

ORIGINAL ARTICLE

Effectiveness of managed gene flow in reducing genetic divergence associated with captive breeding

Charles D. Waters,^{1,*} Jeffrey J. Hard,² Marine S. O. Briec,¹ David E. Fast,³ Kenneth I. Warheit,⁴ Robin S. Waples,² Curtis M. Knudsen,⁵ William J. Bosch³ and Kerry A. Naish^{1,*}

1 School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

2 Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA

3 Yakama Nation Fisheries, Toppenish, WA, USA

4 Washington Department of Fish and Wildlife, Olympia, WA, USA

5 Oncohr Consulting, Olympia, WA, USA

Keywords

captive breeding, domestication selection, genome-wide survey, managed gene flow, supportive breeding.

*Correspondence

Charles D. Waters, School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98195-5020, USA.

Tel.: +1 206 616 3148;

fax: +1 206 543 5728;

e-mail: cwaters8@uw.edu

Kerry A. Naish, School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98195-5020, USA.

Tel.: +1 206 221 6375;

fax: +1 206 543 5728;

e-mail: knaish@uw.edu

Received: 2 July 2015

Accepted: 2 September 2015

doi:10.1111/eva.12331

Introduction

The genetic risks associated with captive breeding are widely recognized. Rearing conditions may result in the relaxation of natural selective pressures and impose artificial selection favoring individuals suited for captivity instead of the natural environment (Frankham 2008). Additionally, genetic drift and inbreeding may reduce genetic diversity within captive populations (Fraser 2008; Naish et al. 2008). Thus, captive and wild populations may become genetically divergent over time, and individuals raised in captivity for supplementation purposes may be less effective in contributing to the population's long-term

Abstract

Captive breeding has the potential to rebuild depressed populations. However, associated genetic changes may decrease restoration success and negatively affect the adaptive potential of the entire population. Thus, approaches that minimize genetic risks should be tested in a comparative framework over multiple generations. Genetic diversity in two captive-reared lines of a species of conservation interest, Chinook salmon (*Oncorhynchus tshawytscha*), was surveyed across three generations using genome-wide approaches. Genetic divergence from the source population was minimal in an integrated line, which implemented managed gene flow by using only naturally-born adults as captive broodstock, but significant in a segregated line, which bred only captive-origin individuals. Estimates of effective number of breeders revealed that the rapid divergence observed in the latter was largely attributable to genetic drift. Three independent tests for signatures of adaptive divergence also identified temporal change within the segregated line, possibly indicating domestication selection. The results empirically demonstrate that using managed gene flow for propagating a captive-reared population reduces genetic divergence over the short term compared to one that relies solely on captive-origin parents. These findings complement existing studies of captive breeding, which typically focus on a single management strategy and examine the fitness of one or two generations.

viability than their wild-born counterparts (Fraser 2008; Jule et al. 2008; Laikre et al. 2010). Yet, captive breeding may be the only feasible tool for species recovery and conservation, particularly as rates of extinction increase (Pimm et al. 2006; Burkhead 2012). Therefore, there is a pressing need to develop and test approaches that minimize these risks over sufficient generations for population recovery to occur (Williams and Hoffman 2009). Theoretical treatments (Lynch and O'Hely 2001; Duchesne and Bernatchez 2002; Ford 2002) suggest that intentional promotion of gene flow from wild to captive populations may mitigate genetic divergence caused by genetic drift, inbreeding, and selection during captive breeding (Moberg et al. 2005;

Frankham 2008; Paquet et al. 2011). This strategy, which we refer to here as managed gene flow, should be investigated in empirical settings to determine whether the approach reduces genetic changes that could adversely affect fitness and limit restoration success.

Supportive breeding is a form of captive breeding intended to enhance the sizes of populations in their natural environment (Ryman and Laikre 1991). In such programs, a fraction of a population is taken into captivity for reproduction, and their progeny are released back into the natural environment to join wild-born conspecifics. This method avoids captive rearing for the full life cycle and might aid *in situ* population recovery if captive offspring have high reintroduction success. Supportive breeding using hatcheries has become an important component of many recovery plans for anadromous Pacific salmon on the West Coast of North America, where wild populations have steadily declined due to anthropogenic disturbances (National Research Council 1996). Historically, many hatcheries used returning hatchery fish for broodstock, but potential divergence of these fish from the wild population might result in decreased fitness following interbreeding in the natural environment (Busack and Currens 1995; Camp-ton 1995). Recent management reforms aimed at mitigating negative effects of captive rearing have led to the widespread use of locally-derived broodstock and implementation of managed gene flow between the hatchery and naturally-derived components (Mobrand et al. 2005; Paquet et al. 2011).

There are several concerted research efforts aimed at understanding the impacts of supportive breeding on wild populations and the effectiveness of recent hatchery reforms. Pedigree-based studies have provided evidence of reduced reproductive success of hatchery-origin fish spawning in the natural environment relative to that of wild fish in the same cohort (Araki et al. 2007; Milot et al. 2013; Christie et al. 2014). However, the significance of reduced fitness of first-generation hatchery fish to the future viability of supported populations is unclear. The captive environment typically facilitates the reproduction of most individuals regardless of their fitness in the natural environment (Neff et al. 2011), but subsequent selection on F_1 offspring released in the wild might remove maladaptive individuals (Baskett and Waples 2013) and reduce their effects on population fitness. Alternatively, ongoing broodstock collection and reduced fitness of F_1 hatchery offspring might have long-term negative demographic and genetic effects on the wild component (Ryman and Laikre 1991; Ford 2002; Baskett and Waples 2013). It is also challenging to measure the relative fitness of the offspring of hatchery fish that spawned in the wild (the F_2 generation, born in the wild) because matings that produce no offspring are typically unobserved (Araki et al. 2009), and subsequent

analyses can have reduced power to detect an effect (Christie et al. 2014). Finally, many pedigree studies are necessarily opportunistic because they often rely on established systems, and specific outcomes might depend on local management objectives. It is therefore important to develop complementary experimental methods that examine the consequences of captive rearing on genetic diversity in hatchery-produced fishes, and investigate the potential for managed gene flow to reduce their genetic divergence from wild fish.

A powerful way to evaluate the use of managed gene flow in supportive breeding would be to study a system that is comparative in nature, namely one that maintains a captive line that has been deliberately kept separate from wild individuals, and a second line where wild individuals are used exclusively as captive broodstock. Variation in each generation of the captive lines could be compared to the original founding population using genome-wide surveys, which offer a means to broadly estimate rates of genetic change. Such a comparison would provide insight into the range of possible outcomes of managed gene flow and refine understanding on the optimal longevity of supportive breeding programs.

A hatchery program at the Cle Elum Supplementation and Research Facility (CESRF) was initiated in 1997 in response to declining anadromous spring Chinook salmon returning to the Yakima River, a tributary of the Columbia River, USA (Fig. 1). Wild adults were collected for founding broodstock from the upper Yakima River population from 1997 to 2002. Beginning with brood year 2002, the hatchery population was divided into a segregated (SEG) line, which was not allowed to interbreed with the source population, and an integrated (INT) line, which was allowed to spawn in the river (Fig. 2). All first-generation hatchery fish from the integrated line were allowed to spawn naturally. The proportion of hatchery fish from the integrated line spawning in the natural environment has varied between 0.2 and 0.76 (mean = 0.56) from 2001 to 2013 (Fast et al. 2015). There is indirect evidence that these hatchery fish successfully contributed to the natural population: hatchery- and natural-origin fish had similar distributions on the spawning grounds (Dittman et al. 2010), and redd (nest) abundance and spatial distribution has increased (Fast et al. 2015). Fish from the integrated and segregated hatchery lines are raised in the same facility but are differentially marked for external identification. The two lines have been reared for three generations, and DNA samples have been collected from every adult fish used as broodstock since the inception of the program (fish are intercepted at Roza Dam, Fig. 1). Thus, this system provides an extraordinary opportunity to experimentally evaluate efforts to reduce genetic change in captive management.

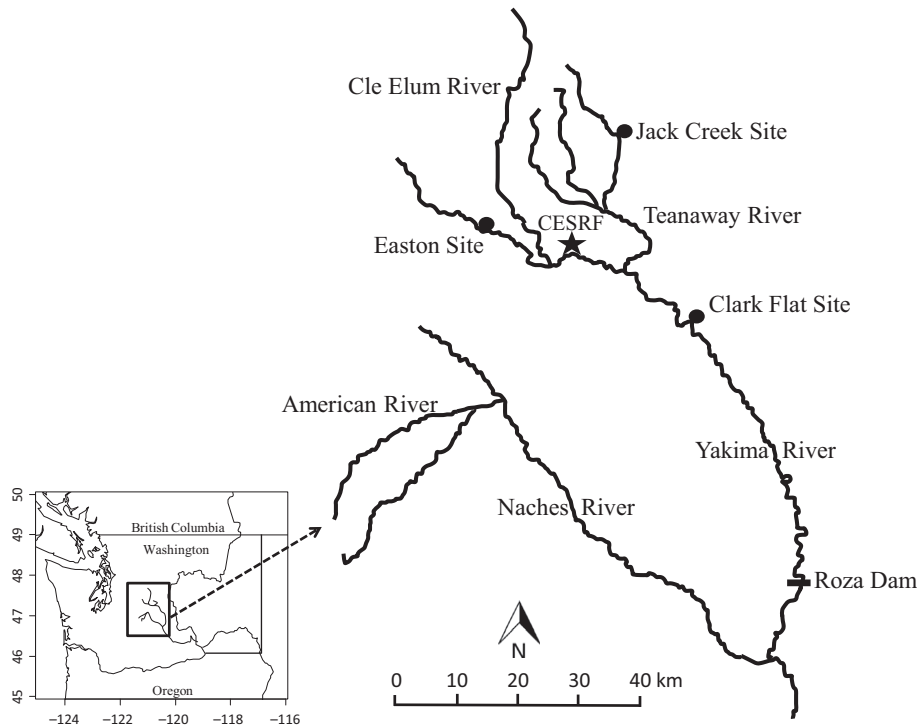


Figure 1 Map of the Yakima River system. The three spring Chinook salmon populations spawn in the American River, the Naches River, and the upper Yakima River above Roza Dam. The upper Yakima population is the target of the Cle Elum Supplementation and Research Facility (CESRF). All adults returning to the upper Yakima River are sampled at Roza Dam and allowed to spawn naturally (natural-origin and integrated line fish) or are removed from the system (all segregated line fish). Spawning and rearing for the hatchery lines occurs at CESRF. Prior to outmigration in spring, juveniles are transferred to the Easton, Jack Creek, and Clark Flat acclimation sites, where they are held for approximately two months before volitional release.

The aim of our study was to determine whether managed gene flow between natural and captive environments was effective at reducing genetic divergence relative to a wild founding population over three generations in a Chinook salmon supportive breeding program. We used a population genomic approach to survey genetic variation at 9410 polymorphic restriction site-associated DNA (RAD) markers, including 4405 markers that were anchored to a dense linkage map representing all 34 chromosomes. These markers were first used to test whether genetic divergence occurred in the integrated and segregated hatchery lines over each generation since founding, and to determine whether managed gene flow was effective at reducing differentiation. We also investigated potential mechanisms underlying observed changes. Changes due to genetic drift were examined using estimates of effective number of breeders. Molecular signatures of adaptive divergence, possibly indicative of domestication selection, were detected using multiple outlier analyses. By combining these results, we empirically validated an approach aimed at minimizing potential deleterious genetic effects of supportive breeding, and identified possible causes that might lead to reduced fitness of captive-reared individuals.

Materials and methods

Study system

The Yakima River is a tributary of the Columbia River located in south-central and eastern Washington State, USA (Fig. 1). The basin is home to three genetically distinct populations of spring, stream-type Chinook salmon (*Oncorhynchus tshawytscha*) in the upper Yakima River, the Naches River, and the American River (Fig. 1; Busack and Marshall 1991). Adult Chinook salmon return to the basin in spring, spawn in fall, and their offspring spend an entire year in freshwater before migrating to the ocean.

Annual returns of anadromous wild spring Chinook salmon to the Yakima River averaged approximately 1600 individuals in the 1980s and 1990s (Fig. 3). These population sizes represent a decline of 98% from estimates of historical returns (ref. in Lichatowich and Mobrand 1995). The hatchery program at CESRF was explicitly designed to test whether supportive breeding could increase harvest and production in the upper Yakima River Chinook population while minimizing ecological and genetic risks associated with captive rearing (RASP (Regional Assessment of Supplementation Planning) 1992).

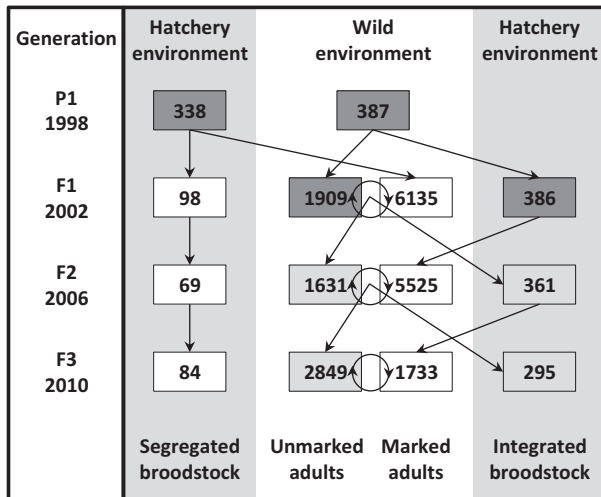


Figure 2 Schematic illustrating the initiation (the founding P_1 generation) and subsequent broodstock management (F_1 – F_3 generations) for the integrated and segregated hatchery lines of anadromous Chinook salmon that were surveyed. Numbers given in each box are the numbers of spawners (wild environment) and the number of broodstock (hatchery environment) for each brood year surveyed. Linear arrows denote the contribution of wild spawners or hatchery broodstock to the subsequent generation. Circular arrows represent unobserved mating between wild-born (unmarked) and hatchery-born (marked) spawners in the wild environment. Dark gray boxes represent wild adults, light gray boxes represent natural-origin adults with hatchery, wild, or hybrid ancestry, and white boxes represent adults born in the hatchery. Only brood years sampled are illustrated, but the same design is implemented each year. Chinook salmon are semelparous but have overlapping generations – approximately 80–90% of adults at CESRF are four-year-olds.

Returning wild adults from the upper Yakima population were collected for broodstock from 1997 to 2002 as they passed the Roza Dam Adult Monitoring Facility (Roza Dam, Fig. 1). Adults were spawned at CESRF, and eggs and fry were reared at the facility for approximately 16 months. Juveniles were then transferred to three acclimation sites, which were designed to limit returns to the hatchery facility, expand the spatial influence of supplementation efforts, and permit related research on homing (Fig. 1). Following an acclimation period of two months, fish were allowed to volitionally begin their migration to the ocean. Approximately 80–90% of upper Yakima River Chinook spend 2 years in the ocean and return to freshwater at age 4 to reproduce (Knudsen et al. 2006).

First-generation hatchery fish began returning as adults in large numbers in 2001 and were allowed to spawn naturally. In 2002, CESRF spawned both wild and returning hatchery-origin adults to create the segregated (SEG) and integrated (INT) hatchery lines (Fig. 2). The segregated line relied solely on returning hatchery-origin adults for broodstock, and no fish were allowed to spawn naturally. In contrast,

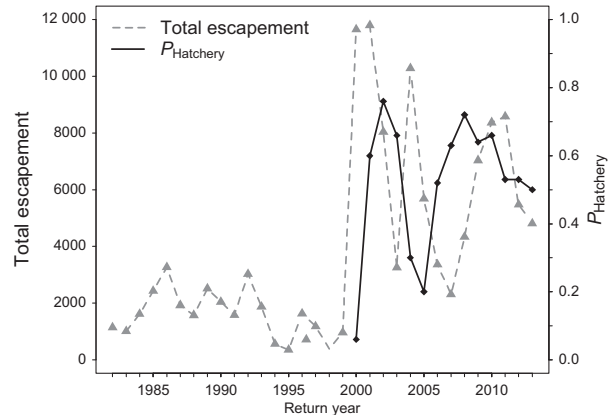


Figure 3 Total annual escapement of adult Chinook salmon in the upper Yakima river (gray, dashed line) and the proportion of spawners that are of hatchery origin (black line).

broodstock for the integrated line comprised only natural-origin fish, and all returning adults were free to spawn in the river. Natural-origin fish are those from the upper Yakima River population; however, beginning in 2005, returning adults were no longer 100% ‘wild’ due to possible influence from naturally spawning hatchery fish (Fig. 2).

Sample collection

Tissue samples for DNA (operculum or axillary process) were collected from all fish during spawning at CESRF by facility staff or the Washington Department of Fish and Wildlife and stored in 100% ethanol in Olympia, WA. We subsampled tissues ($n = 681$) from the 1998 wild founders (2nd founding year; P_1 Founders) and hatchery brood years 2002 (F_1 Wild and F_1 Hatchery), 2006 (F_2 INT and F_2 SEG), and 2010 (F_3 INT and F_3 SEG; Fig. 2). As most individuals mature at age 4, these brood years represent four generations of adults. We considered each generation of each hatchery line as separate ‘populations’ for the purpose of this study.

DNA sequencing and genotyping

DNA was extracted using DNeasy Blood & Tissue kits (Qiagen, Valencia, CA, USA) following the animal tissue protocol. Restriction site-associated (RAD) libraries (Baird et al. 2008) were prepared, with 24–36 individuals per lane, using the restriction enzyme *Sbf1* and sequenced using the Illumina HiSeq 2000 platform.

RAD sequences were processed with *Stacks* (v. 1.09, Catchen et al. 2013). Reads were demultiplexed and trimmed to 74 base pairs in length because sequencing errors increased after that length. Samples that showed signs of contamination were removed from the analysis. All

reads were then aligned to a reference database for Chinook salmon comprising 48 528 nonduplicated RAD loci, including 7146 mapped loci (Briec et al. 2014), using the 'best' option in *Bowtie* (v. 0.12.8, Langmead et al. 2009) and allowing up to three nucleotide mismatches. If a read aligned to multiple loci in the baseline, then those loci were omitted from downstream analyses. Following the *Bowtie* alignment, monomorphic and polymorphic loci were identified in *Stacks* using the bounded-error SNP calling model with default error rates and a minimum stack depth of 10 reads. Loci were then filtered for those with two alleles. In an effort to correct for bias in genotype calls due to differences in read depth between two alleles at a locus, all bi-allelic loci were re-genotyped for each individual with a custom Python script. This script designates a genotype as heterozygous if both alleles had a minimum depth of two and a combined depth >10 reads (Briec et al. 2014). Next, loci were filtered and retained if they had a minor allele frequency ≥ 0.05 in at least one population. These criteria were used to reduce error associated with genotyping duplicated loci that have been retained in the salmonid genome following a whole duplication event, and to provide stringency for the detection of signatures of selection. Individual samples were removed if they had $\geq 50\%$ missing genotypes across all filtered loci.

Heterozygosity and population differentiation

Observed (H_o) and expected (H_e) heterozygosity for each population, and tests for Hardy–Weinberg equilibrium (HWE) at each locus, were computed in the R-package *adegenet* (v. 1.3-9; Jombart 2008). Tests for HWE were conducted using the Monte Carlo procedure and 1×10^5 permutations. The expected false discovery rate (q -value) for each locus was computed to correct for multiple testing using the R-package *qvalue* (v. 1.28.0; Storey 2002). A locus was deemed significantly out of HWE when the q -value was < 0.05 . Loci out of HWE, however, were retained for analyses because they may be of interest to present and future studies.

Two approaches were used to measure genetic change between each of the lines across generations. First, genetic differentiation, F_{ST} , was calculated between populations and at each locus using Weir and Cockerham's unbiased estimator in *Genepop* (v. 4.1, Weir and Cockerham 1984; Raymond and Rousset 1995). Significance of pairwise population F_{ST} comparisons was determined from tests of genotypic differentiation performed in *Genepop* using default parameters.

Second, a discriminant analysis of principal components (DAPC) was used to visualize temporal changes in genetic relationships between the lines and to identify loci that contribute most to population separation. Prin-

cipal components analysis (PCA) is not optimal for the analysis of population structure because it maximizes total genetic variation but does not account for underlying genetic structure (Jombart et al. 2010). On the other hand, discriminant analysis (DA) is affected by correlations between variables (loci) and requires the number of variables to be less than the number of objects (individuals) (Borcard et al. 2011). These constraints are problematic for analysis of genomic data because some loci are physically linked, and the number of loci is typically much greater than the number of individuals. DAPC overcomes these drawbacks by combining PCA and DA. First, PCA is used to identify synthetic, uncorrelated variables (principal components or PCs) that maximize the amount of genetic variation explained in the data. PCs are then retained as variables for DA; the actual number of PCs retained can be chosen to ensure that the number of variables is less than the number of objects. DA is subsequently performed on the retained PCs to explore divergence between groups.

DAPC was conducted in the R-package *adegenet* using all individuals and loci. As the analysis requires a complete data set, missing values were replaced by the mean frequency of the corresponding allele, computed on the whole set of individuals. To avoid over-fitting the discriminant functions, the *optim.a.score* function was used to identify the optimal number of principal components to retain in the first step of the analysis based on the difference between observed and random discrimination.

Loci that contribute most to group separation, which we refer to as discriminatory loci, were identified using the *snpzip* function. The contributions of loci to the DAPC, or loadings, were first used to compute a distance matrix between loci. A hierarchical clustering analysis using the median clustering method was then performed on the distance matrix to separate loci into those that contribute to group divergence and those that do not. Discriminatory loci were compared to explicit tests for molecular signatures of selection to provide support for loci consistent with adaptive divergence.

Effective number of breeders

Effective number of breeders, N_b , was compared between hatchery lines as a proxy metric for estimating the effects of genetic drift in each population. N_E Estimator (v. 2.01, Do et al. 2014) was used to obtain estimates of N_b by the linkage disequilibrium (LD) and temporal methods. To reduce potential bias due to selection, loci that were identified as outliers by F_{TEMP} and *Bayescan*, as well as DAPC discriminatory loci, were omitted. For each year sampled, we used only four-year-old adults, which represented a single cohort of individuals and thus provided the most appropriate measure of N_b .

For the LD estimates, we assumed random mating and used an allele frequency restriction of 0.05. Bias due to physical linkage was removed by excluding calculations between pairs of loci on the same chromosome (Larson et al. 2014). Estimates of N_b and 95% parametric confidence intervals were subsequently calculated for all populations (Larson et al. 2014). However, estimates of N_b obtained by the LD method may be biased by overlapping generations and fluctuating population size (Waples et al. 2014). We accordingly adjusted our estimates of N_b , as the study population was subject to both sources of potential bias (Fig. 3), based on principles in Waples et al. (2014). Briefly, annual census data were first grouped into generations comprising four years each, as 80–90% of this population matures at age four (Knudsen et al. 2006). For each generation, we calculated the harmonic mean of total number of spawners per year (N_{census}), as low return years within a generation have a disproportionate effect on N_b , and multiplied the mean by four to obtain total number of spawners per generation (N_{gen}). We then calculated the weighted harmonic mean of total spawners for the previous three generations (Wt. N_{gen}), as past demographic history influences LD estimates of N_b (Waples et al. 2014). Previous generations, starting from the most recent, received weights of 1/2, 1/4, and 1/8 because half of existing LD decays each generation (Waples et al. 2014). We then divided Wt. N_{gen} by N_{census} for the year of interest. This ratio approximates the ratio of N_e/N_b , assuming N_e/N_{gen} is proportional to N_b/N_{census} . As N_b estimates are a function of the harmonic mean of N_e and true N_b (Waples et al. 2014), this ratio was then used to calculate bias and adjust our N_b estimates. Full calculations are shown in Table S6.

For the temporal method, N_b and 95% confidence intervals were calculated using the P_1 founders and three hatchery generations based on all non-outlier loci. We followed the Jorde and Ryman method, Plan II sampling, and used an allele frequency restriction of 0.05.

Detection of outlier loci consistent with adaptive divergence

Given that F_{ST} outlier tests are known to exhibit a high rate of false-positive results (Lotterhos and Whitlock 2014), three independent tests for genomic regions indicative of diversifying selection were conducted to explore the potential role of selection in divergence of the two lines from the founder population. In our final interpretation, emphasis was placed on outlier regions that were identified by multiple tests, and were significantly divergent across multiple generations.

F_{TEMP} method

F_{TEMP} (Therkildsen et al. 2013) is designed to detect selection in a single population sampled across multiple generations by simulating genetic drift over time. The method is a

modification of the widely used F_{dist} approach (Beaumont and Nichols 1996), and is based on an island model with drift but no migration. We parameterized simulations of genetic drift using estimates of N_b derived from the study populations, and explicitly identified loci that exceeded neutral expectations (i.e., consistent with signatures of adaptive divergence) in each hatchery line. Drift was simulated at 1×10^6 loci for four generations. The model was first run using N_b estimates obtained by the temporal method using all loci (N_b values of 459 and 61 for the integrated and segregated hatchery lines, respectively). Yet, N_b estimates can themselves be biased by loci under selection. Therefore, F_{TEMP} outlier loci identified in the first run, as well as *Bayescan* outlier and DAPC discriminatory loci, were removed, and second estimates of N_b were calculated using only non-outlier loci. Estimates of N_b using non-outlier loci were 466 and 69 for the integrated and segregated hatchery lines, respectively, and final F_{TEMP} simulations incorporated these N_b estimates. However, to test the sensitivity of our results to N_b , we also ran F_{TEMP} simulations using the following estimates: (i) N_b estimates obtained from individuals of all ages (N_b of 388 and 59 for INT and SEG lines, respectively) instead of those calculated using only 4-year-olds; (ii) estimates of N_e instead of N_b (approx. $4*N_b$, Waples 2002), and (iii) N_b estimates for the opposite hatchery line (i.e., N_b of 466 for the SEG line and 69 for the INT line).

Sample sizes (n) per generation were also parameterized in the F_{TEMP} simulations, where sample sizes equaled the harmonic means of the number of individuals sequenced in each generation (61 and 57 for the integrated and segregated lines), but randomly varied between $0.8*n$ and n to account for variation in the number of individuals genotyped per locus per generation. One thousand final F_{TEMP} simulations were performed to account for variability between runs. To correct for multiple testing and control the false discovery rate (Storey 2002), the q -value of each locus was computed from its corresponding P -value using the R-package *qvalue*. A locus was considered significant if its q -value was <0.05 in at least 95% of the final simulations. Significant loci showed greater temporal allelic variation than is expected under genetic drift alone.

Bayescan

We also employed a Bayesian approach to identify outlier loci using the program *Bayescan* (v. 2.1; Foll and Gaggiotti 2008). However, this approach assumes that the sampled populations evolved independently from a common ancestral population; thus, it is not ideally suited for temporal data. In addition, the power of *Bayescan* to detect outlier loci is reduced with few populations (<6), low sample sizes (<30), and low neutral F_{ST} values (<0.01) (Foll and Gag-

giotti 2008). Despite these limitations, we used *Bayescan* because it has lower type I and type II errors than other available methods (De Mita et al. 2013). The program was run with all populations combined with default parameters. A locus was considered to be an outlier if its q -value was <0.05 in each of five independent runs.

Test for outlier regions of the genome: sliding window analysis

Moving averages of pairwise F_{ST} values between each hatchery line and the P_1 founders at mapped markers were calculated using a kernel smoothing sliding window approach (Hohenlohe et al. 2010; Briec et al. 2015). Each chromosome was divided into 100 windows spanning 18 cM each; these parameters were optimized to detect small regions of divergence but minimize background noise due to sampling variance. Contributions of individual loci to the F_{ST} average were weighted by their distance from the center of the window. A null distribution with 95% confidence intervals was calculated for each window using 1×10^6 replicates of randomly sampling F_{ST} values from all mapped loci, with replacement, where sample size was based on the number of loci within each window. The sliding window analysis was performed separately on the F_1 , F_2 , and F_3 generations of each hatchery line. Regions of significantly elevated divergence were identified as those where the moving average exceeded the 95% confidence interval of the null distribution.

Outlier alignment and gene function

All loci located within the genomic regions that exhibited overlap among the three tests of adaptive divergence (summarized in Table 1) were aligned to the rainbow trout genome (Berthelot et al. 2014) using *Bowtie* (Langmead et al. 2009) to identify genes associated with divergence and potential targets of domestication selection. *BLAST2GO* (Conesa et al. 2005; Götz et al. 2008) was then used to conduct a *BLAST* (Altschul et al. 1990) search on the NCBI nr public database for the rainbow trout gene coding

sequences that were identified. We used an e -value threshold of 1×10^{-10} in *BLAST* searches. Gene ontology (GO) terms, GO Slim terms, and functions associated with each gene were subsequently identified.

Results

DNA sequencing and genotyping

We identified 9410 bi-allelic RAD loci, including 4405 loci that aligned to the Chinook salmon linkage map and had a minor allele frequency >0.05 in at least one population (Table S1). A total of 413 individuals were genotyped at $>50\%$ of these loci and retained for analysis (Table S1). Sample sizes for the P_1 founders and each generation of integrated and segregated hatchery lines are summarized in Table S2.

Heterozygosity and population differentiation

Observed (H_o) and expected (H_e) heterozygosity were similar between the integrated and segregated hatchery lines within each generation, although a small decrease was observed over time (Table S2). All populations had $<3\%$ of loci that deviated from Hardy–Weinberg equilibrium. Loci that deviated from HWE were included in further analyses because they may be of interest to this study.

Population genetic differentiation, F_{ST} , for each of the generations compared to the P_1 founders was low in all pairwise comparisons, ranging from 0 to 0.0108 (Table S3). The F_1 hatchery fish and segregated line steadily diverged from the P_1 founders over time; the F_2 and F_3 SEG populations were significantly differentiated from the founders ($F_{ST} = 0.0049$ and 0.0108 respectively, $P < 0.001$). In contrast, the integrated line did not exhibit the same temporal trend of increasing genetic divergence. However, F_{ST} compared to the P_1 founders was significant for the F_2 and F_3 INT populations ($F_{ST} = 0.0026$ and 0.0022 , respectively, $P < 0.001$). F_{ST} between the F_3 SEG and F_3 INT populations was also significant ($F_{ST} = 0.0093$, $P < 0.001$), indicating divergence in the two hatchery lines.

Table 1. Genomic regions that exhibited overlap among all three tests of adaptive divergence. The table indicates the number of generations (Gens) that a portion or all of each region was identified as a region of high divergence by the sliding window (SW) analyses for the integrated (INT) and segregated (SEG) lines, and the number of outlier loci within each region that were identified by F_{TEMP} or *Bayescan*.

Region	Chromosome	Map position (cM)	Gens SW INT	Gens SW SEG	$F_{TEMPINT}$	$F_{TEMPSEG}$	Bayescan
1	Ots04	37.52–43.15		2		2	2
2	Ots05	87.43–95.05	1	3		1	1
3	Ots06	94.64–100.65	1	1		1	1
4	Ots11	57.77–71.46	2	3	1	2	1
5	Ots12	26.14–35.29	1	2		2	2
6	Ots15	132.33–139.26	1	1		1	1
7	Ots20	93.65–99.49	1	2		2	1

The first step of the discriminant analysis of principal components (DAPC) identified 58 PCs to retain, which explained 29.3% of the observed genetic variation among individuals. A plot of the kernel density estimates of individuals along the first discriminant function, which explained 63.7% of the retained variation, showed separation of the F_1 hatchery fish and segregated line from the P_1 founders (Fig. 4). The segregated line became more divergent over time; in comparison, the F_1 wild group and integrated hatchery line clustered closely with the P_1 founders. The second discriminant function explained 13.6% of the retained variation and, to a minor extent, represented temporal variability within the two lines. We identified 25 loci that contributed most to separation along the first discriminant function based on hierarchical clustering analysis (termed discriminatory loci; Table S8). A majority of the discriminatory loci (17 of 25) did not overlap with outlier loci identified by explicit tests of selection and thus were inferred to be selectively neutral.

Estimates of effective number of breeders

Estimates of the effective number of breeders, N_b , obtained by the LD method for each sample year largely reflected the effective number of parents that produced the samples (Waples 2005), but also had contributions from previous generations (Waples et al. 2014). Bias-adjusted estimates of N_b from the natural population revealed a 3–4 fold increase following establishment of the integrated hatchery line (Fig. 5; Table S5a), although this period also experienced

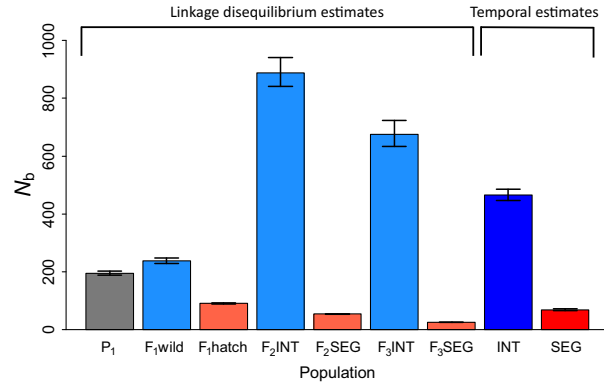


Figure 5 Effective number of breeders, N_b , and 95% confidence intervals estimated by the linkage disequilibrium (LD) and temporal methods for the P_1 Founders (gray) and integrated (blue) and segregated (red) hatchery lines. The LD method allows estimation of N_b for every generation, while the temporal method yields a single estimate for the entire sampling period. LD estimates are adjusted for physical linkage and potential bias caused by overlapping generations and fluctuating population size.

higher adult numbers returning to the basin (Fig. 3). Estimates of N_b in the segregated line declined steadily over time and, in the latter two generations, were over an order of magnitude lower than in the integrated line (Fig. 5; Table S5a), suggesting that genetic drift may have driven the observed genetic divergence. Differences between the two lines were not unexpected, as the average broodstock sizes were 363 (SD = 56) and 85 (SD = 15) for the integrated and segregated lines, respectively. Interestingly, wild

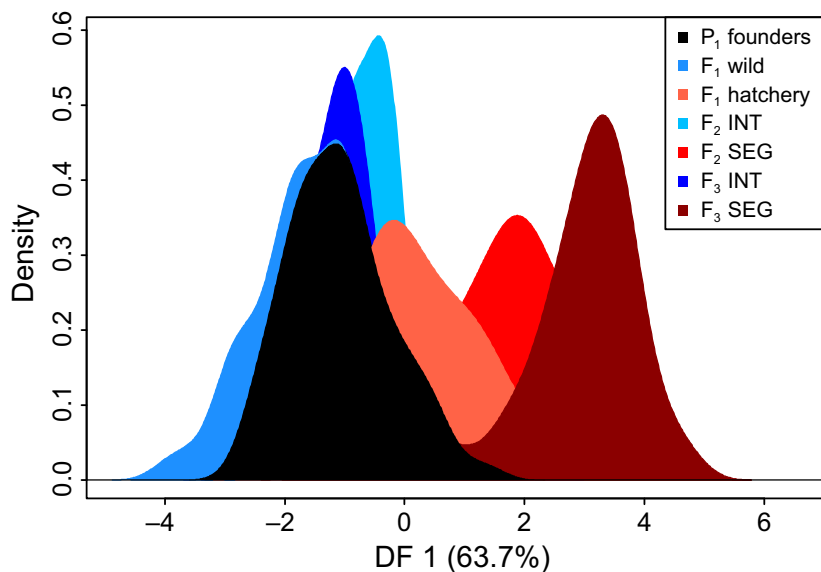


Figure 4 Density plot of individuals from the wild founders (P_1 Founders, black) and three generations of the integrated (INT, blue colors) and segregated (SEG, red colors) hatchery lines along the first discriminant function from the discriminant analysis of principal components (DAPC).

fish that were spawned in the hatchery in 1998 had a lower N_b estimate than the wild fish that reproduced in the river, despite having similar census sizes and the same population history (Table S5a).

The temporal method provided single estimates of N_b for the integrated and segregated lines across the sampled generations (Fig. 5; Table S5b). Temporal estimates closely agreed with the harmonic means of the LD estimates and showed that the integrated line had a larger N_b .

Detection of outlier loci and chromosomal regions of high divergence

F_{TEMP} identified 228 temporal outlier loci in the segregated line and 80 in the integrated line (Table S8). Thirty-two loci were identified as outliers in both lines. F_{TEMP} results did not change considerably when run with the N_b estimates obtained from individuals of all ages. Running F_{TEMP} with estimates of N_e instead of N_b did not alter results for the integrated line, but it significantly increased the number of temporal outliers in the segregated line. In addition, there were 369 F_{TEMP} outliers in the segregated line but just 48 outliers in the integrated line when the N_b estimates for one hatchery line were used for the opposite line. These results indicated that the numbers of outlier loci detected in the segregated line were consistently higher than in the integrated line across a range of N_b values, and that using the single cohort N_b estimates instead of N_e was a conservative choice.

The second analysis, conducted using *Bayescan*, identified 75 outlier loci across all combined populations (Table S8). As the *Bayescan* analysis required combining populations, outlier loci could not be attributed to a specific hatchery line. However, there was considerable overlap with F_{TEMP} results. Twenty-eight *Bayescan* outlier loci were also F_{TEMP} outliers in both hatchery lines, while an additional 17 and 22 *Bayescan* loci were F_{TEMP} outliers in only the integrated or segregated lines, respectively.

Lastly, sliding window analyses conducted on the 4405 mapped loci identified several regions of significantly elevated F_{ST} values in both hatchery lines (Table S9). However, regions of high divergence in the segregated line were more temporally stable than those in the integrated line; that is, more regions were consistently identified across the F_1 , F_2 , and F_3 generations. Five genomic regions had significantly elevated F_{ST} values in two of the three generations within the segregated line, and four regions were significantly high in every generation. In contrast, the integrated line contained three regions that were significantly elevated in two generations, and no region was divergent across all generations. Furthermore, plots of per locus F_{ST} values for markers positioned along the chromosome revealed that

divergence from the P_1 founders increased over time at some regions in the segregated line, and these regions also contained loci that were identified as outliers by other tests (Fig. 6). This directional, temporal trend was not observed in the integrated line, although a region on Ots11 was significantly divergent in both hatchery lines in the F_1 and F_3 generations.

Overall, seven genomic regions exhibited overlap among all three tests of diversifying selection in the segregated line (Table 1). One region showed consensus across tests in the integrated line, but this was also shared by the segregated line. These results, in conjunction with the temporal increase in F_{ST} at some regions, suggest that selection might also have contributed to divergence in the segregated line.

Outlier alignment and gene function

Forty-three mapped loci were located within the seven genomic regions that exhibited overlap among the three tests of adaptive divergence, and 25 of these loci aligned to the rainbow trout (RBT) genome. Twenty-two loci, including five outliers, were located within or near annotated RBT genes. While the RBT genes had an array of functions, commonalities among genes were observed. Three genes, including one associated with an outlier locus, were related to the protein ubiquitin, which is important for protein degradation and immune response. Two other genes were also related to the function and response of the immune system. Three genes, with two linked to outlier loci, were involved in the utilization, breakdown, and reception of lipids, while another two genes, including one near an outlier locus, were essential for the production of ribose, a sugar that can be used for energy or the synthesis of nucleotides. Three genes were linked to neurotransmission, including one gene that coded for receptors of gamma-aminobutyric acid (GABA), a major inhibitory transmitter in the central nervous system, and a second gene that has been shown to induce clustering of GABA receptors. Lastly, two of the RBT genes, including one from an outlier locus, were linked to the development of photoreceptors and eye pigmentation. The shared functions of many genes provided insight into the potential targets of domestication selection. Alignment results, as well as gene functions with references, are summarized in Table S10.

Discussion

The aim of this study was to evaluate the effectiveness of managed gene flow to reduce genetic divergence in supportive breeding programs. This goal was achieved by comparing genome-wide diversity in three generations of

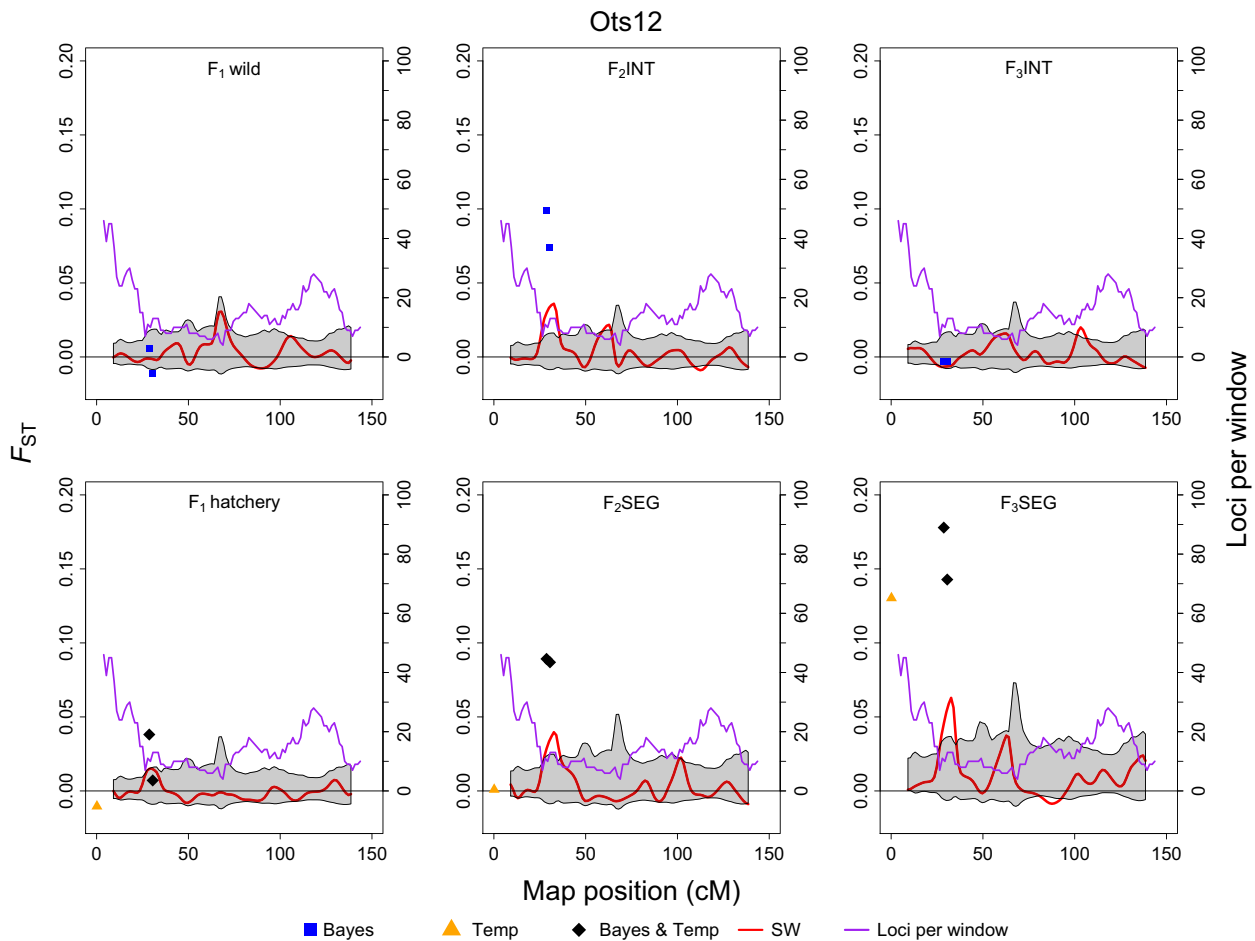


Figure 6 Loci and regions of the genome showing signatures of adaptive divergence, based on pairwise F_{ST} compared to the P_1 founders, on chromosome Ots12 for the integrated (top panel) and segregated (bottom panel) hatchery lines through the F_1 , F_2 , and F_3 generations. Blue squares are loci that were identified as outliers with *Bayescan*, orange triangles are outliers identified by F_{TEMP} , and black diamonds are loci identified by both *Bayescan* and F_{TEMP} . The red line represents the kernel smoothed moving average of F_{ST} and the gray-shaded area is the 95% confidence interval. The purple line shows the number of loci within each sliding window of the moving average.

integrated and segregated Chinook salmon hatchery lines to their founding wild population. Genetic distance measures at 9410 loci showed that divergence in the segregated line was significant by the second generation. Much of this change can be ascribed to genetic drift, as suggested by the small numbers of effective breeders. However, we also found evidence for a temporal trend in divergence at specific genomic regions, consistent with domestication selection. In contrast, genetic divergence in the corresponding integrated line was marginal over three generations, suggesting that the use of natural-origin broodstock was effective at reducing genetic change in the short term. By comparing contrasting regimes, the results illustrate the range of possible outcomes that may occur when using deliberate gene flow to mitigate genetic impacts in captive breeding programs. This information is therefore highly

relevant when weighing the relative risks of different captive rearing strategies.

Like many such studies on supportive breeding programs (Araki et al. 2007; Hess et al. 2012; Anderson et al. 2013; Christie et al. 2014), this study does not have a contemporary control population to which the hatchery lines can be compared. After the first hatchery generation, the upper Yakima population was no longer ‘wild’ due to influence from the integrated line. Therefore, ongoing evolutionary processes in the natural environment cannot be readily discriminated from those in the hatchery. However, the comparisons to the wild founding population strongly suggest that divergence in the segregated line was primarily due to repeated exposure to the hatchery environment, as differentiation did not occur in the integrated line.

It is also important to note that the contribution of integrated hatchery fish to natural population productivity has not been fully quantified. Higher adult returns in the natural population were observed following initiation of the hatchery (Fig. 3), but this increase might be attributed to supplementation, a favorable shift in environmental conditions (Mantua et al. 1997; Fast et al. 2015), or a combination of factors (Scheuerell et al. 2015). Previous studies suggest that hatchery-reared individuals tend to be less fit than naturally-born fish (Araki et al. 2007; Milot et al. 2013; Christie et al. 2014), but successfully spawning hatchery fish can still add to natural production. In this system, there are three lines of evidence that suggest that hatchery fish contributed to population productivity. First, the proportion of hatchery fish spawning in the wild reached 70–75% in some years (Figs 2 and 3) and, given these proportions, it is highly likely that these fish supported subsequent generations. Second, nest (redd) abundance and spatial distribution significantly increased following supplementation efforts, an increase that exceeded numbers observed in a nearby unsupplemented river (Fast et al. 2015). Third, comparisons of female reproductive traits (Knudsen et al. 2008) and breeding success experiments within an artificial stream (Schroder et al. 2008, 2010) revealed little difference between wild and first-generation hatchery salmon, although offspring of wild females had significantly higher survival to the fry stage (5.6%, $P = 0.04$; Schroder et al. 2008). A full understanding of the contribution of hatchery-reared fish to population productivity will be gained through complementary pedigree-based studies.

An ideal experiment would have compared lines with equal numbers of broodstock. For example, in theoretical treatments on the effects of different management strategies on genetic change in supportive breeding programs, Duchesne and Bernatchez (2002) found that increasing the census size of the captive population had the most influence on reducing the inbreeding coefficient compared to other tactics such as changing the proportion of wild individuals in the broodstock or modifying the numbers of captive fish released. Here, the numbers of adults spawned in the segregated line were typically a quarter of those in the integrated line, an outcome of the fact that the Cle Elum facility serves broader restoration goals beyond the experiment. This difference in broodstock size partly explains the divergence from the founder population observed in the segregated line, and it is likely that simply increasing the broodstock size would have reduced divergence. However, one of the major objectives of managed gene flow is to minimize genetic drift. This goal is partly achieved through taking advantage of both wild and hatchery production, effectively increasing the ‘rearing space’ of supportive breeding programs. Therefore, the effective size of integrated programs

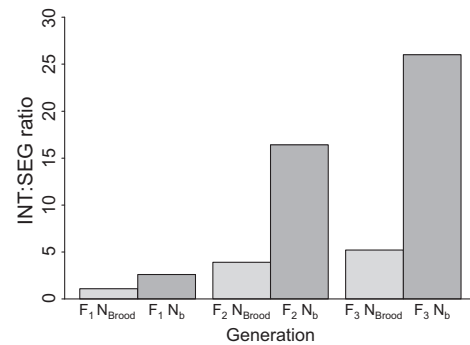


Figure 7 The integrated:segregated ratios of broodstock sizes (N_{brood} INT/ N_{brood} SEG) and effective number of breeders (N_b INT/ N_b SEG) for the F_1 , F_2 , and F_3 hatchery generations. The N_b ratios were consistently higher than the N_{brood} ratios, indicating that differences in N_b between the hatchery lines were primarily due to managed gene flow and not differences in broodstock sizes.

can be expected to exceed comparable segregated programs in cases where the hatchery or the wild component does not become a ‘sink’ for the ‘source’ population.

Evaluations of the broodstock census size and effective number of breeders across the integrated and segregated lines provide insight into the relative importance of gene flow in mitigating genetic divergence from the founder population. The ratio of broodstock census sizes (N_{brood} INT/ N_{brood} SEG) for the two lines can be compared to the ratio of N_b estimates (N_b INT/ N_b SEG) in each generation (Table S7). These ratios would be approximately equal if the differences in N_b estimates solely reflected differences in broodstock size. However, the ratios of N_b INT/ N_b SEG were larger than the ratios of N_{brood} INT/ N_{brood} SEG, particularly in the F_2 and F_3 generations (Table S7, Fig. 7). This simple comparison suggests that managed gene flow disproportionately increased N_b in the integrated line and reduced genetic divergence by incorporating the naturally spawning segment of the population.

There is an interesting aspect of this study that illustrates the temporal variation in the effects of hatchery rearing on N_b . The 2002 (F_1 period) estimates reflect the N_b of fish that spawned in 1998 (P_1 period). In the P_1 period, approximately equal numbers of wild fish spawned in the river and in the hatchery, yet the N_b of fish in the river was over two times greater than the N_b of fish in the hatchery. The result suggests that in this generation, the hatchery environment caused higher variance in reproductive success and decreased N_b , which is consistent with other studies of effective size in hatcheries (reviewed in Naish et al. 2008; Christie et al. 2012; Naish et al. 2013). In contrast, the ratio N_b/N decreased in the F_2 and F_3 generations of the integrated line despite higher N_b estimates relative to the segregated line, suggesting that ecological influences such as

higher competition on the spawning grounds (Fleming and Gross 1994) or limited carrying capacity might have influenced this ratio as the population increased in size. Overall, these results emphasize the dynamic impacts of both natural processes and hatchery rearing on effective size, the significance of multigenerational monitoring, and the need to understand the factors influencing the success of supportive breeding programs as a whole.

Multiple lines of evidence and consistent temporal trends in the segregated line provided compelling support for some genomic regions involved in domestication. In contrast, outlier regions detected in the integrated line had minimal overlap among tests and lacked temporal trends. Tests for F_{ST} outliers are known to exhibit high false-positive rates (Lotterhos and Whitlock 2014), but they continue to be used because they serve as useful starting points for further investigations (Storz 2005). Here, we reduced the likelihood of false-positive results using three independent tests for outliers (e.g., de Villemereuil et al. 2014), including one specifically designed to detect temporal divergence. Signatures of domestication selection are also difficult to detect, particularly in early generations (Lotterhos and Whitlock 2014; Mäkinen et al. 2015), and power is reduced with few populations in the analyses (Karlsson and Moen 2010). However, the results presented here are comparative and point toward greater and more rapid changes in the segregated line. More outliers were observed in the line with a smaller effective size, an interesting finding considering that selection is expected to be more efficient in larger populations (Wright 1931). Thus, the strength of selection must be greater in the hatchery than in the natural environment. This result has been observed in other small, fragmented populations (Koskinen et al. 2002). Finally, we note one outlier region that was divergent in the F_1 and F_3 generations of both hatchery lines. As this region was an outlier in the F_1 wild group, the results might indicate a false positive or the action of natural selection. However, the region was also an outlier in the F_1 hatchery and F_3 groups, possibly providing evidence of selection in the hatchery at early life stages that has affected both lines. We cannot rule out additional regions of undetected selection in both lines, as it is currently not possible to survey the whole genome in Chinook salmon. Despite these potential limitations, outlier regions identified in the segregated line are supported by multiple lines of evidence and provide valuable insight for the genetic basis of adaptation to captivity.

Alignment of loci within outlier regions to the rainbow trout genome and classification of gene functions identified possible targets of domestication selection. Multiple genes were related to the regulation and response of the immune system, which may be affected by the higher rearing densities in hatcheries. Many gene products, including those

associated with outlier loci (ubiquitin, alpha-beta hydrolase, and ribokinase), were connected to the processing of proteins, lipids, and sugars and may be affected by hatchery feeding regimes. For example, in rainbow trout, the ubiquitin-proteasome pathway, which facilitates protein degradation, was affected by starvation (Martin et al. 2002) and feeding status (Seiliez et al. 2008). Additionally, differential gene expression was observed between groups of rainbow trout that were fed different diets (Panserat et al. 2009). Two other gene products were important for the development of photoreceptors and eye pigmentation (salmon are visual feeders). Together, these functions suggest that domestication selection may be occurring in the hatchery environment due to differences in food availability and composition.

The two GABA-related genes, one that coded for GABA receptors and a second that induces clustering of GABA receptors, suggest another potential target of domestication selection. GABA has been shown to stimulate the secretion of two gonadotropins, GTH-1 and GTH-2, in rainbow trout (Mañanos et al. 1999), and these gonadotropins have been shown to stimulate gonad growth in juveniles (Suzuki et al. 1988). Domestication selection acting on this pathway may explain changes in rates of early male maturation commonly observed in hatchery populations of Chinook salmon (Larsen et al. 2010; Harstad et al. 2014). Furthermore, early male maturation is positively correlated with cumulative growth (size at release, Harstad et al. 2014). Domestication selection likely targets polygenic traits, and here, there is evidence that such selection may affect both the GABA-related genes and the genes that process proteins, lipids, and sugars.

The results reported here add to the growing number of studies that have, to date, typically focused on the relative performance of one or two generations of captive individuals released into the wild, and only under a single broodstock management regime. In salmon species, a few such studies documented minimal differences between hatchery and wild fish, including comparable reproductive success (Anderson et al. 2013), breeding success in an artificial channel (Schroder et al. 2008, 2010), reproductive traits (Knudsen et al. 2008), and spawning distributions (Dittman et al. 2010). However, many studies have found significant differences in hatchery fish, including reduced response or increased vulnerability to predation (Fritts et al. 2007), lower survival (McGinnity et al. 2003), differences in growth rate and morphology (McGinnity et al. 2003; Busack et al. 2007), and reduced reproductive success (Araki et al. 2007, 2009; Christie et al. 2014). Yet, the genetic basis of these differences, and the rate at which they occur, remain unclear. Using an experimental system that included contrasting lines founded from a single population and sampled extensively over multiple generations, we

began to address these uncertainties by estimating the rate of genetic change in hatchery fish with unprecedented resolution. While the magnitude of divergence observed in the segregated line was small, it is important to consider that this change occurred after only three generations of captive rearing. If the trend of divergence continues, then F_{ST} could equal or exceed values typically observed between natural Chinook salmon populations (e.g., Briec et al. 2015) within another three to five generations. Thus, we consider the rate at which genetic differentiation occurred to be noteworthy and of significant interest to conservation programs. We also identified processes driving change in the segregated line, including the first possible indicators consistent with domestication selection observed in Pacific salmon hatchery populations.

Determining the impact of different captive rearing approaches on population productivity in species of conservation interest is central to measuring their success, but empirical studies on such species are usually constrained by the natural systems in which they operate. An ideal experiment would demonstrate that integrated fish have greater fitness than segregated fish in the natural environment. However, permitting spawning of the segregated fish in the same environment as the integrated fish would promote gene flow between the lines, and negate both the experiment and the concurrent restoration efforts in the Chinook salmon system we have studied. Therefore, it was necessary to use proxies for fitness by identifying signatures of selection and measuring effective population size, consistent with best practices in genetic monitoring (Schwartz et al. 2007). These measures can be further enhanced by integrating genome-wide association studies with phenotypic measures of fitness.

This study is the first to empirically compare multi-generational consequences of alternative management approaches in a nonmodel species of conservation interest using genome-wide surveys, and adds to the range of available tools for assessing risks associated with supportive breeding. Multigenerational observations are advantageous because processes occurring in the wild, including natural selection, could mitigate or exacerbate the effects of captive rearing over time. Our findings provide the first empirical demonstration that using natural-born parents in captive breeding programs is more effective at reducing genetic divergence than using only captive-born individuals. Interpreting these findings in a management context can be challenging on one hand, because determining 'acceptable' levels of divergence depends in part on a clear link between the molecular measures we report and population productivity. On the other hand, the study is comparative in nature and considers a range of possible outcomes, an approach that is highly relevant to risk assessment in realistic scenarios

(Waples 1999; Duchesne and Bernatchez 2002; Ford 2002; Waples and Drake 2004; Naish et al. 2008). The experimental system we evaluated therefore provides guidance on 'best practices' in supportive breeding. Conservation plans should thus consider establishing captive breeding programs using wild broodstock before the source population becomes too small, so that the benefits of starting such a program outweigh the demographic and genetic costs of removing individuals from the natural environment.

Acknowledgements

We thank everyone who was involved in establishing CESRF and shaping its research direction, including Levi George, Melvin Sampson, Steve Schroder, Craig Busack, past and present members of the Independent Scientific Review Panel, and the Yakama Nation Tribal. We also recognize the following individuals for project development, broodstock collection and sampling, and laboratory assistance: Michael Ford, all Yakama Nation and Washington Department of Fish and Wildlife personnel at the Roza Dam Adult Monitoring Facility and CESRF, Isadora Jimenez-Hidalgo, Katrina van Raay, and Daniel Drinan. Finally, we thank two anonymous referees and the Editor in Chief, Louis Bernatchez, for their insightful comments. Funding for this study was provided by NOAA Fisheries/Federal Columbia River Power System (FCRPS) Biological Opinion Remand Funds (to K.A.N. and J.J.H.), Washington Sea Grant (Award NA14OAR4170078 to K.A.N.), and the Marine Technology Society Graduate Student, John G. Peterson Endowed, Richard T. Whiteleather, and John N. Cobb Scholarships awarded to C.D.W.

Data archiving statement

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.v33f0>.

Literature cited

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.
- Anderson, J. H., P. L. Faulds, W. I. Atlas, and T. P. Quinn 2013. Reproductive success of captively bred and naturally spawned Chinook salmon colonizing newly accessible habitat. *Evolutionary Applications* 6:165–179.
- Araki, H., B. Cooper, and M. S. Blouin 2007. Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. *Science* 318:100–103.
- Araki, H., B. Cooper, and M. S. Blouin 2009. Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. *Biology Letters* 5:621–624.

- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* 3:7.
- 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.
- Beaumont, M. A., and R. A. Nichols 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences* 263:1619–1626.
- Berthelot, C., F. Brunet, D. Chalopin, A. Juanchich, M. Bernard, B. Noel, P. Bento, et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* 5:10.
- Borcard, D., F. Gillet, and P. Legendre 2011. *Numerical Ecology with R*. Springer Verlag, New York.
- Brieuc, M. S. O., C. D. Waters, J. E. Seeb, and K. A. Naish 2014. A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3-Genes Genomes. Genetics* 4:447–460.
- Brieuc, M. S. O., K. Ono, D. P. Drinan, and K. A. Naish 2015. Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* 24:2729–2746.
- Burkhead, N. M. 2012. Extinction rates in North American freshwater fishes, 1900–2010. *Bioscience* 62:798–808.
- Busack, C. A., and K. P. Currens 1995. Genetic risks and hazards in hatchery operations: fundamental concepts and issues. *American Fisheries Science Symposium* 15:71–80.
- Busack, C., and A. Marshall 1991. Genetic analysis of YFP Chinook salmon stocks. In *Yakima hatchery experimental design*. Progress Report to Bonneville Power Administration, edited by S. Phelps and D. Seiler. Portland, Oregon.
- Busack, C., C. M. Knudsen, G. Hart, and P. Huffman 2007. Morphological differences between adult wild and first-generation hatchery upper Yakima River spring Chinook salmon. *Transactions of the American Fisheries Society* 136:1076–1087.
- Campton, D. E. 1995. Genetic effects of hatchery fish on wild populations of Pacific salmon and steelhead: what do we really know? *Transactions of the American Fisheries Society* 15:337–353.
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22:3124–3140.
- Christie, M. R., M. L. Marine, R. A. French, R. S. Waples, and M. S. Blouin 2012. Effective size of a wild salmonid population is greatly reduced by hatchery supplementation. *Heredity* 109:254–260.
- Christie, M. R., M. J. Ford, and M. S. Blouin 2014. On the reproductive success of early-generation hatchery fish in the wild. *Evolutionary Applications* 7:883–896.
- Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
- De Mita, S., A. C. Thuillet, L. Gay, N. Ahmadi, S. Manel, J. Ronfort, and Y. Vigouroux 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* 22:1383–1399.
- Dittman, A. H., D. May, D. A. Larsen, M. L. Moser, M. Johnston, and D. Fast 2010. Homing and spawning site selection by supplemented hatchery- and natural-origin Yakima river spring Chinook salmon. *Transactions of the American Fisheries Society* 139:1014–1028.
- Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden 2014. N_EESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources* 14:209–214.
- Duchesne, P., and L. Bernatchez 2002. An analytical investigation of the dynamics of inbreeding in multi-generation supportive breeding. *Conservation Genetics* 3:47–60.
- Fast, D. E., W. J. Bosch, M. V. Johnston, C. R. Strom, C. M. Knudsen, A. L. Fritts, G. M. Temple, et al. 2015. A synthesis of findings from an integrated hatchery program after three generations of spawning in the natural environment. *North American Journal of Aquaculture* 77:377–395.
- Fleming, I. A., and M. R. Gross 1994. Breeding competition in a Pacific salmon (coho: *Oncorhynchus kisutch*): measures of natural and sexual selection. *Evolution* 48:637–657.
- Foll, M., and O. Gaggiotti 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–993.
- Ford, M. J. 2002. Selection in captivity during supportive breeding may reduce fitness in the wild. *Conservation Biology* 16:815–825.
- Frankham, R. 2008. Genetic adaptation to captivity in species conservation programs. *Molecular Ecology* 17:325–333.
- Fraser, D. J. 2008. How well can captive breeding programs conserve biodiversity? A review of salmonids. *Evolutionary Applications* 1:535–586.
- Fritts, A. L., J. L. Scott, and T. N. Pearsons 2007. The effects of domestication on the relative vulnerability of hatchery and wild origin spring Chinook salmon (*Oncorhynchus tshawytscha*) to predation. *Canadian Journal of Fisheries and Aquatic Sciences* 64:813–818.
- Götz, S., J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36:3420–3435.
- Harstad, D. L., D. A. Larsen, and B. R. Beckman 2014. Variation in minijack rate among hatchery populations of Columbia River basin Chinook salmon. *Transactions of the American Fisheries Society* 143:768–778.
- Hess, M. A., C. D. Rabe, J. L. Vogel, J. J. Stephenson, D. D. Nelson, and S. R. Narum 2012. Supportive breeding boosts natural population abundance with minimal negative impacts on fitness of a wild population of Chinook salmon. *Molecular Ecology* 21:5236–5250.
- Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics* 6:23.
- Jombart, T. 2008. *adeigenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405.
- Jombart, T., S. Devillard, and F. Balloux 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:15.
- Jule, K. R., L. A. Leaver, and S. E. G. Lea 2008. The effects of captive experience on reintroduction survival in carnivores: a review and analysis. *Biological Conservation* 141:355–363.
- Karlsson, S., and T. Moen 2010. The power to detect artificial selection acting on single loci in recently domesticated species. *BMC Research Notes* 3:232.
- Knudsen, C. M., S. L. Schroder, C. A. Busack, M. V. Johnston, T. N. Pearsons, W. J. Bosch, and D. E. Fast 2006. Comparison of life history traits between first-generation hatchery and wild upper Yakima river

- spring Chinook salmon. *Transactions of the American Fisheries Society* **135**:1130–1144.
- Knudsen, C. M., S. L. Schroder, C. Busack, M. V. Johnston, T. N. Pearsons, and C. R. Strom 2008. Comparison of female reproductive traits and progeny of first-generation hatchery and wild upper Yakima river spring Chinook salmon. *Transactions of the American Fisheries Society* **137**:1433–1445.
- Koskinen, M. T., T. O. Haugen, and C. R. Primmer 2002. Contemporary fisherian life-history evolution in small salmonid populations. *Nature* **419**:826–830.
- Laikre, L., M. K. Schwartz, R. S. Waples, N. Ryman, and M. W. G. Ge 2010. Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. *Trends in Ecology & Evolution* **25**:520–529.
- Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:10.
- Larsen, D. A., B. R. Beckman, and K. A. Cooper 2010. Examining the conflict between smolting and precocious male maturation in spring (stream-type) Chinook salmon. *Transactions of the American Fisheries Society* **139**:564–578.
- Larson, W. A., L. W. Seeb, M. V. Everett, R. K. Waples, W. D. Templin, and J. E. Seeb 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications* **7**:355–369.
- Lichtowich, J. A., and L. E. Moberg 1995. Analysis of Chinook salmon in the Columbia River from an ecosystem perspective. In *Research Report to Bonneville Power Administration*. Portland, Oregon.
- Lotterhos, K. E., and M. C. Whitlock 2014. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* **23**:2178–2192.
- Lynch, M., and M. O'Hely 2001. Captive breeding and the genetic fitness of natural populations. *Conservation Genetics* **2**:363–378.
- Mäkinen, H., A. Vasemägi, P. McGinnity, T. F. Cross, and C. R. Primmer 2015. Population genomic analyses of early-phase Atlantic Salmon (*Salmo salar*) domestication/captive breeding. *Evolutionary Applications* **8**:93–107.
- Mañanos, E. L., I. Anglade, J. Chyb, C. Saligaut, B. Breton, and O. Kah 1999. Involvement of gamma-aminobutyric acid in the control of GTH-1 and GTH-2 secretion in male and female rainbow trout. *Neuroendocrinology* **69**:269–280.
- Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society* **78**:1069–1079.
- Martin, S. A. M., S. Blaney, A. S. Bowman, and D. F. Houlihan 2002. Ubiquitin-proteasome-dependent proteolysis in rainbow trout (*Oncorhynchus mykiss*): effect of food deprivation. *Pflügers Archiv-European Journal of Physiology* **445**:257–266.
- McGinnity, P., P. Prodohl, K. Ferguson, R. Hynes, N. O'Maoileidigh, N. Baker, D. Cotter, et al. 2003. Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society B-Biological Sciences* **270**:2443–2450.
- Milot, E., C. Perrier, L. Papillon, J. J. Dodson, and L. Bernatchez 2013. Reduced fitness of Atlantic salmon released in the wild after one generation of captive breeding. *Evolutionary Applications* **6**:472–485.
- Moberg, L. E., J. Barr, L. Blankenship, D. E. Campton, T. T. P. Evelyn, T. A. Flagg, C. V. W. Mahnken, et al. 2005. Hatchery reform in Washington State: principles and emerging issues. *Fisheries* **30**:11–23.
- Naish, K. A., J. E. Taylor, P. S. Levin, T. P. Quinn, J. R. Winton, D. Huppert, and R. Hilborn. 2008. An evaluation of the effects of conservation and fishery enhancement hatcheries on wild populations of salmon. In *Advances in Marine Biology*. San Diego, CA: Elsevier Academic Press Inc.
- Naish, K. A., T. R. Seamons, M. B. Dauer, L. Hauser, and T. P. Quinn 2013. Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (*Oncorhynchus mykiss*) population released in the wild. *Molecular Ecology* **22**:1295–1309.
- National Research Council 1996. *Upstream: Salmon and Society in the Pacific Northwest*. National Academy Press, Washington, D.C.
- Neff, B. D., S. R. Garner, and T. E. Pitcher 2011. Conservation and enhancement of wild fish populations: preserving genetic quality versus genetic diversity. *Canadian Journal of Fisheries and Aquatic Sciences* **68**:1139–1154.
- Panserat, S., G. A. Hortopan, E. Plagnes-Juan, C. Kolditz, M. Lansard, S. Skiba-Cassy, D. Esquerre et al. 2009. Differential gene expression after total replacement of dietary fish meal and fish oil by plant products in rainbow trout (*Oncorhynchus mykiss*) liver. *Aquaculture* **294**:123–131.
- Paquet, P. J., T. Flagg, A. Appleby, J. Barr, L. Blankenship, D. Campton, M. Delarm, et al. 2011. Hatcheries, conservation, and sustainable fisheries-achieving multiple goals: results of the Hatchery Scientific Review Group's Columbia river basin review. *Fisheries* **36**:547–561.
- Pimm, S., P. Raven, A. Peterson, C. H. Sekercioglu, and P. R. Ehrlich 2006. Human impacts on the rates of recent, present, and future bird extinctions. *Proceedings of the National Academy of Sciences of the United States of America* **103**:10941–10946.
- RASP (Regional Assessment of Supplementation Planning). 1992. *Supplementation in the Columbia River basin summary report series*. In *Final Report to Bonneville Power Administration, Project 85-62*. Portland, Oregon.
- Raymond, M., and F. Rousset 1995. Genepop (version 1.2) - Population-genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**:248–249.
- Ryman, N., and L. Laikre 1991. Effects of supportive breeding on the genetically effective population size. *Conservation Biology* **5**:325–329.
- Scheuerell, M. D., E. R. Buhle, B. X. Semmens, M. J. Ford, T. Cooney, and R. W. Carmichael 2015. Analyzing large-scale conservation interventions with Bayesian hierarchical models: a case study of supplementing threatened Pacific salmon. *Ecology and Evolution* **5**:2115–2125.
- Schroder, S. L., C. M. Knudsen, T. N. Pearsons, T. W. Kassler, S. F. Young, C. A. Busack, and D. E. Fast 2008. Breeding success of wild and first-generation hatchery female spring Chinook salmon spawning in an artificial stream. *Transactions of the American Fisheries Society* **137**:1475–1489.
- Schroder, S. L., C. M. Knudsen, T. N. Pearsons, T. W. Kassler, S. F. Young, E. P. Beall, and D. E. Fast 2010. Behavior and breeding success of wild and first-generation hatchery male spring Chinook salmon spawning in an artificial stream. *Transactions of the American Fisheries Society* **139**:989–1003.
- Schwartz, M. K., G. Luikart, and R. S. Waples 2007. Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution* **22**:25–33.
- Seiliez, I., S. Panserat, S. Skiba-Cassy, A. Fricot, C. Vachot, S. Kaushik, and S. Tesseraud 2008. Feeding status regulates the polyubiquitination step of the ubiquitin-proteasome-dependent proteolysis in rainbow trout (*Oncorhynchus mykiss*) muscle. *Journal of Nutrition* **138**:487–491.

- Storey, J. D. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**:479–498.
- Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**:671–688.
- Suzuki, K., H. Kawachi, and Y. Nagahama 1988. Isolation and characterization of two distinct gonadotropins from chum salmon pituitary glands. *General and Comparative Endocrinology* **71**:292–301.
- Therkildsen, N. O., J. Hemmer-Hansen, T. D. Als, D. P. Swain, M. J. Morgan, E. A. Trippel, S. R. Palumbi, et al. 2013. Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. *Molecular Ecology* **22**:2424–2440.
- de Villemeureuil, P., E. Fricot, E. Bazin, O. Francois, and O. E. Gaggiotti. 2014. Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology* **23**:2006–2019.
- Waples, R. S. 1999. Dispelling some myths about hatcheries. *Fisheries* **24**:12–21.
- Waples, R. S. 2002. Effective size of fluctuating salmon populations. *Genetics* **161**:783–791.
- Waples, R. S. 2005. Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? *Molecular Ecology* **14**:3335–3352.
- Waples, R. S., and J. Drake 2004. Risk/Benefit considerations for marine stock enhancement: a Pacific salmon perspective. In K. M. Leber, S. Kitada, H. L. Blankenship, and T. Svåsand, eds. *Stock Enhancement and Sea Ranching: Developments, Pitfalls and Opportunities*. Blackwell Publishing Ltd, Oxford, UK.
- Waples, R. S., T. Antao, and G. Luikart 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics* **197**:769–780.
- Weir, B. S., and C. C. Cockerham 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**:1358–1370.
- Williams, S. E., and E. A. Hoffman 2009. Minimizing genetic adaptation in captive breeding programs: a review. *Biological Conservation* **142**:2388–2400.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97–159.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Lists of the 9410 RAD loci and 413 individuals, with their populations, analyzed in this study.

Table S2. Description of wild founders and each generation of the integrated (INT) and segregated (SEG) hatchery lines, including year sampled, the number of adults used as broodstock, the number of broodstock sampled in this study, the final sample size retained for each population after removing individuals with >50% missing genotypes, observed (H_o) and expected (H_e) heterozyosity, and the number of loci that deviated from Hardy-Weinberg equilibrium after correcting for multiple tests.

Table S3. Pairwise population differentiation (F_{ST}) compared to the P_1 Founders and between the two lines within each generation.

Table S4. F_{ST} per locus for each population when compared to the P_1 Founders.

Table S5. (a) Spawning groups and census sizes (N_{census}) to which estimates of effective number of breeders, N_b , apply, the proportions of N_{census} that were of hatchery origin ($P_{hatchery}$), and the years and populations from which N_b estimates were obtained using the linkage disequilibrium (LD) method. N_{b_LD} (95% CI) are the LD estimates with 95% confidence intervals, Bias is the approximate bias in the LD estimates resulting from overlapping generations and fluctuating population sizes, N_{b_Adj} (95% CI) are the bias-adjusted estimates with 95% confidence intervals, and N_{b_Adj}/N_{census} is the estimated ratio of successful spawners to potential spawners. (b) Effective number of breeders (95% CI) for both hatchery lines from 1998–2010 estimated using the temporal method.

Table S6. Bias calculations for N_b estimates generated by the LD method in *NeEstimator* for the (a) river (wild and integrated line) and (b) hatchery (segregated line).

Table S7. Estimated contribution of naturally spawning fish (i.e. managed gene flow) to N_b in the integrated line.

Table S8. Discriminatory loci identified for DF1 of the DAPC, and outlier loci identified by *Bayescan* and F_{TEMP} .

Table S9. Regions identified by the sliding window analyses as showing significantly elevated divergence from the P_1 founders.

Table S10. (a) Alignments to the rainbow trout (RBT) genome, BLAST results, gene functions, and gene ontologies (GO Slim Terms) for all mapped loci within the seven genomic regions that exhibited overlap among the three tests of adaptive divergence (Region of Overlap, described in Table 1). (b) Complete references, listed in alphabetical order, for the gene functions described in Table S10a.