Measuring reproductive interaction between hatchery-origin and wild steelhead (*Oncorhynchus mykiss*) from northern Puget Sound populations potentially affected by segregated hatchery programs.

Kenneth I Warheit, Ph.D.

Washington Department of Fish and Wildlife
Molecular Genetics Laboratory
600 Capitol Way N.
Olympia, WA 90501-1091
kenneth.warheit@dfw.wa.gov

Unpublished Final Report
Washington Department of Fish and Wildlife
October 10, 2014

Preamble

Purpose of report. The purpose of this report is two-fold. First, I provide a description of and justification for the methods for assessing the potential for reproductive interaction between hatchery-origin and wild steelhead (*Oncorhynchus mykiss*) from northern Puget Sound populations. Second, I present the empirical results from these methods for the Green, Snohomish, Stillaguamish, Skagit, and Nooksack river basins. Furthermore, this report is intended as an addendum to and part of the following Hatchery Genetic Management Plans (HGMPs, 2014 versions): Soos Creek (Green River) Hatchery Winter Steelhead Program (Segregated), Snohomish Basin Winter Steelhead Hatchery Program (Segregated), Whitehorse Ponds (Stillaguamish River) Winter Steelhead Hatchery Program (Segregated), and Kendall Creek (Nooksack River) Winter Steelhead Hatchery Program (Segregated). Although there is no HGMP associated with the Skagit River, I include in this report results from the Skagit River populations to establish a genetic baseline, and for comparative purposes. This report is not intended to be a comprehensive treatment of hatchery-wild introgression or an assessment of the usefulness of any specific statistical technique beyond what I present here. The data and methods in this report will be submitted eventually as manuscripts for publication in peer-reviewed journals.

History of hatchery programs, and the need for and problems with genetic monitoring. Current steelhead hatchery programs in north Puget Sound are designed as segregated programs where the hatchery and wild populations are deliberately kept separate by restricting the hatchery broodstock to hatchery-origin individuals[1], and by limiting the reproductive interaction between hatchery-origin and wild fish that spawn naturally. The primary purpose of these segregated hatchery programs is to create or maintain a recreational harvest fishery, and they are managed so that the hatchery-origin run is sufficient to accommodate both the next generation broodstock (i.e., adults returning to the hatchery) and harvest opportunities in the river. This means hatchery-origin adult fish may be present in the river away from the hatchery and if not harvested will attempt to spawn naturally, creating the potential for reproductive interactions with wild fish. In Puget Sound the segregated hatchery programs use either the early winter or early summer populations. Neither population is native to the river systems in which they are released; the origin of the early winter stock is mostly from the now-extinct Chambers Creek population (south Puget Sound), and the origin of the early summer stock is the Skamania Hatchery, Washougal River (lower Columbia River tributary) (Crawford 1979). Since their inception (1950s) the early winter hatchery (EWH) and early summer hatchery (ESH) populations have been managed for accelerated juvenile growth and early adult return and spawn timing (Crawford 1979), so that these hatchery populations now return to fresh water and spawn several months earlier than the native fish in the basins. Therefore, hatchery-origin and wild populations that spawn naturally are presumed to be kept separate reproductively by their difference in return and spawn timing. Hatchery-origin individuals that successfully spawn naturally will produce offspring that are phenotypically indistinguishable from offspring produced from wild fish (e.g., no clipped adipose fin). We refer to these fish as hatchery-lineage

[1] Hatchery-origin individuals are identified by a clipped or missing adipose fin

fish[2]. Furthermore, protracted spawning seasons by hatchery- origin/lineage and wild populations may result in an overlap in spawning creating the opportunity for these "segregated" populations to hybridize. Hybrid individuals would be phenotypically indistinguishable from hatchery-lineage and wild fish.

Hatchery managers of segregated programs are required to monitor the presence of hatchery fish that spawn naturally, and the reproductive interaction between hatchery and wild populations. Since steelhead spawning is difficult to monitor in Puget Sound, and hatchery-lineage and hybrid fish are phenotypically indistinguishable from wild fish, WDFW has elected to use genetic data as the primary tool for monitoring reproductive interaction between hatchery and wild populations. The efficacy of such a monitoring program depends on how well genetic data can differentiate hatchery-lineage, hybrid, and wild fish sampled from the natural-origin (adipose fin intact) population. The EWH source population (Chambers Creek, south Puget Sound), and the wild steelhead populations in north Puget Sound share a common ancestor, perhaps as recent as 13 – 16,000 ybp (date of retreat of Puget Lobe of the Cordilleran ice sheet; Porter and Swanson (1998), Mosher and Hewitt (2004)), while the recent common ancestor between these Puget Sound populations, as a group, and the ESH source population (Washougal River, lower Columbia River) occurred earlier in time. Present-day genetic similarity between the wild steelhead populations in north Puget Sound and the EWH or ESH populations is a function of their recency of common ancestry, natural gene flow, and human facilitated gene flow through hatchery practices (Figure 1). The utility of genetic data to identify the occurrence of hatchery-lineage, hybrid, and wild steelhead fish spawning naturally will depend on our ability to statistically differentiate genetic similarity resulting from recency of common ancestry and natural gene flow, from that resulting from human facilitated gene flow (Figure 1). In this report I describe methods to make these distinctions and to evaluate present-day co-occurrence and interbreeding between steelhead wild populations and segregated hatchery programs.

Organization of report. This report is organized in three Sections. In Section 1, I describe the limits of the program *Structure* (Pritchard et al. 2000, Falush et al. 2003, Hubisz et al. 2009) for identifying known pure (hatchery and wild) and admixed (F1 hybrid) individuals from a series of simulated populations composed of individuals from closely related taxonomic units (e.g., EWH, ESH, and wild steelhead individuals). I used the programs *MS* (Hudson 2002) and *Matlab* (MathWorks 2012) to create these simulated populations, which I parameterized in part with empirical data from north Puget Sound steelhead hatchery programs and natural-origin samples. From the *Structure* analyses of the simulated populations I calculate assignment errors rates. In Section 2, I propose a likelihood method to adjust the proportions from *Structure* to account for these assignment errors and tested the method using the simulated populations from Section 1. I use measures of error and bias to evaluate the accuracy of the *Structure* and likelihood-adjusted proportions. Finally, in Section 3, I use the methods described in Section 2 to provide estimates of the proportion of hatchery-lineage, F1 hybrid, and wild individuals in sample collections

[2] Fish that originated from a river, rather than a hatchery, are natural-origin fish. The parents of natural-origin fish can either be hatchery-origin, natural-origin, or mixed hatchery- and natural-origin (i.e., the fish is a hybrid). To avoid confusing the origin of a fish (i.e., where it was born), from its ancestry (i.e., the identity of its parents), in this document I refer to natural-origin fish (intact adipose fin) with hatchery-origin parents as hatchery-lineage, natural-origin fish with natural-origin parents as wild-lineage or simply wild, and natural-origin fish with one hatchery-origin parent and one wild parent as a hybrid. Therefore, a wild fish may have a hatchery ancestor at some point in its lineage more distant than its parents (e.g., great grandparents), and the term hybrid is limited to the F1 generation.

(Operational Units) and Demographically Independent Populations (DIPs; PSSTRT 2013) from the Green, Snohomish, Stillaguamish, Skagit, and Nooksack river basins. These proportions are summarized using the Proportion Effective Hatchery Contribution (PEHC) statistic, which I defined in Section 1.

Section 1

INTRODUCTION

Since hatchery programs were first implemented in the 1950s wild steelhead populations in north Puget Sound have been subjected to both natural- and human-facilitated gene flow. This means that the genetic similarities among hatchery-origin and wild steelhead from north Puget Sound will be the result of a combination natural processes (e.g., common ancestry and natural gene flow through migration), and human-facilitated gene flow, in particular immigration from early winter hatchery (EWH) and early summer hatchery (ESH) segregated programs. To document the genetic effects resulting solely from the segregated hatchery programs on wild populations we need to differentiate genetic similarity resulting from common ancestry and natural gene flow, from that associated with the hatchery programs (Figure 1). I am concerned here only with identifying hatchery-lineage, F1 hybrid, and wild individuals, that is, individuals with two hatchery-origin parents, one hatchery-origin parent and one wild parent, and two wild parents, respectively (see Preamble), associated with EWH and ESH programs. Consequently, I require a statistical method that will partition "genomes" of individuals from sample collections into three groups (EWH-lineage, ESH-lineage, and wild), allowing for individuals to be admixed (i.e., portions of genomes assigned to more than one group) or hybrids between these groups. I assume that assignment error will exist for populations with recent common ancestry or those with extensive natural gene flow where an individual is assigned to a population that is different from its source population, or an individual is assigned as a hybrid when it is actually a purebred, or vice versa. Since I am concerned only with the parental generation of the individuals in the sample, and because the source populations for the hatchery programs are either extinct (EWH; Chambers Creek) or in the lower Columbia River (ESH), I assume that assignment error from present-day natural immigration is nil and is primarily due to recency of common ancestry among the wild, EWH, and ESH populations. Therefore, if assignment error is mainly a function of common ancestry among the populations, and if I can adjust assignment proportions to correct for this assignment error, I assume that the now-adjusted proportions will reflect mostly human-facilitated gene flow resulting from hatchery practices. That is, natural-origin individuals assigned as hatchery-lineage had two hatchery-origin parents, as hybrids had one hatchery-origin parent and one wild parent, and as wild had two wild parents. To determine assignment error rates resulting only from common ancestry I would need genetic samples from north Puget Sound populations that were not subjected to human-facilitated gene flow, or from affected populations collected prior to the beginning of the segregated hatchery programs (time "C" in Figure 1). However, all north Puget Sound steelhead populations potentially have been exposed to at least one segregated hatchery program, and I do not have available to me a collection of samples from a time before the segregated programs. Therefore, in the absence of empirical data, I modeled pre-hatchery and hatchery-effected populations using simulation populations.

In this Section I developed a two-phased model divided temporally by the date in which the segregated hatchery programs were first implemented in north Puget Sound (time "C" in Figure 1). For the pre-hatchery phase of the model I simulated three populations representing wild Chambers Creek, wild Washougal River (lower Columbia River), and wild north Puget Sound

steelhead populations using the program *MS* (Hudson 2002). I then generated F1 hybrids from pairs of each of these populations and subjected samples from this dataset to a series of analyses using the program *Structure*. From these analyses I determine the assignment error based on the known identity of each individual from the simulated populations. For the second phase of the model (hatchery phase) I simulated a set of hatchery- and natural-origin populations modeled after steelhead populations in north Puget Sound that may have existed in 2008, after the initiation of the segregated hatchery programs but prior to the changes in WDFW's management of these segregated hatchery programs implemented in 2009. As such, these simulated populations are intended to include among-population genetic similarities resulting from common ancestry and natural gene flow, as simulated in the pre-hatchery phase of the model, and genetic similarities from human facilitated gene flow associated with the segregated hatchery programs ("Matlab Simulated Populations" arrow in Figure 1). I used the program *Structure* to identify natural-origin individuals with two hatchery-origin parents (i.e., hatchery-lineage) and one natural-origin parent and one hatchery-origin parent (i.e., F1 hybrid), and two natural-origin parents (i.e., wild), and assessed the *Structure* assignment error based on the known identity of these simulated individuals.

METHODS

Pre-hatchery phase

Simulated populations from the program MS

The program *MS* generates genetic data using coalescent trees that approximate evolution under a Wright-Fisher model (*sensu* Crow and Kimura 1970) based on a series of user-defined parameters. The intent of these simulated populations was to model the empirical genetic structure among a wild north Puget Sound (NPS), Chambers Creek (CC; representing source of EWH programs), and lower Columbia River (LC; representing source of ESH programs) populations that existed prior to the start of the segregated hatchery programs in north Puget Sound. I constructed these populations using parameters values that resulted in (1) hierarchical relationships among the populations, with the Puget Sound populations (wild NPS and CC) sharing a more recent common ancestor than either have with the lower Columbia River (LC) population, (2) genetic diversity within each simulated population that was similar to that within empirical wild, EWH, and ESH populations based on DNA samples analyzed in Section 3, measured using effective number of alleles (A_e), observed heterozygosity (H_o), and unbiased expected heterozygosity (uH_e), each calculated using *GenAlEx* (Peakall and Smouse 2006, 2012), (3) genetic differentiation between each pair of populations that was similar to that between empirical wild, EWH, and ESH population, measured using F_{ST} (Weir and Cockerham 1984) calculated using a *Matlab* custom script, and (4) the number of single nucleotide polymorphism (SNP) loci as in the empirical datasets. Ideally, the genetic diversity within and differentiation among these simulated populations should match those that existed empirically at time "C" in Figure 1. However, since the Chambers Creek population is now extinct, and I am unaware of any north Puget Sound steelhead populations that have not potentially been affected by either a EWH or ESH program, there were no appropriate populations from which I could have parameterized *MS* or to which I could have compared my simulated populations. As such, I used as a surrogate the diversity and differentiation data from existing steelhead populations in

north Puget Sound (Supplemental Tables S2, S3)[3]. These existing populations would differ from the ideal populations that occurred at time "C" in Figure 1 in showing genetic similarity from human facilitated gene flow, in addition to that from common ancestry and natural gene flow (Figure 1). To determine if the simulated populations provided an appropriate representation of populations that may have existed at time "C" in Figure 1, I compared the genetic diversity and differentiation statistics between the simulated and empirical populations. To calculate A_e, H_o, uH_e, and F_{ST} in the empirical populations I aggregated all collections into Demographically Independent Populations (DIPs, PSSTRT 2013), and included all individuals from each collection (see Section 3).

Input parameters for *MS*. *MS* is a Monte-Carlo program and therefore the same set of parameter values can produce marginally different results each run. For this analysis all loci were considered to be unlinked but present on a single chromosome, and I generated a total of 1500 chromosomes (i.e., haploid individuals), 500 chromosomes per population. Although Puget Sound steelhead populations have declined during the past 100 years (Gayeski et al. 2011, PSSTRT 2013), I set the population growth parameter to zero because I wanted a balanced number of samples from each population and given this restriction, this parameter would have had little effect on the population diversity and differentiation. I set the current diploid population size (N_o in *MS*) to 20,000 (Gayeski et al. 2011) and assumed that the split between south Puget Sound (e.g., CC) and north Puget Sound populations to be around 16,000 ybp (Porter and Swanson 1998, Mosher and Hewitt 2004) or 4000 generations. This resulted a NPS – CC population splitting parameter $t = 0.05$. The Puget Sound – lower Columbia (LC) splitting parameter, and the migration matrix parameters were set through trial and error, as I attempted to achieve the appropriate differentiation among the three populations, based on present-day empirical data. The final Puget Sound – lower Columbia population splitting parameter $t = 0.20$ resulted in a divergence time approximately 64,000 ybp or 16,000 generations, assuming N_o = 20,000 (Table S1). I generated loci by setting the mutation parameter to 200 (see below). Assuming N_o = 20,000, a mutation parameter equal to 200 resulted in a neutral mutation rate = 0.0025, and produced 3236 segregating sites or loci. Although the number of SNP loci needed was only 183, based on the number of loci used in the empirical data sets, I wanted to generate several thousand loci to mimic the now routine empirical procedure of generating thousands of SNPs through a next-generation sequencing processes, and reducing this large set down to approximately 192 SNPs. Finally, I allowed for recombination and used as initial parameter value the example value in the *MS* manual, but iteratively altered the recombination and the mutation parameters, until I generated the appropriate genetic diversity within each of the three populations. All *MS* parameter values are summarized in Table S1.

Output from *MS* and subsequent population modeling. Based on the selected parameters and their values *MS* generated 500 haploid genotypes with 3236 unlinked loci from each of the three populations (1500 total individuals). I randomly selected 183 of the 3236 unlinked loci, and based on the allele frequencies calculated from the 500 haploid genotypes I created 500 diploid samples each for the NPS, CC, and LC populations. These samples represented the "founding" populations for the point in time just prior to start of the segregated hatchery programs (i.e., just

[3] The collection data (source of samples and sample sizes) for the hatchery- and natural-origin samples from north Puget Sound are presented in Section 3. All Supplemental Tables are prefixed by an "S")

prior to "C" in Figure 1). To generate a dataset that includes individuals from the NPS, CC, and LC populations, and their pairwise F1 hybrids, I first randomly selected from each founding population 200 of the 500 individuals, randomly divided each set into female and male subsets, produced "gametes" for each individual, and randomly paired female and male individuals, without replacement, within each population (i.e., monogamous mating) to produce 100 diploid individuals for each population. I repeated this process four times, the first time to produce the NPS, CC, and LC individuals in the dataset, and each subsequent time to produce the parents for a specific F1 hybrid combination. I produced F1 hybrids by pairing, without replacement, between each population (i.e., monogamous mating) to produce 100 diploid individuals each for hybrid group (CC-NPS, LC-NPS, CC-LC). This process generated a dataset with 600 individuals, 100 in each category. I repeated this entire post-*MS* process 50 times, and chose from the 50 different datasets, the 10 datasets (iterations) that best modeled the empirical datasets (Tables S4 – S5). I used a combination of *Perl* and *Matlab* custom scripts to generate all simulated populations derived from the *MS* output.

Determining Structure thresholds and assignment error from simulated populations

The program *Structure* is one of the most widely used programs for inferring population structure (see Gilbert et al. 2012 for summary of its use), and has also been used for detecting hybrid individuals, frequently between wild and domestic populations (e.g., Norén et al. 2005, Kidd et al. 2009, Sanz et al. 2009, Marie et al. 2011, Lamaze et al. 2012, Seamons et al. 2012, Harbicht et al. 2014). *Structure* makes use of each individual's multilocus genotype to infer population structure (e.g., hatchery versus wild), given an *a priori* assumed number of groups or populations (k). The program will probabilistically assign individuals to populations, or if the admixture option is used, will assign a portion of an individual's genome to populations.

I ran *Structure* on each of the ten simulated datasets using two different protocols. First, I used the default mode whereby *Structure* considered only genetic information to form groups. Second, I used the USEPOPINFO model whereby *Structure* considered both the genetic information and the prior population source information. This model is useful to test for migrants or hybrids if the predefined populations are usually correct (Pritchard et al. 2010). I limited the use of prior population data to only the NPS, CC, and LC individuals. The hybrid individuals were grouped based only on their genetic information. For both set of protocols, I used the admixture ancestry model, allowing for the identification of admixed or hybrid individuals, k = 2 – 3 groups (two groups for identifying Puget Sound versus lower Columbia; three groups for identifying NPS, CC, and LC populations), and three iterations for each simulated dataset and k. Initial runs were set at both 50,000 burn-in and 100,000 data collection chains (designated here, 50/100), and 5,000 burn-in and 50,000 data collection chains (5/50). Both burn-in and collection chain sets of parameters provided the same results, so all subsequent runs were kept at the shorter 5/50 chains. Therefore, I ran *Structure* six times (k = 2 and 3, with three iterations) for each of ten datasets for a total of 60 runs per model, times two protocol sets (i.e., with or without prior population data), for a total of 120 *Structure* runs.

Q-score thresholds. My main target was k = 3 as my input was three populations (NPS, CC, and LC) and their hybrids. Since I selected the admixture option, *Structure* partitioned a portion of each simulated individual's genome into each of the three populations. These portions are

represented by Q-scores and run from 0 (0% of the individual's genome) to 1 (100% of the individual's genome). The k = 3 groups were standardized across the three iterations so that Assignment Group 1, 2, and 3 represented the NPS, CC, and LC populations, respectively. For each individual I used mean Q-values for each group across the three iterations, and assigned each individual into one of seven assignment groups, representing the three source populations plus their F1 hybrids, and a No Call group where an individual could not be assigned with confidence (Figure 2). To establish the assignment regions in Figure 2, I varied the threshold value (i.e., the lines that define the assignment regions) from 0.05 to 0.50 in 0.01 intervals, and used the following protocols: (1) identified an individual as Pop 1 (i.e., NPS) if the Q-scores for Pop 2 (CC) and Pop 3 (LC) were both less than the threshold value; (2) identified an individual as a hybrid between Pop 1 and Pop 2 (CC-NPS) if the Q-score for Pop 3 was less than the threshold value and Q-scores for Pop 1 and Pop 2 were both greater than the threshold value, and (3) identified an individual as No Call (i.e., not assigned to one of the six source populations) if the Q-scores for all three populations were greater than or equal to the threshold value. Assignments to Pop 2 and Pop 3, and their hybrids, were assigned in like fashion. To select the appropriate threshold value I needed to strike a balance between the overall error rate and the number of No Calls. I defined overall error rate (OER) as

$$OER = 1 - \left(\frac{N_C}{N_A}\right), \qquad \text{Equation 1}$$

with N_C = number of individuals correctly assigned, and N_A = total number of individuals assigned. A low threshold value would constrain the assignment regions, lowering the OER, which is good, but will produce an extensive No Call region resulting in most of the individuals not being assigned. Conversely, a less stringent threshold will reduce the No Call region, as in Figure 2, but will potentially increase the OER. I considered achieving a low OER a higher priority than attempting to reduce the number of No Calls. Therefore, I calculated a mean squared error (MSE) that includes both assignment error and a weighted error associated with the numbers of No Calls, and selected the threshold value that produced the lowest MSE, defined as

$$MSE = \frac{\sum_{i=1}^{n}\left(\hat{X}_i - X_i\right)^2 + w\left(\hat{Y} - Y\right)^2}{n}, \qquad \text{Equation 2}$$

with $\hat{X}_i$ = number of individuals assigned to category i (one of the six assignment categories), X_i = true number of individuals in category i, here 1000 (N = 100 per category for each dataset, pooling all 10 datasets), $\hat{Y}$ = number of unassigned individuals (No Calls), Y = true number of No Calls, here zero (all individuals were from a true source category), w = weight, here 0.10, and n = total number of categories, here seven, including No Calls. I reduced the square error associated with No Calls to 10% of that of square errors associated with the assignment as a way of prioritizing more the assignment error rate than No Calls. The choice of 10% was arbitrary.

Structure error rates. In a typical *Structure* analysis (i.e., not based on modeled or simulated populations), the source group (or population) is not known, and is typically what *Structure* is being used to estimate. For the simulated populations I know the identity of each individual, so I can evaluate the efficacy of any *Structure* run by measuring assignment error rates. There are three types of error associated with *Structure* assignments: (1) the overall error (OER), as

defined above; (2) source error rate (SER) is the frequency at which a source category is incorrectly assigned, either as a proportion of the total in that source category (e.g., proportion incorrectly assigned out of 1000 simulated individuals across the 10 iterations), or as a percentage of the total assigned (i.e., removing the unassigned or No Call individuals). The source error rates are the column errors in Tables 1 and 2; and (3) assignment error rate (AER) is the proportion of individuals incorrectly assigned to a specific assignment group (row assignment errors in Table 1 and 2). An assignment group is proportionally over- or under-estimated by *Structure* when the total assigned to that category is greater than, or less than, respectively, the total that should be assigned (N = 1000) minus the unassigned (No Call) from that source category (Table 1 and 2).

Hatchery phase

Simulated hatchery- and natural-origin populations from north Puget Sound

The purpose of the hatchery-phase of the model was to simulate a present-day system with a wild steelhead population and two segregated hatchery programs, each with a portion of their returning hatchery-origin adults straying away from the hatchery and spawning naturally. I created five spawning situations: one each for the two segregated hatcheries, and three natural spawning conditions where the first included only wild fish, the second included only hatchery-origin fish, and the third where the wild and hatchery-origin fish overlapped and spawned randomly. Although adult females were subjected to natural selection to reduce their fecundity, this model was not intended to evaluate the relative reproductive success of hatchery, hybrid, or wild individuals, or to test different models of relative fitness (e.g., Ford 2002, Baskett and Waples 2013).

To simulate hatchery- and natural-origin steelhead populations from north Puget Sound that existed after the beginning of the segregated hatchery programs in 1950s (time "C" in Figure 1) but prior to the changes in WDFW's management of these segregated hatchery programs in 2009, I used a custom *Matlab* script, which I parameterized using values from WDFW unpublished data, WDFW HGMPs, or the published literature (Table S6). The script starts with the same 1500 haploid genotypes and 3236 unlinked loci I generated using the program *MS* for the pre-hatchery phase of the model . From these haploid genotypes and unlinked loci I founded the ESH, EWH, and wild populations with 5000 diploid individuals each and 183 randomly selected loci. The basic plan for this simulation (Figure S1) was to randomly mate individuals as monogamous pairs, generate eggs per female, and convert eggs to smolts. The hatchery-origin smolts were randomly partitioned into two groups; one released on-station (from the hatchery) and the other released off-station (away from the hatchery). All smolts released on-station either died or returned to the hatchery as hatchery-origin adults. Off-station smolts either died or returned as hatchery-origin adults to spawn naturally either with wild adults to produce F1 hybrid fish, or with other hatchery-origin adults to produce hatchery-lineage fish. The surviving natural-origin smolts, regardless of their lineage, were pooled together as adults and randomly divided into the two spawning groups: one group that spawned only with other wild fish, and another group that spawned randomly with the hatchery-origin fish discussed above. To each female fish I assigned a fitness value that was equal to their percent wild, based on their pedigree, but I set the minimum fitness value to 0.084, based on relative reproductive success values in Araki et al. (2008); median value from Table 1, steelhead – nonlocal). That is, I gave

to each hatchery-origin female that spawned naturally a fitness value of 0.084 rather than zero. Selection affected egg survival only. For each female I multiply her fecundity, which was a random draw from a Poisson distribution, by her fitness, thereby reducing the number of her eggs based on her percent wild. I implemented selection at this point in the program to reduce the computational burden of producing hundreds of thousands of eggs each generation, and not as a hypothesis as to which life history stage would experience natural selection. I ran each iteration of the simulation for 12 generations (approximate number of generations between the start of the segregated hatchery programs and 2008), and for sets of iterations I changed the proportion of hatchery-origin and wild fish that would overlap in their spawning, and potentially interact reproductively. Although I held constant at10% that proportion for the ESH population, I varied the EWH and wild populations proportions using 10%, 50%, and 100% overlap. Therefore, there were nine combinations of overlap (Table 3). For each generation and iteration I recorded and saved the identity, pedigree, and genotype of every hatchery- or natural-origin smolt, and following the 12th generation I randomly selected without replacement 75 natural-origin adults as the "genetic sample" for that population. I also selected all the hatchery-origin individuals that returned to either EWH or ESH following the 12th generation. I repeated each population overlap combination set three times to produce a total of 27 simulated data sets each composed of individuals from the natural-origin, EWH, and ESH populations. Prior to analyzing each population with the program *Structure* I removed from each population all but one individual from each full-sibling family, based on the individuals' pedigree (see Section 3 for discussion of removing full sibling individuals from analyses).

Statistical Analyses

I conducted the same set of statistical analyses on the 27 hatchery phase simulated natural-origin populations as I did on the 10 pre-hatchery phase simulated populations described above. I used the program *Structure* to estimate the composition of each of the simulated populations. I ran *Structure* with and without prior population source information (proportion of source of individuals for each collection are the columns under Pedigree in Table 3). For both protocols, I set runs at 5,000 burn-in and 50,000 data collection chains, used the admixture ancestry model, k = 2 – 3 groups, and three iterations for each simulated dataset and k. Although I ran *Structure* using individuals from natural-origin, EWH, or ESH populations, I reported results only from the targeted natural-origin population, since in empirical datasets these individuals would be unmarked and therefore of unknown identity. I pooled the results for the natural-origin individuals from all 27 analyses and calculated error associated with *Structure* assignments, as overall assignment error (OER), source error rate (SER), and assignment error rate (AER), each defined above. I summarized the hatchery influence on each of the 27 hatchery phase natural-origin populations using two statistics. First, hatchery-wild hybridization (introgression) was assessed simply as the proportion of the population composed of F1 hybrids, either EWH-wild or ESH-wild hybrids. Second, I calculated the Proportion Effective Hatchery Contribution (PEHC), which measures the proportion of the sampled population that is derived from early winter or early summer hatchery-origin parents, and was calculated as:

$$PEHC_W = \frac{2EWH + EWH \sim Wild}{2} \qquad \text{Equation 3}$$

$$PEHC_S = \frac{2ESH + ESH \sim Wild}{2}, \qquad \text{Equation 4}$$

with $PEHC_W$ and $PEHC_S$ being the proportion effective hatchery contribution for the early winter and early summer hatchery programs, respectively, *EWH* and *ESH* equal to the proportion of individuals from a population assigned to the EWH- or ESH-lineage, respectively, and *EWH~Wild* and *ESH~Wild* equal to the proportion of individuals from a population assigned to the EWH-Wild or ESH-Wild hybrid category, respectively.

RESULTS AND DISCUSSION

Pre-hatchery phase: Genetic diversity within and differentiation between simulated populations from the program MS

For the simulated populations to appropriately represent NPS, CC, and LC populations that may have existed just prior to the start of the segregated hatchery programs, their hierarchical relationships, within population genetic diversity, and pairwise genetic differentiation should be similar to the empirical populations used to parameterize the *MS* model. To determine if the simulated populations were appropriate, I compared the simulated and empirical populations' hierarchical relationships, within population genetic diversity, and pairwise genetic differentiation using box plots (Figure 3; see also Tables S2-S5) and principal component analyses (Figure 4, Figures S2 and S3). The empirical populations showed lower within population diversity (Figure 3) and overall genetic variance (Figure 4) than did the simulated populations, possibly reflecting the added human facilitated gene flow in the present-day empirical populations. There were no differences between the simulated and empirical populations in their pairwise differentiation or in their hierarchical relationships (i.e., Puget Sound populations more similar to each other than to the lower Columbia populations). Since the overall ordination of the 20 PCAs were nearly coincident (i.e., there were little multivariate differences between the simulated and empirical populations, Figure 4), I considered the populations simulated using *MS* to be appropriate models for the NPS, CC, and LC populations that may have existed just prior to the start of the segregated hatchery programs.

Pre-hatchery phase: Determining Structure thresholds and assignment error from simulated populations

Structure in default mode – no prior population information. Pooling results for all 10 simulated populations (i.e., total N = 6000), the Q-value threshold that minimized mean squared error (MSE) was 0.22, which coincidentally resulted in an OER = 0.22 and 372 No Calls (Table 1). That is, when the Q-value threshold was set at 0.22, 1256 of the 5628 individuals assigned were assigned to an incorrect source population, and I attributed this error to the genetic similarity resulting from common ancestry and gene flow. Category-specific assignment error rates (AERs) were not equal among the assigned categories. Generally, there were higher AERs for the hybrid categories than the pure categories, with the CC-NPS hybrids showing the highest AER (37%); the lowest AER belonged to LC, followed by NPS (Table 1). There was a positive

bias (overestimation) in the assignment error for CC-NPS, and negative biases (underestimation) for the assignment error for CC and NPS (Table 1), indicating that more CC and wild individuals were incorrectly assigned as CC-NPS hybrids than there were CC-NPS incorrectly assigned as either CC or NPS (Figure S4). The source error rates (SER) for assigned individuals followed the same pattern as that for the AERs; higher SERs for the hybrid categories than the pure categories, CC-NPS with the highest SER (31%) and LC with the lowest SER (7%) (Table 1, Figure S4). As expected, due to their more recent common ancestry and higher levels of natural gene flow, there were higher AERs and SERs among the Puget Sound categories, than any assignment associated with a LC source category. There were also considerably more unassigned (No Call) individuals among the hybrids than the pure individuals.

Structure using USEPOPINFO model – prior population source information used. When using prior population source information for pure individuals the Q-value threshold that minimized mean squared error was 0.27, which resulted in an OER = 0.09 and 550 No Calls (Table 2). That is, using prior population information when running *Structure* resulted in a substantial reduction in the overall error rate, compared with not using prior population information. Likewise, except for LC, the AERs were lower with prior than without prior population information (compare AERs in Tables 1 and 2). Although LC had the lowest AER without the priors, with priors LC had the largest AER, with nearly 10% of the LC-NPS and CC-LC hybrids incorrectly assigning to LC. Not surprisingly, making use of the prior population data resulted in very few wild, CC, and LC individuals being incorrectly assigned (Table 2, Figure S5). Furthermore, the No Call rate was zero for all three pure categories, 8% for CC- NPS, but quite high for the LC-NPS, and CC-LC hybrid categories (Figure S5).

Summary. The two different *Structure* models (with and without prior population source information) produced dramatically different error rates. Although when using prior population information the number of unassigned individuals (No Calls) was nearly 3% greater, the OER was 13% less, compared with not using prior population information. However, the prior population source model is only useful when prior information is available, such as in simulated data where the identity of all individuals is known. Such is not the case with empirical data. Regardless of which model is used, the *Structure*-based error rates for these pre-hatchery phase simulated populations are most-likely due to genetic similarity among the populations resulting from common ancestry and natural gene flow.

Hatchery phase: Simulated natural-origin populations from north Puget Sound

The purpose of the hatchery phase of the model was to simulate present-day natural-origin steelhead populations from north Puget Sound. The genetic similarity between these populations and EWH and ESH populations is a function of natural gene flow (prior to the extirpation of the Chambers Creek population), common ancestry, and human-facilitated gene flow resulting from hatchery practices. Therefore, these hatchery phase populations differed from the pre-hatchery phase populations discussed above in having the added component of human-facilitated gene flow (Figure 1). Table 3 describes the 27 natural-origin hatchery phase collections simulated using schema in Figure S1 and parameters in Table S6. Regardless of the percent overlap between the EWH and Wild populations each of the 27 simulated natural-origin collections was dominated by Wild fish, with Wild proportions ranging from 0.72 to 1.00, with a mean = 0.95.

Nine of the 27 collections were composed entirely of Wild fish. The number of true hybrids, based on the pedigree of the samples within each collection ranged from 0.00 to 0.05 (mean = 0.01) for EWH-Wild and 0.00 to 0.03 (mean = 0.00) for ESH-Wild hybrids (Table 3), while true PEHC ranged from 0.00 to 0.24 (mean = 0.04) for the EWH program and 0.00 to 0.02 (mean = 0.00) for the ESH program (Table 4). The percent overlap set that produced the highest $PEHC_W$ was {0.5, 1.0, 0.1}, followed by {1.0, 1.0, 0.1} and {1.0, 0.5, 0.1}, for {Wild, EWH, ESH}, respectively. Nevertheless, given the model's stochasticity, and a random sample of only 75 individuals (minus all but one individual from full-sibling families, if present), there was considerable variation in the population proportions and PEHC within each percent overlap set.

Hatchery phase: Structure results

As with the pre-hatchery phase of the model, I subjected the populations here to two types of *Structure* analyses, with and without prior population source information. I used as the *Structure* thresholds the values calculated for the pre-hatchery phase of the model (0.22 and 0.27 for the no-priors and priors analyses, respectively). These thresholds minimize OER for the pre-hatchery phase populations, based on assignment errors associated with common ancestry and natural gene flow only, and was used as part of the assignment adjustment procedure in Sections 2 and 3. The overall error rate (OER) for the hatchery phase no-priors analysis was 0.18 (Table 5), compared with 0.04 when using prior source population information (Table 6). These OERs generally reflect the error associated with the Wild fish since the source for an average of 95% of the individuals was Wild (Tables 3, 5-6). For the no-priors analysis, most of the EWH- and ESH-lineage individuals assigned correctly (SER = 0.08 and 0.00, respectively). However, assignment error rates (AER) were extremely high for all categories except Wild (Table 5). Over half of the 114 individuals assigned to the EWH-lineage were not truly EWH-lineage (nearly half were Wild), and nearly all of the 259 individuals assigned as hybrids where actually Wild fish. That is, there was a strong positive bias using *Structure* to overestimate the proportion of hatchery-lineage and hybrid fish, with nearly 20% of the 1793 Wild fish being incorrectly assigned as hatchery-lineage or hybrid fish (Table 5, Figure 5). If an individual's Q-score for the Wild category in *Structure* accurately represented the percent Wild of that's individual's genome, there should be a high correlation between that Q-score and the percent Wild based on the individual's pedigree (diagonal line, Figure 5). In fact, the relationship is quite poor, even for the set of individuals that are greater than 90% wild, based on their pedigree; *Structure* assigned these nearly pure Wild fish to all seven categories, including No Call (Figure 5). In other words, *Structure* did a poor job of sorting individuals into their correct pedigree-based source category, and the analysis resulted in a relatively low OER only because the collections were composed nearly completely of Wild fish.

The priors analysis presented a different set of results (Table 6, Figure 6). First, the OER was only 4%. Second, the SERs and AERs were high and low, respectively, for the hatchery-lineage categories, which is opposite of what occurred in the no-priors analysis. Of the 61 EWH-lineage individuals 23 assigned as EWH-Wild hybrid and 15 assigned as Wild. Although 16 of the 18 EWH-Wild hybrids assigned as Wild, there were twice as many fish assigned as EWH-Wild hybrids as there should be because of the incorrect assignment of Wild and EWH-lineage fish to the EWH-Wild category (Table 6). Finally, there were no unassigned fish in the priors analysis, with all individuals' Q-values having little influence from a third group in the k = 3 analyses.

The greater overall accuracy of the priors analysis, compared with the no-priors analysis can be seen by comparing Figures 5 and 6. However, what is also apparent from Figure 6 is the high SER for EWH-lineage fish, where roughly equal number of EWH-lineage fish were assigned to EWH-lineage, EWH-Wild hybrid, and Wild categories (see also Table 5), and the fact that EWH-lineage and EWH-Wild hybrids are assigned as Wild fish at a relatively high proportion. Both the no-priors and priors analyses overestimated the number of hybrid fish; however the overestimate for the no-priors analysis was nearly six times that of the priors analysis.

In the no-priors analysis for every simulated natural-origin collection $PEHC_W$ was greater than the true $PEHC_W$, and the average $PEHC_W$ value was three times the true value (Table 4, Figure 7, upper left). This resulted from the overestimate of both EWH-lineage and EWH-Wild hybrids Figure 7, upper right) by Wild fish being assigned as either a hatchery or hybrid fish. In the priors analysis there were fewer Wild fish assigned as either hatchery-lineage or hybrids, and roughly 64% of the EWH-lineage fish incorrectly assigned to a category other than EWH-lineage (Table 6). This resulted in an underestimate of the number of EWH-lineage fish, and an underestimate of $PEHC_W$ for 22 of the 27 simulated natural-origin collections (Table 4, Figure 8, upper left). Despite this underestimate, $PEHC_W$ from priors analysis had a smaller mean squared error and bias compared with that from the no-priors analysis.

CONCLUSIONS

To determine the number of hatchery-lineage and hatchery-wild F1 hybrids among unmarked (natural-origin) steelhead in Puget Sound requires a reliable method to differentiate hatchery-lineage and hybrid individuals from each other and from their genetically similar wild individuals. One of the goals of this section was to evaluate the efficacy of the program *Structure* to correctly identify hatchery-lineage, hybrid, and wild fish in a collection of present-day steelhead populations in north Puget Sound, simulated using a two-phased model. When I used *Structure* without prior population information, it underestimated the number of wild fish and overestimated the number of hatchery-lineage and hatchery-wild hybrid fish, resulting in an overestimate of PEHC and introgression. Therefore, based on simulated data, *Structure*, in no-prior population information mode, overestimated the genetic effects of hatchery-origin fish on the wild steelhead population. I obtained nearly the opposite result when using *Structure* with prior population information. Here, *Structure* overestimated the number of wild fish, while underestimating the number of hatchery-lineage fish, resulting in an underestimate of PEHC, and an underestimate of the genetic effects of hatchery-origin fish on the wild steelhead population. Although the underestimation bias for PEHC from the priors analysis was considerably lower than the overestimation bias from the no-priors analysis, as a precautionary measure it is more prudent to overestimate than underestimate an effect. Furthermore, the priors analysis is based on having prior source-identity information, which is available for simulated data but generally not available for empirical data.

For the other goal of this section I used the simulated populations from the pre-hatchery phase of the model to established *Structure* thresholds that minimize *Structure* assignment error resulting from genetic similarity associated with common ancestry and natural gene flow only. I used these thresholds to objectively define the Q-value boundaries for identifying hatchery-lineage, hybrid, and wild fish. I will use the pre-hatchery phase assignment errors in a likelihood-based

procedure, described in Section 2 and implemented in Sections 2 and 3, to adjust *Structure* proportions to factor out the effects of common ancestry and natural gene flow, thereby providing a more accurate estimate of the genetic effects of hatchery-origin fish on the wild steelhead population.

Section 2

INTRODUCTION

In Section 1 I described briefly the history of the segregated steelhead hatchery programs in north Puget Sound, and suggested that the genetic similarities among natural-origin hatchery-lineage, hybrid, and wild fish are the result of a combination of common ancestry, natural gene flow that occurred in the past, and human facilitated present-day gene flow from segregated hatcheries. I showed that there is error in identifying genetically hatchery-lineage, hybrid, and wild fish using the program *Structure*, and depending on how *Structure* is implemented, that this error may underestimate or greatly overestimate the degree to which hatchery-origin fish affect wild populations. I suggested that this assignment error, for the most part, was associated with close genetic similarity among the fish resulting from common ancestry. In Section 2 I propose that if we account statistically for this assignment error resulting from common ancestry we can adjust the *Structure* proportions to more accurately describe the composition of natural-origin collections. I present here a likelihood-based procedure (Warheit and Knapp, in prep) that performs such an adjustment. Furthermore, in this Section I test the efficacy of this likelihood method using the 27 hatchery phase simulated collections from Section 1. I compare PEHC and introgression (F1 hybrids) with and without the likelihood adjustment using measures of mean squared error and bias (defined below) to determine if the likelihood procedure improves the accuracy of the proportion estimates of natural-origin collections.

METHODS

Adjusting structure assignments using known assignment errors from pre-hatchery phase of model

When *Structure* assigns an individual (or portion of that individual's genome, with the admix model) to a group or groups, we assume that the assignment reflects the true identity of that individual. However, the program *Structure* is known to make incorrect assignments under a variety of conditions (e.g., Vähä and Primmer 2006, Anderson and Dunham 2008, Kalinowski 2011, Seamons et al. 2012), and there is known *Structure* assignment error here with my simulated populations (Tables 1-2, 5-6). Given this assignment error, we need more confidence that when using *Structure* the assigned groups correspond to source groups, so when we summarize *Structure* assignments into assigned group proportions, these group proportions are accurate frequencies of each group's occurrence in the population. That is, when we analyze a dataset of individuals of unknown identity and *Structure* assigns 10% of these individuals as hybrids, for example, we need confidence that this assignment proportion reflects a correct occurrence of hybrids in the river. I propose here a likelihood approach (Warheit and Knapp, in prep) to adjust the *Structure*-based summary assignment proportions, given assignment error (Table 1, 2), to more accurately characterize the source proportions. I outline in detail this approach below (see also Figure S6), but the procedure can be summarized as a series of random resampling to generate an estimate of slope, intercept, and variance for a regression line. These parameters are then used in the normal regression likelihood function (Equation 5), with the maximum likelihood providing the adjusted proportion for that source category. I then use log-

likelihood ratios to calculate confidence intervals for the adjusted source proportion. The likelihood adjustments to source proportions are calculated separately for each of the six assignment categories (two hatchery-lineages, three hybrids, and one wild category, as in Figure 2), and then normalized by dividing each adjusted source proportion by the sum of all source proportions, ensuring that the sum of all proportions equals 1.00. In what follows I itemize the steps showing as an example the adjustment to the EWH-Wild hybrid category.

Step 1. Run *Structure* and summarize assignments into the six assignment categories. Remove individuals in the No Call category. Calculate relative frequencies for each of the six categories, and based on relative frequencies, expand sample to a larger size, here I use 1000. The target category is EWH-Wild, with an expanded count of 208.

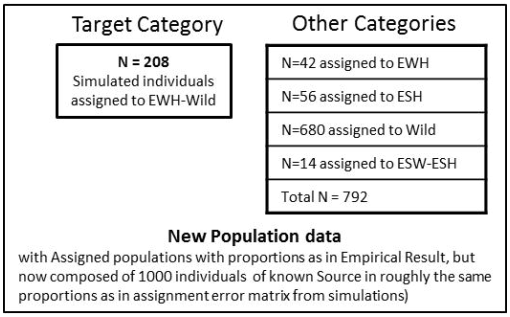
Assigned Category	Assignments from Structure	Proportions	Expand to N = 1000
EWH	3	0.04	**42**
EWH-Wild	15	0.21	**208**
ESH-Wild	4	0.06	**56**
ESH	0	0.00	**0**
Wild	49	0.68	**680**
EWH-ESH	1	0.01	**14**
N	72		**1000**

Step 2. Generate simulated data and calculate assignment error matrix (e.g., Tables 1, 2). The error matrix is essentially a lookup table, where you look up the number of source individuals that were assigned to the target category. For example, our target category, EWH-Wild, had 1162 individuals assigned to it, of which only 625 were actually EWH-Wild source individuals. That is, each of the 1162 individuals assigned to EWH-Wild have source names attached to them, 216 are EWH, 625 are EWH-Wild, and so on, reading along the EWH-Wild row.

Assigned	Source						
	EWH	EWH-Wild	ESH-Wild	ESH	Wild	EWH-ESH	TOTAL
EWH	655	113	4	1	5	66	**844**
EWH-Wild	216	625	58	1	229	33	**1162**
ESH-Wild	0	11	514	24	26	37	**612**
ESH	0	0	87	839	0	159	**1085**
Wild	5	153	79	0	693	1	**931**
EWH-ESH	61	18	40	98	3	507	**727**

Step 3. Randomly select, with replacement, 208 individuals (the target category expanded count) from the 1162 simulated individuals assigned to the EWH-Wild category (designated here "Target Category"). On average 54% (625/1162, from the assignment error matrix) of the 208 individuals (112) are EWH-Wild source individuals, the remaining 46% are from other source categories. Repeat process for the other assigned categories. For example, for the Wild category, randomly select with replacement, 680 individuals (Wild expanded count from Step 1) from the 931 simulated individuals assigned to the Wild category. On average 16% (153/931, from the assignment error matrix) of the 931 individuals (149) are actually EWH-Wild source

individuals. All randomly selected individuals from the five other non-target categories are compiled together (as Other Category) but kept separate from the randomly selected individuals from target Category. The result from this step is a new dataset composed of 1000 randomly selected individuals with known source categories.



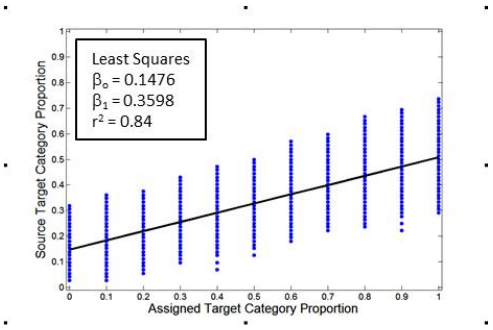
Step 4. The assignment error matrix is based on a *Structure* analysis where there were equal numbers of individuals from each of the six source categories (N = 100 for each, see Section 1). These six categories most-likely do not occur in a river or in a sample in equal proportions. In fact, it is these relative proportions that we are attempting to estimate. To simulate different relative proportions between the Target Category and Other Categories, construct a series of new datasets, composed of the original sample size from Step 1 (for this example, N = 72), by randomly selecting specific number of Target and Other Category individuals from the 1000 individuals compiled at Step 3. For example, a dataset that simulates 0% Target individuals and 100% Other individuals is composed of 72 randomly selected individuals from the N = 792 Other Category. A dataset that simulates 60% Target individuals and 40% Other individuals is composed of 43 randomly selected individuals from the N = 208 Target category and 29 randomly selected individuals from the N = 792 Other Category. This process is repeated for 0% to 100% Target Category, at 10% intervals, for a total of 11 new datasets, each composed of 72 individuals of know source category.

Proportion of Target in new collection	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
# of randomly drawn Target samples	0	8	15	22	29	36	43	50	57	64	72
# of randomly drawn Other samples	72	64	57	50	43	36	29	22	15	8	0

Step 5. For each of these 11 new datasets, count the number of individuals whose source category is the same as the Target assigned category (for this example, EWH-Wild). Convert counts to relative frequencies. Repeat Step 4 and this step multiple times. For this analysis I repeated the process 10,000 times to produce a 10,000 x 11 matrix composed of relative frequencies of individuals whose source is the Target category.

Proportion of Target in new collection	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Proportion of Target Source individuals	0.10	0.21	0.18	0.22	0.19	0.32	0.35	0.46	0.49	0.49	0.54

Step 6. To quantify the relationship between the assigned Target category proportion (0% - 100% at 10% intervals) and the source Target category proportion (from Step 5), conduct a least squares analysis and record the intercept (β_o) and slope (β_1) of the regression line.



Step 7. To estimate the adjusted assigned Target category's relative frequency from the unadjusted assigned Target category's relative frequency, use the likelihood function for the normal regression,

$$\mathcal{L}\left(\beta_0, \beta_1, \sigma^2 \mid Y, X\right) = \frac{1}{\sqrt{(2\pi\sigma_i^{\,2})}} e^{-\frac{1}{2\sigma^2}\left(Y-(\beta_0+\beta_1 X_i)\right)^2}$$

Equation 5

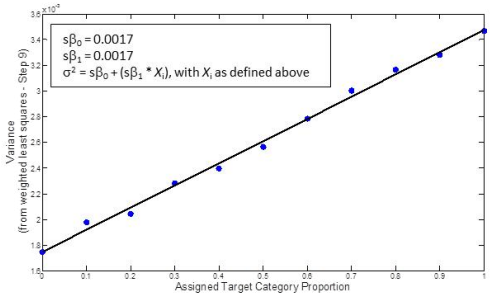
with,

Y = empirical assigned Target proportion (from Step 1; here 0.21),

X_i = assigned Target proportions as in least squares regression in Step 6, except here from i= 0 to 1.0 at 0.001 intervals,

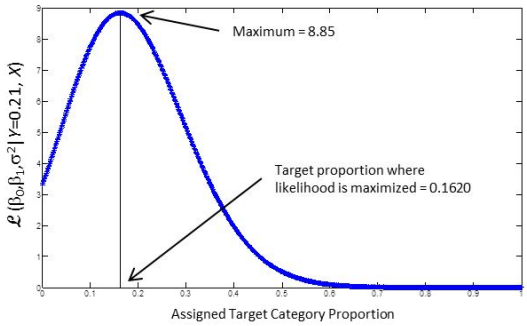
β_0 and β_1 from least squares regression (Step 6),

σ^2 from least squares regression of the variance from Step 6 against the assigned Target category proportions, and calculated as: $\sigma^2 = s\beta_o + (s\beta_1 * X_i)$, with $s\beta_o$ and $s\beta_1$ from regression, and X_i as above.

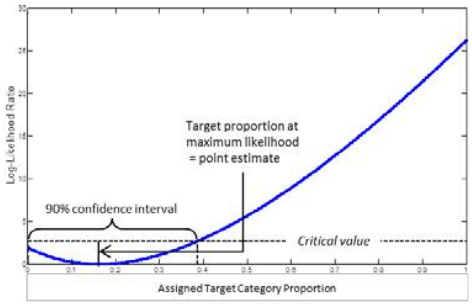


Step 8. Use the function from Step 7 to calculate likelihoods for each X_i. The adjusted assigned Target proportion, given an assigned Target proportion from *Structure* is the proportion where the likelihood is maximized (for this example, maximum likelihood is 8.85, with an adjusted assigned Target proportion = 0.16, correcting downward the assigned proportion of 0.21 (from

Step 1). In other words, given the assigned proportions from *Structure* and the assignment error matrix from the simulations, the Target proportion is adjusted downward from 21% to 16%.



Step 9. Calculate confidence intervals (here, 90% CI) for the point estimate (i.e., the proportion where the likelihood is maximized) using the log-likelihood ratio test. That is, determine what likelihood values were not significantly different from the maximum likelihood value. Here alpha is defined as 0.10, with the critical value approximated using chi-square with 1 degree of freedom. The confidence interval is defined as the range of likelihoods that are not significantly different from the maximum, given alpha. The 90% CI range is defined as those likelihoods that fall below the critical value, and the end points of the range are the smallest and largest likelihoods within that range. The flatter the log-likelihood ratio curve, the broader the confidence interval.



Step 10 (not shown in Figure S6). Since each Target category is adjusted separately, the now adjusted source proportions, across all categories, may not sum to 1.00, as they should. Therefore, normalize proportions by dividing each adjusted source proportion by the sum of all source proportions.

RESULTS

Likelihood Adjustments to the Structure Proportions

For the no-priors analysis there were four simulated collections where the 90% confidence interval range (CI) for the likelihood adjustment of the EWH-Wild hybrid proportion exceeded 0.25. These adjusted proportions are highlighted in bold in Table S7, and should be treated as

uncertain (actual 90% CIs are not shown). The PEHC values for these collections are highlighted in red in Figure 7, and they too should be treated as uncertain. Overall, the likelihood procedure adjusted downward the frequency of hybrids, increased the frequency of the Wild category and had little or no effect to the frequency of the hatchery-lineage categories (Table S7). As a result, the likelihood procedure correctly adjusted downward the unadjusted values for both hatchery-wild introgression and PEHC (Table 4). The adjustment to the hatchery-wild introgression values greatly reduced the mean squared error from the unadjusted values, and when I removed those proportions with uncertain adjustments, the likelihood adjustment provides an unbiased estimate of introgression (Figure 7). Because the likelihood procedure had little effect on the hatchery-lineage proportions, the downward adjustment to $PEHC_W$ was influenced mostly by the change in the EWH-Wild proportion (Table S7). Although the $PEHC_W$ mean squared error was reduced from the unadjusted value, mean squared error was not completely removed and the adjusted results still showed a positive bias, especially for the higher $PEHC_W$ scores (Figure 7).

The likelihood procedure had little effect on the unadjusted proportions for the priors analysis, especially for $PEHC_W$ (Table 4, Figure 8, Table S8). Although the error and bias for the unadjusted EWH-Wild introgression results were small, the likelihood procedure did provide some adjustment, and, as with the no-priors analysis, resulted in an unbiased estimate of introgression (Figure 8). The likelihood adjustment resulted in a poorer fit for $PEHC_W$ than the unadjusted value (Figure 8). After the likelihood adjustment, there was little difference between the no-priors and priors $PEHC_W$ value mean squared errors and biases, except that in the no-priors analysis $PEHC_W$ was overestimated and in the priors analysis $PEHC_W$ was underestimated (Figures 7, 8).

DISCUSSION

The hatchery-phase simulated collections provided a mechanism to test the accuracy of *Structure* to estimate the relative proportions of hatchery-lineage, hybrid, and wild fish in a collection, and the efficacy of the likelihood procedure to adjust the proportions to improve the accuracy of the estimated proportions. There was high assignment error when using the no-priors *Structure* procedure (see Section 1), which was improved to a large degree when I applied the likelihood adjustments. The likelihood procedure provided an unbiased estimate for the F1 hybrid proportions, and therefore an unbiased estimate for the amount of hatchery-wild introgression represented by each of the collections. However, the likelihood procedure provided little or no adjustment to the hatchery-lineage proportions in the collections, and although $PEHC_W$ was correctly adjusted downward, the estimate was not unbiased. The likelihood procedure used the pre-hatchery simulated population assignment error matrix (Table 1), and although the error matrix showed high assignment error for both CC and CC-NPS hybrids, there was strong positive bias to the CC-NPS hybrids assignment (i.e., more individuals assigned as CC-NPS hybrids than there should have been), but only weak negative bias to the CC assignment. There was also a negative bias to the NPS assignment. The strong positive bias to the CC-NPS hybrid assignment was the reason the likelihood procedure adjusted downward the EWH-Wild proportions in all 27 collections (Table S7). The negative bias for the NPS assignment was greater than that for the CC assignment, and the likelihood procedure correctly adjusted upwards the Wild proportions in all 27 collections, but adjusted upwards the EWH-lineage proportions for

only eight collections, while it adjusted downward for three collections, and kept the same proportion for 16 of the collections (Table S7). The assignment error for the priors *Structure* procedure was considerably lower than that for the no-priors produce (Tables 1, 2) and therefore the proportion estimates for the hatchery phase collections required and received little adjustment from the likelihood procedure.

Overall, for both the no-priors and priors analyses the likelihood procedure worked as it was intended; however, the efficacy of the procedure will depend on the accuracy of the assignment error matrix. In terms of which analysis provided the best estimate for PEHC and introgression within a collection, I considered error, bias, and the usefulness of the likelihood procedure. Clearly, the priors analysis provided more accurate assignments (lower MSE) than the no-priors analysis when using *Structure* without the likelihood adjustments (Figures 7, 8). However, following the likelihood adjustments, there was no difference between the two analyses in measuring the hybrid proportions (considering only the no-priors analyses that did not include the uncertain adjustments), and therefore both analyses provided an unbiased estimate of introgression. There were only small differences in error and bias between the no-priors and priors analyses in the likelihood adjusted PEHC measure. The main difference between the two analyses was that the PEHC estimate was an underestimate for the priors analysis, and an overestimate for the no-priors analysis. If one were to use the precautionary approach to manage steelhead segregated hatchery programs in north Puget Sound, it would be better to use a statistical method that results in an overestimate rather than an underestimate of a potentially harmful effect, such as hatchery fish contributing to the next generation of the natural-origin population. Furthermore, the priors analysis requires knowing the source-identity of each of the individuals. This information is available for simulated data, but generally not available for empirical data, and is frequently what you are trying to estimate. For these reasons, to estimate the relative proportions of hatchery-lineage, hybrid, and wild individuals in a population, I used the no-priors *Structure* analysis in combination with the likelihood adjustment to the population proportions with the empirical data in Section 3.

Section 3

INTRODUCTION

In Sections 1 and 2 I introduced and justified a series of statistical analyses that would estimate the relative proportions of hatchery-lineage, F1 hybrid, and wild individuals in a collection. The likelihood procedure I described and tested in Section 2 provided an unbiased estimate of the F1 hybrid proportion in the collection, and therefore an unbiased estimate of hatchery-wild introgression. However, the procedure, as implemented in Section 2, using an assignment error matrix developed in Section 1 did not sufficiently adjust the hatchery-lineage proportions and therefore, the estimate of PEHC may be upwardly biased, or an overestimate of the true PEHC. Based on the results in Section 2, I determined that the error and bias in PEHC was small enough to warrant the continued use of both the program *Structure* and the likelihood procedure to adjust the *Structure* proportions. The fact that PEHC is potentially overestimated in empirical datasets would allow for a precautionary approach to the management of steelhead segregated hatchery programs.

In this Section I implement the methods described in Section 2 to estimate the proportion of hatchery-lineage, F1 hybrid, and wild individuals in samples from the Green, Snohomish, Stillaguamish, Skagit, and Nooksack river basins. I conducted this analysis at two organizational levels. First, I aggregated samples from the same river basin, depending on collection date, life history stage (juvenile versus adult), and run-timing, into Operational Units (OUs). Second, I aggregated the OUs into Demographically Independent Populations (DIPs; PSSTRT 2013). As I did with the hatchery-phase simulated collections in Section 2, I summarize the relative contribution of hatchery-origin fish into the natural-origin population (OU or DIP) using two statistics: Introgression, which is simply the proportion of F1 hybrids, and the proportion effective hatchery contribution (PEHC) defined in Section 1.

METHODS

Samples

All samples used in this analysis were fin tissue samples archived in the Washington Department of Fish and Wildlife Molecular Genetics Laboratory (WDFW-MGL) tissue collection. Each sample collection is accessioned with a WDFW-MGL code, and most collections are associated with field collection data that includes collection year and location, age or life stage of individual samples, collection dates, origin (hatchery [adipose fin absent] versus natural [adipose fin present]), and a presumed run timing (Table 7). My focus here was wild and hatchery steelhead collections in north Puget Sound, so I limited collections for genotyping and analysis to those located in the Green, Snohomish, Stillaguamish, Skagit, Samish, and Nooksack river basins. I combined wild samples with similar collection year and dates, life stage, origin, and presumed run timing into collection aggregates, which I called Operational Units (OUs) (Table S9). Operational Units were the primary unit for analyses. Operational Units were combined into NOAA PSSTRT designated Demographically Independent Populations (DIPs) (PSSTRT 2013), based on the OUs' location and presumed run timing (Table S9). All hatchery collections

(Tables 7, S10) were limited to steelhead segregated programs, which in Puget Sound include early winter hatchery (EWH) programs, which were derived primarily from wild winter steelhead from Chambers Creek, Puget Sound, Washington, and early summer hatchery (ESH) programs, which were derived initially from wild summer steelhead from the Washougal River (Skamania Hatchery), lower Columbia River, Washington.

Genotypes

I used two 96 single nucleotide polymorphism (SNP) panels, for a total of 192 SNPs, to genotype all samples (Table S11). These panels, designated by WDFW-MGL as Panels E and F, were designed for the purpose of genotyping *O. mykiss* (steelhead, and rainbow and redband trout) samples throughout Washington State as baseline samples for genetic stock identification (GSI), population differentiation, and hatchery management, and not specifically for measuring hatchery-wild introgression. That is, SNP loci were selected primarily for describing within population genetic diversity and among population differentiation for anadromous and resident populations of *O. mykiss* statewide, and not specifically for their ability to differentiate groups of individuals within populations in north Puget Sound. Three of 192 SNPs were designed to identify pure cutthroat trout (*O. clarki*), or cutthroat – steelhead hybrids, both of which occur in north Puget Sound river systems, and could be phenotypically confused with pure steelhead samples, especially juvenile samples.

Sample and genotype quality assurance (QA)

For the entire dataset, I removed loci if there was no variation across all individuals, or if fewer than 80% of the individuals were scored. For individual basin analyses I removed a locus if it was not scored for an entire OU. After surveying for cutthroat trout alleles, I removed the three loci designed to identify pure cutthroat trout or cutthroat – steelhead hybrids. Samples were removed for two reasons: (1) if the sample was scored with one or more cutthroat alleles at any one of the three designated cutthroat specific loci; and (2) if more than one-third ($N \geq 63$) of the loci were missing . Additionally, each OU was analyzed using the program *COLONY* (Wang 2004, Wang and Santure 2009, Wang 2012, 2013) to estimate full-sibling families. Each OU was run separately. Each *COLONY* run was a short run, with both parents selected as polygamous, without inbreeding, with the combined pairwise-full likelihood method (Wang 2012) and medium precision. Allele frequencies were set as unknown and were not updated. For the *Structure* analyses only, I removed all but one randomly chosen individual from each full-sibling group estimated by *COLONY*. The presence of family groups in a data set violates the model assumptions in *Structure* and can affect the program's results by inflating k (i.e., by conflating population and family group structure) (Falush et al. 2003, Anderson and Dunham 2008, Garza et al. 2014). Following the *Structure* analyses, I reinstated all individuals into the appropriate assigned group based on the assignment of that full-sibling group's representative in the data set. Therefore, the hatchery-lineage, hybrid, and wild proportions from *Structure*, and the adjusted proportions from the likelihood procedure, reflect the full data set that includes all members of each full-sibling group. I used the full data set for all other statistical analyses (e.g., PCA, F_{ST}).

Aggregating OUs into Demographically Independent Populations (DIPs)

I aggregated OUs into their respective DIP (Table S9) by adding together the frequencies in each *Structure* assigned category (i.e., the frequency of individuals in the EWH-lineage, ESH-lineage, wild, and F1 hybrid categories) across all OUs within each DIP. For DIPs where spawning distributions are not evenly distributed among the contributing OUs, I weighted (i.e., multiplied) each of the OU's frequencies for each category by the OU's estimated spawning proportion within the DIP (Table S12). Then for each category I added these products across all OUs contributing to that DIP. Because I applied a weight to each frequency the sum of the weighted products for each category was less than it would be if I had not applied the weights. To adjust the category sums so that the sample size for the entire DIP would equal the sum of the sample sizes of the contributing OUs, I multiplied each category sum by the ratio of the unadjusted to adjusted sum of the contributing OUs across all categories.

Statistical Analyses

To determine the relative proportion of hatchery-lineage, hybrid, and wild individuals in each OU and DIP I used the same *Structure* and likelihood methods that I described and implemented in Sections 1 and 2. However, in this Section I limited the *Structure* analyses to the default model where *Structure* considered only genetic information to form groups, and ignored the prior population source information (the "no-priors" analysis, as labeled in Section 2). Each *Structure* analysis included one natural-origin collection (OU or DIP), an EWH collection, and an ESH collection (i.e., three "populations"). The Skagit and Nooksack Rivers do not have an ESH program. Therefore, to include ESH fish in those analyses (to compare with the native summer populations, or to look for ESH hatchery-lineage strays), I used the Reiter Ponds collection from the Snohomish River. Similarly, although the Stillaguamish River does have both EWH and ESH programs, no samples from these programs were available to me for this project. Therefore, I used the Tokul Creek and Reiter Ponds (Snohomish River) collections for this analysis. To evaluate the latent population structure among hatchery and natural-origin collections, I conducted principal component analyses (total individual correlation matrix of allele frequencies) on the entire dataset, and on each river basin's data set, including all samples. Unless otherwise indicated, I conducted all statistical analyses in *Matlab* using custom scripts.

RESULTS AND DISCUSSION

Samples and Loci

Technicians from the WDFW Molecular Genetics Laboratory genotyped 1787 natural-origin fish from 48 collections and 33 OUs, and 464 hatchery-origin fish from eight collections and six hatchery programs (Table 7, S9, S10). I removed from the 1787 natural-origin fish, 34 and 48 individuals due to their incomplete genotypes and presence of cutthroat alleles, respectively. For the *Structure* analyses only I also removed a total 188 samples from 22 OUs (27 collections; range = 1 – 35 individuals removed per collection) where I detected full-sibling families, retaining only one member of each family. Across all collections the median value for the number of related individuals removed was one. If I limited the calculation to the 27 collections with related individuals the median increased to three. As expected, I found more full-siblings

within the juvenile collections (N = 164), than among the other life stages. These 164 individuals represented 21% of all juvenile samples genotyped, while the full-sibling proportion of the adult samples was only 3%. Of the 84 individuals I removed from the hatchery collections, 80 were removed as full-siblings, and four removed due to the presence of cutthroat alleles. Among the hatcheries, Kendall Creek – early winter collection had the most number of full-siblings (33, 33%), while Soos Creek – early winter collection had the fewest (2, 5%). Across all hatchery- and natural-origin OUs, I removed 16% of the individuals and used N = 2165 for all statistical analyses except *Structure*, and N = 1897 for the *Structure* analyses.

Three of the 189 steelhead-specific loci were removed from all samples because fewer than 80% of all individuals had a usable genotype (N = 2) or the locus was monomorphic (N = 1). Of the remaining 186 loci, 180 (Nooksack and Samish), 182 (Snohomish and Skagit), and 183 (Green) and 184 (Stillaguamish) loci were used for basin-specific analyses, and 178 loci were used when all collections were combined (Table S11).

Genetic similarities within and between hatchery and natural-origin collections

I used a series of principal component analyses (PCAs) to provide a general understanding of the genetic similarities among the OUs, among the hatchery collections, and between the hatchery- and natural-origin samples (Figures 9, 10). With the exception of the Soos Creek EWH collection, the EWH programs (Tokul Creek, Marblemount, and Kendall Creek) showed nearly identical 90% confidence ellipse for PC1 and PC2 (Figure 9) and low F_{ST} values (mean pairwise $F_{ST} = 0.006$), suggesting that these hatchery programs have retained genetic similarity due to their Chambers Creek common ancestry. Likewise, the two ESH programs (Soos Creek and Reiter Ponds) also showed similar 90% confidence ellipse for PC1 and PC2 (Figure 9) and low F_{ST} values (pairwise $F_{ST} = 0.008$), showing their retained genetic similarity from Skamania Hatchery (Washougal River) common ancestry. The close genetic similarity among the three EWH programs and between the two ESH programs supports my use of the out-of-basin Reiter Ponds data for the Nooksack, Skagit, and Stillaguamish analyses, and the out-of-basin Tokul Creek data for the Stillaguamish analysis. As shown in Figure 9, and later discovered from the *Structure* analyses, the Soos Creek EWH collection is composed approximately one-third of ESH individuals.

There appears to be little structure among the natural-origin individuals from all river basins. Although the 90% confidence ellipse is broader than those for the hatchery collections, most natural-origin individuals are concentrated toward the middle of Figure 9. There is also overlap between the natural-origin ellipse and each of the hatchery ellipses, many natural-origin individuals occurring within the hatchery ellipses, and no overlap between the EWH and ESH ellipses. This suggests that some individuals are genetically more similarity to one or more of the hatchery collections than they are to the aggregate natural-origin collection, although for the most part there is genetic differentiation between the natural- and hatchery-origin individuals. The similarity between some natural-origin individuals and a hatchery collection is not uniform across all river basins (Figure 10). For the Skagit River and especially Nooksack River there is clear separation between nearly all the natural-origin individuals and both EWH and ESH. This contrasts sharply with the Snohomish River where there appears to be two distinct, but loose clusters, with considerable overlap between the natural-origin fish and both EWH and ESH

(Figure 10). Here, PC1 appears to divide the collections into an ESH (i.e., lower Colombia ancestry) group to the right and an EWH (Puget Sound ancestry) group to the left. The difference between the Snohomish River, on one hand, and the Skagit and Nooksack Rivers, on the other hand, suggests that there has been greater genetic interaction between wild and hatchery-origin fish in the Snohomish River than in either the Skagit or Nooksack Rivers. The Stillaguamish River appears intermediate between the Snohomish River, and the Skagit and Nooksack Rivers. Finally, the Green River is difficult to interpret for two reasons. First, as discussed above, the Soos Creek EWH collection is composed partially of ESH individuals. Second, there is a tight cluster of individuals that overlaps with that portion of the EWH ellipse opposite to the ESH ellipse (more negative on PC 1) and a loose set of individuals with more extreme negative values on PC2. This suggests that they may be structure among the Green River natural-origin collections, with one group showing genetic interaction with the EWH program.

Structure and likelihood analyses: Composition of natural-origin collections

Of the 198 likelihood adjustments of hatchery-lineage, hybrid, and wild proportions from *Structure* across all 33 OUs nine showed 90% CI ranges greater than 0.25 (Tables S13, S15, S17, S19, S21). I consider uncertain these adjusted proportions, the original *Structure* proportions, and all PEHC values that made use of these proportions. The source category with the most number of uncertain estimates was EWH-Wild hybrid, with five of the nine broad 90% CI ranges, or 15% of all EWH-Wild hybrid estimates. This category also had the highest mean 90% CI range (0.14), compared with a range of 0.03, 0.06, 0.03, 0.10, and 0.01 for EWH-lineage, ESH-Wild hybrid, ESH-lineage, Wild, and EWH-ESH hybrid categories, respectively. The CC-NPS hybrid category for the pre-hatchery phase simulated populations, which represented the EWH-Wild hybrid category, had the highest assignment error rate (0.37) among all the other categories (Table 1). With or without the likelihood adjustment, assignments to the EWH-Wild hybrid category are the most uncertain among all assignments. There were no DIP proportions with 90% CI ranges greater than 0.25, although the Samish River OU/DIP had a range of 0.24 for the EWH-Wild hybrid category.

F1 Hybrid (introgression) and PEHC values for all DIPs are in Table 8 and for OUs in Tables S14, S16, S18, S20, and S22. Since there were no DIP proportions with 90% CI ranges greater than 0.25, there were no introgression and PEHC values that I treated as uncertain, although there were five $PEHC_W$ 90% CI ranges that exceeded 0.10, and one range (Samish River) was 0.17 (Table 8). There were two $PEHC_S$ 90% CI ranges that exceeded 0.10, and one range (Tolt River Summer-Run) was 0.21 (Table 8). All but two each of the original unadjusted *Structure*-based $PEHC_W$ and $PEHC_S$ were within the 90% CI of their adjusted PEHC, although many were only marginally within that interval (i.e., at one or the other extreme value of the range). The likelihood adjustment resulted in a decrease in most (11 of 15) $PEHC_W$ estimates from the original unadjusted *Structure*-based $PEHC_W$, while the remaining four of the 15 $PEHC_W$ remained unchanged. Seven of the 15 $PEHC_S$ estimates remained unchanged from the original unadjusted *Structure*-based $PEHC_S$, while six decreased and two increased (North Fork Skykomish Summer-Run and Tolt River Summer-Run; see below) (Table 8).

In what follows I provide more detailed results and discussion for each of the river basins.

Green River. I genotyped 162 samples from four collections, aggregated into three OUs and one DIP (Tables S9, S13, S14). The collections are not geographically diverse, with both adult collections sampled at Soos Creek Hatchery (but not part of the segregated hatchery programs), and the juvenile collections from a single smolt trap just upriver from the hatchery. The two adult OUs were composed mostly of wild fish, but the juvenile OU was genetically affected by both the EWH and ESH programs, with 9% F1 $Hybrid_W$, 15% $PEHC_W$, and 2% $PEHC_S$. (Table S14). The 2004 adult collection appears to have been influenced more by the ESH program than the EWH program (Table S14), although $PEHC_W$ for the DIP as a whole was 6%, compared with 1% for $PEHC_S$ (Table 8). If hatchery-lineage or F1 hybrids are at a selective disadvantage compared with wild fish, you would expect to find higher F1 hybrid and PEHC values among younger age classes than for the adults. The higher F1 $Hybrid_W$ and $PEHC_W$ values for the juvenile OU, compared with the adult OUs, may reflect a higher mortality rate for fish with EWH ancestry than those with wild ancestry, although these data where not collected to test this hypothesis, and with this specific data set I am not comparing juveniles and adults from the same cohort.

Snohomish River. The Snohomish River provided more natural-origin samples (392), OUs (9), and DIPs (5), than any other river basin (Table S9). Samples were collected from the Skykomish, Snoqualmie, and Pilchuck rivers, tributaries of the Snohomish River. Samples were also taken from geographic areas presumed to have either winter-run or summer-run populations. EWH (Tokul Creek, Snoqualmie River) and ESH (Reiter Ponds, Skykomish River) programs are present in the Snohomish River. The steelhead collections from the Snohomish River present a complex association of genetic relationships with individuals from putative native summer-runs resembling each other, or individuals from native winter-run, EWH, ESH, lower Columbia River (but not ESH), or no other group (summarized in Figure 18). The unadjusted proportions from *Structure*, and the likelihood adjusted proportions for each OU and DIP are shown in Table S15 and the F1 Hybrid and PEHC values are in Table 8, S16. The North Fork Skykomish summer-run OUs and DIP were dominated by ESH (lower Columbia River) lineage fish, with 77%-100% identified as pure ESH-lineage. Wild fish were not present in two of the three OUs, and the third OU consisted of only 8% wild individuals. Considering all samples from the North Fork Skykomish DIP, the $PEHC_S$ was 95%, while the $PEHC_W$ was only 1%. This DIP appears to be mostly a feral population that originated from lower Columbia hatchery transplants from the Reiter Ponds program.

The other summer-run DIP in the Snohomish is from the Tolt River, a tributary to the Snoqualmie River, and an area to which Reiter Ponds ESH smolts were released through the 2008 release year (average number of annual releases 2003-2008 = 51,566 smolts, WDFW unpublished data). Four juvenile collections from the Tolt River were genotyped, each representing their own OU: North Folk and South Fork Tolt, and above and below putative barriers to winter-run steelhead spawning. The Snoqualmie River Winter-run DIP includes winter-run fish from the Tolt River. Therefore, to form this DIP, I aggregated the NF and SF Tolt below barrier OUs with the Snoqualmie River Winter-run OU. The North Fork Tolt River – above barrier OU was not aggregated into the Tolt River Summer-run DIP because its results were uncertain (see below), and therefore, the Tolt River Summer-run DIP was composed

entirely of the South Fork Tolt – above barrier OU. There was a high occurrence of ESH-lineage fish in the Tolt system, in particular, in the above-barriers OUs, and especially in the South Fork (Table S15). For the South Fork Tolt River – above barrier OU/DIP, 94% of the collection was composed of ESH fish, split roughly half each between pure ESH-lineage and ESH-Wild F1 hybrids. This resulted in a 69% $PEHC_S$, and 51% F1 $Hybrid_S$ (Tables 8, S16). There was also evidence for winter-run wild and EWH-Wild F1 hybrids in the South Fork Tolt, above the putative barrier. Like the North Fork Skykomish Summer-Run DIP, the South Fork Tolt River – above barrier OU/DIP appears to have a large feral population presumably derived from releases from Reiter Ponds ESH program.

The North Fork Tolt River – above barrier also had a large ESH component. However, there are two problems with the results from the North Fork Tolt River – above barrier OU. First, the OU is composed of only 18 samples. Second, the likelihood adjustment to the *Structure* proportions for half the assignment categories had broad confidence intervals, which makes the adjustments uncertain (Table S15). But, the adjustment to the pure ESH-lineage proportion was not uncertain and was equal to 24%. The OUs below the barriers in the North and South Fork Tolt rivers had higher proportions of wild fish, 97% in the North Fork, and 55% in the South Fork. The remaining proportion from the North Fork (3%) was ESH-lineage fish, while South Fork showed a mix of hatchery-wild hybrids (21%) and ESH-lineage fish (24%).

In addition to the Tolt River – below barrier OUs, the Snoqualmie River Winter-Run DIP also included adult samples from the mainstem Snoqualmie River, represented by a single OU. The Snoqualmie River OU was composed mostly (90%) of wild fish, with an additional 2% EWH-lineage. The original *Structure* EWH-wild hybrid proportion was 17%; however, the likelihood adjustment to that proportion (8%) had a 90% CI range greater than 0.25, making this adjustment and the original *Structure* proportion uncertain. Snoqualmie River Winter-Run DIP, as a whole, was affected nearly equally by the EWH and ESH programs with 4% $PEHC_W$ and 3% $PEHC_S$, while F1 $Hybrid_W$ was 6% (Table 8).

The remaining two OUs and DIPs in the Snohomish River have similar genetic composition. Both the Skykomish Winter adult OU (representing the Snohomish / Skykomish R Winter-Run DIP), and the Pilchuck River OU/DIP were composed mostly of wild fish (95% for both OUs), followed by ESH-lineage fish (5% and 3%, respectively). The Pilchuck River also had 1% EWH-lineage fish. There were no hybrid fish in either OU/DIP.

In summary, the Reiter ESH program appears to have had a larger contribution than the Tokul Creek EWH program to the genetic structure of the natural-origin populations in the Skykomish and Snoqualmie Rivers (and their tributaries), and in the Pilchuck River. All DIPs had $PEHC_S$ values greater than zero, ranging from 3 – 95%, while $PEHC_W$ values ranged 0 – 4%, with three DIPs having values equal to 1%. F1 $Hybrid_S$ was limited to the Tolt River DIP (51%), while the largest F1 $Hybrid_W$ value was 6% for the Snoqualmie River winter-run DIP.

Stillaguamish River. The Stillaguamish River was the poorest represented system in terms of samples available for this project. In addition to having no hatchery-origin samples from the Stillaguamish hatchery programs (Tokul Creek and Reiter Ponds were used as EWH and ESH

surrogates), samples from a known wild winter-run collection currently do not exist. All natural-origin samples were summer-run samples from Canyon and Deer Creek, or smolt samples from a trap in the mainstem below the confluence of the North and South Forks, which could contain samples from multiple DIPs. Finally, except for one adult fish sampled in 2012 and seven in 2013 from Deer Creek, all remaining samples (255) were taken from juveniles or smolts (Tables 7, S9). Since the smolt trap collected samples from a variety of locations, I did not aggregate the results from this OU in either the Canyon Creek or Deer Creek DIPs. Although all smolts from the trap included in the OU were unmarked, 14% assigned to ESH-lineage, and 9% assigned as ESH-wild hybrids (Table S17). The remaining samples (76%) assigned as wild. Both the Canyon and Deer creek DIPs were all or mostly composed of summer-run wild fish (100% and 97%, respectively), with the remaining 3% from Deer Creek assigned as ESH-wild hybrids. Based on the limited samples available, there were no EWH-lineage fish or EWH-wild hybrids in any of the OUs, and the only hatchery presence in the system was from the ESH program affecting both the smolt and Deer Creek OUs (Table 8, S18).

Skagit River. The Skagit River samples (N = 333) were composed of 15 collections, aggregated into eight OUs and three DIPs. There was one summer-run OU (Finney Creek summers), and all samples, except from the Nookachamps OU, were from adults (Tables 7, S9). Compared with the Snohomish River, natural-origin samples from the Skagit system showed less influence from either the Marblemount EWH program, or an out-of-basin ESH program (Tables 8, S19, S20, Figure 10). Roughly 5% of the Finney Creek summer OU assigned to ESH-lineage, despite the fact that there are no ESH programs in the Skagit River. If ESH-origin fish are getting into the Skagit basin, they are most-likely straying from either the Whitehorse (Stillaguamish) or Reiter Ponds (Snohomish) ESH programs. The Finney Creek summer OU also showed evidence of EWH influence, with 3% assigned to EWH-lineage, and 7% as EWH-wild hybrids. However, the 90% CI range for the likelihood adjustment to the hybrid and wild categories was greater than 0.25, and therefore, I consider these assignment proportions uncertain. By contrast, the Finney Creek winter OU assigned as 98% wild and 2% EWH-lineage.

Unadjusted proportions from *Structure* showed that 15% of the Cascade OU was composed of EWH-wild hybrids, but zero contribution from EWH-lineage fish. As with the samples from Finney Creek summer OU, the 90% CI range for the likelihood adjustment to the ESH-Wild category was greater than 0.25, and therefore, I consider these assignment proportions uncertain. However, the point estimate for the adjusted EWH-wild hybrids assignment was 6% (Table S19), which, if correct, would be largest hatchery effect for all OUs in the Skagit. The Finney Creek and Cascade OUs were aggregated with the upper Skagit River OUs forming the Mainstem Skagit R Summer- and Winter-Run DIP. This DIP was composed of 96% wild fish, and 2% EWH- and 1% ESH-lineage fish. Because the samples sizes from the Cascade and Finney Creek summer OUs were small compared with that from the upper Skagit OU (Table S9), and since both OUs are a small component of the total spawning population in the DIP, (Table S12), when aggregated with the upper Skagit OU, the DIP showed no F1 $Hybrid_W$ introgression, but 2% $PEHC_W$ and 1% $PEHC_S$ (Table 8).

The Sauk R Summer- and Winter-Run DIP (aggregated from Sauk and Suiattle OUs) and Nookachamps River OU/DIP were both composed mostly of wild fish (96% and 98%,

respectively), with 4% and 2% of the fish assigned as EWH-lineage, respectively. Neither DIP showed evidence of hybridization, so the $PEHC_W$ value for both DIPs reflected the EWH-lineage proportion only (Table 8).

In summary, samples from the Skagit River showed evidence of hatchery influence around 2 – 4% $PEHC_W$, with the contribution from both EWH-lineage fish and EWH-wild hybrids. The areas with the largest hatchery influence were Finney Creek (summer population), and Cascade River, the location of the Marblemount EWH program. There was also evidence that ESH-origin fish strayed into Finney Creek and the upper Skagit River. These results are consistent with an earlier analysis of hatchery-wild introgression in the Skagit River (Warheit 2013). The earlier work was conducted with a limited set of microsatellite loci and used unadjusted (and therefore biased) *Structure* proportions only. In that report's conclusions I state "The SPAN microsatellite loci lack sufficient power to reliably quantify Marblemount Hatchery (Chambers Creek-origin) introgression into the wild Skagit River winter steelhead populations, or reliably identify pure unmarked hatchery or hatchery-ancestry fish using the program STRUCTURE. However, under some reasonable assumptions, the Finney Creek adult and juvenile populations appeared to have a higher level of hatchery-wild introgression than all other wild populations" (Warheit 2013:119). In this earlier study, the *Structure* analyses did not included samples from an ESH program or from the Finney Creek summer OU.

Nooksack and Samish rivers. I included in the Nooksack analyses the Samish River OU/DIP because until 2009 EWH smolts from Kendall Creek (North Fork Nooksack River) were released into the Samish River (WDFW unpublished data). In addition to the 84 Samish River OU samples, the Nooksack River analyses included 283 samples from 10 collections aggregated into six OUs and two DIPs (Tables 7, S9). As with the Skagit River analysis, the Nooksack River appears relatively devoid of hatchery influence (Tables 8, S21, S22; Figure 10). Three of the five Nooksack winter-run OUs were composed entirely of wild fish. The remaining two OUs were composed of at least 95% wild fish, with remaining fish assigning as EWH-lineage, ESH-wild hybrids, or EWH-ESH hybrids (Table S21). However, when all five OUs are aggregated into a single DIP (Nooksack R Winter-Run), there is no hatchery signal, with zero F1 $Hybrids_W$ and zero $PEHC_W$ (Table 8). The South Fork Nooksack summer-run OU/DIP assigned as wild individuals, with no influence from either the EWH program or an out-of-basin ESH program. As discussed above, the Samish River OU/DIP was composed of 11% EWH-wild hybrids and 1% EWH-lineage fish. This equates to 6% $PEHC_W$ (Tables 8, S21, S22).

CONCLUSIONS

Based on samples used in these analyses, segregated steelhead hatchery programs in the Green, Snohomish, Stillaguamish, Skagit, and Nooksack River basins have affected the genetic structure of the natural-origin populations in those rivers. However, this effect varied among the river systems, and is best summarized in Figure 10. The Stillaguamish, Green, and especially the Snohomish basins showed higher levels of F1 hybridization (introgression) and PEHC than the Skagit and Nooksack Rivers. But, even within these more heavily affected rivers, there was spatial and temporal variation to the hatchery influence. For example, there was larger hatchery signal among the smolt collections in the Green River than among the adult collections, and in

the Stillaguamish River there was small to no hatchery influence among the juvenile summer-run OUs, but a large ESH effect on the smolt collection. In the Snohomish River, the Reiter Ponds ESH program affected the natural-origin population more than the Tokul Creek EWH program, and may have resulted in extensive feral summer-run populations in the North Fork Skykomish and the Tolt Rivers.

Acknowledgements

I thank Sarah Bell (WDFW-MGL) for all laboratory work associated with this project, and Todd Kassler (WDFW-MGL) for overall project management and sample curation. I also thank all members of the WDFW-MGL, in particular Todd Seamons, for discussions that helped focused this project. Todd Seamons, Maureen Small (WDFW-MGL), and Craig Busack (NOAA Fisheries) read several version of this report and provided helpful comments and edits. Mike Ford, Jeff Hard, Paul Moran, Jim Myers, Robin Waples, and Gary Winans (all NOAA Fisheries) provided many helpful comments on an earlier version of this report, and through those comments help shaped this current version. Todd Seamons also conducted the Colony analyses for this project. Shannon Knapp (formerly WDFW, now New Mexico State University) designed the first version of the likelihood method to adjust *Structure* proportions. For sample collection I thank the field staffs from Seattle City Light and WDFW's Region 4, especially Dave Pflug, Ed Conner, and Liz Ablow (SCL), Brett Barkdull and Jenni Whitney (WDFW), and also Ned Currence (Nooksack Indian tribe). This technical report is part of a larger effort to develop Hatchery Genetic Management Plans (HGMPs) for Puget Sound steelhead hatchery programs. The development of many of the concepts in this report was part of a coordinated effort among Jim Scott, Annette Hoffmann, Kelly Cunningham, Brian Missildine, and Beata Dymowska. Finally, this project was funded in part by Seattle City Light, and the Washington General and Wildlife Funds.

Literature Cited

Abadia-Cardoso, A., A. J. Clemento, and J. C. Garza. 2011. Discovery and characterization of single-nucleotide polymorphisms in steelhead/rainbow trout, Oncorhynchus mykiss. Molecular Ecology Resources **11**:31-49.

Aguliar, A., and J. C. Garza. 2008. Isolation of 15 single nucleotide polymorphisms from coastal steelhead, *Oncorhynchus mykiss* (Salmonidae). Molecular Ecology Resources **8**:659-662.

Anderson, E. C., and K. K. Dunham. 2008. The influence of family groups on inferences made with the program Structure. Molecular Ecology Resources **8**:1219-1229.

Araki, H., B. A. Berejikian, M. J. Ford, and M. S. Blouin. 2008. Fitness of hatchery-reared salmonids in the wild. Evolutionary Applications **1**:342-355.

Baskett, M. L., and R. S. Waples. 2013. Evaluating Alternative Strategies for Minimizing Unintended Fitness Consequences of Cultured Individuals on Wild Populations. Conservation Biology **27**:83-94.

Brunelli, J. P., G. H. Thorgaard, R. F. Leary, and J. L. Dunnigan. 2008. Single-nucleotide polymorphisms associated with allozyme differences between inland and coastal rainbow trout. Transactions of the American Fisheries Society **137**:1292-1298.

Campbell, N. R., and S. R. Narum. 2009. Identification and characterization ofheat shock response–related single-nucleotide polymorphisms in *O. mykiss* and *O. tshawytscha*. Molecular Ecology Resources **9**:1460-1559.

Crawford, B. A. 1979. The origin and history of the trout brood stocks of the Washington Department of Game. Olympia, WA.

Crow, J. F., and M. Kimura. 1970. An Introduction to Population Genetics Theory. Harper and Row, New York.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567-1587.

Finger, A., M. R. Stephens, N. W. Clipperton, and B. May. 2009. Six diagnostic single nucleotide polymorphism markers for detecting introgression between cutthroat and rainbow trouts. Molecular Ecology Resources **9**:759-763.

Ford, M. J. 2002. Selection in captivity during supportive breeding may reduce fitness in the wild. Conservation Biology **16**:815-825.

Garza, J. C., E. A. Gilbert-Horvath, B. C. Spence, T. H. Williams, H. Fish, S. A. Gough, J. H. Anderson, D. Hamm, and E. C. Anderson. 2014. Population Structure of Steelhead in Coastal California. Transactions of the American Fisheries Society **143**:134-152.

Gayeski, N., B. McMillian, and P. Trotter. 2011. Historical abundance of Puget Sound steelhead, *Oncorhynchus mykiss*, estimated from catch record data. Canadian Journal of Fisheries and Aquatic Sciences **68**:498-510.

Gilbert, K. J., R. L. Andrew, D. G. Bock, M. T. Franklin, N. C. Kane, J. S. Moore, B. T. Moyers, S. Renaut, D. J. Rennison, T. Veen, and T. H. Vines. 2012. Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. Molecular Ecology **21**:4925-4930.

Hansen, M. H. H., S. Young, H. B. H. Jorgensen, C. Pascal, M. Henryon, and J. Seeb. 2011. Assembling a dual purpose TaqMan-based panel of single-nucleotide polymorphism

markers in rainbow trout and steelhead (Oncorhynchus mykiss) for association mapping and population genetics analysis. Molecular Ecology Resources **11**:67-70.
Harbicht, A. B., M. Alshamlih, C. C. Wilson, and D. J. Fraser. 2014. Anthropogenic and habitat correlates of hybridization between hatchery and wild brook trout. Canadian Journal of Fisheries and Aquatic Sciences **71**:688-697.
Hubisz, M., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weal population structure with the assistance of sample group information. Molecular Ecology Resources **9**:1322-1332.
Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18**:337-338.
Kalinowski, S. T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity **106**:625-632.
Kidd, A. G., J. Bowman, D. Lesbarrères, and A. I. Schulte-Hostedde. 2009. Hybridization between escaped domestic and wild American mink (Neovison vison). Molecular Ecology **18**:1175-1186.
Lamaze, F. C., C. Sauvage, A. Marie, D. Garant, and L. Bernatchez. 2012. Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (Salvelinus fontinalis). Molecular Ecology **21**:2877-2895.
Limborg, M. T., S. M. Blankenship, S. F. Young, F. M. Utter, L. W. Seeb, M. H. H. Hansen, and J. E. Seeb. 2012. Signatures of natural selection among lineages and habitats in Oncorhynchus mykiss. Ecology and Evolution **2**:1-18.
Marie, A. D., L. Bernatchez, and D. Garant. 2011. Empirical assessment of software efficiency and accuracy to detect introgression under variable stocking scenarios in brook charr (Salvelinus fontinalis). Conservation Genetics **12**:1215-1227.
MathWorks. 2012. Matlab R2012B. Natick, MA.
McGlaufin, M., M. J. Smith, J. T. Wang, S. F. Young, N. Chen, Y. C. Lee, C. Pascal, L. W. Seeb, J. Steven, and J. E. Seeb. 2010. High resolution melting analysis for the discovery of novel SNPs in rainbow (*Oncorhynchus mykiss*) and cutthroat trout (*Oncorhynchus clarki*) for species identification. Transactions of the American Fisheries Society **139**:676-684.
Mosher, D. C., and A. T. Hewitt. 2004. Late Quaternary deglaciation and sea-level history of eastern Juan de Fuca Strait, Cascadia. Quaternary International **121**:23-39.
Norén, K., L. Dalén, K. Kvaløy, and A. Angerbjörn. 2005. Detection of farm fox and hybrid genotypes among wild arctic foxes in Scandinavia. Conservation Genetics **6**:885-894.
Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.
Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - An update. Bioinformatics **28**:2537-2539.
Porter, S. C., and T. W. Swanson. 1998. Radiocarbon age contraints on rates of advance and retreat of the Puget Lobe of the Cordilleran Ice Sheet during the last glaciation. Quaternary Research **50**:205-213.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.
Pritchard, J. K., X. Wen, and D. Falush. 2010. Documentation for *structure* software: Version 2.3.

PSSTRT. 2013. Identifying historical populations of steelhead within the Puget Sound Distinct Population Segment. Final Review Draft.
Quinn, T. P. 2005. The Behavior and Ecology of Pacific Salmon and Trout. University of Washington, Seattle.
Sánchez, C. C., T. P. L. Smith, R. T. Wiedmann, R. L. Vallejo, M. Salem, J. Yao, and C. E. Rexroad. 2009. Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. Bmc Genomics **10**:559.
Sanz, N., R. M. Araguas, R. Fernández, M. Vera, and J. L. Garcia-Marin. 2009. Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. Conservation Genetics **10**:225-236.
Seamons, T. R., L. Hauser, K. A. Naish, and T. P. Quinn. 2012. Can interbreeding of wild and artificially propagated animals be prevented by using broodstock selected for a divergent life history? Evolutionary Applications **5**:705-719.
Sprowles, A. E., M. R. Stephens, N. W. Clipperton, and B. May. 2006. Fishing for SNPs: a targeted locus approach for Single Nucleotide Polymorphism discovery in rainbow trout. Transactions of the American Fisheries Society **135**:1698-1721.
Stephens, M. R., N. W. Clipperton, and B. May. 2009. Subspecies-informative SNP assays for evaluating introgression between native golden trout and introduced rainbow trout. Molecular Ecology Resources **9**:339-343.
Vähä, J.-P., and C. R. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology **15**:63-72.
Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics **166**:1963-1979.
Wang, J. 2012. Computationally efficient sibship and parentage assignment from multilocus marker data. Genetics **191**:183-194.
Wang, J. 2013. An improvement on the maximum likelihood reconstruction of pedigrees from marker data. Heredity **111**:165-174.
Wang, J., and A. W. Santure. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. Genetics **181**:1579-1594.
Warheit, K. I. 2013. Introgressive Hybridization. Pages 88-119 *in* D. Pflug, editor. Ecological, Genetic and Productivity Consequences of Interactions between Hatchery and Natural-Origin Steelhead of the Skagit Watershed. Report to Saltonstall-Kennedy Grant Program (NOAA).
Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.

Tables

Table 1. Distribution of assignments from *Structure* and associated assignment error rates for the pre-hatchery phase simulated populations, using *Structure's* default no prior population information mode. The *Structure* assigned categories are the rows and the source categories are the columns. Each source category consists of 1000 individuals; 100 individuals each from the ten model iterations. Chambers Creek, Lower Columbia, and North Puget Sound simulated populations represent the early winter hatchery (EWH), early summer hatchery (ESH), and north Puget Sound (Wild) populations that existed just prior to the beginning of the hatchery programs in the 1950s (point "C" in Figure 1). The overall error rate (OER) is 0.22.

	Source Category								
Assigned Category	Chambers Creek (CC)	Hybrid: CC - NPS	Hybrid: LC - NPS	Lower Columbia (LC)	North Puget Sound (NPS)	Hybrid: –CC - LC	Total Assigned	Total that should be assigned minus No Call	Assignment Error Rate
Chambers Creek (CC)	**762**	147	4	0	2	45	960	988	0.21
Hybrid: CC - NPS	164	**659**	25	0	164	28	1040	953	0.37
Hybrid: LC - NPS	0	12	**644**	48	60	75	839	856	0.23
Lower Columbia (LC)	0	0	73	**924**	0	74	1071	998	0.14
North Puget Sound (NPS)	1	118	45	0	**762**	2	928	988	0.18
Hybrid: CC - LC	61	17	65	26	0	**621**	790	845	0.21
No Call	12	47	144	2	12	155	372	-	-
Total Source	1000	1000	1000	1000	1000	1000	6000	-	0.22
No Call Rate	0.01	0.05	0.14	0.00	0.01	0.16	0.06		
Source Error Rate Total	0.24	0.34	0.36	0.08	0.24	0.38	0.27		
Source Error Rate Assigned Only	0.23	0.31	0.25	0.07	0.23	0.27	0.22		

Table 2. Distribution of assignments from *Structure* and associated assignment error rates for the pre-hatchery phase simulated populations, using *Structure's* prior population information mode. The *Structure* assigned categories are the rows and the source categories are the columns. Each source category consists of 1000 individuals; 100 individuals each from the ten model iterations. Chambers Creek, Lower Columbia, and North Puget Sound simulated populations represent the early winter hatchery (EWH), early summer hatchery (ESH), and north Puget Sound (Wild) populations that existed just prior to the beginning of the hatchery programs in the 1950s (point "C" in Figure 1). The overall error rate (OER) is 0.09.

	Source Category								
Assigned Category	Chambers Creek (CC)	Hybrid: CC - NPS	Hybrid: LC - NPS	Lower Columbia (LC)	North Puget Sound (NPS)	Hybrid: –CC - LC	Total Assigned	Total that should be assigned minus No Call	Assignment Error Rate
Chambers Creek (CC)	**991**	46	0	0	0	16	1053	1000	0.06
Hybrid: CC - NPS	4	**824**	28	0	4	46	906	919	0.09
Hybrid: LC - NPS	0	7	**599**	2	3	52	663	785	0.10
Lower Columbia (LC)	0	0	95	**995**	0	97	1187	1000	0.16
North Puget Sound (NPS)	0	29	21	0	**993**	1	1044	1000	0.05
Hybrid: CC - LC	5	13	42	3	0	**534**	597	746	0.11
No Call	0	81	215	0	0	254	550	-	-
Total Source	1000	1000	1000	1000	1000	1000	6000	-	0.09
No Call Rate	0.00	0.08	0.22	0.00	0.00	0.25	0.09		
Source Error Rate Total	0.01	0.18	0.40	0.01	0.01	0.47	0.18		
Source Error Rate Assigned Only	0.01	0.10	0.24	0.01	0.01	0.28	0.09		

Table 3. Description of the 27 hatchery phase natural-origin collections simulated using schema in Figure S1 and parameters in Table S6. Percent overlap is the portion of each population (Wild, EWH, ESH) that interacted reproductively with the other populations. N is sample size of the collection, and the columns under Pedigree describe the composition of each collection after 12 simulated generations.

Simulated Population	Percent Overlap			N	Pedigree				
	Wild	EWH	ESH		EWH Lineage	Hybrid: EWH-Wild	Hybrid: EWS-Wild	ESH Lineage	Wild
Sim0.1_0.1_0.1_0.87551	0.1	0.1	0.1	59	0.03	0.00	0.00	0.02	0.95
Sim0.1_0.1_0.1_0.87994	0.1	0.1	0.1	72	0.00	0.00	0.00	0.00	1.00
Sim0.1_0.1_0.1_0.88945	0.1	0.1	0.1	70	0.00	0.00	0.00	0.00	1.00
Sim0.1_0.5_0.1_0.25404	0.1	0.5	0.1	62	0.03	0.02	0.02	0.00	0.94
Sim0.1_0.5_0.1_0.26156	0.1	0.5	0.1	71	0.00	0.03	0.00	0.00	0.97
Sim0.1_0.5_0.1_0.27494	0.1	0.5	0.1	71	0.00	0.01	0.00	0.00	0.99
Sim0.1_1.0_0.1_0.23617	0.1	1	0.1	70	0.06	0.01	0.00	0.01	0.91
Sim0.1_1.0_0.1_0.93021	0.1	1	0.1	72	0.00	0.03	0.00	0.00	0.97
Sim0.1_1.0_0.1_0.93805	0.1	1	0.1	75	0.01	0.01	0.00	0.00	0.97
Sim0.5_0.1_0.1_7257	0.5	0.1	0.1	73	0.00	0.01	0.01	0.00	0.97
Sim0.5_0.1_0.1_73675	0.5	0.1	0.1	74	0.00	0.00	0.00	0.00	1.00
Sim0.5_0.1_0.1_74815	0.5	0.1	0.1	66	0.00	0.00	0.03	0.00	0.97
Sim0.5_0.5_0.1_78476	0.5	0.5	0.1	70	0.14	0.01	0.00	0.00	0.84
Sim0.5_0.5_0.1_81303	0.5	0.5	0.1	71	0.00	0.00	0.00	0.00	1.00
Sim0.5_0.5_0.1_83379	0.5	0.5	0.1	74	0.00	0.00	0.00	0.00	1.00
Sim0.5_1.0_0.1_29295	0.5	1	0.1	65	0.23	0.02	0.02	0.02	0.72
Sim0.5_1.0_0.1_30511	0.5	1	0.1	70	0.07	0.00	0.00	0.00	0.93
Sim0.5_1.0_0.1_31937	0.5	1	0.1	73	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.1_0.1_34109	1	0.1	0.1	68	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.1_0.1_34805	1	0.1	0.1	71	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.1_0.1_35596	1	0.1	0.1	70	0.01	0.00	0.00	0.00	0.99
Sim1.0_0.5_0.1_36349	1	0.5	0.1	64	0.14	0.02	0.00	0.02	0.83
Sim1.0_0.5_0.1_37503	1	0.5	0.1	74	0.00	0.00	0.00	0.01	0.99
Sim1.0_0.5_0.1_38213	1	0.5	0.1	70	0.01	0.01	0.00	0.00	0.97
Sim1.0_1.0_0.1_38892	1	1	0.1	63	0.17	0.05	0.00	0.02	0.76
Sim1.0_1.0_0.1_39511	1	1	0.1	72	0.00	0.00	0.00	0.00	1.00
Sim1.0_1.0_0.1_4006	1	1	0.1	73	0.00	0.03	0.00	0.00	0.97
Mean					0.03	0.01	0.00	0.00	0.95

Table 4. True, unadjusted, and likelihood adjusted PEHC for each of the 27 hatchery phase simulated natural-origin collections, using *Structure's* default no prior and prior population information modes. The True PEHC are based on the pedigree proportions in Table 3. The No Priors and Priors PEHC are based on the *Structure* proportions in Supplemental Tables S7 and S8, respectively. The unadjusted PEHC are calculations based on *Structure* assignments only (Section 1). PEHC values in bold typeface are for those collections where one or more likelihood adjusted proportions were uncertain due to large confidence intervals in the adjustment (see Figure 7 and Supplemental Table S7).

Simulated Population	N	TRUE		*Structure:* No Priors				*Structure:* Priors			
				Unadjusted		Adjusted		Unadjusted		Adjusted	
		PEHCw	PEHCs	PEHCw	PEHCs	PEHCw	PEHCs	PEHCw	PEHCs	PEHCw	PEHCs
Sim0.1_0.1_0.1_0.87551	59	0.034	0.017	0.136	0.034	0.107	0.043	0.025	0.025	0.028	0.031
Sim0.1_0.1_0.1_0.87994	72	0.000	0.000	0.035	0.035	0.030	0.014	0.007	0.000	0.000	0.000
Sim0.1_0.1_0.1_0.88945	70	0.000	0.000	0.071	0.014	0.015	0.015	0.007	0.000	0.000	0.000
Sim0.1_0.5_0.1_0.25404	62	0.040	0.008	0.065	0.024	0.028	0.020	0.024	0.008	0.021	0.000
Sim0.1_0.5_0.1_0.26156	71	0.014	0.000	0.129	0.021	**0.073**	0.001	0.000	0.000	0.000	0.000
Sim0.1_0.5_0.1_0.27494	71	0.007	0.000	0.143	0.000	0.072	0.000	0.007	0.000	0.000	0.000
Sim0.1_1.0_0.1_0.23617	70	0.064	0.014	0.109	0.029	0.095	0.021	0.057	0.014	0.057	0.015
Sim0.1_1.0_0.1_0.93021	72	0.014	0.000	0.063	0.069	0.021	0.053	0.014	0.000	0.001	0.000
Sim0.1_1.0_0.1_0.93805	75	0.020	0.000	0.041	0.020	0.008	0.018	0.013	0.007	0.017	0.000
Sim0.5_0.1_0.1_7257	73	0.007	0.007	0.063	0.014	0.000	0.000	0.000	0.000	0.000	0.000
Sim0.5_0.1_0.1_73675	74	0.000	0.000	0.122	0.020	0.051	0.018	0.000	0.000	0.000	0.000
Sim0.5_0.1_0.1_74815	66	0.000	0.015	0.023	0.031	0.000	0.018	0.000	0.000	0.000	0.000
Sim0.5_0.5_0.1_78476	70	0.150	0.000	0.243	0.007	0.204	0.000	0.071	0.000	0.060	0.000
Sim0.5_0.5_0.1_81303	71	0.000	0.000	0.125	0.029	**0.093**	0.019	0.000	0.000	0.000	0.000
Sim0.5_0.5_0.1_83379	74	0.000	0.000	0.178	0.000	0.116	0.000	0.007	0.000	0.000	0.000
Sim0.5_1.0_0.1_29295	65	0.238	0.023	0.269	0.077	0.262	0.065	0.185	0.015	0.179	0.020
Sim0.5_1.0_0.1_30511	70	0.071	0.000	0.136	0.021	0.085	0.001	0.029	0.000	0.016	0.000
Sim0.5_1.0_0.1_31937	73	0.000	0.000	0.157	0.021	**0.103**	0.002	0.007	0.000	0.000	0.000
Sim1.0_0.1_0.1_34109	68	0.000	0.000	0.103	0.029	**0.051**	0.020	0.000	0.007	0.000	0.000
Sim1.0_0.1_0.1_34805	71	0.000	0.000	0.085	0.021	0.017	0.000	0.000	0.000	0.000	0.000
Sim1.0_0.1_0.1_35596	70	0.014	0.000	0.123	0.014	0.057	0.000	0.014	0.000	0.001	0.000
Sim1.0_0.5_0.1_36349	64	0.148	0.016	0.195	0.031	0.187	0.018	0.063	0.016	0.057	0.022
Sim1.0_0.5_0.1_37503	74	0.000	0.014	0.056	0.042	0.000	0.049	0.000	0.000	0.000	0.000
Sim1.0_0.5_0.1_38213	70	0.021	0.000	0.064	0.043	0.008	0.018	0.000	0.000	0.000	0.000
Sim1.0_1.0_0.1_38892	63	0.198	0.016	0.336	0.016	0.308	0.019	0.063	0.016	0.061	0.010
Sim1.0_1.0_0.1_39511	72	0.000	0.000	0.064	0.014	0.000	0.000	0.000	0.000	0.000	0.000
Sim1.0_1.0_0.1_4006	73	0.014	0.000	0.082	0.034	0.016	0.034	0.007	0.007	0.000	0.000
Mean		0.04	0.00	0.12	0.03	0.07	0.02	0.02	0.00	0.02	0.00

Table 5. Distribution of assignments from *Structure* and associated assignment error rates for pooled 27 hatchery phase simulated natural-origin collections, using *Structure's* default no prior population information mode. The *Structure* assigned categories are the rows and the source categories are the columns. Source identity is based on individual's pedigree. The overall error rate (OER) is 0.18.

	Source Category								
Assigned Category	EWH Lineage	Hybrid: EWH Lineage - Wild	Hybrid: ESH Lineage - Wild	ESH Lineage	Wild	Hybrid: EWH - ESH	Total Assigned	Total that should be assigned minus No Call	Assignment Error Rate
EWH Lineage	**56**	3	0	0	55	0	114	61	0.51
Hybrid: EWH Lineage - Wild	3	**7**	0	0	199	0	209	18	0.97
Hybrid: ESH Lineage - Wild	0	0	**1**	0	49	0	50	5	0.98
ESH Lineage	0	0	1	**6**	17	0	24	6	0.75
Wild	0	8	3	0	**1454**	0	1465	1778	0.01
Hybrid: EWH - ESH	2	0	0	0	4	**0**	6	0	-
No Call	0	0	0	0	15	0	15	-	-
Total Source	61	18	5	6	1793	0	1883	-	0.18
No Call Rate	0.00	0.00	0.00	0.00	0.01	-	0.01		
Source Error Rate Total	0.08	0.61	0.80	0.00	0.19	-	0.19		
Source Error Rate Assigned Only	0.08	0.61	0.80	0.00	0.18	-	0.18		

Table 6. Distribution of assignments from *Structure* and associated assignment error rates for pooled 27 hatchery phase simulated natural-origin collections, using *Structure's* prior population information mode. The *Structure* assigned categories are the rows and the source categories are the columns. Source identity is based on individual's pedigree. The overall error rate (OER) is 0.04.

	Source Category								
Assigned Category	EWH Lineage	Hybrid: EWH Lineage - Wild	Hybrid: ESH Lineage - Wild	ESH Lineage	Wild	Hybrid: EWH - ESH	Total Assigned	Total that should be assigned minus No Call	Assignment Error Rate
EWH Lineage	**22**	0	0	0	0	0	22	61	0.00
Hybrid: EWH Lineage - Wild	23	**2**	0	0	11	0	36	18	0.94
Hybrid: ESH Lineage - Wild	1	0	**1**	1	4	0	7	5	0.86
ESH Lineage	0	0	0	**4**	0	0	4	6	0.00
Wild	15	16	4	1	**1778**	0	1814	1796	0.02
Hybrid: EWH - ESH	0	0	0	0	0	**0**	0	0	-
No Call	0	0	0	0	0	0	0	-	-
Total Source	61	18	5	6	1793	0	1883	-	0.04
No Call Rate	0.00	0.00	0.00	0.00	0.00	-	0.00		
Source Error Rate Total	0.64	0.89	0.80	0.33	0.01	-	0.04		
Source Error Rate Assigned Only	0.64	0.89	0.80	0.33	0.01	-	0.04		

Table 7. Geographic and temporal scope, and biological and management descriptors of wild (natural-origin or unmarked) and hatchery-origin steelhead collections used in this study. In the Presumed Run Timing column, EWH = early winter hatchery program; ESH = early summer hatchery program.

Basin	Subbasin	Collection Code	Collection Year	Life Stage	Collection Dates	Origin	Presumed Run Timing
Nooksack	Mainstem	11NW	2011	adult	Dec 2010 - Jan 2011	wild	winter
Nooksack	Mainstem	12MP	2012	adult	Dec 2011 - Jan 2012	wild	winter
Nooksack	Mainstem	13GC	2013	adult	Dec 2012 - Jan 2013	wild	winter
Nooksack	Northfork	12MQ	2012	adult	Feb - April	wild	winter
Nooksack	Northfork	09MN	2009	juvenile	Fall	wild	winter
Nooksack	Northfork	10PY	2010	juvenile	unknown	wild	winter
Nooksack	Southfork	12CF	2012	adult	Feb - March	wild	winter
Nooksack	Southfork	09LQ	2012	juvenile	June - August	wild	unknown
Nooksack	Southfork	10GX	2010	adult	Sept - October	wild	summer
Nooksack	Southfork	11GO	2011	adult	August - October	wild	summer
Nooksack	Kendall Creek Hat.	01GA	2001	adult broodstock		hatchery	EWH
Samish	Samish	08BN	2008	adult	Feb - April	wild	winter
Samish	Samish	12AP	2012	adult	Feb - March	wild	winter
Skagit	Cascade	12DA	2012	adult	May	wild	winter
Skagit	Finney Creek	10CQ	2010	adult	March - May	wild	winter
Skagit	Finney Creek	11BK	2011	adult	March - May	wild	winter
Skagit	Finney Creek	12FT	2012	adult	November	wild	summer
Skagit	Suiattle	10AQ	2010	adult	March - April	wild	winter
Skagit	Suiattle	11BM	2011	adult	April	wild	winter
Skagit	upper Skagit	08DQ	2008	adult	Feb - May	wild	winter
Skagit	upper Skagit	09BN	2009	adult	April	wild	winter
Skagit	upper Skagit	10AO	2010	adult	March - May	wild	winter
Skagit	upper Skagit	11BI	2011	adult	April - May	wild	winter
Skagit	upper Skagit	10NI	2010	adult	Nov 2010 - Jan 2011	wild	winter
Skagit	Nookachamps	12AO	2012	Juv. (adult = 2)	March, May	wild	winter
Skagit	Sauk	09DU	2009	adult	March - April	wild	winter
Skagit	Sauk	10AR	2010	adult	Feb - May	wild	winter
Skagit	Sauk	11BN	2011	adult	April - May	wild	winter
Skagit	Marblemount Hat.	08LF	2008	adult broodstock		hatchery	EWH
Skagit	Marblemount Hat.	09CF	2009	adult broodstock		hatchery	EWH
Skagit	Marblemount Hat.	10AN	2010	adult broodstock		hatchery	EWH
Stillaguamish	Canyon Creek	13KA	2013	juvenile	October	wild	summer
Stillaguamish	Deer Creek	95CG	1995	juvenile	unknown	wild	summer
Stillaguamish	Deer Creek	12FL	2012	adult	July	wild	summer
Stillaguamish	Deer Creek	13GE	2013	adult	October	wild	summer
Stillaguamish	Deer Creek	13KB	2013	juvenile	Sept - October	wild	summer
Stillaguamish	mixed	06BY	2006	smolt	unknown	wild	mixed
Snohomish	NF Skykomish	04HN	2004	juvenile	unknown	wild	summer
Snohomish	NF Skykomish	12FK	2012	adult	August - September	wild	summer
Snohomish	NF Skykomish	13GF	2013	adult	July - August	wild	summer
Snohomish	NF Skykomish	13LJ	2013	juvenile	October	wild	summer
Snohomish	Pilchuck River	12MN	2012	adult	Feb - April	wild	winter
Snohomish	Skykomish mainstem	13GH	2013	adult	Feb - April	wild	winter
Snohomish	NF Tolt (Snoqualmie)	11IW	2011	juvenile	September	wild	summer
Snohomish	NF Tolt (Snoqualmie)	12IS	2012	juvenile	September	wild	winter
Snohomish	SF Tolt (Snoqualmie)	10IX	2010	juvenile	September	wild	winter
Snohomish	Snoqualmie	13BC	2013	adult	Feb - April	wild	winter
Snohomish	SF Tolt (Snoqualmie)	10IW	2010	juvenile	September	wild	summer
Snohomish	Reiter Ponds Hat.	01GG	2001	adult broodstock		hatchery	ESH
Snohomish	Tokul Creek Hat.	01GC	2001	adult broodstock		hatchery	EWH
Green	Mainstem	04AY	2004	adult	unknown	wild	winter
Green	Mainstem	07CO	2007	smolt	unknown	wild	winter
Green	Mainstem	08EF	2008	smolt	May - June	wild	winter
Green	Soos Creek	13EH	2013	adult	March - April	wild	winter
Green	Soos Creek	03LZ	2003	adult broodstock		hatchery	EWH
Green	Soos Creek	03MA	2003	adult broodstock		hatchery	ESH

Table 8. Unadjusted and likelihood adjusted F1 Hybrid (introgression) and PEHC values for each of the DIPs. Below the adjusted F1 Hybrid and PEHC values are the 90% confidence intervals. The StillaguamishRSmoltTrap06 OU is presented below with the DIPs because there are no other collections from the Stillaguamish that may contain winter steelhead. See Tables S13 (Green R.), S15 (Snohomish R.), S17 (Stillaguamish R.), S19 (Skagit R.), and S21 (Nooksack R.) for unadjusted and likelihood adjusted proportions for each of the OUs and DIPs, and Tables S14 (Green R.), S16 (Snohomish R.), S18 (Stillaguamish R.), S20 (Skagit R.), and S22 (Nooksack R.) for the unadjusted and likelihood adjusted F1 Hybrids and PEHC values for each of the OUs.

PSSTRT DIP	River Basin	N	Unadjusted				Adjusted			
			F1 Hyb_{winter}	F1 Hyb_{summer}	$PEHC_{winter}$	$PEHC_{summer}$	F1 Hyb_{winter}	F1 Hyb_{summer}	$PEHC_{winter}$	$PEHC_{summer}$
Green River Winter-Run	Green/Duwamish	165	0.13 -	0.00 -	0.13 -	0.01 -	0.00 (0.00 - 0.09)	0.00 (0.00 - 0.00)	0.06 (0.03 - 0.13)	0.01 (0.01 - 0.02)
North Fork Skykomish Summer-Run	Snohomish	145	0.02 -	0.03 -	0.01 -	0.93 -	0.03 (0.01 - 0.05)	0.00 (0.00 - 0.02)	0.01 (0.01 - 0.03)	0.95 (0.94 - 0.96)
Tolt River Summer-Run	Snohomish	74	0.03 -	0.45 -	0.01 -	0.66 -	0.02 (0.00 - 0.06)	0.51 (0.42 - 0.60)	0.01 (0.00 - 0.03)	0.69 (0.59 - 0.80)
Snoqualmie River Winter-Run	Snohomish	166	0.07 -	0.00 -	0.04 -	0.04 -	0.06 (0.00 - 0.15)	0.00 (0.00 - 0.02)	0.04 (0.00 - 0.12)	0.03 (0.03 - 0.06)
Snohomish / Skykomish R Winter-Run	Snohomish	21	0.00 -	0.00 -	0.00 -	0.05 -	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	0.05 (0.03 - 0.09)
Pilchuck R Winter-Run	Snohomish	49	0.10 -	0.02 -	0.07 -	0.03 -	0.00 (0.00 - 0.13)	0.00 (0.00 - 0.05)	0.01 (0.00 - 0.12)	0.03 (0.02 - 0.08)
Canyon Creek Summer-Run	Stillaguamish	96	0.03 -	0.00 -	0.02 -	0.00 -	0.00 (0.00 - 0.03)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.02)	0.00 (0.00 - 0.00)
Deer Creek Summer-Run	Stillaguamish	157	0.03 -	0.07 -	0.01 -	0.04 -	0.00 (0.00 - 0.02)	0.03 (0.00 - 0.07)	0.00 (0.00 - 0.01)	0.02 (0.00 - 0.03)
StillaguamishRSmoltTrap06	Stillaguamish	86	0.08	0.11	0.05	0.18	0.00 (0.00 - 0.08)	0.09 (0.04 - 0.15)	0.00 (0.00 - 0.07)	0.18 (0.13 - 0.25)
Mainstem Skagit R Summer- and Winter-Run	Skagit	185	0.09 -	0.03 -	0.08 -	0.02 -	0.00 (0.00 - 0.04)	0.00 (0.00 - 0.02)	0.02 (0.00 - 0.06)	0.01 (0.01 - 0.03)
Sauk R Summer- and Winter-Run	Skagit	115	0.03 -	0.00 -	0.05 -	0.00 -	0.00 (0.00 - 0.03)	0.00 (0.00 - 0.00)	0.04 (0.02 - 0.07)	0.00 (0.00 - 0.00)
Nookachamps Creek Winter-Run	Skagit	45	0.05 -	0.00 -	0.05 -	0.00 -	0.00 (0.00 - 0.08)	0.00 (0.00 - 0.00)	0.02 (0.00 - 0.10)	0.00 (0.00 - 0.00)
Nooksack R Winter-Run	Nooksack	246	0.06 -	0.02 -	0.04 -	0.01 -	0.00 (0.00 - 0.02)	0.00 (0.00 - 0.01)	0.00 (0.00 - 0.02)	0.00 (0.00 - 0.01)
South Fork Nooksack R Summer-Run	Nooksack	66	0.02 -	0.00 -	0.01 -	0.00 -	0.00 (0.00 - 0.05)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.02)	0.00 (0.00 - 0.00)
Samish R Winter-Run	Samish	87	0.19 -	0.03 -	0.13 -	0.02 -	0.11 (0.00 - 0.24)	0.00 (0.00 - 0.04)	0.06 (0.00 - 0.17)	0.00 (0.00 - 0.02)

Figures

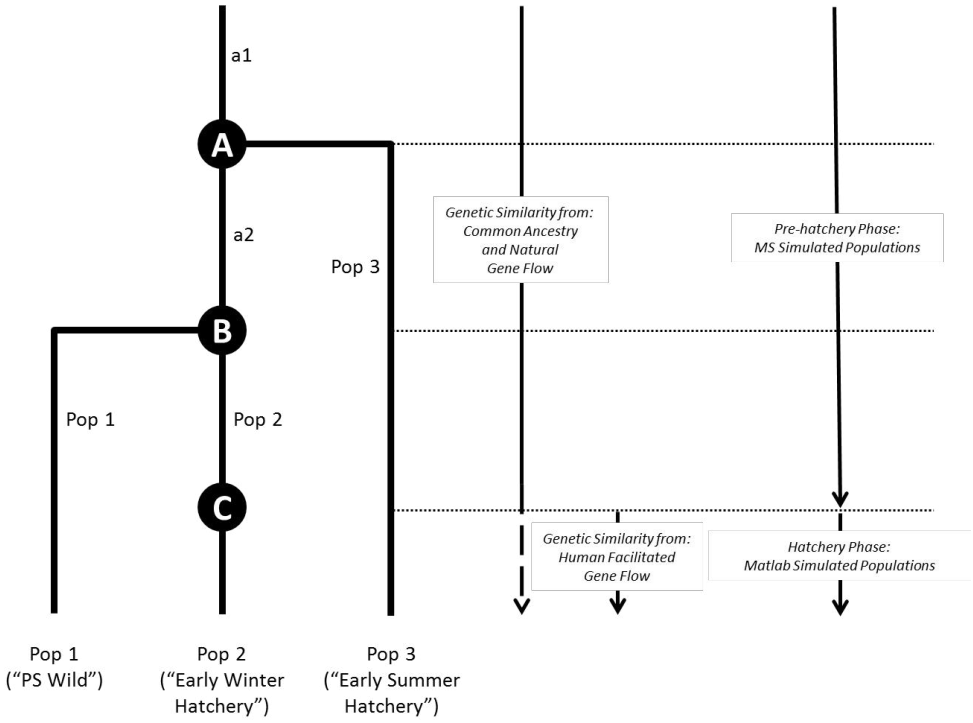


Figure 1. The hierarchical relationship among the wild (Pop 1), and early winter (Pop 2) and early summer (Pop 3) hatcheries, and temporal division between the simulated populations modeled using the programs *MS* (Hudson 2002) and *Matlab* (MathWorks 2012). Simulated populations are hierarchically related. Single population existed at some time in the past (Pop a1). At time A, Pop a1 instantaneously split into two populations (Pop a2 and Pop 3) of the same size. The two populations were demographically stable and exchanged immigrants. At time B, Pop a2 instantaneously split into two populations (Pop 1 and Pop 2) of the same size. The two populations were demographically stable and exchanged immigrants. Time C represents the start of the segregated hatchery programs in Puget Sound, and the beginning of human facilitated gene flow among the three populations. Genetic similarity among all populations up to time C was a result of common ancestry and natural gene flow. There is little to no natural gene flow among the populations now, but similarity from common ancestry still exists (hence the broken line starting at time C). See Tables S1 and S6 and Figure S1 for *MS* and *Matlab* model parameters and schema.

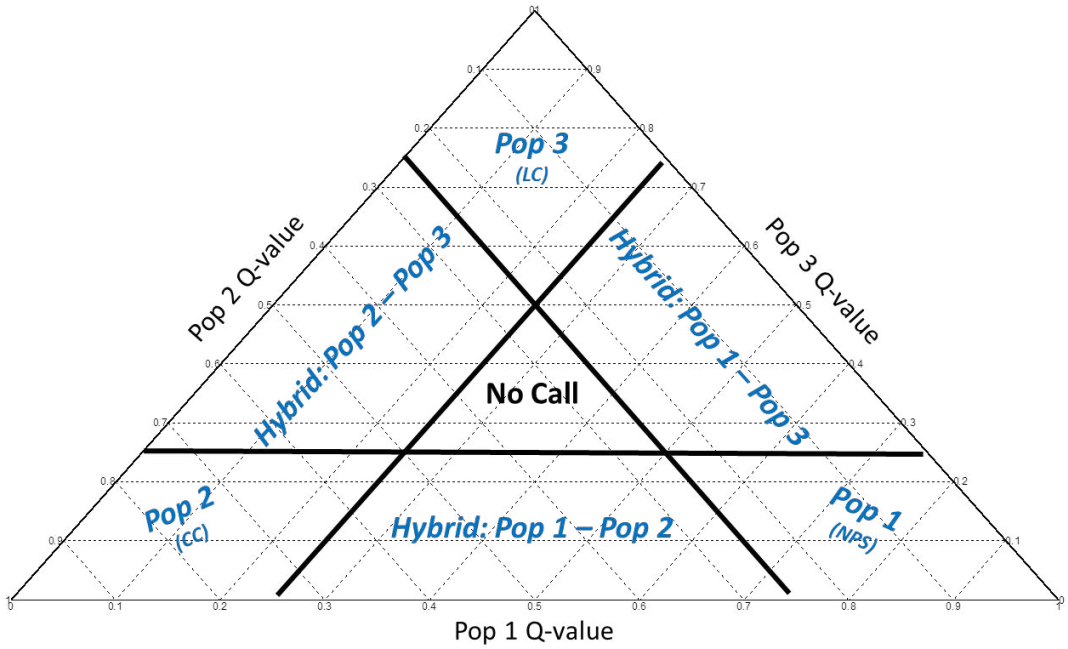


Figure 2. Ternary diagram indicating the program *Structure*'s k = 3 assignment regions and thresholds. Assignment thresholds (thick solid black lines) were set to minimize mean squared errors (Equation 2, see text).

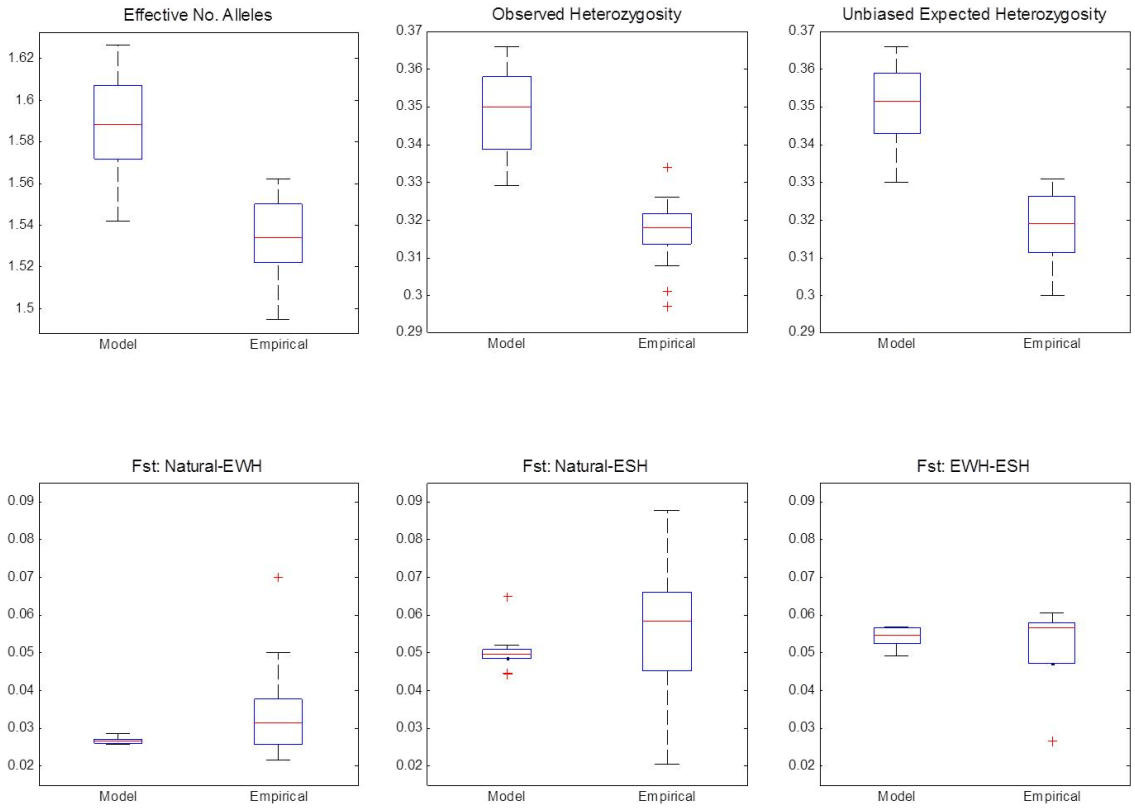


Figure 3. Box plots comparing the genetic diversity (top row) and pairwise population differentiation (bottom row) in the empirical (Tables S2, S3) and simulated (model) (Tables S4, S5) populations. Horizontal line in each box corresponds to the median value, lower and upper bounds of the box are the first and third quartile, respectively, and the "whisker" tips cover approximately 99% of the data, if the data were normally distributed.

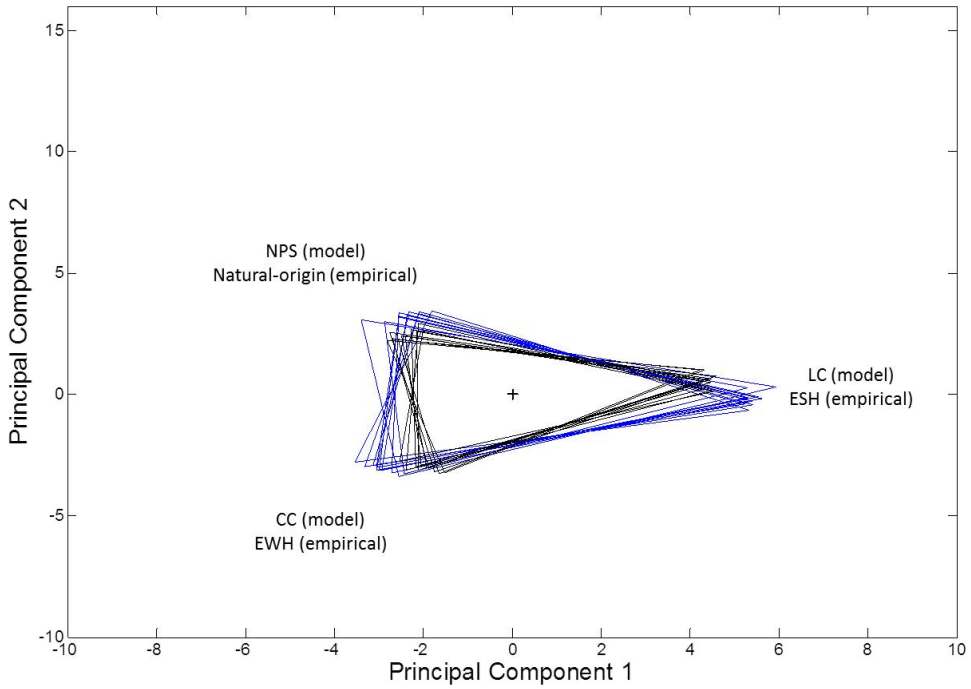


Figure 4. To compare directly the ordination of the three populations in Figures S2 and S3, each of the 20 principal component analysis (10 simulated populations [Figure S2] and 10 empirical populations [Figure S3]) are represented as overlaid triangles with their vertices located at the centroids for each of the populations, and plotted using the axes from Figure S3. Blue triangles for the simulated populations (iterations) in Figure S2, and black triangles for empirical populations in Figure S3. The triangles occupy similar multivariate space and are nearly superimposed. The empirical-based triangles occupy slightly less area than the simulation-based triangles, reflecting greater overlap among their clusters (i.e., smaller Mahalanobis D^2), but the overall ordination of the 20 PCAs are nearly coincident.

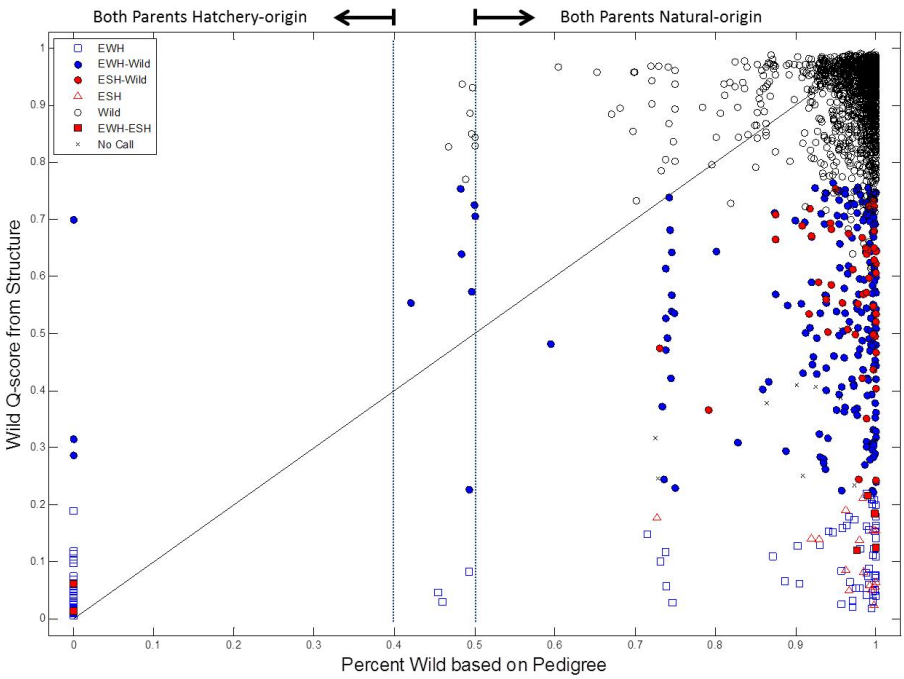


Figure 5. Q-scores for the Wild assignment category using *Structure's* default no prior population information mode, for all natural-origin individuals in the pooled 27 hatchery phase simulated collections, as a function of the individual's true percent wild, based on its pedigree (through 12 generations). Each point is an individual's Q-score with the symbol representing the individual's *Structure* k = 3 assignment (see legend; rows in Table 5). The true identity of each individual is based on the identity of that individual's parents. Individuals with two natural-origin parents are defined as Wild (right portion of plot), with one each natural- and hatchery-origin parent as a hybrid (narrow center portion of plot), and with two hatchery-origin parents as hatchery-lineage (left portion of plot; see text, especially footnote 2 in Preamble). There is a direct correspondence between this figure and Table 5. For example, there are 1793 Wild individuals (Wild Source column in Table 5) defined by having two natural-origin parents and are shown above in the right portion of the plot. These Wild individuals were assigned using *Structure* into every assignment category, including No Call (distribution of assignments for Wild individuals are the Wild Source column in Table 4, and the different symbols shown above in the right portion of the plot). If there was a perfect relationship between percent Wild based on pedigree and percent Wild based on Q-scores all symbols would align along the solid diagonal line.

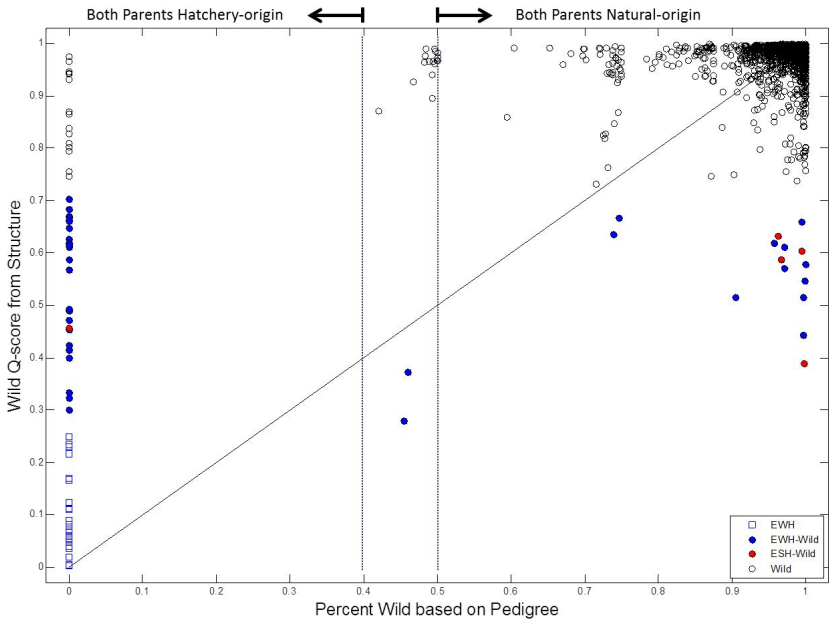


Figure 6. Q-scores for the Wild assignment category using *Structure's* prior population information mode, for all natural-origin individuals in the pooled 27 hatchery phase simulated collections, as a function of the individual's true percent wild, based on its pedigree (through 12 generations). See Figure 5 for details and Table 6 for comparisons.

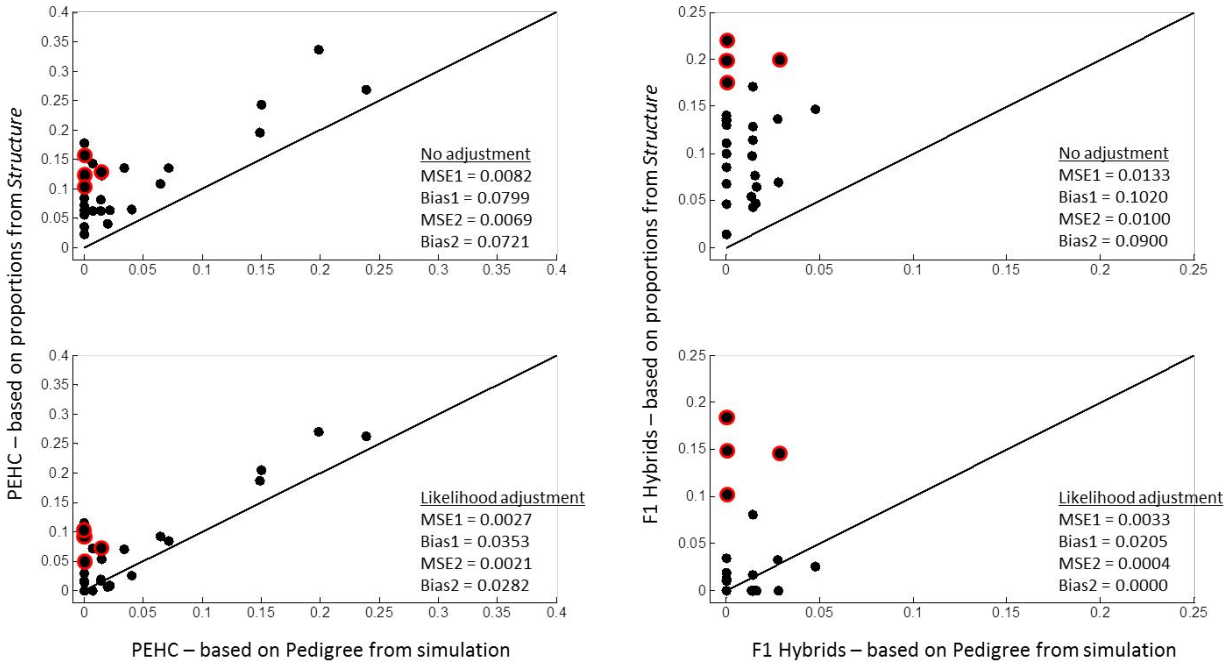


Figure 7. Comparison of $PEHC_W$ (left) and F1 $Hybrids_W$ (introgression) (right) based on *Structure* proportions without (top) and with (bottom) likelihood adjustment, as a function of the true $PEHC_W$ (Table 4) for the 27 hatchery phase simulated natural-origin collections, using *Structure's* default no prior population information mode. Points highlighted in red are those collections where one or more likelihood adjusted proportions were uncertain due to large confidence intervals in the estimate (see Methods, Section 2). MSE1 and Bias1 refer to mean squared error and bias using all collections, and MSE2 and Bias2 refer to mean squared error and bias with the collections with uncertain estimates removed from the calculations. Diagonal lines represent perfect correlations between estimated and true $PEHC_W$, and estimated and true F1 Hybrids. If all collections fell on the diagonal lines, there would be zero MSE and bias. Bias is calculated as the mean of the unsquared errors.

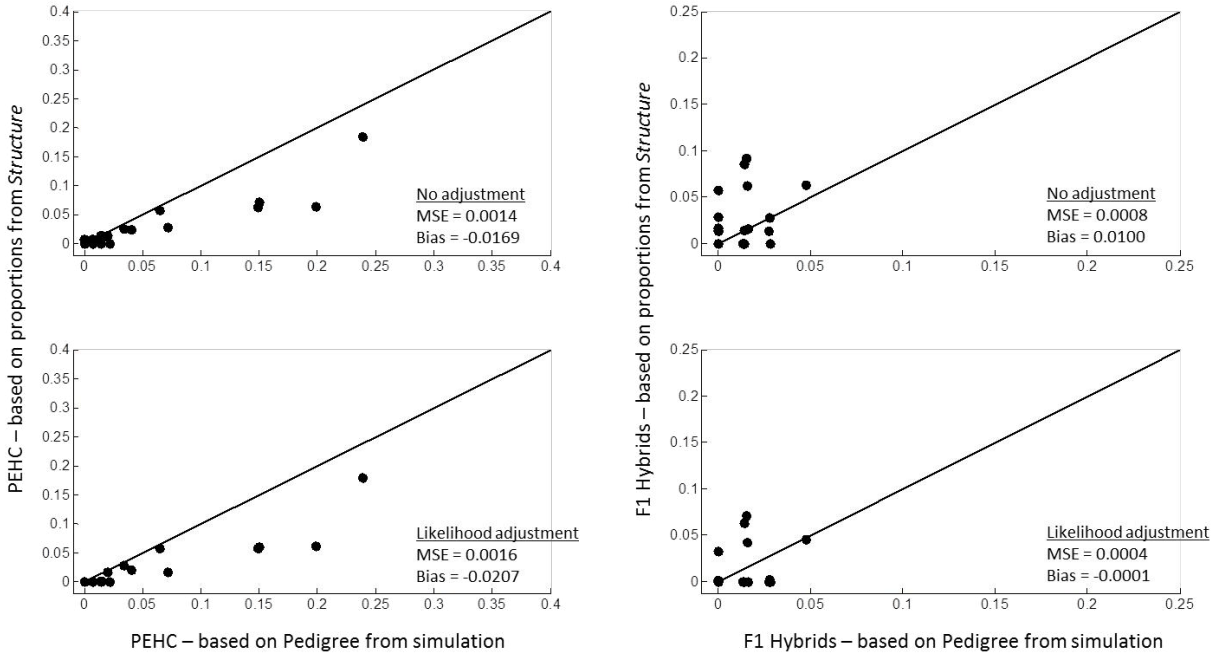


Figure 8. Comparison of $PEHC_W$ (left) and F1 $Hybrids_W$ (introgression) (right) based on *Structure* proportions without (top) and with (bottom) likelihood adjustment, as a function of the true $PEHC_W$ (Table 4) for the 27 hatchery phase simulated natural-origin collections, using *Structure's* prior population information mode. MSE is the mean squared error, and bias is calculated as the mean of the unsquared errors. Diagonal lines represent perfect correlations between estimated $PEHC_W$ and true $PEHC_W$. If all collections fell on the diagonal lines, there would be zero MSE and bias.

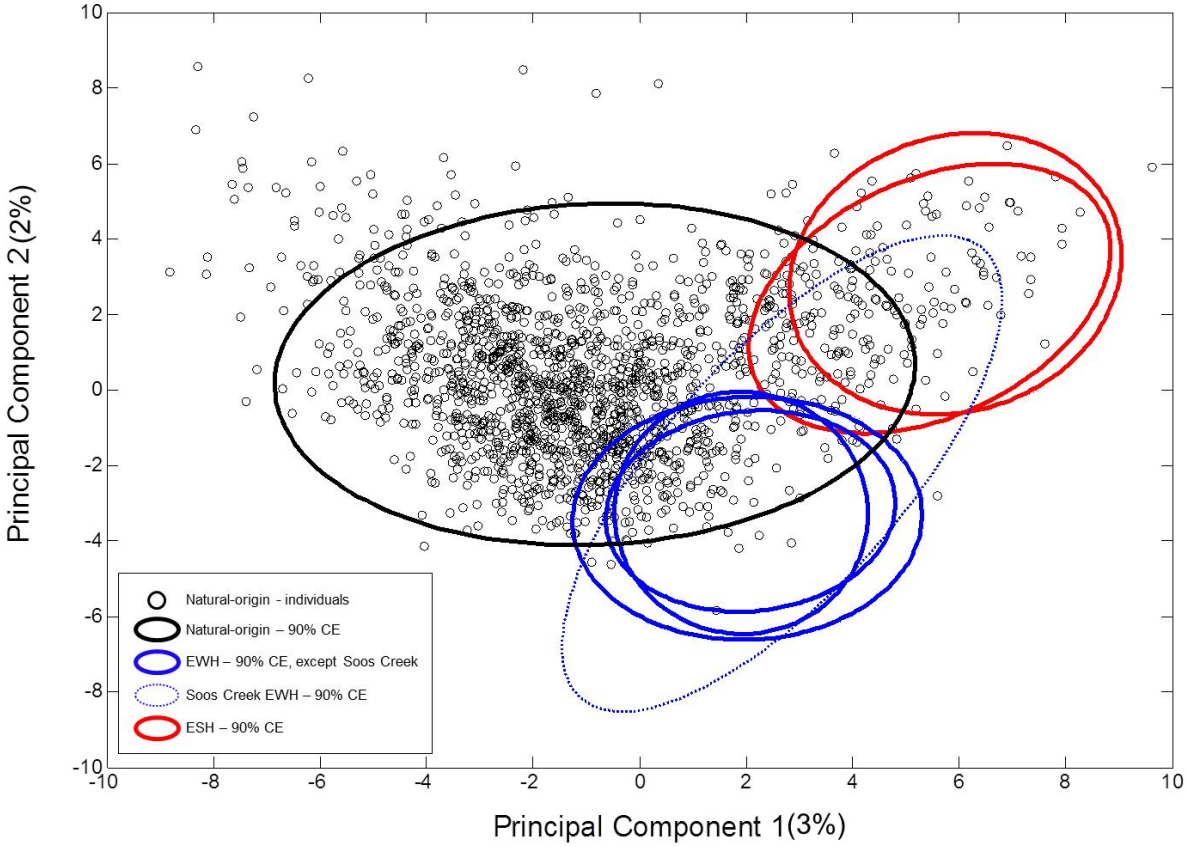


Figure 9. Principal component analysis using the pooled data set from all river basins. Individual data for hatchery programs are not shown, but each hatchery program is represented by its 90% confidence ellipse for scores on the first two components.

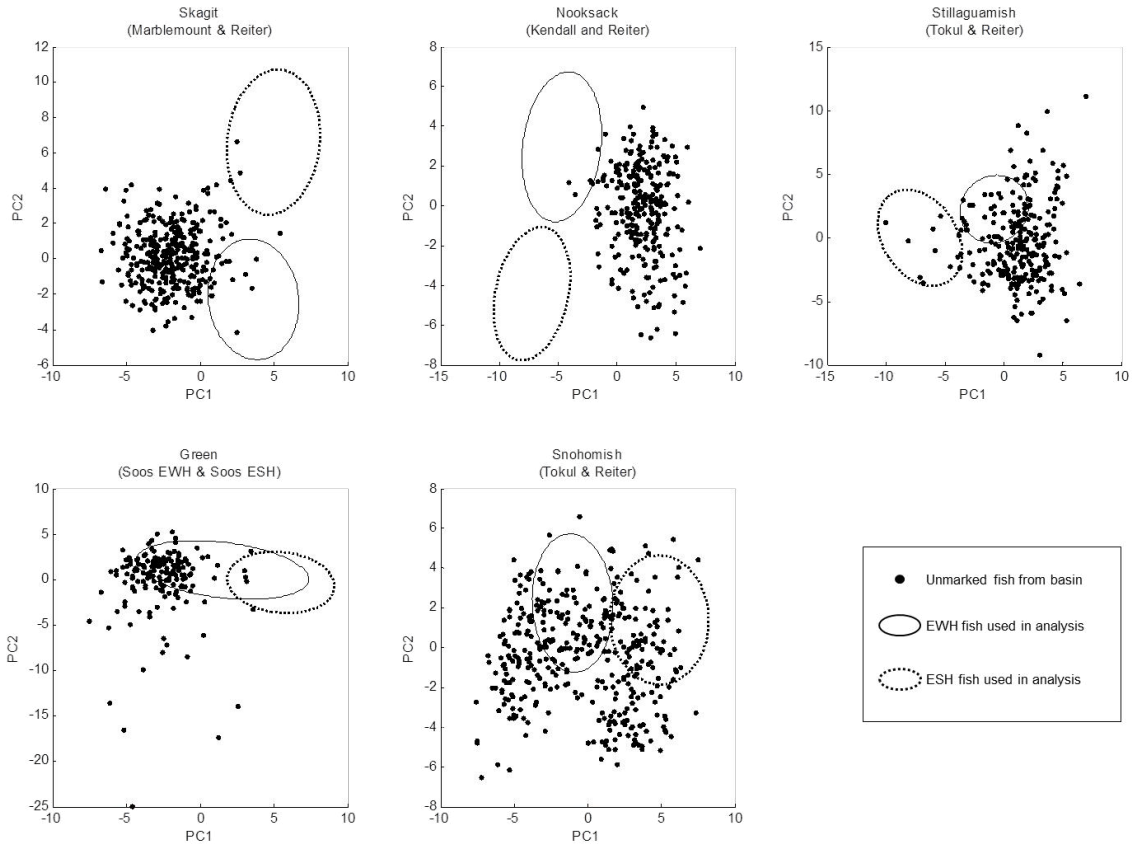


Figure 10. Principal component analyses using pooled data sets each from the Skagit, Nooksack, Stillaguamish, Green, and Snohomish river. Since there are no ESH programs in the Skagit or Nooksack Rivers, I used the Reiter Ponds data for those analyses. Likewise, there were no available samples for the EWH and ESH programs in the Stillaguamish River, and I used Tokul Creek and Reiter Ponds data for that analysis.

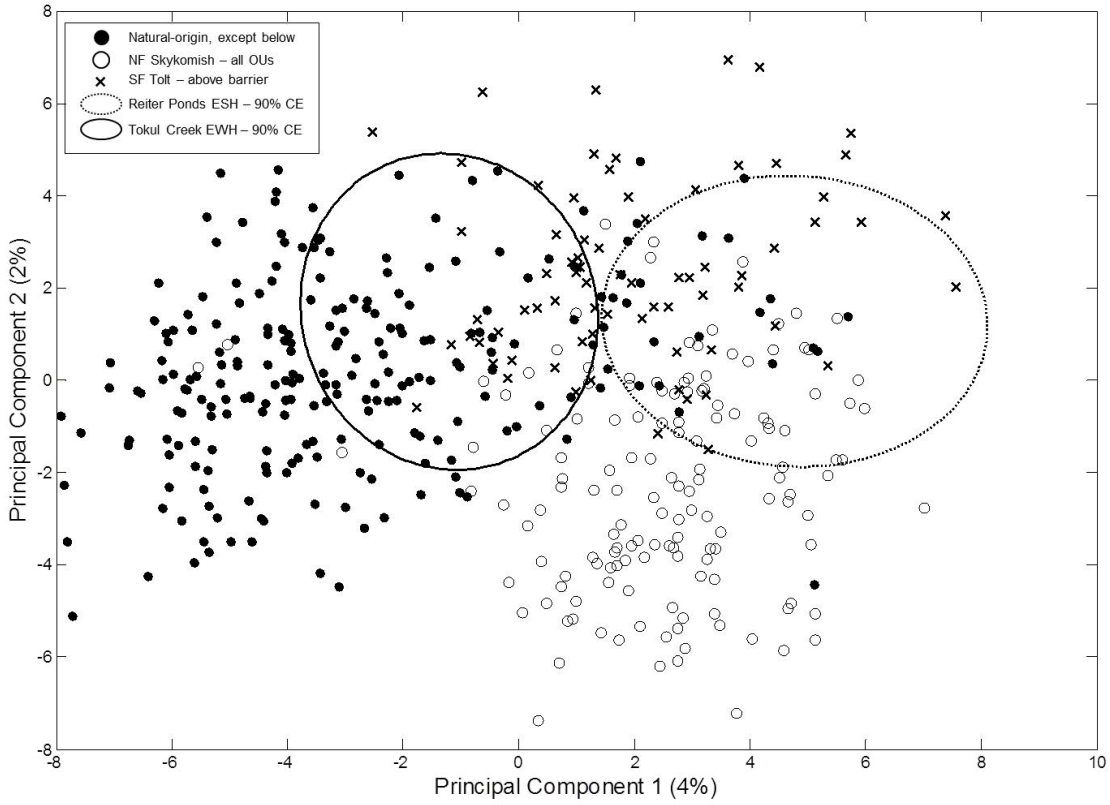


Figure 11. Principal component analysis for samples from the Snohomish River. Individual data for hatchery programs are not shown, but each hatchery program is represented by the individuals' 90% confidence ellipse for scores on the first two components.

Supplemental Tables and Figures

Table S1. Parameters used in the program *MS* (Hudson 2002) to simulate populations in the pre-hatchery phase of the model. N_o = current diploid population size.

Parameter	Parameter Switch	Values	Notes
# samples	nsam	1500	Number of chromosome copies (i.e., individuals). See –I below
# reps	nrep	1	Number of repeated samples (i.e., independent iterations under same conditions)
Mutation (θ)	-t θ	200	$\theta = (4N_o\mu) = 200$. With μ = neutral mutation rate for entire locus. Generates molecular diversity
Recombination (ρ)	-r ρ nsites	250 2501	$\rho = (4N_o r) = 250$. With *r* = probability of cross-over. nsites (~ number of basepairs @ locus where recombination can occur) = 2501. Generates molecular diversity
Population growth	-G α	0	Specifies population size at t = 4 N_o generation before present: $N(t) = N_o \exp^{-\alpha t}$. $\alpha = 0$ is stable population size
Spatial structure	-I npop n1 n2 n3	3 500 500 500	Number of populations. 3 populations each with 500 individuals sampled
Migration matrix	-ma (m_{ij})	xx 18 6 18 xx 5 6 5 xx	Elements of matrix = m_{ij}, fraction of subpopulation i equal to j. Generates population differentiation
Population splitting	-ej *t*, pop_i, pop_j	0.05 1 2	Population splitting = $(4N_o t)$ = generations, with t = 0.05, for split between pop1 and pop2 (i.e., split among Puget Sound populations). Generates population differentiation.
Population splitting	-ej *t*, pop_i, pop_j	0.20 3 2	Population splitting = $(4N_o t)$ = generations, with t = 0.20, for split between pop1&2 and pop3 (i.e., split between Puget Sound and Lower Columbia populations). Generates population differentiation.

Table S2. Diversity statistics from empirical data sets used to parameterize the pre-hatchery phase modeled populations simulated in the program *MS* (see Table 1). A_e = effective number of alleles, H_o = observed heterozygosity, and uH_e = unbiased expected heterozygosity. See Section 3 for description of DIPs and empirical data sets.

Basin	DIP or Hatchery	Category	A_e	H_o	uH_e
Green	Soos Creek - early summer	ESH	1.53	0.32	0.32
Snohomish	Reiter Ponds - early summer	ESH	1.52	0.32	0.31
		Mean	1.53	0.32	0.31
Green	Soos Creek - early winter	EWH	1.52	0.31	0.31
Nooksack	Kendall Creek - early winter	EWH	1.50	0.30	0.30
Skagit	Marblemount - early winter	EWH	1.51	0.33	0.30
Snohomish	Tokul Creek - early winter	EWH	1.51	0.30	0.31
		Mean	1.51	0.31	0.31
Green	Green River Winter-Run	Natural-origin	1.55	0.31	0.33
Nooksack	Nooksack Winter-Run	Natural-origin	1.56	0.32	0.33
Nooksack	SF Nooksack Summer-Run	Natural-origin	1.54	0.32	0.32
Skagit	Mainstem Skagit Summer- & Winter-Runs	Natural-origin	1.56	0.32	0.33
Skagit	Nookachamps Winter-Run	Natural-origin	1.56	0.33	0.33
Skagit	Sauk Summer- & Winter-Runs	Natural-origin	1.54	0.32	0.32
Snohomish	NF Skykomish Summer-Run	Natural-origin	1.55	0.32	0.32
Snohomish	Pilchuck Winter-Run	Natural-origin	1.53	0.31	0.32
Snohomish	Snohomish/Skykomish Winter-Run	Natural-origin	1.53	0.32	0.32
Snohomish	Snoqualmie Winter-Run	Natural-origin	1.55	0.32	0.32
Snohomish	Tolt Summer-Run	Natural-origin	1.53	0.32	0.32
Stillaguamish	Canyon Creek Summer-Run	Natural-origin	1.53	0.32	0.31
Stillaguamish	Deer Creek Summer-Run	Natural-origin	1.55	0.32	0.33
		Mean	1.55	0.32	0.32
		Overall Mean	1.54	0.32	0.32

Table S3. Pairwise population differentiation (F_{ST}, *sensu* Weir and Cockerham 1984) from empirical data sets used to parameterize the pre-hatchery phase modeled populations simulated in the program *MS* (see Table 1). See Section 3 for description of DIPs and empirical data sets.

Basin	Pair		Category	F_{ST}
	DIP or Hatchery	Hatchery		
Green	Soos Creek - early winter	Soos Creek - early summer	EWH-ESH	0.027
Nooksack	Kendall Creek - early winter	Reiter Ponds - early summer	EWH-ESH	0.061
Skagit	Marblemount - early winter	Reiter Ponds - early summer	EWH-ESH	0.054
Snohomish	Tokul Creek - early winter	Reiter Ponds - early summer	EWH-ESH	0.056
Stillaguamish	Tokul Creek - early winter	Reiter Ponds - early summer	EWH-ESH	0.057
			Mean	0.051
Green	Green River Winter-Run	Soos Creek - early summer	ESH-Wild	0.049
Nooksack	Nooksack Winter-Run	Reiter Ponds - early summer	ESH-Wild	0.073
Nooksack	SF Nooksack Summer-Run	Reiter Ponds - early summer	ESH-Wild	0.088
Skagit	Mainstem Skagit Summer- & Winter-Runs	Reiter Ponds - early summer	ESH-Wild	0.053
Skagit	Nookachamps Winter-Run	Reiter Ponds - early summer	ESH-Wild	0.060
Skagit	Sauk Summer- & Winter-Runs	Reiter Ponds - early summer	ESH-Wild	0.063
Snohomish	NF Skykomish Summer-Run	Reiter Ponds - early summer	ESH-Wild	0.023
Snohomish	Pilchuck Winter-Run	Reiter Ponds - early summer	ESH-Wild	0.069
Snohomish	Snohomish/Skykomish Winter-Run	Reiter Ponds - early summer	ESH-Wild	0.059
Snohomish	Snoqualmie Winter-Run	Reiter Ponds - early summer	ESH-Wild	0.045
Snohomish	Tolt Summer-Run	Reiter Ponds - early summer	ESH-Wild	0.020
Stillaguamish	Canyon Creek Summer-Run	Reiter Ponds - early summer	ESH-Wild	0.066
Stillaguamish	Deer Creek Summer-Run	Reiter Ponds - early summer	ESH-Wild	0.057
Stillaguamish	Stillaguamish Smolt - Aggregate	Reiter Ponds - early summer	ESH-Wild	0.038
			Mean	0.054
Green	Green River Winter-Run	Soos Creek - early winter	EWH-Wild	0.023
Nooksack	Nooksack Winter-Run	Kendall Creek - early winter	EWH-Wild	0.050
Nooksack	SF Nooksack Summer-Run	Kendall Creek - early winter	EWH-Wild	0.070
Skagit	Mainstem Skagit Summer- & Winter-Runs	Marblemount - early winter	EWH-Wild	0.027
Skagit	Nookachamps Winter-Run	Marblemount - early winter	EWH-Wild	0.033
Skagit	Sauk Summer- & Winter-Runs	Marblemount - early winter	EWH-Wild	0.030
Snohomish	NF Skykomish Summer-Run	Tokul Creek - early winter	EWH-Wild	0.038
Snohomish	Pilchuck Winter-Run	Tokul Creek - early winter	EWH-Wild	0.028
Snohomish	Snohomish/Skykomish Winter-Run	Tokul Creek - early winter	EWH-Wild	0.026
Snohomish	Snoqualmie Winter-Run	Tokul Creek - early winter	EWH-Wild	0.022
Snohomish	Tolt Summer-Run	Tokul Creek - early winter	EWH-Wild	0.033
Stillaguamish	Canyon Creek Summer-Run	Tokul Creek - early winter	EWH-Wild	0.034
Stillaguamish	Deer Creek Summer-Run	Tokul Creek - early winter	EWH-Wild	0.038
Stillaguamish	Stillaguamish Smolt - Aggregate	Tokul Creek - early winter	EWH-Wild	0.022
			Mean	0.034

Table S4. Diversity statistics from three pre-hatchery phase modeled populations simulated using the program *MS*. A_e = effective number of alleles, H_o = observed heterozygosity, and uH_e = unbiased expected heterozygosity. Each iteration was an independent simulation in *MS*, with each simulation using the same set of parameters in Table S1. Chambers Creek, Lower Columbia, and North Puget Sound simulated populations represent the early winter hatchery (EWH), early summer hatchery (ESH), and north Puget Sound (Wild) populations that existed just prior to the beginning of the hatchery programs in the 1950s (point "C" in Figure 1).

Iteration	Category	A_e	H_o	uH_e
Iter3	Lower Columbia	1.61	0.36	0.36
Iter4	Lower Columbia	1.61	0.35	0.36
Iter7	Lower Columbia	1.61	0.36	0.36
Iter13	Lower Columbia	1.62	0.37	0.37
Iter18	Lower Columbia	1.61	0.36	0.36
Iter25	Lower Columbia	1.58	0.35	0.35
Iter26	Lower Columbia	1.56	0.34	0.34
Iter27	Lower Columbia	1.54	0.33	0.33
Iter38	Lower Columbia	1.58	0.35	0.35
Iter50	Lower Columbia	1.56	0.34	0.34
	Mean	1.59	0.35	0.35
Iter3	Chambers Creek	1.63	0.36	0.37
Iter4	Chambers Creek	1.61	0.36	0.36
Iter7	Chambers Creek	1.59	0.34	0.35
Iter13	Chambers Creek	1.60	0.36	0.36
Iter18	Chambers Creek	1.60	0.36	0.36
Iter25	Chambers Creek	1.59	0.35	0.35
Iter26	Chambers Creek	1.55	0.34	0.34
Iter27	Chambers Creek	1.56	0.34	0.34
Iter38	Chambers Creek	1.55	0.33	0.33
Iter50	Chambers Creek	1.56	0.34	0.34
	Mean	1.58	0.35	0.35
Iter3	North Puget Sound	1.60	0.36	0.35
Iter4	North Puget Sound	1.60	0.35	0.35
Iter7	North Puget Sound	1.61	0.35	0.36
Iter13	North Puget Sound	1.59	0.35	0.35
Iter18	North Puget Sound	1.61	0.36	0.36
Iter25	North Puget Sound	1.61	0.36	0.36
Iter26	North Puget Sound	1.58	0.35	0.35
Iter27	North Puget Sound	1.59	0.35	0.35
Iter38	North Puget Sound	1.57	0.34	0.34
Iter50	North Puget Sound	1.57	0.35	0.34
	Mean	1.59	0.35	0.35
	Overall Mean	1.59	0.35	0.35

Table S5. Pairwise population differentiation (F_{ST}, *sensu* Weir and Cockerham 1984) for the three pre-hatchery phase modeled populations simulated using the program *MS*. . Each iteration was an independent simulation in *MS*, with each simulation using the same set of parameters in Table S1. Chambers Creek (CC), Lower Columbia (LC), and North Puget Sound (NPS) simulated populations represent the early winter hatchery (EWH), early summer hatchery (ESH), and north Puget Sound (Wild) populations that existed just prior to the beginning of the hatchery programs in the 1950s (point "C" in Figure 1).

Iteration	Category	F_{ST}
Iter3	CC-LC	0.054
Iter4	CC-LC	0.057
Iter7	CC-LC	0.057
Iter13	CC-LC	0.056
Iter18	CC-LC	0.054
Iter25	CC-LC	0.057
Iter26	CC-LC	0.053
Iter27	CC-LC	0.049
Iter38	CC-LC	0.056
Iter50	CC-LC	0.050
	Mean	0.054
Iter3	LC-NPS	0.049
Iter4	LC-NPS	0.051
Iter7	LC-NPS	0.065
Iter13	LC-NPS	0.052
Iter18	LC-NPS	0.050
Iter25	LC-NPS	0.049
Iter26	LC-NPS	0.044
Iter27	LC-NPS	0.049
Iter38	LC-NPS	0.045
Iter50	LC-NPS	0.050
	Mean	0.050
Iter3	CC-NPS	0.027
Iter4	CC-NPS	0.026
Iter7	CC-NPS	0.027
Iter13	CC-NPS	0.026
Iter18	CC-NPS	0.027
Iter25	CC-NPS	0.026
Iter26	CC-NPS	0.027
Iter27	CC-NPS	0.028
Iter38	CC-NPS	0.026
Iter50	CC-NPS	0.026
	Mean	0.027

Table S6. Parameters used to simulate the hatchery phase natural-origin, EWH, and ESH populations in north Puget Sound, modeled after empirical populations and hatchery practices that existed prior to 2009. See Figure S1.

Parameter	Parameter Value	Description	Target Population	Source
$Escape_W$	~ Normal (μ=1696, σ=834)	Natural-origin escapement	Wild	WDFW unpublished data, Snoqualmie, 2001-2010
ESS_H	~ Poisson (λ = 0.73)	Hatchery Egg-to-smolt survival	All hatchery	Tokul Creek HGMP using total number of releases and egg take goal
ESS_W	~ Poisson (λ = 0.014)	Egg-to-smolt survival	Wild	Quinn (2005, Table 15-1)
ET_{ESH}	235,000	Egg take goal	ESH	WDFW HGMP
ET_{EWH}	280,000	Egg take goal	EWH	WDFW HGMP
FEC_H	~ Poisson (λ = 3500)	Female fecundity	All hatchery	WDFW HGMP
FEC_W	~ Poisson (λ = 4293)	Female fecundity	Wild	Quinn (2005, Table 15-1)
FW	(FW_W♀Parent + FW_W♂Parent) / 2	Fitness weight: If FW < FWH then FW = FWH else FW=FW	All	Calculated
FW_H	0.084	Fitness weight of pure hatchery individuals spawning naturally. Sets lowest value for fitness	All hatchery	mean from Araki et al. (2008, Table 1; steelhead - nonlocal)
FW_W	percent wild based on pedigree	Fitness weight of wild individuals	Wild	Calculated
HOB_{ESH}	2 * (ET_{ESH} / FEC_H)	Number of hatchery-origin broodstock	ESH	Calculated
HOB_{EWH}	2 * (ET_{EWH} / FEC_H)	Number of hatchery-origin broodstock	EWH	Calculated
Ngen	12	Number of simulated generations	All	Modeled
$OffRet_{ESH}$	($OnRet_{ESH}$ / $OnSR_{ESH}$) * $OffSR_{ESH}$	Off-station adult returns	ESH	Calculated
$OffRet_{EWH}$	($OnRet_{EWH}$ / $OnSR_{EWH}$) * $OffSR_{EWH}$	Off-station adult returns	EWH	Calculated
$OffSR_{ESH}$	~ Normal (μ=39,195, σ=18,142)	Off-station smolt releases	ESH	WDFW unpublished data, Snohomish-Skykomish, 2001-2010
$OffSR_{EWH}$	~ Normal (μ=48,562, σ=18,136)	Off-station smolt releases	EWH	WDFW unpublished data, Snoqualmie, 2001-2010
OffStray	1	Off-station stray rate	All hatchery	Modeled
$OnRet_{ESH}$	~ Normal (μ=133, σ=59)	On-station adult returns	ESH	WDFW unpublished data, Snohomish-Skykomish, 2001-2010
$OnRet_{EWH}$	~ Normal (μ=717, σ=331)	On-station adult returns	EWH	WDFW unpublished data, Snoqualmie, 2001-2010
$OnSR_{ESH}$	~ Normal (μ=167,790, σ=33,018)	On-station smolt releases	ESH	WDFW unpublished data, Snohomish-Skykomish, 2001-2010
$OnSR_{EWH}$	~ Normal (μ=143,288, σ=21,662)	On-station smolt releases	EWH	WDFW unpublished data, Snoqualmie, 2001-2010
OnStray	0	On-station stray rate	All hatchery	Modeled
OvL_{ESH}	0.1	Natural-spawning overlap with wild pop	ESH	Modeled
OvL_{EWH}	{0.1, 0.5, 1.0}	Natural-spawning overlap with wild pop	EWH	Modeled
OvL_W	{0.1, 0.5, 1.0}	Natural-spawning overlap with hatchery-origin/lineage pop	Wild	Modeled
Sample	75	Number of individuals sampled from populations	All	Modeled
SASurv	~ Poisson (λ = 0.13)	Smolt-to-adult survival	All	Quinn (2005, Table 15-1)
$TotSR_{ESH}$	$OnSR_{ESH}$ + $OffSR_{ESH}$	Total smolt releases	ESH	Calculated
$TotSR_{EWH}$	$OnSR_{EWH}$ + $OffSR_{EWH}$	Total smolt releases	EWH	Calculated

Table S7. Unadjusted and likelihood adjusted proportions for each of the 27 hatchery phase simulated natural-origin collections, using *Structure's* default no prior population information mode. Proportions in bold typeface are those where the range of the 90% confidence interval for the likelihood adjustment exceeded 0.25, reducing confidence in the estimate of the proportion. The unadjusted proportions are calculations based on *Structure* assignments only (Section 1). The *Structure* likelihood adjustment procedure is described in Section 2. See Table 4 and Figure 7.

		Unadjusted					Adjusted				
Simulated Collections	N	EWH Lineage	Hybrid: EWH-Wild	Hybrid: ESH-Wild	ESH Lineage	Wild	EWH Lineage	Hybrid: EWH-Wild	Hybrid: ESH-Wild	ESH Lineage	Wild
Sim0.1_0.1_0.1_0.87551	59	0.10	0.07	0.00	0.03	0.80	0.11	0.00	0.00	0.04	0.85
Sim0.1_0.1_0.1_0.87994	71	0.03	0.01	0.04	0.01	0.90	0.03	0.00	0.00	0.01	0.96
Sim0.1_0.1_0.1_0.88945	70	0.03	0.09	0.00	0.01	0.87	0.02	0.00	0.00	0.02	0.97
Sim0.1_0.5_0.1_0.25404	62	0.03	0.06	0.02	0.02	0.87	0.03	0.00	0.00	0.02	0.95
Sim0.1_0.5_0.1_0.26156	70	0.03	0.20	0.04	0.00	0.73	0.00	**0.15**	0.00	0.00	0.85
Sim0.1_0.5_0.1_0.27494	70	0.06	0.17	0.00	0.00	0.77	0.03	0.08	0.00	0.00	0.89
Sim0.1_1.0_0.1_0.23617	69	0.09	0.04	0.03	0.01	0.83	0.10	0.00	0.00	0.02	0.88
Sim0.1_1.0_0.1_0.93021	72	0.03	0.07	0.08	0.03	0.79	0.02	0.00	0.05	0.03	0.90
Sim0.1_1.0_0.1_0.93805	74	0.01	0.05	0.01	0.01	0.91	0.01	0.00	0.00	0.02	0.97
Sim0.5_0.1_0.1_7257	72	0.01	0.10	0.03	0.00	0.86	0.00	0.00	0.00	0.00	1.00
Sim0.5_0.1_0.1_73675	74	0.05	0.14	0.01	0.01	0.78	0.04	0.02	0.00	0.02	0.92
Sim0.5_0.1_0.1_74815	65	0.00	0.05	0.03	0.02	0.91	0.00	0.00	0.00	0.02	0.98
Sim0.5_0.5_0.1_78476	70	0.19	0.11	0.01	0.00	0.69	0.20	0.00	0.00	0.00	0.80
Sim0.5_0.5_0.1_81303	68	0.01	0.22	0.03	0.01	0.72	0.00	**0.19**	0.00	0.02	0.80
Sim0.5_0.5_0.1_83379	73	0.11	0.14	0.00	0.00	0.75	0.11	0.01	0.00	0.00	0.88
Sim0.5_1.0_0.1_29295	65	0.23	0.08	0.06	0.05	0.58	0.26	0.00	0.03	0.05	0.66
Sim0.5_1.0_0.1_30511	70	0.09	0.10	0.04	0.00	0.77	0.08	0.00	0.00	0.00	0.91
Sim0.5_1.0_0.1_31937	70	0.06	0.20	0.04	0.00	0.70	0.03	**0.15**	0.00	0.00	0.82
Sim1.0_0.1_0.1_34109	68	0.01	0.18	0.03	0.01	0.76	0.00	**0.10**	0.00	0.02	0.88
Sim1.0_0.1_0.1_34805	71	0.01	0.14	0.04	0.00	0.80	0.00	0.03	0.00	0.00	0.97
Sim1.0_0.1_0.1_35596	69	0.06	0.13	0.03	0.00	0.78	0.05	0.01	0.00	0.00	0.94
Sim1.0_0.5_0.1_36349	64	0.17	0.05	0.03	0.02	0.73	0.19	0.00	0.00	0.02	0.79
Sim1.0_0.5_0.1_37503	72	0.00	0.11	0.00	0.04	0.85	0.00	0.00	0.00	0.05	0.95
Sim1.0_0.5_0.1_38213	70	0.00	0.13	0.06	0.01	0.80	0.00	0.02	0.01	0.01	0.96
Sim1.0_1.0_0.1_38892	61	0.26	0.15	0.00	0.02	0.57	0.30	0.03	0.00	0.02	0.66
Sim1.0_1.0_0.1_39511	70	0.01	0.10	0.03	0.00	0.86	0.00	0.00	0.00	0.00	1.00
Sim1.0_1.0_0.1_4006	73	0.01	0.14	0.01	0.03	0.81	0.00	0.03	0.00	0.03	0.93
Mean		0.06	0.11	0.03	0.01	0.79	0.06	0.03	0.00	0.02	0.89

Table S8. Unadjusted and likelihood adjusted proportions for each of the 27 hatchery phase simulated natural-origin collections, using *Structure's* prior population information mode. The unadjusted proportions are calculations based on *Structure* assignments only (Section 1). The *Structure* likelihood adjustment procedure is described in Section 2. See Table 4 and Figure 8.

Simulated Collections	N	Unadjusted					Adjusted				
		EWH Lineage	Hybrid: EWH-Wild	Hybrid: ESH-Wild	ESH Lineage	Wild	EWH Lineage	Hybrid: EWH-Wild	Hybrid: ESH-Wild	ESH Lineage	Wild
Sim0.1_0.1_0.1_0.87551	59	0.02	0.02	0.02	0.02	0.93	0.03	0.00	0.01	0.03	0.94
Sim0.1_0.1_0.1_0.87994	72	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	1.00
Sim0.1_0.1_0.1_0.88945	70	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	1.00
Sim0.1_0.5_0.1_0.25404	62	0.02	0.02	0.02	0.00	0.95	0.02	0.00	0.00	0.00	0.98
Sim0.1_0.5_0.1_0.26156	71	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim0.1_0.5_0.1_0.27494	71	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	1.00
Sim0.1_1.0_0.1_0.23617	70	0.06	0.00	0.00	0.01	0.93	0.06	0.00	0.00	0.01	0.93
Sim0.1_1.0_0.1_0.93021	72	0.00	0.03	0.00	0.00	0.97	0.00	0.00	0.00	0.00	1.00
Sim0.1_1.0_0.1_0.93805	75	0.01	0.00	0.01	0.00	0.97	0.02	0.00	0.00	0.00	0.98
Sim0.5_0.1_0.1_7257	73	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim0.5_0.1_0.1_73675	74	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim0.5_0.1_0.1_74815	66	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim0.5_0.5_0.1_78476	70	0.03	0.09	0.00	0.00	0.89	0.03	0.06	0.00	0.00	0.91
Sim0.5_0.5_0.1_81303	71	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim0.5_0.5_0.1_83379	74	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	1.00
Sim0.5_1.0_0.1_29295	65	0.14	0.09	0.00	0.02	0.75	0.14	0.07	0.00	0.02	0.77
Sim0.5_1.0_0.1_30511	70	0.00	0.06	0.00	0.00	0.94	0.00	0.03	0.00	0.00	0.97
Sim0.5_1.0_0.1_31937	73	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.1_0.1_34109	68	0.00	0.00	0.01	0.00	0.99	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.1_0.1_34805	71	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.1_0.1_35596	70	0.00	0.03	0.00	0.00	0.97	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.5_0.1_36349	64	0.03	0.06	0.00	0.02	0.89	0.04	0.04	0.00	0.02	0.90
Sim1.0_0.5_0.1_37503	74	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim1.0_0.5_0.1_38213	70	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim1.0_1.0_0.1_38892	63	0.03	0.06	0.03	0.00	0.87	0.04	0.05	0.02	0.00	0.90
Sim1.0_1.0_0.1_39511	72	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
Sim1.0_1.0_0.1_4006	73	0.00	0.01	0.01	0.00	0.97	0.00	0.00	0.00	0.00	1.00
Mean		0.01	0.02	0.00	0.00	0.96	0.01	0.01	0.00	0.00	0.97

Table S9. Collection data, sorted in same order as in Table 7. Collections were aggregated into Operational Units (OUs), which were the primary units for analysis. OUs were aggregated into Demographically Independent Populations (DIPs; PSSTRT 2013), which are the primary management units. Total N is the number of samples genotyped; *Structure* N is the number of samples used for the *Structure* analyses per OU. Samples were removed if they were missing more than one-third of loci, showed at least one cutthroat allele, or, for the *Structure* analyses, if they were part of a full-sibling group (one individual from the group was retained).

Code	PSSTRT DIP	Operational Unit	Total N	Removed Missing Loci	Removed Cut. Alleles	Removed Full-sibling	*Structure* N
11NW	Nooksack R Winter-Run	MainstemNookEarlyAd	24	0	0	1	23
12MP	Nooksack R Winter-Run	MainstemNookEarlyAd	22	0	1	0	21
13GC	Nooksack R Winter-Run	MainstemNookEarlyAd	12	0	0	0	12
12MQ	Nooksack R Winter-Run	NFNooksackAd	50	0	0	3	47
09MN	Nooksack R Winter-Run	NFNooksackJuv	61	0	23	19	19
10PY	Nooksack R Winter-Run	NFNooksackJuv	2	1	0	0	1
12CF	Nooksack R Winter-Run	SFNooksackWinterAd	42	0	0	0	42
09LQ	Nooksack R Winter-Run	SFNooksackSeineJuv	58	0	0	4	54
10GX	South Fork Nooksack R Summer-Run	SFNooksackSummerAd	36	0	0	1	35
11GO	South Fork Nooksack R Summer-Run	SFNooksackSummerAd	31	1	0	1	29
08BN	Samish R Winter-Run	SamishRiver	41	1	0	2	38
12AP	Samish R Winter-Run	SamishRiver	47	0	0	1	46
12DA	Mainstem Skagit R Summer- and Winter-Run	CascadeRiverwinteradultSTHD	13	0	0	0	13
10CQ	Mainstem Skagit R Summer- and Winter-Run	FinneyCreekAdults	22	0	0	0	22
11BK	Mainstem Skagit R Summer- and Winter-Run	FinneyCreekAdults	31	0	0	1	30
12FT	Mainstem Skagit R Summer- and Winter-Run	FinneyCreeksummerSTHD	26	0	0	4	22
10AQ	Sauk R Summer- and Winter-Run	SuiattleAdults	17	0	0	1	16
11BM	Sauk R Summer- and Winter-Run	SuiattleAdults	34	0	0	1	33
08DQ	Mainstem Skagit R Summer- and Winter-Run	upperSkagitRiverAdults	20	1	0	0	19
09BN	Mainstem Skagit R Summer- and Winter-Run	upperSkagitRiverAdults	10	0	0	0	10
10AO	Mainstem Skagit R Summer- and Winter-Run	upperSkagitRiverAdults	25	0	0	0	25
11BI	Mainstem Skagit R Summer- and Winter-Run	upperSkagitRiverAdults	34	0	2	0	32
10NI	Mainstem Skagit R Summer- and Winter-Run	upperSkagitRiverlargeresidentOmykiss	8	0	1	0	7
12AO	Nookachamps Creek Winter-Run	NookachampsCreekjuvenileOmykiss	50	1	4	3	42
09DU	Sauk R Summer- and Winter-Run	SaukRiver	17	0	0	0	17
10AR	Sauk R Summer- and Winter-Run	SaukRiver	24	2	0	1	21
11BN	Sauk R Summer- and Winter-Run	SaukRiver	25	0	0	1	24
13KA	Canyon Creek Summer-Run	CanyonCreekSummerJuv	100	2	2	35	61
95CG	Deer Creek Summer-Run	DeerCreekJuveniles95	48	0	0	20	28
12FL	Deer Creek Summer-Run	DeerCreekSummerAdult	1	0	0	0	1
13GE	Deer Creek Summer-Run	DeerCreekSummerAdult	7	0	0	0	7
13KB	Deer Creek Summer-Run	DeerCreekSummerJuv13	101	0	0	21	80
06BY	NA (sample is aggregate)	StillaguamishRiverSmoltTrap	94	1	5	2	86
04HN	North Fork Skykomish Summer-Run	NFSkyJuv04	47	11	1	6	29
12FK	North Fork Skykomish Summer-Run	NFSkySumAd1213	10	0	0	0	10
13GF	North Fork Skykomish Summer-Run	NFSkySumAd1213	4	0	0	0	4
13LJ	North Fork Skykomish Summer-Run	NFSkySumJuv2013	100	2	2	12	84
12MN	Pilchuck R Winter-Run	PilchuckR12	50	1	0	1	48
13GH	Snohomish / Skykomish R Winter-Run	SkyWinAd13	21	0	0	0	21
11IW	Snoqualmie River Winter-Run	NFToltAboveJuv11	25	0	0	7	18
12IS	Snoqualmie River Winter-Run	NFToltBelowJuv11	50	0	1	3	46
10IX	Snoqualmie River Winter-Run	SFToltBelowJuv10	75	6	1	10	58
13BC	Snoqualmie River Winter-Run	SnoqualmieWinAd13	24	0	0	0	24
10IW	Tolt River Summer-Run	SFToltAboveJuv10	75	0	1	24	50
04AY	Green River Winter-Run	GreenR04	49	0	1	3	45
07CO	Green River Winter-Run	GreenRJuv0708	39	4	3	0	32
08EF	Green River Winter-Run	GreenRJuv0708	54	0	0	0	54
13EH	Green River Winter-Run	GreenRWildWinterBroodstock13	31	0	0	0	31

Table S10. Geographic and temporal scope, biological and management descriptors, and sample size of hatchery-origin collections used in this study. All samples were collected from segregated hatchery programs, either early winter (i.e., Chambers Creek – origin), or early summer (i.e., Skamania – origin), designated here as Operational Units, which were the primary units for analysis. Total N is the number of samples genotyped; *Structure* N is the number of samples used for the *Structure* analyses per OU. Samples were removed if they were missing more than one-third of loci, showed at least one cutthroat allele, or, for the *Structure* analyses, if they were part of a full-sibling group (one individual from the group was retained).

Basin	Hatchery/Program	Code	Collection Year	Life Stage	Origin	Program Type	OperationalUnit	Total N	Removed Poor Genotype	Removed Cut. Alleles	Removed Relatedness	Structure N
Nooksack	Kendall Creek - early winter	01GA	2001	broodstock	hatchery	segregated	Kendall	100	0	0	33	67
Skagit	Marblemount - early winter	08LF	2008	broodstock	hatchery	segregated	MarblemountHatcheryAdults	44	0	4	3	37
Skagit	Marblemount - early winter	09CF	2009	broodstock	hatchery	segregated	MarblemountHatcheryAdults	54	0	0	3	51
Skagit	Marblemount - early winter	10AN	2010	broodstock	hatchery	segregated	MarblemountHatcheryAdults	53	0	0	9	44
Snohomish	Reiter Ponds - early summer	01GG	2001	broodstock	hatchery	segregated	ReiterPonds	39	0	0	3	36
Snohomish	Tokul Creek - early winter	01GC	2001	broodstock	hatchery	segregated	TokulHatchery	40	0	0	3	37
Green	Soos Creek - early winter	03LZ	2003	broodstock	hatchery	segregated	SoosChambers03	44	0	0	2	42
Green	Soos Creek - early summer	03MA	2003	broodstock	hatchery	segregated	SoosSkamania03	90	0	0	24	66

Table S11. SNP loci, with WDFW identifier, assay names, and reference for locus – source. Samples were genotyped using all loci. Loci were removed from analyses for a variety of reasons (see text). A check mark indicated that that locus was used for all analyses in that basin.

WDFW Nickname	Assay name	Reference	Database-wide Status	Genotyped In:						
				Green	Snohomish	Stillaguamish	Skagit	Samish	Nooksack	Cross-Basin
AOmy005	Omy_aspAT-123	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy010	Omy_CRB2677.106	13	Omykiss Genotyping	✓	✓	✓	✓			
AOmy014	Omy_e1-147	13	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy015	Omy_gdh-271	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy016	Omy_GH1P1_2	2	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy021	Omy_LDHB-2_e5	2	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy023	Omy_MYC_2	2	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy026	Omy_myoD.178	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy027	Omy_nkef-241	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy028	Omy_nramp-146	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy029	Omy_Ogo4.212	4	Omykiss Genotyping			✓	✓			
AOmy042	Omy_BAC-F5.284	9	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy047	Omy_u07-79-166	9	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy048	Omy_113490-159	1	Omykiss Genotyping	✓						
AOmy049	Omy_114315-438	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy051	Omy_121713-115	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy056	Omy_128693-455	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy058	Omy_130524-160	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy059	Omy_187760-385	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy061	Omy_96222-125	1	Omykiss Genotyping	✓	✓	✓	✓			
AOmy062	Omy_97077-73	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy065	Omy_97954-618	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy067	Omy_aromat-280	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy068	Omy_arp-630	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy072	Omy_cd59b-112	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy073	Omy_colla1-525	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy074	Omy_cox2-335	16	Removed - too few individuals scored							
AOmy078	Omy_g1-103	14	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy079	Omy_g12-82	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy081	Omy_gh-475	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy082	Omy_gsdf-291	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy084	Omy_hsc715-80	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy087	Omy_hsp47-86	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy088	Omy_hsp70aPro-329	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy089	Omy_hsp90BA-193	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy091	Omy_IL17-185	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy092	Omy_IL1b-163	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy094	Omy_inos-97	16	Removed - No Variation							
AOmy095	Omy_mapK3-103	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy096	Omy_mcsf-268	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy100	Omy_nach-200	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy105	Omy_OmyP9-180	13	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓

Table S11. Continued

WDFW Nickname	Assay name	Reference	Database-wide Status	Genotyped In:						
				Green	Snohomish	Stillaguamish	Skagit	Samish	Nooksack	Cross-Basin
AOmy107	Omy_Ots249-227	4	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy108	Omy_oxct-85	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy110	Omy_star-206	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy111	Omy_stat3-273	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy113	Omy_tlr3-377	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy114	Omy_tlr5-205	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy117	Omy_u09-52-284	9	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy118	Omy_u09-53-469	9	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy120	Omy_u09-54.311	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy123	Omy_u09-55-233	9	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy125	Omy_u09-56-119	9	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy129	Omy_BAMBI4-238	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy132	Omy_G3PD_2.246	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy134	Omy_Il-1b-028	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy137	Omy_u09-61.043	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy144	Omy_UT16_2.173	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy147	Omy_U11_2b.154	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy149	Omy_gluR-79	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy152	Omy_SECC22b-88	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy173	BH2VHSVip10	11	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy174	OMS00003	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy176	OMS00013	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy177	OMS00018	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy179	OMS00041	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy180	OMS00048	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy181	OMS00052	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy182	OMS00053	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy183	OMS00056	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy184	OMS00057	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy185	OMS00061	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy186	OMS00062	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy187	OMS00064	12	Omykiss Genotyping			✓		✓	✓	
AOmy189	OMS00071	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy190	OMS00072	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy191	OMS00078	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy192	OMS00087	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy193	OMS00089	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy194	OMS00090	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy195	OMS00092	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy197	OMS00103	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy198	OMS00105	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy199	OMS00112	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy200	OMS00116	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy201	OMS00118	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓

Table S11. Continued

WDFW Nickname	Assay name	Reference	Database-wide Status	Genotyped In:						
				Green	Snohomish	Stillaguamish	Skagit	Samish	Nooksack	Cross-Basin
AOmy202	OMS00119	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy203	OMS00120	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy204	OMS00121	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy205	OMS00127	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy206	OMS00128	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy207	OMS00132	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy208	OMS00133	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy209	OMS00134	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy210	OMS00153	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy211	OMS00154	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy212	OMS00156	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy213	OMS00164	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy214	OMS00169	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy215	OMS00175	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy216	OMS00176	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy218	OMS00180	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy220	Omy_1004	8	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy221	Omy_101554-306	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy222	Omy_101832-195	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy223	Omy_101993-189	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy225	Omy_102505-102	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy226	Omy_102867-443	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy227	Omy_103705-558	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy228	Omy_104519-624	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy229	Omy_104569-114	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy230	Omy_105075-162	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy231	Omy_105385-406	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy232	Omy_105714-265	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy233	Omy_107031-704	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy234	Omy_107285-69	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy235	Omy_107336-170	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy237	Omy_107806-34	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy238	Omy_108007-193	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy239	Omy_109243-222	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy240	Omy_109525-403	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy241	Omy_110064-419	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy242	Omy_110078-294	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy243	Omy_110362-585	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy244	Omy_110689-148	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy246	Omy_111084-526	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy247	Omy_111383-51	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy248	Omy_111666-301	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy249	Omy_112301-202	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy250	Omy_112820-82	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓

Table S11. Continued

WDFW Nickname	Assay name	Reference	Database-wide Status	Genotyped In:						
				Green	Snohomish	Stillaguamish	Skagit	Samish	Nooksack	Cross-Basin
AOmy252	Omy_114976-223	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy253	Omy_116733-349	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy254	Omy_116938-264	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy255	Omy_117259-96	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy256	Omy_117286-374	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy257	Omy_117370-400	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy258	Omy_117540-259	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy260	Omy_117815-81	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy261	Omy_118175-396	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy262	Omy_118205-116	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy263	Omy_118654-91	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy265	Omy_120255-332	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy266	Omy_128996-481	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy267	Omy_129870-756	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy268	Omy_131460-646	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy269	Omy_98683-165	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy270	Omy_cyp17-153	16	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy271	Omy_ftzf1-217	16	Omykiss Genotyping	✓	✓	✓				
AOmy272	Omy_GHSR-121	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy273	Omy_metA-161	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy274	Omy_UBA3b	8	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy275	M09AAC.055	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy276	M09AAE-082	15	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy277	OMGH1PROM1-SNP1	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy279	OMS00015	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy280	OMS00024	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy283	OMS00070	12	Omykiss Genotyping	✓						
AOmy284	OMS00074	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy285	OMS00096	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy286	OMS00111	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy288	OMS00149	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy289	OMS00173	12	Removed - too few individuals scored							
AOmy290	Omy_105105-448	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy291	Omy_110201-359	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy292	Omy_128923-433	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy293	Omy_anp-17	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy294	Omy_bcAKala-380rd	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy295	Omy_cin-172	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy296	Omy_ndk-152	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy297	Omy_nips-299	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy298	Omy_ntl-27	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy299	Omy_rbm4b-203	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy300	Omy_sys1-188	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy301	Omy_txnip-343	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓

Table S11. Continued

WDFW Nickname	Assay name	Reference	Database-wide Status	Genotyped In:						
				Green	Snohomish	Stillaguamish	Skagit	Samish	Nooksack	Cross-Basin
AOmy302	Omy_vamp5-303	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy303	Omy_vatf-406	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy305	OMS00077	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy306	OMS00101	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy311	Omy_G3PD_2-371	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy320	Omy_redd1-410	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy322	Omy_srp09-37	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy324	Omy1011	8	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy326	OMS00068	12	Omykiss Genotyping		✓	✓	✓	✓	✓	
AOmy327	OMS00079	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy328	OMS00106	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy329	OMS00179	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy331	Omy_114587-480	1	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy335	OMS00017	12	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
AOmy341	Omy_metB-138	5	Omykiss Genotyping	✓	✓	✓	✓	✓	✓	✓
ASpI001	Ocl_Okerca	10	Omykiss-Oclarki introgression ID only							
ASpI014	Omy_F5_136	6	Omykiss-Oclarki introgression ID only							
ASpI018	Omy_Omyclmk436-96	5	Omykiss-Oclarki introgression ID only							
			Total	183	182	184	182	180	180	178

1 Abadia-Cardoso et al. 2011
2 Aguilar and Garza 2008
3 Brunelli et al. 2008
4 Campbell and Narum 2009
5 CRITFC - N Campbell unpubl.
6 Finger et al. 2009
7 NOAA – JC Garza unpubl.
8 Hansen et al. 2011
9 Limborg et al. 2011
10 McGlauflin et al. 2010
11 UW – C Pascal and M Hansen unpubl.
12 Sánchez et al. 2009
13 Sprowles et al. 2006
14 Stephens et al. 2009
15 WDFW - S. Young unpubl.
16 WSU-J. DeKoning unpubl.

Table S12. Operational Units' estimated spawning proportion within specific DIPs (A. Hoffmann, WDFW, pers. comm. 2014)

PSSTRT DIP	Operational Unit	Proportion of Spawning within DIP
Snoqualmie River Winter-Run	NFToltAbove&BelowJuv11	0.0411
Snoqualmie River Winter-Run	SFToltBelowJuv10	0.0589
Snoqualmie River Winter-Run	SnoqualmieWinAd13	0.9000
Mainstem Skagit R Summer- and Winter-Run	CascadeRiverwinteradultSTHD	0.1194
Mainstem Skagit R Summer- and Winter-Run	upperSkagitRiverAdults	0.7384
Mainstem Skagit R Summer- and Winter-Run	FinneyCreekSummer&WinterAdults	0.1422
Sauk R Summer- and Winter-Run	SaukRiver	0.7149
Sauk R Summer- and Winter-Run	SuiattleAdults	0.2851
Nooksack R Winter-Run	MainstemNookEarlyAd	0.2754
Nooksack R Winter-Run	SFNooksackWinterAd	0.2484
Nooksack R Winter-Run	NFNooksackAd&Juv	0.4762

Table S13. Unadjusted and likelihood adjusted proportions for OUs and DIPs from the Green River, using *Structure's* default no prior population information mode. Below the adjusted proportions are the 90% confidence intervals.

	Unadjusted						Adjusted					
Unit	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid
Operational Unit												
GreenRAd04	0.02	0.10	0.00	0.02	0.85	0.00	0.01	0.00	0.00	0.03	0.96	0.00
	-	-	-	-	-	-	(0.00 - 0.05)	(0.00 - 0.12)	(0.00 - 0.00)	(0.02 - 0.05)	(0.88 - 0.96)	(0.00 - 0.00)
GreenRJuv0708	0.12	0.17	0.00	0.01	0.69	0.01	0.11	0.09	0.00	0.02	0.79	0.00
	-	-	-	-	-	-	(0.06 - 0.16)	(0.00 - 0.21)	(0.00 - 0.00)	(0.01 - 0.02)	(0.71 - 0.88)	(0.00 - 0.03)
GreenRAd13	0.00	0.03	0.00	0.00	0.94	0.03	0.00	0.00	0.00	0.00	0.96	0.04
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.09)	(0.00 - 0.00)	(0.00 - 0.00)	(0.91 - 0.96)	(0.01 - 0.08)
DIP - All samples												
Green River Winter-Run	0.07	0.13	0.00	0.01	0.78	0.01	0.06	0.00	0.00	0.01	0.92	0.00
	-	-	-	-	-	-	(0.03 - 0.09)	(0.00 - 0.09)	(0.00 - 0.00)	(0.01 - 0.02)	(0.86 - 0.98)	(0.00 - 0.02)

Table S14. Unadjusted and likelihood adjusted F1 Hybrid (introgression) and PEHC for OUs and DIPs from the Green River. Below the adjusted F1 Hybrid and PEHC values are the 90% confidence intervals. See Table S13 for proportions.

Unit	Unadjusted				Adjusted			
	F1 $Hybrids_{winter}$	F1 $Hybrids_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$	F1 $Hybrids_{winter}$	F1 $Hybrids_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$
Operational Unit								
GreenRAd04	0.10	0.00	0.07	0.02	0.00	0.00	0.01	0.03
	-	-	-	-	(0.00 - 0.12)	(0.00 - 0.00)	(0.00 - 0.12)	(0.02 - 0.05)
GreenRJuv0708	0.17	0.00	0.20	0.01	0.09	0.00	0.15	0.02
	-	-	-	-	(0.00 - 0.21)	(0.00 - 0.00)	(0.06 - 0.26)	(0.01 - 0.02)
GreenRAd13	0.03	0.00	0.02	0.00	0.00	0.00	0.00	0.00
	-	-	-	-	(0.00 - 0.09)	(0.00 - 0.00)	(0.00 - 0.04)	(0.00 - 0.00)

Table S15. Unadjusted and likelihood adjusted proportions for OUs and DIPs from the Snohomish River, using *Structure's* default no prior population information mode. Below the adjusted proportions are the 90% confidence intervals. Proportions in bold typeface are those where the range of the 90% confidence interval for the likelihood adjustment exceeded 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted						Adjusted					
	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid
Operational Unit												
NFSkyAd1213	0.00	0.07	0.00	0.93	0.00	0.00	0.00	0.11	0.00	0.89	0.00	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.06 - 0.25)	(0.00 - 0.00)	(0.81 - 0.89)	(0.00 - 0.00)	(0.00 - 0.00)
NFSkyJuv04	0.00	0.06	0.12	0.74	0.09	0.00	0.00	0.07	0.08	0.77	0.08	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.02 - 0.15)	(0.00 - 0.18)	(0.68 - 0.88)	(0.04 - 0.15)	(0.00 - 0.00)
NFSkyJuv13	0.00	0.00	0.00	0.98	0.00	0.02	0.00	0.00	0.00	1.00	0.00	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.00)	(0.00 - 0.00)	(0.99 - 1.00)	(0.00 - 0.00)	(0.00 - 0.02)
NFToltAboveJuv11	0.00	0.23	0.15	0.23	0.38	0.00	0.00	**0.25**	**0.14**	0.24	**0.37**	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	**(0.07 - 0.53)**	**(0.04 - 0.29)**	(0.16 - 0.34)	**(0.21 - 0.54)**	(0.00 - 0.00)
NFToltBelowJuv11	0.00	0.08	0.02	0.02	0.88	0.00	0.00	0.00	0.00	0.03	0.97	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.10)	(0.00 - 0.05)	(0.02 - 0.05)	(0.90 - 0.97)	(0.00 - 0.00)
PilchuckR12	0.02	0.10	0.02	0.02	0.84	0.00	0.01	0.00	0.00	0.03	0.95	0.00
	-	-	-	-	-	-	(0.00 - 0.06)	(0.00 - 0.13)	(0.00 - 0.05)	(0.02 - 0.05)	(0.85 - 0.95)	(0.00 - 0.00)
SFToltAboveJuv10	0.00	0.03	0.45	0.44	0.07	0.01	0.00	0.02	0.51	0.44	0.04	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.06)	(0.42 - 0.60)	(0.38 - 0.50)	(0.00 - 0.09)	(0.00 - 0.03)
SFToltBelowJuv10	0.02	0.12	0.14	0.22	0.48	0.03	0.00	0.09	0.12	0.24	0.55	0.00
	-	-	-	-	-	-	(0.00 - 0.04)	(0.00 - 0.20)	(0.05 - 0.19)	(0.20 - 0.29)	(0.47 - 0.64)	(0.00 - 0.05)
SkyWinAd13	0.00	0.00	0.00	0.05	0.95	0.00	0.00	0.00	0.00	0.05	0.95	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.00)	(0.00 - 0.00)	(0.03 - 0.09)	(0.88 - 0.95)	(0.00 - 0.00)
SnoqWinAd13	0.04	0.17	0.00	0.00	0.79	0.00	0.02	**0.08**	0.00	0.00	0.90	0.00
	-	-	-	-	-	-	(0.00 - 0.11)	**(0.00 - 0.32)**	(0.00 - 0.00)	(0.00 - 0.00)	(0.74 - 0.92)	(0.00 - 0.00)
DIP - All samples												
North Fork Skykomish Summer-Run	0.00	0.02	0.03	0.92	0.02	0.01	0.00	0.03	0.00	0.95	0.02	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.01 - 0.05)	(0.00 - 0.02)	(0.94 - 0.95)	(0.01 - 0.04)	(0.00 - 0.02)
Tolt River Summer-Run	0.00	0.03	0.45	0.44	0.07	0.01	0.00	0.02	0.51	0.44	0.04	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.06)	(0.42 - 0.60)	(0.38 - 0.50)	(0.00 - 0.09)	(0.00 - 0.03)
Snoqualmie River Winter-Run	0.01	0.07	0.00	0.04	0.89	0.00	0.01	0.06	0.00	0.03	0.89	0.00
	-	-	-	-	-	-	(0.00 - 0.04)	(0.00 - 0.15)	(0.00 - 0.02)	(0.03 - 0.05)	(0.82 - 0.96)	(0.00 - 0.01)
Snohomish / Skykomish R Winter-Run	0.00	0.00	0.00	0.05	0.95	0.00	0.00	0.00	0.00	0.05	0.95	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.00)	(0.00 - 0.00)	(0.03 - 0.09)	(0.88 - 0.95)	(0.00 - 0.00)
Pilchuck R Winter-Run	0.02	0.10	0.02	0.02	0.84	0.00	0.01	0.00	0.00	0.03	0.95	0.00
	-	-	-	-	-	-	(0.00 - 0.06)	(0.00 - 0.13)	(0.00 - 0.05)	(0.02 - 0.05)	(0.85 - 0.95)	(0.00 - 0.00)

Table S16. Unadjusted and likelihood adjusted F1 Hybrid (introgression) and PEHC for OUs and DIPs from the Snohomish River. Below the adjusted F1 Hybrid and PEHC values are the 90% confidence intervals. Bold typeface for F1 Hybrid and PEHC values indicate that the contributing proportions from Table S15 were uncertain as a result of their 90% confidence interval exceeding 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted				Adjusted			
	F1 $Hybrid_{winter}$	F1 $Hybrid_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$	F1 $Hybrid_{winter}$	F1 $Hybrid_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$
Operational Unit								
NFSkyAd1213	0.07	0.00	0.04	0.93	0.11	0.00	0.05	0.89
	-	-	-	-	(0.06 - 0.25)	(0.00 - 0.00)	(0.03 - 0.13)	(0.81 - 0.89)
NFSkyJuv04	0.06	0.12	0.03	0.79	0.07	0.08	0.03	0.81
	-	-	-	-	(0.02 - 0.15)	(0.00 - 0.18)	(0.01 - 0.08)	(0.68 - 0.97)
NFSkyJuv13	0.00	0.00	0.00	0.98	0.00	0.00	0.00	1.00
	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.00)	(0.00 - 0.00)	(0.99 - 1.00)
NFToltAboveJuv11	0.23	0.15	0.12	0.31	**0.25**	**0.14**	**0.13**	**0.31**
	-	-	-	-	**(0.07 - 0.53)**	**(0.04 - 0.29)**	**(0.03 - 0.26)**	**(0.18 - 0.49)**
NFToltBelowJuv11	0.08	0.02	0.04	0.03	0.00	0.00	0.00	0.03
	-	-	-	-	(0.00 - 0.10)	(0.00 - 0.05)	(0.00 - 0.05)	(0.02 - 0.08)
PilchuckR12	0.10	0.02	0.07	0.03	0.00	0.00	0.01	0.03
	-	-	-	-	(0.00 - 0.13)	(0.00 - 0.05)	(0.00 - 0.12)	(0.02 - 0.08)
SFToltAboveJuv10	0.03	0.45	0.01	0.66	0.02	0.51	0.01	0.69
	-	-	-	-	(0.00 - 0.06)	(0.42 - 0.60)	(0.00 - 0.03)	(0.59 - 0.80)
SFToltBelowJuv10	0.12	0.14	0.08	0.28	0.09	0.12	0.04	0.30
	-	-	-	-	(0.00 - 0.20)	(0.05 - 0.19)	(0.00 - 0.14)	(0.23 - 0.38)
SkyWinAd13	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.05
	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.00)	(0.00 - 0.00)	(0.03 - 0.09)
SnoqWinAd13	0.17	0.00	0.13	0.00	**0.08**	0.00	**0.06**	0.00
	-	-	-	-	**(0.00 - 0.32)**	(0.00 - 0.00)	**(0.00 - 0.27)**	(0.00 - 0.00)

Table S17. Unadjusted and likelihood adjusted proportions for OUs and DIPs from the Stillaguamish River, using *Structure's* default no prior population information mode. Below the adjusted proportions are the 90% confidence intervals. Proportions in bold typeface are those where the range of the 90% confidence interval for the likelihood adjustment exceeded 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted						Adjusted					
	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid
Operational Unit												
CanyonCreekSuJuv13	0.00	0.03	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.03)	(0.00 - 0.00)	(0.00 - 0.00)	(0.98 - 1.00)	(0.00 - 0.00)
DeerCreekSuAd1213	0.00	0.00	0.13	0.00	0.88	0.00	0.00	0.00	**0.13**	0.00	0.87	0.00
							(0.00 - 0.00)	(0.00 - 0.00)	**(0.00 - 0.34)**	(0.00 - 0.00)	(0.71 - 0.87)	(0.00 - 0.00)
DeerCreekSuJuv13	0.00	0.01	0.03	0.00	0.96	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.03)	(0.00 - 0.03)	(0.00 - 0.00)	(0.98 - 1.00)	(0.00 - 0.00)
DeerCreekSuJuv95	0.00	0.06	0.15	0.00	0.77	0.02	0.00	0.00	0.14	0.00	0.85	0.02
							(0.00 - 0.00)	(0.00 - 0.08)	(0.07 - 0.21)	(0.00 - 0.00)	(0.74 - 0.91)	(0.00 - 0.05)
StillaguamishRSmoltTrap06	0.01	0.08	0.11	0.13	0.64	0.02	0.00	0.00	0.09	0.14	0.77	0.00
							(0.00 - 0.03)	(0.00 - 0.08)	(0.04 - 0.15)	(0.11 - 0.17)	(0.69 - 0.85)	(0.00 - 0.04)
DIP - All samples												
Canyon Creek Summer-Run	0.00	0.03	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.03)	(0.00 - 0.00)	(0.00 - 0.00)	(0.98 - 1.00)	(0.00 - 0.00)
Deer Creek Summer-Run	0.00	0.03	0.07	0.00	0.90	0.01	0.00	0.00	0.03	0.00	0.97	0.00
							(0.00 - 0.00)	(0.00 - 0.02)	(0.00 - 0.07)	(0.00 - 0.00)	(0.95 - 0.97)	(0.00 - 0.01)

Table S18. Unadjusted and likelihood adjusted F1 Hybrid (introgression) and PEHC for OUs and DIPs from the Stillaguamish River. Below the adjusted F1 Hybrid and PEHC values are the 90% confidence intervals. Bold typeface for F1 Hybrid and PEHC values indicate that the contributing proportions from Table S17 were uncertain as a result of their 90% confidence interval exceeding 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted				Adjusted			
	F1 $Hybrids_{winter}$	F1 $Hybrids_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$	F1 $Hybrids_{winter}$	F1 $Hybrids_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$
Operational Unit								
CanyonCreekSuJuv13	0.03	0.00	0.02	0.00	0.00	0.00	0.00	0.00
					(0.00 - 0.03)	(0.00 - 0.00)	(0.00 - 0.02)	(0.00 - 0.00)
DeerCreekSuAd1213	0.00	0.13	0.00	0.06	0.00	**0.13**	0.00	**0.06**
					(0.00 - 0.00)	**(0.00 - 0.34)**	(0.00 - 0.00)	**(0.00 - 0.17)**
DeerCreekSuJuv13	0.01	0.03	0.00	0.01	0.00	0.00	0.00	0.00
					(0.00 - 0.03)	(0.00 - 0.03)	(0.00 - 0.01)	(0.00 - 0.01)
DeerCreekSuJuv95	0.06	0.15	0.03	0.07	0.00	0.14	0.00	0.07
					(0.00 - 0.08)	(0.07 - 0.21)	(0.00 - 0.04)	(0.04 - 0.11)
StillaguamishRSmoltTrap06	0.08	0.11	0.05	0.18	0.00	0.09	0.00	0.18
					(0.00 - 0.08)	(0.04 - 0.15)	(0.00 - 0.07)	(0.13 - 0.25)

Table S19. Unadjusted and likelihood adjusted proportions for OUs and DIPs from the Skagit River, using *Structure's* default no prior population information mode. Below the adjusted proportions are the 90% confidence intervals. Proportions in bold typeface are those where the range of the 90% confidence interval for the likelihood adjustment exceeded 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted						Adjusted					
	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid
Operational Unit												
CascadeRWiAd	0.00	0.15	0.00	0.00	0.85	0.00	0.00	**0.06**	0.00	0.00	0.94	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	**(0.00 - 0.38)**	(0.00 - 0.00)	(0.00 - 0.00)	(0.75 - 0.94)	(0.00 - 0.00)
FinneyCreekWiAd	0.02	0.02	0.00	0.00	0.96	0.00	0.02	0.00	0.00	0.00	0.98	0.00
	-	-	-	-	-	-	(0.01 - 0.05)	(0.00 - 0.04)	(0.00 - 0.00)	(0.00 - 0.00)	(0.95 - 0.98)	(0.00 - 0.00)
FinneyCreekSuAd	0.04	0.15	0.04	0.04	0.73	0.00	0.03	**0.07**	0.00	0.05	**0.85**	0.00
	-	-	-	-	-	-	(0.00 - 0.10)	**(0.00 - 0.30)**	(0.00 - 0.10)	(0.03 - 0.09)	**(0.69 - 0.96)**	(0.00 - 0.00)
NookachampsCreekJuv	0.02	0.05	0.00	0.00	0.93	0.00	0.02	0.00	0.00	0.00	0.98	0.00
	-	-	-	-	-	-	(0.00 - 0.06)	(0.00 - 0.08)	(0.00 - 0.00)	(0.00 - 0.00)	(0.93 - 0.98)	(0.00 - 0.00)
SaukRAd	0.05	0.03	0.00	0.00	0.92	0.00	0.05	0.00	0.00	0.00	0.95	0.00
	-	-	-	-	-	-	(0.03 - 0.09)	(0.00 - 0.04)	(0.00 - 0.00)	(0.00 - 0.00)	(0.91 - 0.95)	(0.00 - 0.00)
SuiattleRAd	0.00	0.02	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	-	-	-	-	-	-	(0.00 - 0.00)	(0.00 - 0.06)	(0.00 - 0.00)	(0.00 - 0.00)	(0.97 - 1.00)	(0.00 - 0.00)
upperSkagitRAd	0.04	0.09	0.04	0.01	0.81	0.01	0.02	0.00	0.00	0.01	0.96	0.00
	-	-	-	-	-	-	(0.00 - 0.06)	(0.00 - 0.08)	(0.00 - 0.04)	(0.00 - 0.03)	(0.87 - 0.97)	(0.00 - 0.03)
upperSkagitRLargeResident	0.00	0.14	0.00	0.00	0.86	0.00	0.00	**0.05**	0.00	0.00	0.95	0.00
							(0.00 - 0.00)	**(0.00 - 0.53)**	(0.00 - 0.00)	(0.00 - 0.00)	(0.70 - 0.95)	(0.00 - 0.00)
DIP - All samples												
Mainstem Skagit R Summer- & Winter-Run	0.03	0.09	0.03	0.01	0.83	0.01	0.02	0.00	0.00	0.01	0.96	0.00
	-	-	-	-	-	-	(0.00 - 0.04)	(0.00 - 0.04)	(0.00 - 0.02)	(0.01 - 0.02)	(0.92 - 0.96)	(0.00 - 0.02)
Sauk R Summer- and Winter-Run	0.04	0.03	0.00	0.00	0.94	0.00	0.04	0.00	0.00	0.00	0.96	0.00
	-	-	-	-	-	-	(0.02 - 0.06)	(0.00 - 0.03)	(0.00 - 0.00)	(0.00 - 0.00)	(0.95 - 0.96)	(0.00 - 0.00)
Nookachamps Creek Winter-Run	0.02	0.05	0.00	0.00	0.93	0.00	0.02	0.00	0.00	0.00	0.98	0.00
	-	-	-	-	-	-	(0.00 - 0.06)	(0.00 - 0.08)	(0.00 - 0.00)	(0.00 - 0.00)	(0.93 - 0.98)	(0.00 - 0.00)

Table S20. Unadjusted and likelihood adjusted F1 Hybrid (introgression) and PEHC for OUs and DIPs from the Skagit River. Below the adjusted F1 Hybrid and PEHC values are the 90% confidence intervals. Bold typeface for F1 Hybrid and PEHC values indicate that the contributing proportions from Table S19 were uncertain as a result of their 90% confidence interval exceeding 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted				Adjusted			
	F1 $Hybrid_{winter}$	F1 $Hybrid_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$	F1 $Hybrid_{winter}$	F1 $Hybrid_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$
Operational Unit								
CascadeRWiAd	0.15	0.00	0.08	0.00	**0.06**	0.00	**0.03**	0.00
	-	-	-	-	**(0.00 - 0.38)**	(0.00 - 0.00)	**(0.00 - 0.19)**	(0.00 - 0.00)
FinneyCreekWiAd	0.02	0.00	0.03	0.00	0.00	0.00	0.02	0.00
	-	-	-	-	(0.00 - 0.04)	(0.00 - 0.00)	(0.01 - 0.07)	(0.00 - 0.00)
FinneyCreekSuAd	0.15	0.04	0.12	0.06	**0.07**	0.00	**0.06**	0.05
	-	-	-	-	**(0.00 - 0.30)**	(0.00 - 0.10)	**(0.00 - 0.25)**	(0.03 - 0.14)
NookachampsCreekJuv	0.05	0.00	0.05	0.00	0.00	0.00	0.02	0.00
	-	-	-	-	(0.00 - 0.08)	(0.00 - 0.00)	(0.00 - 0.10)	(0.00 - 0.00)
SaukRAd	0.03	0.00	0.06	0.00	0.00	0.00	0.05	0.00
	-	-	-	-	(0.00 - 0.04)	(0.00 - 0.00)	(0.03 - 0.11)	(0.00 - 0.00)
SuiattleRAd	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	-	-	-	-	(0.00 - 0.06)	(0.00 - 0.00)	(0.00 - 0.03)	(0.00 - 0.00)
upperSkagitRAd	0.09	0.04	0.08	0.03	0.00	0.00	0.02	0.01
	-	-	-	-	(0.00 - 0.08)	(0.00 - 0.04)	(0.00 - 0.10)	(0.00 - 0.05)
upperSkagitRLargeResident	0.14	0.00	0.07	0.00	**0.05**	0.00	**0.02**	0.00
	-	-	-	-	**(0.00 - 0.53)**	(0.00 - 0.00)	**(0.00 - 0.27)**	(0.00 - 0.00)

Table S21. Unadjusted and likelihood adjusted proportions for OUs and DIPs from the Nooksack and Samish rivers, using *Structure's* default no prior population information mode. Below the adjusted proportions are the 90% confidence intervals. Proportions in bold typeface are those where the range of the 90% confidence interval for the likelihood adjustment exceeded 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted						Adjusted					
	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid	EWH Lineage	EWH-Wild Hybrid	ESH-Wild Hybrid	ESH Lineage	Wild	EWH-ESH Hybrid
Operational Unit												
NFNookackAd12	0.00	0.04	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.07)	(0.00 - 0.00)	(0.00 - 0.00)	(0.96 - 1.00)	(0.00 - 0.00)
NFNooksackJuv0910	0.00	0.05	0.00	0.00	0.95	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.09)	(0.00 - 0.00)	(0.00 - 0.00)	(0.95 - 1.00)	(0.00 - 0.00)
MainstemNooksackEarlyAd	0.04	0.05	0.04	0.00	0.86	0.02	0.03	0.00	0.00	0.00	0.95	0.02
							(0.00 - 0.07)	(0.00 - 0.06)	(0.00 - 0.05)	(0.00 - 0.00)	(0.89 - 0.95)	(0.00 - 0.05)
SFNooksackSeineJuv09	0.00	0.12	0.02	0.00	0.86	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.15)	(0.00 - 0.04)	(0.00 - 0.00)	(0.92 - 1.00)	(0.00 - 0.00)
SFNooksackSuAd1011	0.00	0.02	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.05)	(0.00 - 0.00)	(0.00 - 0.00)	(0.97 - 1.00)	(0.00 - 0.00)
SFNooksackWiAd12	0.02	0.02	0.05	0.00	0.90	0.00	0.03	0.00	0.00	0.00	0.97	0.00
							(0.01 - 0.06)	(0.00 - 0.07)	(0.00 - 0.08)	(0.00 - 0.00)	(0.92 - 0.97)	(0.00 - 0.00)
SamishRAd0812	0.03	0.19	0.03	0.00	0.74	0.00	0.01	0.11	0.00	0.00	0.88	0.00
							(0.00 - 0.05)	(0.00 - 0.24)	(0.00 - 0.04)	(0.00 - 0.00)	(0.79 - 0.98)	(0.00 - 0.00)
DIP - All samples												
Nooksack R Winter-Run	0.01	0.06	0.02	0.00	0.91	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.01)	(0.00 - 0.02)	(0.00 - 0.01)	(0.00 - 0.00)	(0.99 - 1.00)	(0.00 - 0.01)
South Fork Nooksack R Summer-Run	0.00	0.02	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00	1.00	0.00
							(0.00 - 0.00)	(0.00 - 0.05)	(0.00 - 0.00)	(0.00 - 0.00)	(0.97 - 1.00)	(0.00 - 0.00)
Samish River	0.03	0.19	0.03	0.00	0.74	0.00	0.01	0.11	0.00	0.00	0.88	0.00
							(0.00 - 0.05)	(0.00 - 0.24)	(0.00 - 0.04)	(0.00 - 0.00)	(0.79 - 0.98)	(0.00 - 0.00)

Table S22. Unadjusted and likelihood adjusted F1 Hybrid (introgression) and PEHC for OUs and DIPs from the Nooksack and Samish rivers. Below the adjusted F1 Hybrid and PEHC values are the 90% confidence intervals. Bold typeface for F1 Hybrid and PEHC values indicate that the contributing proportions from Table S21 were uncertain as a result of their 90% confidence interval exceeding 0.25, reducing confidence in the estimate of the proportion.

Unit	Unadjusted				Adjusted			
	F1 $Hybrid_{winter}$	F1 $Hybrid_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$	F1 $Hybrid_{winter}$	F1 $Hybrid_{summer}$	$PEHC_{winter}$	$PEHC_{summer}$
Operational Unit								
NFNookackAd12	0.04	0.00	0.02	0.00	0.00	0.00	0.00	0.00
					(0.00 - 0.07)	(0.00 - 0.00)	(0.00 - 0.04)	(0.00 - 0.00)
NFNooksackJuv0910	0.05	0.00	0.03	0.00	0.00	0.00	0.00	0.00
					(0.00 - 0.09)	(0.00 - 0.00)	(0.00 - 0.05)	(0.00 - 0.00)
MainstemNooksackEarlyAd	0.05	0.04	0.06	0.02	0.00	0.00	0.03	0.00
					(0.00 - 0.06)	(0.00 - 0.05)	(0.00 - 0.10)	(0.00 - 0.03)
SFNooksackSeineJuv09	0.12	0.02	0.06	0.01	0.00	0.00	0.00	0.00
					(0.00 - 0.15)	(0.00 - 0.04)	(0.00 - 0.07)	(0.00 - 0.02)
SFNooksackSuAd1011	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00
					(0.00 - 0.05)	(0.00 - 0.00)	(0.00 - 0.02)	(0.00 - 0.00)
SFNooksackWiAd12	0.02	0.05	0.04	0.02	0.00	0.00	0.03	0.00
					(0.00 - 0.07)	(0.00 - 0.08)	(0.01 - 0.10)	(0.00 - 0.04)
SamishRAd0812	0.19	0.03	0.13	0.02	**0.11**	0.00	**0.06**	0.00
					(0.00 - 0.24)	(0.00 - 0.04)	**(0.00 - 0.17)**	(0.00 - 0.02)

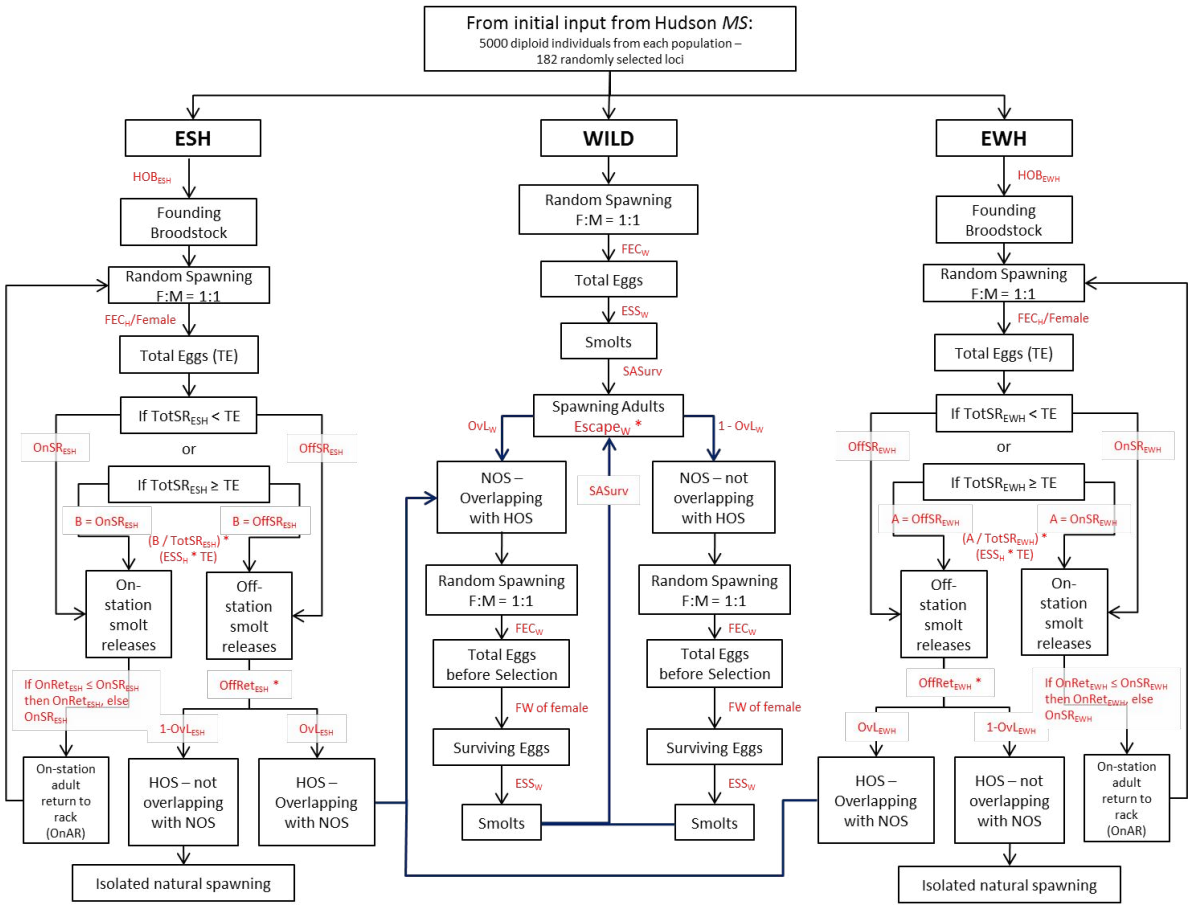


Figure S1. Schema representing the model to simulate hatchery phase natural-origin, EWH, and ESH populations in north Puget Sound, modeled after empirical populations and hatchery practices that existed prior to 2009. Wild, EWH, and ESH populations follow separate pathways, connected at the overlap between wild and hatchery individuals spawning naturally (blue arrows). Since EWH and ESH are segregated programs there are no feedbacks from the Wild pathway to the hatchery pathways. Text in red are the parameters in Table S6. See text for more details.

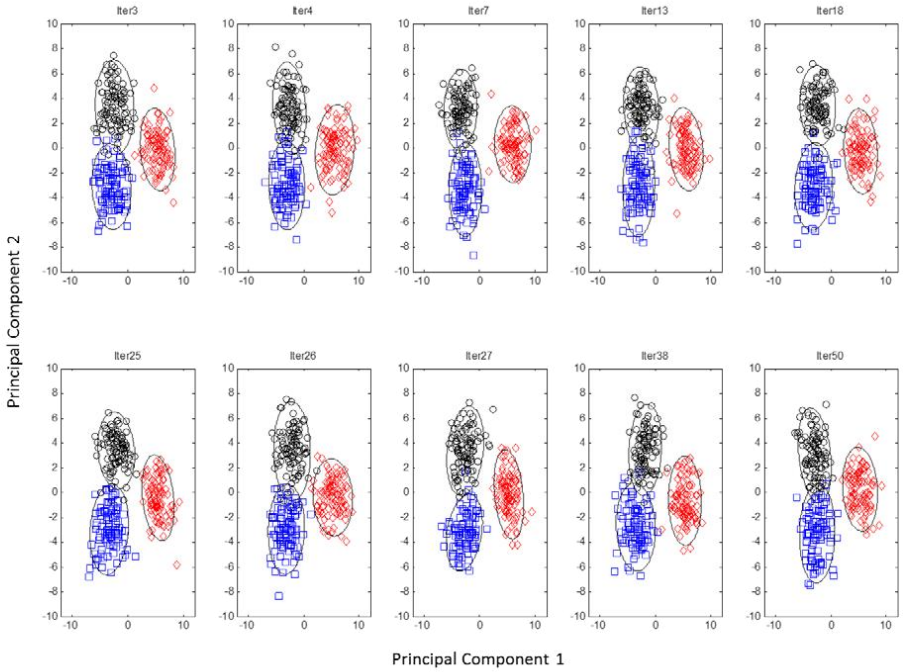


Figure S2. Principal component analyses for the 10 iterations (pre-hatchery phase simulated population sets). Black circles = "NPS", blue squared = "CC", and red diamonds = "LC" populations. Ellipses represent 95% confidence ellipse of principal component scores across components 1 and 2. The first two principal components accounted for approximately 7% of the total molecular variation for all iterations except Iter7 where it accounted for 8%. These two components clearly separated the NPS, CC, and LC populations, but appropriately showed less differentiation between the Puget Sound populations (NPS and CC) than that between these populations and the LC population (mean Mahalanobis D^2 = 3.16, 3.55, 3.59 for CC – NPS, LC – NPS, and CC – LC, respectively; all D^2 significantly > 0). The within population diversity, represented by the spread of points within each cluster and their 95% confidence ellipses, was similar among populations within an iteration, and across all iterations.

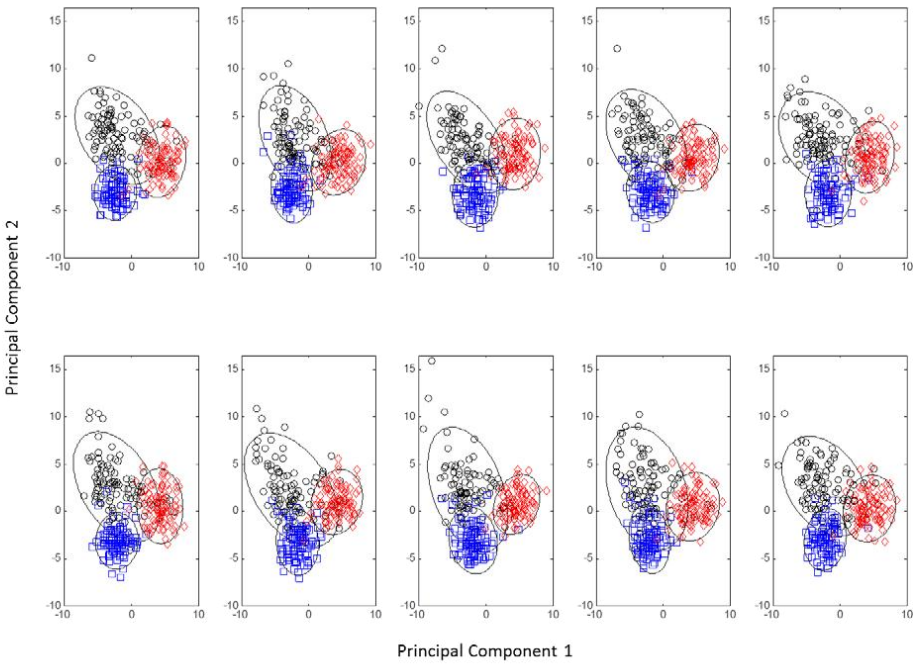


Figure S3. Principal component analyses for 10 populations composed of natural-origin, EWH, and ESH individuals from empirical data sets. To construct these 10 populations I pooled all samples from the empirical data into natural-origin, EWH, and ESH bins (Table S2) and randomly drew, without replacement, 100 samples from each bin (300 individuals in total). I repeated this procedure 10 times to generate 10 independent principal component analyses. For these analyses I removed the Soos Creek - early winter samples as that collection was composed of roughly one-third Soos Creek - early summer fish (see Section 3). As with the principal component analyses on the modeled populations (Figure S2) the first two principal components accounted for approximately 7% of the total molecular variation for all analyses. Although there was clear separation between the wild, EWH, and ESH populations (mean D^2 = 2.56, 2.72, 3.32 for EWH – natural-origin, ESH – natural-origin, and EWH – ESH, respectively; all D^2 significantly > 0), this separation was less pronounced than with the simulated populations, most likely reflecting the added human facilitated gene flow that may have occurred in these present-day empirical populations. Black circles = natural-origin, blue squared = EWH, and red diamonds = ESH populations. Ellipses represent 95% confidence ellipse of principal component scores across components 1 and 2.

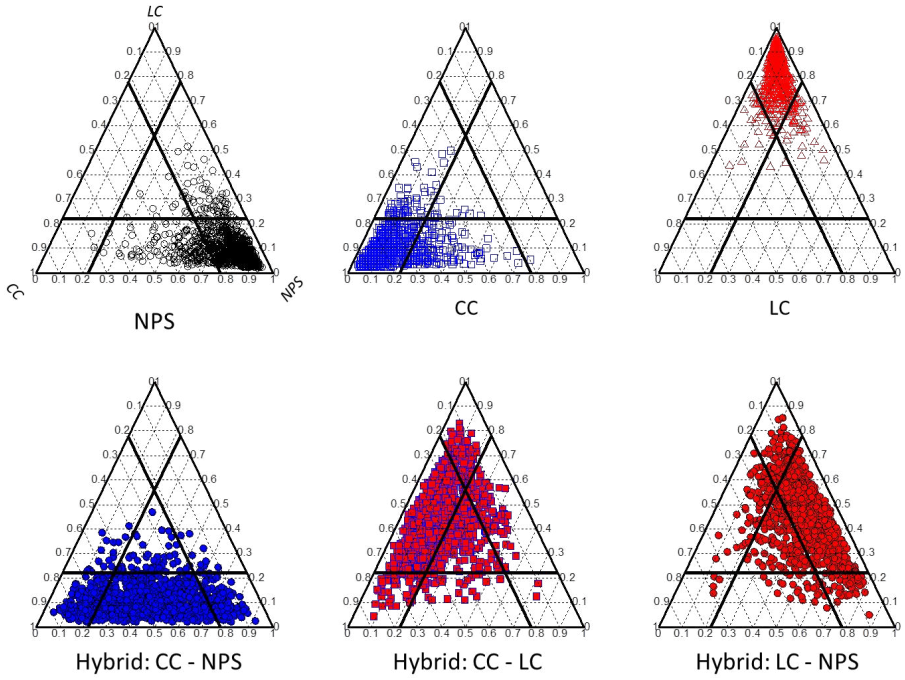


Figure S4. Distribution of the NPS (bottom axis on each plot), CC (left axis), and LC (right axis) assigned category Q-scores from k = 3 *Structure* analysis of the pre-hatchery phase simulated populations, using the default no prior population information mode, for the six source categories (one ternary plot per source category). Each simulated individual's Q-score is represented by a symbol (e.g., NPS individuals are shown in the upper left ternary plot as black circles). Correct assignment for NPS would be in the lower right corner of the upper left ternary plot, for CC in the lower left corner of the upper middle plot, and for LC in the upper corner of the upper right plot. Correct assignments for each of the hybrid source categories would be in the respective trapezoid-shaped polygon. For example, correct assignments for Hybrid: CC – NPS would be in the lower trapezoid of the lower left plot. The triangular area within the middle of each ternary plot is the No Call zone where individuals are not identified. See Figure 2. The assignments using threshold value = 0.22 (solid black lines) for the source categories are indicated in Table 1.

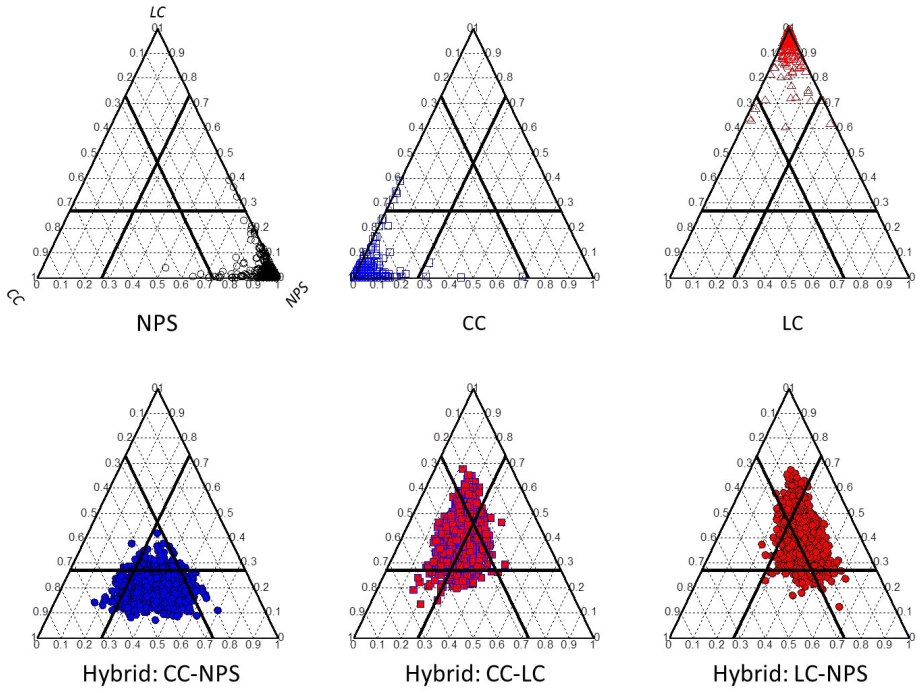


Figure S5. Distribution of the NPS (bottom axis on each plot), CC (left axis), and LC (right axis) assigned category Q-scores from k = 3 *Structure* analysis of the pre-hatchery phase simulated populations, using the prior population information mode, for the six source categories (one ternary plot per source category). Each simulated individual's Q-score is represented by a symbol (e.g., NPS individuals are shown in the upper left ternary plot as black circles). Correct assignment for NPS would be in the lower right corner of the upper left ternary plot, for CC in the lower left corner of the upper middle plot, and for LC in the upper corner of the upper right plot. Correct assignments for each of the hybrid source categories would be in the respective trapezoid-shaped polygon. For example, correct assignments for Hybrid: CC – NPS would be in the lower trapezoid of the lower left plot. The triangular area within the middle of each ternary plot is the No Call zone where individuals are not identified. See Figure 2. The assignments using threshold value = 0.27 (solid black lines) for the source categories are indicated in Table 2.

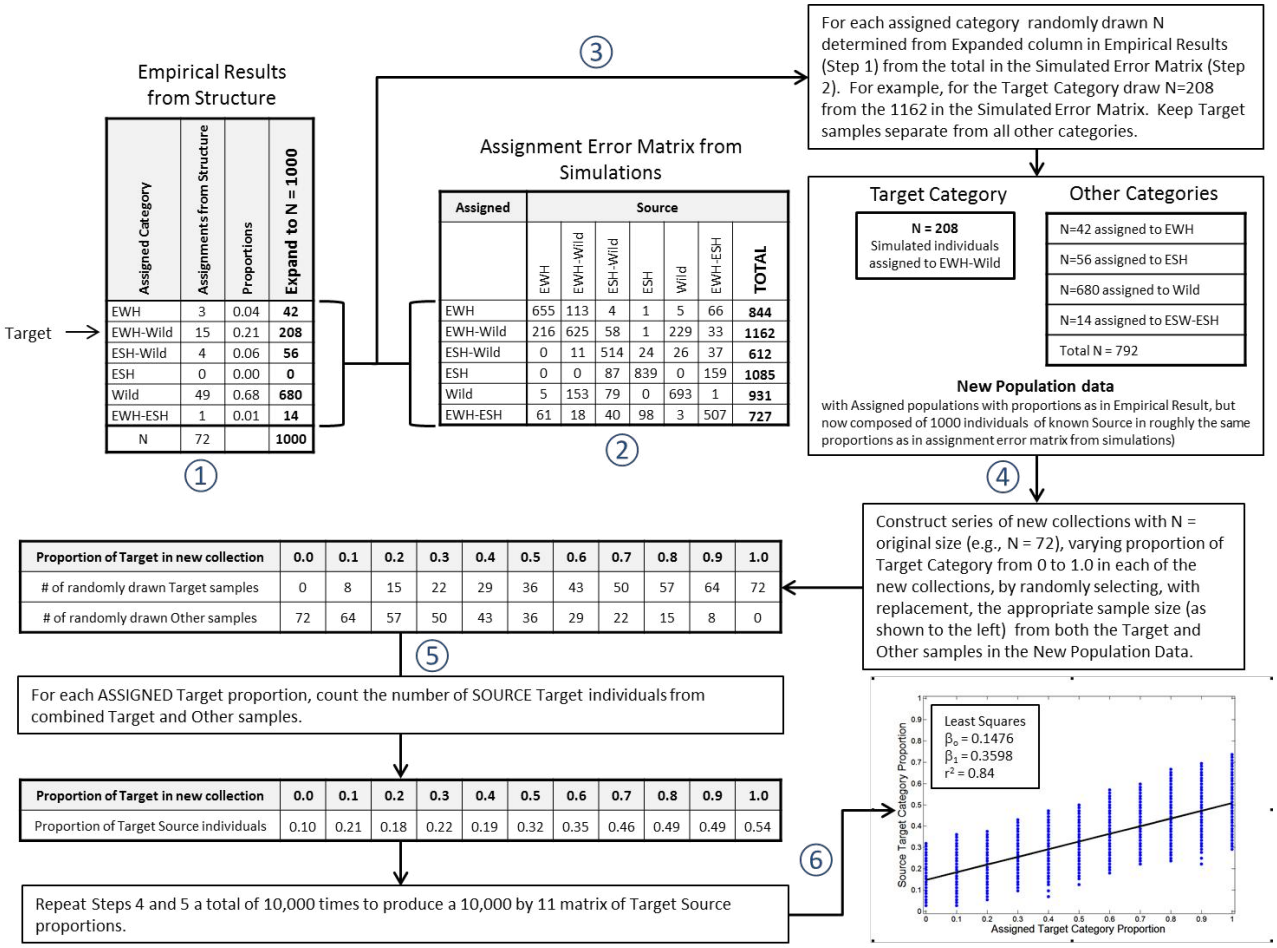


Figure S6. Likelihood-based procedure to adjust *Structure* results to account for common ancestry between the hatchery populations and wild populations. Numbered circles are procedure steps explained more fully in the text.

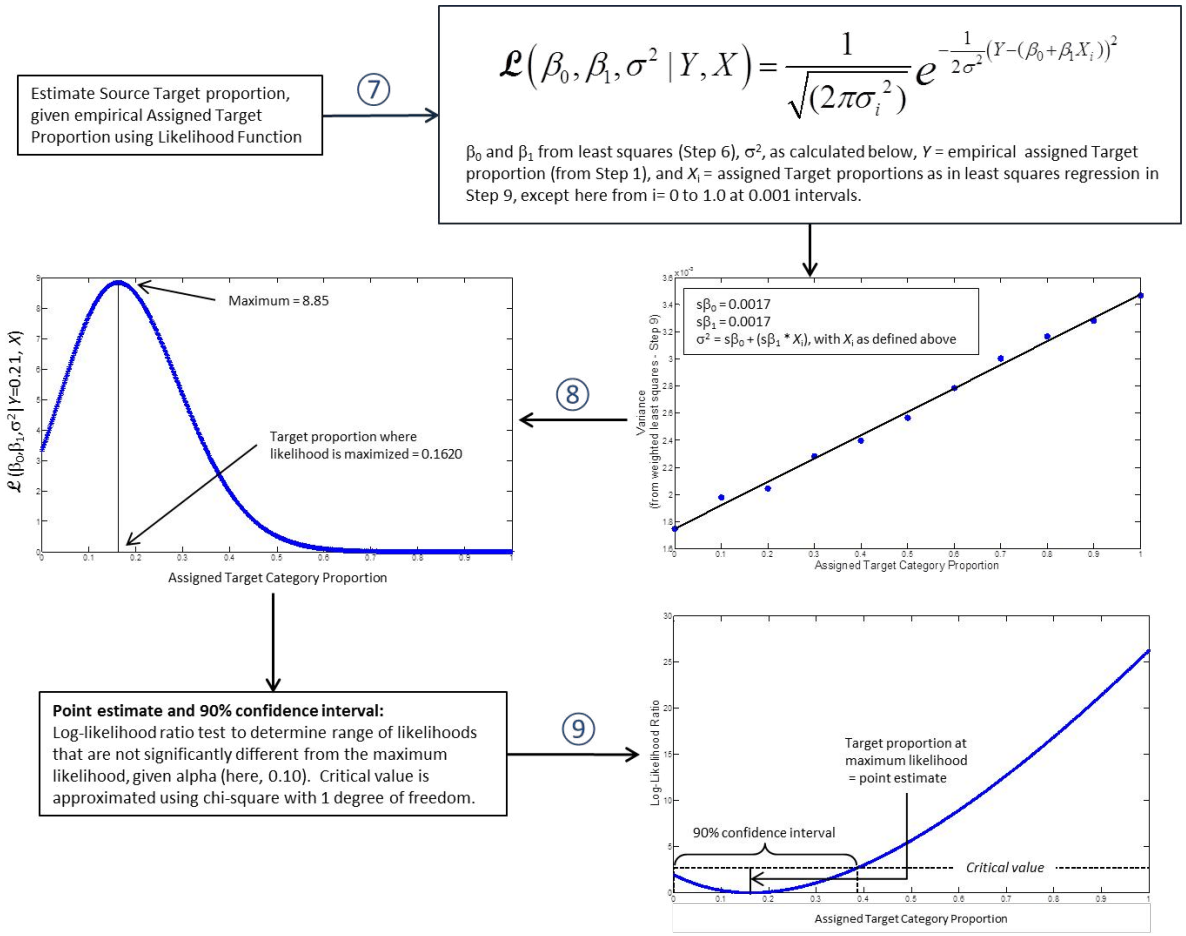


Figure S6. Continued