idaho Department of Fish and Game's (IDFG) long-range goal of its anadromous fish program, consistent with basin-wide mitigation and recovery efforts, is to preserve Idaho's salmon and steelhead runs and recover them to benefit all users (IDFG 2007). Fisheries management requires an understanding of how salmonid populations function as well as regular status assessments to achieve these goals (McElhany et al. 2000). Estimates of abundance, combined with sex and age information over time, allow estimation of population growth rates; and both abundance and productivity metrics provide indicators of the resiliency and viability of populations. Estimates of these metrics at the genetic stock or MPG level is information that fisheries managers can use to achieve sustainable harvest of larger populations, while protecting weaker stocks and the biodiversity within them.

However, population level or MPG assessments of abundance and productivity for ESA-listed Snake River steelhead and spring/summer Chinook Salmon can be particularly difficult (see Report Introduction). Specific data on Snake River steelhead and Chinook Salmon MPGs and populations are lacking, particularly key parameters such as population abundance, age composition, genetic diversity, recruits per spawner, and survival rates (ICTRT 2003). Genetic Stock Identification is one potential means for estimating these parameters at a finer-scale (e.g., MPG, genetic stock [reporting group], or population). Genetic Stock Identification uses multilocus genotype data from reference populations (representing potential contributing stocks) as a baseline and a complimentary set of genotype data from mixtures of fish of unknown origin to estimate stock proportions within the mixture and to estimate stock of origin of individual fish (Shaklee et al. 1999). Section 2 of Vu et al. (2015) presents the SNP baselines used for GSI in the Snake River basin. Here we use complementary sets of genotype data from adults sampled at the Lower Granite Dam (LGR) adult trap and juveniles sampled at the LGR juvenile bypass facility to estimate the genetic stock of origin of upstream migrating adults and emigrating juveniles. We then provide life history diversity (sex, length, age, migration timing) information of individuals assigning to the various Snake River genetic stocks.

In this report, we present individual genetic assignments and life history diversity information for SY2016-SY2017 adults and MY2016-MY2017 juveniles (both steelhead and Chinook Salmon) sampled at LGR.

METHODS

Adult Trap Operations

Detailed methods for operation of the LGR adult trap can be found in Camacho et al. (2017) and citations within. Briefly, adult steelhead and spring/summer Chinook Salmon migrating upstream past LGR may be intercepted at a trapping facility, located on the adult fish ladder above the counting window, according to a predetermined sampling rate. A committee of collaborating management agencies determines the trap sampling rates that achieve sample requirements for multiple projects and balance fish handling concerns. Sample rates are typically 10–20%. The sample rate determines how long a trap gate remains open four times per hour; the trap is operational 24 hours per day.

Juvenile Trap Operations

Detailed methods for operation of the LGR juvenile trap can be found in Camacho et al. (2018b) and citations within. The juvenile trap is located on the LGR juvenile bypass system. The trap captures a systematic sample of fish by operating two trap gates according to a predetermined sample rate. The sample rate determines how long the trap gates remain open, up to six times per hour. The trap is operational 24 hours per day and fish are processed every morning. Sample rate is predetermined daily to collect 250-750 fish per day (all species combined) and is based on the expected number of fish entrained in the bypass system that day.

Fish Handling Protocols (Adults and Juveniles)

Fish handling procedures are detailed in Camacho et al. (2017) for adults and Camacho et al. (2018b) for juveniles (and citations within both reports). Fish captured at the LGR adult or juvenile trap are anesthetized; identified to species; examined for external marks, tags, and injuries; scanned for an internal CWT or PIT-tag; and measured for fork length (FL). All fish are examined for the presence (unclipped) or absence (clipped) of the adipose fin and classified to putative origin (hatchery or natural). All natural-origin fish have an unclipped adipose fin because they spend their entire life cycle in the natural environment. Most hatchery-origin fish have a clipped adipose fin. However, some hatchery fish may be released with an unclipped adipose fin for supplementation or tribal harvest opportunities. Thus, unclipped fish are also examined for a CWT or a PIT-tag. The presence of a CWT definitively identifies an unclipped fish as hatchery origin. For unclipped steelhead, hatchery origin may also be determined by the presence of dorsal and/or ventral fin erosion, which is assumed to occur only in hatchery-reared steelhead (Latremouille 2003). Captured fish determined to be putatively natural-origin or unclipped hatchery with no CWT are sampled for scales (for age; except juvenile Chinook) and tissue (for sex and genotype data). For juveniles, fish bearing PIT tags and/or diseased or injured fish were omitted from the subsample, as were Chinook deemed to be yearling fall Chinook based on external morphology (Tiffan et al. 2000).

Scales were taken from above the lateral line and posterior to the dorsal fin. Samples were stored in coin envelopes for transport to the IDFG ageing laboratory in Nampa, Idaho. Tissue samples were taken from a small clip of the anal fin. Tissues were stored in a vial with 200-proof non-denatured ethyl alcohol for transport to the IDFG Eagle Fish Genetics Laboratory. Gender was not visually determined at the trap, but was assessed using Y-specific genetic assays (Campbell et al. 2012). After processing, all fish were returned to the fish ladder to resume upstream migration (adults) or the bypass system to resume downstream migration (juveniles).

Scale Age Protocol

Protocols for determining a fish's age from scales are detailed in Wright et al. (2015).

Genetics Laboratory Protocol

Laboratory protocols for DNA extraction, amplification, and SNP genotyping are detailed in Section 2 of Vu et al. (2015). MY2016 steelhead and Chinook Salmon juveniles and SY2016 steelhead and Chinook Salmon adults were processed at the CRITFC Genetics Lab in Hagerman, Idaho. MY2017 steelhead and Chinook Salmon juveniles and SY2017 steelhead

and Chinook Salmon adults were processed at IDFG's Eagle Fish Genetics Lab (EFGL) in Eagle, Idaho.

Parentage-Based Tagging

Beginning in 2008, parentage-based tagging (PBT; Anderson and Garza 2005) has been used to genetically tag nearly all hatchery-origin steelhead in the Snake River basin (Steele et al. 2013a, 2013b). This genetic tagging technique is accomplished by genotyping all parental broodstock each spawn year, thereby allowing any offspring to be assigned back to their parents and identifying the hatchery of origin and age of offspring. The implementation of PBT provides an alternative to coded-wire tags (CWT) for identifying the origin and age of fish harvested in mixed-stock fisheries or that stray into natural spawning areas.

We conducted PBT analysis for all unclipped juvenile fish sampled in MY2016-MY2017 and adult fish sampled in SY2016-SY2017. In using PBT to evaluate all the fish, we are better able to identify putative natural-origin (unclipped, unmarked) fish that are truly of hatchery origin. Any individuals identified as unmarked hatchery origin adults with a PBT were removed from the dataset before performing GSI and evaluating life history diversity of genetic stocks.

Genetic Stock Identification

Individual assignment (IA) tests were conducted on all unclipped juveniles and adults that did not receive a PBT assignment in MY2016-2017 and SY2016-2017 using the Snake River SNP baselines v3.1 described in Section 2 of Vu et al. (2015). SNP allele frequency estimates from baseline collections serve as the reference information for IA tests. Fish sampled at the LGR adult and juvenile trapping facilities were genotyped at the same SNPs and multilocus genotype data were used to assign individual fish back to their estimated population (and genetic stock) of origin (Pella and Milner 1987; Shaklee et al. 1999). In IA, the probability that each fish originates from a baseline population is calculated based on the likelihood that the individual's genotype belongs to that population, given baseline allele frequency estimates. Individual population estimates were first calculated and then summed into genetic stock estimates (allocate-sum procedure; Wood et al. 1987). Genetic stocks (aka reporting groups) are assemblages of reference (baseline) populations grouped primarily by genetic and geographic similarities and secondarily by political boundaries and/or management units (Ackerman et al. 2011a). IA procedures assign an individual's genotype to the reporting group from which it is most likely to have originated.

Ten genetic stocks were used for natural-origin steelhead IA analyses. Genetic stocks include: 1) UPSALM: upper Salmon River; 2) MFSALM: Middle Fork Salmon River (including Chamberlain and Bargamin creeks); 3) SFSALM: South Fork Salmon River; 4) LOSALM: lower Salmon River; 5) UPCLWR: upper Clearwater River (Lochsa and Selway rivers); 6) SFCLWR: South Fork Clearwater River (including Clear Creek); 7) LOCLWR: lower Clearwater River; 8) IMNAHA: Imnaha River; 9) GRROND: Grande Ronde River; and 10) LSNAKE: Asotin Creek and tributaries to the Snake River downstream of the Clearwater River confluence.

Seven natural-origin Chinook Salmon genetic stocks were used during IA analyses. Genetic stocks include: 1) UPSALM: upper Salmon River; 2) MFSALM: Middle Fork Salmon River; 3) CHMBLN: Chamberlain Creek; 4) SFSALM: South Fork Salmon River; 5) HELLSC: an aggregate reporting group that includes the Little Salmon, Clearwater, Grande Ronde, and Imnaha rivers; 6) TUCANO: Tucannon River, and 7) FALL: Snake River fall Chinook Salmon. Three collections of Snake River fall Chinook Salmon (see Table 2 in Ackerman et al. 2012) are

included in the SNP baselines (FALL genetic stock); we are able to identify fall Chinook within mixtures of spring/summer Chinook with 100% accuracy.

After performing IA, we estimated genetic stock compositions of all samples analyzed and evaluated life history diversity for each genetic stock. We summarize results for eight sample groups:

- SY2016 steelhead adults
- MY2016 steelhead juveniles
- SY2016 Chinook adults
- MY2016 Chinook juveniles
- SY2017 steelhead adults
- MY2017 steelhead juveniles
- SY2017 Chinook adults
- MY2017 Chinook juveniles

RESULTS

We inventoried a total of 22,643 samples from SY2016-SY2017 adults and MY2016-MY2017 juveniles from LGR (Table 2). All inventoried samples were queued for genotyping. We failed to successfully genotype 257 (1.1%) of the queued samples. We assigned 5,103 (22.8%) samples to hatchery parents in our PBT baseline despite all samples being collected from fish with intact adipose fins. Results for the remaining 17,283 samples on which we performed IA are summarized below and in Tables 2-6.

SY2016-SY2017 Steelhead Adults

Of the 5,237 unclipped adult steelhead sampled in SY2016, 4,830 (92.2%) were phenotypically identified as natural-origin because they had no apparent dorsal or ventral fin erosion (Table 2). The remaining 407 (7.8%) fish were identified as hatchery origin due to the presence of dorsal and/or ventral fin erosion. All 5,237 samples collected in SY2016 were queued for genotyping, and we successfully genotyped 5,149 (98.3%) of these samples. We assigned 436 (9.2%) of the 4,744 successfully genotyped phenotypically natural-origin fish to hatchery parents, and 375 (92.6%) of the 405 successfully genotyped phenotypically hatchery fish to hatchery parents. The remaining successfully genotyped phenotypically natural-origin (4,308) and phenotypically hatchery (30) fish that did not assign to hatchery parents were all assigned to a genetic stock via IA.

Life history diversity information (sex, length, and ocean age) for the 4,308 unclipped steelhead adults sampled in SY2016 that were assigned a genetic stock is summarized in Table 3. Of the 4,308 assigned a genetic stock, 606 (14.1%) assigned to UPSALM, 290 (6.7%) to MFSALM, 165 (3.8%) to SFSALM, 145 (3.4%) to LOSALM, 454 (10.5%) to UPCLWR, 271 (6.3%) to SFCLWR, 465 (10.8%) to LOCLWR, 328 (7.6%) to IMNAHA, 1,074 (24.9%) to GRROND, and 510 (11.8%) to LSNAKE.

Of the 4,731 unclipped adult steelhead sampled in SY2017, 3,951 (83.5) were phenotypically identified as natural-origin because they had no apparent dorsal or ventral fin erosion (Table 2). The remaining 780 (16.5%) fish were identified as hatchery origin due to the presence of dorsal and/or ventral fin erosion. All 4,731 samples collected in SY2017 were queued for genotyping, and we successfully genotyped 4,713 (99.6%) of these samples. We

assigned 851 (21.6%) of the 3,935 successfully genotyped phenotypically natural-origin fish to hatchery parents, and 727 (93.4%) of the 778 successfully genotyped phenotypically hatchery fish to hatchery parents. The remaining successfully genotyped phenotypically natural-origin (3,084) and phenotypically hatchery (51) fish that did not assign to hatchery parents were all assigned to a genetic stock via IA.

Life history diversity information (sex, length, and ocean age) for the 3,084 unclipped steelhead adults sampled in SY2017 that were assigned a genetic stock is summarized in Table 3. Of the 3,084 assigned a genetic stock, 232 (7.5%) assigned to UPSALM, 200 (6.5%) to MFSALM, 126 (4.1%) to SFSALM, 32 (1.0%) to LOSALM, 432 (14.0%) to UPCLWR, 212 (6.9%) to SFCLWR, 223 (7.2%) to LOCLWR, 199 (6.5%) to IMNAHA, 1,008 (32.7%) to GRROND, and 420 (13.6%) to LSNAKE.

MY2016-MY2017 Steelhead Juveniles

All 1,243 unclipped juvenile steelhead samples for MY2016 were queued for genotyping (Table 2). A total of 1,231 (99.0%) juveniles were genotyped successfully, with 31 (2.5%) assigning back to hatchery parents. The remaining 1,200 (97.5%) were assigned a genetic stock via IA.

Life history diversity information for the 1,200 emigrating steelhead smolts that were assigned a genetic stock is summarized in Table 4. Of the 1,200 steelhead smolts assigned a genetic stock, 191 (15.9%) assigned to UPSALM, 114 (9.5%) to MFSALM, 78 (6.5%) to SFSALM, 52 (4.3%) to LOSALM, 138 (11.5%) to UPCLWR, 93 (7.8%) to SFCLWR, 99 (8.3%) to LOCLWR, 108 (9.0%) to IMNAHA, 210 (17.5%) to GRROND, and 117 (9.8%) to LSNAKE.

All 1,516 unclipped juvenile steelhead samples for MY2017 were queued for genotyping (Table 2). A total of 1,474 (98.1%) juveniles were genotyped successfully, with 20 (1.4%) assigning back to hatchery parents. The remaining 1,454 (98.6%) were assigned a genetic stock via IA.

Life history diversity information for the 1,454 emigrating steelhead smolts that were assigned a genetic stock is summarized in Table 4. Of the 1,454 steelhead smolts assigned a genetic stock, 343 (23.6%) assigned to UPSALM, 68 (4.7%) to MFSALM, 26 (1.8%) to SFSALM, 24 (1.7%) to LOSALM, 129 (8.9%) to UPCLWR, 126 (8.7%) to SFCLWR, 120 (8.3%) to LOCLWR, 102 (7%) to IMNAHA, 352 (24.2%) to GRROND, and 164 (11.3%) to LSNAKE.

SY2016-SY2017 Chinook Salmon Adults

We inventoried 4,538 unclipped adult Chinook Salmon samples for SY2016 and all of them were queued for genotyping (Table 2). Of those, 4,489 (98.9%) were genotyped successfully, of which 1,409 (31.4%) were assigned back to hatchery parents and 3,080 (68.6%) were assigned back to a genetic stock via IA.

Life history diversity information for the 3,080 Chinook Salmon adults that were assigned to a genetic stock is summarized in Table 5. Of the 3,080 samples, 614 (19.9%) assigned to UPSALM, 448 (14.5%) to MFSALM, 51 (1.7%) to CHMBLN, 359 (11.7%) to SFSALM, 1,254 (40.7%) to HELLSC, 8 (0.3%) to TUCANO, and 346 (11.2%) to FALL.

We inventoried 2,060 unclipped adult Chinook Salmon samples for SY2017 and all of them were queued for genotyping (Table 2). Of those, 2,055 (99.8%) were genotyped

successfully, of which 828 (40.3%) were assigned back to hatchery parents and 1,227 (59.7%) were assigned back to a genetic stock via IA.

Life history diversity information for the 1,227 Chinook Salmon adults that were assigned to a genetic stock is summarized in Table 5. Of the 1,227 samples, 193 (15.7%) assigned to UPSALM, 165 (13.4%) to MFSALM, 28 (2.3%) to CHMBLN, 206 (16.8%) to SFSALM, 479 (39.0%) to HELLSC, 9 (0.7%) to TUCANO, and 147 (12.0%) to FALL.

MY2016-MY2017 Chinook Salmon Juveniles

We inventoried 1,516 unclipped juvenile Chinook Salmon for MY2016. Of the 1,516 juvenile Chinook Salmon inventoried, all were queued for genotyping and 1,475 (97.3%) of those genotyped successfully (Table 2). Of the juveniles genotyped, 171 (11.6%) were assigned back to hatchery parents and the remaining 1,304 (88.4%) were assigned a genetic stock via IA.

Life history diversity information for the 1,304 Chinook Salmon smolts assigned a genetic stock is summarized in Table 6. Of the 1,304 Chinook Salmon smolts assigned a genetic stock, 206 (15.8%) assigned to UPSALM, 226 (17.3%) to MFSALM, 29 (2.2%) to CHMBLN, 212 (16.3%) to SFSALM, 590 (45.2%) to HELLSC, 5 (0.4%) to TUCANO, and 36 (2.8%) to FALL.

We inventoried 1,816 unclipped juvenile Chinook Salmon for MY2017. Of the 1,816 juvenile Chinook Salmon inventoried, all were queued for genotyping and 1,800 (99.1%) of those genotyped successfully (Table 2). Of the juveniles genotyped, 255 (14.2%) were assigned back to hatchery parents and the remaining 1,545 (85.8%) were assigned a genetic stock via IA.

Life history diversity information for the 1,545 Chinook Salmon smolts assigned a genetic stock is summarized in Table 6. Of the 1,545 Chinook Salmon smolts assigned a genetic stock, 295 (19.1%) assigned to UPSALM, 237 (15.3%) to MFSALM, 11 (0.7%) to CHMBLN, 285 (18.4%) to SFSALM, 710 (46.0%) to HELLSC, 0 (0.0%) to TUCANO, and 7 (0.5%) to FALL.

DISCUSSION

Adult steelhead and spring/summer Chinook Salmon are intercepted at the LGR adult trapping facility at approximately a 10–20% trapping rate. Tissue samples are taken from trapped fish as part of this project to estimate abundance and life history diversity metrics at the genetic stock and/or MPG scale. This work allows estimation of abundance and productivity by the Idaho Department of Fish and Game for both steelhead and Chinook Salmon at the genetic stock scale across the entire Snake River basin. These metrics are critical components of VSP monitoring and are reported in the wild adult and juvenile steelhead and Chinook Salmon abundance and composition reports (e.g., Camacho et al. 2017; Camacho et al. 2018a; Camacho et al. 2018b).

Trapped adult fish are also PIT tagged by the Integrated Status and Effectiveness Monitoring Project (ISEMP; BPA Project 2003-017-00). Detection of these PIT-tagged throughout the Snake River basin are used in a Bayesian branching model to provide reliable and unbiased estimates of abundance at the tributary scale (QCI 2013; See et al. 2016). A multi-agency collaboration has been initiated to utilize information generated from these two innovative technologies (SNP genotyping and PIT-tag detections). The goal of this collaboration

is to synthesize available data regarding abundance, life-history diversity, and genetic structure and diversity of Snake River steelhead and spring/summer Chinook Salmon. This information is available from the PIT tagging and biological sampling of adults at LGR and the subsequent detection of those adults throughout the Snake River basin. We discuss where the results of this collaboration are reported in Section 4.

The Columbia River Inter-Tribal Fish Commission conducts PBT and GSI of adult steelhead and Chinook Salmon at Bonneville Dam to estimate stock composition and abundance and to evaluate life history information for stocks migrating above Bonneville Dam. In the future, we intend to combine information from GSI at both LGR and Bonneville Dam to evaluate straying and survival between the two dams for both species. Further, we will evaluate adults captured in the Zone 6 fishery (between Bonneville Dam and McNary Dam) using a combination of PBT and GSI. The above information combined will also greatly assist run reconstruction efforts.

Parentage-based tagging is another important genetic technology implemented on all fish with intact adipose fins sampled at LGR prior to GSI. Using this technology we can remove unmarked, untagged hatchery origin individuals from the natural-origin sample used to estimate abundance at the genetic stock, MPG, population, and/or subpopulation levels. Failing to remove these unidentified hatchery individuals will result in overestimating abundance of natural-origin stocks. This overestimate is likely the largest potential source of bias in abundance estimation within the Snake River basin. We illustrate the importance of this filtering step with returning steelhead. Hatchery steelhead can potentially be identified through the presence of dorsal and/or ventral fin erosion that is assumed absent from natural-origin individuals (Latremouille 2003) in addition to any physical marks (e.g., coded wire tags, adipose fin clip) applied at the hatchery. Despite this additional "mark," using PBT we identified an additional 9.2% to 21.7% of the phenotypically natural-origin steelhead trapped at LGR as hatchery origin in SY2016-2017 (Table 2). Because there is no additional phenotypic hatchery "mark" for Chinook Salmon the rate of adipose-intact hatchery fish identified via PBT only is higher than for steelhead. Thus, the application of PBT is instrumental in accurately estimating abundance of natural-origin stocks in the Snake River basin.

Continuation of GSI efforts at LGR will allow us to 1) monitor genetic structure and diversity throughout the basin over time, and 2) estimate productivity parameters and related life history diversity information for genetic stocks throughout the Snake River basin.

SECTION 4: SUMMARIZE LIFE HISTORY AND GENETIC DIVERSITY OF NATURAL-ORIGIN STEELHEAD AND SPRING/SUMMER CHINOOK SALMON THAT ARE DETECTED AT PIT TAG DETECTIONS SYSTEMS IN THE SNAKE RIVER BASIN SY2016-2017

Powell et al. (2017) synthesized life history and genetic diversity of 15 of the 24 extant independent steelhead populations within the Snake River DPS, 17 of the 28 extant Snake River spring/summer Chinook Salmon populations, and 4 extirpated Snake River spring/summer Chinook Salmon populations for SY2010–2015. Powell and others are presenting this information for SY2016-2017 in a companion report titled “Abundance, life history and genetic diversity of natural-origin steelhead and spring/summer Chinook Salmon detected at instream PIT-tag detection systems in the Snake River basin.”

ACKNOWLEDGEMENTS

The authors would like to thank the many individuals who contributed time and expertise towards implementing this project (alphabetical):

IDFG Eagle Fish Genetics Lab:

- Amanda Boone (IDFG)
- Grant Bruner (IDFG)
- Dan Eardley (IDFG)
- Kelly Heindel (IDFG)
- Dylan Kovis (PSMFC)
- Laura Redfield (PSMFC)
- Lynn Schrader (IDFG)
- Thea Wickersham (PSMFC)

Idaho Department of Fish and Game:

- Carlos Camacho
- Tim Copeland
- J. Lance Hebdon
- Bill Schrader

Columbia River Inter-Tribal Fish Commission:

- Nate Campbell
- Stephanie Harmon
- Nick Hoffman
- Vanessa Jacobson
- Amanda Matala
- Andrew Matala
- Lori Maxwell
- Megan Moore
- Jeff Stephenson

Quantitative Consultants, Inc. (QCI)

- Chris Beasley
- Mike Ackerman
- Kevin See

NOAA Northwest Fisheries Science Center:

- Darren Ogden

Oregon Department of Fish and Wildlife

Washington Department of Fish and Wildlife

US Fish and Wildlife Service

Nez Perce Tribe

Shoshone-Bannock Tribes

Bonneville Power Administration

- Dorothy Welch
- Barbara Shields

Primary funding for this project comes from the Bonneville Power Administration (Project #2010-026-00).

LITERATURE CITED

Ackerman, M. W., C. Habicht, and L. W. Seeb. 2011a. Single-Nucleotide Polymorphisms (SNPs) under Diversifying Selection Provide Increased Accuracy and Precision in Mixed-Stock Analyses of Sockeye Salmon from the Copper River, Alaska. Transactions of the American Fisheries Society 140:865-881. doi: 10.1080/00028487.2011.588137

Ackerman, M. W., J. McCane, C. A. Steele, M. R. Campbell, A. P. Matala, J. E. Hess, and S. R. Narum. 2011b. Chinook and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam. Idaho Department of Fish and Game Report 11-113. Annual Report, BPA Project 2010-026-00.

Ackerman, M. W., J. McCane, C. A. Steele, M. R. Campbell, A. P. Matala, J. E. Hess, and S. R. Narum. 2012. Chinook and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam. Idaho Department of Fish and Game Report 12-15. Annual Report, BPA Project 2010-026-00.

Ackerman, M. W., N. Vu, C. A. Steele, J. McCane, M. R. Campbell, A. P. Matala, J. E. Hess, and S. R. Narum. 2014. Chinook and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam. Idaho Department of Fish and Game Report 14-01. Annual Report, BPA Project 2010-026-00.

Ackerman, M. W., N. Vu, K. K. Wright, J. McCane, M. R. Campbell, J. E. Hess, and S. R. Narum. 2016. Chinook and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam. Idaho Department of Fish and Game Report 16-03. Annual Report, BPA Project 2010-026-00.

Alexander, D. H., J. Novembre, and K Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655-1664.

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

Anderson, E. C., and J. C. Garza. 2005. A description of full parentage genotyping [online]. Report submitted to the Pacific Salmon Commission, Vancouver, British Columbia. Available from http://swfsc.noaa.gov/publications/FED/00675.pdf

Anderson, E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 65(7):1475-1486.

Anderson, E. C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Molecular Ecology Resources 10(4):701-710.

Beacham, T. D., S. Pollard, and K. D. Le. 2000. Microsatellite DNA population structure and stock identification of steelhead trout (Oncorhynchus mykiss) in the Nass and Skeena Rivers in northern British Columbia. Marine Biotechnology 2(6):587-600.

Beacham, T. D., J. R. Candy, K. J. Supernault, T. Ming, B. Deagle, A. Schulze, D. Tuck, K. H. Kaukinen, J. R. Irvine, K. M. Miller, and R. E. Withler. 2001. Evaluation and application

of microsatellite and major histocompatibility complex variation for stock identification of coho Salmon in British Columbia. Transactions of the American Fisheries Society 130(6):1116-1149.

Benjamini, Y., and D. Yekutieli. 2001. The control of false discovery rate in multiple testing under dependency. The Annals of Statistics 29:1165-1188.

Brown, E. 2016. Extended spawning ground surveys conducted for spring-summer Chinook salmon in the Chamberlain Creek population, Idaho, during 2015. *in* Stiefel, C., B. Barnett, K. K. Wright, K. A. Apperson, M. J. Belnap, R. Hand, M. Peterson, E. Ziolkowski, E. Brown, L. Hebdon, and W. C. Schrader. Idaho adult Chinook Salmon monitoring. Annual report 2015. Idaho Department of Fish and Game Report 16-12.

Busby, P. J., T. C. Wainwright, G. J. Bryant, L. J. Lierheimer, R. S. Waples, F. W. Waknitz, and I. L. Lagomarsino. 1996. Status review of (West Coast) steelhead from Washington, Idaho, Oregon, and California. National Marine Fisheries Technical Memorandum NMFS-NWFSC-27. Seattle.

Camacho, C. A., K. K. Wright, J. Powell, W. C. Schrader, T. Copeland, M. W. Ackerman, M. Dobos, M. P. Corsi, and M. R. Campbell. 2017. Wild adult steelhead and Chinook salmon abundance and composition at Lower Granite Dam, spawn years 2009-2016. Cumulative Report 2009 through 2016. Idaho Department of Fish and Game Report 17-06.

Camacho, C. A., J. Powell, M. Davison, M. E. Dobos, W. C. Schrader, T. Copeland, and M. R. Campbell. 2018a. Wild adult steelhead and Chinook Salmon abundance and composition at Lower Granite Dam, spawn year 2017. Annual Report. Idaho Department of Fish and Game Report 18-06.

Camacho, C. A., K. K. Wright, J. Powell, W. C. Schrader, T. Copeland, M. W. Ackerman, M. E. Dobos, M. Davison, M. P. Corsi, and M. R. Campbell. 2018b. Wild juvenile steelhead and Chinook salmon abundance and composition at Lower Granite Dam, migratory years 2010 and 2017. Cumulative Report 2010 through 2017. Idaho Department of Fish and Game Report 18-05.

Campbell, M. R., C. C. Kozfkay, T. Copeland, W. C. Schrader, M. W. Ackerman, and S. R. Narum. 2012. Estimating abundance and life history characteristics of threatened wild Snake River steelhead using genetic stock identification. Transactions of the American Fisheries Society. 141(5):1310-1327.

Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology resources 15:855-867.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis models and estimation procedures. American Journal of Human Genetics 19:233-257.

Cornuet, J. M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153:1989-2000.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximization likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39:1-38.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package), version 3.5c. Department of Genome Sciences, University of Washington, Seattle.

Ford, M. J., A. Albaugh, K. Barnas, T. Cooney, J. Cowen, J. Hard, R. Kope, M. M. McClure, P. McElhany, J. Myers, N. Sands, D. Teel, and L. A. Weitkamp. 2011. Status review update for Pacific salmon and steelhead listed under the Endangered Species Act: Pacific Northwest. US Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-113.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data Analysis, 2nd ed. Chapman & Hall/CRC Press, Boca Raton, Florida.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science 7:457-511.

Good, T. P., R. S. Waples, and P. B. Adams. 2005. Updated status of federally listed ESUs of West Coast Salmon and steelhead. U.S. Department of Commerce, NOAA Technical Memorandum, NMFS-NWFSC-66. 598 pp.

Goudet J., and T. Jombart. 2015. hierfstat: Estimation and Tests of Hierarchical F-Statistics. R package version 0.04-22. https://CRAN.R-project.org/package=hierfstat

Graffelman, J. 2015. Exploring Diallelic Genetic Markers: The Hardy-Weinberg Package. Journal of Statistical Software 64:1-22.

Graffelman, J., and J. Morales-Camarena. 2008. Graphical tests for Hardy-Weinberg Equilibrium based on the ternary plot. Human Heredity 65:77-84.

Habicht, C., L. W. Seeb, K. W. Myers, E. V. Farley, and J. E. Seeb. 2010. Summer-fall distribution of stocks of immature sockeye Salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. Transactions of the American Fisheries Society 139(4):1171-1191.

Hasselman, D. J., E. C. Anderson, E. E. Argo, N. D. Bethoney, S. R. Gephard, D. M. Post, B. P. Schondelmeier, T. F. Schultz, T. V. Willis, and E. P. Palkovacs. 2016. Genetic stock composition of marine bycatch reveals disproportional impacts on depleted river herring genetic stocks. Canadian Journal of Fisheries and Aquatic Sciences 73:951-963.

Hess, J. E., N. R. Campbell, A. P. Matala, S. R. Narum. 2013. 2012 Annual Report: Genetic Assessment of Columbia River Stocks. U.S. Dept. of Energy Bonneville Power Administration Report Project #2008-907-00.

Hess, J. E., N. R. Campbell, A. P. Matala, D. J. Hasselman, S. R. Narum. 2016. 2015 Annual Report: Genetic assessment of Columbia River stocks. U.S. Dept. of Energy Bonneville Power Administration Report Project #2008-907-00.

ICTRT (Interior Columbia Basin Technical Recovery Team). 2003. Independent Populations of Chinook, steelhead and sockeye for listed Interior Columbia Basin ESUs. Interior Columbia Basin Technical Recovery Team Report. July 2003.

ICTRT. 2007. Viability criteria for application to interior Columbia Basin salmonid USUs. Interior Columbia Basin Technical Recovery Team. Technical Review Draft. March 2003.

ICTRT. 2010. Current status reviews: Interior Columbia Basin salmon ESUs and steelhead DPSs. Vol. 1. Snake River ESUs/DPS. 786pp. + attachments

IDFG (Idaho Department of Fish and Game). 2007. Fisheries management plan 2007-2012. IDFG, Boise.

Jombart, T. 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Jombart T. and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070-3071.

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.

Kalinowski, S. T., K. R. Manlove, and M. L. Taper. 2007. ONCOR: A computer program for genetic stock identification. Montana State University, Bozeman. Available at http://www.montana.edu/kalinowski/software/oncor.html.

Latremouille, D. N. 2003. Fin erosion in aquaculture and natural environments. Reviews in Fisheries Science 11:315-335.

Matala, A. P., M. W. Ackerman, M. R. Campbell, and S. R. Narum. 2014. Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes. Evolutionary Applications 7:682-701.

McElhany, P., M. H. Ruckelshaus, M. J. Ford, T. C. Wainwright, and E. P. Bjorkstedt. 2000. Viable Salmonid populations and the recovery of evolutionary significant units. U.S. Department of Commerce, NOAA Technical Memo. NMFS-NWFSC-42, 156 p.

Moran, P. 2003. Genetic structure of Oncorhynchus mykiss populations in the Grande Ronde River, Imnaha River, and adjacent regions of the Snake River basin. Final report submitted to the U.S. Fish and Wildlife Service, Lower Snake River Compensation Plan Office, Boise, Idaho, in partial fulfillment of Contract No. 14110-1-H070. 28p. + Appendices.

Narum, S. R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics 7:783–787.

Narum, S. R., J. J. Stephenson, and M. R. Campbell. 2007. Genetic variation and structure of Chinook Salmon life history types in the Snake River. Transactions of the American Fisheries Society 136(5):1252-1262.

Narum, S. R., D. Hatch, A. J. Talbot, P. Moran, and M. S. Powell. 2008. Iteroparity in complex mating systems of steelhead Oncorhynchus mykiss (Walbaum). Journal of Fish Biology 72(1):45-60.

Nielsen, J. L., A. Byrne, S. L. Graziano, and C. C. Kozfkay. 2009. Steelhead Genetic Diversity at Multiple Spatial Scales in a Managed Basin: Snake River, Idaho. North American Journal of Fisheries Management 29(3):680-701.

NMFS (National Marine Fisheries Service). 2007. Biological Opinion - Remand Draft. Consultation on Remand for Operation of the Federal Columbia River Power System, 11 Bureau of Reclamation Projects in the Columbia Basin and ESA Section 10(a)(1)(A) Permit for Juvenile Fish Transportation Program (Revised and reissued pursuant to court order, NWF v. NMFS, Civ. No. CV 01-640-RE (D. Oregon)). National Marine Fisheries Service (NOAA Fisheries) - Northwest Region, Seattle. October 2007.

Paetkau, D., W. Calvert, I. Stirling, and C. Strobeck. 1995. Microsatellite analysis of population structure in Canadian polar bears. Molecular Ecology 4:347-354.

Pella, J., and M. Masuda. 2006. The Gibbs and split-merge sampler for population mixture analysis from genetic data with incomplete baselines. Canadian Journal of Fisheries and Aquatic Sciences 63: 576-596.

Pella, J., and M. Masuda. 2005. Classical discriminant analysis, classification of individuals, and source population composition of mixtures. *In* Stock identification methods: applications in fishery science. *Edited by* S. Cadrin, K. Friedland, and J. Waldman. Academic Press, New York. pp. 517–552.

Pella, J. J., and G. B. Milner. 1987. Use of genetic marks in stock composition analysis. Pages 274-276 *in* N. Ryman and F. Utter, editors. Population genetics and fisheries management. University of Washington Press, Seattle.

Powell, J. H., N. Vu, J. McCane, M. W. Ackerman, M. R. Campbell, D. J. Hasselman, and S. R. Narum. 2017. Chinook and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam. Idaho Department of Fish and Game Report 17-02. Annual Report, BPA Project 2010-026-00.

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

QCI (Quantitative Consultants, Inc.). 2013. Integrated status and effectiveness monitoring project: Salmon subbasin cumulative analysis report. Quantitative Consultants, Inc. Annual report 2012, BPA Project 2003-017-00.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A. 2012. Tree Figure Drawing Tool (FigTree) Version 1.4.0. Institute of Evolutionary Biology, University of Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the USA 94:9197–9221.

Robards, M. D., and T. P. Quinn. 2002. The migratory timing of adult summer-run steelhead in the Columbia River over six decades of environmental change. Transactions of the American Fisheries Society 131(3):523-536.

Satterthwaite, W. H., M. S. Mohr, M. R. O'Farrell, E. C. Anderson, M. A. Banks, S. J. Bates, M. R. Bellinger, L. A. Borgerson, E. D. Crandall, J. C. Garza, B. J. Kormos, P. W. Lawson, and M. L. Palmer-Zwahlen. 2014. Use of genetic stock identification data for comparison of the ocean spatial distribution, size at age, and fishery exposure of an untagged stock and its indicator: California coastal versus Klamath River Chinook Salmon. Transactions of the American Fisheries Society 143:117-133.

See, K., R. Orme, R. Kinzer, and M. W. Ackerman. 2016. PIT tag based escapement estimates to Snake Basin population. Prepared by ISEMP for the Bonneville Power Administration. Published by QCI. Available: https://www.researchgate.net/publication/303289911_PIT_Tag_Based_Escapement_Estimates_to_Snake_Basin_Populations. (December 2016).

Shaklee, J. B., T. D. Beacham, L. Seeb, and B. A. White. 1999. Managing fisheries using genetic data: case studies from four species of Pacific Salmon. Fisheries Research 43:45-78.

Smith, C. T., W. D. Templin, J. E. Seeb, and L. W. Seeb. 2005. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of US and Canadian Chinook Salmon caught in Yukon River fisheries. North American Journal of Fisheries Management 25(3):944-953.

Smith, C. T. 2007. Feasibility of genetic stock ID of Chinook Salmon sampled at Lower Granite Dam. U. S. Fish and Wildlife Service, Abernathy Fish Technology Center Report.

Smouse, P. E., R. S. Waples, and J. A. Tworek. 1990. A genetic mixture analysis for use with incomplete source population data. Canadian Journal of Fisheries and Aquatic Sciences 47: 620–634.

Stark, E. J., C. Bretz, A. Byrne, P. Cleary, T. Copeland, L. Denny, R. Engle, T. Miller, S. Rosenberg, E. R. Sedell, G. E. Shippentower, and C. Warren. 2016. Snake River Basin steelhead 2013/2014 Run Reconstruction. Report to Bonneville Power Administration, Portland, Oregon.

Steele, C. A., M. W. Ackerman, M. A. Hess, N. R. Campbell, S. R. Narum, and M. R. Campbell. 2013a. A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. Canadian Journal of Fisheries and Aquatic Science 70: 1046-1054.

Steele, C., J. McCane, M. Ackerman, M. Campbell, M. Hess, and S. Narum. 2013b. Parentage based tagging of Snake River hatchery steelhead and Chinook Salmon. Idaho Department of Fish and Game Report 13-23. Annual report 2013. BPA Project 2010-031-00.

Stiefel, C., B. Barnett, K. K. Wright, K. A. Apperson, M. J. Belnap, R. Hand, M. Peterson, E. Ziolkowski, E. Brown, L. Hebdon, and W. C. Schrader. 2016. Idaho adult Chinook Salmon monitoring. Annual report 2015. Idaho Department of Fish and Game Report 16-12.

Thurow, R. F. 1985. Middle Fork Salmon River Fisheries Investigations. Job Completion Report, Project F-73-R-6.

Thurow, R. F. 2000. Dynamics of Chinook Salmon populations within Idaho's Frank Church Wilderness - implications for persistence in McCool, S.F., and others, v3. Wilderness as a place for scientific inquiry: Proceedings of the wilderness science in a time of change conference, Missoula, Montana, May 23-27, 1999, U.S. Forest Service RMRS-P-15-VOL-3, p. 143-151.

Tiffan, K. F., D. W. Rondorf, R. D. Garland, and P. A. Verhey. 2000. Identification of fall versus spring Chinook Salmon migrating through the lower Snake River based on body morphology. Transactions of the American Fisheries Society 129:1389-1395.

Vu, N., M. W. Ackerman, K. K. Wright, J. McCane, M. R. Campbell, J. E. Hess, and S. R. Narum. 2015. Chinook and steelhead genotyping for genetic stock identification at Lower Granite Dam. Idaho Department of Fish and Game Report 15-02. Annual Report, BPA Project 2010-026-00.

Waples, R. S. 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity 89:438-450.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358-1370.

Wood, C. C., S. McKinnell, T. J. Mulligan, and D. A. Fournier. 1987. Stock identification with the maximum-likelihood mixture model: sensitivity analysis and application to complex problems. Canadian Journal of Fisheries and Aquatic Sciences 44:866-881.

Wright, K., W. Schrader, L. Reinhardt, K. Hernandez, C. Hohman, and T. Copeland. 2015. Process and methods for assigning ages to anadromous salmonids from scale samples. Idaho Department of Fish and Game Report 15-03.

TABLES

Table 1. Summary of GSI baselines and analysis methods employed to assign individual steelhead and Chinook Salmon to their reporting unit of origin from 2009–2017.

Year	Species	Baseline Version	Loci	Individuals	Collections	Populations	Reporting Units	Assignment Method	Citation
2009	STHD	1.0	191	2514	52	49	10	*ONCOR* Individual Assignment	Ackerman et al. (2011)
	CHNK	1.0	141	2390	54	32	5		
2010	STHD	1.0	191	2514	52	49	10		
	CHNK	1.0	141	2390	54	32	5		
2011	STHD	2.0	191	4145	83	63	10	*gsi_sim* Mixture Modeling	Ackerman et al. (2012)
	CHNK	2.0	191	3392	111	39	7		
2012	STHD	2.0	191	4116	85	63	10		Ackerman et al. (2014)
	CHNK	2.0	191	3327	110	39	7		
2013	STHD	3.0	185	8028	139	68	10		Vu et al. (2015)
	CHNK	3.0	180	6151	151	57	7		
2014	STHD	3.1	185	6150	136	66	10	*gsi_sim* Individual Assignment	Ackerman et al. (2016)
	CHNK	3.1	180	4604	148	46	7		
2015	STHD	3.1	185	6150	136	66	10		Powell et al. (2017)
	CHNK	3.1	180	4604	148	46	7		
2016	STHD	3.1	179	5967	136	45	10		This report
	CHNK	3.1	173	4356	148	30	7		
2017	STHD	3.1	179	5967	136	45	10		
	CHNK	3.1	173	4356	148	30	7		

Table 2. Summary of SY2016-SY2017 adult and MY2016-MY2017 juvenile steelhead and Chinook Salmon samples from Lower Granite Dam (LGR). Summary includes the number of samples that arrived from LGR (inventoried) and the number inventoried that were queued for genotyping. Of queued samples, we show the number that genotyped successfully and the number that failed genotyping. For samples that genotyped successfully, we show the number that had a parentage based tag (PBT) and the number that were assigned a genetic stock based on individual assignment (IA) using SNP baselines v3.1

Sample Group	Total Samples Inventoried	Samples Queued for Genotyping	Failed Genotyping (NG)	Successfully Genotyped	PBT Assignments	GSI Assignments
Steelhead						
SY2016 Adults (Natural-origin Phenotype)	4,830	4,830	86 (1.8%)	4,744 (98.2%)	436 (9.2%)	4,308 (90.8%)
SY2016 Adults (Hatchery Phenotype)	407	407	2 (0.5%)	405 (99.5%)	375 (92.6%)	30 (7.4%)
MY2016 Juveniles	1,243	1,243	12 (1.0%)	1,231 (99.0%)	31 (2.5%)	1,200 (97.5%)
SY2017 Adults (Natural-origin Phenotype)	3,951	3,951	16 (0.4%)	3,935 (99.6%)	852 (21.7%)	3,083 (78.3%)
SY2017 Adults (Hatchery Phenotype)	780	780	2 (0.3%)	778 (99.7%)	727 (93.4%)	51 (6.6%)
MY2017 Juveniles	1,502	1,502	28 (1.9%)	1,474 (98.1%)	20 (1.4%)	1,454 (98.6%)
Chinook						
SY2016 Adults	4,538	4,538	49 (1.1%)	4,489 (98.9%)	1,409 (31.4%)	3,080 (68.6%)
MY2016 Juveniles	1,516	1,512	41 (2.7%)	1,475 (97.3%)	171 (11.6%)	1,304 (88.4%)
SY2017 Adults	2,060	2,060	5 (0.2%)	2,055 (99.8%)	828 (40.3%)	1,227(59.7%)
MY2017 Juveniles	1,816	1,816	16 (0.9%)	1,800 (99.1%)	255 (14.2%)	1,545 (85.8%)
TOTAL:	22,643	22,643	257 (1.1%)	22,386 (98.9%)	5,104 (22.8%)	17,282 (77.2%)

Table 3. Summary of 7,391 Lower Granite Dam (LGR) adult steelhead samples from SY2016-SY2017 assigned to a genetic stock using individual assignment based on Snake River steelhead SNP baseline v3.1. Summaries of life history diversity information (sex, length, and ocean age) for each genetic stock are shown. The 'Other' saltwater age category includes fish that were not queued to be aged, fish that could not be aged, and fish with spawn checks.

				Sex					Length							Ocean (Saltwater) Age						
				Frequency			Percentage		Mean Length (cm FL) by Ocean Age			Frequency		Percentage		Frequency				Percentage		
SY	Genetic Stock	Total Assignments	% Stock Composition	F	M	U	F	M	1	2	3	A-Run	B-Run	A-Run	B-Run	1	2	3	Other	1	2	3
2016	UPSALM	606	14%	420	186	0	69%	31%	58.7	68.3	75	601	3	100%	0%	174	219	1	212	44%	56%	0%
	MFSALM	290	7%	219	71	0	76%	24%	61	72.6	82	251	39	87%	13%	52	120	3	115	30%	69%	2%
	SFSALM	165	4%	118	47	0	72%	28%	64.1	75.8	85	112	53	68%	32%	27	102	1	35	21%	78%	1%
	LOSALM	145	3%	100	45	0	69%	31%	58.9	69	-	142	3	98%	2%	46	51	0	48	47%	53%	0%
	UPCLWR	454	11%	315	139	0	69%	31%	63.7	77.3	84.2	279	175	61%	39%	39	198	4	213	16%	82%	2%
	SFCLWR	271	6%	177	94	0	65%	35%	63.2	77.2	88.5	159	112	59%	41%	21	170	4	76	11%	87%	2%
	LOCLWR	465	11%	313	152	0	67%	33%	60	70.4	-	446	19	96%	4%	104	200	0	161	34%	66%	0%
	IMNAHA	328	8%	222	106	0	68%	32%	58.7	68.6	-	325	3	99%	1%	108	149	0	71	42%	58%	0%
	GRROND	1074	25%	714	360	0	66%	34%	58.6	68.3	-	1065	9	99%	1%	275	448	0	351	38%	62%	0%
	LSNAKE	510	12%	339	171	0	66%	34%	58.7	68.5	78	500	10	98%	2%	124	196	1	189	39%	61%	0%
SY2016 Total:		*4308*		*2937*	*1371*	*0*	*68%*	*32%*	*59.4*	*71.1*	*83.9*	*3880*	*426*	*90%*	*10%*	*970*	*1853*	*14*	*1471*	*34%*	*65%*	*0%*
2017	UPSALM	232	8%	156	72	4	68%	32%	57	69.7	-	222	8	97%	3%	27	123	0	82	18%	82%	0%
	MFSALM	200	6%	162	35	3	82%	18%	62.2	73.5	-	145	53	73%	27%	6	112	0	82	5%	95%	0%
	SFSALM	126	4%	103	21	2	83%	17%	56	77.4	82.6	65	61	52%	48%	1	99	5	21	1%	94%	5%
	LOSALM	32	1%	18	12	2	60%	40%	50.5	71.5	-	28	4	88%	12%	2	17	0	13	11%	89%	0%
	UPCLWR	432	14%	335	84	13	80%	20%	57	78.3	86	177	248	42%	58%	1	282	4	145	0%	98%	1%
	SFCLWR	211	7%	152	54	5	74%	26%	64.5	78.4	84.6	94	113	45%	55%	2	150	5	54	1%	96%	3%
	LOCLWR	223	7%	168	52	3	76%	24%	57.7	72.3	-	185	34	84%	16%	10	129	0	84	7%	93%	0%
	IMNAHA	199	6%	150	48	1	76%	24%	56.9	69	-	194	4	98%	2%	29	124	0	46	19%	81%	0%
	GRROND	1008	33%	742	238	28	76%	24%	59.1	70	84	951	42	96%	4%	114	534	1	359	18%	82%	0%
	LSNAKE	420	14%	285	127	8	69%	31%	58.9	69.7	-	397	20	95%	5%	52	239	0	129	18%	82%	0%
SY2017 Total:		*3083*		*2271*	*743*	*69*	*75%*	*25%*	*58.5*	*72.7*	*84.3*	*2458*	*587*	*81%*	*19%*	*244*	*1809*	*15*	*1015*	*12%*	*87%*	*1%*

Table 4. Summary of 2,654 Lower Granite Dam (LGR) juvenile steelhead samples from MY2016-MY2017 assigned to a genetic stock using individual assignment based on Snake River steelhead SNP baseline v3.1. Summaries of life history diversity information (sex, length, and freshwater age) for each genetic stock are shown. The 'Other' freshwater age category includes fish that were not queued to be aged or could not be aged.

				Sex					Length	Freshwater Age										
				Frequency			Percentage			Frequency						Percentage				
MY	**Genetic Stock**	**Total Assignments**	**% Stock Composition**	**F**	**M**	**U**	**F**	**M**	**Mean Length (mm FL)**	**1**	**2**	**3**	**4**	**5**	**Other**	**1**	**2**	**3**	**4**	**5**
2016	UPSALM	191	16%	109	82	0	57%	43%	177	13	88	49	2	0	39	9%	58%	32%	1%	0%
	MFSALM	114	10%	77	37	0	68%	32%	178	1	31	60	12	0	10	1%	30%	58%	12%	0%
	SFSALM	78	6%	52	26	0	67%	33%	183	0	20	44	9	1	4	0%	27%	59%	12%	1%
	LOSALM	52	4%	35	17	0	67%	33%	175	1	26	19	2	0	4	2%	54%	40%	4%	0%
	UPCLWR	138	12%	99	39	0	72%	28%	178	1	42	71	20	0	4	1%	31%	53%	15%	0%
	SFCLWR	93	8%	56	37	0	60%	40%	169	5	63	14	1	0	10	6%	76%	17%	1%	0%
	LOCLWR	99	8%	68	31	0	69%	31%	176	7	53	23	2	0	14	8%	62%	27%	2%	0%
	IMNAHA	108	9%	74	34	0	69%	31%	180	6	55	29	2	0	16	7%	60%	32%	2%	0%
	GRROND	210	18%	142	68	0	68%	32%	184	12	107	55	2	0	34	7%	61%	31%	1%	0%
	LSNAKE	117	10%	68	49	0	58%	42%	178	7	64	27	2	0	17	7%	64%	27%	2%	0%
MY2016 Total:		*1,200*		*780*	*420*	*0*	*65%*	*35%*	*178*	*53*	*549*	*391*	*54*	*1*	*152*	*5%*	*52%*	*37%*	*5%*	*0%*
2017	UPSALM	343	24%	208	129	6	62%	38%	174	60	218	52	3	1	9	18%	65%	16%	1%	0%
	MFSALM	68	5%	46	21	1	69%	31%	176	2	28	33	1	0	4	3%	44%	52%	2%	0%
	SFSALM	26	2%	15	11	0	58%	42%	180	0	8	16	2	0	0	0%	31%	62%	8%	0%
	LOSALM	24	2%	16	8	0	67%	33%	167	2	16	5	0	0	1	9%	70%	22%	0%	0%
	UPCLWR	129	9%	73	55	1	57%	43%	171	5	81	32	7	0	4	4%	65%	26%	6%	0%
	SFCLWR	126	9%	82	42	2	66%	34%	166	19	89	12	0	0	6	16%	74%	10%	0%	0%
	LOCLWR	120	8%	64	54	2	54%	46%	169	22	79	15	0	0	4	19%	68%	13%	0%	0%
	IMNAHA	102	7%	70	30	2	70%	30%	172	10	71	18	1	0	2	10%	71%	18%	1%	0%
	GRROND	352	24%	229	119	4	66%	34%	175	38	232	66	6	0	10	11%	68%	19%	2%	0%
	LSNAKE	164	11%	89	70	5	56%	44%	173	28	107	21	2	0	6	18%	68%	13%	1%	0%
MY2017 Total:		*1,454*		*892*	*539*	*23*	*62%*	*38%*	*173*	*186*	*929*	*270*	*22*	*1*	*46*	*13%*	*66%*	*19%*	*2%*	*0%*

Table 5. Summary of Lower Granite Dam (LGR) adult Chinook Salmon samples from SY2016-SY2017 assigned to a genetic stock using individual assignment based on Snake River Chinook Salmon SNP baseline v3.1. Summaries of life history diversity information (sex, length, and ocean age) for each genetic stock are shown. MJ = minijack.

				Sex						Ocean (Saltwater) Age									Length			
				Frequency			Percentage			Frequency					Percentage					Mean Length (cm FL) by Ocean Age			
SY	Genetic Stock	Total Assignments	% Stock Composition	F	M	U	F	M	MJ	1	2	3	U	MJ	1	2	3	MJ	1	2	3
	UPSALM	614	20%	354	254	6	58%	42%	0	12	270	232	100	0%	2%	53%	45%	-	51.2	74.3	87.6
	MFSALM	448	15%	259	189	0	58%	42%	0	14	172	137	125	0%	4%	53%	42%	-	53	73.6	87.7
	CHMBLN	51	2%	25	24	2	51%	49%	0	4	22	6	19	0%	12%	69%	19%	-	51	74.9	83
2016	SFSALM	359	12%	190	167	2	53%	47%	0	14	203	100	42	0%	4%	64%	32%	-	55.2	74.7	86.7
	HELLSC	1,254	41%	687	561	6	55%	45%	0	34	704	239	277	0%	3%	72%	24%	-	53.7	72.5	84.5
	TUCANO	8	0%	6	2	0	75%	25%	0	0	5	1	2	0%	0%	83%	17%	-	-	71	85
	FALL	346	11%	183	157	6	54%	46%	5	21	79	82	159	3%	11%	42%	44%	37.6	54.3	79.5	85.7
SY2016 Total:		*3,080*		*1,704*	*1,354*	*22*	*56%*	*44%*	*5*	*99*	*1,455*	*797*	*724*	*0%*	*4%*	*62%*	*34%*	*37.6*	*53.5*	*73.7*	*86.3*
	UPSALM	193	16%	105	88	0	54%	46%	0	45	69	75	4	0%	24%	37%	40%	-	49.1	78.6	93.3
	MFSALM	165	13%	55	110	0	33%	67%	0	58	71	29	7	0%	37%	45%	18%	-	50.1	75.4	93
	CHMBLN	28	2%	13	15	0	46%	54%	0	8	14	4	2	0%	31%	54%	15%	-	52.9	75.6	89.5
2017	SFSALM	206	17%	82	123	1	40%	60%	0	77	95	28	6	0%	38%	48%	14%	-	53.3	77.2	89.4
	HELLSC	479	39%	195	284	0	41%	59%	0	123	276	65	15	0%	27%	59%	14%	-	50.4	74.3	88.8
	TUCANO	9	1%	2	7	0	22%	78%	0	3	5	0	1	0%	38%	62%	0%	-	45.7	69.6	-
	FALL	147	12%	81	66	0	55%	45%	8	6	58	67	8	6%	4%	42%	48%	36.6	55.5	77.4	86.1
SY2017 Total:		*1227*		*533*	*693*	*1*	*43%*	*57%*	*8*	*320*	*588*	*268*	*43*	*1%*	*27%*	*50%*	*23%*	*36.6*	*51.0*	*75.7*	*89.9*

Table 6. Summary of Lower Granite Dam (LGR) juvenile Chinook Salmon samples from MY2016-MY2017 assigned to a genetic stock using individual assignment based on Snake River Chinook Salmon SNP baseline v3.1. Summaries of life history diversity information (sex and length) by genetic stock are shown. Freshwater age is not summarized because scales were not collected from juvenile Chinook Salmon at Lower Granite Dam.

				Sex					Length
				Frequency			Percentage		
MY	Genetic Stock	Total Assignments	% Stock Composition	F	M	U	F	M	Mean Length (mm FL)
	UPSALM	206	16%	114	92	0	55%	45%	113
	MFSALM	226	17%	123	103	0	54%	46%	109
	CHMBLN	29	2%	20	9	0	69%	31%	107
2016	SFSALM	212	16%	92	120	0	43%	57%	110
	HELLSC	590	45%	307	283	0	52%	48%	112
	TUCANO	5	0%	3	2	0	60%	40%	117
	FALL	36	3%	23	12	1	66%	34%	122
MY2016 Total:		*1304*		*682*	*621*	*1*	*52%*	*48%*	*111*
	UPSALM	295	19%	157	138	0	53%	47%	110
	MFSALM	237	15%	114	123	0	48%	52%	106
	CHMBLN	11	1%	4	7	0	36%	64%	104
2017	SFSALM	285	18%	148	137	0	52%	48%	106
	HELLSC	710	46%	407	302	1	57%	43%	111
	TUCANO	0	0%	0	0	0	-	-	-
	FALL	7	0%	4	3	0	57%	43%	119
MY2017 Total:		*1545*		*834*	*710*	*1*	*54%*	*46%*	*109*

FIGURES

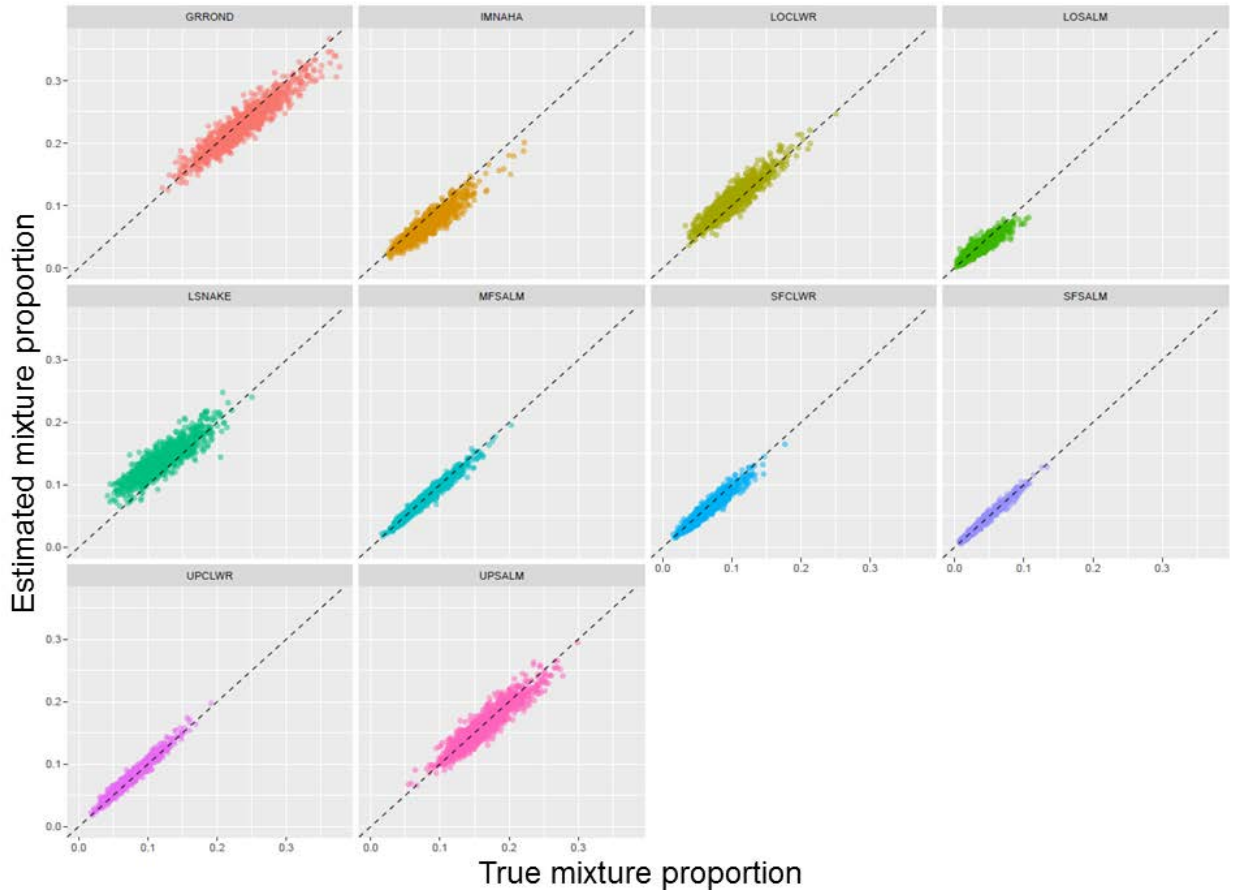

Figure 1. The observed bias in the estimated mixing proportions of steelhead reporting units due to imbalanced numbers of populations calculated using leave-one-out cross validation in the R package *rubias*. The true simulated mixture proportions are plotted on the x-axis and the estimated proportions are plotted on the y-axis. One-to-one lines are added as reference with positive bias in estimated mixture proportions appearing as points above the line and negative bias as points below the line.

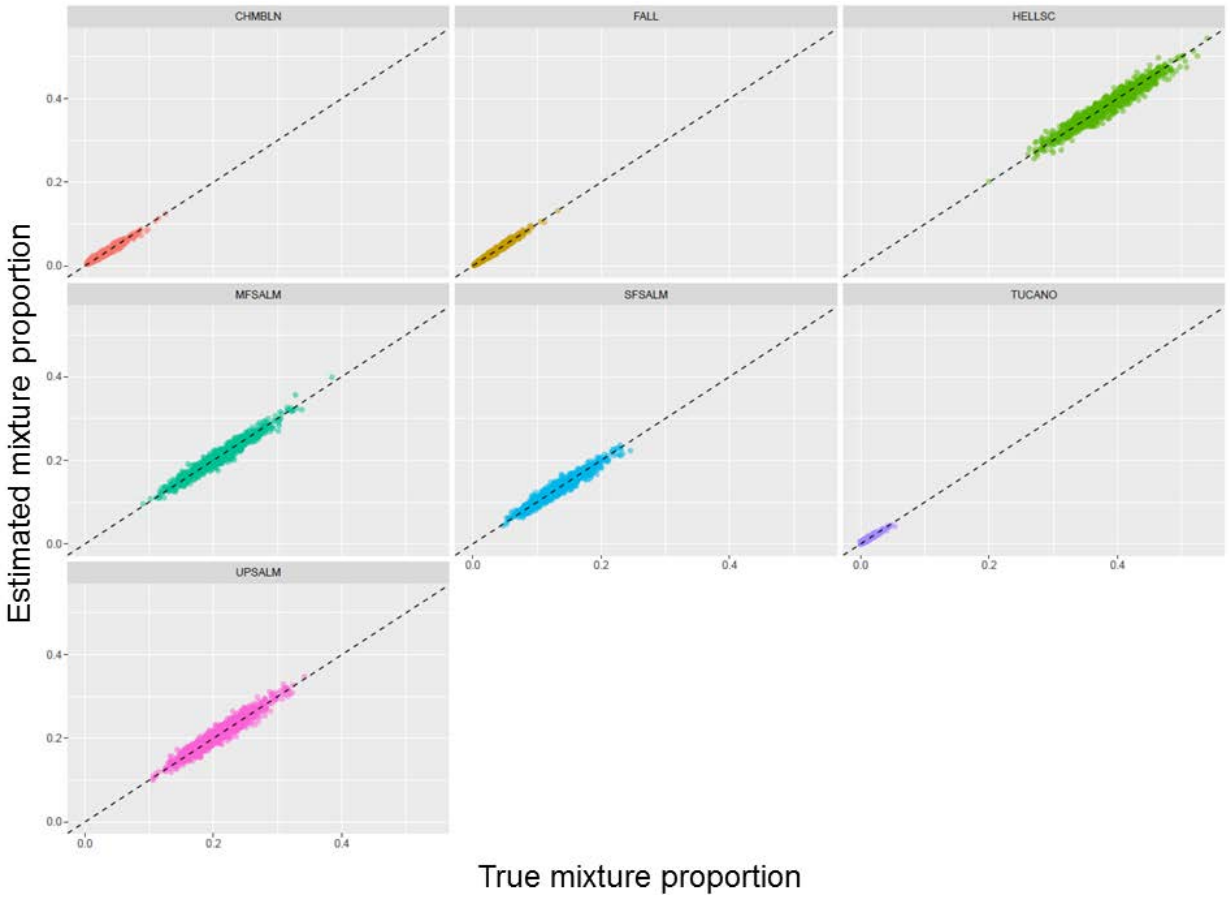

Figure 2. The observed bias in the estimated mixing proportions of Chinook Salmon reporting units due to imbalanced numbers of populations calculated using leave-one-out cross validation in the R package *rubias*. The true simulated mixture proportions are plotted on the x-axis and the estimated proportions are plotted on the y-axis. One-to-one lines are added as reference with positive bias in estimated mixture proportions appearing as points above the line and negative bias as points below the line.

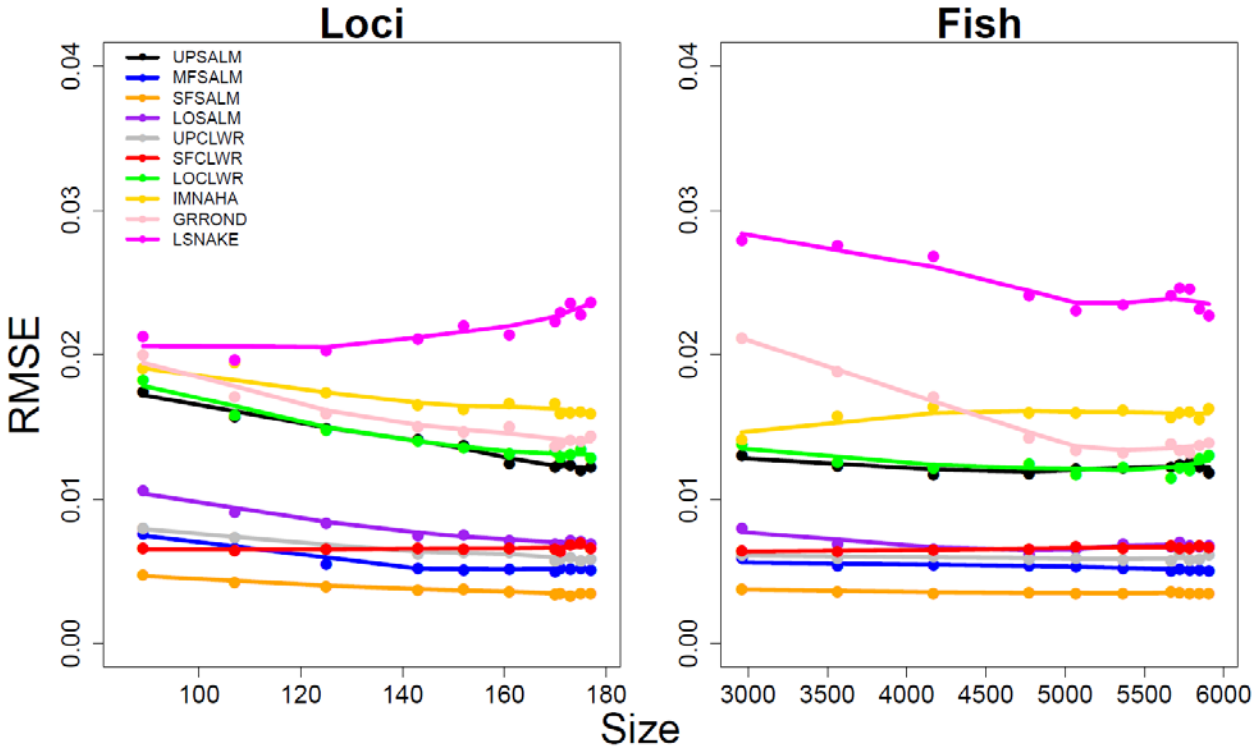

Figure 3. The root mean squared error of the estimated mixing proportions of Snake River steelhead reporting units at various baseline sizes. We plot baseline size in total number of loci or fish on the x-axis and the root mean squared error in the estimated mixture proportions on the y-axis. Reporting units are color-coded as reported in the legend at the top of the left panel.

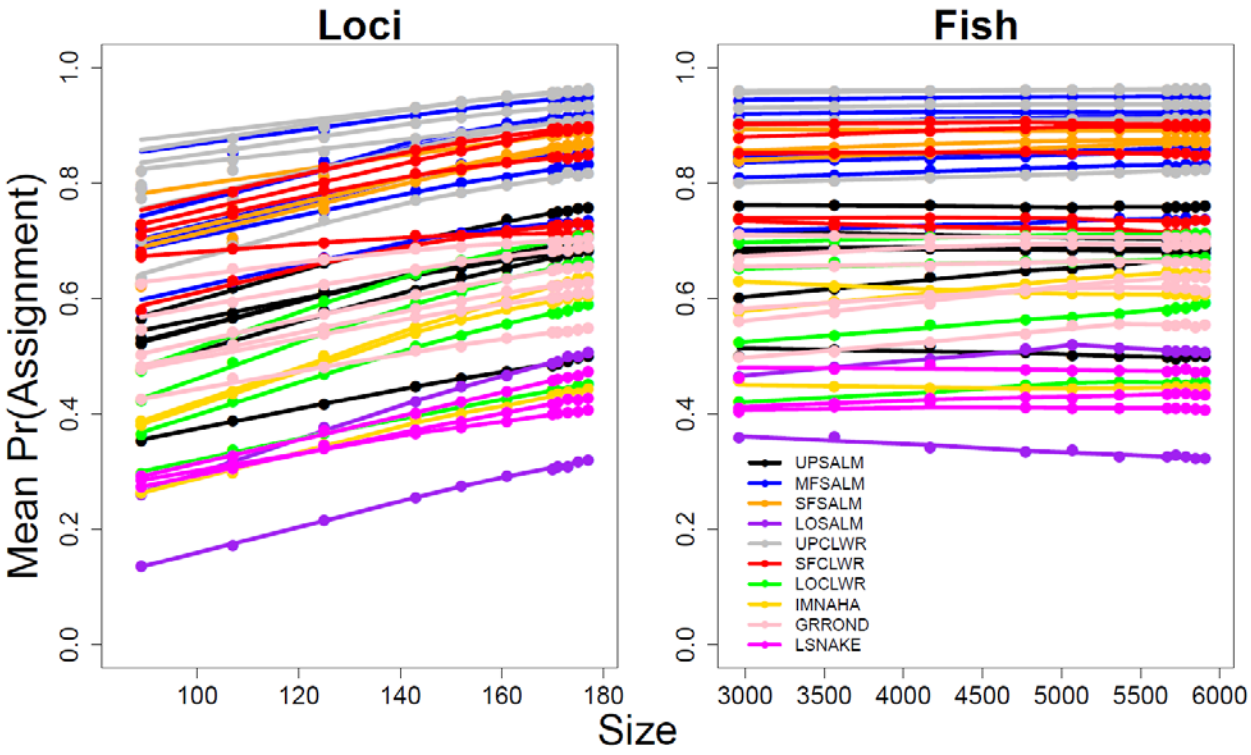

Figure 4. The estimated mean posterior probability of assignment to an individual population within a reporting unit in the Snake River steelhead GSI baseline version 3.1. We plot the baseline size in total number of loci or fish on the x-axis and the mean individual posterior probability of assignment on the y-axis. Reporting units are color-coded as reported in the legend at the bottom of the right panel.

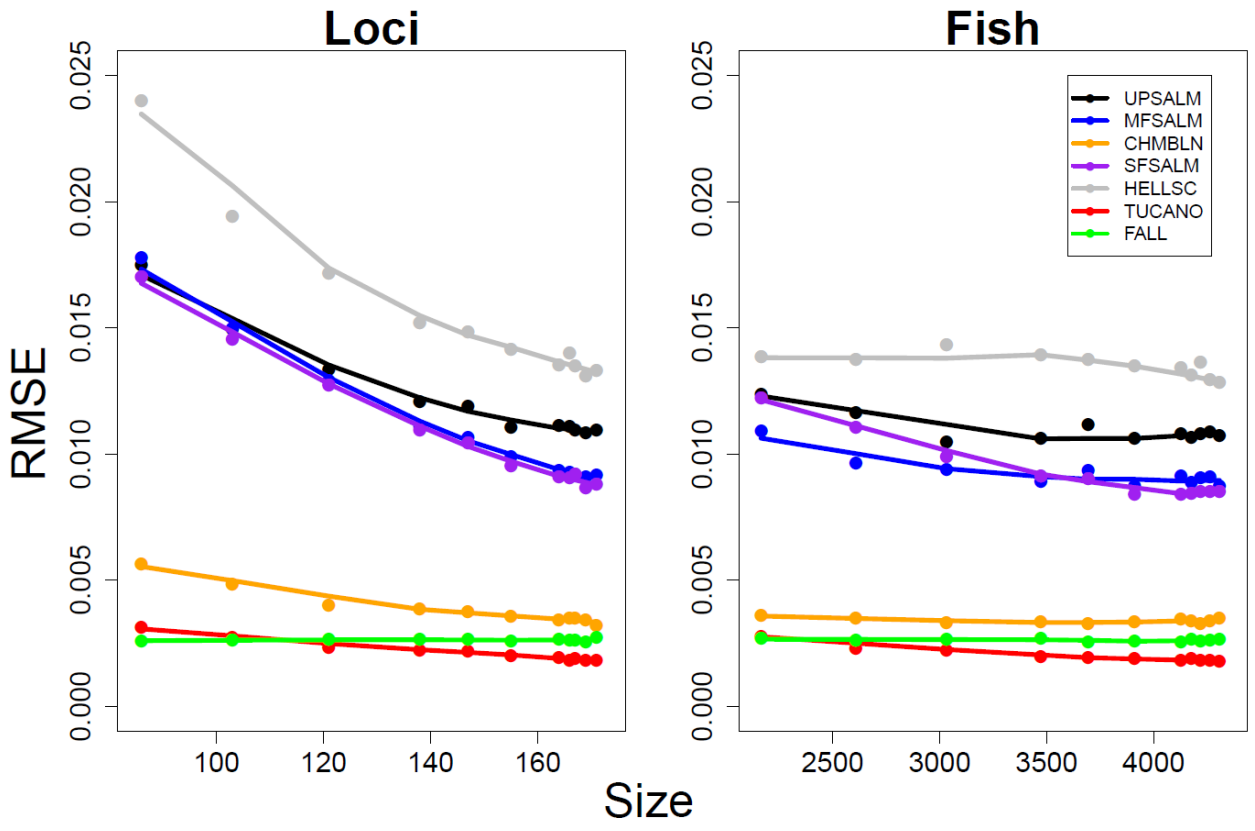

Figure 5. The root mean squared error of the estimated mixing proportions of Snake River Chinook Salmon reporting units at various baseline sizes. We plot baseline size in total number of loci or fish on the x-axis and the root mean squared error in the estimated mixture proportions on the y-axis. Reporting units are color-coded as reported in the legend at the top of the right panel.

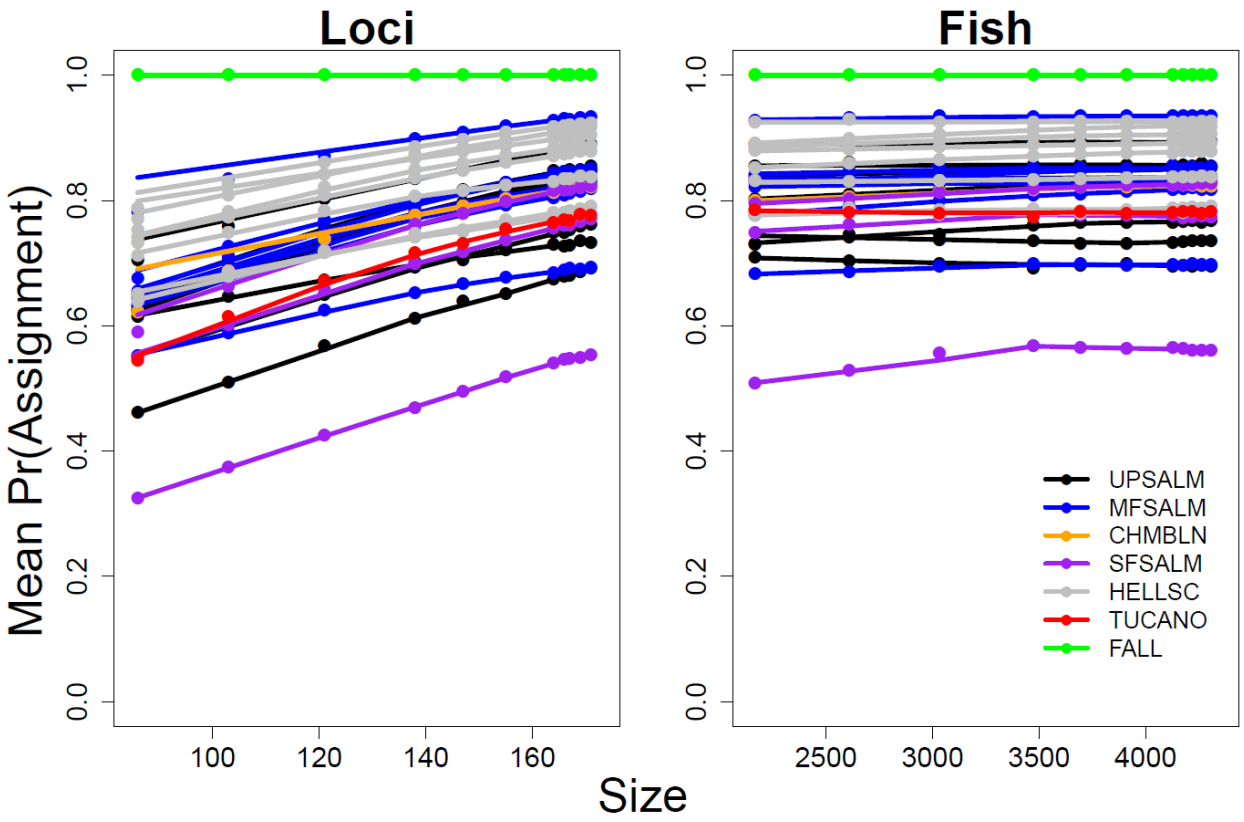

Figure 6. The estimated mean posterior probability of assignment to an individual population within a reporting unit in the Snake River Chinook Salmon GSI baseline version 3.1 baseline sizes. We plot the baseline size in total number of loci or fish on the x-axis and the mean individual posterior probability of assignment on the y-axis. Reporting units are color-coded as reported in the legend at the bottom of the right panel.

APPENDIX A

INTRODUCTION

Multiple statistical models have been developed to classify individuals sampled from a mixture to their source populations (see Pella and Masuda [2005] for a review). For genetic based assignments, the general process involves genotyping individuals collected directly from the source population and using these genotypes to estimate allele frequencies in a set number of baseline populations (Anderson et al. 2008). These baseline allele frequencies are then used to calculate the probability that a given individual's genotype came from a sample in the baseline (Anderson et al. 2008). Methods are also available that allow for the assignment of individuals when no baseline samples have been collected (e.g., the STRUCTURE algorithm developed by Pritchard et al. 2000). When baseline samples are not available, samples from the mixture are used to estimate allele frequencies in the source populations under the assumption of Hardy-Weinberg Linkage Equilibrium (Pella and Masuda 2006). When baseline samples are available, a decision must be made as to how the allele frequencies estimated in these baseline samples will be handled. If conditional estimation is used, these frequencies are fixed and treated as being measured without error (Pella and Masuda 2005). In unconditional estimation the mixture sample is allowed to influence baseline allele frequencies based on the assignment of individuals back to a population of interest (Pella and Masuda 2005). Pella and Masuda (2005) state that unconditional estimation makes more use of information contained in the mixture sample than conditional estimation. Smouse et al. (1990) indicate that unconditional estimation is preferable when baseline samples are small (e.g., >100 individuals), because it is unlikely that the allele frequency distributions in these samples are estimated without error.

The program *gsi_sim* (Anderson et al. 2008, Anderson 2010) is used to assign individuals collected at Lower Granite Dam to their stock of origin. Depending on the method employed in *gsi_sim* there are two unknown quantities that are estimated. One is the mixture proportions of the samples collected at Lower Granite Dam from given reporting unit. The other is the population of origin for an individual sample. There are three options for how *gsi_sim* estimates these quantities. Satterthwaite et al. (2014) state that Gibbs sampling is used to estimate the mixing proportions in the sample and the probability that an individual came from a specific population simultaneously using the likelihood outlined in Rannala and Mountain (1997, hereafter referred to as Bayesian). Alternatively, the expectation maximization (EM) algorithm (Dempster et al. 1977) can be used to estimate the unconditional likelihood of mixing proportions and individual population of origin as described by Smouse et al. (1990) and (Satterthwaite et al. 2014, hereafter referred to as MLE). Finally, the multinomial likelihood of the observed genotype arising from each population (calculated following Paetkau et al. 1995 as reported in Satterthwaite et al. 2014) can be scaled to sum to one with the probability of assignment equal to the scaled likelihood. This is the same as the frequency method employed by the program Geneclass (Cornuet et al. 1999, hereafter referred to as Geneclass).

In addition to estimating the mixture proportions and individual assignments at the population level, *gsi_sim* also estimates these quantities for user specified reporting units. These reporting units are groups of populations specified by the user. To calculate the probability of membership to a reporting unit, individual population estimates are first calculated and then summed (allocate-sum procedure; Wood et al. 1987). Anderson et al. (2008) found higher accuracy for estimating reporting unit contributions than individual population contributions when increased divergence is available between reporting units. For example, if populations in the South Fork Clearwater River are more similar to each other than to populations from other watersheds there is increased accuracy by grouping these populations into a single reporting unit.

In this appendix, we investigate the accuracy and precision of individual reporting unit assignments and genetic stock mixture proportions estimated using the different methods available in the program *gsi_sim*.

METHODS

Prior to spawn and migration year 2017, GSI assignments were made based on the probability of reporting unit membership calculated from the Geneclass method. We tested for differences in the assignment of individuals from our baseline collections due to assignment methods provided by *gsi_sim* (Anderson et al. 2008, Anderson et al. 2010) for both the steelhead baseline version 3.1 and Chinook Salmon baseline version 3.1. We initially randomly removed 40% of the individuals from each baseline population (62 populations in the steelhead baseline, and 40 populations in the Chinook Salmon baseline). We then randomly sampled individuals from the removed populations in proportion to their reporting unit's contribution to the average mixture observed passing LGR SY2013–SY2015. Thus, we built mixtures from the removed individuals that mimicked the average mixture observed during the previous three spawning migrations of adult steelhead and Chinook Salmon. We assigned these mixtures of individuals back to the reduced baseline with the Bayesian, MLE, or Geneclass methods in gsi_sim. For the Bayesian assignments, three individual chains were run for 300,000 total iterations, discarding the first 50,000 iterations. Convergence was assessed by visual inspection of the posterior distributions and trace plots of the mixture proportions in the sample (π), as well as calculating the Gelman-Rubin statistic (Gelman and Rubin [1992] as presented in Gelman et al. 2004), and effective number of draws (Gelman et al. 2004) for this parameter. We calculated cluster membership both proportional to the probability of membership in a reporting unit and based on the maximum probability of membership rule currently employed. We then repeated this test 99 more times, resulting in 100 random samples from the baseline.

We assessed assignment accuracy four ways. First, we calculated bias in the estimated mixing proportions. We also assessed accuracy of the estimated mixing proportions by calculating the mean squared error (MSE) of the estimates of the mixing proportions across simulations. Bias and MSE were calculated for all individual assignment methods, as well as for the mixing proportions estimated from the mixture model step in both the Bayesian and unconditional MLE methods. We also assessed the goodness-of-fit of estimated mixing proportions using X^2 tests for all individual assignment methods. Finally, we calculated the proportion of samples that self-assigned to the correct reporting unit under each individual assignment method.

RESULTS

Individual assignment probabilities within baseline populations were consistent between all methods (Table A1), but the maximum individual assignment method did produce slightly higher self-assignment rates in both Chinook Salmon and steelhead (Table A1). The MLE method had the highest self-assignment rates of all three methods (Table A1). However, proportional estimation often resulted in less biased (Figures A1 and A2), more accurate (Figures A3 and A4) estimates of the true proportion of fish from each reporting unit in both Chinook Salmon and steelhead. In general proportional estimation also produced less divergent estimates of the overall mixture proportions (Table A2). This was not the case for the Geneclass method in either Chinook Salmon or steelhead (Table A2). However, proportional estimation was the only method that did not produce a single divergent estimate of the mixture proportions (Table A2).

Assignment probabilities within baseline populations were similar among all methods for steelhead (Table A3). In general the assignment of individuals was increased in the Grand Ronde River and lower Snake River reporting units when using the Bayesian and MLE estimates, but was reduced in the lower Salmon River, lower Clearwater River, and Imnaha River reporting units (Table A3).

Assignment probabilities within baseline populations were also similar among all methods for Chinook Salmon (Table A4). The assignment of individuals was increased in the Hells Canyon reporting unit when using the Bayesian and MLE estimates, but was reduced in the upper Salmon River, Chamberlain Creek, and Tucannon River reporting units (Table A4). No spring/summer Chinook Salmon were assigned to the fall reporting unit, and no fall Chinook Salmon were assigned to a spring/summer reporting unit regardless of estimation method.

DISCUSSION

We observed that all methods used in the *gsi_sim* program appropriately characterize the underlying mixture proportions over Lower Granite Dam. The Bayesian and MLE estimators assigned samples from reporting units with low differentiation and high stock composition proportions with higher fidelity (e.g., Grande Ronde River steelhead and Hells Canyon Chinook Salmon). Estimating mixture proportions was improved by using proportional allocation of individuals. Beginning in spawn and migration year 2017 we will individually assign genetic stock using the full MLE method of Smouse et al. (1990) as implemented in *gsi_sim* (Anderson et al. 2008; Anderson 2010). Going forward we will also explore methods for incorporating proportional allocation into abundance estimation.

Table A1. The weighted average proportion of Chinook Salmon and steelhead baseline samples assigned to their reporting unit using a given method.

		CHNK	STHD
Maximum Individual Assignment	Bayesian	0.859	0.685
	MLE	0.859	0.686
	Geneclass	0.854	0.682
Proportional Individual Assignment	Bayesian	0.818	0.630
	MLE	0.819	0.632
	Geneclass	0.809	0.627

Table A2. The proportion of simulations with estimated Chinook Salmon and steelhead mixture proportions statistically different (α=0.05) from the true simulated value.

		CHNK	STHD
Maximum Individual Assignment	Bayesian	0.06	0.17
	MLE	0.06	0.13
	Geneclass	0.08	0.04
Proportional Individual Assignment	Bayesian	0.00	0.11
	MLE	0.00	0.12
	Geneclass	0.16	0.11

Table A3. The proportion of steelhead baseline samples assigned to their reporting unit using a given method. Cells are color-coded with warmer colors (red) indicating higher self-assignment rates and cooler colors (blue) indicating lower rates.

Reporting Unit	Collection	n	Maximum Individual Assignment			Proportional Individual Assignment		
			Bayesian	MLE	Geneclass	Bayesian	MLE	Geneclass
UPSALM	Sawtooth Weir	2784	0.802	0.801	0.739	0.714	0.712	0.643
	Valley Cr.	2445	0.752	0.753	0.710	0.671	0.671	0.612
	WF Yankee Fork	3071	0.781	0.778	0.715	0.685	0.683	0.615
	Herd Cr.	2200	0.700	0.698	0.708	0.658	0.660	0.667
	Morgan Cr.	1600	0.749	0.753	0.820	0.686	0.693	0.749
	Pahsimeroi R.	2600	0.740	0.747	0.767	0.672	0.676	0.691
	Lemhi R.	2300	0.738	0.743	0.775	0.655	0.659	0.690
	NF Salmon R.	2600	0.530	0.542	0.543	0.466	0.473	0.475
MFSALM	Capehorn/Marsh Cr.	1500	0.748	0.751	0.797	0.726	0.728	0.770
	Elk/Bear Valley Cr.	1300	0.963	0.963	0.968	0.951	0.951	0.962
	Sulphur Cr.	691	0.942	0.939	0.952	0.907	0.906	0.923
	Rapid R. (MF Salmon)	563	0.863	0.865	0.885	0.840	0.841	0.863
	Pistol Cr.	446	0.814	0.812	0.830	0.792	0.792	0.823
	Loon Cr.	1000	0.920	0.922	0.943	0.891	0.895	0.923
	Camas Cr.	800	0.928	0.931	0.960	0.901	0.907	0.941
	upper Big Cr.	665	0.971	0.971	0.979	0.948	0.948	0.963
	lower Big Cr.	1035	0.831	0.832	0.855	0.784	0.785	0.816
	Chamberlain Cr.	1400	0.829	0.830	0.864	0.797	0.799	0.838
SFSALM	upper Salmon R. mainstem	600	0.867	0.872	0.915	0.843	0.850	0.888
	Johnson Cr.	1207	0.881	0.881	0.891	0.835	0.837	0.857
	EF SF Salmon R.	593	0.933	0.933	0.943	0.901	0.902	0.919
	Lake Cr.	674	1.000	1.000	1.000	0.988	0.988	0.984
	Lick Cr.	832	0.851	0.849	0.861	0.770	0.771	0.791
	Secesh R.	1194	0.903	0.903	0.907	0.871	0.871	0.875
LOSALM	Boulder Cr./Rapid R.	2800	0.522	0.524	0.542	0.471	0.474	0.493
	Slate Cr.	1400	0.191	0.248	0.390	0.186	0.230	0.347
UPCLWR	upper Lochsa R.	1200	0.933	0.936	0.948	0.925	0.926	0.938
	Lake Cr. (Lochsa R.)	434	0.979	0.979	0.986	0.960	0.961	0.974
	Fish Cr.	966	0.914	0.916	0.916	0.900	0.901	0.908
	upper Selway R.	1268	0.973	0.973	0.976	0.964	0.965	0.968
	Whitecap Cr.	1032	0.984	0.982	0.981	0.967	0.967	0.968
	Bear Cr. (Selway R.)	700	0.970	0.971	0.981	0.960	0.962	0.971
	middle Selway R.	1300	0.942	0.942	0.947	0.912	0.914	0.927
	Three Links Cr.	773	0.996	0.996	0.992	0.990	0.990	0.988
	Gedney Cr.	464	0.905	0.901	0.894	0.858	0.857	0.865
	Ohara Cr.	763	0.696	0.695	0.710	0.640	0.642	0.660
SFCLWR	Crooked R.	2400	0.912	0.911	0.910	0.883	0.883	0.884
	Newsome Cr.	1800	0.902	0.900	0.927	0.878	0.879	0.901
	Tenmile Cr. (Clearwater R.)	900	0.674	0.679	0.733	0.666	0.669	0.724
	Clear Cr.	800	0.709	0.711	0.746	0.683	0.689	0.729
	Lolo Cr.	1700	0.846	0.847	0.872	0.817	0.817	0.845
LOCLWR	WF Potlatch R.	1600	0.710	0.714	0.699	0.635	0.642	0.647
	EF Potlatch R.	3000	0.727	0.730	0.743	0.674	0.678	0.692
	Little Bear Cr.	2880	0.657	0.659	0.664	0.592	0.594	0.591
	Big Bear Cr.	1820	0.586	0.584	0.595	0.535	0.536	0.539
	Lapwai R.	3000	0.453	0.460	0.528	0.407	0.412	0.472
IMNAHA	Gumboot/Mahogany Cr.	2000	0.538	0.553	0.631	0.482	0.496	0.566
	Little Sheep Cr.	3334	0.781	0.781	0.757	0.689	0.689	0.660
	Big Sheep Cr.	3366	0.589	0.594	0.598	0.518	0.521	0.516
	Lightning Cr.	1500	0.310	0.347	0.458	0.295	0.329	0.427
GRROND	upper Grande Ronde R.	2617	0.587	0.586	0.538	0.505	0.501	0.452
	Catherine Cr.	3683	0.596	0.589	0.549	0.518	0.514	0.471
	Little Minam R.	2000	0.690	0.688	0.710	0.649	0.647	0.671
	Wallowa R.	2968	0.564	0.555	0.514	0.492	0.488	0.454
	Lostine R.	1832	0.774	0.771	0.772	0.686	0.683	0.678
	Wenaha R.	7700	0.708	0.701	0.635	0.616	0.611	0.547
	Menatchee Cr.	3000	0.654	0.646	0.663	0.583	0.580	0.586
	Elk Cr. (Joseph Cr.)	1824	0.884	0.879	0.874	0.812	0.810	0.805
	Joseph Cr.	2176	0.558	0.549	0.509	0.494	0.489	0.456

Table A3. Continued

Reporting Unit	Collection	n	Maximum Individual Assignment			Proportional Individual Assignment		
			Bayesian	MLE	Geneclass	Bayesian	MLE	Geneclass
LSNAKE	Asotin Cr.	6900	0.422	0.424	0.370	0.385	0.385	0.336
	Alpowa Cr.	3500	0.463	0.470	0.436	0.415	0.420	0.387
	Tucannon R.	3800	0.398	0.402	0.350	0.372	0.374	0.336

Table A4. The proportion of Chinook Salmon baseline samples assigned to their reporting unit using a given method. Cells are color-coded with warmer colors (red) indicating higher self-assignment rates and cooler colors (blue) indicating lower rates.

Reporting Unit	Collection	n	Maximum Individual Assignment			Proportional Individual Assignment		
			Bayesian	MLE	Geneclass	Bayesian	MLE	Geneclass
UPSALM	Sawtooth Weir	5100	0.812	0.811	0.817	0.740	0.741	0.751
	Valley Cr.	2800	0.875	0.876	0.888	0.829	0.830	0.840
	WF Yankee Fork	2100	0.901	0.901	0.930	0.851	0.852	0.884
	upper Salmon R. mainstem	2300	0.919	0.919	0.917	0.850	0.852	0.851
	Herd Cr.	2701	0.924	0.924	0.912	0.874	0.874	0.856
	EF Salmon R.	5199	0.887	0.887	0.880	0.831	0.831	0.817
	Pahsimeroi R.	2600	0.876	0.876	0.903	0.826	0.827	0.859
	upper Lemhi R.	2700	0.739	0.739	0.784	0.714	0.714	0.751
	NF Salmon R.	1600	0.569	0.579	0.738	0.522	0.534	0.667
MFSALM	Marsh Cr.	3233	0.769	0.770	0.780	0.729	0.729	0.727
	Capehorn Cr.	3167	0.950	0.949	0.941	0.903	0.903	0.890
	Elk Cr.	3843	0.884	0.884	0.884	0.851	0.851	0.848
	Bear Valley Cr.	2257	0.899	0.899	0.896	0.837	0.837	0.832
	Sulphur Cr.	3800	0.967	0.966	0.971	0.936	0.937	0.938
	Loon Cr.	2700	0.828	0.829	0.849	0.771	0.773	0.807
	Camas Cr.	3000	0.852	0.852	0.878	0.800	0.801	0.832
	upper Big Cr.	1600	0.822	0.825	0.859	0.777	0.780	0.815
	lower Big Cr.	3900	0.727	0.726	0.742	0.675	0.675	0.691
CHMBLN	Chamberlain Cr. (pre-2008)	1250	0.726	0.724	0.730	0.694	0.694	0.704
	Chamberlain Cr. (post-2008)	2650	0.887	0.886	0.897	0.875	0.876	0.883
SFSALM	Summit/Lake Cr.	3980	0.878	0.878	0.876	0.845	0.846	0.832
	Secesh R.	4320	0.836	0.836	0.815	0.790	0.791	0.768
	Johnson Cr.	4500	0.746	0.746	0.765	0.707	0.708	0.721
	SF Salmon R. mainstem	4600	0.502	0.503	0.502	0.461	0.463	0.459
HELLSC	Rapid R.	3700	0.934	0.935	0.916	0.901	0.902	0.880
	Crooked Fork Cr.	1146	0.853	0.855	0.810	0.820	0.821	0.784
	Powell Weir	1254	0.754	0.758	0.728	0.707	0.712	0.671
	Red R.	2917	0.891	0.892	0.864	0.863	0.864	0.825
	Crooked R. Weir	2689	0.904	0.905	0.896	0.867	0.868	0.844
	Newsome Cr.	3294	0.934	0.934	0.908	0.898	0.898	0.872
	Lolo Cr.	3600	0.863	0.864	0.838	0.829	0.830	0.798
	Imnaha R. (2008, 2010)	3900	0.821	0.822	0.789	0.760	0.762	0.729
	upper Grande Ronde R.	1710	0.865	0.863	0.782	0.810	0.809	0.738
	Catherine Cr.	5590	0.892	0.893	0.868	0.861	0.861	0.825
	Minam R.	5300	0.957	0.957	0.926	0.910	0.910	0.865
	Wallowa R.	1488	0.909	0.909	0.890	0.895	0.894	0.861
	Lostine R.	7012	0.938	0.937	0.912	0.902	0.903	0.866
	Wenaha R.	7000	0.932	0.932	0.896	0.903	0.902	0.857
TUCANO	Tucannon R.	900	0.762	0.764	0.820	0.746	0.749	0.813
FALL	Fall	3600	1.000	1.000	1.000	1.000	1.000	1.000

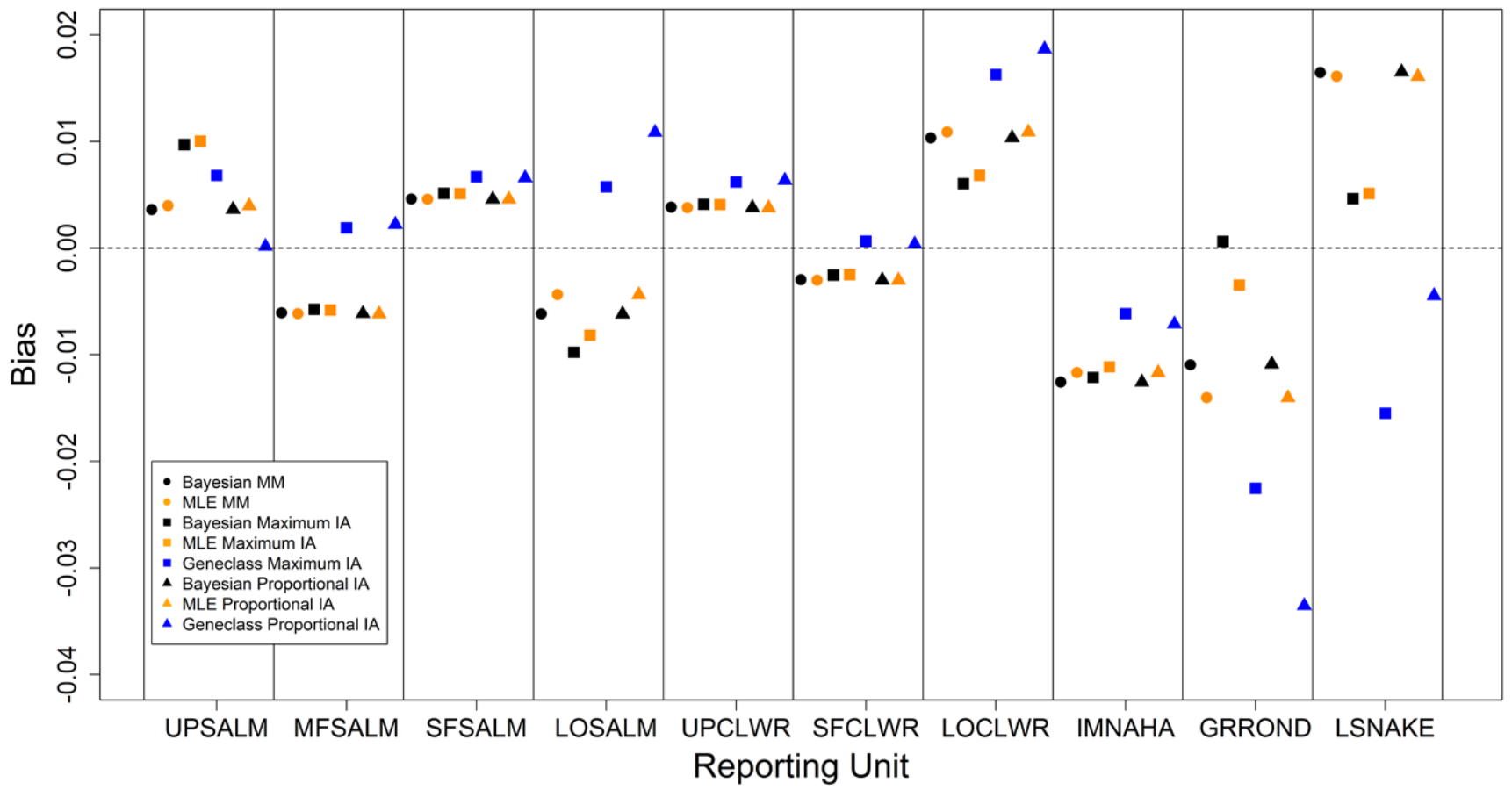

Figure A1. The observed bias in the estimated mixing proportions of steelhead reporting units from self-assignment tests using a given assignment method.

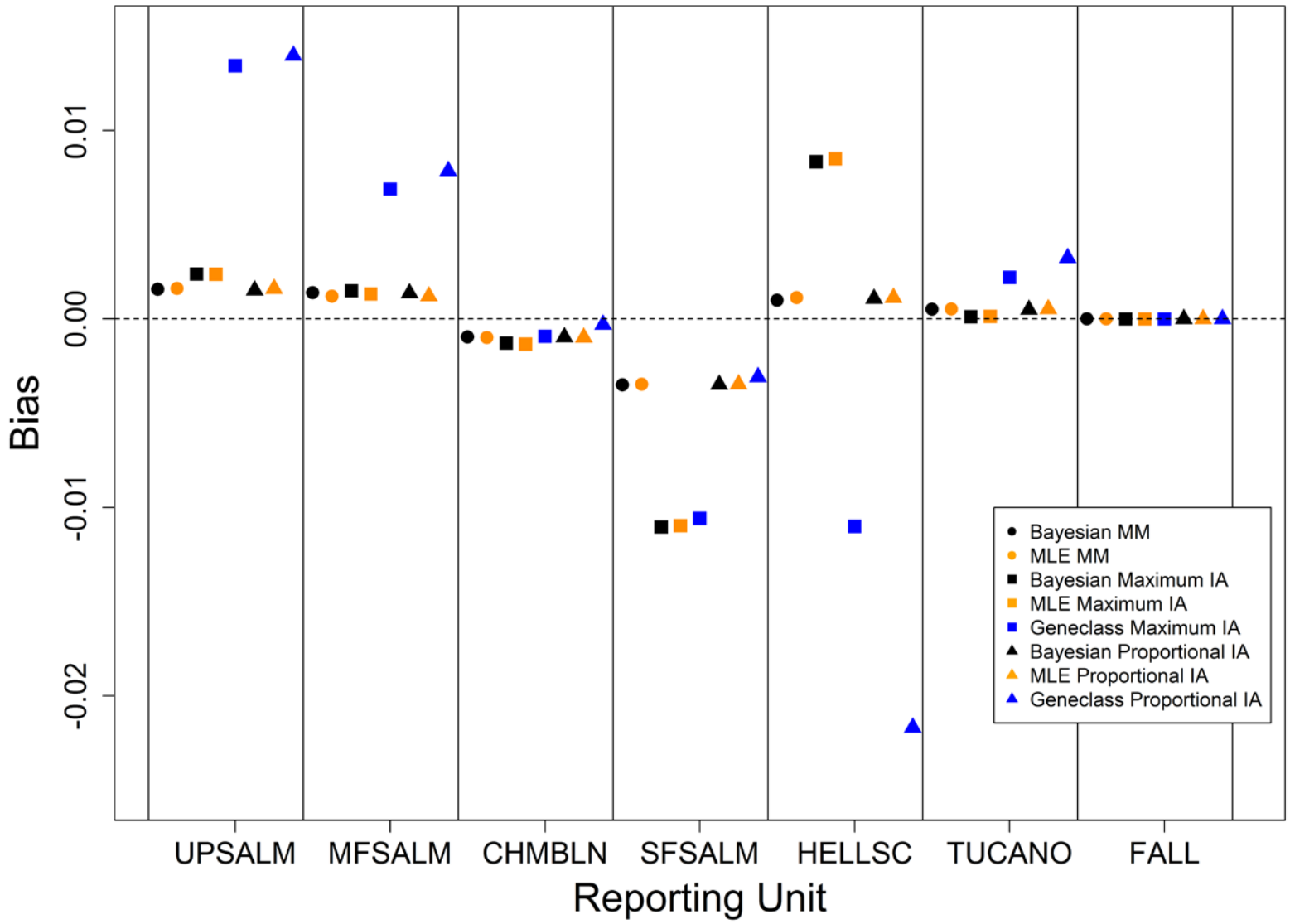

Figure A2. The observed bias in the estimated mixing proportions of Chinook Salmon reporting units from self-assignment tests using a given assignment method.

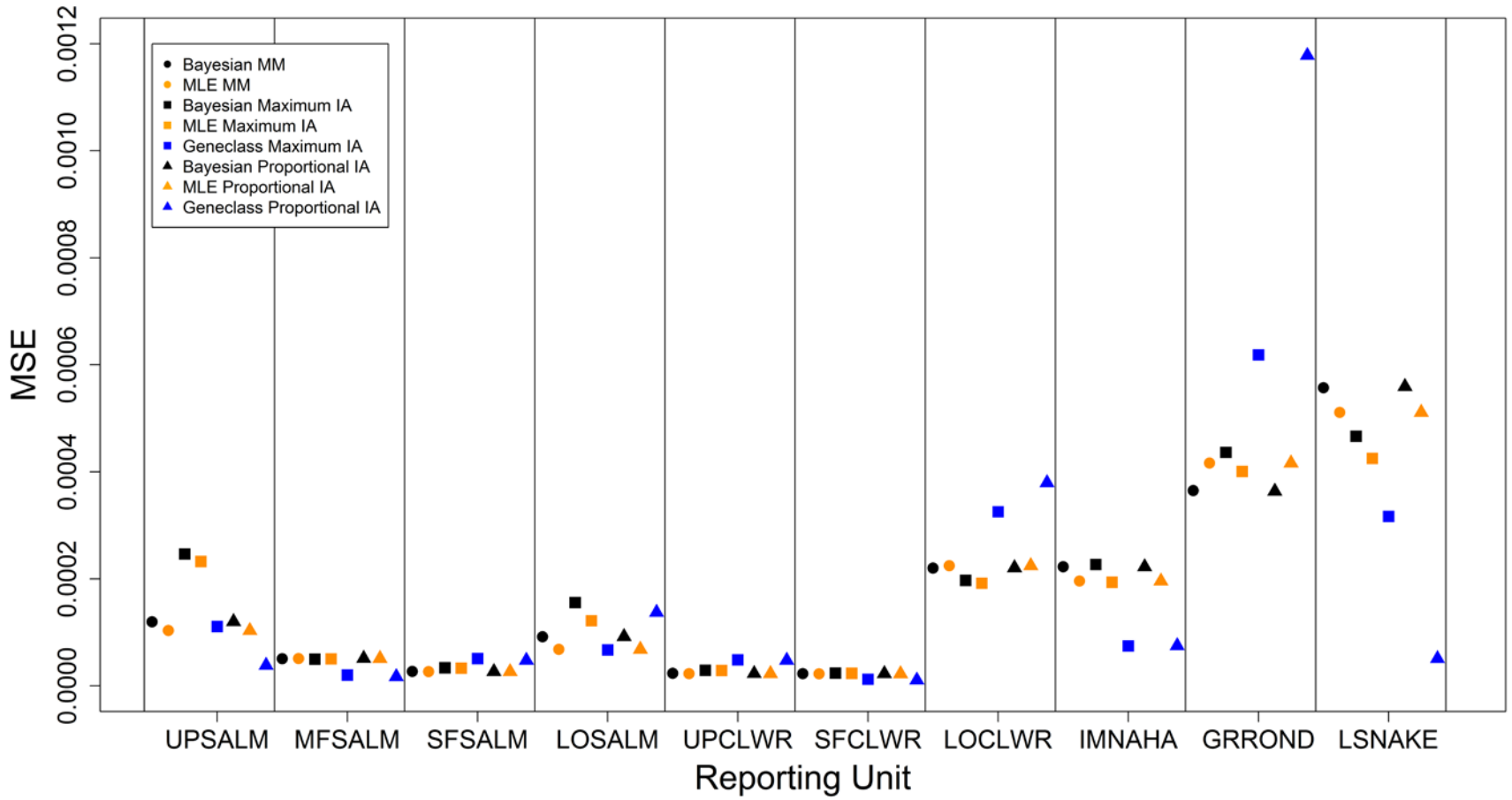

Figure A3. The mean squared error (MSE) of the estimated mixing proportions of steelhead reporting units from self-assignment tests using a given assignment method.

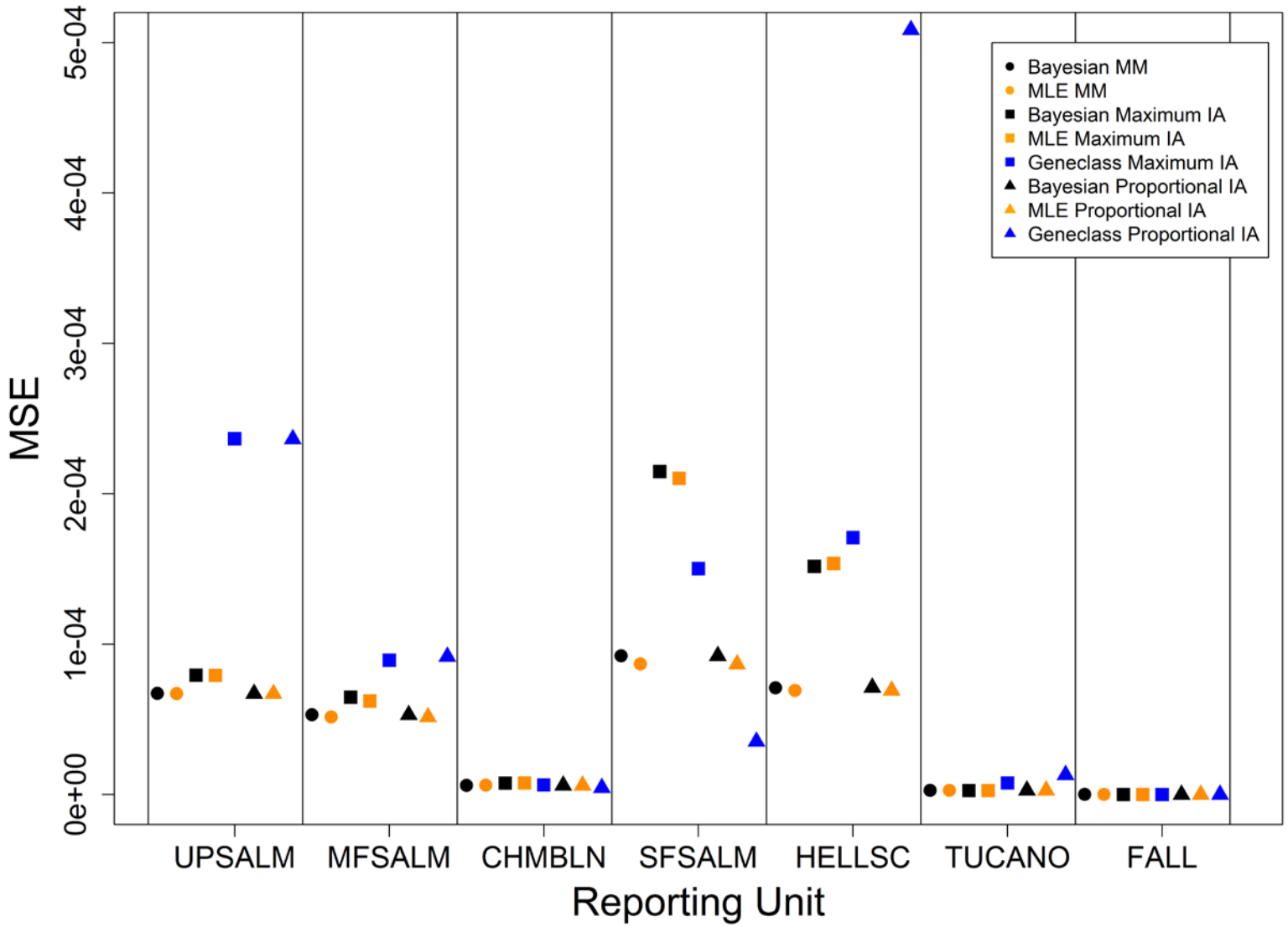

Figure A4. The mean squared error (MSE) of the estimated mixing proportions of Chinook Salmon reporting units from self-assignment tests using a given assignment method.

APPENDIX B

INTRODUCTION

The Chamberlain Creek drainage comprises the largest spawning area of the lowermost population of Chinook Salmon in the Middle Fork Salmon River MPG. Chinook Salmon in the Chamberlain Creek drainage are genetically distinct from other populations in the Snake River basin (Ackerman et al. 2012). Viability criteria for the Middle Fork Salmon River MPG specify that Chamberlain Creek needs to be viable because of its geographic position and genetic differentiation (ICTRT 2007). The Chamberlain Creek population has been the most robust population of Chinook Salmon in the Middle Fork Salmon MPG since listing (see Table 21 in Ford et al. 2011), with estimates of population abundance derived from the spring/summer Chinook Salmon GSI stock composition at LGD exceeding the Interior Columbia Basin Technical Recovery Team (ICTRT) minimum threshold abundance of 500 adults during 2009-2014 (Camacho et al. 2017).

Despite these robust abundance estimates relative to other populations in the Middle Fork Salmon River MPG, high levels of uncertainty exist in the current estimates of productivity (ICTRT 2010). One factor identified as contributing to this uncertainty was that redd counts occurred in index reaches covering a small portion, or 9%, of the total weighted spawning habitat in the watershed (ICTRT 2010). To address this uncertainty, IDFG initiated additional spawning ground survey data collection efforts during 2013-2015 (Stiefel et al. 2016). The primary goal of this effort was to validate the current redd count expansion method that uses single-pass index transects established in 1957. Results showed that the standard index survey expansions were very similar to the extended survey expansions (Brown 2016). However, in 2014 and 2015 GSI estimates of spawner abundance at LGD were approximately half the size of the redd count expansions (Brown 2016). This discrepancy in the abundance of Chinook Salmon in Chamberlain Creek brought into question the accuracy of individual assignments to the Chamberlain Creek reporting unit in the Chinook Salmon Snake River GSI SNP baseline v3.1.

In this appendix, we test the accuracy of individual genetic stock assignments of juvenile and adult Chinook Salmon sampled directly from the Chamberlain Creek watershed. We also estimate the genetic divergence between collections used in the Chinook Salmon Snake River SNP GSI baseline v3.1 and recently sampled juvenile and adult collections. These tests are undertaken to investigate whether our baseline is representative of individual Chinook Salmon rearing in, or returning to, the Chamberlain Creek watershed.

METHODS

We genotyped 319 individual Chinook Salmon collected between 2013 and 2016 in the Chamberlain Creek drainage at 298 SNP loci (Table B1, Powell et al. 2017). Genotyping was performed using the GT-seq platform (Campbell et al. 2015) on an Illumina NextSeq 500 DNA sequencer (Illumina, San Diego, California). Prior to analysis, individual genotypes were trimmed to the 173 SNP loci used in the construction of the Chinook Salmon Snake River GSI baseline version 3.1 (Vu et al. 2015). Unless otherwise stated, analyses were completed in R version 3.4.1 (R Core Team 2017).

When testing population differentiation with allele frequency based statistics, sampling sites can falsely exhibit significant genetic differentiation when large numbers of related juveniles have been inadvertently sampled (Allendorf and Phelps 1981). This increased Type I error rate has been observed in salmonid studies and has been termed the Allendorf-Phelps

effect (*sensu* Waples 1998). To avoid a positive bias in our estimates of population differential we initially assigned individuals in the 2016 juvenile collections from the Chamberlain Creek drainage to full sibling families using the program Colony version 2.0.6.2 (Jones and Wang 2010). These assignments assumed polygamous mating systems for both males and females. Full sibling families were subsequently thinned to a single individual for calculation of observed and expected heterozygosity, the percent of loci that were polymorphic in a sample, tests for Hardy-Weinberg equilibrium, and estimation of F_{ST} with Weir and Cockerham's θ (Weir and Cockerham 1984). We randomly sampled a single individual from each family group one thousand times, and used the mean of the distribution of estimated parameter values for analyses.

We tested for Hardy-Weinberg equilibrium and calculated single locus F_{IS} values using the R package HardyWeinberg version 1.5.6 (Graffelman and Morales-Camarena 2008; Graffelman 2015). Statistical significance was calculated following a within population false discovery rate correction of Benjamini and Yekutieli (2001) as described in Narum (2006) for 173 simultaneous tests. We used the R package hierfstat version 0.04-22 (Goudet and Jombart 2015) to estimate F_{ST} with Weir and Cockerham's θ (Weir and Cockerham 1984) for our four collections of interest and all pairs of populations in Chamberlain Creek, Middle Fork Salmon River, and South Fork Salmon River reporting units in the Chinook Salmon Snake River GSI baseline v3.1.

We tested the assignment of individuals to the Chinook Salmon Snake River SNP GSI baseline v3.1 using the program *gsi_sim* (Anderson et al. 2008; Anderson 2010). To perform this test we assigned all collections to the baseline in isolation. Next, assignments were made by adding the collection of interest to the SY2016 sample of natural-origin Chinook Salmon sampled at Lower Granite Dam. The assignments including the larger mixture of adults passing Lower Granite Dam were completed to investigate whether success was affected by including individuals in a larger mixture of natural-origin fish. All assignments were made using the full expectation-maximization algorithm maximum likelihood estimates of Smouse et al. (1990).

We used two other methods of genetic clustering to assess the relationship between individuals from the four collections of interest and baseline populations from Chamberlain Creek, the Middle Fork Salmon River, and the South Fork Salmon River. First, Individuals were clustered using discriminate analysis of principle components (Jombart et al. 2008) using the R package adegenet version 1.3-1 (Jombart 2008; Jombart and Ahmed 2011). We used a-score optimization to determine that the first 64 principle components should be retained in the final analysis. Next, we assigned portions of each individual's genome to Hardy-Weinberg linkage equilibrium populations using the program Admixture version 1.3 (Alexander et al. 2009).

A neighbor-joining (NJ) tree was constructed with the four Chamberlain Creek drainage collections and the Chinook Salmon Snake River GSI baseline version 3.1. This NJ tree was based on pairwise Cavalli-Sforza Edwards chord distances (Cavalla-Sforza and Edwards 1967) calculated using GENDIST (PHYLIP v3,5; Felsenstein 1993). Pairwise genetic distances were used to construct NJ trees in NEIGHBOR (PHYLIP v3.5). NJ trees were visualized using FigTree v1.4.0 (Rambaut 2012, available at: http://tree.bio.ed.ac.uk/software/figtree/), with bootstrap values added in Microsoft PowerPoint (Microsoft Corporation, Redmond, Washington).

RESULTS

Juvenile samples collected in Chamberlain Creek in 2016 contained 18 full sibling families ranging in size from 2 to 9 members. Fifty-one of 98 genotyped individuals were assigned to a full sibling family group, and the estimated effective number of breeders that gave rise to this sample was 46 individuals (95% confidence interval from 32 to 71 individuals). Juveniles collected in West Fork Chamberlain Creek in 2016 were assigned to 11 full sibling family groups ranging in size from 2 to 23 members. Sixty of 87 genotyped individuals in this collection were assigned to a family group, and the estimated effective number of breeders that gave rise to this sample was 8 individuals (95% confidence interval from 4 to 24 individuals).

Nine loci were out of Hardy-Weinberg equilibrium following a within population false discovery rate correction in at least one of the four collections from Chamberlain Creek. The locus *Ots_MHC1* exhibited an excess of homozygotes in three of four sample collections. No other locus exhibited a departure from Hardy-Weinberg equilibrium in more than one collection. We observed that an excess of homozygotes was responsible for all six loci out of Hardy-Weinberg equilibrium in West Fork Chamberlain Creek. This pattern of excess homozygosity is not unexpected given the small number of effective breeders that was estimated to give rise to this sample collection.

We observed similar levels of observed heterozygosity, expected heterozygosity, and the percent of loci that are polymorphic among the study collections from Chamberlain Creek and the Chamberlain Creek populations included in the Chinook Salmon Snake River GSI baseline version 3.1 (Table B2). West Fork Chamberlain Creek exhibited reduced genetic variation relative to other collections in the Chamberlain Creek Basin (Table B2), which is not unexpected given the small estimated number of effective breeders.

We observed less genetic differentiation among populations in the Chamberlain Creek drainage than among populations in the Middle Fork Salmon River or South Fork Salmon River drainages (Table B3). Except for the sample from West Fork Chamberlain Creek, little genetic divergence was observed among collections in the Chamberlain Creek drainage (Table B4).

Individuals from recent collections in the Chamberlain Creek drainage assigned with high fidelity to the Chamberlain Creek reporting unit both when tested as a single collection (Table B5) and as part of a larger mixture of fish migrating over Lower Granite Dam (Table B6).

A discriminant analysis of principle components indicated the presence of four genetic clusters in Chinook Salmon from Chamberlain Creek, the Middle Fork Salmon River, and the South Fork Salmon River drainages. These four genetic clusters grouped samples from the Chamberlain Creek drainage, the South Fork Salmon River drainage, the Middle Fork Salmon River drainage above Loon Creek, and the Middle Fork Salmon River drainage below Loon Creek (Figure B1).

An Admixture analysis indicated the presence of two ancestral populations of Chinook Salmon from the Chamberlain Creek, Middle Fork Salmon River, and the South Fork Salmon River drainages. These two clusters separated individuals from the Chamberlain Creek drainage from individuals sampled in the South Fork and Middle Fork Salmon River drainages (Figure B2).

A phylogenetic analysis indicated that all individuals sampled from the Chamberlain Creek drainage represent a monophyletic clade (Figure B3). However, this clade forms a polytomy with populations from the South Fork Salmon River drainage (Figure B3).

We also did not observe bias in the estimated mixture proportions from Chamberlain Creek (Figure 2) at returns commonly observed passing Lower Granite Dam between SY2009 and SY2015 (2%-4.6%; Ackerman et al. 2012, 2014, 2016; Vu et al. 2015, Powell et al. 2017).

DISCUSSION

Analyses indicated that the Chamberlain Creek reporting unit in the Chinook Salmon Snake River GSI baseline v3.1 is faithfully assigning current collections of individuals. The only discrepancy that was apparent from this analysis was that not all juveniles or carcasses collected in the Chamberlain Creek drainage assigned to that reporting unit when these samples were included with the SY2016 returns to Lower Granite Dam. However, leave-one-out cross validation tests implemented in the R package rubias (E. Anderson unpublished, available at: https://github.com/eriqande/rubias) indicate that overall the estimated proportion of fish originating from the Chamberlain Creek drainage is unbiased in these samples (Figure 2).

While there appears to be divergence within the Chamberlain Creek drainage (Table B4), this is based on a single sample of juveniles that were likely produced from a small number of parents. Therefore, more sampling from West Fork Chamberlain Creek would be necessary before we can conclusively decide to split Chamberlain Creek and West Fork Chamberlain Creek into two populations in a future version of the Chinook Salmon Snake River GSI baseline.

Table B1. Summary of carcass and juvenile Chinook Salmon sampled from Chamberlain Creek and West Fork Chamberlain Creek. Summary includes the number of samples inventoried in the lab, the number inventoried that were queued for genotyping. Of queued samples, we report the number that genotyped successfully and the number that failed genotyping.

Sample Group	Total Samples Inventoried	Samples Queued for Genotyping	Failed Genotyping (NG)	Successfully Genotyped
2013 Chamberlain Cr. carcasses	54	46	11 (23.9%)	35 (76.1%)
2014 Chamberlain Cr. carcasses	74	73	41 (56.2%)	32 (43.8%)
2016 Chamberlain Cr. juveniles	100	100	2 (2%)	98 (98%)
2016 WF Chamberlain Cr. juveniles	100	100	13 (13%)	87 (87%)
TOTAL:	328	319	67 (21%)	252 (79%)

Table B2. Observed heterozygosity, expected heterozygosity and the percent of polymorphic loci in carcass and juvenile Chinook Salmon sampled from Chamberlain Creek and West Fork Chamberlain Creek relative to Chinook Salmon Snake River GSI baseline version 3.1 collections.

Sample Group	H_O	H_E	Polymorphic Loci
Chamberlain Cr. pre-2008*	0.210	0.208	76.9%
Chamberlain Cr. post-2008*	0.200	0.203	84.4%
2013 Chamberlain Cr. carcasses	0.209	0.198	74.6%
2014 Chamberlain Cr. carcasses	0.200	0.199	74.0%
2016 Chamberlain Cr. juveniles	0.205	0.204	78.8%
2016 WF Chamberlain Cr. juveniles	0.198	0.195	72.2%

* Populations in the Chinook Salmon Snake River GSI baseline version 3.1

Table B3. Average pairwise F_{ST} among populations within and between reporting units in the Chinook Salmon Snake River GSI baseline version 3.1. All carcass and juvenile samples collected in Chamberlain Creek and West Fork Chamberlain Creek were included in the Chamberlain Creek reporting unit.

	CHMBLN	SFSALM	MFSALM
CHMBLN	0.013		
SFSALM	0.043	0.015	
MFSALM	0.053	0.028	0.021

Table B4. Pairwise values of Weir and Cockerham's θ among populations within the Chamberlain Creek reporting unit in the Chinook Salmon Snake River GSI baseline version 3.1 and carcass and juvenile samples collected in Chamberlain Creek and West Fork Chamberlain Creek.

	Baseline pre-08	Baseline post-08	2013 Carcass	2014 Carcass	2016 Juveniles	2016 WF Juveniles
Baseline pre-08	-					
Baseline post-08	0.009	-				
2013 Carcass	0.004	0.004	-			
2014 Carcass	0.014	0.010	0.008	-		
2016 Juveniles	0.010	0.009	0.005	0.010	-	
2016 WF Juveniles	0.023	0.023	0.022	0.024	0.024	-

Table B5. Proportion of carcass and juvenile samples collected in Chamberlain Creek and West Fork Chamberlain Creek that assigned to given reporting units in the Chinook Salmon Snake River GSI baseline version 3.1.

Sample Group	UPSALM	MFSALM	CHMBLN	SFSALM	HELLSC	TUCANO	FALL
2013 Chamberlain Cr. carcasses	0.029	0.000	0.914	0.057	0.000	0.000	0.000
2014 Chamberlain Cr. carcasses	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2016 Chamberlain Cr. juveniles	0.000	0.000	0.969	0.000	0.031	0.000	0.000
2016 WF Chamberlain Cr. juveniles	0.000	0.000	1.000	0.000	0.000	0.000	0.000
TOTAL:	0.004	0.000	0.976	0.008	0.012	0.000	0.000

Table B6. Proportion of carcass and juvenile samples collected in Chamberlain Creek and West Fork Chamberlain Creek that assigned to given reporting units in the Chinook Salmon Snake River GSI baseline version 3.1 when included with all SY2016 returning natural-origin Chinook Salmon at Lower Granite Dam.

Sample Group	UPSALM	MFSALM	CHMBLN	SFSALM	HELLSC	TUCANO	FALL
2013 Chamberlain Cr. carcasses	0.086	0.000	0.857	0.000	0.057	0.000	0.000
2014 Chamberlain Cr. carcasses	0.000	0.000	0.938	0.031	0.031	0.000	0.000
2016 Chamberlain Cr. juveniles	0.010	0.010	0.878	0.020	0.082	0.000	0.000
2016 WF Chamberlain Cr. juveniles	0.000	0.000	0.966	0.034	0.000	0.000	0.000
TOTAL:	0.016	0.004	0.913	0.024	0.044	0.000	0.000

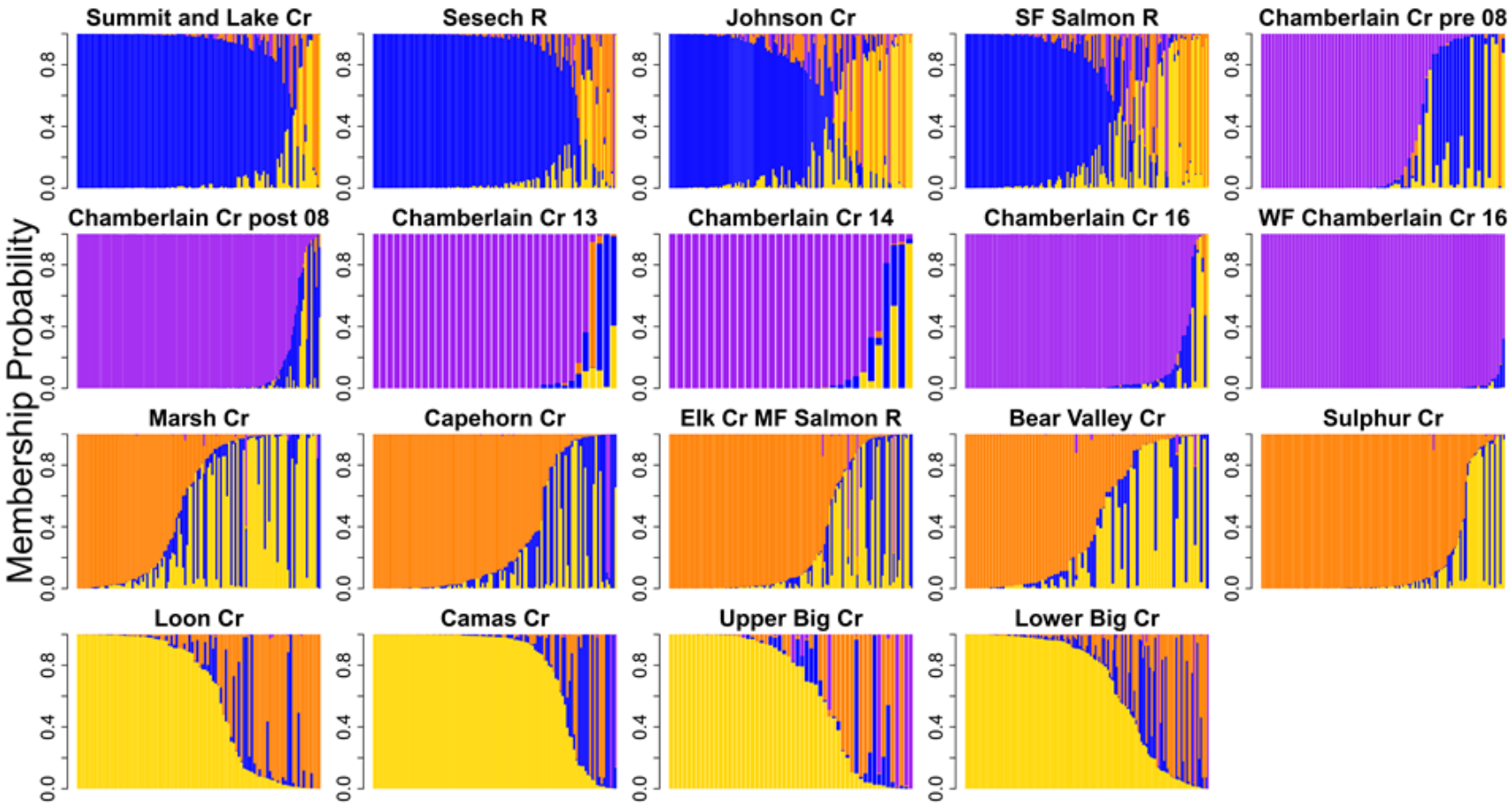

Figure B1. Bar charts of individual membership probability to the four genetic clusters identified in a discriminant analysis of principal components. Each individual is plotted as a single vertical line, and each genetic cluster is color-coded. Individuals are grouped by either 1) their sample date and location in the Chamberlain Creek drainage, or 2) their inclusion in a population in the Chinook Salmon Snake River GSI baseline v3.1. Individuals have been sorted based on cluster membership probability. Clusters generally separated individuals sampled from the South Fork Salmon River drainage (blue), the Chamberlain Creek drainage (purple), the upper Middle Fork Salmon River drainage (orange), and the lower Middle Fork Salmon River drainage (yellow).

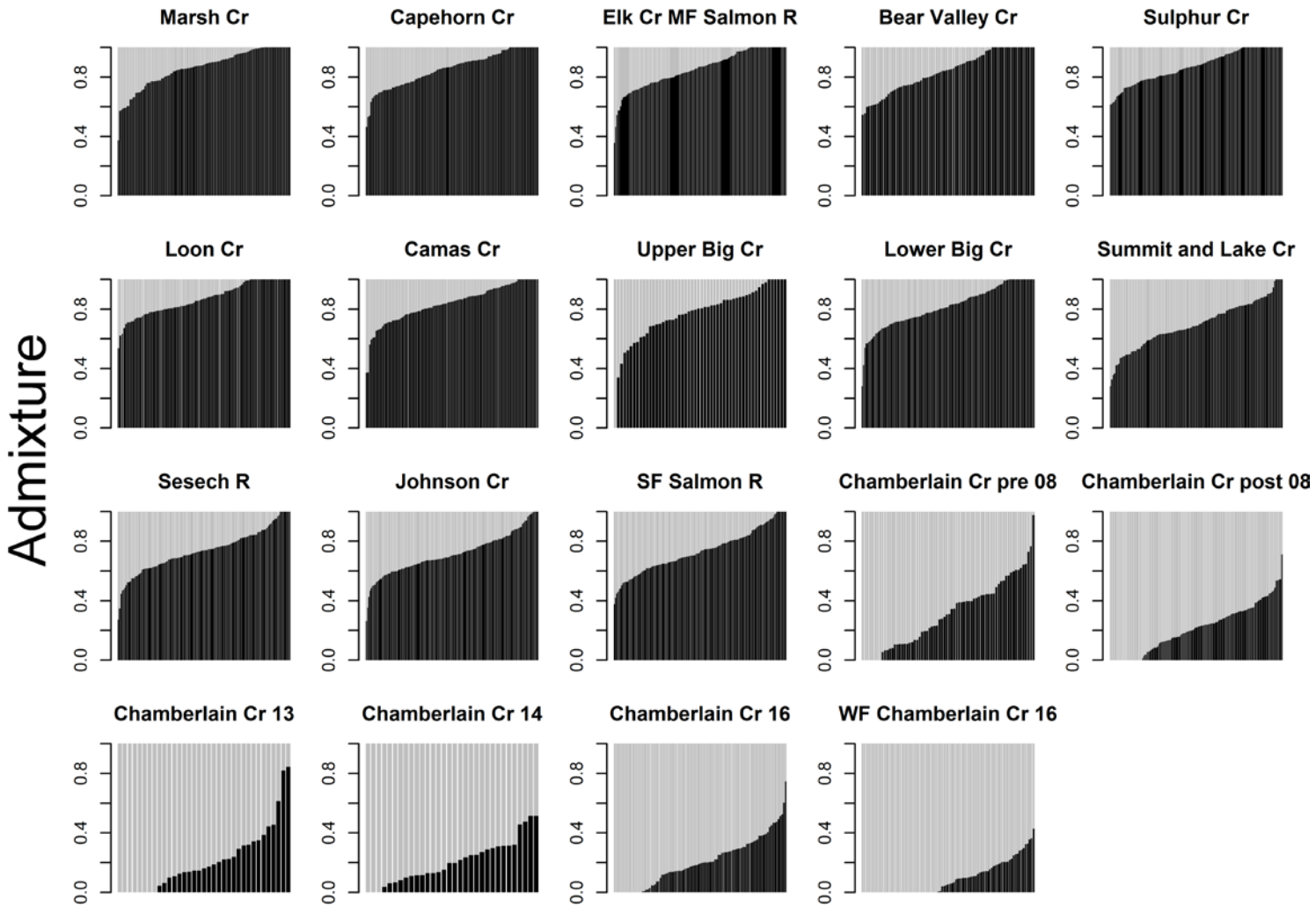

Figure B2. Bar charts of the proportion of an individual's genome that came from the two Hardy-Weinberg linkage equilibrium genetic clusters identified in the program Admixture. Each individual is plotted as a single vertical line, and each genetic cluster is color-coded. Individuals are grouped by their sample date and location in the Chamberlain Creek drainage, or their inclusion in a population in the Chinook Salmon Snake River GSI baseline v3.1. Individuals have been sorted based on proportion admixture, with clusters generally separating individuals sampled in the South and Middle Fork Salmon River drainages (black) from individuals sampled in the Chamberlain Creek drainage (gray).

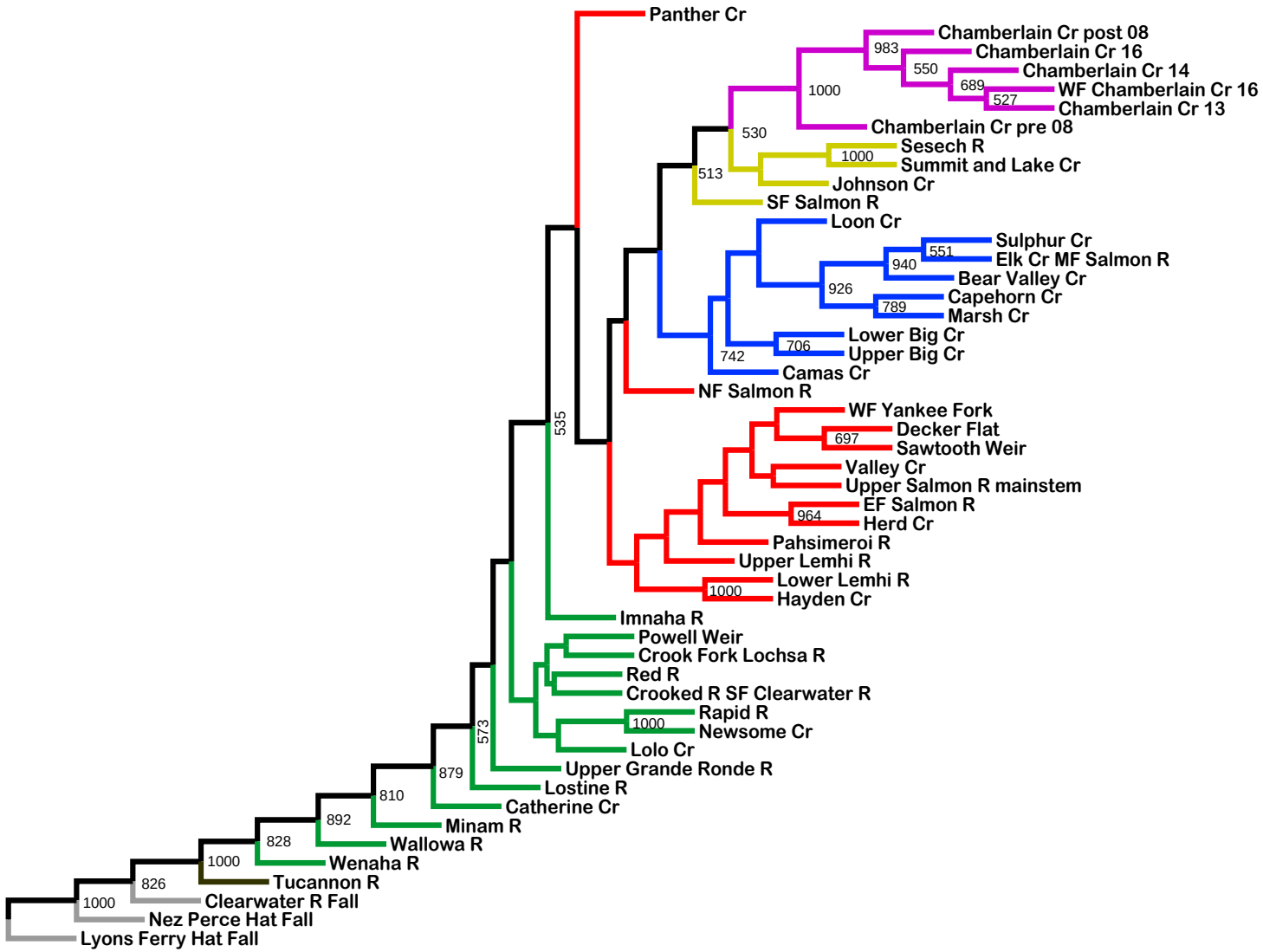

Figure B3. Neighbor-joining (NJ) Tree based on Cavalli-Sforza Edwards chord distance for Chinook Salmon Snake River GSI baseline v3.1 collections and fish collected in the Chamberlain Creek drainage. GSI reporting units are color-coded.

Prepared by:

John H. Powell
Fishery Biologist 2

Ninh Vu
Fishery Biologist 2

Matthew R. Campbell
Fisheries Genetics Program Coordinator

Approved by:
IDAHO DEPARTMENT OF FISH AND GAME

Daniel J. Schill
Fisheries Research Manager

James P. Fredericks, Chief
Bureau of Fisheries