# Assessing Effects of Supplementation on Fitness of Sockeye Salmon in Auke Creek, AK

**Final Report for Phase 1**
**1 June 2014 - 30 June 2017**

prepared by

Megan V. McPhee, Associate Professor
College of Fisheries and Ocean Sciences
University of Alaska Fairbanks
17101 Point Lena Loop Road
Juneau, Alaska 99801 USA
mvmcphee@alaska.edu
+1 (907) 796-5464

31 July 2017

**Table of Contents**

## List of Tables



## List of Figures

## List of Appendices

**Executive Summary**
Three years of experimental hatchery supplementation were conducted (2011-2013) in order to quantify fitness differences between wild and hatchery origin sockeye salmon within the US-Canada transboundary region subject to the Pacific Salmon Treaty. All adult salmon ascending the weir at Auke Creek, Juneau, Alaska, have been sampled for genetic tissue since 2008. A panel of SNP and single-tandem repeat (STR) loci are used to assign adult offspring back to parents, allowing for a full enumeration of fitness, defined as the number of returning adult offspring per parent. By the end of Phase 1 of this long-term project, we 1) demonstrated the effectiveness of parentage assignment with genomic resources that existed at the project's inception; 2) reduced the panel of markers from 96 to 47 SNPs and from 12 to 9 STRs; 3) genotyped all adults returning 2012 and 2013, backfilled missing genotypes from return year 2010, and genotyped a subset of adults returning in 2014 and 2015 with high value for our project; and 5) assigned 2014 and 2015 adults back to candidate parents. We were able to assign ~ 95% to one or both parents in our dataset. A small proportion of returning adults (0.7% in 2014 and 0.4% in 2015) assigned with high probability to no parents in our dataset, and analysis of the principal component scores suggested that these individuals likely strayed into Auke Creek from other source populations.

**Introduction**
The overarching goal of this project is *to quantify differences in fitness between wild and hatchery-origin sockeye salmon Oncorhynchus nerka* in Auke Creek, Alaska, using parentage-based tagging over three generations of experimental hatchery supplementation. Secondary goals of this research are to test for second-generation differences in fitness between wild and hatchery-origin individuals that spawn naturally, and to quantify changes in genetic diversity and population structure in the wild sockeye salmon population as a result of three generations of hatchery supplementation. Results of this study will provide information critical for assessing the relative costs and benefits of hatchery supplementation in managing sockeye salmon populations subject to the Pacific Salmon Treaty (PST).

Evaluating the relative fitness (survival and mating success) of hatchery and wild salmonids is important for several reasons. First, if hatchery individuals have low fitness, hatchery supplementation has a low return on investment and represents fishery management resources that might be better directed elsewhere. Furthermore, in cases where hatchery-origin fish are allowed to interbreed with wild members of the supplemented population, low relative fitness of hatchery-wild hybrid fish can reduce the mean fitness of the supplemented population (Araki et al. 2009). High relative fitness of hatchery fish can increase the variance in reproductive success, thereby reducing the genetically effective size of the supplemented population (Ryman and Laikre 1991, Wang and Ryman 2001). Consequently, adequate monitoring of the effectiveness of supplementation programs requires quantification of the relative fitness of hatchery and wild individuals (Fraser 2008).

Quantifying the relative fitness of hatchery and wild salmonids has become feasible with the advent of parentage-based tagging (PBT; Anderson and Garza 2006). This method uses genotypic data to assign offspring to parents, and therefore can be used to track the relative fitness of hatchery and wild individuals by quantifying the number of offspring in families of each parental pair. PBT has rapidly become the method of choice for studying fitness differences between hatchery and wild salmonids, but the effect of hatchery breeding on fitness is highly variable among species and populations (reviewed in Araki and Schmid 2010). For example, Araki et al. (2007) detected reduced fitness after a single generation of captive breeding in the Hood River population of steelhead trout *O. mykiss*, and subsequent interbreeding between hatchery and wild individuals appeared to reduce the mean fitness of the population (Araki et al. 2009; but see Kitada et al. 2011). Conversely, Hess et al. (2012) found no evidence for reduced fitness of hatchery fish in a supplemented population of Chinook salmon *O. tshawytscha.* In general, variable results may be due to differences in broodstock provenance (local or non-local), hatchery practices, species characteristics, and length of time hatchery populations are reared prior to release (Berejikian et al. 2009), and the genetic history of the wild stock (Willoughby and Christie 2017). Consequently, the best source of information on the effects of supplementation in transboundary sockeye salmon systems will be a study done in this region and on sockeye salmon.

To date, no study other than this one has attempted to quantify fitness differences between hatchery and wild sockeye salmon. Auke Creek provides the ideal setting in which to measure changes in fitness resulting from hatchery supplementation of sockeye salmon for a number of reasons: 1) Auke Creek is located on the road system in Juneau, Alaska, making field sampling cost-effective; 2) the weir at Auke Creek allows researchers to sample all adults returning to spawn in the Auke Creek drainage - this is critical as it allows near-complete genotypic sampling of all parents and offspring; 3) the Auke Creek sockeye salmon population has historically been relatively small (from 2008 through 2015, the average return has been 2,666 adults, ranging from 1,264 to 4,720; J. Joyce and S. Vulstek, NOAA, unpubl.), making complete enumeration of the reproductive success of hatchery and wild individuals computationally feasible; 4) a downstream smolt weir in addition to the upstream adult weir makes further exploration of potential mechanisms for fitness differences possible; and 5) results from this study are likely to be applicable to other sockeye salmon hatchery projects within the geographical region subject to the PST.

Enhancement is an on-going and treaty-mandated component of sockeye salmon production in the areas affected by the PST. In particular, sockeye salmon enhancement plays a large role in transboundary river issues relevant to the PST (e.g., TCTR 1989, 2012). There are enhancement goals for both Stikine and Taku river sockeye salmon populations, and these programs allow for harvest sharing between the U.S. and Canada. In the Stikine River, there are specific penalties if either party fails to meet enhancement obligations. In the Taku River, harvest shares are a

function of enhancement success, and thus either party is penalized for failing to meet enhancement obligations. In addition, recent deliberations of the Pacific Salmon Commission attest to the concerns that biologists and managers have over the genetic effects of hatchery management. However, no study to date has adequately addressed concerns over fitness differences between hatchery and wild sockeye salmon or their F2 crosses (or in any Pacific salmon species in British Columbia and Alaska), in part because of the long-term commitment needed to fully address the question, but also because of the logistical challenges in conducting such research. The weir at Auke Creek provides an unprecedented opportunity to address this critical issue.

This project addresses Strategic Objective 3 of the Northern Fund, "Recognizing that a carefully designed enhancement program would contribute significantly to the restoration of depressed natural stocks and assist the Parties in achieving optimum production." This project has direct relevance to the Comprehensive Salmon Enhancement Plan for Southeast Alaska, which states that the purpose of the state's enhancement program is "to benefit the public by providing additional harvest opportunities to regional salmon fisheries *without adversely affecting natural stocks*" (ADF&G 2004, p. 1; italics added). Alaska's Genetics Policy for fishes acknowledges that the policy is constrained by the "limited amount of information available on the genetic impacts of salmon enhancement on wild stocks" (Davis et al. 1985, p. 1). Similarly, Canada's Policy for Conservation of Wild Pacific Salmon (Fisheries and Oceans Canada, 2005) calls for a biological risk framework for assessing impacts of enhancement on wild stocks, but little data exist to fully implement such assessments. The analysis proposed is providing crucial empirical data to inform implementation of policies in both Alaska and British Columbia. Successful implementation of each proponent's conservation and management policies is essential for the success of the PST; in addition, effective enhancement and supplementation activities are directly related to harvest and harvest allocation under the PST.

**Objectives for Phase 1**

1) to demonstrate the accuracy of parentage assignment in Auke Creek sockeye salmon with existing genomic resources;
2) to identify a cost-effective genetic marker panel for parentage assignment;
3) to complete genotyping of adults from 2012-2015 and back-fill missing genotypes from return year 2010;
4) to assign adults returning in years 2014 and 2015 to candidate parents in unsupplemented years 2008-2010 and supplemented years 2011-2012

Phase 1 was divided into three years. Objectives 1 and 2 were accomplished during the first year of the project (2014-2015). Years 2 and 3 of the project were spent continuing to genotype returning adult sockeye salmon and back-filling missing genotypes (Objective 3), and assigning adult offspring to candidate parents as genotypes became available (Objective 4).

**Methods**

The overall approach of the study is to sample all adult sockeye salmon ascending the weir at Auke Creek; use genotypic information to assign adult offspring back to parents; and then use this information to quantify the relative fitness of hatchery- and natural-origin sockeye salmon over three experimental brood years (2011-2013).

*Hatchery Supplementation*

Experimental supplementation of sockeye salmon took place in Auke Creek over three successive brood years (2011-2013), and was conducted under a separate contract with the Northern Fund (J. Joyce, NOAA, principal investigator). During each brood year, ~30 females and ~15 males were taken from the returning wild sockeye salmon adult population for use in hatchery crosses. These adults were sampled for genetic tissue (by removing an axillary process), and then held in tanks at the weir facility in water from a deep lake intake, mimicking their natural tendency to hold deep in the lake prior to final maturation and spawning. Adults were spawned in September. Cross details are provided in Appendix A. Embryos were incubated over the winter, ponded in early spring, and released into the lake as young-of-the-year fry. Based on scale analysis of Auke Creek sockeye salmon samples (J. Joyce and S. Taylor, NOAA, unpubl. data), adults are expected to return 3-6 years after being spawned. Therefore, we expect to see first-generation hatchery individuals returning during the years 2014-2019 (and second-generation individuals from 2017-2025; Figure 1).

*Adult sampling and genotyping*

Each summer, all adult sockeye salmon were visually identified as male or female (although field identification is not 100% accurate; J. Joyce, NOAA, unpubl.) and sampled for genetic tissues as they were passed upstream through the Auke Creek weir. Tissue samples was sent to the ADF&G Gene Conservation Laboratory (GCL) for genotyping. Genomic DNA was extracted using DNeasy ® 96 Tissue Kit (Qiagen). In the first year (2014-15), samples were genotyped at 12 short tandem repeat (STR; also known as microsatellites) and 93[1] single nucleotide polymorphism markers (SNPs). After accomplishing Objective 2 (identifying a cost-effective marker panel), we reduced this to 9 STRs and 48 SNPs (see Results).

STRs were amplified and electrophoresed at ADF&G, sent to UAF for genotypic determination ('scoring') in years 1 and 2, and scored at ADF&G in year 3. Amplification was carried out in 10 µl reaction volumes [10 mM Tris-HCl, 50 mM KCl, 0.2 mM each dNTP, 0.5 units Taq DNA polymerase (Promega, Madison, WI)] using an Applied Biosystems (AB, Foster City, CA) thermocycler. Multiplexes, including specific cycling parameters and locus-specific primer concentrations, are defined in Appendix B. PCR fragments were analyzed on an AB 3730

[1] Note that the full SNP panel originally consisted of 96 SNPs, but 3 of these were mitochondrial SNPs, which because they are inherited maternally, do not lend themselves well to parentage assignment so were not analyzed.

capillary DNA sequencer. A 96-well reaction plate was loaded with 0.5 ul PCR product along with 0.5 ul of GS500LIZ (AB) internal lane size standard and 9.0 ul of Hi-Di (AB). PCR bands were visualized and separated into bin sets using AB GeneMapper software v4.0. Automated binning was subsequently confirmed or corrected ('scored') manually.

SNP assays were conducted at ADF&G. Extracted DNA was loaded into two Fluidigm® 192.24 Dynamic Arrays in a post-PCR laboratory at ADF&G. Groups of 192 samples and 24 assays were systematically combined into 4,608 parallel reactions on each array. Each reaction was a mixture of 4µl of assay mix (1x DA Assay Loading Buffer (Fluidigm), 10x TaqMan® SNP Genotyping Assay (Applied Biosystems), and 2.5x ROX (Invitrogen)) and 5µl of sample mix (1x TaqMan® Universal Buffer (Applied Biosystems), 0.05x AmpliTaq® Gold DNA Polymerase (Applied Biosystems), 1x GT Sample Loading Reagent (Fluidigm) and 60-400ng/µl DNA) combined in a 7.2nL chamber. Thermal cycling was performed on an Eppendorf IFC Thermal Cycler as follows: 70°C for 30 min for "Hot-Mix" step, initial denaturation of 10 min at 96°C followed by 40 cycles of 96° for 15 sec and 60° for 1 min. The Dynamic Arrays was read on a Fluidigm® EP1™ System or BioMark™ System after amplification and scored using Fluidigm® SNP Genotyping Analysis software.

Assays that failed to amplify on the Fluidigm system were reanalyzed on the Applied Biosystems platform. Each reaction on this platform was performed in 384-well reaction plates in a 5µL volume consisting of 5-40ng/µl of template DNA, 1x TaqMan® Universal PCR Master Mix (Applied Biosystems), and 1x TaqMan® SNP Genotyping Assay (Applied Biosystems). Thermal cycling was performed on a Dual 384-Well GeneAmp® PCR System 9700 (Applied Biosystems) as follows: an initial denaturation of 10 min at 95°C followed by 50 cycles of 92°C for 1s and annealing/extension temperature for 1 min. The plates were scanned on an Applied Biosystems Prism 7900HT Sequence Detection System after amplification and scored using Applied Biosystems' Sequence Detection Software (SDS) version 2.2.

*Quality control of genotypic data*

Quality control analysis (QC) was conducted at ADF&G to identify laboratory errors and to quantify our genotyping error rate (which is necessary for downstream parentage analysis; see below). The QC analyses were performed by staff not involved in the original genotyping. ADF&G staff re-extracted 8% of sample individuals and assayed them for the same markers assayed in the original round of genotyping. The discrepancy rate (which identified DNA extraction, assay plate, and genotyping errors) was calculated as the number of conflicting genotypes divided by the total number of genotypes compared. The discrepancy rate was then divided by two to give the genotyping error rate.

*Parentage analysis*

Parentage analysis consists of using genotypic information to identify, from a pool of candidate parents, the true parents of a given individual. This analysis is fundamental to quantifying fitness differences between hatchery and wild sockeye salmon, as we have defined individual fitness for this study as the number of returning adult offspring (over all potential return years) produced by a focal individual. The earliest we will be able to estimate individual fitness is after the 2017 adult return year (for brood year 2011; see Figure 1), so here we limit describe our methods for parentage assignment. A full treatment of methods for estimating fitness is provided in Appendix C.

Parentage analysis relies on simple Mendelian inheritance to assign offspring to parents based on a combination of exclusion and probabilistic methods. For this study, we used the program *FRANz* (Riester et al. 2009) to assign offspring to parents. Its algorithm uses Mendelian principles in combination with prior information (e.g., genotyping error rate) to determine the maximum likelihood pedigree (i.e., set of parent-offspring triads that best fits the observed genotypic data). *FRANz* was our method of choice because it is computationally feasible (Ford et al. 2012, Kodama et al. 2012), can handle both SNP and STR data, and is expected to perform well in situations such as this study, where we are able to sample a very high proportion (approaching 100%, depending on rates of straying and residualism) of potential parents (Almudevar and LaCombe 2012).

*Objective 1: demonstrate the accuracy of parentage assignment in Auke Creek sockeye salmon with existing genomic resources*

Two classes of error in parentage assignment can contribute towards biases in estimates of relative fitness: 'type A', where one fails to assign an offspring to a true parent that is in the parental dataset; and 'type B', where one assigns an offspring to an untrue parent whether or not the true parent is sampled (Araki and Blouin 2005). We wished to know how accurate we could expect parentage assignment to be based on genomic resources that existed at the inception of this project, so we estimated Type A and B error rates with the initial 93-SNP and 12-STR panels. For type A errors, we genotyped 200 hatchery fry from brood year 2013 (sampled in spring of 2014 prior to release) and estimated the rate at which they failed to assign to the pool of known hatchery parents from that brood year. For type B errors, we determined the rate at which offspring were assigned to a set of individuals that could not be their parents. This was conducted in two different ways. First, we attempted to assign adults returning in year 2009 as the "offspring" of adults returning in subsequent years 2011 (which could not possibly be their parents). Second, we attempted to assign the 200 hatchery fry from brood year 2013 to the wild parents that spawned in 2013 (i.e., we removed the known hatchery parents from the pool of potential parents). Although the offspring sample size was low, the candidate parent pool was from the same return year as the true parents, thus providing a more realistic view of parentage assignment.

*Objective 2: identify a cost-effective genetic marker panel for parentage assignment*

Based on considerations of both Type A and Type B error rates, we sought to identify a smaller marker panel that could be used to accomplish the project goals at a reduced cost per sample. For parentage assignment, the most valuable markers are those with the highest effective number of alleles (A. Gharrett, pers. comm.) for STRs, and for SNPs, those loci with a minor allele frequency (MAF) approaching 0.5 (Thompson 1975, Anderson and Garza 2006). To identify a cost-effective panel, we sequentially removed STR markers with lower effective number of alleles and SNPs (in genotyping units of 24 loci) with low MAF. Markers were removed as genotyping units (i.e., multiplex units for STRs or sets of 24 for SNPs). At each sequential removal, we recalculated assignment type A and B error rates (as described previously). If a marker contributed little to improved accuracy, we assessed the benefits of removing it based on the unit of genotyping. For example, STR loci are assayed in multiplexes (and SNPs in sets of 24), so removing a few loci might do little to reduce laboratory costs, but considerable savings might be achieved by removing an entire multiplex.

*Objective 4: assign adults returning in years 2014 and 2015 to candidate parents in unsupplemented years 2008-2010 and supplemented years 2011-2012*

At the time of this report, none of the experimental brood years had complete returns (the remaining brood of 2011 is expected back at the end of 2017). However, we had genotypes from 'high value' individuals from 2014 (n = 535; where 'high value' was defined as the youngest individuals, i.e., those identified in the field as jacks), and ~3,500 individuals from 2015 including high-value samples.

We conducted parentage assignment on each return year of offspring separately, back to parent years 2008 - 2011 for 2014 offspring and 2009-2012 for 2015 offspring. We had low estimated genotyping error rates (< 0.002 for both SNPs and STRs), so we set fairly stringent parameters for assignment, allowing a maximum of one Mendelian incompatibility between each dyad or triad assignments. We set the maximum number of potential candidate parents (*nmax*) as the total number of adults counted at the weir summed over all possible parent brood years, plus 5% to account for count error, potential residual males (i.e., those that mature without leaving freshwater), or the unaccounted parents of individuals straying into Auke Creek from other sockeye salmon populations. We initially conducted a number of FRANz runs under varying *nmax* and found that values above the total count plus 5% did not change assignment results (data not shown), suggesting that our approach to setting *nmax* was reasonable. We accepted assignments, either to a parent in the candidate set or not to a parent, that had a posterior probability > 0.9; assignments with posterior probability < 0.9 were treated as missing parentage data.

We tallied the number of returning offspring in 2015 that assigned back to wild versus hatchery parents from the 2011 brood year (too few fish returning in 2014 originated from the 2011 brood year). For individuals that assigned back to at least one parent in the candidate set and also had scale-based age information (n = 302 in 2014 and n = 270 in 2015), we compared the ages based on scales to those based on parentage assignment. Nonrandom distributions of discrepancy rates with respect to freshwater or saltwater age determination would suggest that the age error was more likely due to error in scale reading than error in parentage assignment.

Individuals that assigned with confidence (posterior probability > 0.9) back to no parents in the candidate parent set could result from missing genotypic data for both parents in the Auke Creek population, or they could be individuals that strayed into Auke Creek from another population. To evaluate the latter possibility, we conducted a principal components analysis (PCA) on all individual genotypes from return years 2008 - 2015 (conducted in the R package adegenet, v. 2.0.1; Jombart et al. 2008). We then compared the scores for the first three principal components for individuals returning in 2014 and 2015 who assigned back to at least one parent in the candidate set to those who assigned with high confidence to no parents in the candidate set; lack of overlap would support the interpretation that these individuals were strays from outside the Auke Creek population.

**Results and Discussion**

Numbers of sockeye salmon returning to Auke Creek since 2008 have ranged from 1,1262 in 2008 to 4,720 in 2015 (Figure 2). The proportion of jacks by return year has ranged from less than 1% in 2016 to just over 5% in 2011 and 2013.

*Objective 1: demonstrate the accuracy of parentage assignment in Auke Creek sockeye salmon with existing genomic resources;*

*Objective 2: identify a cost-effective genetic marker panel for parentage assignment*

We assessed accuracy of the full panel of markers for Objective 1, and then used this information to compare various reduced marker panels for Objective 2. Therefore, we report the results for both Objective 1 and 2 together. Previous analyses (E. Anderson, NOAA, unpubl; P. Barry, UAF, unpubl.) indicated that a panel of 48 SNPs, when combined with 10-12 STRs, was acceptable for type A and B errors; 24 SNPs was too few. Hence, for the Type A analyses we focused on comparing panels of 93 and 48 SNPs (i.e., the 48 loci with the highest MAF of the 93 nuclear SNPs from ADF&G's full 96-SNP panel) and varying combinations of STR loci, based on their multiplex status (shown in Appendix B). For the Type B comparisons, we report only results from 48-SNP panels, as return years 2011 and 2009 were not genotyped at the full 93-SNP panel (although return year 2013, as well as the resulting hatchery fry, were).

To analyze the effect of dropping STR loci on Type A error rates (using the 2013 hatchery fry data), we pessimistically assumed that ~10% of the true parents had gone unsampled. Results for the Type A error rate are reported in Table 1. Dropping all 12 STR loci yielded an unacceptably high Type A error rate (19.4% with 93 SNPs) or lack of convergence (with 48 SNPs). Adding only the most variable STR locus (*Ssa419*) to the SNPs resulted in a Type A error rate of 1.5% and 3%, with 93 and 48 SNPs, respectively. Any other combination of STR loci we tested had a Type A error rate of 0%.

We expect Type B error rates to be more sensitive to the proportion of parents unsampled, because assuming a high proportion of parents are in the data set increased the probability that an individual was assigned to an incorrect parent rather than going unassigned. Therefore, we initially compared Type B error rates (using 2009 "offspring" assigned to 2011 "parents") over three different percentages of missing parents in the candidate set: 10%, 25%, and 50%. Preliminary analysis showed no change in rank order of Type B error rate of STR combinations over the three percentages, so for the 2013 fry analysis we only reported results for 10% missing parents. Based on results from the Type A analyses, we tested STRs in combination with the 48-SNP panel only (i.e., we did not test the 93-SNP panel, because the additional 45 SNPs did not improve Type A error). For the Type B analyses we used an empirically derived genotyping error rate of 0.0025, which was based on the average of the SNP genotyping error rate over four return years and a pessimistic estimate of the STR genotyping error rate.

Type B error rates in the 2009-to-2011 tests (Table 2) ranged from 5.5% (48 SNPs and all 12 STRs) to 14.3% (48 SNPs and only *Ssa419*, the most variable but also most error-prone STR). These error rates were considered quite pessimistic, as no true parents were included in the data set and siblings were possibly present in 2009-to-2011 comparisons, increasing the likelihood of false assignment (Olsen et al. 2001). In the hatchery fry tests, error rates ranged from 0% to 6.5% depending on locus combination (Table 2). Results of the hatchery fry tests were considered somewhat more optimistic than the 2009-to-2011 tests, because siblings were less likely to complicate the analyses. Furthermore, we did not include the more realistic situation of attempting to assign these individuals to potential parents from multiple brood years. However, it is worth noting that the LOD scores for correct assignments in the Type A tests (Table 1) exceeded those from the comparable Type B tests (Table 2) by an order of magnitude, suggesting that if the correct parents are indeed in the dataset (which is highly likely given our comprehensive adult sampling) then the correct assignment is far more likely. It is also worth noting that in practice we would not accept assignments with low posterior probability for triad assignments (< 0.90), and in practice such an assignment would be categorized as "unknown" (i.e., missing parentage data), rather than assigned or unassigned.

Based on the results described above, we concluded that the locus panel that best optimizes the trade-offs between the benefits of more information from more loci and the costs of additional time and materials consists of 48 SNPs and the STR panel without M3 (a multiplex reaction of *Oki16*, *Omy77*, and *Ots103*). Dropping one of the four STR multiplexes (from 12 to 9 loci) reduces the costs of the project going forward.

*Objective 3) complete genotyping of adults from 2012-2015 and back-fill missing genotypes from return year 2010*

Our pace of genotyping for the project varied depending on adult return rates and funding levels over the course of the project. By the the end of Phase 1, we completed genotyping for return years 2008-2013 and focused genotyping efforts for years 2014 and 2015 on higher-value samples. We genotyped all jacks from return years 2014-2015, as they were more likely than older fish to have originated from the experimental brood years 2011-2013 (see Figure 1). We also genotyped all individuals that were sampled for age and length, as those fish are useful for comparing ages based on scales to those based on parentage assignment. Older individuals and those not sampled for age and length from return year 2014 were considered low priority. Thus far, we have received sufficient funds to genotype almost 75% of the fish returning in 2015. Although we have subjected 100% of the returning-adult tissue samples from 2008- 2013 (and a lower proportion of returning adults in 2014 and 2015 per above discussion) to DNA extraction and genotyping, some individuals, or some loci within some individuals, failed to be genotyped correctly. Table 3 summarizes status of the genotyping by the end of Phase 1, including rates of missing genotypes by return year, and for experimental supplementation brood years 2011-2013 by reproductive type (wild vs. hatchery broodstock). We are still filling in a few important gaps in the dataset, including five SNPs missing from 2013.

*Objective 4) assign adults returning in years 2014 and 2015 to candidate parents in unsupplemented years 2008-2010 and supplemented years 2011-2012*

We attempted to assign 535 individuals returning in 2014 to potential parents from 2008-2011, and 3,503 individuals returning in 2015 to parents in years 2009-2012. These assignments are summarized in Table 4. We were able to assign 94.4% and 95.8% of these individuals to one or both parents from Auke Creek. In both years, the majority of fish originated from brood year 2009, reflecting the effect of the large escapement that year (4,016 fish), even though our sampling of offspring in 2014 was skewed young (targeting jacks) and sampling of the 2015 offspring thus far was also slightly skewed young.

In the 2015 return year, we detected a considerable number of hatchery-origin offspring from brood year 2011 (68 returning adults). For brood year 2011, hatchery parents yielded an average

of 1.7 returning four-year-old offspring per spawner while wild parents yielded an average of 0.09 four-year-old offspring per spawner. If hatchery fish have younger maturing offspring (e.g., Ford et al. 2012), the difference in yield between wild and hatchery parents will diminish as more wild offspring from the 2011 brood year return in later years. Until brood-year returns are complete, this quantitative comparison cannot be used to draw conclusions about relative fitness of wild and hatchery origin fish, other than to say that we do not see indications that hatchery fish had substantially lower fitness than wild fish for brood year 2011.

Agreement between individual ages as determined by scales and as determined by parentage assignment was greater in 2014 (90.7 %) than in 2015 (70.7%; Table 5). While discrepancies could result from errors in either parentage assignment or scale age designation, analysis of discordant individuals suggested that error in determination of freshwater age was a more likely cause for the poor concordance in 2015. In the 2015 cohort, individuals designated as FW1 (i.e., thought to have spent only one year in freshwater based on scales) were more likely to disagree with parentage-based age than to agree, and were also proportionately overrepresented in age-discordant individuals compared to concordant ones (Kolmogorov-Smirnov test, $p < 0.0001$). This was not the case in 2014 (Figure 3; Kolmogorov-Smirnov test, $p = 0.71$). There was no difference in the distribution of saltwater-age designations between age-concordant and discordant samples in either 2014 or 2015 (Figure 4). It is therefore possible that many fish in 2015 were designated as FW1 when in fact they spent two years in freshwater.

Our results provide compelling evidence that sockeye salmon straying into Auke Creek can be identified with genotypic information. A principal components analysis on all individual genotypes from return years 2008 - 2015 yielded a dense cluster of individuals, presumably spawned in Auke Creek, as well as a scattering of individuals falling outside the central cluster (Figure 5). These outlier points could be individuals possessing particularly rare multilocus genotypes, or they could be individuals that strayed in from other source populations. Given the very large sample size (N = 13,600), the former explanation seems unlikely. Alternatively, we found strong support for the idea that these outlier individuals were primarily strays, because in both 2014 and 2015, all of the individuals that assigned with high confidence ($p > 0.9$) to no parents in the candidate had outlier principal components scores (Figure 6). Incidentally, all of the individuals chosen for hatchery broodstock in years 2011-2013 had PC scores well inside the primary cluster (data not shown), reflecting their Auke Creek ancestry. This is useful for the primary goal of this study, as it has been recently demonstrated that ancestral history can bias estimates of relative reproductive success between hatchery and wild fish (Willoughby and Christie 2017).

**Conclusions and Next Steps**

In conclusion, at the end of Phase 1 we have met the four objectives defined at the outset, with the recognition that there are a few gaps in the genotypes dataset that we are still working to fill in.

Reaching the ultimate goal of the project is a long-term commitment, as the oldest offspring from the last experimental brood year (2013) will not return until summer of 2019. During Phase 1 we demonstrated the effectiveness of the approach, and we have taken steps to increase the cost-effectiveness of the project going forward.

A study of this size generates a large amount of data. Genotypes from this project are housed in ADF&G's LOKI database, but a relational database for housing demographic information as well as parentage assignment/pedigree information is also necessary. In consultation with fisheries geneticists carrying out comparable projects (Eric Anderson, NOAA Southwest Fisheries Science Center; Kyle Shedd, Tyler Dann, and Eric Lardizabal, ADF&G Gene Conservation Lab), we have created a Filemaker Pro database, housed at UAF, that store assignments and connect this information to genotypes, demographic information, and run parameters for parentage assignment. The database design is shown in Appendix D.

**References**


ADFG 2004. Comprehensive salmon enhancement plan for Southeast Alaska: Phase III. Alaska Department of Fish and Game, Juneau, AK. www.adfg.alaska.gov/static/fishing/PDFs/hatcheries/plans/se_comprehensivesalmonplan_p3.pdf

Almudevar, A., and J. LaCombe. 2012. On the choice of prior density for the Bayesian analysis of pedigree structure. Theoretical Population Biology 81:131–43.

Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567–2582.

Araki, H., and M. S. Blouin. 2005. Unbiased estimation of relative reproductive success of different groups: evaluation and correction of bias caused by parentage assignment errors. Molecular Ecology 14:4097–4109.

Araki, H., and C. Schmid. 2010. Is hatchery stocking a help or harm? Aquaculture 308:S2–S11.

Araki, H., W. R. Ardren, E. Olsen, B. Cooper, and M. S. Blouin. 2007. Reproductive success of captive-bred steelhead trout in the wild: evaluation of three hatchery programs in the Hood river. Conservation Biology 21:181–90.

Araki, H., B. Cooper, and M. S. Blouin. 2009. Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. Biology Letters 5:621–4.

Berejikian, B. A., D. M. Van Doornik, J. A. Scheurer, and R. Bush. 2009. Reproductive behavior and relative reproductive success of natural- and hatchery-origin Hood Canal summer chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 66:781–789.

Davis, B., B. Allee, D. Amend, B. Bachen, B. Davidson, T. Gharrett, S. Marshall, and A. Wertheimer. 1985. Alaska Department of Fish and Game Genetic Policy. www.adfg.alaska.gov/static/fishing/PDFs/research/genetics_finfish_policy.pdf

Fisheries and Oceans Canada. 2005. Canada's policy for conservation of wild Pacific salmon. Vancouver, B. C. http://www.pac.dfo-mpo.gc.ca/fm-gp/species-especes/salmon-saumon/wsp-pss/index-eng.htm

Ford, M., A. Murdoch, and S. Howard. 2012. Early male maturity explains a negative correlation in reproductive success between hatchery-spawned salmon and their naturally spawning progeny. Conservation Letters 5:450–458.

Fraser, D. J. 2008. How well can captive breeding programs conserve biodiversity? A review of salmonids. Evolutionary Applications 1:535-586.

Hess, M. A., C. D. Rabe, J. L. Vogel, J. J. Stephenson, D. D. Nelson, and S. R. Narum. 2012. Supportive breeding boosts natural population abundance with minimal negative impacts on fitness of a wild population of Chinook salmon. Molecular Ecology 21:5236–50.

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

Kitada, S., H. Kishino, and K. Hamasaki. 2011. Bias and significance of relative reproductive success estimates based on steelhead trout (*Oncorhynchus mykiss*) data : a Bayesian meta-analysis. Canadian Journal of Fisheries and Aquatic Sciences 68:1827-1835.

Kodama, M., J. J. Hard, and K. A. Naish. 2012. Temporal variation in selection on body length and date of return in a wild population of coho salmon, *Oncorhynchus kisutch*. BMC Evolutionary Biology 12:116.

Olsen, J. B., C. Busack, J. Britt, P. Bentzen. 2011. The aunt and uncle effect: an empirical evaluation of the confounding influence of full sibs of parents on pedigree reconstruction. J. Hered. 92:243-247.

Riester, M., P. F. Stadler, and K. Klemm. 2009. FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics 25:2134–2139.

Ryman, N., and L. Laikre. 1991. Effects of supportive breeding on the genetically effective population size. Conservation Biology 5:325–329.

TCTR. 1989. Progress report and recommendations for joint Canada/U. S. sockeye salmon enhancement projects in the Stikine and Taku river drainages. Pacific Salmon Commission Joint Transboundary Technical Committee, Report TCTR (89)-2. http://www.psc.org/publications_tech_techcommitteereport.htm#TCTR

TCTR. 2012. Salmon management and enhancement plans for the Stikine, Taku, and Alsek rivers, 2012. Pacific Salmon Commission Joint Transboundary Technical Committee, Report TCTR (12)-1. http://www.psc.org/publications_tech_techcommitteereport.htm#TCTR

Thompson, E. A. 1975. The estimation of pairwise relationships. Annals of Human Genetics 39:173-188.

Wang, J., and N. Ryman. 2001. Genetic effects of multiple generations of supportive breeding. Conservation Biology 15:1619–1631.

**Table 1.** Type A error rates estimated by comparing hatchery fry (N=200) to potential parents (hatchery, N = 42 and wild, N = 1,985) from the 2013 brood year, using 93 or 48 SNPs and varying combinations of STR loci. Mean posterior probability (post. prob.) was the average for for triad assignments (offspring to both parents). Multiplexes (e.g., M3) are shown in Appendix B.

| STR panel | # STR | N incorrect (1 parent) | N incorrect (2 parents) | Mean LOD (SD) | Mean post. prob. | Type A error rate |
|---|---|---|---|---|---|---|
| *93 SNPs* | | | | | | |
| All STRs | 12 | 0 | 0 | 72.7 (11.5) | 1.000 | 0 |
| No *Ssa419* | 11 | 0 | 0 | 67.5 (10.8) | 0.999 | 0 |
| Only *Ssa419* | 1 | 2 | 1 | 25.1 (6.4) | 0.976 | 0.01 |
| No M3 | 8 | 0 | 0 | 56.1 (9.6) | 0.997 | 0 |
| No M1 | 5 | 0 | 0 | 51.4 (9.5) | 0.998 | 0 |
| M2 only | 4 | 0 | 0 | 34.9 (7.8) | 0.995 | 0 |
| M1 + *Ssa419* | 6 | 0 | 0 | 46.3 (8.5) | 0.999 | 0 |
| No M3, no *One8* | 8 | 0 | 0 | 51.1 (9.2) | 0.998 | 0 |
| No STRs | 0 | 26 | 13 | 20.4 (5.2) | 0.866 | 0.13 |
| | | | | | | |
| *48 SNPs* | | | | | | |
| All 12 STRs | 12 | 0 | 0 | 68.5 (9.9) | 1.000 | 0 |
| No *Ssa419* | 11 | 0 | 0 | 63.3 (9.3) | 0.998 | 0 |
| Only *Ssa419* | 1 | 6 | 0 | 20.9 (4.5) | 0.944 | 0.015 |
| No M3 | 8 | 0 | 0 | 52.0 (8.0) | 0.996 | 0 |
| No M1 | 5 | 0 | 0 | 47.3 (7.8) | 0.996 | 0 |
| M2 only | 4 | 0 | 0 | 30.7 (6.2) | 0.985 | 0 |
| M1 + *Ssa419* | 6 | 0 | 0 | 42.1 (6.7) | 0.996 | 0 |
| No M3, no *One8* | 8 | 0 | 0 | 46.9 (7.7) | 0.993 | 0 |
| No STRs | 0 | FRANz did not converge | | | | |

**Table 2.** Results of Type B error rate estimates for N = 400 'offspring' from 2009 and N = 400 'parents' from 2011 with differing proportions of adults (10% and 50%) missing from the candidate parent pool, and for hatchery fry (N = 200) to wild parents (N = 1,985) assuming 10% missing parents. Mean LOD is for assigned individuals only; mean posterior probability is for triad assignments (offspring to both parents), averaged over all assignments. The 48-SNP panel was used with varying combinations of STR loci, which are shown ranked by Type B error rate. Multiplexes (e.g, M3) are defined in Appendix B.

| % missing parents | STR combination | N incorrect (1 parent) | N incorrect (2 parents) | Mean LOD (SD) | Mean post. prob. | Type B error rate |
|---|---|---|---|---|---|---|
| *2009 'offspring' to 2011 'parents'* | | | | | | |
| 10% | All 12 STRs | 20 | 1 | 5.97 (4.30) | 0.945 | 0.055 |
| | No M3 | 28 | 1 | 5.96 (3.67) | 0.924 | 0.075 |
| | No *Ssa419* | 35 | 1 | 5.89 (3.17) | 0.909 | 0.093 |
| | No STRs | 20 | 11 | 7.65 (2.91) | 0.617 | 0.105 |
| | M2 only | 41 | 2 | 5.56 (1.70) | 0.859 | 0.113 |
| | No M1 | 44 | 1 | 4.92 (1.81) | 0.888 | 0.115 |
| | M1 + *Ssa419* | 48 | 2 | 6.15 (3.11) | 0.831 | 0.130 |
| | Only *Ssa419* | 45 | 6 | 6.34 (2.64) | 0.714 | 0.143 |
| | | | | | | |
| 50% | All 12 STRs | 21 | 1 | 5.97 (4.30) | 0.945 | 0.058 |
| | No M3[a] | 28 | 1 | 5.96 (3.67) | 0.924 | 0.075 |
| | No *Ssa419* | 35 | 1 | 5.89 (3.17) | 0.909 | 0.093 |
| | No STRs | 19 | 11 | 7.80 (2.83) | 0.619 | 0.103 |
| | No M1 | 42 | 1 | 5.02 (1.80) | 0.888 | 0.110 |
| | M2 only | 41 | 2 | 5.56 (1.70) | 0.858 | 0.113 |
| | M1 + *Ssa419* | 47 | 2 | 6.19 (3.12) | 0.829 | 0.128 |
| | Only *Ssa419* | 44 | 6 | 6.37 (2.66) | 0.714 | 0.140 |
| | | | | | | |
| *Hatchery fry to wild parents* | | | | | | |
| 10% | All 12 STRs | 0 | 0 | -- | 1.000 | 0 |
| | No *Ssa419* | 0 | 0 | -- | 1.000 | 0 |
| | No M3[a] | 0 | 0 | -- | 1.00 | 0 |
| | No M1[b] | 0 | 0 | -- | 0.999 | 0 |
| | M1 + *Ssa419* | 1 | 0 | 3.45 (--) | 0.995 | 0.025 |
| | No STRs | 0 | 8 | 1.45 (2.94) | 0.514 | 0.040 |
| | Only *Ssa419* | 0 | 12 | 8.56 (2.99) | 0.524 | 0.060 |
| | M2 only[c] | 18 | 4 | 4.94 (1.82) | 0.903 | 0.065 |

**Table 3**. Progress in genotyping adult sockeye salmon (return years 2008-2015) at the end of Phase 1 (end of project year, 2016-17).

| Return Year | Breeding type | Counted at Weir | Number genotyped by proportion of marker panel | | | | |
|---|---|---|---|---|---|---|---|
| *Years prior to supplementation* | | | 100% | 99-75% | 74-50% | 49-25% | < 25% |
| 2008 | Wild | 1,264 | 0 | 1229 | 16 | 2 | 15 |
| 2009 | Wild | 4,064 | 3451 | 477 | 37 | 22 | 74 |
| 2010 | Wild | 2,000 | 1191 | 564 | 13 | 4 | 290 |
| | | | | | | | |
| *Supplementation years* | | | | | | | |
| 2011 | Wild | 2,382 | 2090 | 261 | 13 | 2 | 15 |
| | Hatchery | 41 | 36 | 3 | 0 | 0 | 2 |
| 2012 | Wild | 1,537 | 1113 | 315 | 20 | 6 | 68 |
| | Hatchery | 32 | 24 | 7 | 0 | 0 | 1 |
| 2013 | Wild | 2,010 | 0[a] | 2001 | 3 | 0 | 1 |
| | Hatchery | 50 | 0[a] | 50 | 0 | 0 | 0 |
| | | | | | | | |
| *Years following supplementation* | | | | | | | |
| 2014 | Natural | 3,443 | 480 | 55 | 0 | 1 | 0 |
| 2015 | Natural | 4,720 | 2251 | 1242 | 6 | 1 | 4 |

[a]currently missing data for 5 SNPs: *One_dds_529, One_DDX5_86, One_GTHa, One_psme2_354, One_U1215_82*

**Table 4.** Summary of parentage assignment for 2014 and 2015 adult offspring, showing the number assigned to two, one, or neither parents in the candidate set, or not assigned with confidence to any parents. UA indicates an assignment made with high confidence (posterior probability > 0.9) to no parent in the dataset; NA indicates that an assignment, either to a parent in the dataset or not, could not be made with confidence (posterior probability < 0.9).

| Return year | Brood year | Two parents | One parent | UA/UA | UA/NA | NA/NA |
|---|---|---|---|---|---|---|
| 2014 (N = 535) | 2008 | 34 | 3 | -- | -- | -- |
| | 2009 | 339 | 44 | -- | -- | -- |
| | 2010 | 19 | 21 | -- | -- | -- |
| | 2011 | 36 | 9 | -- | -- | -- |
| | Unknown brood year | -- | -- | 4 | 22 | 4 |
| | | | | | | |
| 2015 | 2009 | 2,085 | 253 | -- | -- | -- |
| (N = 2,503) | 2010 | 582 | 161 | -- | -- | -- |
| | 2011 | 217 | 54 | -- | -- | -- |
| | 2012 | 2* | 2 | -- | -- | -- |
| | Unknown brood year | -- | -- | 9 | 120 | 18 |

**Table 5.** Comparisons between ages as determined by scale age versus parentage assignment for individuals sampled for scales in 2014 and 2015. Discrepancy values represent the difference in years, with negative values indicating that scale age was younger than parentage-based age.

| Return Year: | 2014 | | 2015 | |
|---|---|---|---|---|
| Discrepancy | N | % | N | % |
| Agreed | 274 | 90.7 | 191 | 70.7 |
| -2 | -- | -- | 3 | 1.1 |
| -1 | 10 | 3.3 | 70 | 25.9 |
| +1 | 17 | 5.6 | 6 | 2.2 |
| +2 | 1 | 0.3 | -- | -- |
| Total | 302 | | 270 | |

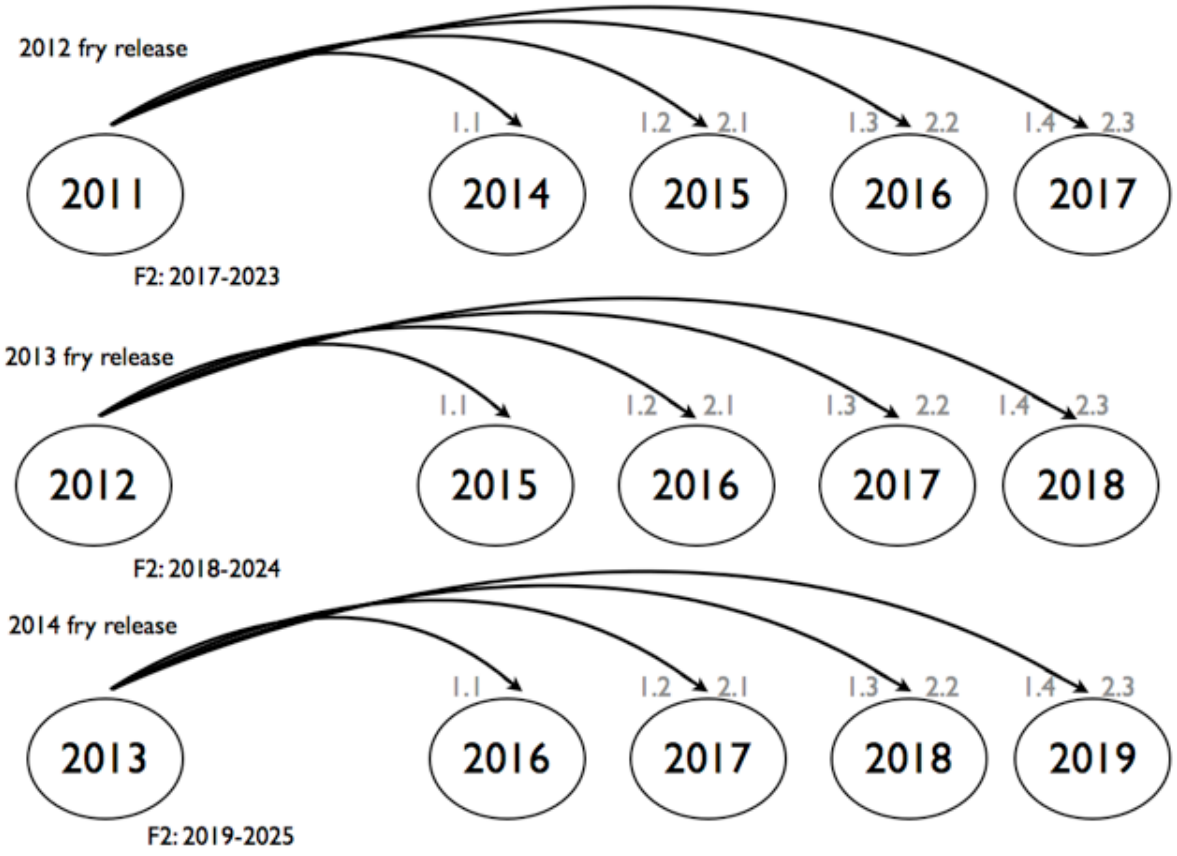


**Figure 1.** Schematic showing potential return years of first-generation adults resulting from brood years 2011-2013 (fry releases 2012-2014, respectively). Numbers above return-year ovals represent age of returns (e.g., 1.1 = one winter in freshwater, one winter in saltwater, age 3). One more cycle would be required to compare fitness between wild-spawning individuals of hatchery and wild origin, as denoted by F2 return years.

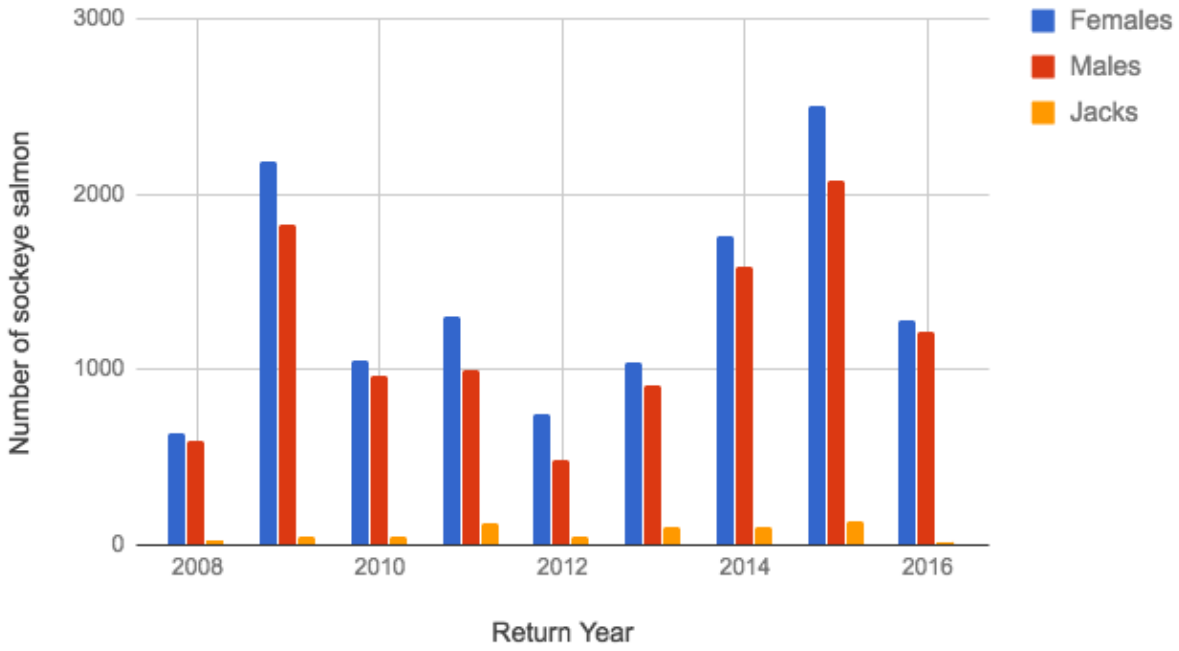


**Figure 2**. Numbers of adult sockeye salmon returning to the Auke Creek weir, 2008-2016, by field-called identity (female, non-jack males and jacks).

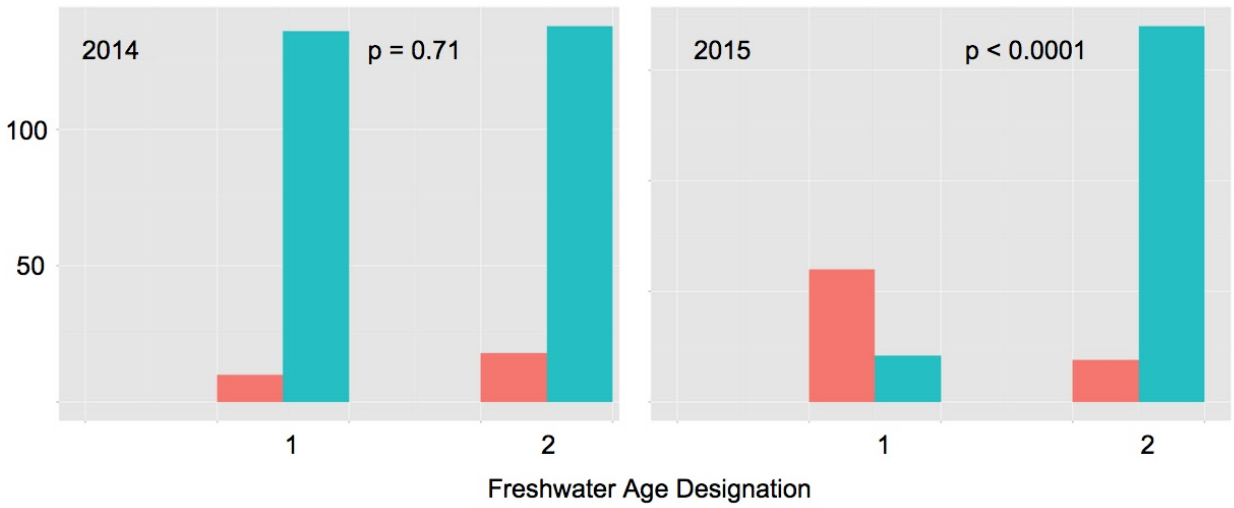


**Figure 3.** Numbers of individuals whose scale-based age disagreed (red) or agreed (turquoise) with parentage-based age by freshwater scale age designation (1 or 2 years in fresh water), for return years 2014 and 2015. P-values are for bootstrap Kolmogorov-Smirnov tests for differences in distribution of freshwater ages by concordance between scale- and parentage-based ages.

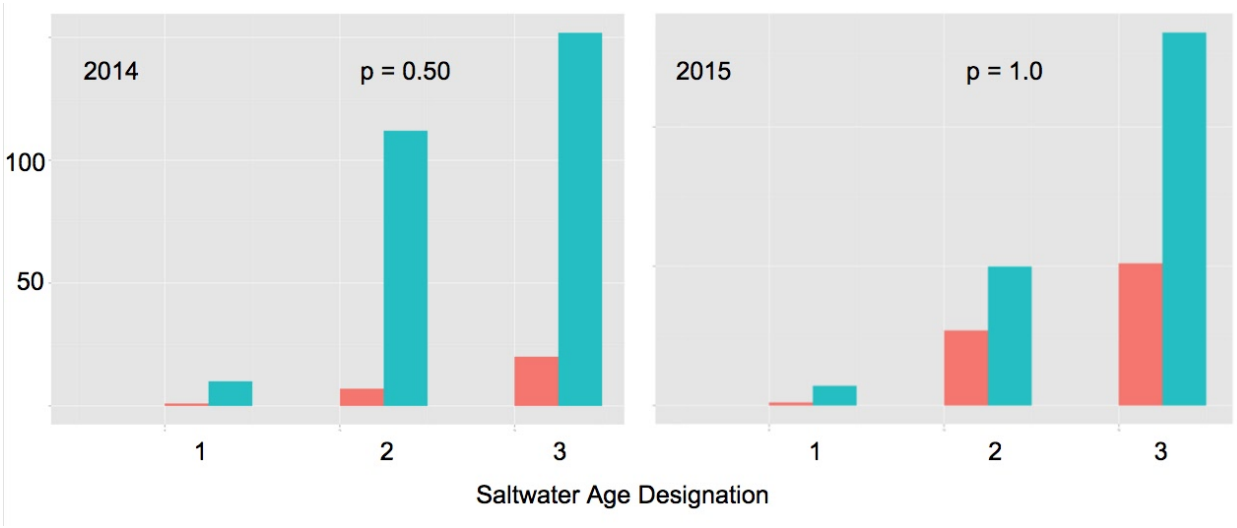


**Figure 4.** Numbers of individuals whose scale-based age disagreed (red) or agreed (turquoise) with parentage-based age by saltwater scale age designation, for return years 2014 and 2015. P-values are for bootstrap Kolmogorov-Smirnov tests for differences in distribution of saltwater ages by concordance between scale- and parentage-based ages.

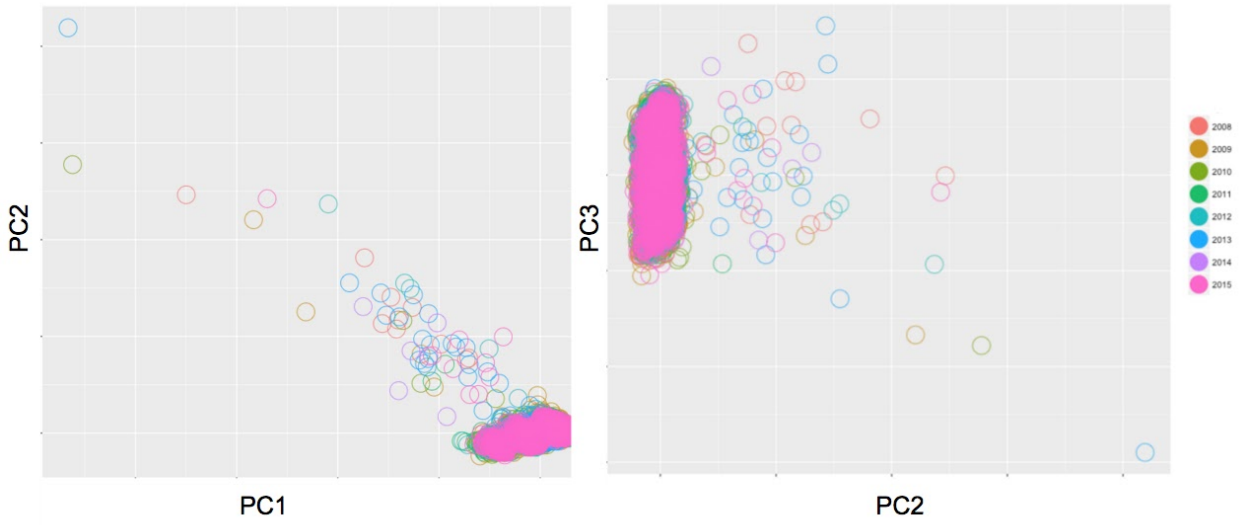


**Figure 5.** Pairwise scatterplots of principal components scores for first three axes (PC1, PC2, and PC3) for individual genotypes from return years 2008 - 2015.


**Figure 6.** Pairwise scatterplot of PC scores (PC1 and PC2) of individuals that assigned with high probability to no parent in the candidate set (turquoise circles) compared to those that assigned to at least one parent in the candidate set (red open circles). Return years 2014 and 2015 are depicted separately.

**Appendix A.** Individuals (identified by Fish ID) involved in crosses at the Auke Creek Hatchery in brood years 2011-2013. ADFG&G species, location, and year ('SILLY') codes are provided by year; the entire identity of genetic sample is the SILLY Code plus Fish ID; e.g., SAUKE11_8312. Individuals with an asterisk are missing > 75% of genotypic data. In 2012, some females were crossed with more than one male.

| Brood Year | | | Brood Year | | | Brood Year | |
|---|---|---|---|---|---|---|---|
| 2011 (SAUKE11) | | | 2012 (SAUKE12) | | | 2013 (SAUKE13) | |
| Female | Male | | Female | Male(s) | | Female | Male |
| 8312 | 8301 | | 11001 | 11003 | | 263898 | 263931 |
| 8313 | 8301 | | 11002 | 11003 | | 263899 | 263931 |
| 8314 | 8302 | | 11004 | 11006 | | 263900 | 263932 |
| 8315 | 8302 | | 11005 | 11006 | | 263901 | 263932 |
| 8316 | 8303 | | 11007 | 11009 | | 263902 | 263933 |
| 8317 | 8303 | | 11008 | 11009 | | 263903 | 263933 |
| 8318 | 8304 | | 11010 | 11012 | | 263904 | 263934 |
| 8319 | 8304 | | 11011 | 11012 | | 263905 | 263934 |
| 8320 | 8305 | | 11013 | 11015 | | 263906 | 263935 |
| 8321 | 8305 | | 11014 | 11015 | | 263907 | 263935 |
| 8322 | 8305 | | 11016 | 11012/11015 | | 263908 | 263936 |
| 8323 | 8306 | | 11017 | 11012/11015 | | 263909 | 263936 |
| 8324 | 8306 | | 11018 | 11020 | | 263910 | 263937 |
| 8325 | 8306 | | 11019 | 11020 | | 263911 | 263937 |
| 8326* | 8307 | | 11021 | 11023 | | 263912 | 263938 |
| 8327 | 8307 | | 11022 | 11023 | | 263913 | 263938 |
| 8328 | 8307 | | 11024 | 11020/11023 | | 263914 | 263939 |
| 8329 | 8308 | | 11025 | 11020/11023 | | 263915 | 263939 |
| 8330 | 8308 | | 11026 | 11028/11029 | | 263916 | 263940 |
| 8331 | 8308 | | 11027 | 11028/11029 | | 263917 | 263940 |
| 8332 | 8309 | | 11030 | 11028/11029 | | 263918 | 263941 |
| 8333 | 8309 | | 11031 | 11028/11029 | | 263919 | 263941 |
| 8334 | 8309 | | 11032 | 11029 | | 263920 | 263942 |
| 8335* | 8310 | | | | | 263921 | 263943 |
| 8336 | 8310 | | | | | 263929 | 263945 |
| 8337 | 8310 | | | | | | |
| 8338 | 8311 | | | | | | |
| 8339 | 8311 | | | | | | |
| 8340 | 8311 | | | | | | |
| 8341 | 8303 | | | | | | |

**Appendix B.** Details of STR genotyping conditions, including their PCR multiplex and electrophoresis multi- or pseudo-plex status.

| PCR Multiplex | Annealing temp. (°C) | Number cycles | Locus | [primer] (μM) | Capillary multiplex |
|---|---|---|---|---|---|
| M1 | 55 | 35 | *Oki100* | 0.20 | 1 |
| | | | *One102* | 0.10 | |
| | | | *One109* | 0.10 | |
| | | | *Oki1a* | 0.06 | |
| | | | *Oki1b* | 0.06 | |
| M2 | 56 | 35 | *Oki10* | 0.20 | 2 |
| | | | *One114* | 0.35 | |
| M3 | 57 | 36 | *Ssa419* | 0.20 | |
| M4 | 60 | 35 | *One8* | 0.20 | |
| M5 | 46 | 35 | *Oki16* | 0.20 | 3 |
| | | | *Omy77* | 0.20 | |
| | | | *Ots103* | 0.20 | |

**Appendix C.** Methods for estimating relative fitness.

Fitness can be defined as the number of an individual's offspring that survive to adulthood (Allendorf and Luikart 2003), and thus encompasses both reproductive success (mating and fertilization) as well as the survival of resulting offspring. We will quantify individual fitness as the number of adult offspring assigned to a focal adult individual. Relative fitness (RF) will be estimated (after Araki et al. 2007) as the ratio of mean individual fitness of hatchery individuals ($F_H$) to the mean fitness of wild individuals ($F_W$):

$$RF = \frac{\hat{F}_H}{\hat{F}_W} = \frac{\sum_{i=1}^{N_H} A_{iH} / P_H}{\sum_{i=1}^{N_W} A_{iW} / P_W} \qquad \text{(eqn 1)}$$

where A is number of offspring assigned and P is number of brood-year adults (potential parents), H denotes hatchery, and W denotes wild.

Araki and Blouin (2005) noted that (eqn 1) would be biased towards unity in the presence of high rates of two forms of error in parentage assignment: failure to assign offspring to its true parent (type A error) and assignment of offspring to a false parent (type B error). They showed that this bias could be corrected for as:

$$RF_{UNB} = \frac{\hat{F}_H - \left(\frac{S-A}{P}\right)\left(\frac{\hat{b}}{1-\hat{b}}\right)}{\hat{F}_W - \left(\frac{S-A}{P}\right)\left(\frac{\hat{b}}{1-\hat{b}}\right)} \qquad \text{(eqn 2)}$$

where S is the total number of offspring sampled, A is the total number of offspring assigned to parents, P is the total number of potential parents sampled, and $\hat{b}$ is the estimated type B error rate. In the Auke Creek system, the proportion of un-sampled parents is likely to be very low, so Araki and Blouin's (2005) estimator may not be entirely appropriate. We have derived an alternative method of estimating $RF_{UNB}$ which we will compare to estimates from (eqn 2); it is derived as follows:

1. Starting with Araki and Blouin's (2005) unbiased estimator:

$$RF_{UNB} = \frac{\hat{F}_H - \left(\frac{S-A}{P}\right)\left(\frac{\hat{b}}{1-\hat{b}}\right)}{\hat{F}_W - \left(\frac{S-A}{P}\right)\left(\frac{\hat{b}}{1-\hat{b}}\right)}$$

2. Given that nearly 100% of the potential parents will be sampled, P ≈ S, usually P > S. And S – A should be << S. So,
(S – A)/P < 1: call it ε.

3. $\hat{b}$ should also be << 1 and $\frac{\hat{b}}{1-\hat{b}} < 1$: call it δ.

4. That means that $\left(\frac{S-A}{P}\right)\left(\frac{\hat{b}}{1-\hat{b}}\right) = \varepsilon\delta << 1$: call it η.

5. We can rewrite $RF_{UNB}$:

$$RF_{UNB} = \frac{\hat{F}_H - \left(\frac{S-A}{P}\right)\left(\frac{\hat{b}}{1-\hat{b}}\right)}{\hat{F}_W - \left(\frac{S-A}{P}\right)\left(\frac{\hat{b}}{1-\hat{b}}\right)} = \frac{\hat{F}_H - \eta}{\hat{F}_W - \eta} = \frac{\frac{\hat{F}_H - \eta}{\hat{F}_W}}{\frac{\hat{F}_W - \eta}{\hat{F}_W}} = \frac{\frac{\hat{F}_H}{\hat{F}_W} - \frac{\eta}{\hat{F}_W}}{\frac{\hat{F}_W}{\hat{F}_W} - \frac{\eta}{\hat{F}_W}} = \frac{\frac{\hat{F}_H}{\hat{F}_W} - \frac{\eta}{\hat{F}_W}}{1 - \frac{\eta}{\hat{F}_W}}$$

For Auke Creek sockeye salmon, it is likely that $\hat{F}_W >> \eta$; that means that $\frac{\eta}{\hat{F}_W}$ <<1.

We can use the approximation:

$\frac{1}{1-a} \approx 1 + a$, where $|a|$ <<1 to again rewrite $RF_{UNB}$:

$$RF_{UNB} = \frac{\frac{\hat{F}_H}{\hat{F}_W} - \frac{\eta}{\hat{F}_W}}{1 - \frac{\eta}{\hat{F}_W}} \approx \left(\frac{\hat{F}_H}{\hat{F}_W} - \frac{\eta}{\hat{F}_W}\right)\left(1 + \frac{\eta}{\hat{F}_W}\right) = \frac{\hat{F}_H}{\hat{F}_W} + \frac{\hat{F}_H}{\hat{F}_W}\frac{\eta}{\hat{F}_W} - \frac{\eta}{\hat{F}_W} - \left(\frac{\eta}{\hat{F}_W}\right)^2$$

The second order term should be negligible, so:

$$RF_{UNB} \approx \frac{\hat{F}_H}{\hat{F}_W} + \frac{\hat{F}_H}{\hat{F}_W}\frac{\eta}{\hat{F}_W} - \frac{\eta}{\hat{F}_W} = \frac{\hat{F}_H}{\hat{F}_W}\left(1 + \frac{\eta}{\hat{F}_W}\right) - \frac{\eta}{\hat{F}_W}$$

We will use a two-tailed permutation test to assess the probability of obtaining a difference (absolute value) in fitness between wild and hatchery fish greater or equal to the observed difference under the null hypothesis $F_W = F_H$, employing the bias correction of Araki and Blouin (2005):

$$\hat{F}_H - \hat{F}_W = \left(F_H - F_W\right)(1 - a) \qquad \text{(eqn 3)}$$

where $a$ is the type A error rate (estimated as $\hat{a}$).

Because RF is a ratio of two small numbers, it will be very sensitive to sources of variation (such as interannual variability), and could be biased either upwards or downwards even if biases in the

numerator and denominator are corrected. Once we have a full accounting of returning offspring from brood years 2011-2013, we will employ the methods of Kitada et al. (2012) to calculate the relative bias:

$$\frac{E(\hat{R})}{R} = \exp\left(\frac{\sigma_H^2}{P_H} - r\sqrt{\frac{\sigma_H^2}{P_H}\frac{\sigma_W^2}{P_W}}\right) \quad \text{(eqn 4)}$$

where $\sigma_X$ is the variance in $\log(F_X)$ and $r$ is the correlation between $\log(F_H)$ and $\log(F_W)$ over three brood years. We will then calculate the posterior probability that RF is greater or less than unity under the Bayesian framework outlined in Kitada et al. (2012):

$$P_{diff} = P\left(RF \neq 1 \mid \hat{R}F_3\right) \quad \text{(eqn 5)}$$

where $\hat{R}F_3$ is the RF estimated over three consecutive brood years (2011-2013).

We will conduct sensitivity analyses to determine how much error in $\hat{a}$ and $\hat{b}$ would revise our conclusions. Additionally, we will explore the degree to which fractional assignment (dividing progeny among multiple parents proportional to their likelihood), as opposed to binomial assignment (assigned or not), affects our estimates of relative fitness, as binomial assignment based on a likelihood threshold has been shown to bias fitness comparisons in wild and hatchery salmonids (Ford and Williamson 2010).

**Appendix D.** Components of relational database for the Auke Creek sockeye salmon fitness study.

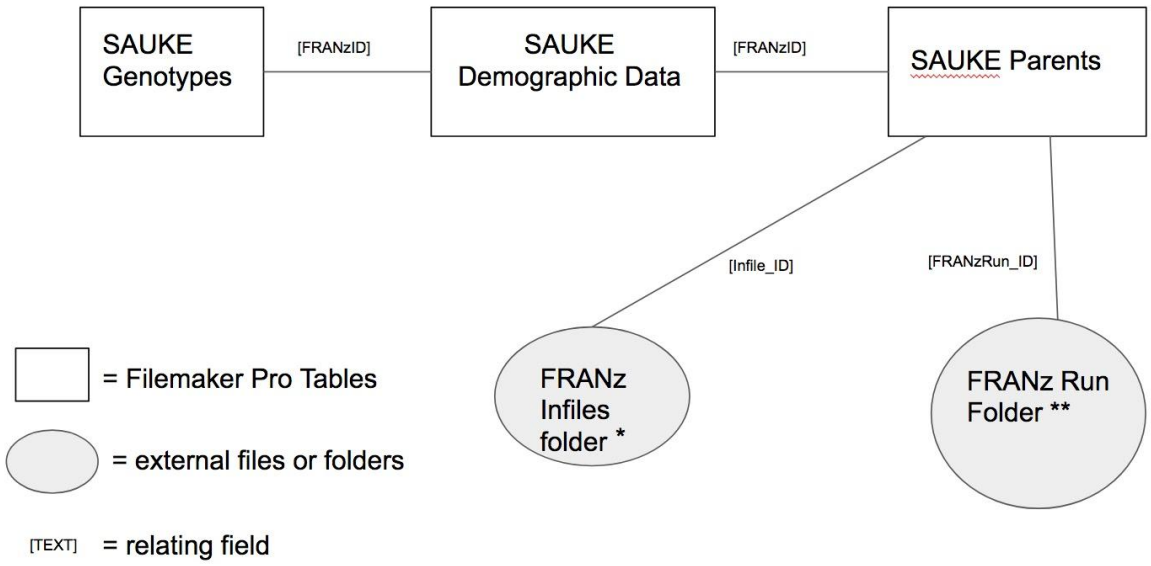


Figure D1. Structure of relational database, showing relationship among tables within database and external files.

Each table contains a common identifier, the FRANzID. This identifier is 10-character name used to identify individuals in infiles for the parentage assignment program FRANz. The first two characters identify the year that individual returned; the third character identifies whether that fish spawned in the wild (w) or in the hatchery (h), and the fourth character identifies sex as identified in the field (m, f, or u for unknown). The last six characters correspond with the fish identification number used by ADF&G Gene Conservation Laboratory (GCL), with zeros used as placeholders before the identification number if that number is fewer than six characters long.

The SAUKE Genotypes table contains the FRANzID, the SILLY code + ID used by the GCL, plus the two-allele genotypes for each locus in a single field (e.g., 210/214 for an STR; A/G for a SNP). The SAUKE Demographic Data table contains information about individual fish, such as their field-identified sex, date of capture at the weir, type as a spawning adult (wild, hatchery, or mortality), and where applicable, their length (mid-eye to fork) and scale age (e.g., 1.3 for a fish that spent one year in freshwater and three years at sea). The SAUKE Parents database provides information about parentage assignment. Each table row consists of an offspring, identified by its FRANzID, the identity of its mother, identity of its father, the posterior probabilities for each parent assignment, the identity of the infile used by FRANz in its parent assignment, and the identity of a folder containing FRANz run parameters and output. FRANz infiles and run folders are archived should they need to be referred back to.