

Patterns of genomic variation in Coho salmon following reintroduction to the interior Columbia River

Nathan R. Campbell¹  | Cory Kamphaus² | Keely Murdoch² | Shawn R. Narum¹ 

¹Columbia River Inter-Tribal Fish Commission, Hagerman, ID, USA

²Yakama Nation Fisheries, Yakima, WA, USA

Correspondence

Nathan R. Campbell, Columbia River Inter-Tribal Fish Commission, Hagerman, ID, USA.
Email: camn@critfc.org

Funding information

Bonneville Power Administration, Grant/Award Number: 1996-040-00

Abstract

Coho salmon were extirpated in the mid-20th century from the interior reaches of the Columbia River but were reintroduced with relatively abundant source stocks from the lower Columbia River near the Pacific coast. Reintroduction of Coho salmon to the interior Columbia River (Wenatchee River) using lower river stocks placed selective pressures on the new colonizers due to substantial differences with their original habitat such as migration distance and navigation of six additional hydropower dams. We used restriction site-associated DNA sequencing (RAD-seq) to genotype 5,392 SNPs in reintroduced Coho salmon in the Wenatchee River over four generations to test for signals of temporal structure and adaptive variation. Temporal genetic structure among the three broodlines of reintroduced fish was evident among the initial return years (2000, 2001, and 2002) and their descendants, which indicated levels of reproductive isolation among broodlines. Signals of adaptive variation were detected from multiple outlier tests and identified candidate genes for further study. This study illustrated that genetic variation and structure of reintroduced populations are likely to reflect source stocks for multiple generations but may shift over time once established in nature.

KEYWORDS

Coho salmon, genetic adaptation, genetic monitoring, RAD sequencing, reintroduction

1 | INTRODUCTION

Reintroduction of animals into habitats they once populated is a conservation practice that is being utilized with increasing frequency (e.g., Seddon, Armstrong, & Maloney, 2007). As human influences on habitats have been the root cause of decline, extirpation, or extinction of various species (Dobson et al., 2006; Estes et al., 2011), changes in management policies such as environmental regulations, harvest restrictions, limitations on development, and habitat restoration have made reintroduction a possibility for some species (Corlett, 2015; Wallace, Clark, & Reading, 2002). Many reintroduction efforts, however, fail in their objective to establish viable naturally reproducing populations (Griffith, Scott, Carpenter, & Reed, 1989). The success

or failure of reintroduction efforts is dependent on multiple factors including suitability of habitat to support reintroduced species, proper management following reintroduction, and the ability of reintroduced populations to adapt to new habitat (Seddon et al., 2007). Genetic diversity within reintroduced populations provides capacity for adaptive changes to the new environment as beneficial combinations of gene variants become enriched in subsequent generations. Therefore, factors which would reduce genetic diversity such as a small initial reintroduced population (founder effect), high mortality rates following reintroduction (genetic bottleneck), or the breeding of related animals (inbreeding) are taken into careful consideration when planning and managing reintroduction efforts (Haig, Ballou, & Derrickson, 2014; Jamieson, 2011).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Anadromous fishes such as salmonids utilize various habitats throughout their life cycles (small streams, large rivers, estuaries, and the ocean), and alterations to any of them may lead to population declines or extirpation (Gustafson et al., 2007). However, reintroduction efforts for salmon typically focus on spawning habitat as homing behavior (philopatry) is the basis for population structure in these species (Hendry, Castric, Kinnison, & Quinn, 2004; Keefer & Caudill, 2014). In order to establish a new population, juvenile salmon must imprint on a new stream before they begin their migration to the ocean in order to return to the same location as an adult (Quinn 1993; Lohmann, Putman, & Lohmann, 2008; Johnstone, Lubieniecki, Koop, & Davidson, 2011). High mortality rates prior to adulthood can mean replacement levels may not be met for newly established naturally spawning salmon populations. For these reasons, many populations of salmon are kept viable using hatchery supplementation to produce enough juvenile fish to maintain adult returns at or above replacement (Fast et al., 2015; Ford, Murdoch, Hughes, Seamons, & LaHood, 2016; Hess et al., 2012).

In the Columbia River drainage in the Pacific Northwest of the United States, Coho salmon historically utilized many of the upriver tributaries as viable spawning habitat. However, as with other salmonid species, large portions of Coho salmon spawning habitat were either blocked by impassable dams or otherwise rendered unusable by this species. Early mitigation efforts to restore salmonid species through hatchery programs in the Columbia River system focused mainly on Chinook salmon, Sockeye salmon, and steelhead, while little effort was made toward restoring Coho salmon populations (Kareiva, Marvier, & McClure, 2000; Paquet et al., 2011; Williams, 2008). As a result, by the 1980s, all Coho salmon populations above Bonneville Dam (river mile 145) had been extirpated (Galbreath, Bisbee, Dompier, Kamphaus, & Newsome, 2014). In the late 1990s and early 2000s, the Yakama, Umatilla, and Nez Perce tribes began hatchery-based reintroduction programs in several mid and upper Columbia River tributaries (Yakima R., Wenatchee R., Methow R., and Clearwater R.) using offspring from lower Columbia River adult fish as their source stock.

In the Wenatchee River, Coho salmon from lower Columbia River stocks were reintroduced in order to establish a locally adapted population (Bosch et al., 2007; Galbreath et al., 2014). Releases of juvenile smolts (1.5 year old) in 1999, 2000, and 2001 produced the first generation of 3-year-old adult returns in 2000, 2001, and 2002 (Galbreath et al., 2014; Murdoch, Prevatte, & Kamphaus, 2005). Although first-generation fish in 2000 and 2002 required supplementation with lower Columbia River juveniles in order to meet production goals, the program quickly transitioned to exclusively using fish returning to the Wenatchee system as broodstock. Selective forces acting on the new population were especially significant due to freshwater migration distances more than 10 times that of their source stock for both emigrating smolts and returning adult fish (Wenatchee River [756 km migration]; source stock [70 km migration]). Phenotypic differences correlating with freshwater migration distances in anadromous fish have been previously reported (Crossin et al., 2004; Jonsson & Jonsson, 2006). Energetic demands and exposure to stress related to increased migration distance are hypothesized to create selective

pressure on the new Coho salmon population which may manifest as significant population-wide changes in allele frequencies within areas of the genome.

In this study, we examine patterns of neutral and adaptive genetic variation in this reintroduced population of Coho salmon over the course of four generations. To screen for signals of genetic adaptation, SNP markers throughout the genome were examined for significant changes in allele frequency over multiple generations to identify outlier loci that exhibited evidence for divergent selection as opposed to genetic drift. As Coho salmon populations can display distinct temporal structure following three-year age classes (Smith et al., 2015), this provided the opportunity to test for consistent outlier results among biological replicates within the same study system.

2 | METHODS

2.1 | Tissue samples

During the initial reintroduction (1999–2001), juvenile smolts raised in lower river hatcheries were released into the Wenatchee River. In later years, adult fish returning to the Wenatchee River were used to produce locally acclimated juvenile Coho salmon. Tissue samples (fin clips) were collected from these adult broodstock from 2000 to 2011 for DNA extraction and genotyping. Supplementation with lower river stocks was used in 2000 and 2002 but in subsequent years adult fish returning to the Wenatchee River system became suitably abundant to meet hatchery production and escapement goals (Figure 1). A total of 664 samples were included for genotyping, with four generations of fish represented from each of three broodlines (Figure 1). However, samples were not available from one generation (year 2003) of broodline “A.” Broodline groups were labeled “A” (2000, 2006, and 2009), “B” (2001, 2004, 2007, and 2010), and “C” (2002, 2005, 2008, and 2011) and will be referred to as such hereafter (Figure 1).

2.2 | Library preparation and genotyping

Genomic DNA was extracted from fin tissue using Qiagen DNeasy 96 kits and quantified with a Tecan m200 96-well plate reader fluorimeter using Invitrogen Quant-iT™ pico green reagent. The DNA concentration of each sample was determined by the standard curve method using samples of known DNA concentration. Restriction site-associated DNA (RAD-seq; Miller et al. 2007) libraries were prepared by first digesting 500 ng of each sample with *Sbf*I-HF (NEB; New England Biolabs, Ipswich, MA, USA) followed by heat inactivation of the enzyme. The remainder of the preparation followed the instructions of previously published studies (Hecht, Campbell, Holecek, & Narum, 2013; Miller et al., 2012). In brief, adapters containing six base barcodes were ligated onto the *Sbf*I cut sites using T4 ligase (NEB). Uniquely barcoded samples were then combined into pools of 48 individual samples per pool for further preparation. Three hundred microliters of each DNA pool was sheared to an average size range of 500 base pairs using a Bioruptor sonicator (Diagenode). The sheared DNA was then purified using Qiagen MinElute columns, size selected using

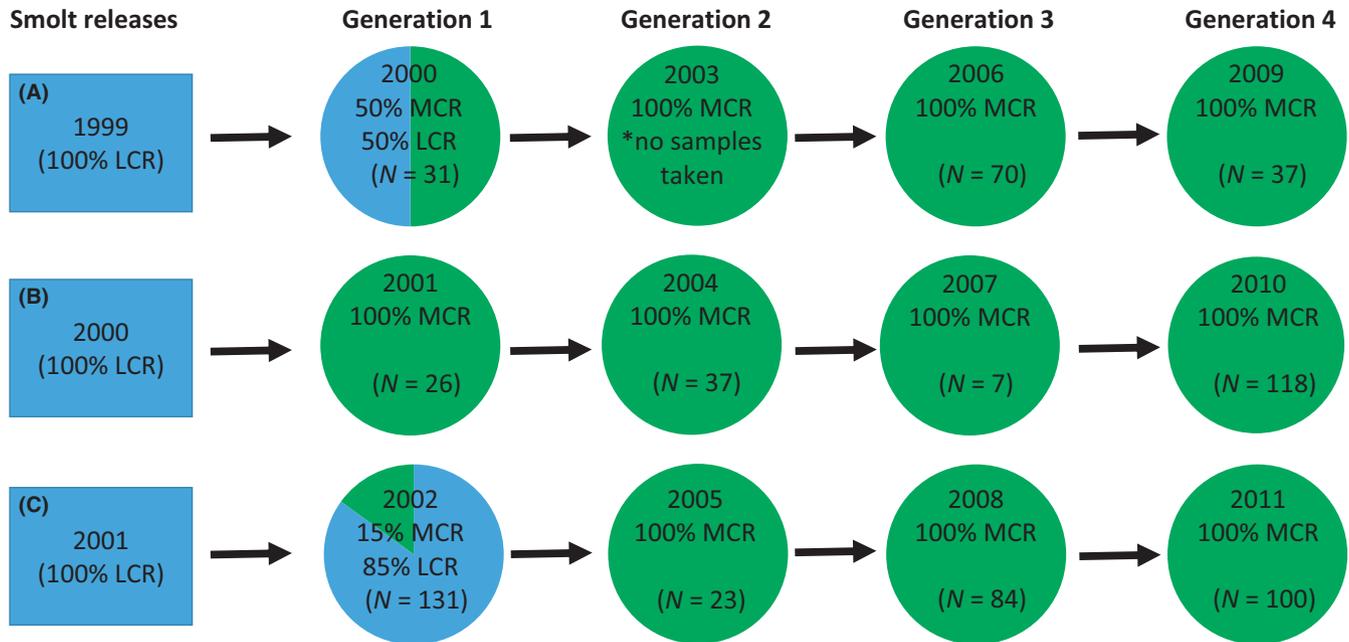


FIGURE 1 Coho broodstock samples. This graphic outlines the reintroduction program in the Wenatchee River. Reintroduction was started by releasing lower Columbia River (LCR) smolts beginning in 1997. Returning two- to three-year-old fish (middle Columbia River; MCR) were used as broodstock in subsequent years with some reliance on LCR stock in generation 1. Numbers in parentheses indicate the successfully genotyped fish for each broodyear. The A, B, and C broodlines identified by principal coordinate analysis are denoted

Agencourt Ampure XP magnetic beads to a size range of 200 to 700 base pairs, blunt-ended using a quick blunting kit (NEB), A-tailed using Klenow fragment DNA polymerase (NEB), second adapter added using T4 ligase (NEB), amplified with 14 cycles of PCR using Phusion Master Mix (NEB), and purified with Agencourt Ampure XP beads. Final library concentrations were quantified using 2× Sybr Green Master Mix on a QuantStudio 6 (Thermo Fisher) qPCR instrument and normalized to a concentration of 5 ng/μl. Each library was then sequenced with an Illumina HiSeq 1500 instrument for 100 cycles from the read 1 primer site (single read). A single library was chosen to run as a paired-end run with 100 cycles in each direction to create longer contigs for individual loci later.

Genotypes were generated from the raw sequencing data using the de novo pipeline in the program Stacks v1.03 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Raw sequences were trimmed to 80 bases, quality filtered, and binned into fastq files for each individual sample using the “process_radtags” script. In samples where an insufficient number of sequencing reads were collected, additional libraries were created to attempt to collect enough sequencing data for adequate genotyping. Where applicable, sequencing reads from the same individual (but different libraries) were concatenated. Individual samples with less than 1M total reads were immediately excluded from the dataset. The average number of raw reads for the remaining samples was 2.9M but varied greatly among samples (1M–14M). As higher numbers of raw reads increase the likelihood that reads with convergent sequencing errors will be combined into a stack, the minimum stack depth (m parameter) used with the “ustacks” program was scaled according to raw read numbers. The m parameter was set to the nearest integer using the formula (raw reads in millions * 2). The

distance allowed between stacks (M parameter) was set at 2. The cstacks portion of the pipeline to build the catalog of SNPs included a total of 22 ascertainment samples, with two individual samples from each year class of the three broodlines that had between 2.5M and 3.5M reads per sample. The number of mismatches allowed between sample tags when building the catalog was set to 2 (n parameter). The sstacks module was then run with all samples using the newly created catalog, and the output data were then compiled into a raw vcf (variant call format) file using the “populations” module (rather than outputting to a mysql database). SNP loci were then filtered directly from the vcf output using a custom perl script which enforced a minor allele frequency cutoff of 5.0%. Samples with less than 75% of the genotypes from the remaining SNP loci were then removed from the dataset followed by excluding any remaining SNP loci with less than 80% of the genotypes within the remaining samples. In order to account for SNP loci with potential null alleles and paralogous sequent variations (PSVs), deviation from Hardy–Weinberg expectations was then tested for all remaining loci using all the remaining samples as a single population. Loci with corrected *p*-values less than the BY-FDR (Benjamini & Yekutieli False Discovery Rate; Benjamini & Yekutieli, 2001) corrected value ($\alpha = 0.01$; FDR corrected *p*-value = 1.8×10^{-6}) were removed from the dataset.

2.3 | Testing for SNP loci under selection

The remaining samples and loci were converted to GenePop files and imported into Excel using the GenAIEx plug-in (Peakall & Smouse, 2006). The collections from each year were used as independent populations for the purposes of identifying possible underlying genetic

structure. A principal components analysis (PCA) was conducted using pairwise genetic distance data generated using GenAlEx. Each of the three broodlines was treated as biological replicates of the same reintroduction experiment. Genotype data were split into three datasets with each broodline composed of four generations of adult broodstock fish.

In order to protect against false positives in outlier tests (Lotterhos & Whitlock, 2014; Narum & Hess, 2011), three different methods were used to identify potential candidates for divergent selection within each broodline. As each broodline was tested separately, there was minimal underlying population structure that could bias outlier results. The methods included linear regression of allele frequency changes, and two standard applications of outlier tests implemented with software programs Lositan (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008), and Bayescan 2.0 (Foll & Gaggiotti, 2008). In order for loci to be considered candidates under divergent selection in our final outlier results, they had to be significant in at least two of the three methods as applied below.

A linear regression method was used as an initial test for loci with significant changes in allele frequency across generations within each broodline. To distinguish potential loci under selection from those experiencing random genetic drift, a significance threshold was established to identify SNPs with a greater change in allele frequency than expected by chance for neutral markers. Rather than simply applying the frequentist approach of classifying all slopes outside of an arbitrary cutoff value as outliers (i.e., two standard deviations) which will always return approximately 5% of the loci, we applied a different strategy which would allow the possibility that no loci were significant. To identify outliers that were considered statistically significant from neutral loci, a Q-Q test was performed using the slopes for each locus for each of the broodlines. A chi-square test was performed for each locus using the difference between the expected slope (based on Q-Q linearity) and observed slope at each of the loci. Loci with p -values of less than .01 were designated as outliers. This cutoff value corresponded to a deviation from Q-Q linearity of 3.56 \times the average deviation across all loci.

A second test for loci under selection utilized the program Lositan (Antao et al., 2008). This program examines the relationship between F_{ST} and heterozygosity (H_e) at each locus and identifies loci outside of the specified confidence interval for neutral loci as outliers. This program was executed separately for each of the three broodline datasets specifying samples from each generation as a population using the "infinite alleles" mutation model. The confidence interval was set to 0.95, the false discovery rate was set to 0.1, the subsample size was set to 50, the number of simulations was set to 50,000, and the program was set to first calculate a neutral mean F_{ST} . Samples from the 2007 year class were excluded from the analysis due to small sample size ($N = 7$).

The program Bayescan was used as a third approach to identify loci under selection within the datasets (Foll & Gaggiotti, 2008). For these analyses, the 2007 data were not omitted due to the reported lack of bias with small sample sizes when using this method. A custom perl script was used to convert the GenePop formatted genotype files

for each of the three broodlines to Bayescan formatted input files. The program was run using default parameters using the SNP genotypes as regular codominant data.

The locations of all RAD loci used in the analyses were mapped to the Okis_V1 reference genome assembly (GenBank assembly accession: GCA_002021735.1). As the traditional RAD library preparation method uses sonication to randomly shear the DNA following P1 adapter ligation, the resulting library constructs are of varying length. Some of the R2 reads will overlap with the R1 reads allowing their assembly into longer contigs. Using the RAD tag sequences for each locus, we used a series of custom perl scripts to sample paired-end data for assembly into longer contigs using the CAP3 program (Huang & Madan, 1999). The program bowtie2 (Langmead & Salzberg, 2012) was used to align RAD sequences to the two reference genomes in order to determine the most likely genome locations of each of the outlier loci. Genome positions with a mapQ score of less than 10 were discarded and when a locus mapped to multiple positions, all possible locations were retained. Another custom perl script was used to identify the three nearest genes within 100,000 bases of the mapped outlier loci using the annotation file for the reference assembly.

3 | RESULTS

Tissue samples collected over the course of the reintroduction period varied greatly in quality and many samples failed to produce genotype data meeting the minimum requirements for inclusion in the final dataset. Poor genotyping was generally the result of low numbers of sequencing reads for many of the individual samples. After the initial screening of 12 libraries containing between 81 and 96 individuals each, 192 samples were targeted for additional sequencing (two HiSeq lanes) in order to boost the total number of raw sequencing reads to the target range of 2–4M reads. However, only 68% of the resequenced samples produced enough combined reads for the genotypes to pass filtering. In total, 664 samples were utilized for the analyses of the 1,103 attempted for RAD genotyping after quality filtering of the genotype data. Quality filtering of the SNP loci identified using the denovo Stacks pipeline left a total of 5,392 markers in the dataset.

Temporal genetic structure among the three broodlines of reintroduced fish was evident from PCA plots that revealed three distinct clusters (Figure 2). Each cluster corresponded to a broodline from one of the initial return years (2000, 2001, and 2002) and their descendants. This was likely the result of genetic drift between year classes within source stocks due to a lack of intergenerational gene flow (i.e., spawning is limited to the 3-year-old age class) as has been observed previously in hatchery-reared stocks of Coho salmon (Smith et al., 2015). Broodline "B" was the most distinct from the other two broodlines in the PCA plot (Figure 2). As each of these lineages was genetically distinct, each was treated as a separate dataset for remaining analyses. Each dataset was composed of the initial founder stock and 2–3 generations of descendants (no samples were collected in brood year 2003).

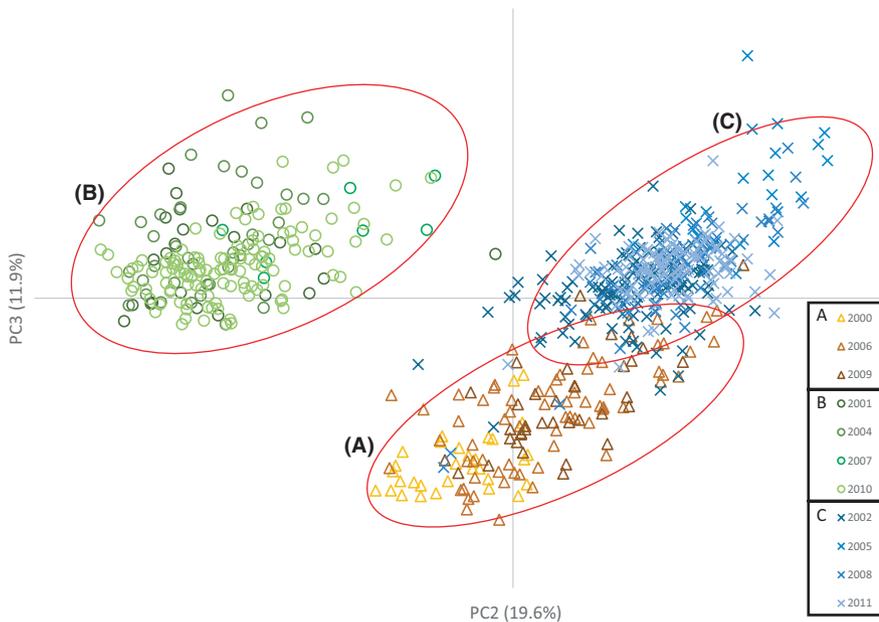


FIGURE 2 Principal coordinate analysis reveals genetic distinctions between brood years in reintroduced Wenatchee River Coho salmon. These clusters were termed the A, B, and C broodlines

The first method used for detecting loci under selection was to inspect the genotype data for significant changes in allele frequency over time. Analysis of allele frequency trends showed a normal distribution of slopes within each of the three broodline datasets (Figure 3a). Broodlines A and B, however, showed a broader distribution that is likely reflective of lower accuracy in the allele frequency estimates in those datasets due to low representation within some brood years (Figure 1). Loci under selection were expected to have the steepest slopes compared to those under random genetic drift as reflected by the linearity of ordered loci in Q-Q plots (Figure 3b). Deviation from linearity at the posterior end of the distribution was observed for each broodline which revealed loci with slopes that were greater than would be expected for normally distributed data (Figure 3b). Chi-square analysis was used to identify a significance cutoff ($\alpha = 0.01$) of $3.56\times$ the average deviation from linearity for each dataset (Figure 3c). A total of 169 loci were identified as significant outliers using this method, and nine outlier loci were observed in common among two broodlines. No loci were identified as outliers in all three broodlines using this method (Data S4).

The program Lositan identified 179 loci as being significant candidates for directional selection (Figure 4). Of the 179 significant loci, six were identified in two broodlines (Coho_64855-71, Coho_87190-34, Coho_69193-61, Coho_27527-68, Coho_68125-68, Coho_115799-69), and a single locus (Coho_92766-45) was identified as an outlier in all three broodlines (Data S4).

Results from Bayescan identified a total of 14 loci as being under selection (Figure 5), with two loci (Coho_64855-71 and Coho_87190-34) identified in two broodlines (Data S4). No loci were identified as outliers in all three broodlines. All but two loci identified using this analysis method were also identified as significant with one or both of the other methods used to detect loci under selection (Figure 6a). Across outlier methods, a total of 25 loci were in common in two broodlines, and one marker in all three broodlines (Figure 6b).

Extended consensus sequences of 5,009 of the 5,392 RAD loci mapped to at least one location within the Okis_V1 genome assembly (Data S5 and S6). Sequences of the 328 candidate loci identified as being under selective pressure by at least one of the nine outlier analyses (three broodlines \times three analysis methods) were aligned to the *O. kisutch* (Okis_V1) reference genome. Of the 37 highly significant candidate loci (significant in at least two of the outlier tests), 31 mapped to at least one location within the *O. kisutch* genome. No significant clustering of the mapped loci was observed within specific regions of the genome (Figure 7). The three nearest genes within a window of 100,000 bases of the genome locations of the 31 highly significant RAD tags were identified and 25 of these loci returned 108 nearby genes. These nearby genes, their genomic location, and their distances from the marker were compiled into a table and are included as supplemental information (Data S3). The most significant marker mapping to a single location in the genome assembly was Coho_64855-71 which lies within intron 6 of the sodium channel and clathrin linker 1-like gene (four reported isoforms: XP_020324990.1–XP_020324993.1). This gene resides within an unanchored scaffold (MPKV01000936.1–54,120 bases) which contains no other RAD loci.

4 | DISCUSSION

This study provided an opportunity to evaluate neutral and adaptive genetic variation in Coho salmon throughout the process of colonization to a new habitat following reintroduction from source stocks. Genetic analyses revealed that reintroduction of Coho salmon in the Wenatchee River began with three genetically distinct broodlines. Through the first adult returns and the subsequent three generations, there was evidence of limited gene flow between these stocks. Throughout their natural range, Coho salmon typically have overlapping generation times as adults return to spawn primarily as three year

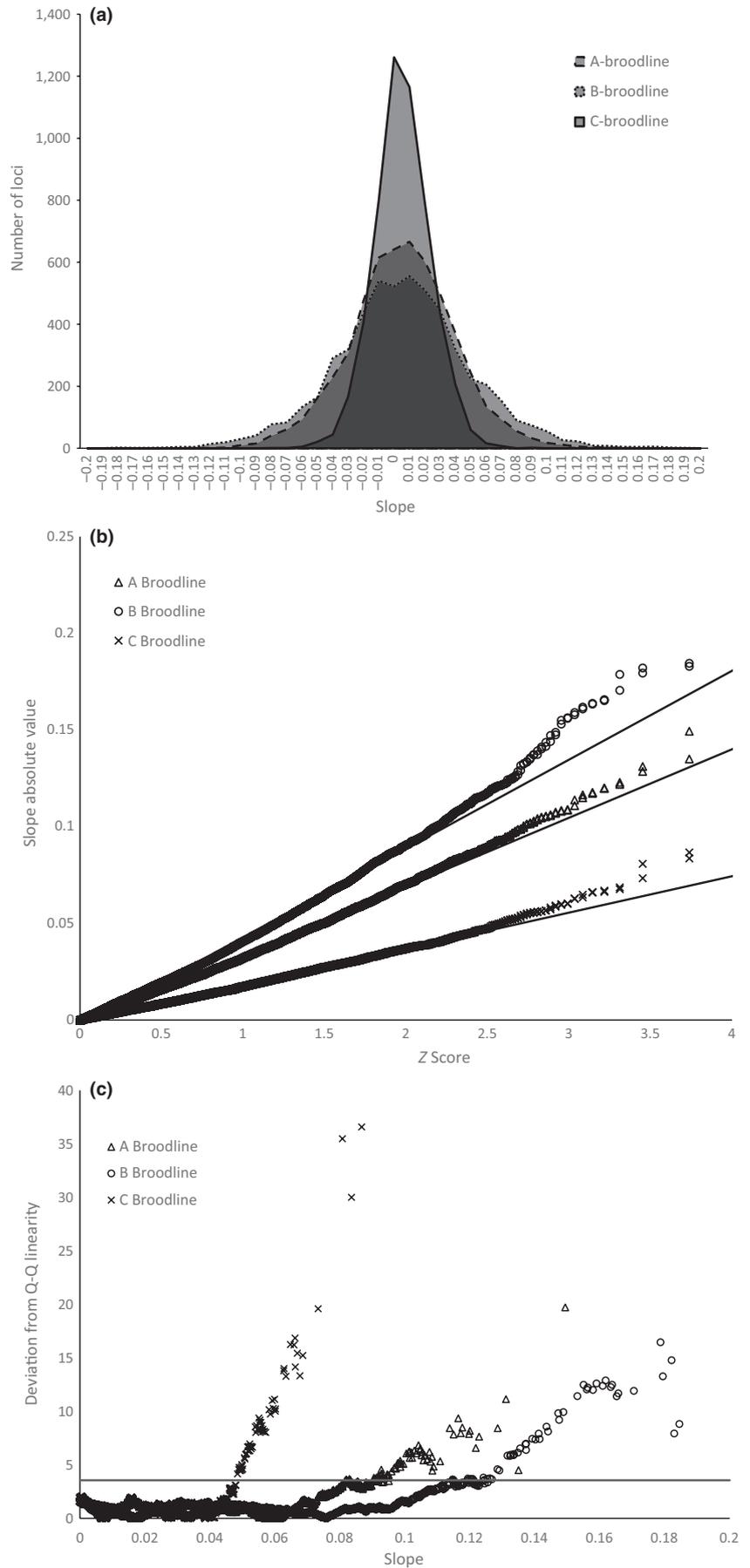


FIGURE 3 Histogram of slopes derived from linear regression of allele frequency differences at each of the 5,392 SNP loci over four generations of reintroduced fish (a). A Q-Q plot showing the expected linear distribution of slopes given a normal distribution for the number of data points observed and the deviation from linearity on the tails of the distribution for each of the three broodlines (b). For each of the observed data points, the deviation from the Q-Q plot linearity was calculated and plotted against the slope (c). The cutoff line in the figure was set to a significance threshold determined by chi-square test (threshold equivalent to a *p*-value of .01)

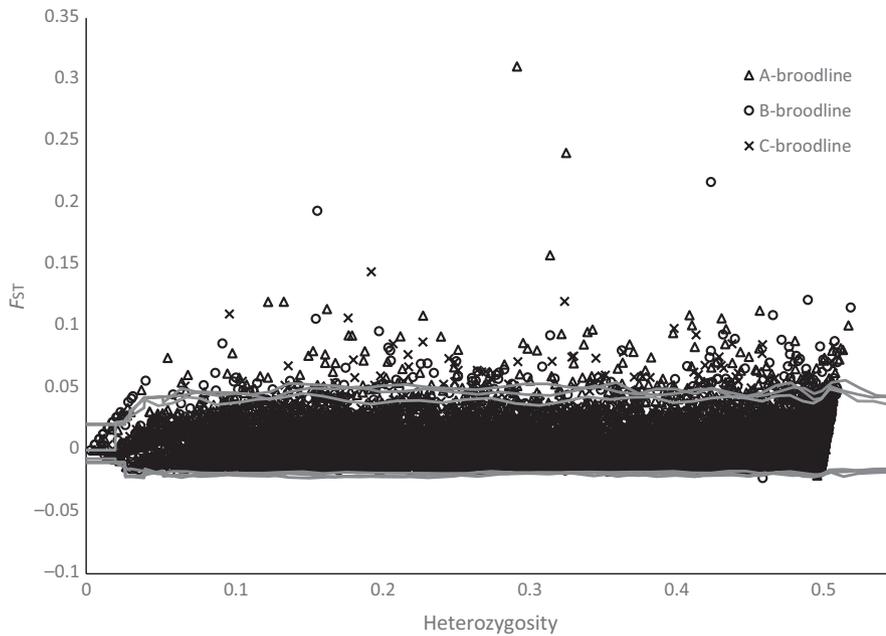


FIGURE 4 F_{ST} outlier analysis plot using Lositan. The gray lines represent the 95% confidence interval for putatively neutral loci for each of the three broodlines analyzed

olds, but two-year-old jacks (early maturing males) are also common and provide gene flow across generations (Beacham, Wetklo, Deng, & MacConnachie, 2011; Sandercock, 1991). However, hatchery programs of Coho salmon often limit the use of jacks within broodstock and distinct broodlines that follow three-year intervals can develop (Conrad, Gilbert-horvath, & Carlos, 2013; Smith et al., 2015). This phenomenon of distinct temporal structure between broodlines was apparent at the initiation of the reintroduction program in the Wenatchee River and reflects spawning practices from source hatcheries used for reintroduction. However, distinct temporal structure persisted among broodlines in the Wenatchee River over subsequent generations despite increased potential for gene flow from jacks through natural spawning and inclusion in the local broodstock program. Broodline

“B” was the most distinct from the other two broodlines and may have trait characteristics (e.g., differential run-timing) that maintains isolation among broodlines as has been observed in other studies of this species (Smith et al., 2015). However, run-timing and other trait characteristics were not available for individuals in this study.

The three distinct Coho salmon broodlines in this system provided the opportunity to examine biological replicates for signals of local adaptation over four generations and this paradigm was the basis for our analysis strategy for detecting SNP loci under selection following reintroduction. In total, 37 loci were identified as significant outliers in at least two of the nine tests and were considered the most likely candidate loci, leaving another 291 loci identified as significant in only a single test. While mapping these likely candidate

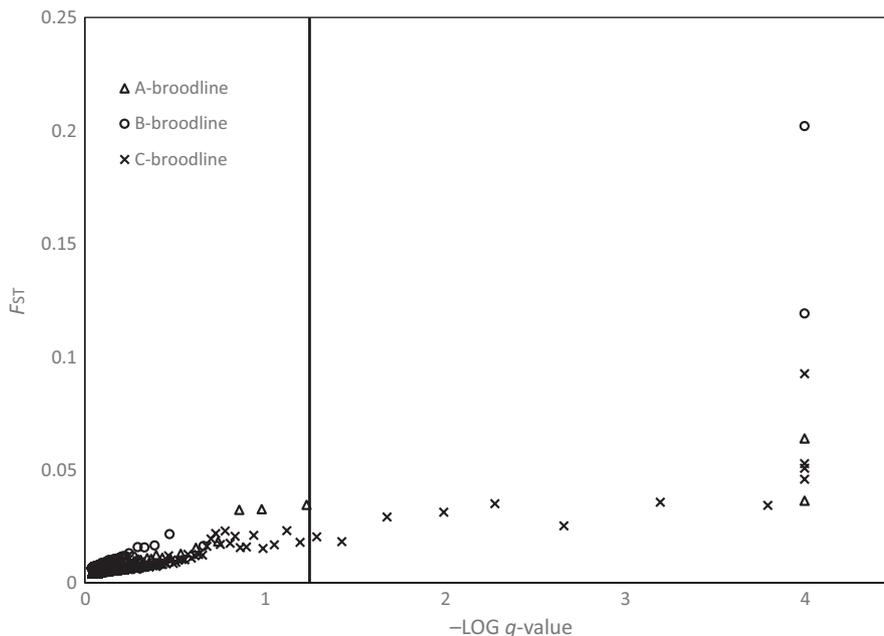


FIGURE 5 Bayescan analysis plot showing outlier loci for each of the three broodlines

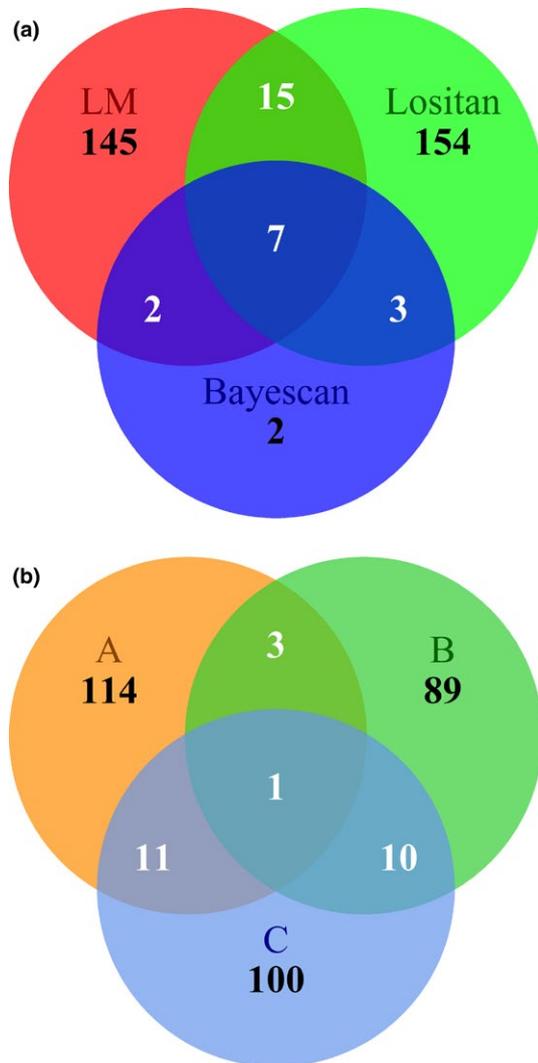


FIGURE 6 Venn diagram indicating loci identified as being under selection using the three analysis methods (a). Loci identified within and between the three broodlines are illustrated in the second Venn diagram (b)

loci revealed many genes in this list were involved in processes that are potentially beneficial for adaptation to a long-distance migration, the lack of a significant genomic “hotspot” and the low density of SNP markers used in this study was not sufficient to speculate which nearby genes are responsible for the adaptive signal. Similarly, RAD sequencing is a reduced representation approach for genome scans and therefore may miss regions of adaptive variation in species with rapid linkage decay and small linkage blocks (Catchen et al., 2017; Lowry et al., 2017; McKinney, Larson, Seeb, & Seeb, 2017; Tiffin & Ross-Ibarra, 2014). Therefore, we interpret these candidate loci cautiously until further studies with more extensive genomic coverage can more confidently map genetic adaptation. The most significant RAD locus within our dataset (Coho_64855-71) mapped within an intron of a sodium channel gene suggesting variations within this gene may be important for long-distance migration. However, for the reasons described above, this result will require further validation.

In this study, we utilized multiple tools developed for detection of adaptive loci between populations in order to reduce false positives in our dataset. Specifically, we required outliers to be based on replicates and overlapping results among methods in order to be considered as candidate loci (François, Martins, Caye, & Schoville, 2016; Lotterhos & Whitlock, 2015). This approach proved necessary as each outlier method that we implemented included caveats that could have led to elevated signals of false positives if interpreted alone. The linear method for detection of loci under selection relies on the accuracy of the allele frequency estimates at each generation in order to provide accurate data points. In some year classes, only a few individual samples remained after quality filtering, and allele frequency estimates would be expected to be less accurate. A reflection of this inaccuracy could manifest in the range of slopes seen in each of the broodlines with respect to their sample sizes. Although each broodline showed a range of slopes with a normal distribution centering on zero, a broad range of slopes was observed in the “A” and “B” broodlines when compared to the “C” broodline which contained the most samples per generation. Likewise, the “A” and “B” broodlines showed higher degrees of deviation from linearity in the Q-Q plots. This could also be an indication of higher degrees of genetic drift within the “A” and “B” broodlines manifesting as generalized genome-wide changes in allele frequency. However, the pattern of higher slope ranges with lower sample numbers indicates that this is more likely the result of inaccurate allele frequency estimates in year classes with small sample numbers. Although there is more statistical noise in the “A” and “B” broodline datasets, outlier loci still deviated significantly from expectations and many of the outlier loci were also identified as significant using the other analysis methods.

Similar to the linear model, Lositan is also sensitive to small sample sizes due to inaccuracies in allele frequencies used to calculate F_{ST} and heterozygosity. For the “B” broodline analysis, the 2007 samples ($N = 7$) were removed due to unrealistically high F_{ST} estimates when included. As F_{ST} is a measure of genetic distance between collections (usually populations) and our collections are different year classes of the same population, the average F_{ST} should be very low. This is indeed what is observed in the data with only a few outliers exceeding the 95% confidence interval for neutral loci. Nevertheless, disproportionately high F_{ST} may be reported at some loci due to small sample numbers and possibly low genotyping percentage within one or more collections resulting in false positives. The false-positive rate for this method may even exceed that of the linear method as inaccuracies at a single collection would have a greater effect on the analysis than would a single erroneous data point used in linear regression.

Unlike the other two methods used for detection of SNP loci under selection, Bayescan incorporates uncertainties in allele frequency estimation based on sample size and often has less false positives than other outlier tests (De Mita et al., 2013; Lotterhos & Whitlock, 2014; Narum & Hess, 2011). Not surprisingly, Bayescan also returned the fewest number of SNP loci under selection ($N = 16$). However, of these loci, only two had not also been identified using

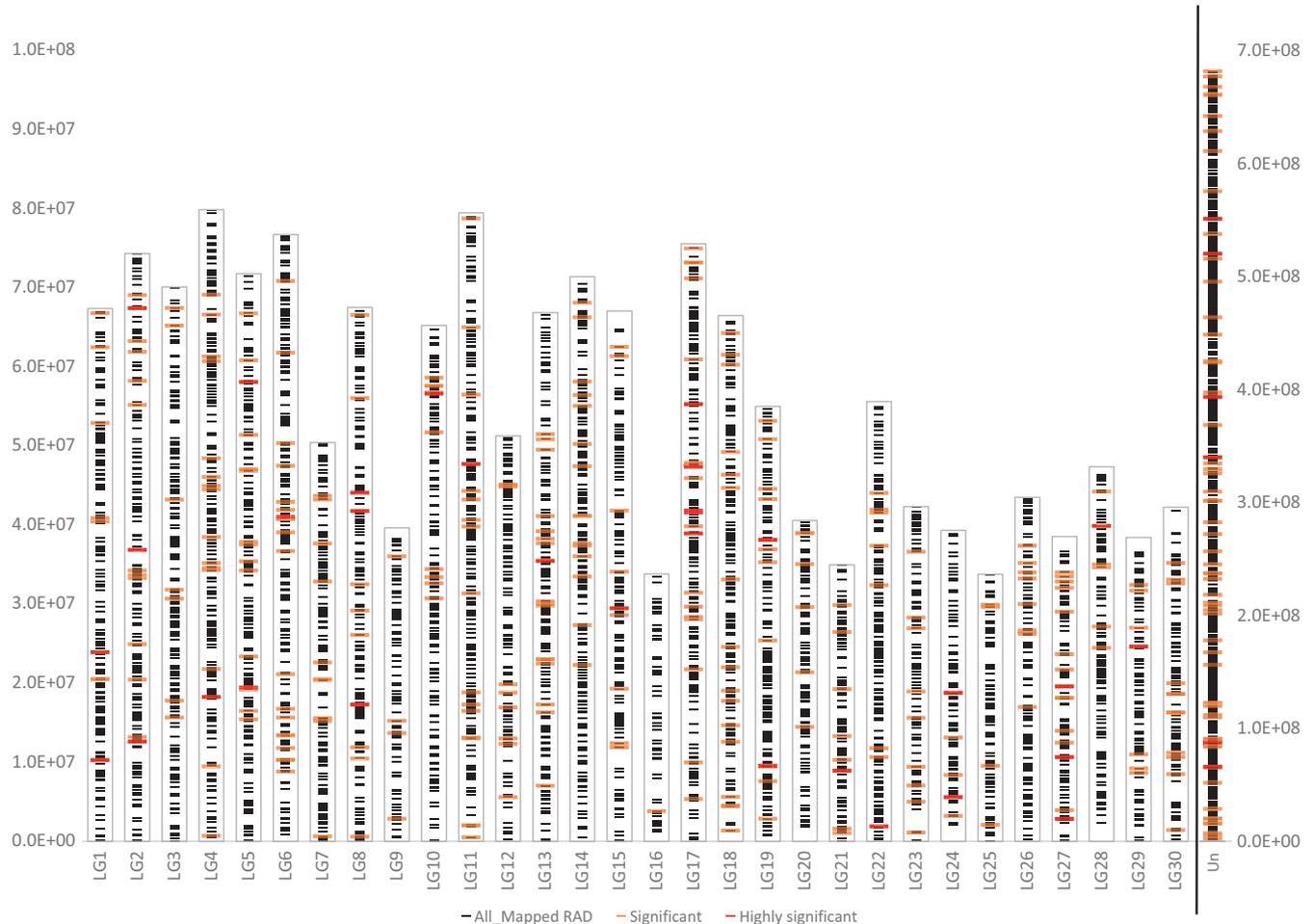


FIGURE 7 Illustrates the genome positions of 5,009 of the 5,392 RAD loci used in the study (small black bands). Significant and highly significant loci are indicated by the larger orange and red bands, respectively. The 21,216 unplaced scaffolds represent approximately 27% of the total genome size and are therefore presented on a secondary axis on the far right

the linear method or Lositan revealing agreement between the analysis methods.

In conclusion, our study identified patterns of neutral variation that reflected degrees of reproductive isolation among three broodlines that have been perpetuated from source stocks. While these broodlines have increased opportunity for gene flow in nature from overlapping age classes (Smith et al., 2015), it is uncertain how much time will be necessary to reduce or eliminate temporal variation in this system. Additionally, patterns of adaptive variation indicate many candidate loci with signals of divergent selection following reintroduction of Coho salmon to new spawning habitat. This genome scan for adaptive loci was independent of a-priori selection of candidate markers or association with phenotypic traits. Tracking genetic variation over four generations following reintroduction to a new habitat revealed a total of 328 unique loci that showed a significant signal, 26 were identified as likely outliers due to consistent results in at least two of the broodlines, and 27 were positive in two or more analysis methods, lending some credibility that the identified loci are not simply noise (i.e., false positives). Although the marker density was insufficient to confidently implicate specific adaptive genes, a single potentially adaptive gene is reported. These data indicate

that the genomes of these reintroduced Coho salmon are changing in response to their new spawning habitat and increased migration distance.

ACKNOWLEDGMENTS

We would like to thank Stephanie Harmon for preparing the RAD sequencing libraries, Amanda Matala for performing the Illumina sequencing runs, and the various Yakama Nations biologists who collected tissues and field data for the samples used in our study. This research was funded by a grant from Bonneville Power Administration (Funding project 1996-040-00).

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

NRC analyzed the sequencing data, conducted the statistical analyses, and wrote the manuscript. CK is the lead biologist overseeing

the reintroduction of Coho salmon in the middle Columbia River, he coordinated the collection of the samples, consulted on the study design, and edited the manuscript. KM is also a lead biologist with Yakama Nation fisheries implementing the Coho reintroduction strategy in the Wenatchee River and edited the manuscript. SRN designed the study, oversaw the analysis of the data, and edited the manuscript.

DATA AVAILABILITY

Raw sequencing data are available in the NCBI sra database (Accession number: BioProject Accession numbers PRJNA141569 and PRJNA414579 – in process will be made available upon acceptance). Supplemental data include RAD tags with SNP loci, SNP genotypes, annotations of genes near significant SNPs, outlier test results, extended consensus sequences from RAD loci used in the study, and sam formatted alignment file with genome locations for RAD loci mapped to the Okis_V1 genome.

ORCID

Nathan R. Campbell  <http://orcid.org/0000-0002-3061-5693>

Shawn R. Narum  <http://orcid.org/0000-0002-7470-7485>

REFERENCES

- Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, *9*, 323.
- Beacham, T. D., Wetklo, M., Deng, L., & MacConnachie, C. (2011). Coho Salmon population structure in North America determined from microsatellites. *Transactions of the American Fisheries Society*, *140*, 253–270.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*, 1165–1188.
- Bosch, W. J., Newsome, T. H., Dunnigan, J. L., Hubble, J. D., Neeley, D., Lind, D. T., ... Blodgett, J. W. (2007). Evaluating the feasibility of reestablishing a Coho Salmon population in the Yakima River, Washington. *North American Journal of Fisheries Management*, *27*, 198–214.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140.
- Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365.
- Conrad, J. L., Gilbert-horvath, E. A., & Carlos, J. (2013). Genetic and phenotypic effects on reproductive outcomes for captive-reared coho salmon, *Oncorhynchus kisutch*. *Aquaculture*, *404–405*, 95–104.
- Corlett, R. T. (2015). The Anthropocene concept in ecology and conservation. *Trends in Ecology & Evolution*, *30*, 36–41.
- Crossin, G. T., Hinch, S. G., Farrell, A. P., Higgs, D. A., Lotto, A. G., Oakes, J. D., & Healey, M. C. (2004). Energetics and morphology of sockeye salmon: Effects of upriver migratory distance and elevation. *Journal of Fish Biology*, *65*, 788–810.
- De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*, 1383–1399.
- Dobson, A., Lodge, D., Alder, J., Cumming, G. S., Keymer, J., McGlade, J., ... Xenopoulos, M. A. (2006). Habitat loss, trophic collapse, and the decline of ecosystem services. *Ecology*, *87*, 1915–1924.
- Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., ... Wardle, D. A. (2011). Trophic downgrading of planet earth. *333(6040)*, 301–306.
- Fast, D. E., Bosch, W. J., Johnston, M. V., Strom, C. R., Knudsen, C. M., & Fritts, A. L. (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.
- Foll, M., & Gaggiotti, O. (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., Murdoch, A. R., Hughes, M. S., Seamons, T. R., & LaHood, E. S. (2016). Broodstock history strongly influences natural spawning success in hatchery steelhead (*Oncorhynchus mykiss*). *PLoS ONE*, *11*, 1–20.
- François, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*, 454–469.
- Galbreath, P. F., Bisbee, M. A. Jr, Dompier, D. W., Kamphaus, C. M., & Newsome, T. H. (2014). Extirpation and tribal reintroduction of coho salmon to the interior Columbia River basin. *Fisheries*, *39*, 77–87.
- Griffith, B., Scott, J. M., Carpenter, J. W., & Reed, C. (1989). Translocation as a species conservation tool: Status and strategy. *Science (Washington)*, *245*, 477–480.
- Gustafson, R. G., Waples, R. S., Myers, J. M., Weitkamp, L. A., Bryant, G. J., Johnson, O. W., & Hard, J. J. (2007). Pacific salmon extinctions: Quantifying lost and remaining diversity. *Conservation Biology*, *21*, 1009–1020.
- Haig, S. M., Ballou, J. D., & Derrickson, S. R. (2014). Management options for preserving genetic diversity: Of Guam reintroduction rails to the wild. *Conservation Biology*, *4*, 290–300.
- Hecht, B. C., Campbell, N. R., Holecek, D. E., & Narum, S. R. (2013). Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, *22*, 3061–3076.
- Hendry, A. P., Castric, V., Kinnison, M. T., & Quinn, T. P. (2004). The evolution of philopatry and dispersal: Homing versus straying in salmonids. In Hendry A. P., Stearns S. C. (Eds.), *Evolution illuminated: Salmon and their relatives* (pp. 52–91). Oxford: Oxford University Press.
- Hess, M. A., Rabe, C. D., Vogel, J. L., Stephenson, J. J., Nelson, D. D., & Narum, S. R. (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.
- Huang, X., & Madan, A. (1999). CAP 3: A DNA sequence assembly program. *Genome Research*, *9*, 868–877.
- Jamieson, I. G. (2011). Founder effects, inbreeding, and loss of genetic diversity in four avian reintroduction programs. *Conservation Biology*, *25*, 115–123.
- Johnstone, K. A., Lubieniecki, K. P., Koop, B. F., & Davidson, W. S. (2011). Expression of olfactory receptors in different life stages and life histories of wild Atlantic salmon (*Salmo salar*). *Molecular Ecology*, *20*, 4059–4069.
- Jonsson, B., & Jonsson, N. (2006). Life-history effects of migratory costs in anadromous brown trout. *Journal of Fish Biology*, *69*, 860–869.
- Kareiva, P., Marvier, M., & McClure, M. (2000). Recovery and management options for spring/summer Chinook Salmon in the Columbia River Basin. *Science*, *290(5493)*, 977–979.
- Keefer, M. L., & Caudill, C. C. (2014). Homing and straying by anadromous salmonids: A review of mechanisms and rates. *24(1)*, 333–368.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

- Lohmann, K. J., Putman, N. F., & Lohmann, C. M. F. (2008). Geomagnetic imprinting: A unifying hypothesis of long-distance natal homing in salmon and sea turtles. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 19096–19101.
- Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, 23(9), 2178–2192.
- Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24(5), 1031–1046.
- Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 17, 142–152.
- McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: Comment on Breaking RAD by Lowry et al. (2016). *Molecular Ecology Resources*, 17(3), 362–365.
- Miller, M. R., Brunelli, J. P., Wheeler, P. A., Liu, S., Rexroad, C. E., Palti, Y., ... Thorgaard, G. H. (2012). A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology*, 21, 237–249.
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.*, 17, 240–248.
- Murdoch, K. C., Prevalte, S., & Kamphaus, C. (2005). Mid-Columbia Coho reintroduction feasibility study: Monitoring and evaluation. Annual Report to the Bonneville Power Administration, Project 1996-040-00 Portland Oregon.
- Narum, S. R., & Hess, J. E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11, 184–194.
- Paquet, P. J., Flagg, T., Appleby, J., Barr, L., Blankenship, D., Campton, M., ... Smith, S. (2011). Hatcheries, conservation, and sustainable fisheries—achieving multiple goals: Results of the hatchery scientific review group's Columbia River Basin review. *Fisheries*, 36(11), 547–561.
- Peakall, R., & Smouse, P. E. (2006). genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6, 288–295.
- Quinn, T. P. (1993). A review of homing and straying of wild and hatchery produced salmon. *Fisheries Research*, 18(1–2), 29–44.
- Sandercock, F. K. (1991). Life history of coho salmon. In C. Groot, & L. Margolis (Eds.), *Pacific salmon life histories* (pp. 345–445). Vancouver, BC: University of British Columbia Press.
- Seddon, P. J., Armstrong, D. P., & Maloney, R. F. (2007). Developing the science of reintroduction biology. *Conservation Biology*, 21, 303–312.
- Smith, C. T., Baumsteiger, J., Ardren, W. R., Dettlaff, Y., Hawkins, D. K., & Van Doornik, D. M. (2015). Eliminating variation in age at spawning leads to genetic divergence within a single Salmon population. *Journal of Fish and Wildlife Management*, 6(1), 4–18.
- Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, 29, 673–680.
- Wallace, R. L., Clark, T. W., & Reading, R. P. (2002). *An interdisciplinary approach to endangered species recovery: Concepts, applications, cases*. Endangered Species Update Special Issue: School of Natural Resources and Environment, The University of Michigan, 19, 67–203.
- Williams, J. G. (2008). Mitigating the effects of high-head dams on the Columbia River, USA: Experience from the trenches. *Hydrobiologia*, 609(1), 241–251.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Campbell NR, Kamphaus C, Murdoch K, Narum SR. Patterns of genomic variation in Coho salmon following reintroduction to the interior Columbia River. *Ecol Evol.* 2017;7:10350–10360. <https://doi.org/10.1002/ece3.3492>