Parentage-based tagging in three western Washington Chinook salmon hatcheries: a proof of concept study

<u>For</u>

PSC Southern Boundary Restoration and Enhancement Fund

Contract Number – Pacific Salmon Commission # SF-2008-I-8
WDFW Contract No. CAPS 08-1246

Sewall F. Young[1], Denise K. Hawkins[2], and Kenneth I. Warheit

Washington Department of Fish and Wildlife, Fish Science Division, 600 Capitol Way N., Olympia, WA 98501

July 2011

[1] Corresponding author: Sewall.Young@dfw.wa.gov
[2] Currently at U.S. Fish and Wildlife Service, Abernathy Fish Technology Center, Longview, WA

Abstract

Parentage-based tagging (PBT), an application of genetic parentage inference to identify the stock and broodyear of origin of fish that do not bear artificial tags, has been suggested as a prospective replacement of or enhancement to coded-wire tags in Chinook salmon catch and abundance assessment. Large-scale validation of PBT in all relevant management scenarios has not been done. This study examines the logistical challenges of sampling the entire spawning populations in large-scale Chinook salmon hatcheries in Washington State and evaluates the reliability of maximum likelihood based parentage assignments in situations where two hatcheries produce Chinook from a common recent genetic lineage but contribute to different spatial strata. We chose George Adams Hatchery on Hood Canal, Washington and Voights Creek Hatchery in South Puget Sound, Washington because both of those Chinook salmon programs were founded with the Green River fall Chinook stock (also known as Soos Creek fall Chinook). We conducted a pilot study, supported by the Washington State General Fund, on the early spawning spring Chinook stock at Marblemount Hatchery in North Puget Sound to test sampling, lab processing and parentage analysis strategies and to evaluate the influence of gamete handling on the mating structure in the hatchery. The PBT-based mating structure analysis revealed that with sequential addition of sperm from primary and backup males, the primary males accounted for > 90% of the offspring produced. Using the experience from the pilot study, in September 2008 we sampled and genotyped 2,159 spawned Chinook at George Adams Hatchery on Hood Canal, Washington and 953 spawned Chinook at Voights Creek Hatchery in South Puget Sound, Washington, the entire populations spawned at the two hatcheries. We used the genotypes of the spawners to simulate sets of offspring from the two stocks with the same expected relatedness as the hatchery production and tested the affects of (1) full-siblings spawning in different allocation strata, (2) incomplete genotypes among the offspring being allocated, (3) genotyping errors in the offspring data, (4) genotyping errors in the parents' data, and (4) incomplete sampling of the adults. Parentage assignment accuracies under all scenarios exceeded 90%. We make the following conclusions concerning the efficacy of using PBT to ascertain maximum likelihood based parentage assignments : First, random imperfections in the data can reduce the apparent likelihoods of valid parent-offspring relationships, but rarely increase the apparent likelihoods of invalid parent-offspring relationships. Second, although there were reductions in likelihoods of valid relationships, the information content of our data set was sufficient to overcome most of these reductions, and for the most-part, we were able to correctly assign simulated individuals to their correct parents.

Introduction

Chinook salmon (*Oncorhynchus tshawytscha)* management in the area covered by the Pacific Salmon Treaty relies on catch and abundance estimates from fisheries and spawning populations covered by the treaty. Those estimates traditionally have been derived with data assembled from coded-wire tag (CWT) recoveries. In response to concerns about the continued viability of the CWT program, the PSC's Chinook Technical Committee assembled a group of scientists to evaluate the future of CWTs in Pacific salmon management and examine alternatives to CWTs. The Expert Panel identified two genetics-based technologies, mixed-stock analyses and parentage analyses, which can identify the origins of fish encountered in areas where stocks are mixed and so might augment or replace CWTs in salmon management (Expert Panel 2005). Genetic methods are attractive because they rely on differences in the states of naturally occurring genetic characteristics that are carried by all individuals; breeding fish "mark" their offspring at fertilization so all progeny are marked and the number of marks recovered from a fishery theoretically equals the number of fish sampled (Waples et al. 2008).

Traditional genetics-based stock mixture analysis methods are well developed (Pella and Masuda 2001) and in common use (Shaklee et al. 1999). However, current genetic baseline data sets and analytical methods do not provide enough resolving power to separate closely related stocks with the precision required for the PSC's cohort analyses (Waples et al. 2008). Furthermore, genetic mixed-stock analysis methods will not provide estimates of cohort abundance unless they are coupled with an age estimation method that has acceptable accuracy and precision (Expert Panel 2005). For example, the widespread hatchery use of fish with Green River ancestry in the Puget Sound basin makes it difficult to differentiate fall Chinook from major regions of Puget Sound and Hood Canal and scale aging is not considered reliable enough by some researchers to use in the cohort reconstructions required by the management model.

Anderson and Garza (2006) proposed the use of genetic parentage analysis to provide stock- and cohort-of-origin identification of hatchery-produced Chinook. The method initially was referred to as Full Parental Genotyping (FPG), but is now commonly called Parentage-Based Tagging (PBT). Parentage-based tagging, as it was originally proposed, would require sampling and genotyping all Chinook spawned in hatcheries to construct a database of prospective parents' genotypes. Tissue samples also would be collected from fish encountered in mixed-stock areas to develop a database of putative offspring genotypes. Genetic parentage analysis methods then would be used to identify the parents of each genotype in the offspring database. The inference of the parent-offspring trios would provide the hatchery stock and the broodyear identities that are required for cohort analyses.

The large-scale use of parentage analysis to identify the origins of fish from multiple source populations has not been tested fully in a complex setting with multiple hatcheries raising related strains derived from a common stock. The scale and scope of real world applications of PBT

require massive sampling and genotyping efforts at all potentially contributing hatcheries, high resolving power in the genetic data, and accurate and robust parentage assignment software.

The Puget Sound Chinook stocks in the Green River lineage provide a good opportunity to test the power of PBT to distinguish closely related stocks. The Green River fall Chinook stock (now known as Soos Creek fall Chinook) provided founders for hatchery programs throughout north and south Puget Sound and Hood Canal. Derived strains from north Puget Sound, south Puget Sound, and Hood Canal cannot be distinguished accurately using traditional genetic mixed-stock analysis methods, but in the current management regime, fishery impacts within those geographic regions must be evaluated. Exploitation and abundance monitoring require identification of the hatchery and brood year of origin of Chinook from the region, so if PBT is to be used as a tool for that monitoring, it must be able to tease apart the origins of fish from hatcheries with a recent common ancestry and it must be able to identify the cohort of each fish in the progeny sample.

The presence of close relatives among the pool of potential parents can bias parentage assignments (Olsen et al. 2001). Allele sharing generally will be more prevalent among relatives in a group of spawners than among unrelated individuals. That can result in relatively large differences in log of odds (LOD) values for non-relatives among the parents and relatively small differences among relatives in the parental pool. Very small differences mean that parentage is ambiguous so we would expect that offspring that do not have aunts or uncles in the parental pool will be assigned to parents at a higher rate than offspring with aunts or uncles in the pool. The potential bias might overestimate the reproductive success of parents that do not have relatives in the pool and underestimate success of parents with close relatives in the pool. That bias is relevant to parentage inferences within a hatchery but is less relevant between hatcheries if they do not exchange broodstock either willfully or by straying.

In this study, we examined the logistical practicality of characterizing complete spawning populations at George Adams Hatchery on Hood Canal and Voights Creek Hatchery in south Puget Sound and developed a protocol for high-throughput sampling that captured precise information about the matings that occurred. We genotyped the samples and used those data to seed a simulation model that assessed the impacts on the accuracy of PBT in Puget Sound, Washington of several factors that are common to population genetics studies including genotyping errors among the offspring and parents, incomplete genotypes among the offspring, and incomplete sampling of the parent population.

As a precursor to the main study, we undertook a pilot study with the Marblemount Hatchery's spring Chinook population with support from the Washington State General Fund. The purpose of the pilot study was to test sampling methods and develop protocols that we could apply to the main study that targeted the fall Chinook populations at George Adams and Voights Creek hatcheries. The pilot study included an evaluation of the fertilization success of males that were used as primary milt donors and as backup males. We present the pilot study here because our

fertilization success results are relevant to concerns about the effect of mating structure variation on parentage assignment success.

Methods

Tissue Sampling

We designed the sampling to minimize the amount of writing that we would need to do at the time of spawning. We pre-labeled our sample containers and prepared data recording forms with few blank fields. Our goal was to identify samples from dam-sire pairs that were spawned together so that we could filter parentage assignments easily to eliminate parent pairs that were not credible. Our sample identifiers were in our laboratory's standard two-part format: a four-character collection identifier with the last two digits of the year (08 in our case) followed by two alphabetic characters with no intrinsic meaning except that the four character code identified a unique collection, and a four-digit individual identification number, in the range 0001 to 9999. We assigned numbers between 0001 and 4999 to dams, and numbers between 5001 and 9999 to sires so that for a dam-sire pair, the sire identification number equaled the dam identification number + 5000. That sample identification scheme encoded the hatchery, sex, and mate for all prospective parents.

Marblemount Hatchery pilot study

We collected samples of several tissues from all individuals that were spawned. We collected axillary processes into manila coin envelopes (57 mm X 90 mm), and portions of pectoral fins, approximately 7 mm^2, into 2 mL screwcap vials from all fish in the pilot study tissue evaluation. The vials were pre-loaded with absolute ethanol and each had a pre-printed, internal label with an individual sample identifier. The axillary process envelopes for a spawned pair were bound together and pre-labeled so that the dam and sire samples were easily identifiable. In addition to assigning females and males to non-overlapping sample number series (see above), we also evaluated sampling different tissues from dams and sires to provide redundant sex identification. We used paring knives to collect dermal scrapes into 57 mm X 90 mm manila envelopes from the bodies of dams posterior to the dorsal fin and above the lateral line, and we collected a small sample of milt (generally < 1 mL) from each male into a WhirlPak® bag at the beginning of the stream as it was stripped. We dried the axillary processes and dermal scrapes overnight in an incubator at 56° C, and then stored them at room temperature until DNA extraction. We kept the milt samples on wet ice until we returned to the laboratory, then we transferred approximately 100 µL to a 2 mL screwcap vial and added ethanol to preserve the samples. We stored the ethanol-treated milt samples at room temperature until DNA extraction. The milt samples identified the donor parent as a male.

We sampled up to 100 eyed embryos from each of 92 incubation trays at Marblemount Hatchery and preserved them in 100% ethanol. The axillary processes, dermal scrapes, milt samples and embryo samples were transported to the Washington Department of Fish and Wildlife's Molecular Genetics Laboratory in Olympia, Washington for genetic analysis.

George Adams and Voights Creek Hatcheries

At George Adams and Voights Creek hatcheries we collected axillary processes from the dams and sires into pre-labeled manila coin envelopes (57mm X 90mm) during spawning when the gametes were extruded into the buckets. We pre-assembled the envelopes in pairs to physically link the samples taken from mated adults, and eliminate the need for labeling the envelopes at the time of sample collection or recording the mated pair identities. After the gametes were mixed and allowed to interact for at least 60 sec, five (George Adams) or ten (Voights Creek) fertilized egg lots were combined into a common bucket. A designated scribe recorded which single-pair gamete lots were combined into larger gamete pools. The paired axillary process samples were dried overnight in a 37° C incubator and then stored at room temperature with a desiccant until DNA extraction.

In addition to systematically collecting paired axillary processes from each mated pair, we also collected pooled male and pooled female samples (bulk samples) each spawning day. The bulk samples represent a trade-off between logistic simplicity and loss of information about spawning structure. After spawning, carcasses of the spawned fish were deposited in totes where they accumulated prior to CWT screening. We clipped the remaining axillary process from each carcass (we took the first axillary processes when the fish were spawned) and placed it in a one-liter vessel with at least a four-fold volumetric excess of absolute ethanol. Axillary processes from females and males were segregated into separately labeled containers.

To represent the post-hatching 2008 Chinook cohort at George Adams Hatchery, we collected into jars containing 100% ethanol, adipose fins that were excised from juvenile Chinook during mass marking in Spring 2009. We extracted DNA from a subset of the fins to verify that we recovered sufficient high-quality DNA to run genotyping assays in the laboratory. All Chinook production from the 2008 eggtake at Voights Creek Hatchery was lost during catastrophic flooding in December 2008 so no juvenile samples other than our eyed egg samples were collected there. The flood loss did not affect our modeling of the Voights Creek population.

Molecular methods

We extracted genomic DNA from the genetic tissue samples with Nucleospin 96 Tissue kits (Machery-Nagel) following the manufacturer's instructions. Embryos were washed in sterile, de-ionized water before DNA extraction to remove exogenous DNA and surface contaminants. The washed embryos were cut in half to expose the embryonic tissue and then processed following the same protocol as the adult tissues. We eluted 100 µL volumes of purified DNA for use in genotyping assays. We knew from our experience with the microsatellite loci used in this study

that we did not need to normalize the DNA extracts to amplify these loci using the polymerase chain reaction (PCR).

We configured six PCR multiplexes (Wenburg et al. 1996) to genotype the broodstock and embryos with the 13 GAPS microsatellite loci (Seeb et al. 2007) (Table 1) and an additional highly polymorphic locus, *Ssa-197*. One PCR primer in each pair was synthesized with an 18 – 20 oligonucleotide 5' tail that enabled incorporation of oligonucleotide dye-vectors in PCR amplicons (Schuelke 2000). Each 5 µL reaction cocktail included 1 µL Promega Go-Taq 5x buffer, 0.5 µL 200 µM dNTP mix, and 0.3 µL 25 mM $MgCl_2$. Promega GoTaq DNA polymerase, PCR primer, and oligonucleotide dye-vector concentrations varied among the multiplexes (available on request). Each reaction contained 1 µL non-normalized genomic DNA template and sterile, de-ionized water to bring the final volume to 5 µL. Thermal cycling conditions included an initial denaturation at 94° for 2 min followed by 3 cycles of 94° C for 30 sec, 60° C for 30 sec, and 72° C for 1 min, and then 36 cycles of 94° C for 30 sec, 50° C for 30 sec, and 72° C for 1 min. After cycling the reactions were held at 72° C for 10 min and then were cooled to 10° C.

Genotyping

Following PCR, the reaction products lengths were estimated by electrophoresis in an Applied Biosystems 3730 DNA Analyzer. The samples were prepared according to the instrument protocols and electophoresed through POP6 polymer in a 50 cm 96-capillary array. The loading cocktail included GeneScan-500 Rox size standard to facilitate accurate PCR amplicon length estimates. Allele identities were assigned with Applied Biosystems' Genemapper v 3.7 software, and multilocus genotypes were compiled in Microsoft Excel 2007.

Genotyping of the Marblemount parents and embryos was done in separate stages. First we screened the parents to identify which mating triads (female x primary and backup males) would yield definitive evidence of paternity with the embryo genotypes. We organized the genotypes of all potential parents in the broodstock into the 156 female - primary male - backup male triads established at fertilization. Genotyping failures reduced the numbers of genotypes we could compare within parental triads; we eliminated triads that did not have at least nine loci in common among the dam and primary and backup males. We ranked the remaining triads according to the numbers of loci at which the primary and backup males shared no alleles. In genotyping stage two, we selected the 58 highest ranked triads and subsampled 20 embryos from each for paternity analyses. We genotyped the embryos following the same protocol we used for the adults.

Data analysis

Relatedness and allele sharing within and among hatcheries

Because genetic estimation of relatedness and parentage analyses are based on expected patterns of genotypic similarities between individuals, it is important to consider the potential effects of inherent similarities of the parental genotypes due to recent common ancestry at the stock and family levels (Olsen et al. 2001, Wang 2008).

The widespread culture of the Green River fall Chinook within the Puget Sound ESU and the resulting genetic similarities between hatchery lines might result in high levels of allele sharing among individuals that return to spawn in different hatcheries. Allele sharing had not been quantified previously for full year-classes of parents in Puget Sound Chinook hatcheries. We used PowerMarker v.3.25 (Liu and Muse 2005) to calculate allele sharing distances within and among the Marblemount, George Adams, and Voights Creek hatcheries.

Simulations

We conducted most of the analyses using simulated progeny generated from the genotypes of adults that we sampled at the three hatcheries in fall 2008. The simulations were performed using a custom program written in VisualBasic 2008 (Microsoft Corportation, Redmond, Washington) and the results were summarized using custom Perl scripts. Every simulated family had a spawning location identifier, a batch identifier that specified a subset of spawners at the location (by spawning day, carcass tote, or specific time interval) and parent pair associated with it in an input file for the model. Those identifiers allowed us to test for losses in assignment accuracy that might be associated with decreasing amounts of information regarding credible parent pairings. The simulation program produced a log of simulated offspring that identified their parents, spawning location, and spawning batch.

We used the hatchery rack return counts from the stocks in the study as the approximate numbers of offspring to simulate so that we generated 19,422 simulated genotypes, 14,319 based on George Adams Hatchery spawners, 3,874 based on Voights Creek Hatchery spawners, and 1,229 based on Marblemount Hatchery spawners. We divided the rack counts by the numbers of female spawners to estimate the mean family size at each hatchery (7.9, 13.7, and 8.4 at the Marblemount, George Adams, and Voights Creek hatcheries, respectively) and used the Poisson function in Excel to generate a distribution of family sizes for each hatchery population. The parental genotypes and the Poisson-distributed family sizes were entered into a two-column list. The order of the parental genotypes was fixed and we permuted order of the family sizes within each hatchery to determine each family size for each simulation. Individual offspring genotypes were created by combining simulated gametes generated from two parents specified in the family size file.

Missing single locus genotypes in simulated offspring were generated by applying a generic missing data rate that was entered at run-time. Genotyping errors also were simulated from a generic genotyping error rate also entered at run-time. The program compared random numbers with the user specified genotyping error rates, rates of missing data from incomplete genotypes, and proportions of ungenotyped parents. Random numbers less than or equal to the specified rate or proportion triggered a routine to generate the appropriate modification of the data. This process inadvertently built in very low background rates of modifications because when the specified rates were zero, random zeros triggered the relevant modification subroutine (i.e. missing data, genotyping error, missing parent).

Our initial intent was to model microsatellite mutations separately from genotyping errors, but simulated microsatellite mutations based on strand-slippage could have resulted in mutant alleles that were not present in the parents' data set. A random mutation model in which any allele can change to any other observed allele likely would present a stronger challenge the parentage inference process because all mutant alleles could be confused with alleles that were observed in the parents' genotypes. Parentage analysis relies on the transmission of parental alleles to offspring so that any inter-generational change in allele identity from an allele observed in a true parent to an allele not observed in a true parent leads to a reduced likelihood of the true parent-offspring relationship. Any process that results in non-parental alleles in the offspring should have similar effects on the parent-offspring relationship likelihoods, so the effect is analogous to a genotyping error.

Simulation scenarios

We designed simulations to evaluate the reliability of parentage assignments under several plausible spawning scenarios and several common departures from ideal genotypic data:

1) *Factorial matings*. Busack and Knudsen (2007) recommended the use of factorial mating designs to increase the effective population size in conservation hatcheries. The offspring produced within a factorial group are closely related. Most of the benefit from factorial matings accrues with 5x5 or 10x10 designs (Busack and Knudsen 2007) so we simulated 5x5 and 10x10 factorial matings. Parents were divided into blocks of 5 or 10 pairs and we specified 25 parent-pairings and 100 parent-pairings respectively for the two mating structures. Numbers of offspring per pair were randomized for each of 20 iterations of the simulator.

2) *Double first cousins*. Closely related parents might pose a challenge for PBT. It is possible, but probably uncommon, for a spawning pair to have a counterpart spawning pair that consists of a full-sibling to each of the first pair. The progeny from the two pairs would be double first cousins. If the two pairs spawn in two different allocation groups separated in space or in different return years, then misallocation of those progeny would represent an erroneous assignment to either the

location or the spawning year. We simulated full-sibling offspring from the genotypes we observed at Marblemount, George Adams, and Voights Creek hatcheries, and then used those full-sibling groups to create clusters of double first cousin families. We randomized the number of offspring per pair for each of 20 iterations of the simulator. We ran these simulations with low rates (0.01 and 0.05) of excluded parents to mimic offspring production from parents from another, unsampled, population that are closely related to sampled parents.

3) *Offspring genotyping error rates*. Our analyses of parent-offspring genotypic mismatches in assignments of known-ancestry embryos from the Marblemount Hatchery spring Chinook population suggested that our true genotyping error rates were low in this study (Table 2). Likelihood-based parentage assignments are designed partly to provide robust inferences even when genotypes are not known without error. We simulated offspring genotyping error rates in which 5, 10, 15 and 20 percent of single-locus genotypes were inconsistent with the genotypes of the parent pair that produced them to assess the reliability of parentage inferences with higher genotyping error rates than we would expect to encounter.

4) *Incomplete offspring genotypes*. Partial genotypes are among the normal deviations from ideal genotypic data. Our enumeration of missing data among the George Adams Hatchery and Voights Creek Hatchery adults were all below 0.02, but we simulated missing data at higher rates 0.05, 0.10, 0.15 and 0.20 to assess the performance of PBT in cases with poor tissue quality. We also simulated incomplete parental genotypes with similar amounts of missing data but the relationship likelihoods depend on the number of compared loci within putative parent-offspring relationships so for a given putative relationship it is not important whether it is the parent or offspring genotype that is incomplete.

5) *Missing parents*. We sampled the entire spawning populations at Marblemount, George Adams and Voights Creek hatcheries in 2008, but it is important to assess how PBT can be expected to perform when parents are missing from the reference data set. Also, genetics based parentage analyses are conducted on natural populations to assess the breeding ecology of those populations, and often the proportion of parents sampled is not known. We simulated a broad range of missing parent proportions 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, and 0.90 to evaluate the effect missing parents have on these analyses.

Parentage inference

In our assessments of parent-offspring assignment accuracy, we recognized five general classes of results: 1) parent-offspring trios assigned and the relationship was true, 2) parent-offspring trios assigned and the relationship was false, 3) single parent-offspring duo assigned and the relationship was true, 4) single-parent-offspring duo assigned and the relationship was false, and 5) no relationship assigned. We defined the reliability of assignments in a particular simulation scenario as the proportion of relationships that were assigned and were true divided by the total number of relationships that were assigned.

Parentage inference for pilot study embryos. We examined the pilot study adult genotypes in each dam – primary male – backup male triad to identify the triads with the fewest shared alleles among the parents. We refer to loci at which the parents share no alleles as definitive loci. Based on those comparisons, we selected 58 families and genotyped 20 embryos from each of those families. We inferred the paternity of embryos from the pilot study by maximum likelihood with the software FRANz v 2 (Riester et al. 2009). Paternity was assigned only when definitive loci were typed in both primary and backup males.

We did not restrict the initial dam-sire assignments for the Marblemount embryos to known mating triads, and we classified the inferred paternity as by primary male, backup male, or other male. Assignments to the 'other male' category were considered erroneous and provided clues to mis-assignment rates and causes (e.g., genotyping errors, sample mislabeling). When clearly erroneous paternity assignments were made, we referred to the spawning documents and the multilocus genotypes of the parents and offspring to infer the causes of the mis-assignments and to correct them.

We tallied primary and backup male paternity in the Marblemount production within each full- or half-sibling family. We calculated the experiment-wide paternity by primary and backup males to assess the overall effectiveness of the two classes of fertilizers. We also identified males that were used as both primary and backup males within the embryo batches that we analyzed and we assessed the cumulative reproductive success for those males as sums of their primary and backup paternities. We reasoned that if differences in male fertility caused variability in their reproductive success then males with above average fertility would have above average success in both the primary and backup roles. We calculated the correlation coefficient between primary and backup male reproductive success to assess the consistency of the males' reproductive performance.

We compared our observed tallies with simulated data distributions to estimate p-values for alternative hypothetical male success scenarios. We assessed the statistical significance of our results by simulating 9,999 hypothetical distributions of paternity by primary and backup males in 58 families. We drew numbers from a random uniform distribution and assigned values that were less than or equal to threshold values of 0.5 or 0.1 as hypothetical backup male paternity.

The 0.5 threshold value simulated the distribution of backup male paternity when fertilization was equally likely by primary and backup males. The 0.1 threshold simulated the distribution of paternity by backup males when primary males accounted for 90% of the paternity. We determined the rank-position of each observed count of backup paternity among the simulated family counts and divided the ranks by 10,000 to determine the quantiles within the simulated distributions that corresponded to our observations. We summed the offspring counts assigned to each male in the primary and backup male roles to estimate the relative reproductive output of each male.

Parentage inference with simulated data: We used FRANz v.2.0 to infer parent-offspring relationships in the simulated populations. The FRANz parentage assignment output files were then processed in the VisualBasic model to add columns with the hatchery, spawning batch, and parent identities that were used to generate each simulated offspring. The FRANz output files included all putative parent-offspring relationships with positive LODs so that there were some offspring with several plausible sets of parents. Those processed parentage assignment files were then filtered with custom Perl scripts that screened out putative parent pairs with conflicting places or times of spawning. We filtered the putative parent pairs to examine the reliability of assignments when parents were evaluated at two levels of aggregation, 1) parents that were spawned in the same place or hatchery, and 2) parents that were spawned in the same batch within a location or hatchery. We did a preliminary assessment of limiting the putative parents to only pairs that were spawned but those data will not be presented here.

Results and Discussion

Spawning protocol

We did not control the gamete collection and fertilization at the hatcheries but the protocols used exemplify some commonly expressed concerns about realized mating structures in hatcheries. All three hatcheries in this study practiced variations on a single-pair mating protocol. The eggs from a single dam were combined with the milt from a single sire and allowed to comingle undisturbed for at least one minute. Single-pair matings were then combined intomulti-pair pools (see below). The pooling addressed two aquacultural concerns: 1) it provided eggs that might initially have been combined with non-motile sperm access to second sperm source, and 2) the pooled, fertilized eggs could be distributed to incubation trays in a way that uses incubation stack space efficiently.

Spawning

<u>Marblemount Hatchery pilot study</u>

Three hundred and twelve spring-run Chinook salmon adults were spawned at Marblemount Salmon Hatchery between 5 and 26 August, 2008. Individual egg lots from 156 females were spawned into separate buckets that were labeled with the females' identities. Milt was stripped from 156 males into two plastic bags each, one to be used as a primary sperm source in fertilization and the other as a backup sperm source. The bags were stored on wet ice for less than two hours before their contents were used to fertilize eggs. Each primary aliquot of milt was added to an unfertilized egg bucket and the combined gametes were mixed gently by hand. The mixed gametes were allowed to interact for at least 60 sec before the second milt aliquot from a different male was added and the gametes were mixed again. The fertilized eggs were immersed in an iodophor solution to water harden them and kill surface pathogens, then each lot was transferred to a separate incubation tray that was loaded into a vertical incubation stack with a constant flow of water. All incubation trays had a fine mesh cover that prevented the passive transfer of embryos into or out of the trays. The identities of the dam, primary male and backup male were recorded on each incubation tray. The embryos were allowed to develop undisturbed for approximately one month until they reached the eyed stage.

<u>George Adams and Voights Creek Hatcheries</u>

Two thousand one hundred and fifty-nine fall-run Chinook were spawned at George Adams Hatchery on six days between 2 and 25 September and 953 were spawned at Voights Creek Hatchery on three days between 16 and 30 September in 2008 (Table 3). Eggs from a single female were stripped into a spawning bucket and the milt from a single male was stripped directly onto the eggs in the bucket. The gametes were mixed gently and allowed to interact for at least 60 sec. The single-pair egg lots then were combined into either five-pair pools (George Adams Hatchery) or ten-pair pools (Voights Creek Hatchery) before iodophor treatment and water hardening.

Tissue collection methods

The screw cap vials with pre-loaded absolute ethanol required more dexterity than the manila envelopes to open, fill, and close, so they were dropped from the sampling procedure and the envelopes became the primary tissue receptacles. The dermal scrapes were more difficult to work with in the laboratory than the axillary processes so we abandoned the scrapes in favor of the axillary processes. The milt samples were easy to collect at the pace of spawning in the Marblemount Hatchery but might have been difficult at the production levels in the other two hatcheries. Also laboratory processing of the milt was difficult because we did not have an optimized extraction protocol. Extraction of DNA from 1 μL of milt produced a large pellet of precipitate and clogged the silica membranes in the extraction columns. Work flow in the laboratory is smoother when following a single set of protocols so we abandoned the milt

collection in favor of axillary processes. We realize that there is a trade-off between absolute sex identification and lab workflow and we could reconsider our decision in the future.

Eggs in six incubation trays from Marblemount failed to develop and embryos from two were damaged in transit from the hatchery to the laboratory. The embryo samples from the remaining 84 parental triads were cataloged and available for analysis. Sixty-two of the 84 parental triads had primary and backup male genotypes that shared no alleles at eight or more loci and therefore would be definitive evidence of paternity. We selected embryos from the 58 triads with the most definitive male genotypes for parentage analyses. We processed 1,135 embryos and determined genotypes for 1,121 at 2 or more loci that unequivocally identified the sire. The genotypes of the remaining 14 embryos were considered unreliable and were excluded from the parentage inference.

Sampling effort

The sampling effort exceeded what was necessary because we did not want to disrupt the normal pace of spawning activities at either facility. The odd numbers reflect samples taken from back-up males in a few instances. We also collected daily bulk samples from all broodstock spawned for each eggtake. The bulk samples were taken from post-spawning carcasses that were being held in totes prior to coded-wire tag screening. The bulk samples were collected by two samplers per day.

Genotyping

We genotyped 2,159 parents from George Adams Hatchery at 14 microsatellite loci and had 62 unscored loci for a failure rate of 0.0021. We had 165 unscored loci among 953 adults from Voights Creek Hatchery genotyped at 14 loci for a failure rate of 0.0124. (Table 3). We estimated the minimum genotyping error rate from the counts of genotypic incompatibilities between genotyped embryos from Marblemount Hatchery and their known parents (Table 4). The estimated genotyping error rates were 0.024 among the adults and 0.002 among the embryos.

Thirteen of the 14 microsatellite loci we used had more than 10 discrete alleles in the animals that we genotyped for this study and 9 loci had 25 or more discrete alleles (Table 1). The expected heterozygosity was above 0.8 for 10 of the 14 loci suggesting that they should provide good individual resolution.

Relatedness and allele sharing

Allele-sharing was highest among George Adams and Voights Creek hatcheries where both stocks were founded from the Green River fall Chinook lineage (Table 4). The allele sharing distance between George Adams Hatchery and Voights Creek Hatchery is similar to the allele sharing values within the two hatcheries.

Parentage assignments in simulated populations

Factorial matings

The proportions of the simulated sets of 5x5 and 10x10 factorial matings that assigned to the correct parent pairs were similar with median proportions from 20 iterations at 0.987 for each design (Table 5). The factorial designs create clusters of half-siblings so the high proportions of correct assignments to parent pairs demonstrates that the microsatellite data set has ample power to assign close relatives to the correct parents. The assignment success was slightly improved when the assignments were filtered to exclude parents that were not in the same batch. Mis-assignments between the relationships within a factorial mating matrix would not result in allocation to the wrong population or cohort, so even at higher rates than we observed in our simulations, they would not confound PBT.

Double-first cousins

We simulated populations that were predominantly composed of double first-cousins. We simulated two scenarios in which spawning pairs were withheld from the reference data set to mimic offspring straying into a PBT population. The median rate of correct assignments to parent pairs with 1% of parents withheld was 0.976 when parent pair filtering was done at the coarse level of requiring the pair to have spawned in the same place (hatchery), and it was 0.977 when parent-pair filtering required pairs to have spawned in the same batch (Table 6). Assignment accuracy declined to 0.941 and 0.942 for place and batch filtering when 5% of parent pairs were withheld (Table 6).

It is worth noting that if 1% of the parent pairs are withheld from the reference set, the maximum attainable correct assignment rate is less than 1.0. Since family sizes were not equal in these simulations, we cannot calculate the maximum attainable rate from the data available. Consequently, the 'correct assignment' rates in Table 8 underestimate the accuracy of the process because they were calculated as the proportions of the whole populations that were assigned correctly without adjusting for unassigned individuals who were correctly not assigned to any parents in the reference data set.

Offspring genotype errors

Offspring genotyping errors simulated at 0.05 and 0.10 did not depress parent-offspring assignment accuracy over 20 iterated populations (Table 7). It might seem counterintuitive that a genotyping error rate as high as 0.10 did not result in poor assignment accuracy, however, the polymorphic information content (PIC) values for most of the loci in this microsatellite data set are above 0.8 and it is likely that random errors in genotypes would decrease the apparent likelihoods of valid parent-offspring relationships but would rarely increase the apparent likelihoods of invalid parent-offspring relationships. The parentage analysis software that we used provides maximum likelihood parentage assignments and the robustness of the assignments

with simulated genotyping errors suggests that the PIC is great enough to overcome the noise that random genotyping errors introduce. We note that genotyping errors at the frequencies that we simulated here would not pass QA/QC screening in our laboratory.

Incomplete offspring genotypes

We simulated populations with missing data ranging from 0.005 to 0.10 to assess the performance of parent-offspring inferences in the presence of incomplete data. The amounts of missing data at the low end of the range are near what we observed in the parent genotypes from George Adams Hatchery and Voights Creek Hatchery. Median assignment accuracy was above 0.99 at all simulated levels of missing offspring data (Table 8). This result suggests that the 14 microsatellite loci included in this study have more than enough information to assign parent-offspring relationships accurately.

Parent genotyping errors

Conceptually, genotyping errors in parents should have a larger effect on parentage assignments than errors in offspring genotypes because they can cause apparent genotypic incompatibilities with multiple offspring. The median correct assignment proportions were above 0.97 for error rates of 0.02 and 0.05 (Table 9). As we discussed above in the section on offspring genotype errors, this suggests again that the suite of microsatellites contains enough information to overcome the noise from the genotyping errors.

Incomplete parent sampling

Our assessment of parent-offspring relationship assignment accuracy across a range of proportions of missing parent genotypes differs from the simulations that included withheld pairs of parents of double first cousins in that in this scenario, parents of simulated offspring were excluded at random from the reference data set. Offspring in this scenario can have 0, 1 or 2 parents in the reference data set. It is not surprising that assignment success is low when the proportions of true parents that are missing from the reference data set are high (Table 10). In fact, with 10% of parents missing, assignment accuracy was ~90% but it declined to ~60% with 50% of the parents missing. It is sobering to us to realize that parentage inferences made in natural salmonid populations are often done without knowledge of the proportion of parents that are missing.

Marblemount Hatchery embryo paternity

Primary males accounted for an overwhelming majority of the fertilizations in our analysis of the Marblemount Hatchery embryos (Table 2). One thousand and forty-three of 1,121 embryos (93%) were assigned to the primary males and 78 embryos were assigned to the backup male. All of the paternity in 32 of the 58 egg lots (55%), and at least 90% of the paternity in 14 other egg lots (24%) was by the primary males (Table 11).

Our paternity assignments revealed that male reproductive success within a parental triad depended on his role as either a primary or backup male. Sperm from primary males had uncontested access to unfertilized ova for at least one minute. Twenty-nine males were used in both primary and backup roles in families that we genotyped (Table 12). They produced 556 embryos, 515 of which (93%) were produced by the primary male. The reproductive success of those 29 males in the primary and backup roles was negatively correlated ($r = -0.18$), high success as a primary male did not mean high success as a backup male, which suggests that having uncontested access to unfertilized ova gave primary males substantial reproductive advantage over backup males. The vast majority of eggs apparently were fertilized before the backup males' milt was introduced. That first-in advantage overwhelmed any intrinsic differences in fertility that might have existed between males.

In the analyses of three egg lots, primary males were not assigned the majority of paternity. In two cases, it is likely that sample labels were swapped after spawning and before the samples were processed in the laboratory (see below). In the third case, M-057 was the most successful male in the backup role, he sired all 20 embryos in egg lot F-058, but he was the second least successful male in the primary role where he sired 14 of the 19 embryos (74%) that were assigned from the F-057 egg lot. Only M-058 was less successful in the primary role. This result was an anomaly in our study and we cannot discount the possibility that it was due to an experimental process error rather than to differenences in male fertility. The most likely experimental source of error is that the backup milt aliquot of M-057 might have been swapped mistakenly with the primary milt aliquot of M-058 when the F-058 egg lot was fertilized. In this scenario, M-057 would have been the primary male used in the F-057 and F-058 egg lots and M-058 would have been the backup male used in the F-058 and F-059 egg lots. It also is possible that M-058 was the primary male for F-058 but was ineffectual in that role. It is important to note, however, that our paternity assignments show that M-058 was not infertile. He was the backup male for egg lot F-059 and he sired one of the 19 F-059 embryos. Our data do not allow us to unequivocally resolve this anomaly.

Sample labeling errors

Paternity assignments for embryo samples F-003 and F-004 are inconsistent with spawning records and suggest misidentification of two adult male samples in the laboratory. Spawning records indicated that males M-003 and M-002 were the primary and backup males respectively

for egg lot F-003, and males M-004 and M-003 were the primary and backup males for egg lot F-004. Our genotypes for males M-003 and M-004 suggested that male M-004 sired all of the F-003 embryos that we genotyped and male M-003 sired all but one of the F-004 embryos. The inferred paternity of F-003 embryos by M-004 is at odds with the spawning records that indicated his milt was not applied to that egg lot. The fertilization of the majority of the F-004 embryos by M-003 in a backup role was possible according to the spawning records but it stands in stark contrast to the prevailing dominance of primary males in this study. We assert that the identities of males M-003 and M-004 likely were switched sometime after spawning so that the primary male for F-003 was labeled in our analyses as M-004 rather than as M-003, and the primary male for F-004 was labeled as M-003 rather than as M-004.

Statistical significance

Paternity by backup males in our study was much rarer than we would expect if primary and backup males had equal fertilization opportunities and equal fertilities. Paternity assignments to backup males in 53 of 58 embryo samples were in the lowest 0.1 % of our simulations of randomly distributed paternity among primary and backup males.

Our analyses show that backup males in the Marblemount Salmon Hatchery's spring Chinook program play a minor role in the genetic dynamics in the population. The low cumulative paternity that we detected by backup males suggests that their use causes little variance in male reproductive success. Our estimates are based on at most 20 embryos from each female so their sensitivity to low frequency contributions is limited, but the pattern of low variance in backup male paternity across 58 replicate samples makes it unlikely that major fluctuations in male reproductive success were commonplace.

Primary males enjoyed a substantial advantage in reproductive success. The consistent pattern that we observed across replicated matings suggests that uncontested access to unfertilized ova contributed greatly to the primary males' high fertilization rates, and that the low rates for backup males were not likely due to intrinsically lower quality gametes or occasional poor mixing of the combined gametes after the addition of the backup milt. The pattern we observed is contrary to expectations if sperm are inactive in ovarian fluid and require water to induce motility. Our result is concordant with previous experiments that showed sperm activation and fertilization could occur in ovarian fluid without the addition of water (Rucker et al. 1960).

We do not know why our Marblemount paternity results show a strong first-in advantage for primary males in contrast to the earlier studies by Gharrett and Shirley (1985) and Withler and Beacham (1994). Those studies carried out fertilizations in more controlled settings than in our study. Perhaps the extra handling that was necessary in those controlled settings diminished the advantage of prolonged, uncontested access to ova that the first aliquots enjoyed in our study. We made no attempt to estimate the effects of varying the interval between the primary and backup fertilizations because our purpose was to evaluate the standard spawning protocol in the

hatchery with respect to its effect on the mating structure. Withler and Beacham (1994) found that variance in sperm quality among males in pooled sperm fertilization scenarios decreased if the mixed milt was held for an hour before use in fertilization.

Recent investigations have shown that social interactions among spawning fish with external fertilization can induce changes in sperm characteristics. Several studies have shown that parameters of sperm quality such as swimming speed, directed swimming, and longevity vary with a male's position in the social hierarchy in Arctic charr (*Salvelinus alpines*) (Rudolfsen et al. 2006) and in the cichlid, *Telmatochromis vittatus* (Ota et al. 2010). Those studies also showed that sperm characteristics can shift quickly when a male's social status changes so that his sperm characteristics are appropriate to his new status. The effects of hatchery induced stresses on gamete quality have not been studied extensively, but in one study rainbow trout (*O. mykiss*) held in small tanks showed a four-fold increase in cortisol levels and a greater than 90% reduction in sperm motility (Bobe and Labbe 2010). The crowding and handling that precede gamete stripping in salmon hatcheries might similarly induce physiological responses to stress, and those responses might be expressed differently in individual fish. If the biochemical response to stress becomes attenuated as time elapses, then the interval between the induction of stress and fertilization trials might influence the outcome of those trials. It could be the case that the fertilization trials performed in previous studies of sperm competition in salmon hatcheries were done with shorter elapsed times between milt stripping from the males and sperm addition to the egg lots than in our study.

This paternity analysis was limited to estimating the effects of backup male use on genetic diversity during early embryonic development in the hatchery. Chinook salmon fecundities in this stock average over 4000 ova per female. Post-hatching factors cause the majority of mortalities in hatchery produced salmonids and differential effects among families, some deterministic and some stochastic, likely cause population variance in family sizes. The relatively small pre-hatching variance that we observed in our study could easily be overwhelmed by modest post-hatching factors.

Conclusions

The proof of concept study carried out at George Adams and Voights Creek hatcheries have provided information on potential error sources and possible solutions at several process steps. Our results indicated that successful assignment of parent-offspring relationships using this set of microsatellites, and therefore successfully PBT results can be obtained even when (1) individuals from hatcheries that share a recent common ancestry are included in the parent-pool, (2) genotypic data are incomplete or some loci are in error, and (3) full siblings spawning in different allocation strata exist, even to the extent where full sibling spawning pairs in the two strata produced double first cousins .

The data and analyses presented here from the Marblemount pilot study provided an assessment of the use of back-up males in hatchery spawning protocols in addition to providing an assessment of logistical concerns that needed to be addressed prior to initiating PBT at a large hatchery. Our parentage assignments for the embryos showed that the primary + backup male fertilization protocol that includes an interval of at least 60 sec between the addition of the primary and backup milt aliquots did not result in skewed rates of paternity or large complements of half-siblings among the offspring. That finding suggests that fertilization occurs quickly after sperm are added to the ova and that fertilization can proceed in the presence of ovarian fluid before water is added to the combined gametes. We also demonstrated that the use of backup males in fertilization can be an effective strategy to prevent egg wastage in a hatchery program without posing a substantial risk to the genetic diversity of the hatchery program. In fact, the slight increases in the variance in male reproductive output that we observed from low-level fertilization by backup males must be evaluated against an alternative large increase in variance in female reproductive output that can occur if some egg lots are combined only with non-motile sperm and are lost completely: variance in female reproductive success is also important to genetic diversity. The fertilization dynamics that we infer from these results also suggest that the pooling of combined gametes at least 60 sec after fertilization into 5-pair or 10-pair egglots at George Adams and Voights Creek hatcheries respectively do not likely reduce the genetic diversity in those populations through sperm competition.

Based on these data, PBT can be a viable method of determining stock- and cohort-of-origin even in when the genotypic data sets include minor imperfections. We note though that parentage assignments in situations where substantial proportions of spawners are not sampled must be interpreted with caution.

Acknowledgements

The main body of this work was supported by the PSC Southern Boundary Restoration and Enhancement Fund. The Washington State General Fund supported the pilot study at Marblemount Hatchery and provided the laboratory and administrative support. Craig Busack was instrumental in generating the conceptual framework, study design and proposal. Norm Switzler provided valuable assistance from the onset of the planning and logistics phase through the sampling and laboratory preparation of sample extracts. Cherril Bowman performed almost all of the molecular laboratory work downstream of DNA purification and helped with the sampling during the pilot study. The tissue collections relied heavily on the support and cooperation of the WDFW Hatcheries Division personnel including Ron Warren; Hatchery Complex Managers Brodie Antipa, Rich Eltrich, Ted Thygesen; and the fish hatchery staffs including Steve Stout, Gary Schooland and crew at Marblemount Hatchery; Ed Jouper, Dan Minton and crew at George Adams Hatchery; and Jill Phillips, Sherman Davis and crew at Voights Creek Hatchery. Many people contributed to the success of the sampling including

citizen volunteers Susan Dean and John Dean, and WDFW staff members Scott Blankenship, Sean Carrell, Cheryl Dean, Alisa Harjo, Kristina Iverson, Todd Kassler, Lucinda Morrow, Shelley Peterson, Jim Pruske, Jennifer Shefler, Maureen Small, and Jennifer Von Bargen.

Literature cited

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

Augerot, X. and D. N. Foley. 2005. Atlas of Pacific Salmon: the first map-based status assessment of salmon in the North Pacific. University of California Press, Berkeley. 150 p.

Bobe, J., and C. Labbe. 2010. Egg and sperm quality in fish. General and Comparative Endocrinology 165(3):535-548.

Busack, C. and C. M. Knudsen. 2007. Using factorial mating designs to increase the effective number of breeders in fish hatcheries. Aquaculture 273: 24-32.

Expert Panel on the Future of the Coded Wire Tag Program for Pacific Salmon. 2005. Report. PacificSalmon Comm. Tech. Rep. No. 18: 230 p.

Gharrett, A.J., Shirley, S.M., 1985. A genetic examination of spawning methodology in a salmon hatchery. Aquaculture 47, 245-256.

Liu, K and S. V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2128~2129

Nehlsen, W., J. E. Williams, and J. A. Lichatowich. 1991. Pacific salmon at the crossroads: stocks at risk from California, Oregon, Idaho, and Washington. Fisheries 16(2): 4–21.

Olsen, J. B., C. Busack, J. Britt, and P. Bentzen. 2001. The aunt and uncle effect: an empirical evaluation of the confounding influence of full sibs of parents on pedigree reconstruction. Journal of Heredity 92: 243-247.

Ota, K., D. Heg, M. Hori, and M. Kohda. 2010. Sperm phenotypic plasticity in a cichlid: a territorial male's counterstrategy to spawning takeover. Behavioral Ecology 21(6):1293-1300.

Pella, J. and M. Masuda. 2001. Bayesian methods foranalysis of stock mixtures from genetic characters. Fishery Bulletin 99: 151–167.

Riester, M., P. F. Stadler, and K. Klemm. 2009. FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics 25(16):2134-2139.

Rucker, R. R., J. F. Conrad, and C. W. Dickeson. 1960. Ovarian Fluid: Its Role in Fertilization. The Progressive Fish-Culturist 22(2):77-78.

Rudolfsen, G., L. Figenschou, I. Folstad, H. Tveiten, and M. Figenschou. 2006. Rapid adjustments of sperm characteristics in relation to social status. Proceedings of the Royal Society B-Biological Sciences 273(1584):325-332.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18(2):233-234.

Seeb, L. W., A. Antonovich, M. A. Banks, T. D. Beacham, M. R. Bellinger, S. M.Blankenship, . R. Campbell, N. A.Decovich, J. C. Garza, C.M. Guthrie III, T. A. Lundrigan, P. Moran, S. R. Narum, J. J. Stephenson, K. J. Supernault, D. J. Teel, W. D. Templin, J. K.Wenburg, S. F. Young, and C. T. Smith. 2007. Development of a Standardized DNA Database for Chinook Salmon. Fisheries 32 (11): 540-552.

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.

Wang, J. 2007. Parentage and sibship exclusions: higher statistical power with more family members. Heredity. 99: 205-217.

Waples, R. S., A. E. Punt, and J. E. Cope. 2008. Integrating genetic data into management of marine resources: how can we do it better? Fish and Fisheries 9: 423-449.

Wenburg, J. K., J. B. Olsen, and P. Bentzen. 1996. Multiplexed systems of microsatellites for genetic analysis in coastal cutthroat trout (Oncorhynchus clarki clarki) and steelhead (Oncorhynchus mykiss). Molecular Marine Biology and Biotechnology 5(4):273-283.

Withler, R.E., 1988. Genetic consequences of fertilizing Chinook salmon (Oncorhynchus-tshawytscha) eggs with pooled milt. Aquaculture 68, 15-25.

Withler, R.E., Beacham, T.D., 1994. Genetic consequences of the simultaneous or sequential addition of semen from multiple males during hatchery spawning of Chinook salmon (Oncorhynchus-tshawytscha). Aquaculture 126, 11-23.

Table 1. Characteristics of the microsatellite loci used in this study that relate to parentage inference. Observed (H_{obs}) and expected (H_{exp}) heterozygosities are from FRANz v 2 (Riester et al. 2009).

Locus	Alleles	H_{obs}	H_{exp}	Exclusion probabilities			Probability of Identity		Missing genotypes - count (proportion)	
				1st parent	2nd parent	Parent pair	Unrelated	Siblings	George Adams	Voights Creek
Ogo-2	18	0.646	0.648	0.267	0.455	0.670	0.148	0.463	3 (0.001)	8 (0.008)
Ogo-4	15	0.794	0.805	0.448	0.622	0.805	0.064	0.364	5 (0.002)	14 (0.015)
Oki-100	31	0.936	0.946	0.803	0.890	0.979	0.006	0.279	7 (0.003)	15 (0.016)
Omm-1080	41	0.961	0.961	0.855	0.922	0.989	0.003	0.270	4 (0.002)	15 (0.016)
Ots-201b	37	0.933	0.938	0.780	0.876	0.974	0.007	0.283	5 (0.002)	13 (0.014)
Ots-208b	34	0.939	0.948	0.810	0.895	0.981	0.005	0.277	8 (0.004)	13 (0.014)
Ots-211	25	0.913	0.923	0.735	0.847	0.963	0.011	0.291	6 (0.003)	14 (0.015)
Ots-212	27	0.858	0.874	0.612	0.760	0.918	0.025	0.319	3 (0.001)	10 (0.010)
Ots-213	41	0.930	0.936	0.772	0.871	0.972	0.008	0.284	5 (0.002)	9 (0.009)
Ots-3M	11	0.686	0.697	0.310	0.497	0.705	0.119	0.431	4 (0.002)	11 (0.012)
Ots-9	7	0.410	0.425	0.093	0.221	0.358	0.374	0.631	3 (0.001)	11 (0.012)
Ots-G474	12	0.719	0.730	0.332	0.507	0.698	0.114	0.414	5 (0.002)	10 (0.010)
Ssa-197	35	0.933	0.941	0.788	0.881	0.976	0.007	0.281	1 (0.000)	9 (0.009)
Ssa-408	46	0.861	0.867	0.588	0.742	0.903	0.030	0.324	3 (0.001)	13 (0.014)

Table 2. Genotypic incompatibilities within known parent-offspring trios reveal genotyping errors. Overall error rates were well below 5%, so the occurrence of multiple offspring with incompatibilities within a few families suggests that one or both of the parents' genotypes has an error. Genotyping errors in one or two individuals in a family of approximately 20 siblings likely are due to errors in the offspring genotype.

	Minimum Genotyping	
	Error	**Error Rate**
Parents	93	0.024
Offspring	23	0.002

Table 3. Numbers of Chinook salmon spawned and sampler staffing at George Adams Hatchery and Voights Creek Hatchery during 2008. We deployed sampling crews large enough to capture spawning information at the highest possible resolution, spawning pairs.

Date	George Adams Hatchery		Voights Creek Hatchery	
	Number spawned	Samplers	Number spawned	Samplers
2-Sep-08	170	6		
5-Sep-08	1004	5		
9-Sep-08				
11-Sep-08	484	6		
16-Sep-08	180	5	192	7
18-Sep-08	151	6		
23-Sep-08			547	8
25-Sep-08	170	5		
30-Sep-08			214	11
Totals	2159	33	953	26

Table 4. Average allele sharing proportions in pairwise comparisons of genotypes of Chinook spawned in 2008 at Marblemount, George Adams, and Voights Creek hatcheries. The proportion of shared alleles between pairs of George Adams and Voights Creek Chinook are close to the pairwise sharing within the George Adams population.

	Marblemount	**George Adams**	**Voights Creek**
Marblemount	0.2289		
George Adams	0.1845	0.2488	
Voights Creek	0.1845	0.2434	0.2538

Table 5. Parentage assignment performance on simulated Chinook salmon offspring under 5x5 and 10x10 factorial mating scenarios. Twenty sets of offspring were assembled under both factorial mating designs. The assignments were executed with two levels of pre-filtering to avoid assignments to pairs of candidate parents that did not spawn in the same hatchery (Place) or during the same time within a hatchery (Batch).

Simulation	5x5 factorial matings		10x10 factorial matings	
	Place	Batch	Place	Batch
Minimum	0.985	0.986	0.985	0.987
5th percentile	0.986	0.987	0.986	0.988
Median	0.987	0.988	0.987	0.989
95th percentile	0.988	0.989	0.988	0.990
Maximum	0.989	0.990	0.988	0.991

Table 6. Parentage assignment performance on simulated groups of double first cousin Chinook salmon offspring. The simulations were designed to assess the potential for misallocation to population aggregates that contain closely related parents. Full-sibling parents could spawn in different allocation groups if some families split between groups when they return to spawn, by straying to a neighboring population or by returning at different ages. A parent generation of full-sibling pairs was simulated from the genotypes of Chinook sampled at Marblemount, George Adams, and Voights Creek hatcheries in August and September 2008. Pairs of simulated full-sibling genotypes were used to generate sets of simulated double-first cousins. We withheld small portions of the simulated parent pairs to mimic 1% and 5% stray rates. Note that with 1% and 5% of the parent pairs missing from the candidate parent pool, the maximum possible proportions of valid parentage assignments are 0.99 and 0.95 respectively. The assignments were executed with two levels of pre-filtering to avoid assignments to pairs of candidate parents that did not spawn in the same hatchery (Place) or during the same time within a hatchery (Batch).

Simulation	1% of parents withheld		5% of parents withheld	
	Place	Batch	Place	Batch
Minimum	0.975	0.975	0.936	0.937
5th percentile	0.975	0.976	0.938	0.938
Median	0.976	0.977	0.941	0.942
95th percentile	0.978	0.979	0.945	0.945
Maximum	0.978	0.979	0.946	0.947

Table 7. Parentage assignment performance on simulated Chinook salmon with errors in the offspring genotypes. The genotyping error rates in the simulations are well above the error rates that we estimated from the counts of mismatching loci in the assignments of embryos from known parents from Marblemount Hatchery. The rates simulated here demonstrate that likelihood-based parentage assignment provides robust estimates within the expected range of offspring genotyping errors. The assignments were executed with two levels of pre-filtering to avoid assignments to pairs of candidate parents that did not spawn in the same hatchery (Place) or during the same time within a hatchery (Batch).

Simulation	5% Error Rate		10% Error Rate	
	Place	Batch	Place	Batch
Minimum	0.993	0.999	0.992	0.999
5th percentile	0.994	0.999	0.992	0.999
Median	0.995	0.999	0.993	0.999
95th percentile	0.996	1.000	0.993	0.999
Maximum	0.996	1.000	0.994	1.000

Table 8. Parentage assignment performance on simulated Chinook salmon offspring with incomplete genotypes. Single-locus genotypes were excluded at random and might overestimate real-world performance if missing data rates differ across loci. The assignments were executed with two levels of pre-filtering to avoid assignments to pairs of candidate parents that did not spawn in the same hatchery (Place) or during the same time within a hatchery (Batch).

Simulation	Missing Data							
	0.50%		1.50%		5%		10%	
	Place	Batch	Place	Batch	Place	Batch	Place	Batch
Minimum	0.995	0.999	0.995	0.999	0.994	0.999	0.991	0.999
5th percentile	0.995	0.999	0.995	0.999	0.994	0.999	0.991	0.999
Median	0.996	0.999	0.996	0.999	0.995	0.999	0.992	0.999
95th percentile	0.996	1.000	0.997	1.000	0.995	1.000	0.993	0.999
Maximum	0.997	1.000	0.997	1.000	0.996	1.000	0.993	1.000

Table 9. Parentage assignment performance on simulated Chinook salmon with errors in the parents' genotypes. Errors in parental genotypes potentially can affect the parent-offspring relationship likelihoods of multiple offspring within a family. The error rates simulated here are similar to our estimates from the parent-offspring mismatches that we attributed to likely parent genotyping errors in our assignments of Marblemount embryos to known parents. The results here suggest that the likelihood-based assignments performed by the software FRANz can be robust to parental genotyping errors within the range of error rates that we can expect to observe and that the microsatellite data set has enough information to overcome the noise of moderately high genotyping error rates among the parents. The assignments were executed with two levels of pre-filtering to avoid assignments to pairs of candidate parents that did not spawn in the same hatchery (Place) or during the same time within a hatchery (Batch).

Simulation	2% Error Rate		5% Error Rate	
	Place	Batch	Place	Batch
Minimum	0.980	0.982	0.976	0.979
5th percentile	0.980	0.982	0.976	0.979
Median	0.983	0.985	0.978	0.981
95th percentile	0.984	0.986	0.980	0.983
Maximum	0.985	0.986	0.981	0.983

Table 10. Parentage assignment performance on simulated Chinook salmon offspring with true parents missing from the candidate parent pool. Parents were excluded at random so that simulated offspring might have 0, 1 or 2 parents in the candidate parent pool. The assignments were executed with two levels of pre-filtering to avoid assignments to pairs of candidate parents that did not spawn in the same hatchery (Place) or during the same time within a hatchery (Batch).

Simulation	Missing Parents									
	10%		20%		30%		40%		50%	
	Place	Batch	Place	Batch	Place	Batch	Place	Batch	Place	Batch
Minimum	0.898	0.932	0.805	0.864	0.727	0.797	0.645	0.718	0.569	0.636
5th percentile	0.898	0.932	0.806	0.864	0.727	0.801	0.647	0.722	0.570	0.638
Median	0.904	0.939	0.818	0.878	0.744	0.812	0.663	0.739	0.589	0.659
95th percentile	0.913	0.944	0.832	0.888	0.756	0.831	0.681	0.755	0.608	0.679
Maximun	0.914	0.945	0.835	0.888	0.757	0.832	0.684	0.756	0.609	0.681

Simulation	Missing Parents							
	60%		70%		80%		90%	
	Place	Batch	Place	Batch	Place	Batch	Place	Batch
Minimum	0.486	0.541	0.398	0.439	0.289	0.310	0.155	0.163
5th percentile	0.488	0.545	0.406	0.450	0.290	0.314	0.158	0.166
Median	0.507	0.566	0.421	0.465	0.309	0.333	0.172	0.180
95th percentile	0.521	0.585	0.433	0.477	0.322	0.345	0.186	0.196
Maximun	0.522	0.585	0.434	0.478	0.333	0.358	0.192	0.202

Table 11. The reproductive output of primary and backup males used to fertilize 58 Chinook salmon egg lots in this study. (a) The overwhelming majority of paternity was by primary males. (b) Distribution of paternity among families. Over half of the embryo lots that we analyzed were sired only by the primary males. At least 90% of the embryos were sired by the primary males in over 75% of the families that we analyzed. The paternity for two families is ambiguous and might reflect data recording errors either in the hatchery or in the laboratory.

(a)

Progeny assigned to:	Progeny count (proportion)
Primary males	1043 (0.93)
Backup males	78 (0.07)
All males	1121 (1.00)

(b)

Paternity by primary males	Count of families (proportion)
100%	32 (0.55)
≥90%	14 (0.24)
≥70%	10 (0.18)
ambiguous	2 (0.03)

Table 12. Paternity in both primary and backup fertilization roles was estimated for 29 males. The primary and backup reproductive outputs were negatively correlated, suggesting that the variances in success that we observed were due to random variation rather that intrinsic superiority of sperm from some males over others. The extreme values for M-057 and M-058 are strikingly inconsistent with our other observations and might have been caused by a mistake at the time of fertilization (see text).

Male ID	Primary successes	Backup successes	Total successes
M-003	19	0	19
M-004	20	1	21
M-005	20	1	21
M-012	18	0	18
M-013	19	0	19
M-014	18	0	18
M-020	16	0	16
M-023	20	0	20
M-024	20	0	20
M-033	16	0	16
M-034	20	0	20
M-035	17	0	17
M-048	18	0	18
M-050	17	0	17
M-057	14	20	34
M-058	0	1	1
M-059	18	2	20
M-060	16	1	17
M-068	20	4	24
M-073	20	0	20
M-074	20	0	20
M-080	20	0	20
M-090	17	0	17
M-091	20	0	20
M-092	20	1	21
M-093	18	1	19
M-094	19	6	25
M-130	18	1	19
M-149	17	2	19
Totals	515	41	556
Mean	17.76	1.41	37.07
Std Dev	3.79	3.82	98.13
Correlation coefficient	-0.18		