U.S. Fish and Wildlife Service

Genetic profile of North Hood Canal coho salmon using RAD sequencing

Final Report FY 2017/18

May 2018

By: Justin Bohling, Jennifer Von Bargen

Abernathy Fish Technology Center
Longview, Washington 98632

Peter Bahls

Northwest Watershed Institute
Port Townsend, WA 98368

Photo Credit: Peter Bahls

Genetic profile of North Hood Canal coho salmon using RAD sequencing

Final Report FY 2017/18

Submitted: 14 May 2018

Justin Bohling* and Jennifer Von Bargen

Abernathy Fish Technology Center
1440 Abernathy Creek Rd.
Longview, WA 98632
(360) 425-6072
*Email: justin_bohling@fws.gov

In cooperation with:

Peter Bahls

Northwest Watershed Institute
3407 Eddy Street
Port Townsend, WA 98368
(360) 385-6786

Background

The Hood Canal in western Washington supports both naturally spawning and hatchery stocks of coho salmon (*Oncorhynchus kisutch*). Since 2001, the natural coho salmon populations in Hood Canal fishery management areas (Figure 1) have been managed by Washington Department of Fish and Wildlife (WDFW) and the Jamestown, Elwha, Port Gamble and Skokomish Tribes (hereafter referred to as the "co-managers"). Annual natural coho salmon escapement estimates for Hood Canal (excluding streams draining to Port Gamble Bay (9A) and Quilcene/Dabob Bay (12A) management units) are produced using a formal assessment procedure defined to be the number of naturally spawning coho salmon for the entire region (PSC 2004; Larry Lestelle pers. comm.). The escapement objectives for coho salmon in 9A and 12A are for the hatchery programs only and there are no natural coho salmon escapement objectives for the streams that flow into management units Area 9A or 12A, which includes Tarboo Creek. Commercial salmon fisheries exist in the northern portion of Hood Canal (Area 12 and Area 12B [Figure 1]) for both coho and chum salmon (*O. keta*) (beginning in mid-September for coho and changing to chum management by mid-October). The coho salmon fishery in Area 12A targets adults produced by Quilcene National Fish Hatchery (NFH), which is located at the confluence of the Big Quilcene River and Penny Creek near Quilcene Bay, from August 21-October 10 (Figure 1). Starting in 2009, the co-managers closed the northern end of Tarboo-Dabob Bay as a conservation measure in order to reduce catching coho that might hold in the bay in late September before entering Tarboo Creek after stream flow rises later in the fall.

Previous research has found Hood Canal coho populations to be genetically distinct from those found in other regions in Washington (Smith et al. 2007; Van Doornik et al. 2007). Within the Hood Canal itself, Smith et al. (2007) detected weak population structure among coho collected from spawning tributaries using a microsatellite panel designed for stock identification across the coho range. Importantly, though, the broodstock from Quilcene NFH was genetically distinct from these populations and only coho from the nearby Little Quilcene River showed substantial hatchery influence. Using a single nucleotide polymorphism (SNP) panel, Paul (2015) found some structure among naturally produced coho but had limited ability to assign individuals to the tributary of origin (51% assignment rate). However, there was high power (89% assignment rate) to distinguish naturally produced coho from the Quilcene NFH stock with this panel.

The limited ability to assign coho to a tributary of origin that was observed in past studies might reflect a high rate of gene flow across the basin, which would create a lack of appreciable genetic structure. Alternatively, the resolution provided by microsatellite and SNP panels might be limited by the power of the genetic markers. New techniques that exploit high-throughput DNA sequencing allow for genotyping thousands of loci offer increased power over previous methods. Such data have provided increased resolution for detecting subtle patterns of population structure, as has been demonstrated with Chinook salmon in Alaska (Larson et al. 2014) and other species of fishes (Corander et al. 2013; Bradbury et al. 2015). Our goal was to use one such technique, Restriction-site Associated DNA sequencing (RADseq), to genotype samples from Hood Canal coho to resolve patterns of population structure. Specifically our objectives were to: 1. Assess where the Hood Canal coho population falls along the continuum ranging from basin-wide panmixia to complete genetic isolation of tributaries, and 2. examine the extent of introgression from hatchery stocks into naturally spawning populations. Addressing these questions provides informtational needs to the co-managers of the coho fishery and also helps assesses genetic risks associated with the coho salmon program at Quilcene NFH.

Methods

Our dataset included samples collected from six Hood Canal tributaries where coho salmon have been documented to spawn (Table 1). For most of these tributaries tissue samples were collected in-stream from either juveniles or out-migrating smolts. The exception was Tarboo Creek: from 2007-2016 scale samples were collected from returning spawning adults. Also included were samples from coho broodstock used for propagation at Quilcene NFH. Another eight adult coho samples collected from the north end Tarboo-Dabob Bay (hereafter Tarboo Bay) in 2012. These individuals were collected during a fishery harvest. One additional sample was collected from Camp Discovery Creek. Both the Tarboo Bay and Camp Discovery Creek samples were included to assess their origin (i.e. which spawning tributary or hatchery they may have originated from). Given that allele frequencies can shift over time and a sample of individuals from a single year may not be representative of a population (Waples & Teel 1990), we included samples from multiple years. This was especially pertinent for Quilcene NFH due to the historical exclusion of two year old adults (commonly called “jacks”) from the breeding population, a series of three genetically distinct broodlines have emerged that return to the hatchery on a three-year cycle (Smith et al. 2007, 2015). We included samples from all three

broodlines and multiple years per broodline to capture the genetic variation within this hatchery population. Samples from jacks were also included.

With RADseq analyses a major issue is ensuring adequate sequence coverage across individuals. Low coverage can bias locus identification and genotype calling (Fountain et al. 2016). Therefore, we were limited in the number of samples we could include in each individual RADseq library. To balance coverage needs and reduce project costs, we included a subset of available samples from each tributary or broodyear.

For several of the collections from tributaries in 2005 and 2006 and Quilcene NFH broodstocks from 2001 to 2010 we had genotypic data from 12 microsatellite loci (Smith et al. 2007; Van Doornik et al. 2007). To minimize the impact of the presence of siblings on the analysis we ran these genotypes through the program Colony (Jones & Wang 2010) to identify potential sibling groups. As a conservative precaution we only included one individual per potential family group in the RADseq libraries. In forming our subsets we attempted to select individuals that were genotyped using a full suite of microsatellite loci, assuming they had high quantities of DNA, and from across the time of a collection season to ensure we captured diversity in run timing.

In addition to our study, the scale samples collected from coho sampled in Tarboo Creek were sent to the WDFW to perform a scale pattern analysis. This was done to assess whether the fish from which each scale came was of hatchery or wild origin. We included samples from carcasses designated as hatchery and wild origin in our libraries in proportions equal to that observed across an entire collection year.

Samples were prepared for RADseq using the protocol developed by Baird et al. (2008). To summarize, genomic DNA from each sample was digested using the *SbfI* restriction enzyme, barcode adapters (P1) were then ligated to the cut sites. Product was pooled into eight libraries with 56 individuals each. Pooled product was then sheared using a Bio-Rupter (Diagenode) with nine cycles of 30 seconds shearing and 90 seconds resting. Fragments are then size selected for 300-600 base pairs using AMPure bead (Beckman-Coulter) size select. We then used the KAPA library prep kit (Roche) to ligate a second adapter (P2) to prepare the fragments for sequencing. The libraries were sent to the Midwest Fisheries Center Whitney Genetics Laboratory for paired-end sequencing on an Illumina NextSeq 500. We sequenced 100 cycles for the forward sequence and 50 sequences for the reverse sequence.

We first demultiplexed the resulting FASTQ files using the *process_radtags* module in Stacks 1.46 (Catchen et al. 2013) and then ran the *clone_filter* module to remove reads derived from PCR clones. Our procedure for processing the RADseq data followed the dDocent bioinformatics pipeline (Puritz et al. 2014). First, reads were quality filtered using Trimmomatic. Both forward and reverse reads were then aligned to a reference coho salmon genome (Rondeau et al. 2017) using the software BWA (Li & Durbin 2009). Variants were called from the alignments using the program FreeBayes (Garrison & Marth 2012). We used VCFtools (Danecek et al. 2011) to filter variants using several criteria. First, we removed all variants missing data for more than 50% of individuals and only retained those that were biallelic (i.e. only two alleles present) and had a quality score $\geq$30. After that initial filtering we further removed variants with a minor allele frequency <0.05 and retained individual genotypes when the sequence depth was three or higher. With the reduced set of variants, we removed individuals from the dataset if they were missing genotypes at more than 90% of the variants. The final filtering step involved retaining variants with less than 70% missing data and removing insertions and deletions.

For each tributary we estimated average heterozygosity (H) and nucleotide diversity (π) as metrics of genetic diversity for each natural spawning population. Heterozygosity reflects the proportion of loci that are heterozygous (i.e. two different alleles) average across individuals. Nucleotide diversity reflects the number of variants detected, scaled by the total length of the coho genome we covered in our sequencing.

Our study included samples collected across multiple years from these tributaries (Table 1). Random fluctuations in allele frequencies can create signals of population structure across years, especially for semelparous reproducers that may have discrete cohorts. Since some samples from tributaries were collected over a decade apart, we examined genetic differentiation across years within individual tributaries. We estimated G_{ST} (Nei & Chesser 1983) within each tributary individually by grouping individuals by the year they were collected. G_{ST} is a measure of genetic differentiation that ranges from zero to one with zero reflecting no differentiation and one complete fixation for different alleles. Then we pooled individuals collected across years to represent each tributary and then estimated G_{ST} between these tributaries. We compared G_{ST} values within these individual tributaries to those between them; all G_{ST} calculations were performed using the using the R package *assigner* (Gosselin et al. 2016).

We also calculated relative migration rates between tributaries based on G_{ST} using the R package *diveRsity* (Keenan et al. 2013; Sundqvist et al. 2016). This method uses measures of genetic differentiation and simulations of shared gene pools to estimate not only the magnitude of gene flow but also directional asymmetries of the flow in a network of populations. It is scaled from zero to one, with higher values reflecting higher genetic exchange. Due to potential for temporal shifts in allele frequencies, we estimated relative migration between tributaries for just 2015, which was the only year for which we had samples from all tributaries. For this analysis we used the samples collected from Tarboo Creek in 2013: these were returning adults that would have corresponded to the brood year that would have produced the out-migrating smolts sampled from the other tributaries in 2015.

Estimating genetic differentiation between tributaries relies on grouping individuals by collection location. True patterns of genetic structure may not follow these designations, however. Therefore, we conducted additional analyses to investigate structure across the entire collection of individuals. The first was a principal component analysis (PCA) based on allele frequencies. For this we included coho samples collected from tributaries and Tarboo Bay and performed the PCA using the *glPca* function in the R package *adegenet* (Jombart 2008). We also did this for just the 2015 samples (2013 in the case of Tarboo Creek). This was accompanied by a Bayesian admixture analysis that identifies distinct genetic clusters among a dataset and estimates membership probabilities to these clusters for each individual. Cluster identification was performed on an incremental basis: various values of K (number of potential clusters in the data) were tested and compared using the likelihood score across ten replicates. The admixture analysis was conducted using the program NGSadmix (Skotte et al. 2013). To run NGSadmix we estimated genotype probabilities using ANGSD (Korneliussen et al. 2014) based on our sequence alignment files produced by BWA.

To test for migration and introgression from the Quilcene NFH coho program, we first measured genetic differentiation (G_{ST}) among Quilcene coho broodstocks across years. Then we estimated G_{ST} between Quilcene and Tarboo Creek collections for every year for which we data from both sites (2007, 2008, 2009, 2010, 2014, 2015, 2016). We only compared coho collections from Quilcene and Tarboo Creek because they represented adult returns to both spawning locations. For these years both sites had similar sample sizes. Other tributaries had collections from fewer years with greater time gaps, reducing the power to make inferences over time. We

also perform a clustering analysis using NGSadmix that included both the collections from the spawning tributaries and Quilcene NFH.

Results

On average our eight libraries produced over 150 million paired-end reads. After demultiplexing and PCR clone removal we had an average of 1,449,448 (SD=881,693) reads per individual. The average number of reads, both forward and reverse, per individual that aligned to the reference coho genome was 1,063,540 (SD=713,553). Following the dDocent pipeline we identified ~8 million variants. Our filtering procedure reduced this number to 32,485 biallelic variants with a minor allele frequency ≥0.05 and having missing data in <30% of individuals. Filtering individuals with more than 90% missing genotypes removed 28 individuals from the dataset: most (19) were scale samples collected from Tarboo Creek but they were distributed across years and did not result in the loss of any particular year class from the overall dataset.

Among coho collected from the tributaries values of H_E and π were similar (Figure 2). Ignoring the individuals from Quilcene NFH, when individuals were grouped by tributary, global G_{ST} was ~0.005. For Big Beef Creek, Little Anderson Creek, Seabeck Creek, and Stavis Creek, the value of G_{ST} between sample years within a tributary was higher than the global estimate across tributaries (Figure 3). This suggests there is more genetic variation between cohorts within tributaries than across tributaries. Only Tarboo Creek produced a lower between-year G_{ST} than the global estimate.

The average relative migration between tributaries based on the 2015 samples was 0.785 with all pairwise values greater than 0.5. When tributaries were plotted in a network diagram weighted by migration values, the overall network showed high levels of gene flow between Big Beef and Seabeck Creeks towards the center of the graph (Figure 4). Coho from Duckabush River displayed the lowest levels of gene flow with other populations.

Regardless of whether samples were combined across years or only 2015 samples were considered, there was substantial overlap between samples collected from different tributaries in principal component space (Figures 5 and 6). No distinct clusters of individuals emerged that corresponded to tributaries. A few individuals in both PCA plots were highly distinct: these likely reflect Quilcene-origin individuals (see below). NGSadmix produced log-likelihood scores from K=1-10 displayed an almost linear increase with each successive value (Figure 7). Typically log-likelihood estimates increase with each value of K but support for particular

clustering patterns is reflected in the size of increase from log-likelihood from one value of *K* to the next. There were no clear disproportionally large increases observed with these data: the linear increase suggests little improvement to the model in adding more genetic clusters. That suggests little support for coho sampled from natural spawning tributaries forming multiple distinct genetic clusters. This was further reflected by the fact that for any value of *K* the clusters that were identified did not correspond to tributaries.

The global value of G_{ST} between years for coho collected at Quilcene NFH was 0.0458. Annual pairwise comparisons of G_{ST} between Quilcene and Tarboo Creek returning adults were all above this value except for 2007. This suggests genetic differentiation between Quilcene hatchery returns and wild-returning adults for any given year was greater than between cohorts propagated at Quilcene NFH. Furthermore, estimates of G_{ST} between collections from Tarboo Creek and Quilcene NFH were greater than between tributaries (Figure 3).

Results from NGSadmix incorporating data from both the Quilcene NFH and tributary samples suggested a complex relationship between the two, depending on the number of clusters assumed. When *K* was set to two, the two clusters that emerged in the data roughly corresponded to the Quilcene and tributary collections with a high degree of shared ancestry (Figure 8). With each successive value of *K* distinct broodlines emerged as clusters from the Quilcene samples. For *K*=3 and 4 many natural-origin individuals had ancestry assigned to these broodline clusters. However, at *K*=5 and 6 this signal disappeared and the individuals collected from tributaries formed two clusters. At these values of *K* only a few individuals captured in the tributaries had evidence of ancestry from these broodlines. Only one individual collected from Tarboo Creek that was determined to be of hatchery-origin based on scale pattern analysis actually clustered with one of the Quilcene broodlines using the RADseq data (Table 2). Most of the eight returning adults collected in Tarboo Bay clustered with the coho samples collected tributaries (Table 3). One individual had evidence of mixed ancestry with about 40% of its ancestry assigned to two Quilcene broodlines.

Discussion

Our results suggest that naturally reproducing coho in this region of the Hood Canal form a metapopulation with high levels of gene flow. The analyses we conducted provided little support for genetic substructuring among these spawning aggregations that correspond to Hood Canal tributaries. These findings corroborate the results of Smith et al. (2007) and Paul (2015)

that were based on different sets of markers. Researchers studying other species have observed increased power to detect population structure using RADseq data compared to microsatellite and assay-based SNP markers (Corander et al. 2013; Larson et al. 2014; Bradbury et al. 2015). However, similar findings across three different datasets suggest the lack of appreciable structure is not due to the limitations of a particular dataset but a consequence of high levels of intra-tributary gene flow. Conducting a variety of analyses (e.g. G_{ST}, PCA, clustering) with different properties that produced congruent findings also supports these claims.

Research in western North America has demonstrated that the extent of genetic differentiation among coho salmon populations varies with spatial scale. Studies examining coho salmon across large regional scales have found populations from major drainage basins to form distinct genetic groups (Small et al. 1998a; Van Doornik et al. 2007; Starks et al. 2015). Within basins, however, differentiation may be less pronounced. Small et al. (1998b) found coho from the Fraser River watershed in British Columbia formed two genetic groups they hypothesized reflect post-glacial patterns of colonization. However, within these two main groups spawning populations were not highly differentiated. Coho returning to watersheds along the southern Oregon Coast displayed weak genetic structure (Johnson & Banks 2008). Similar findings were found for coho returning to central Oregon coastal watersheds (Ford et al. 2004).

The geographic scale of our study of Hood Canal coho salmon is much smaller than these studies. Therefore, our observation of weak genetic structure and high gene flow meshes with findings in other systems. Typical spawning habitat for coho salmon consists of small streams (Behnke 2002) and, correspondingly, spawning aggregations are often small (Ford et al. 2004; Johnson & Banks 2008). At this small scale, persistence of spawning populations is likely maintained by dispersal, creating a metapopulation structure (Young 1999; Schtickzelle & Quinn 2007; Scribner et al. 2017). Just a few migrants per generation can facilitate genetic mixing and inhibit local differentiation if spawning populations are small (Wehrhahn & Powell 1987). Such findings are informative for management in this region. Dispersal and colonization has the potential to seed populations should appropriate habitat be available, suggesting habitat restoration could be a valuable tool for promoting coho viability. However, that requires dispersers to reach these habitats.

Similar to the findings of Smith et al. (2007) and Paul (2015), we found that the hatchery stocks propagated at Quilcene NFH are genetically distinct from naturally produced coho salmon

in Hood Canal. As has been observed in coho hatchery stocks (Smith et al. 2015), we found genetic divergence between broodlines from Quilcene. The levels of divergence exceeded those within tributaries, likely reflecting the historic exclusion of jacks that would have facilitated gene flow between broodlines. However, divergence between coho collected from Quilcene NFH and these tributaries still exceeded differentiation between hatchery broodlines. This suggests that propagation of coho salmon at Quilcene NFH has not led to the genetic swamping of natural spawning aggregations.

The results from the clustering analysis (i.e. NGSadmix) provide a more complicated view of introgression and ancestral relationships between the Quilcene NFH stocks and natural-origin coho. At lower levels of *K* there are high amounts of shared ancestry between these groups, especially between tributaries and the "late" returning Quilcene stock. Historical exclusion of jacks from the Quilcene broodstock resulted in three genetically distinct broodlines that have different return times ("early", "middle", and "late") and return in three year cycles (Smith et al. 2015). The "late" broodline retained a returning timing similar to natural coho populations in that area. As this study demonstrates, these groups appear to have retained genetic similarity as well (the "late" broodline is indicated by the green cluster for *K*=4-6 in Figure 8). However, from *K*=5 upwards this shared ancestry was not observed. This suggests that the clustering algorithm tried to fit patterns of genetic ancestry within the constraints of the number of clusters allowed in the model. In other words, at low levels of *K* the program attempted to assign genetic variation observed in the natural population to the highly distinct broodlines as they emerged as distinct clusters. By *K*=5 and 6 there were additional clusters available to adequately explain the genetic variation observed in natural-origin individuals. In fact, the two clusters that appear in the natural coho at *K=5* and 6 mirror the clustering results performed solely on the natural-origin individuals themselves at *K*=2 (results not shown).

We must emphasize caution in interpreting the subclustering within the natural-origin coho. Assignment of individuals to these two clusters did not correspond with geography or time. Clearly these two clusters are more similar to each than any Quilcene NFH broodline, suggesting it is not introgression. It could be dispersal from neighboring populations that are slightly more distinct or simply be an artifact of the clustering algorithm misappropriating ancestry among a large and diverse sample of individuals. We did not include samples from George Adams State Fish Hatchery (FH) located in the southern portion of Hood Canal. Two coho stocks are raised at

this hatchery. The hatchery raises eggs produced by Quilcene NFH that are then used to stock the Port Gamble tribal net pen rearing program in Port Gamble Bay, which is located at the northern entrance to the Hood Canal. Since they are derived from Quilcene stocks, introgression from the net pen-raised coho would leave a similar genomic signature and thus is unlikely to account for the subclustering. The second is a local-origin segregated stock. Only hatchery-origin coho (which are indicated by double-index tagging) are used in this broodstock; natural-origin individuals returning to the hatchery are killed (Angie Stefani, WDFW pers. comm.). There is a chance this stock could be a source of introgression and this clustering pattern. Smith et al. (2007) found that coho from this hatchery were moderately distinct from natural-origin coho and clustered with the Quilcene NFH stocks. Paul (2015), however, suggested coho from George Adams FH were more similar to natural populations, especially Tarboo Creek. Regardless, considering that this is a segregated stock receiving no natural inputs, we would expect coho propagated at George Adams FH to form distinct genetic clusters similar to what we observed with the Quilcene NFH stocks. However, as our analyses of the natural-origin coho revealed, we did not observe this pattern. Furthermore, the two sub-clusters observed in natural-origin coho at higher values of *K* did not correspond to the WDFW assessment of hatchery origin. The most probable explanation is that these clusters reflect substructure within the natural population. The typical three-year age structure mitigated by stochastic gene flow between two-year jacks displayed by coho salmon can produce broodlines that do not correspond to clear temporal or spatial patterns.

Based on this study, only a small number of individuals sampled in tributaries had evidence of ancestry from Quilcene NFH. Averaged across all individuals and all four broodlines (at K=6 the 2004 "early" stock emerged as a distinct cluster), only 6.4% of the natural population's ancestry were assigned to the Quilcene stocks. Only five individuals collected from tributaries had ancestry coefficients greater than 0.8 for any of the hatchery broodlines, suggesting they are hatchery strays. Surprisingly, four of these individuals were smolts sampled in 2015 from Little Anderson Creek. These were the same four individuals that were highly divergent from the natural-origin individuals based on the PCAs. The low incidence of adults of Quilcene NFH ancestry could be a result of specific management practices in some of these watersheds. On Big Beef Creek, WDFW operates a trap that captures returning adult coho. All adipose-clipped adults captured are euthanized and only unmarked individuals, which are

assumed to be of wild-origin, are passed upstream (Clayton Kinsel, WDFW pers. comm.). In 2013, which would have been the brood that produced the 2015 out-migrating smolts, 709 unclipped adult coho were passed into Big Beef Creek while 79 adipose fin clipped individuals were removed.

There could be multiple explanations for the presence of smolts with Quilcene NFH ancestry in Little Anderson Creek. Perhaps the simplest scenario is escapement of adults propogated at the hatchery into the watershed that then reproduced. However, data from WDFW suggests an alternative (Clayton Kinsel, WDFW pers. comm.). During 2013 and 2016 smolt surveys in Big Beef, Little Anderson, Stavis, and Seabeck Creeks captured hatchery-origin smolts with clipped adipose fins. Coded-wire tag analysis revealed these smolts were released from the Port Gamble net pen program. In 2013 there were large numbers of adipose-clipped smolts observed in these surveys (Big Beef=965, Little Anderson=220, Stavis=670, Seabeck=165), which also corresponded to an early release date from the net pens due to an algal bloom. No adipose-clipped smolts were observed in 2015 but it is possible unmarked smolts could have entered Little Anderson Creek and been sampled by WDFW.

Interestingly, Tarboo Creek, despite its geographic proximity to Quilcene NFH, did not appear to receive any greater inputs of coho with Quilcene ancestry than other tributaries. Ten coho sampled from Tarboo Creek that were included in our RADseq libraries were assigned as hatchery origin by the WDFW scale analysis. Only one of those individuals assigned to a Quilcene broodline using the genetic data; one other appeared to be the progeny of Quilcene and natural parents. The remainder clustered with other natural-origin individuals. This was unexpected and suggests conflict between the two methods in terms of identifying hatchery origin coho salmon. However, it is important to note that Quilcene NFH operates as an integrated fish hatchery in which natural-origin coho with adipose fins are incorporated into broodstock. Also, once hatchery broodstock goals are met returning coho, regardless of natural or hatchery-origin, are passed above the hatchery weir to spawning naturally in the Big Quilcene River. Results from Smith et al. (2007) and Paul (2015) suggested that natural coho collected from Little Quilcene River resembled the Quilcene stocks genetically, likely due to straying from the nearby hatchery. This means it is entirely feasible for individuals that were produced in a natural setting to have been the offspring of parents derived from the Quilcene stocks. Among the coho salmon from Tarboo Creek determined to be of natural-origin by the WDFW scale analysis, none

had strong assignment to the Quilcene broodstocks (ancestry coefficients >0.8). However, three had ancestry coefficients greater than 0.4 and another 11 had values greater than 0.1. These results suggest there is low-level introgression of Quilcene broodstocks into the natural coho population but it is minimal and does not appear to have resulted in the replacement of natural production with Quilcene-origin individuals.

Samples from eight returning adult coho salmon collected from Tarboo Bay were included in this study. Given their proximity to Tarboo Creek, one of our questions was whether these individuals originated from this spawning location. However, given the lack of appreciable genetic structure observed among tributaries, it would be difficult to assess origin using genetic assignment tests. Paul (2015) found self-assignment rates for coho from tributaries in this area to typically be less than 50%. Smith et al. (2007) also reported an overall self-assignment rate of around 41%. Although these levels are greater than expected by random chance, it still suggests that over half the assignments for a given tributary were incorrect if those individuals did indeed originate from them. Considering that both studies sampled age 0 or 1 coho from these tributaries it is highly likely those individuals originated from those watersheds. Our results do suggest that these adult coho collected from Tarboo Bay originated from the natural population and were not derived from the Quilcene broodstocks.

There is one final consideration that must be given to the genetic makeup of Hood Canal coho. A status review of coho salmon conducted by the National Oceanic and Atmospheric Administration (NOAA) noted that the Hood Canal region, especially hatcheries, had received significant numbers of transfers mainly from hatchery coho stocks originating from the Puget Sound/Hood Canal region and outside basins along the West Coast of the United States (Weitkamp et al. 1995). However, subsequent genetic research has demonstrated that Hood Canal coho are genetically distinct compared to other West Coast populations, suggesting transfers did not lead to a total genetic replacement of native population (Smith et al. 2007; Van Doornik et al. 2007). Smith et al. (2007) included coho from several Puget Sound hatcheries in their analysis and although most coho Hood Canal tributaries formed a cluster distinct from the Puget Sound hatcheries, some clustered with the out-of-basin stocks. Whether this reflects out-of-basin introgression or simply low natural divergence between Hood Canal and Puget Sound coho is difficult to determine. Quilcene NFH broodstocks are more similar genetically to naturally-produced coho from the Hood Canal than Puget Sound stocks (Smith et al. 2007, 2015;

Van Doornik et al. 2007). This is expected given that the Quilcene NFH broodstock was founded from local Hood Canal populations and suggests there has been genetic continuity within this section of the basin. Although introgression from these transfers may still be present in the genomes of Hood Canal coho, we do not anticipate it is playing a major role in the current genetic structure of these populations. Adding coho from Puget Sound and other western Washington populations to our RADseq dataset would provide the resolution necessary to discern whether these transfers have had a lasting genomic impact.

Acknowledgements

Bringing together such a large collection of samples was made possible by the efforts of many researchers at various agencies and organizations. Many of the earlier samples used for this study were collected by the US Fish and Wildlife Service Abernathy Fish Technology Center and Quilcene NFH. More recent samples from Quilcene NFH were collected by Dan Magnuson, assistant manager, and we thank him for his annual collection of broodstock samples. Juvenile samples collected from wild tributaries in 2015 were collected by the Washington Department of Fish and Wildlife and we thank Todd Seamons for allowing us to use these samples. We thank the various co-managers of this fishery and Ann Gannam at AFTC for providing comments on draft versions of this report. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

Data and sampling management

Tissue samples and purified extracted DNA for these samples are archived at AFTC and can be provided to other researchers upon request. Raw FASTQ files and filtered VCF files are stored at AFTC and can be provided if requested. A copy of the report is archived in the AFTC database and can be made available to the public upon request.

Suggested citation

Bohling, J.H., J. Von Bargen, P. Bahls. 2018. Genetic profile of Hood Canal coho salmon using RAD sequencing. AFTC Final Report.

References

Baird et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376.

Bradbury, I. R., L. C. Hamilton, B. Dempson, M. J. Robertson, V. Bourret, L. Bernatchez, and E.

Verspoor. 2015. Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. Molecular Ecology 24:5130–5144.

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.

Corander, J., K. K. Majander, L. Cheng, and J. Merilä. 2013. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. Molecular Ecology 22:2931–2940.

Danecek, P. et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Ford, M. J., D. Teel, D. M. Van Doornik, D. Kuligowski, and P. W. Lawson. 2004. Genetic population structure of central Oregon Coast coho salmon (*Oncorhynchus kisutch*). Conservation Biology 5:797–812.

Fountain, E. D., J. N. Pauli, B. N. Reid, P. J. Palsbøll, and M. Z. Peery. 2016. Finding the right coverage: The impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. Molecular Ecology Resources 16:966–978.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907.

Gosselin, T., E. C. Anderson, and I. Bradbury. 2016. assigner: Assignment Analysis with GBS/RAD Data using R. R package version 0.4.1. https://github.com/thierrygosselin/assigner. doi: 10.5281/zenodo.51453.

Johnson, M. A., and M. A. Banks. 2008. Genetic structure, migration, and patterns of allelic richness among coho salmon (*Oncorhynchus kisutch*) populations of the Oregon coast. Canadian Journal of Fisheries and Aquatic Sciences 65:1274–1285.

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

Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10:551–555.

Keenan, K., P. McGinnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl. 2013. diveRsity : An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution 4:782–788.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356.

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.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Nei, M., and R. K. Chesser. 1983. Estimation of fixation indices and gene diversities. Annals of Human Genetics 47:253–259.

Pacific Salmon Commission (PSC). 2004. Pacific Salmon Treaty. Pacific Salmon Commission, Vancouver, BC.

Paul, A. 2015. Genetic differentiation of wild and hatchery-reared Hood Canal Coho Salmon: Evaluation of an initial Single Nucleotide Polymorphism (SNP) marker genetic baseline for stock identification. Final Report prepared by AP Bioresearch for Point No Point Treaty Council.

Puritz, J. B., C. M. Hollenbeck, and J. R. Gold. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431.

Schtickzelle, N., and T. P. Quinn. 2007. A metapopulation perspective for salmon and other anadromous fish. Fish and Fisheries 8:297–314.

Scribner, K. T., C. Soiseth, J. Mcguire, G. K. Sage, L. Thorsteinson, J. L. Nielsen, and E. Knudsen. 2017. Genetic assessment of the effects of streamscape succession on coho salmon *Oncorhynchus kisutch* colonization in recently deglaciated streams. Journal of Fish Biology 91:195–218.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195:693–702.

Small, M. P., R. E. Withler, and T. D. Beacham. 1998a. Population structure and stock identification of British Columbia coho salmon, *Oncorhynchus kisutch*, based on microsatellite DNA variation. Fishery Bulletin 96:843–858.

Small, M. P., T. D. Beacham, R. E. Withler, and R. J. Nelson. 1998b. Discriminating coho salmon (*Oncorhynchus kisutch*) populations within the Fraser River, British Columbia, using microsatellite DNA markers. Molecular Ecology 7:141-155.

Smith, C. T., J. Baumsteiger, M.Diggs, S. Walden, S., and W. Ardren. 2007. Genetic analyses of coho salmon populations in Hood Canal: How similar genetically are hatchery and natural populations? United States Fish and Wildlife Service, Abernathy Fish Technology Center Annual Report 6.

Smith, C. T. et al. 2015. Eliminating variation in age at spawning leads to genetic divergence within a single salmon population. Journal of Fish and Wildlife Management 6:4–18.

Starks, H. A., A. J. Clemento, and J. C. Garza. 2015. Discovery and characterization of single nucleotide polymorphisms in coho salmon, *Oncorhynchus kisutch*. Molecular Ecology Resources 11:50–66.

Sundqvist, L., K. Keenan, M. Zackrisson, P. Prodohl, and D. Kleinhans. 2016. Directional genetic differentiation and asymmetric migration. Ecology and Evolution 6:3461–3475.

Van Doornik, D. M., D. J. Teel, D. R. Kuligowski, C. A. Morgan, and E. Casillas. 2007. Genetic analyses provide insight into the early ocean stock distribution of juvenile coho salmon off the coasts of Washington and Oregon. North American Journal of Fisheries Management 27:220–237.

Waples, R. S., and D. J. Teel. 1990. Conservation genetics of Pacific salmon I. Temporal changes in allele frequency. Conservation Biology 4:144–156.

Wehrhahn, C. F., and R. Powell. 1987. Electrophoretic variation, regional differences, and gene flow in the coho salmon (*Oncorhynchus kisutch*) of southern British Colombia. Canadian Journal of Fisheries and Aquatic Sciences 44:822–831.

Weitkamp, L. A., T. C. Wainwright, G. J. Bryant, G. B. Milner, D. J. Teel, R. G. Kope, and R. S. Waples. 1995. Status Review of Coho Salmon from Washington, Oregon, and California. NOAA Technial Memorandum NMFS-NWFSC-24.

Young, K. A. 1999. Managing the decline of Pacific salmon: metapopulation theory and artificial recolonization as ecological mitigation. Canadian Journal of Fisheries and Aquatic Sciences 56:1700–1706.

Table 1: List of populations included in the RADseq libraries. Each column includes the year of collection and the associated sample size that was collected.

Tributary	2000	2001	2002	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	**Total**
Big Beef Creek					8	9									10		27
Camp Discovery Creek											1						1
Duckabush River															10		10
Little Anderson River						10									10		20
Quilcene NFH	12	18	18	18			18	18	15	17	2			18	18	18	190
Seabeck Creek						10									9		19
Stavis Creek						10									10		20
Tarboo Bay												8					8
Tarboo Creek							28	3	14	3	14	17	7	23	11	5	125

Table 2: Ancestry coefficients for coho salmon collected in the Tarboo Creek watershed determined to be of hatchery-origin by a scale analysis performed by WDFW. This is based on the NGSadmix results. “Middle”, “Late”, and “Early” refer to the three Quilcene NFH broodlines that formed distinct clusters. “Wild1” and “Wild2” refer to the two wild clusters identified.

Individual	Early	Middle	Late	Wild1	Wild2	Year collected	Collection location	Date collected
3370-012	0	0	0	0	1	2007	Tarboo Creek	12/10/2007
3370-369	0	0.572	0	0	0.428	2011	Lower Tarboo Creek	11/29/2011
3370-370	0	0	0	0.422	0.578	2011	Yarr Creek	11/29/2011
3370-374	0	0	0	0	1	2011	East Fork Tarboo Creek	12/7/2011
3370-379	0	0	0	1	0	2011	East Fork Tarboo Creek	12/16/2011
3370-656	0.051	0	0	0.001	0.948	2012	Yarr Creek	12/3/2012
3370-750	0	0	1	0	0	2014	Lower Tarboo Creek	10/22/2014
3370-751	0	0	0	1	0	2014	Upper Tarboo Creek	12/18/2014
3370-787	0	0	0	0.54	0.46	2016	East Fork Tarboo Creek	11/8/2016
3370-788	0	0	0	0	1	2016	Lower Tarboo Creek	11/18/2016

Table 3: Ancestry coefficients for coho salmon collected from Tarboo Bay in 2012 determined to be of hatchery-origin by a scale analysis performed by WDFW. This is based on the NGSadmix results. “Middle”, “Late”, and “Early” refer to the three Quilcene NFH broodlines that formed distinct clusters. “Wild1” and “Wild2” refer to the two wild clusters identified.

Individual	Early	Middle	Late	Wild1	Wild2	Collection location	Date Collected
3370-389	0.003	0	0	0.365	0.631	Tarboo Bay	Sept/22/2012
3370-390	0	0	0	0	1	Tarboo Bay	Sept/22/2012
3370-391	0	0	0	0.263	0.737	Tarboo Bay	Sept/22/2012
3370-392	0	0	0	0	1	Tarboo Bay	Sept/22/2012
3370-393	0	0	0	0	1	Tarboo Bay	Sept/22/2012
3370-394	0	0	0	0	1	Tarboo Bay	Sept/22/2012
3370-395	0.231	0	0.173	0.596	0	Tarboo Bay	Sept/22/2012
3370-396	0	0	0.006	0.339	0.655	Tarboo Bay	Sept/22/2012

Figure 1: Map of Hood Canal and the tributaries from which coho salmon samples were collected. The star indicates the location of Quilcene National Fish Hatchery. Fishery management units are designated by the black lines and the units names are in *italics*.

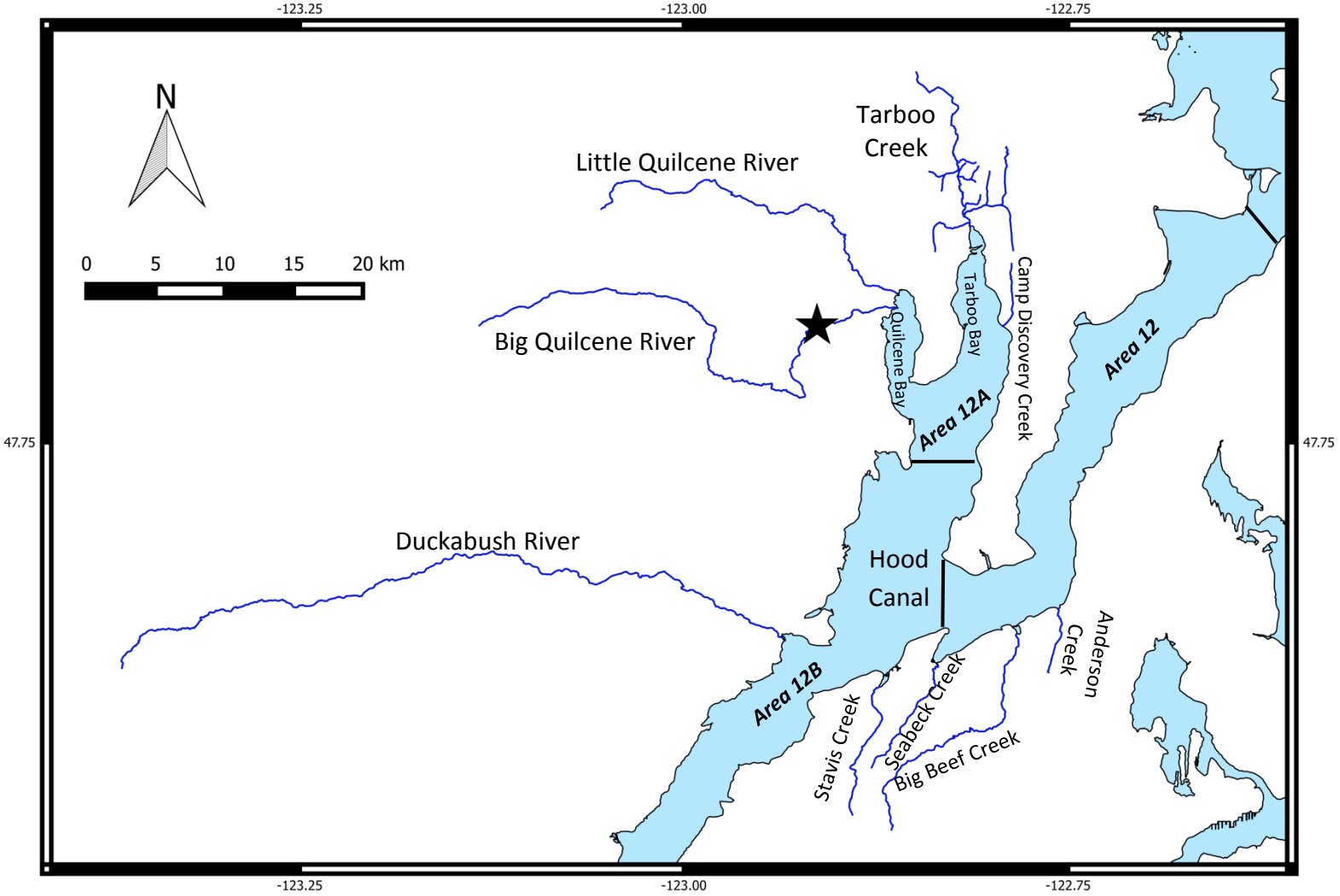


Figure 2: Plots of genetic diversity for coho collected from spawning tributaries in the Hood Canal. Values are based on combining samples across years. A. Barplot of expected heterozygosity (H_E). B. Violin plot of the distribution of nucleotide diversity (π) for individual coho sampled form each tributary. The interior boxplots display the maximum and minimum observed π along with the first and third interquartile. The white dot reflects the median value of π.

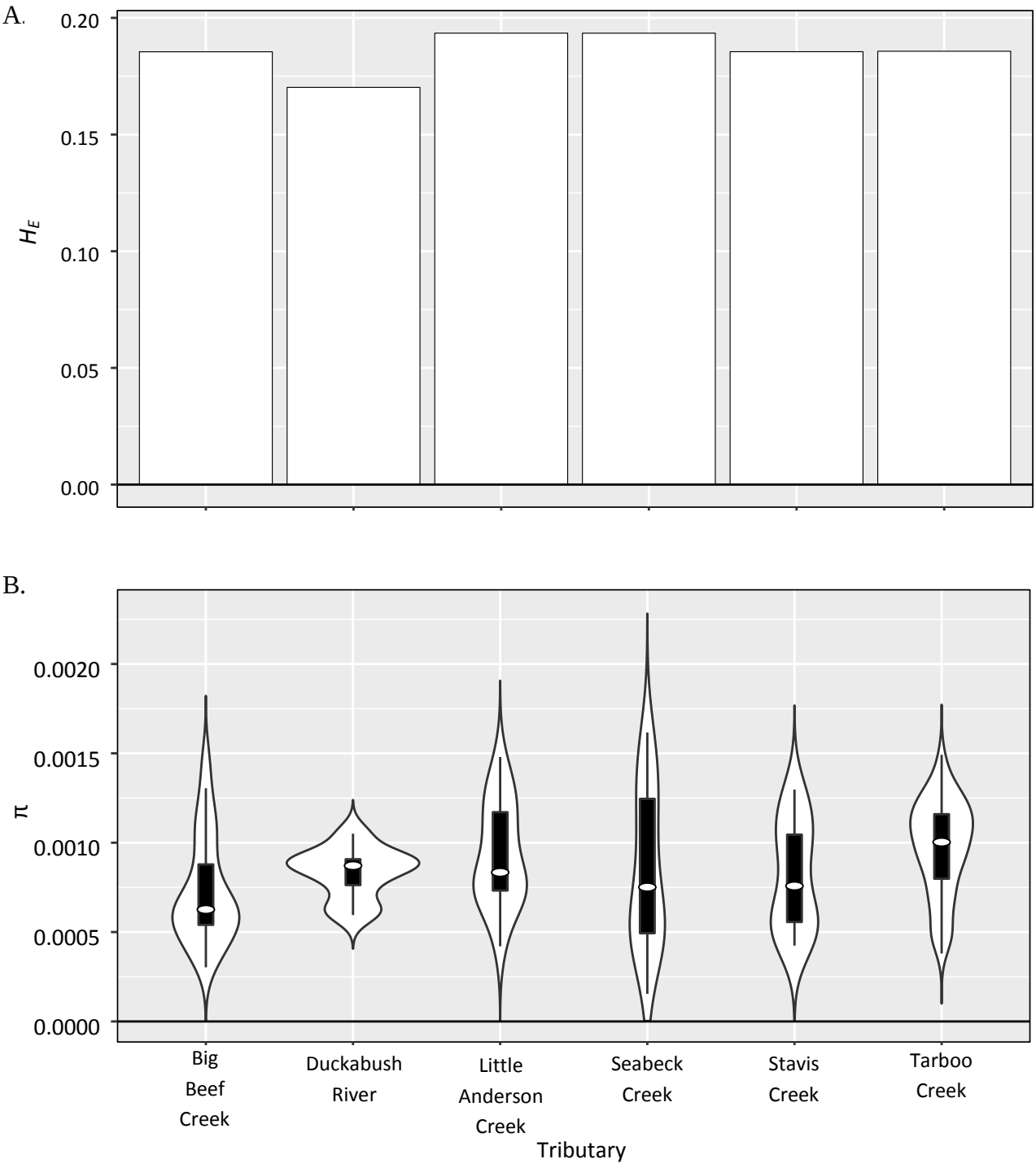


Figure 3: Plot of Nei's G_{ST} for each naturally spawning coho population sampled from the Hood Canal. Estimates for each population are based on differentiation between samples collected during different years. Points reflect the actual point value of G_{ST} and are bounded by 95% confidence intervals. The solid line reflects the global G_{ST} when all samples are grouped by tributary. The dotted lines are the 95% confidence intervals for this estimate.

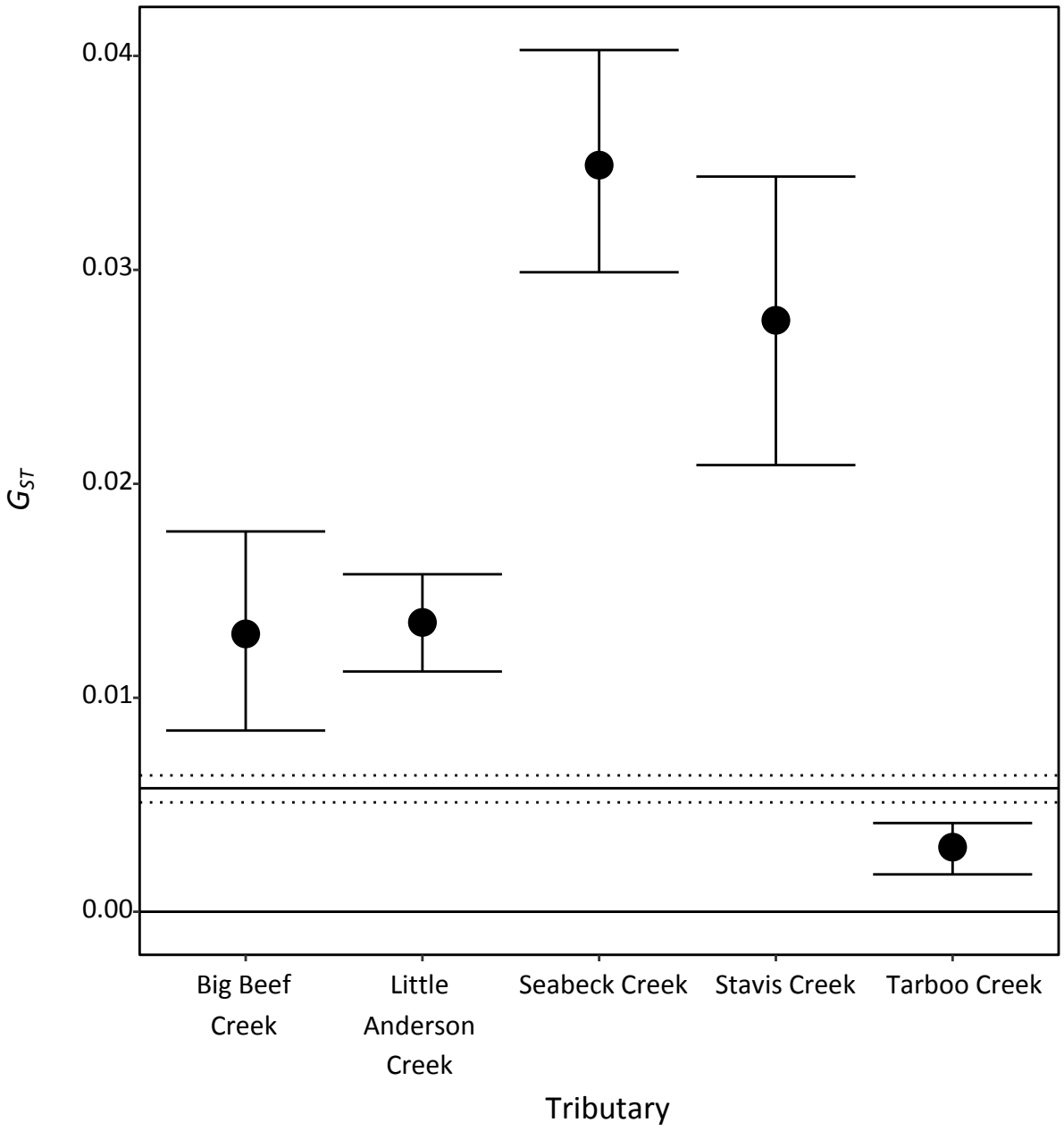


Figure 4: Network diagram of relative migration rates between natural spawning coho populations. Relative migration is based on G_{ST}. The network is edge-weighted so that highly connected nodes are towards the center of the graph. Edges, which reflect directional gene flow between population, are weighted so that more transparent edges reflect lower gene flow. The Tarboo Creek individuals were sampled in 2013: these were returning adults that would have corresponded to the brood year that would have produced the out-migrating smolts sampled from the other tributaries in 2015. Population codes: BIGB=Big Beef Creek, DUCK=Duckabush River, LAND=Little Anderson Creek, SEAB=Seabeck Creek, STAV=Stavis Creek, TARB=Tarboo Creek.

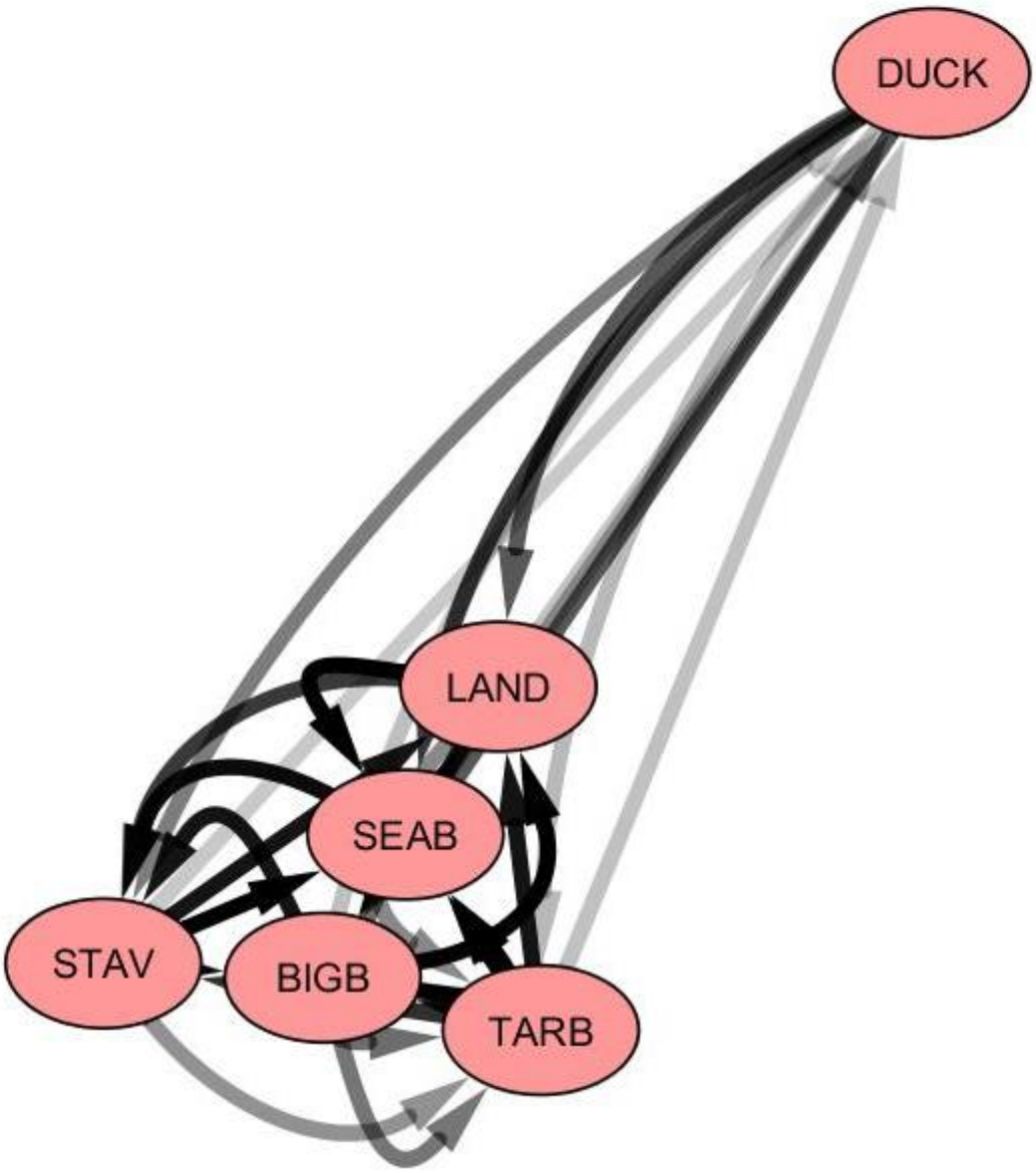


Figure 5: Principal component analysis (PCA) based on allele frequencies for coho sampled in natural spawning tributaries in the Hood Canal. Samples are grouped by the spawning tributary where they were captured and combined across years. The first two PCA axes are displayed.

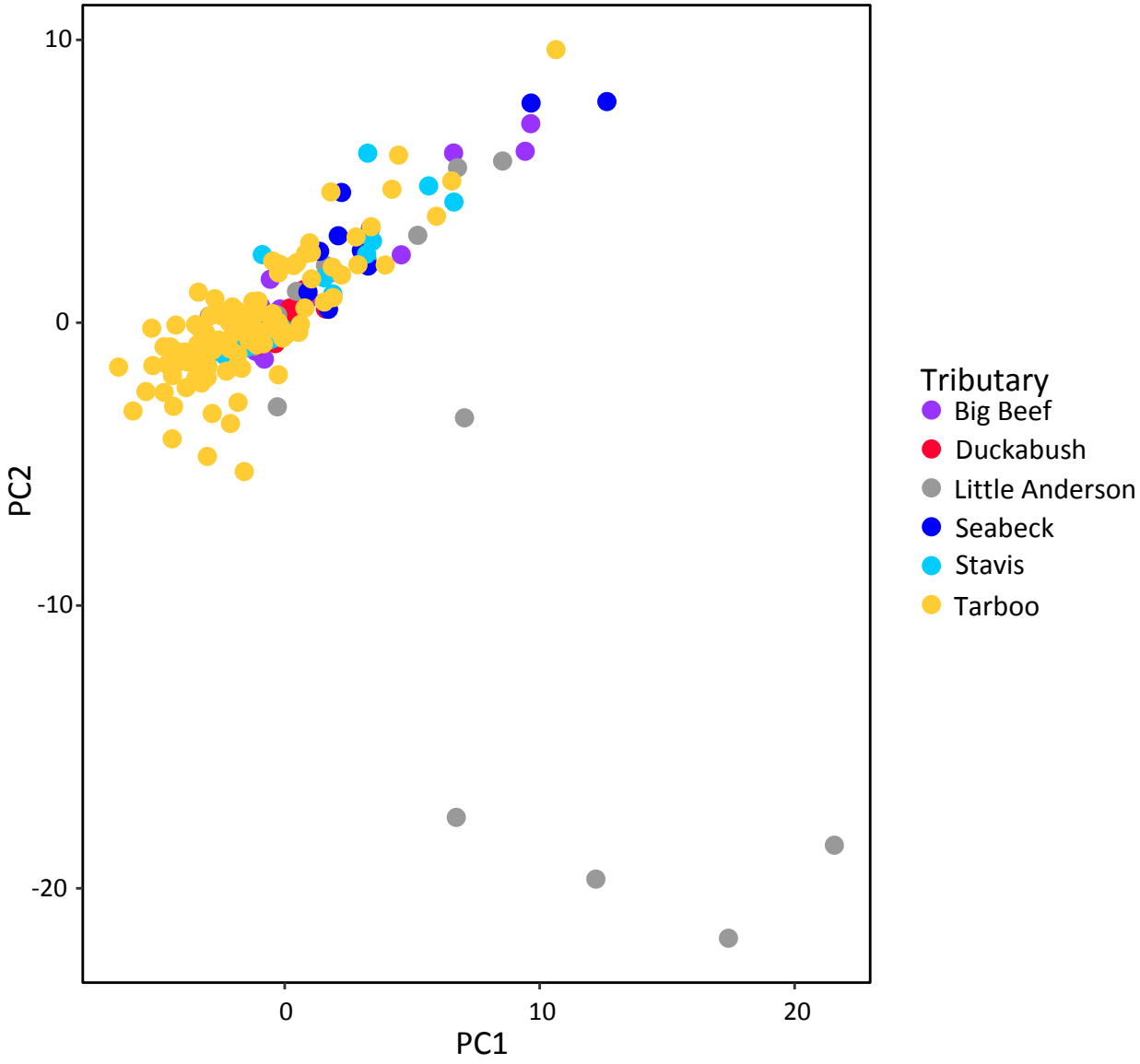


Figure 6: Principal component analysis (PCA) based on allele frequencies for coho sampled in natural spawning tributaries in the Hood Canal for only 2015. Samples are grouped by the spawning tributary that they were captured. The Tarboo Creek individuals were sampled in 2013: these were returning adults that would have corresponded to the brood year that would have produced the out-migrating smolts sampled from the other tributaries in 2015. The first two PCA axes are displayed.

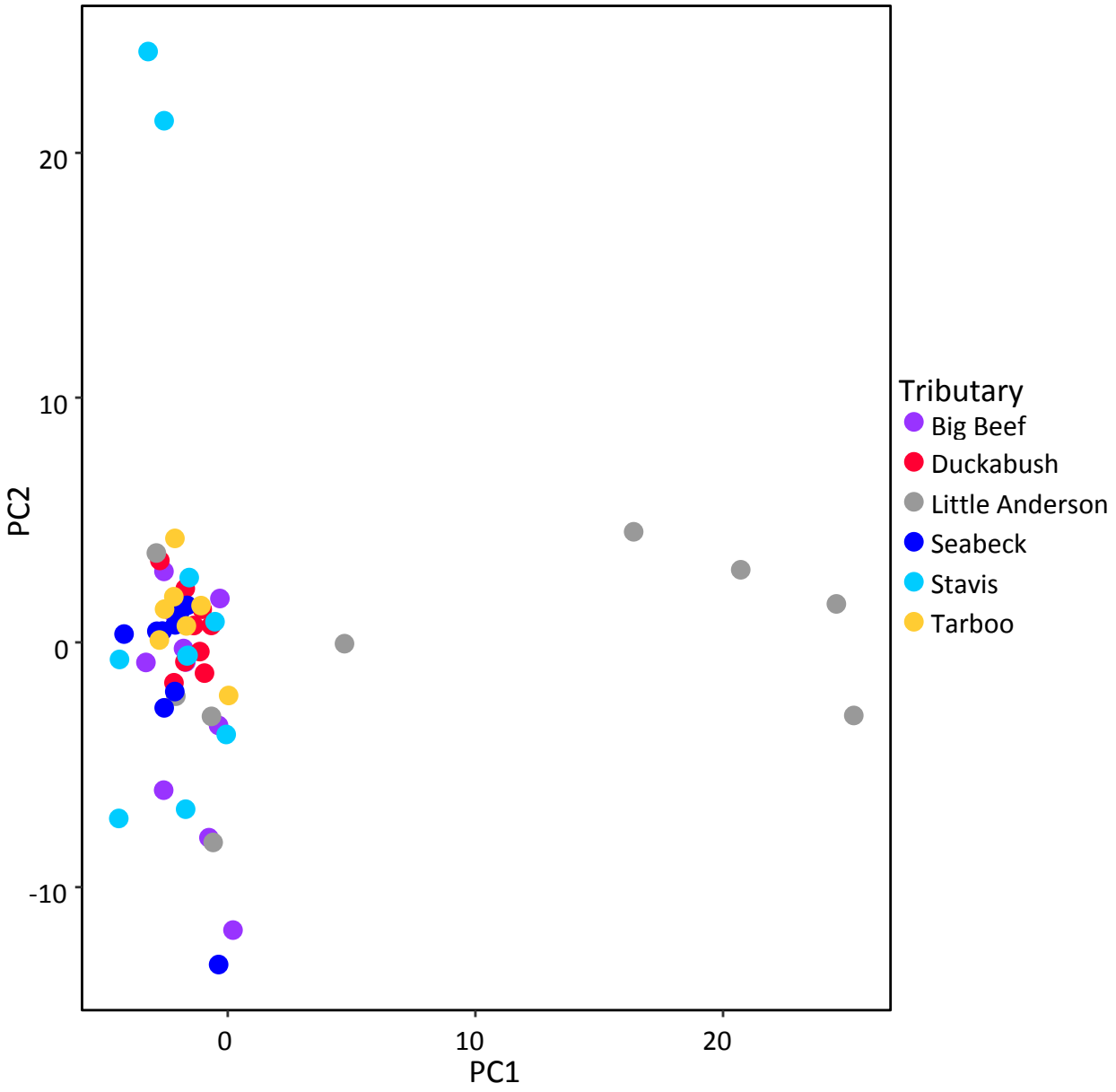


Figure 7: Plot of mean log-likelihood values produced for various levels of *K* with NGSadmix for coho sampled from natural spawning tributaries. Log-likelihood values were averaged across ten replicate runs for each *K* value.

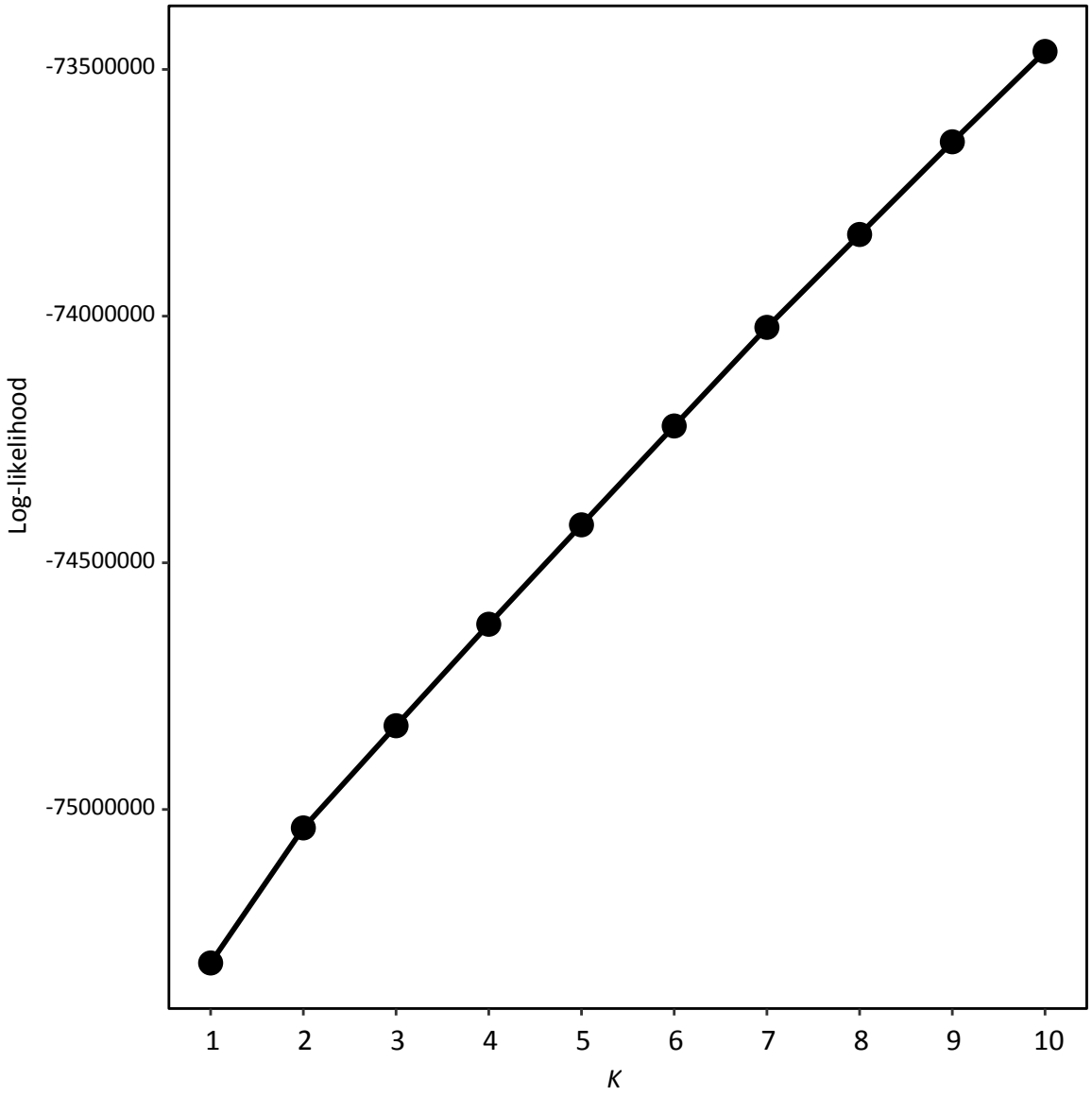
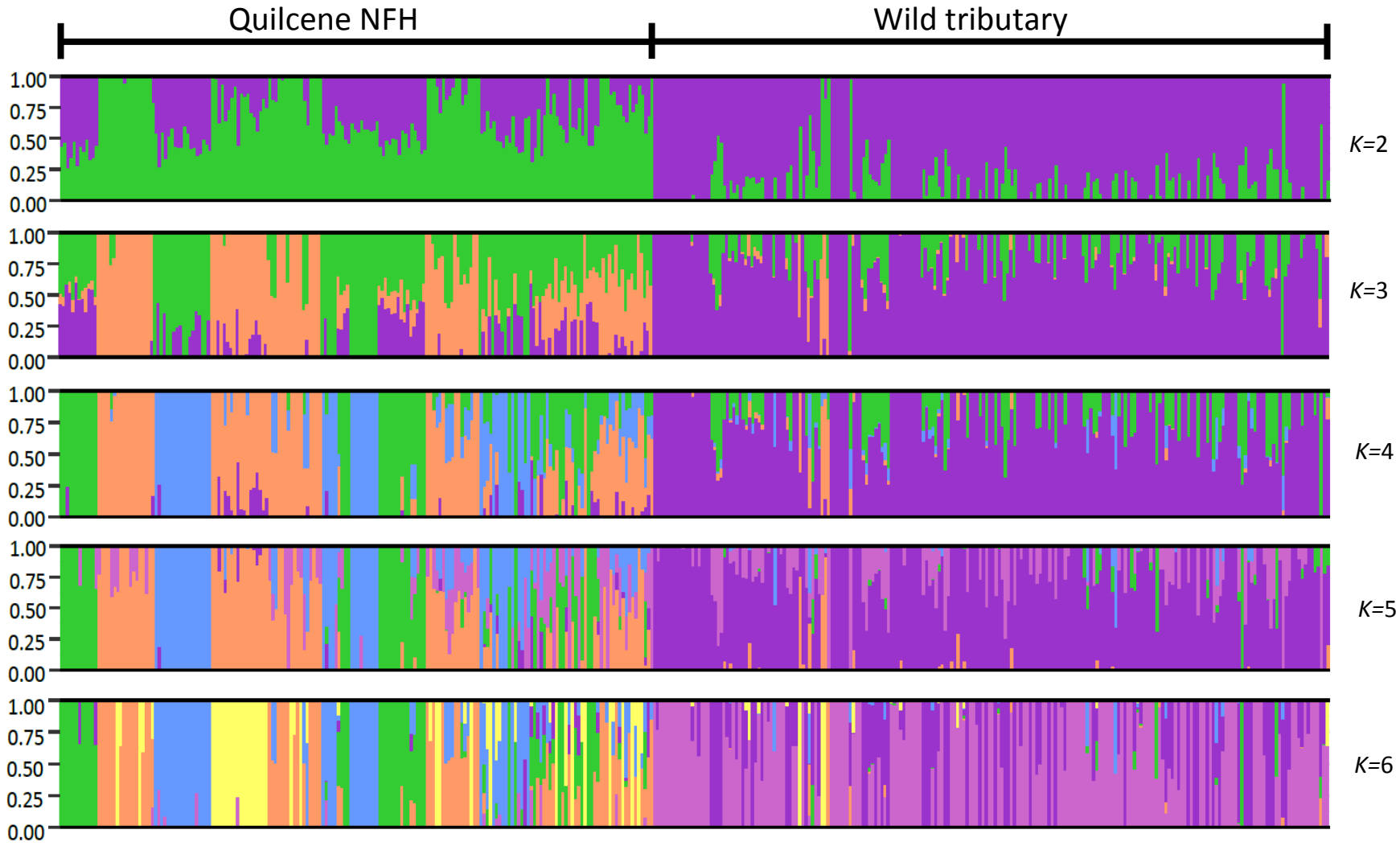


Figure 8: Barplot displaying admixture proportions derived for coho salmon captured at Quilcene NFH and natural spawning tributaries in Hood Canal. Each vertical bar represents an individual salmon and the colors represent distinct genetic clusters (*K*) identified by NGSadmix. Results for *K*=2-6 are displayed.



U.S. Fish & Wildlife Service
Abernathy Fish Technology Center
Longview, Washington, 98632
Phone: 360-425-6072
https://www.fws.gov/aftc/