U.S. Fish and Wildlife Service

Evaluation of genetic population structure and effective population size among populations of redband trout in the Deschutes River, OR

Final Report

June 2017

By Justin Bohling, Jennifer Von Bargen

Abernathy Fish Technology Center
Longview, Washington

Steven Starcevich

Oregon Department of Fish and Wildlife
Native Fish Investigations Program
Corvallis, Oregon

Evaluation of genetic population structure and effective population size among populations of redband trout in the Deschutes River Basin, OR

Draft report submitted

27 April 2017

Final report submitted

19 June 2017

Prepared by:

Justin Bohling and Jennifer Von Bargen

Abernathy Fish Technology Center
US Fish and Wildlife Service
1440 Abernathy Creek Rd.
Longview, WA 98626
360-425-6072 x337
Justin_bohling@fws.gov

In collaboration with:

Steve Starcevich

Oregon Department of Fish and Wildlife
Native Fish Investigations Program
28655 Highway 34
Corvallis, OR 97333
541-231-1806
steve.starcevich@oregonstate.edu

Summary

Redband trout are an iconic species in the Deschutes River basin of central Oregon; however, even with this status there has been little investigation of the genetic characteristics of populations occupying the middle section of the basin from Lake Billy Chinook to Wickiup Dam. Illuminating patterns of genetic diversity, population structure, and effective population size would provide managers with information valuable for managing this species. In a partnership between the US Fish and Wildlife Service's Abernathy Fish Technology Center and the Native Fish Investigations Program at the Oregon Department of Fish and Wildlife (ODFW), we developed the first comprehensive genetic assessment of redband trout from this section of the Deschutes basin. Over 1280 young of the year redband trout collected by ODFW field personnel, along with representatives from several hatchery stocks, were genotyped at a panel of 269 single nucleotide polymorphism loci. We observed high genotyping success rate with an average per locus genotyping rate of 94% (SD±9%). Major breaks in genetic population structure corresponded to significant barriers to fish movement, such as Big Falls and the North Canal Dam. However, there were cryptic patterns of population structure that did not correspond to these types of features. Several areas of the Deschutes exchange substantial numbers of migrants and likely function in a linear stepping-stone model. We estimated the effective number of breeders that produced the cohorts of redband trout we sampled. These estimates were highly variable across the study area. Genetic introgression from a hatchery stock of redband trout originating from the Upper Deschutes River was prevalent in the Fall River tributary, which is currently stocked with this strain. Introgression from this strain was also observed in the main stem Deschutes River in areas adjacent to the mouth of Fall River. Our project provides valuable insights into the genetic characteristics of redband trout from this portion of the Deschutes River, which can inform management and guide conservation activities.

Background

Redband trout (*Oncorhynchus mykiss gairdneri*) are of great interest to fisheries managers across the Intermountain West because they are popular sport fish and considered an icon of many western river systems. However, the species has been heavily impacted by human activity across its range (Schroeder and Hall 2007, Muhlfeld et al. 2015). The Deschutes River in central Oregon is a major tributary of the Columbia River that supports redband populations throughout much of the basin. This watershed has also been heavily impacted by environmental

degradation, impacting the distribution and abundance of redband trout. The Oregon Department of Fish and Wildlife (ODFW) Native Fish Investigations Program has surveyed the middle and upper Deschutes River along with several tributaries to assess the status of redband trout and other salmonids in this watershed. Focusing on the middle section of the Deschutes between Big Falls and Wickiup Dam, ODFW's objectives were to determine the distribution and habitat associations of young trout and utilize molecular methods to evaluate genetic characteristics of these fish. Specifically, ODFW was interested in assessing genetic population structure, effective population size, and genetic introgression from hatchery stocks.

In this middle segment of the Deschutes River basin and its tributaries, the population structure of redband trout has not been studied and management agencies and watershed groups are uncertain if the species should be treated as one or several local populations. Previous research documented genetic structure among specific local populations of redband within the larger Deschutes River system (Currens et al. 1990, Matala et al. 2008, Adams et al. 2015). However, genetic surveys covering large portions of the basin, including the middle portion above Lake Billy Chinook and below Wickiup Dam, are lacking. Also, hatchery-origin coastal rainbow trout have been historically stocked within the basin along with a strain of redband trout derived from Crane Prairie Reservoir in the upper Deschutes basin. Understanding patterns of genetic structure and introgression will allow ODFW to identify units at which to manage the species. Along with field-based estimates of abundance, ODFW has interest in using genetic-based metrics to monitor populations. Effective population size (N_E) and effective number of breeders (N_b) estimated using genetic markers can be used to assess population status.

The genetic samples obtained in this project were used for the following three objectives: 1. Evaluate population structure of redband trout in the middle and upper segments of the main stem Deschutes River, 2. Assess the extent of introgression from hatchery-origin redband and coastal rainbow trout in these populations, and 3. Estimate genetic N_b for identified populations.

Methods

Sampling

ODFW biologists collected juvenile redband trout via electrofishing surveys in lateral habitats of the Deschutes River, between Big Falls (the historical limit of anadromy for *O. mykiss*) and Wickiup Dam, and its major tributaries. For sampling purposes, ODFW divided the study area into 14 reaches (Figure 1). The boundaries between these reaches corresponded to

natural and anthropogenic barriers hypothesized to influence gene flow, tributaries, and confluences with major tributaries. The study area covered a total river and tributary length of 209 km. The Deschutes River tributaries surveyed were Tumalo Creek, Little Deschutes River, Spring River (from which no redband trout were captured) and Fall River. In this report these reaches and tributaries will be referred to as sampling groups.

Fin clips were collected by ODFW biologists and stored in 100% non-denatured ethanol before being sent to Abernathy Fish Technology Center (AFTC). ODFW was interested in estimating metrics from a single cohort of young-of-the-year (YOY) trout. Therefore, a length frequency analysis was conducted to determine appropriate size bins to classify trout into age groups. A total of 1480 redband tissue samples were collected; 1288 of these were classified as YOY fish. Another 35 tissue samples were provided from a hatchery stock of redband trout raised at Wizard Falls Fish Hatchery. This stock, often referred to as "cranebows", was established from native populations in Crane Prairie Reservoir in the upper Deschutes watershed. Currently cranebows are stocked in Fall River and were released in other portions of the upper Deschutes until 2015. We also included individuals from two common hatchery strains of rainbow trout: the Cape Cod strain (n=46) and the Oak Springs strain (n=48). Both of these strains are believed to have been cultured from populations of *O. mykiss* in California. Although these stocks are not currently stocked in the study area, they have been released in the past (E. Moberly, pers. comm.), so we included these samples to test for lingering introgression.

Library preparation and genotyping

DNA was extracted at AFTC. Extractions followed a modified protocol using Qiagen DNeasy Tissue Kits (Qiagen, Inc. Valencia, CA). To produce adequate DNA concentration for high-throughput sequencing, during the final DNA elution step we spun 100 µL of the AE buffer through the filter membrane twice. We used the Genotyping-in-Thousands (GT-seq) approach (Campbell et al. 2015) to genotype our samples using high-throughput sequencing technology. The protocol followed that of Campbell et al. (2015) and used a panel of 269 single nucleotide polymorphism (SNP) loci. These loci were developed for stock ID of Columbia River steelhead. To summarize the process, extracted DNA samples were first cleaned in an ExoSAP reaction (New England Biolabs, Ipswich, MA). Then each sample was amplified in a Polymerase Chain Reaction (PCR) containing the forward and reverse primers for all loci and Qiagen Plus MasterMix. In a new set of 96 well PCR plates, a unique i7 index primer was added to each plate

followed by an aliquot of the amplified PCR product. Each well of these plates then received one of 96 i5 index primers. The combination of the i7 and i5 primers creates a unique series of genetic barcodes to identify each individual in the library. We used SequelPrep Normalization Kits (ThermoFisher Scientific, Grand Island, NY) to normalize the PCR products. Samples from the same plate were then pooled together and subjected to a bead size selection procedure using AgencourtAMPure beads (Beckman Coulter Life Sciences, Indianapolis, IN). We quantified the amount of DNA product using a Kapa qPCR quantification kit (Kapa Biosystems, Wilmington, MA) using four different dilutions (1:1000, 1:2000, 1:4000, 1:8000). Based on the results of the DNA quantification, these per-plate pools were combined and normalized to a 5 nM concentration. The final pooled library was run on an Illumina NextSeq (Illumina, Inc., San Diego, CA) with a 100 cycle mid-output kit. The sequencing was performed at the Columbia River Inter-Tribal Fish Commission Hagerman Genetics Laboratory. Genotyping based on the sequence reads was performed using the scripts outlined in Campbell et al. (2015).

Testing for introgression

Prior to estimating population genetic parameters, we wanted to identify populations and individuals with evidence of introgression from hatchery stocks. First, we performed a correspondence analysis based on allele frequencies to identify wild redband populations that showed similarity with hatchery stocks. Our second analysis used the Bayesian clustering approach implemented by the program STRUCTURE (Pritchard et al. 2000, Falush et al. 2003) to identify admixed fish. This method simultaneously identifies genetic clusters among a group of individuals and probability of assignment to those clusters. We used the correlated allele frequency model, inferring alpha for each population, and allowed for admixture. STRUCTURE was run in parallel using the R 3.2 (R Core Team 2015) package *ParallelStructure* (Besnier and Glover 2013). Because STRUCTURE can be sensitive to sample sizes (Kalinowski 2011), we ran the full dataset (n=1377) and a subset with a maximum of 50 randomly chosen individuals per sampling group (n=657) to confirm the patterns of clustering. K (the number of genetic clusters assumed to be present in the dataset) varied from 1-16 with five replicates per value. We considered individual redband to be admixed between wild and hatchery stocks if they produced ancestry coefficients (or q-values) greater than 0.2. There is debate in the genetics literature regarding acceptable q-value thresholds produced by STRUCTURE for classifying individuals as hybrids (Vähä and Primmer 2006, Bohling et al. 2013). Q>0.2 is a conservative value based on

common practice and is analogous to classifying an individual as a hybrid if 20% of its ancestry assigns to a hatchery stock. Studies suggest that $q>0.2$ is often indicative of true ancestry for a particular group, whereas lower levels can be due to statistical noise (Bohling et al. 2013).

The final method we used to test for hatchery influence was an assignment test approach implemented by the R package *assignPOP* (Chen et al. 2017). In our case, since we did not have a clearly defined wild reference group of redband trout, we used the cross-validation approach in *assignPOP* to estimate self-assignment of individuals back to the designated sampling groups. We used the *k*-fold cross-validation approach with the RandomForest model to estimate membership probabilities to the pre-defined sampling groups and hatchery strains. We expected low self-assignment rates for wild populations due to gene flow and shared ancestry; the main goal was to identify individuals captured in the wild that assigned to the hatchery stocks.

Population genetics

For each sampling group (i.e., reach or tributary capture location) of wild-caught fish, we conducted tests of deviations from Hardy-Weinberg Proportions (HWP) and linkage disequilibrium (LD). HWP is expected in populations exhibiting random breeding; deviations from HWP can be due to population processes such as inbreeding and introgression and marker properties such as genotyping errors. Exact tests of HWP were conducted using GenePop 4.2 (Rousset 2008). For these tests we set our *p*-value threshold for significant deviation of HWP at 0.05. Because we analyzed each population individually at all 262 autosomal loci (see Results), we had a total of 3,406 significance tests for HWP. To avoid inflated Type I error due to multiple tests, we performed several additional analyses to compliment the raw HWP *p*-values. First, we used a cumulative binomial function to estimate whether the number of significant observed fell outside the expected range given the number of tests and $\alpha=0.05$ (Waples 2015). We also adjusted *p*-values using the False Discovery Rate procedure (Narum 2006).

Given the number of HWP deviations we observed, we wanted to discern whether they were due to population-level effects or locus-specific effects. To test for population effects, we compared locus-specific estimates of F_{IS} within all 13 populations and the proportion of positive and negative values. A trend in F_{IS} values in either direction is indicative of population processes that can influence HWP. For the number of positive and negative values we observed, we performed a χ^2 test of equal proportions. Locus-specific effects were determined by estimating the expected number of deviant populations per locus using a binomial distribution and $\alpha=0.05$.

LD is a measure of association between two presumably independent genetic markers in a dataset. Strongly linked markers or populations with significant linkage within them can violate assumptions of some genetic analyses. Processes such as non-random mating, cryptic population structure, inbreeding, and bottlenecks can elevate LD between loci. To test whether sampling groups were in LD, we estimated the Index of Association (I_a) for each individual sampling group using the R package *poppr* (Kamvar et al. 2014). Higher values of I_a suggest higher within-group LD. This test estimates average linkage disequilibrium (r_d) within a population across all pairs of loci and then uses a permutation test to examine whether the observed value deviates from the null hypothesis of no LD. The metric r_d ranges from zero to one with low values indicating weak association between pairs of loci and high values strong association.

Global estimates of observed heterozygosity (H_o), expected heterozygosity (H_e), and heterozygote excess (F_{IS}) for each locus were generated for each sampling group using the R package *diveRsity* (Keenan et al. 2013). F_{IS} ranges between -1 and 1 with values near zero indicating equivalence between the observed number of heterozygote individuals and the expected number of heterozygotes based on allele frequencies. Deviations from zero could be generated by several different population processes. For example, inbreeding, genetic drift, and cryptic population structure can generate positive F_{IS}; negative values can be due to outbreeding, small effective size, or sex-based selection differences. We also calculated the proportion of loci that were polymorphic within each group.

Population structure

To assess genetic structure across the entire dataset we performed several analyses. First, we performed pairwise estimates of genetic differentiation between sampling groups. We used the metric G_{ST} (Nei and Chesser 1983) implemented in *diveRsity* with 95% confidence intervals estimated using 10,000 bootstrap replicates. G_{ST} ranges from 0 to 1, with larger values indicating higher differentiation. Given the dendritic and linear features of a river system, we hypothesized that gene flow and migration should follow a linear stepping stone model. Also, the presence of barriers, whether natural or artificial, should restrict gene flow between adjacent populations. To test these patterns, we estimated directional migration rates using the *divMigrate* function in *diveRsity*, which is based on the method described by Sundqvist et al. (2016). We estimated the relative migration rate matrix, which is scaled from 0 to 1, for pairs of populations using the G_{ST}

metric. For relative migration, higher values indicate relatively greater exchange of genes between groups compared to the entire dataset.

To disentangle the natural structure from the groupings created by the sampling scheme, we first performed a Principal Components Analysis (PCA) based on allele frequencies with all the wild samples. We also conducted a Discriminant Analysis of Principal Components (DAPC)(Jombart et al. 2010) based on allele frequencies, which provide a model-free multivariate perspective on population structure. We used the *find.clusters* function in the R package *adegenet* (Jombart et al. 2008) to identify the grouping with the lowest Bayesian Information Criterion (BIC) score based on *K*-means clustering. Like STRUCTURE, this method determines the optimal clustering pattern in the dataset across a range of *K*-values. We estimated the posterior probability of individual membership to each of these identified clusters, which is analogous to ancestry coefficients for an individual redband. We did this for multiple values of *K* that produced similar BIC scores.

Effective number of breeders

Since redband trout collected for this study were estimated to be of the same age cohort, we estimated the effective number of breeders (N_b) that produced these individuals. In other words, we estimated the number of adult redband trout that produced this cohort based on the genetic data. For simplicity we first grouped individuals based on sampling groups. We estimated N_b using two different frameworks. The first used within-population linkage disequilibrium to estimate effective size ($N_{b\text{-}LD}$) as estimated by the program NeEstimator 2.01 (Do et al. 2014). The second approach used the program COLONY2 (Wang 2012) to identify sibling groups and estimate effective size ($N_{b\text{-}Sibs}$). COLONY2 attempts to identify dyads (i.e. sibling groups) in the dataset and then estimates $N_{b\text{-}Sibs}$ using a re-sampling approach that incorporates family size. Along with $N_{b\text{-}Sibs}$ we estimated family sizes and number of families for each population. For each COLONY2 analysis we performed five independent runs with the full likelihood model.

Based on the results of the population structure analyses, we then grouped individuals based on probable genetic units for estimating N_b. We did this using both NeEstimator and COLONY2. For this we combined reaches 2, 3, 4, and Tumalo Creek; reaches 5, 6, and 7; and reaches 9 and 10.

Results

Sequencing success

In total our GT-seq library produced 139.9 million DNA sequence reads. Removing negative controls, the average number of reads per individual was 97,142 (SD=82,441). On average 41,455 (SD=33,087) of those reads per individual were on-target reads corresponding to the loci in our panel. One locus (Omy_rbm4b-203) did not produce usable genotypes and was removed from the dataset. For the 1,417 individual fish in the library, the mean proportion of loci genotyped was 93.9% (minimum 0%, maximum 100%, SD=13.5%). We decided to retain individuals that were genotyped at over 70% of the loci, which resulted in a dataset containing 1,377 individuals. Among these individuals the mean genotyping success rate per locus was 96% (SD=9.2%). We obtained genotypes for 263 loci (which included a single sex-ID marker) from 74% or more individuals. For the subsequent analyses the sex-ID marker was not included. Therefore, the downstream population genetic analyses were based on 262 loci.

Hatchery introgression

The tests for genetic introgression revealed little genetic influence of out-of-basin hatchery stocks on the wild-caught redband trout. The correspondence analysis (CA) revealed that the Oak Springs and Cape Cod stocks had distinct allele frequencies divergent from wild populations (Figure 2). The Wizard Falls Hatchery cranebow stock was intermediate between wild Deschutes populations and the other out-of-basin hatchery stocks in the CA. Fall River was the only population of redband trout that aligned with the Wizard Falls stock.

Similar patterns were observed with the assignment test. Across all *k*-fold values (2-5), the mean self-assignment rates for the Oak Springs and Cape Cod strains were 96% and 89%, respectively. For Wizard Falls, the mean self-assignment rate was 54.9% with the highest values observed when *k*-fold was five; therefore, we report the results using this parameter value. Out of the 45 individuals of the Cape Cod strain, for 42 the highest probability of assignment was back to this group. For Oak Springs this value was 42 out of 43. In contrast, only 20 out of 34 of the Wizard Falls fish had the highest probability of assignment back to that strain; another 11 assigned to Fall River. Similarly, for Fall River 16 out 36 individuals self-assigned back; 13 were assigned to Wizard Falls. No individuals captured in the wild assigned to either the Cape Cod or Oak Springs groups. Three individuals from Reach 10 (4% of fish from that reach), two from Reach 7 (11%), and four from Reach 8 (15%) were assigned to the Wizard Falls stock.

The STRUCTURE analysis divisions were more clearly defined using the subset of 50 individuals per sampling group; thus, we report those results. The primary division was between the wild Deschutes redband and the two out-of-basin hatchery stocks. At $K=2$ the Wizard Falls stock had almost equal assignment to those two groups. Wild populations formed distinct clusters at higher levels of K; at K=7 the Wizard Falls stock formed a distinct cluster. At K=7 Fall River had the highest proportion of ancestry assigned to the Wizard Falls stock and the highest proportion of individuals with q-values (i.e., ancestry coefficients) greater than 0.2 for that cluster (Table 1). Only reaches 7, 8, 9, 10, and Fall River had any individuals with q>0.2 Wizard Falls ancestry. Excluding Fall River, only five out of the 25 individuals with q>0.2 also produced a value greater than 0.8, suggesting these individuals with moderate values ($0.2<q<0.8$) are hybrids between wild redband trout and Wizard Falls cranebows.

Population genetics

Out of 3,406 tests of HWP, 269 produced p-values that were below the 0.05 threshold (Table 2). Another 299 locus-population pairings could not be tested because the population was fixed (i.e., monomorphic) at that particular locus. Four loci (Omy_BAMBI4, Omy_inos97, Omy_myclar, Omy_Omyclm) were fixed in all sampling groups and the three hatchery strains. One (Omy_RAD768) was fixed in all groups except the Cape Cod strain. Locus Ocl_gshpx3 was fixed in all groups except Reach1. Another locus (OMS00174) was fixed in all groups except the three hatchery strains and Fall River.

With our error rate under a binomial distribution, we expected between 146 and 195 significant tests to occur purely by random chance. The number we observed was much higher. We observed trends in the distribution of HWP deviations across populations and loci. Most sampling groups tended to display a positive F_{IS} skew across loci; however, of the three significant χ^2 tests we observed, two were due to an excess of loci producing negative F_{IS} values (Table 2). At the locus-level, the number of sampling groups out of HWP for a particular locus ranged from 0 to 12 (Figure 3). Assuming a binomial random distribution, we would expect at most four sampling groups to display significant p-values per locus; however, we observed 15 loci with significant values in five or more sampling groups, suggesting some locus-level effects.

When the raw p-values were corrected for multiple tests using the FDR method, the number of significant tests dropped to 69. With this FDR correction, most sampling groups produced less than five significant deviations except for Reach4 (11 deviations), Reach 5 (15),

Reach 6 (8) and Tumalo Creek (13). At the locus level, most loci had deviations in three or fewer sampling groups; four (Omy_986831, Omy_aromat, Omy_GHSR12, Omy_hus152) had significant deviations in four or more populations and two (OMS00018 and OMS00173) had significant deviations in five or more. Most of the deviations observed for these six loci were for the cluster of sampling groups (reaches 4-6 and Tumalo Creek) that had the most overall corrected FDR deviations.

All but one permutation test examining LD produced a *p*-value less than 0.05 and 13 out of 16 sampling groups produced values <0.001 (Table 3), indicating observed associations between loci within sampling groups was greater than expected by chance. However, mean r_d was less than 0.01 for all wild sampling groups except Fall River. Reach7, Reach8, Tumalo Creek, Fall River, and the Little Deschutes were the only wild sampling groups to produce global I_a values >1, indicative of high LD across all loci.

Levels of heterozygosity were relatively even across all 13 sampling groups (Figure 4A): the highest values were observed in Fall River, Reach8, and Reach10 and the lowest values in the Little Deschutes River and Reach5. Every population was fixed for a single allele in less than 15% of loci with the exception of the Little Deschutes River, which had over 22% of loci fixed (Figure 4B). F_{IS} tended to positive, with every sampling group from Reach6 downstream producing 95% confidence intervals that did not overlap zero (Figure 4C). The Little Deschutes River was the only group to produce a negative F_{IS}, but its 95% CI did overlap zero.

Population structure

Average pairwise G_{ST} between sampling groups was 0.03 and no estimate produced a 95% CI that overlapped zero (Figure 5). The highest differentiation was observed between Reach1 and all other populations. The Little Deschutes and Fall River also produced relatively high values when compared to other populations. The lowest values were observed between reaches 2, 3, 4, and Tumalo Creek; reaches 5 and 6; reaches 6 and 7; and reaches 9 and 10. A matrix of all G_{ST} values between pairs of sampling groups is provided in Appendix 1. For relative migration rates, the average value across all pairs of populations was 0.26 with a standard deviation of 0.19. When rates were filtered to display those above 0.5, several groupings emerged (Figure 6A). The lower portion of the system mostly followed a linear model, with the following pairings showing bidirectional migration: reaches 2 and 3, reaches 3 and 4, Reach3 and Tumalo Creek, and Reach4 and Tumalo Creek. Unidirectional migration was

observed from Tumalo Creek to Reach2. Reaches 5 and 6 showed bidirectional migration and unidirectional migration from Reach6 into Reach7. Bidirectional migration was also observed between reaches 9 and 10. When Wizard Falls Hatchery stock was included, bidirectional gene flow was observed between the hatchery and Fall River (Figure 6B). The entire matrix of pairwise migration rates and a network diagram displaying these rates are in Appendix 2.

The first axis of the PCA separated Reach1 from the remaining sampling groups in multivariate space (Figure 7). Along the second axis the remaining groups formed a gradual cline with substantial overlap. On one end of this cline were reaches 2, 3, 4, and Tumalo Creek and on the opposite end reaches 5 and 6. With the DAPC, the lowest BIC score was observed at K=8, although comparable scores were observed for other values (Figure 8). Thus, we report the results from K=6-8. Across all these values of K, Reach1 formed a distinct cluster (Figure 9). Reaches 9 and 10 together also formed a distinct cluster across K-values as did Fall River, with individuals from both of these clusters observed in neighboring sampling groups. From K=6 to K=8 the biggest changes in groupings occurred in the middle sampling groups. At K=6 reaches 2, 3, 4, and Tumalo Creek formed a single distinct cluster. Reach5 formed a cluster as did reaches 6 and 7, but there was substantial migration across these clusters. A small subset of Tumalo Creek redband clustered with reaches 6 and 7 and the Little Deschutes clustered with reaches 2-4.

At K=7 a subset of individuals from Tumalo Creek formed a distinct cluster, with some individuals from reaches 2-4 displaying shared ancestry with this group (Figure 9). The Little Deschutes also formed a distinct cluster with some individuals having shared ancestry with reaches 6 and 7. At K=8 a new cluster was formed by Reach 4 with substantial shared ancestry with reaches 2 and 3. Little Deschutes clustered with the Reaches 2-3 group at this K-value.

Effective number of breeders

The two methods we utilized for estimating the effective number of breeders produced dramatically different values. Estimates from the $N_{b\text{-}LD}$ method were almost always lower than $N_{b\text{-}Sibs}$ and often by an order of magnitude (Figure 10). In general the largest values of N_b were observed in reaches 2, 3, 4, 5, and 6. For these five sampling groups neither $N_{b\text{-}LD}$ nor $N_{b\text{-}Sibs}$ produced a confidence interval that stretched below 200. In contrast the lowest values were observed for Fall River and Reach8. Across the 13 sampling groups, the number of new family groups identified by COLONY2 decreased with increasing sample size (Figure 11).

When sampling groups were combined (reaches 2, 3, and 4; reaches 5 and 6; reaches 9 and 10), trends varied depending on the specific metric. $N_{b\text{-}LD}$ for the combined groupings tended to be intermediate to estimates for the individual sampling groups (Figures 11, 12). The combined estimates also had much narrower confidence intervals. For the Reach2-4 and Reach5-6 groupings, the combined $N_{b\text{-}Sibs}$ estimates were much larger than any of the individual group estimates (with the exception of Reach4). The combined $N_{b\text{-}Sibs}$ estimate for the Reach9-10 grouping was intermediate to the two groupings individually.

Conclusions

Our study provides the most comprehensive survey of redband trout genetics in the Deschutes River basin above barriers to anadromy. With the large number of samples and spatial scale of the study, the use of GT-seq provided an efficient technique to genotype our sample of fish at a large number of loci. The panel we used was originally designed for anadromous steelhead trout (Campbell et al. 2015) but we observed a high genotyping success rate for these resident redband trout populations and most loci were polymorphic. This suggests we have a robust dataset for examining the population genetics of redband trout from this basin. Importantly, by altering our DNA extraction protocol we were able to obtain sufficiently high quantities of DNA required for GT-seq despite using smaller fin clips from YOY fish.

These results provide multiple insights into the genetic diversity and structure of redband trout from this basin. First, we detected no genetic influence from out-of-basin strains (e.g. Oak Springs and Cape Cod strains) of rainbow trout that have been in the past stocked in the basin. However, there is evidence of genetic introgression from the Wizard Falls (i.e., cranebow) strain in the basin. Since the trout sampled for this study were all YOY fish, this implies released cranebows (which are released at a larger size in Fall River) are reproducing in the wild and the presence of trout with admixed ancestry suggests they are interbreeding with wild-origin individuals. Unsurprisingly, Fall River had the greatest amount of influence from the hatchery strain and levels of introgression decreased with further distance from that tributary: reaches 7 and 8 had the highest amounts of cranebow ancestry in the main stem Deschutes River. This suggests that the presence of cranebow ancestry in the main stem is due to recently stocked individuals and/or their offspring dispersing from Fall River, as opposed to lingering residual ancestry from past stocking events. However, these data are not sufficient to conclusively distinguish between these two scenarios.

There were several clear patterns of genetic structure among wild-origin redband. First, Reach1, which is between Steelhead and Big Falls on the Deschutes River, is the most distinct redband population sampled. There were a few suspected immigrants captured in this area, but otherwise it was genetically homogenous with no gene flow from upstream populations. Big Falls was the historic limit of anadromy for *O. mykiss* and continues to serve as barrier to gene flow. There is a remote possibility that the redband we sampled were not native redband but the progeny of recently released anadromous steelhead. With fish passage facilities recently installed at Round Butte Dam, managers have released hundreds of thousands of steelhead smolts above the dam to establish migratory populations (Adams et al. 2015). Since 2012 steelhead adults have also been released above the dam. However, no smolts have been released above Steelhead Falls and no radio tracked adult steelhead has been documented in this section of the river (Becky Burchell, pers. comm.). It is highly unlikely any returning steelhead have contributed progeny to this section of the Deschutes, insinuating that Reach1 contains a highly distinct and isolated population of native redband trout.

Above Big Falls, from a genetic perspective reaches 2, 3, 4, and to a lesser extent Tumalo Creek appear to form an interconnected population. Although some analyses, such as the *K*-means and positive F_{IS}-values, suggested some sub-structuring, there appears to be extensive exchange of individuals and genes among these reaches. This is despite several natural falls that could be barriers to movement. Gene flow appeared to follow a mostly linear model with downstream areas receiving the highest proportion of migrants. Active gene flow likely contributes to the large estimates of N_b for the reaches individually and combined.

North Canal Dam, which is located in the city of Bend and formed the boundary between reaches 4 and 5, was a complete barrier to gene flow. This may change in the future: the North Canal Dam fish ladder was completed in 2017, providing fish passage above the dam for the first time since its construction in 1912. Future genetic analyses will be useful to detect the degree to which gene flow is restored between populations that in this study were genetically distinct, and how quickly. Similar to the lower reaches, reaches 5, 6, and 7 appear to form an interconnected genetic population and have high N_b. There was some sub-structuring between reaches 5 and 6, which could be due to the presence of Lava Island Falls, but gene flow appears to occur between these two groups. Reach5 also had the largest sample size of any sampling group. Some analyses

of population structure can be biased due to sample size inequality (Kalinowski 2011), so there is the potential that the observed sub-structure hierarchy could be an artifact of over-splitting.

Above Reach6 the patterns are less clear. Reaches 9 and 10 had low genetic differentiation and formed a single cluster. Reach7 appeared to share much of its ancestry with Reach6 and Reach8 was more closely associated with Reach9, but both were mixed and were influenced by the cranebow stock. Despite this mixed ancestry, both reaches 7 and 8 produced some of the lowest estimates of N_b. They also had the lowest sample sizes, which could have biased these estimates. An intriguing population was the Little Deschutes River: it was distinct from neighboring reaches and appeared to be more closely related to populations below North Canal Dam. It also had an extremely low effective size (which could be influenced by the low sample size and surveys only in a small portion of this tributary) and low levels of genetic diversity. These lines of evidence suggest that the Little Deschutes River contains a distinct native population that is isolated from main stem populations, but additional sampling would improve the robustness of these conclusions.

Many insights can be drawn from the overall dataset and the various analytical tools we used. The large number of HWP deviations we observed and high within-population LD suggested the designated sampling reaches were not meeting expectations of randomly mating populations. Analyses of population structure and positive F_{IS} values suggested the boundaries of these sampling groups did not correspond to the underlying natural population structure, which is likely causing many of these deviations. It was particularly evident that reaches 2, 3, 4, and Tumalo Creek; reaches 5 and 6; and reaches 9 and 10 form genetic clusters. Individually these sampling groups had the highest number of HWP and LD deviations, but they contained multiple distinct clusters that were exchanging genes. These findings underscore the need to evaluate natural patterns of population structure rather than relying solely on *a priori* designations. Management of this watershed should consider these patterns, for there are clear boundaries to gene flow influencing population dynamics throughout the watershed.

The comparison of the two methods for estimating N_b also revealed interesting patterns. The $N_{b\text{-}Sibs}$ estimate was always higher than $N_{b\text{-}LD}$, often dramatically so. The methods use different frameworks to estimate N_b. Which technique produces the estimate that is most accurate is difficult to determine, especially with an empirical dataset in which the "true" number of breeders is unknown. The $N_{b\text{-}LD}$ estimate is based on random LD that emerges in a single

generation (Waples and Do 2010). However, our sampling groups displayed extensive LD and sub-structuring that can bias point estimates of $N_{b\text{-}LD}$ (Wang et al. 2016). The method is also optimal for use with smaller populations and datasets with multi-allelic marker (Waples and Do 2010, Wang et al. 2016). This was evident from the estimates of $N_{b\text{-}LD}$ when several sampling groups were combined. At the very least combining individuals from multiple sampling groups should have increased the point estimates of N_b, but they were often intermediate to the single sampling group estimates.

Wang (2016) suggested that the $N_{b\text{-}Sibs}$ method is more accurate than $N_{b\text{-}LD}$ based on simulations, especially when sample sizes exceed true N_b. The calculation for $N_{b\text{-}Sibs}$ is based on the number and size of family groups compared to the entire sample, so it is more robust to violations of assumptions of random mating. It should also be robust to increasing population sizes and when we combined individuals from multiple sampling groups $N_{b\text{-}Sibs}$ increased as expected. This method appears to be more sensitive to sample size, especially for this system in which average family sizes were small: the sampling groups with low sample size produced wide error bars. Overall small family sizes across the dataset are likely responsible for the wide error bars observed with $N_{b\text{-}Sibs}$. Given that the designated sampling groups do not contain small, discrete natural populations and gene flow is extensive in the system, the $N_{b\text{-}Sibs}$ estimate likely provides a more robust estimate of the number of breeders than $N_{b\text{-}LD}$.

It should be noted as well that N_E and N_b are not measures of population abundance; at best they are proxies for the absolute number of individuals in the population. However, studies of other freshwater salmonids have shown that N_b is not always correlated with census size and even when there is the correlation can be weak and site-specific (Palstra and Fraser 2012, Whiteley et al. 2015, Ferchaud et al. 2016). While census size can experience dramatic annual fluctuations, N_b can be stable. As Whiteley et al. (2015) observed with eastern brook trout, increases in census size did not produce a comparable increase in the number of breeders. Thus, managers of the Deschutes River system must recognize that N_b is a different metric reflecting different population attributes compared to traditional estimates of census size. The merits of monitoring N_E/N_b for natural population have long been touted by geneticists because it provides a measurement of crucial population processes (Franklin 1980, Schwartz et al. 2007, Frankham et al. 2014) and can be more accurate than general estimates of census size (Whiteley et al. 2015). The differences in estimation and interpretation of N_b compared to traditional census size

estimates must be communicated with managers and stakeholders if it is to be used as a monitoring and management tool.

Patterns of $N_{b\text{-}Sibs}$ estimates are relevant for evaluating redband trout in the middle Deschutes River basin. The conservation genetics community has frequently debated general guidelines for interpreting N_E. One of the earliest recommendations was the 50/500 rule, which proposed that a minimum N_E of 50 was needed to avoid the short-term impacts of inbreeding and a long-term N_E of 500 was needed to preserve evolutionary potential (Franklin 1980). These values have been oft-debated and a recent re-evaluations suggests a 100/1000 rule is more appropriate (Frankham et al. 2014). Some context is needed before evaluating our results against these generalized rules.

First, N_E and N_b, which was what we estimated, are not the same. N_E is a measure of the number of individuals that contribute offspring to a gene pool; N_b is the number of individuals that contribute offspring to a specific cohort. Given that redband trout are iteroperous and individuals from multiple generations likely form the overall spawning population, our estimates of N_b from a single cohort are likely less than the generational N_E size. Also, a single cohort from a single year could be impacted by stochastic processes (e.g. extreme drought, disease, etc.), potentially resulting in estimates of N_b that are not reflective of long-term trends.

Second, the main motivation for developing these guidelines was assessing the viability of isolated individual populations of conservation concern (e.g. endangered species, captive populations). Redband trout in the Deschutes do not appear to function as discrete isolated populations: there is a hierarchical metapopulation structure. Thus, determining an appropriate unit (i.e. spatial scale) to base the 50/500 or 100/1000 rules is difficult. Modeling metapopulation effective size and gene flow would provide a more realistic image of the long-term dynamics of the entire watershed (Hössjer et al. 2014).

Given these considerations, overall the estimates of $N_{b\text{-}Sibs}$ are above thresholds we would expect for viable populations. Unsurprisingly, the aggregate populations that covered multiple sampling groups (reaches 2-4; reaches 5-7) produced values above 1000, suggesting these reaches contain robust redband populations. Several groups, though, warrant additional attention. Although the point estimate of $N_{b\text{-}Sibs}$ for Reach1 was 370 (95% CI 303-457), it is totally isolated from all other reaches and thus unlikely to receive influxes of gene flow. This could be a

problem should the population decline or experience annual fluctuations that reduce effective size and increase inbreeding. This is also true for reaches 9 and 10.

Reaches 7, 8, and Fall River all produced among the lowest $N_{b\text{-}Sibs}$ estimates and also had the highest levels of introgression. If the Wizard Falls cranebow stock is composed of highly related individuals, then their presence in wild populations could deflate $N_{b\text{-}Sibs}$ artificially by forming larger sibling groups. Regardless, the presence of introgressed individuals suggests that any "natural" population of redband in these regions likely exist in low numbers and do not form a substantial breeding population. The Little Deschutes River produced one of the lowest estimates of $N_{b\text{-}Sibs}$. Considering the totality of the genetic data suggests, this is an isolated, genetically distinct population containing a small number of breeders. Although we must consider the low sample size, our results for this population suggest that it is one of the most vulnerable to some of the deleterious effects of low N_E.

Data and sampling management

Tissue samples and purified extracted DNA obtained from ODFW are archived at AFTC and can be provided to other researchers upon request. Genotypes are stored at AFTC and can be provided to ODFW cooperators and the public at large if requested. A copy of the report is archived in the AFTC database and can be made available to the public upon request.

Acknowledgements

We thank ODFW field personnel, especially Beth Bailey, for collecting redband trout samples. Erik Moberly at ODFW and Becky Burchell at Portland General Electric provided valuable information on stocking practices in the basin. Nate Campbell at CRITFC was extremely helpful in providing advice on developing the GT-seq libraries. We also thank Amanda Matala and Shawn Narum at CRITFC for allowing us to run our library on the CRITFC sequencer and providing the genotypes. Christian Smith and Ann Gannam at AFTC provided comments that improved the 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.

Suggested citation

Bohling, J.H., J. Von Bargen, S. Starcevich. 2017. Evaluation of genetic population structure and effective population size among populations of redband trout in the Deschutes River Basin, OR. AFTC Final Report 325.

References

Adams, B., P. DeHaan, and C. Smith. 2015. Genetic determination of stock of origin for *Oncorhynchus mykiss* collected in the Upper Deschutes River Basin. AFTC Final Report 278.

Besnier, F., and K. A. Glover. 2013. ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers. PLoS ONE 8:1–9.

Bohling, J. H., J. R. Adams, and L. P. Waits. 2013. Evaluating the ability of Bayesian clustering methods to detect hybridization and introgression using an empirical red wolf data set. Molecular Ecology 22:74–86.

Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources 15:855–867.

Chen, K-Y., E.A. Marschall, M.G. Sovic., A.C. Fries, H.L. Gibbs, and S.A. Ludsin. 2017. assignPOP: Population assignment using genomic, non-genetic or integrated data in a machine-learning framework. R package version 1.1.2. https://CRAN.R-project.org/package=assignPOP.

Currens, A. K. P., C. B. Schreck, and H. W. Li. 1990. Allozyme and morphological divergence of rainbow trout (Oncorhynchus mykiss) above and below waterfalls in the Deschutes River, Oregon. Copeia 1990:730–746.

Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources 14:209–214.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–87.

Ferchaud, A.-L., C. Perrier, J. April, C. Hernandez, M. Dionne, and L. Bernatchez. 2016. Making sense of the relationships between Ne, Nb and Nc towards defining conservation thresholds in Atlantic salmon (Salmo salar). Heredity 117:268–278

Frankham, R., C. J. A. Bradshaw, and B. W. Brook. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation 170:56–63.

Franklin, I.R., 1980. Evolutionary change in small populations. In: Soulé, M.E., Wilcox, B.A. (Eds.), Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer, Sunderland, MA, pp. 135–149.

Hössjer, O., F. Olsson, L. Laikre, and N. Ryman. 2014. A new general analytical approach for modeling patterns of genetic differentiation and effective size of subdivided populations over time. Mathematical Biosciences 258:113–133.

Jombart, T., I. Ahmed, F. Calboli, A. Cori, T. E. Reiners, P. Solymos, and M. T. Jombart. 2008. Package "adegenet." Bioinformatics Application Note 24:1403–1405.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11:94.

Kalinowski, S. T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 106:625–32.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis

of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.
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.
Matala, A. P., S. Marx, and T. G. Wise. 2008. A genetically distinct wild redband trout (Oncorhynchus mykiss gairdneri) population in Crane Prairie Reservoir, Oregon, persists despite extensive stocking of hatchery rainbow trout (O. m. irideus). Conservation Genetics 9:1643–1652.
Muhlfeld, C. C., S. E. Albeke, S. L. Gunckel, B. J. Writer, B. B. Shepard, and B. E. May. 2015. Status and conservation of Interior Redband Trout in the Western United States. North American Journal of Fisheries Management 35:31–53.
Narum, S. R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics 7:783–787.
Nei, M., and R. K. Chesser. 1983. Estimation of fixation indices and gene diversities. Annals of Human Genetics 47:253–259.
Palstra, F. P., and D. J. Fraser. 2012. Effective/census population size ratio estimation: a compendium and appraisal. Ecology and Evolution 2:2357–65.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–59.
Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106.
Schroeder, R. K., and J. D. Hall, editors. 2007. Redband trout: resilience and challenge in a changing landscape. Oregon Chapter of the American Fisheries Society, Corvallis, OR.
Schwartz, M. K., G. Luikart, and R. S. Waples. 2007. Genetic monitoring as a promising tool for conservation and management. Trends in Ecology & Evolution 22:25–33.
Sundqvist, L., K. Keenan, M. Zackrisson, P. Prodohl, and D. Kleinhans. 2016. Directional genetic differentiation and asymmetric migration. Ecology and Evolution 6:3461–3475.
Vähä, J.-P., and C. R. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology 15:63–72.
Wang, J. 2012. Computationally efficient sibship and parentage assignment from multilocus marker data. Genetics 191:183-194.
Wang, J., E. Santiago, and A. Caballero. 2016. Prediction and estimation of effective population size. Heredity 117:193–206.
Waples, R. S. 2015. Testing for Hardy-Weinberg Proportions: have we lost the plot? Journal of Heredity 106:1–19.
Waples, R. S., and C. Do. 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evolutionary Applications 3:244–262.
Whiteley, A. R., J. A. Coombs, M. Cembrola, M. J. O'Donnell, M. Hudy, K. H. Nislow, and B. H. Letcher. 2015. Effective number of breeders provides a link between interannual variation in stream flow and individual reproductive contribution in a stream salmonid. Molecular Ecology 24:3585–3602.

Table 1: Average proportion of ancestry assigned to the Wizard Falls Hatchery cranebow stock by STRUCTURE for the wild Deschutes populations. These are based on the mean *q*-value at *K*=7 using a subset of 50 individuals per population in the analysis. Reaches and tributaries are organized from furthest downstream to furthest upstream.

Sampling group	Average WFH ancestry	Proportion of individuals *q*>0.2 WFH ancestry
Reach1	0.007	0
Reach2	0.010	0
Reach3	0.007	0
Reach4	0.009	0
Tumalo Creek	0.031	0
Reach5	0.006	0
Reach6	0.016	0
Reach7	0.169	0.21
Little Deschutes	0.008	0
Reach8	0.313	0.50
Fall River	0.734	0.89
Reach9	0.047	0.11
Reach10	0.062	0.05
WFH	0.843	1

Table 2: Results of the exact tests of Hardy-Weinberg Proportions (HWP) for each of the 13 sampling units across 262 autosomal SNP loci. The second column indicates the number of loci that produced a significant deviation at the $p \leq 0.05$ level using exact tests. The third and fourth columns are the counts of loci that produced positive and negative F_{IS} values, respectively. The fifth column is the *p*-value from the χ^2 test of equal proportion between the observed and expected number of F_{IS} values. The sixth column is the number of HWP deviations observed after correcting *p*-values using the False Discovery Rate (FDR). The final column is the number of fixed loci observed in each population.

Sampling group	HWP deviations	F_{IS} positive	F_{IS} negative	χ^2 *p*-value	HWP deviations FDR	Fixed loci
Reach1	27	132	115	0.309	5	17
Reach2	24	126	111	0.363	3	26
Reach3	33	126	113	0.438	1	25
Reach4	28	123	122	1.000	11	17
Tumalo Creek	40	156	94	<0.001	13	12
Reach5	40	130	118	0.446	15	18
Reach6	20	134	116	0.255	8	13
Reach7	8	107	128	0.151	0	30
Little Deschutes	3	74	130	<0.001	0	58
Reach8	8	111	130	0.198	1	22
Fall River	17	108	132	0.121	0	22
Reach9	7	96	139	0.004	0	23
Reach10	14	123	124	0.899	1	16

Table 3: Estimates of linkage disequilibrium for each sampling group across all 262 loci. "I_a" refers to the index of association between loci in a population: larger numbers indicate elevated LD. "r_d" is the mean estimate of linkage disequilibrium between pairs of loci within a population. The final column is the *p*-value of the permutation test examining whether mean r_d differs from the null hypothesis of no LD. Values for the three hatchery strains are included as a comparison.

Sampling group	I_a	r_d	*p*-value
Reach1	0.763	0.003	0.001
Reach2	0.516	0.002	0.381
Reach3	0.567	0.003	0.001
Reach4	0.326	0.001	0.001
Tumalo Creek	1.532	0.007	0.001
Reach5	0.567	0.003	0.001
Reach6	0.499	0.002	0.028
Reach7	1.687	0.008	0.001
Little Deschutes	1.330	0.007	0.001
Reach8	1.770	0.008	0.001
Fall Creek	2.640	0.011	0.001
Reach9	0.739	0.003	0.001
Reach10	0.729	0.003	0.001
Wizard Falls Hatchery	0.767	0.003	0.001
Cape Cod strain	7.031	0.033	0.001
Oak Springs strain	1.848	0.009	0.016

Figure 1: Map of the Deschutes River basin and our associated study area. Included are important features of the watershed such as the starting point for the sampling groups and the point location and number of redband trout sampled for this study. Tributaries of the watershed that were sampled (i.e. mainstem Deschutes River, Tumalo Creek, Fall River, and Little Deschutes River) are denoted in blue; other un-sampled tributaries are in gray.

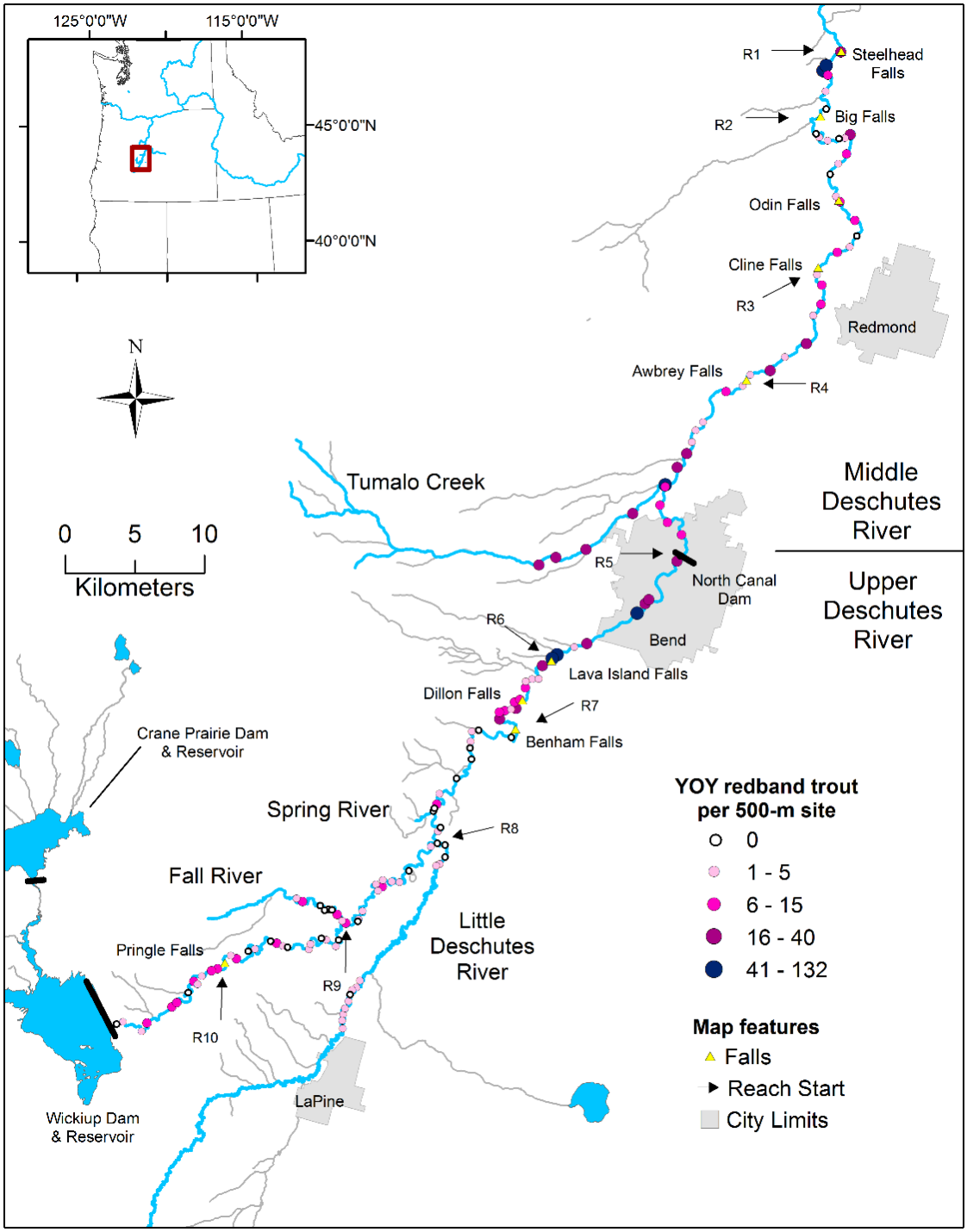


Figure 2: Correspondence analysis based on allele frequencies for the 13 wild Deschutes redband sampling groups and the three hatchery stocks. Wild redband are grouped according to the designated sampling units and each sampling group/stock is represented by a single point. The plot is based on the first two dimensions that explained the greatest amount of variation in the dataset (indicated by the numbers in the parentheses). "WFH"=Wizard Falls Hatchery stock, "OSH"= Oak Spring stock, "CCH"=Cape Cod stock.

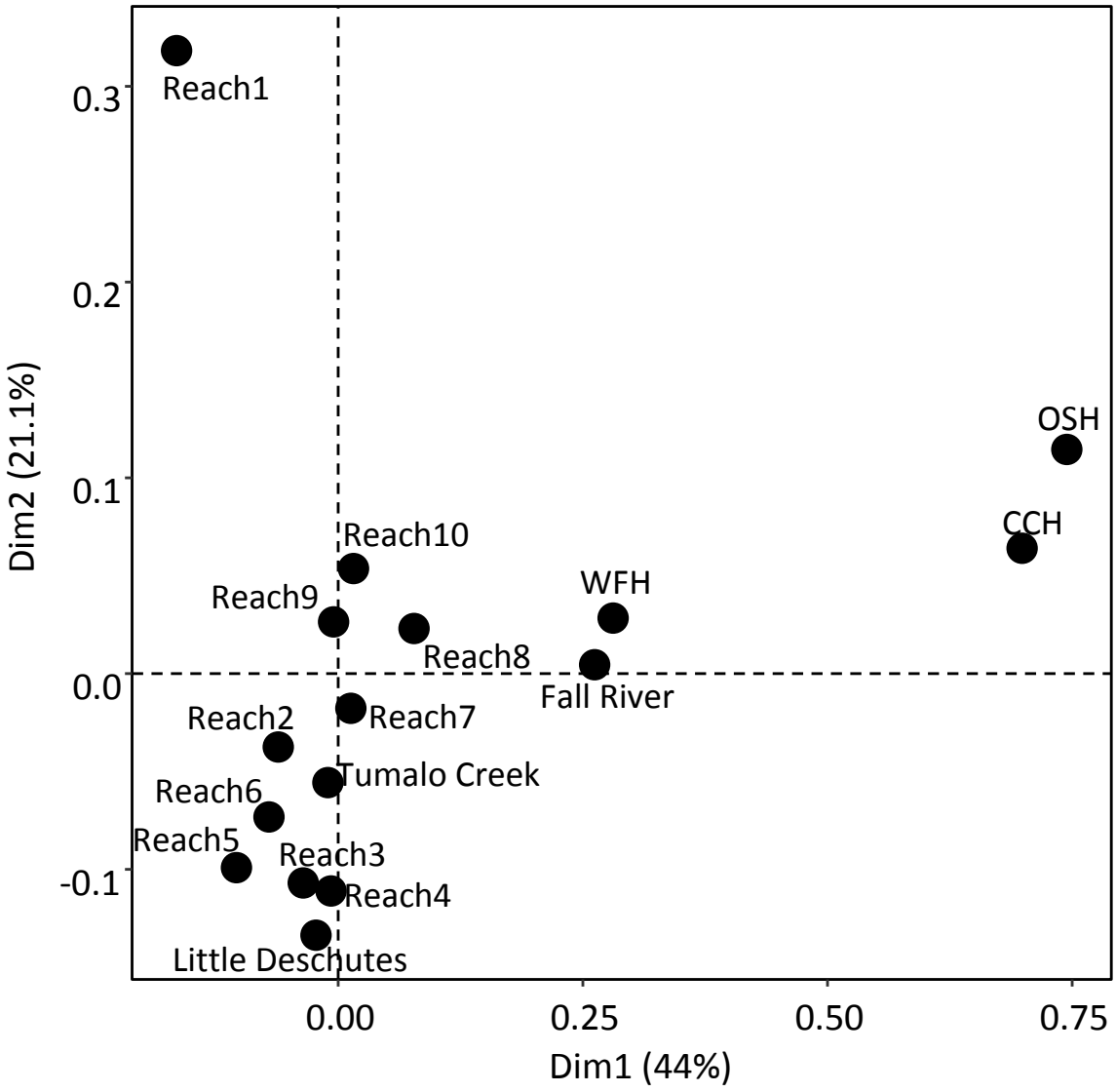


Figure 3: Distribution of the number of populations for which a significant departure from HWP (p<0.05) was observed across the 262 loci examined. These values are compared to expected number of departures under a binomial distribution with α=0.05.

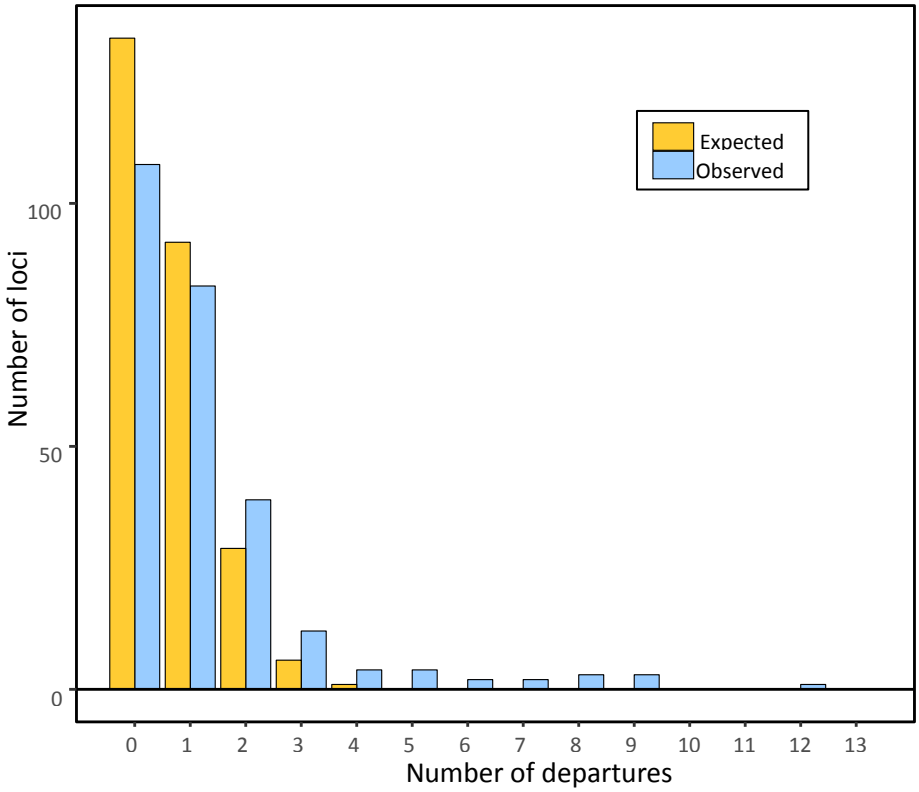


Figure 4: Measures of genetic diversity across the designated sampling units of redband trout from the Deschutes River and its tributaries. A.) Values of expected (H_E) and observed (H_E) heterozygosity. B.) Proportion of loci in the dataset that were fixed (i.e. monomorphic). C.) Population-specific values of F_{IS} and the associated 95% confidence interval based on 1000 bootstraps.

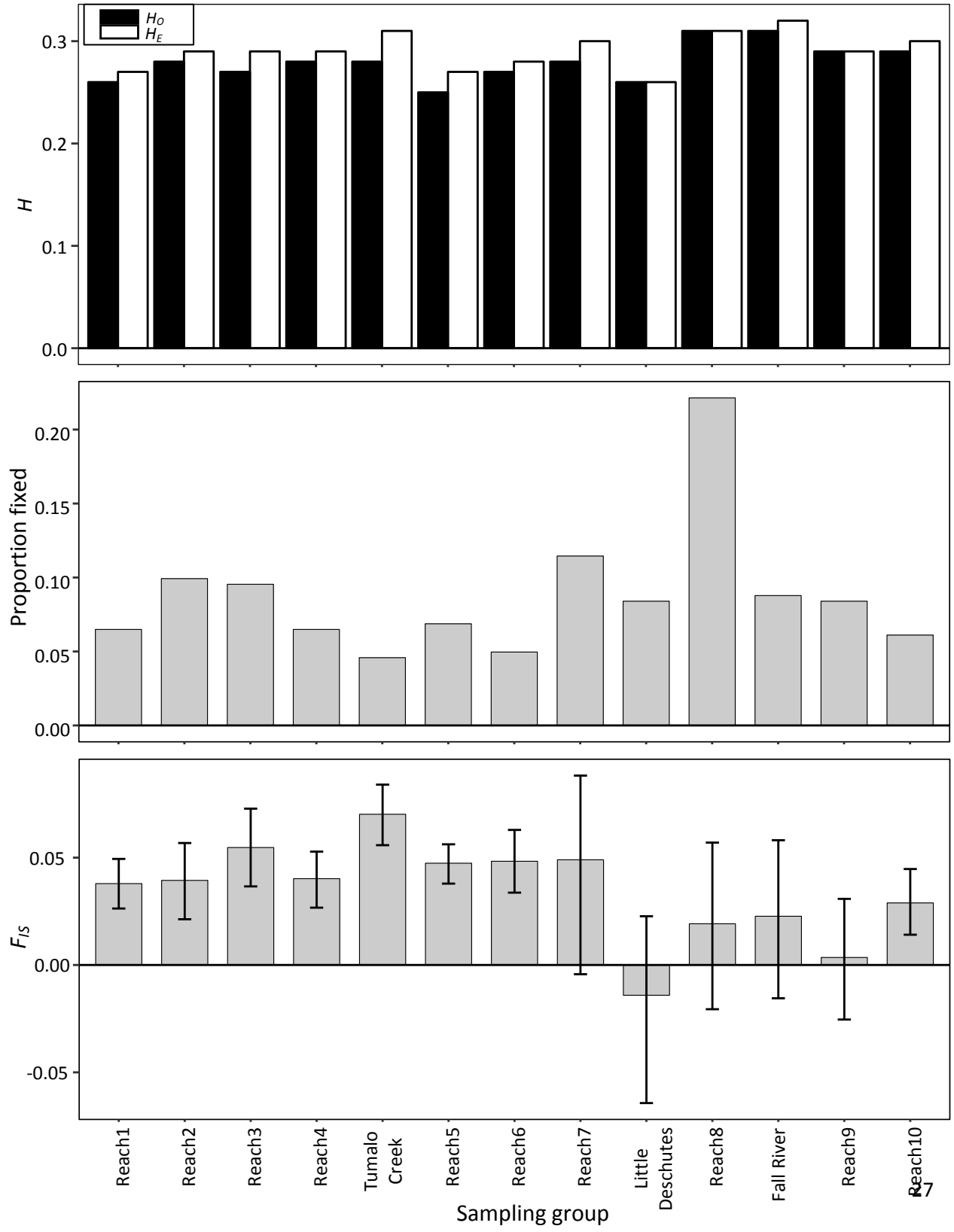


Figure 5: Heatmap of pairwise G_{ST} values between sampling groups. Higher values (darker colors) indicate higher genetic differentiation.

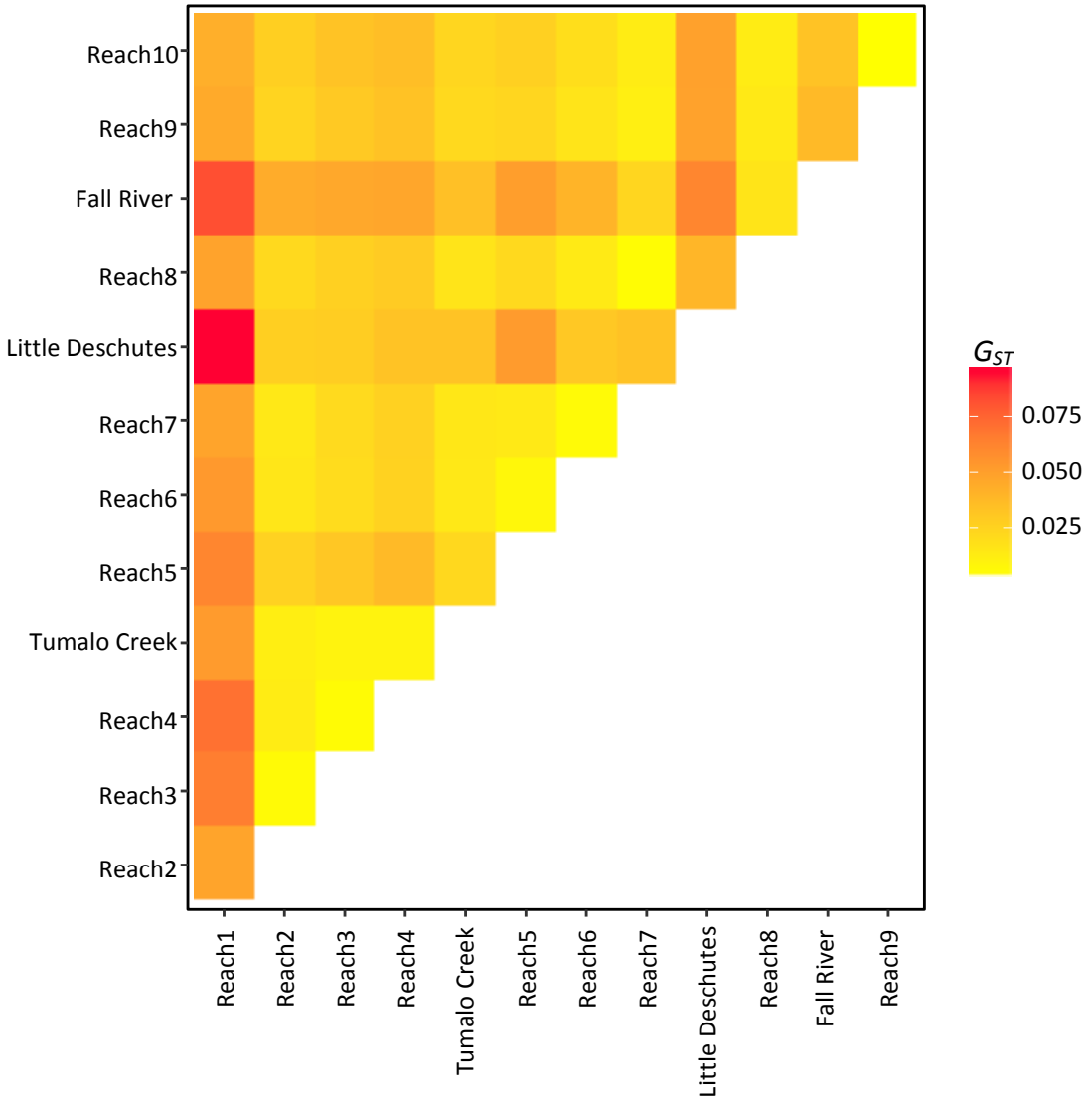
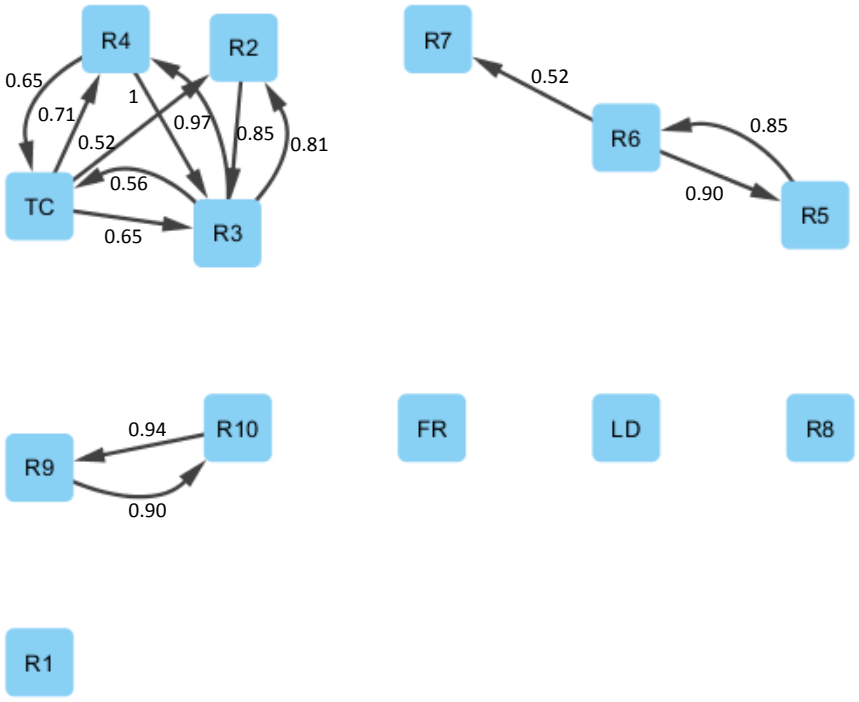


Figure 6: Network plots of migration levels between sampling groups. The arrows refer to the direction of gene flow from one group to another and the edge values reflect the level of gene flow. Higher values mean more sharing of migrants. The overall network was trimmed to only display values with migration levels >0.5. A) Network with only wild sampling groups, B) network with wild sampling groups and Wizard Falls Hatchery. Sampling groups are coded with "R" standing for Reach, "TC" for Tumalo Creek, "LD" for Little Deschutes, "FR" for Fall River, and "WFH" for Wizard Falls Hatchery.

A.



B.

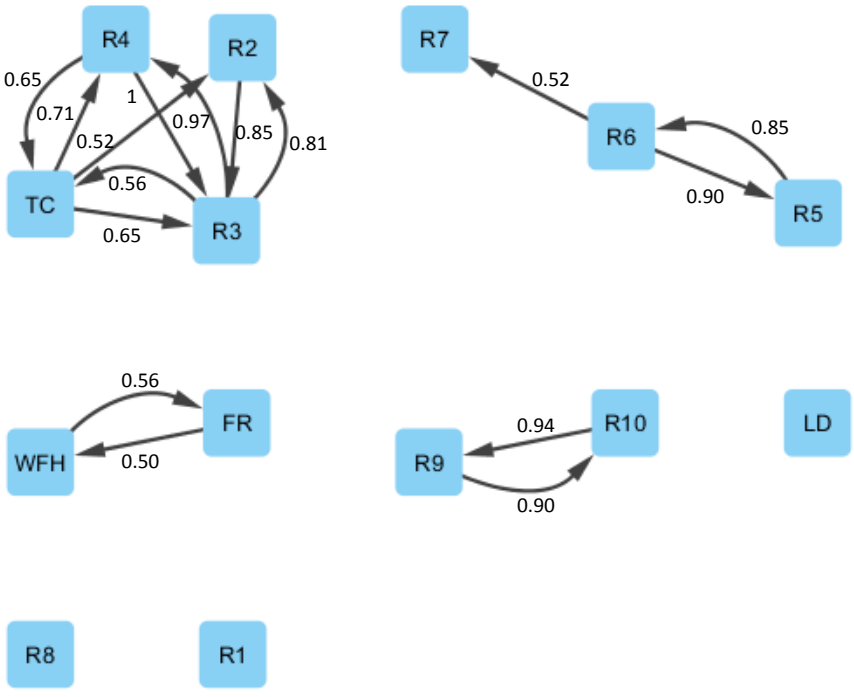


Figure 7: Principal Components Analysis (PCA) of individual redband trout based on allele frequencies. Individuals are color coded by the 13 sampling groups included in this study. The percent of variation explained by the two plotted dimensions is indicated on the horizontal and vertical axes.

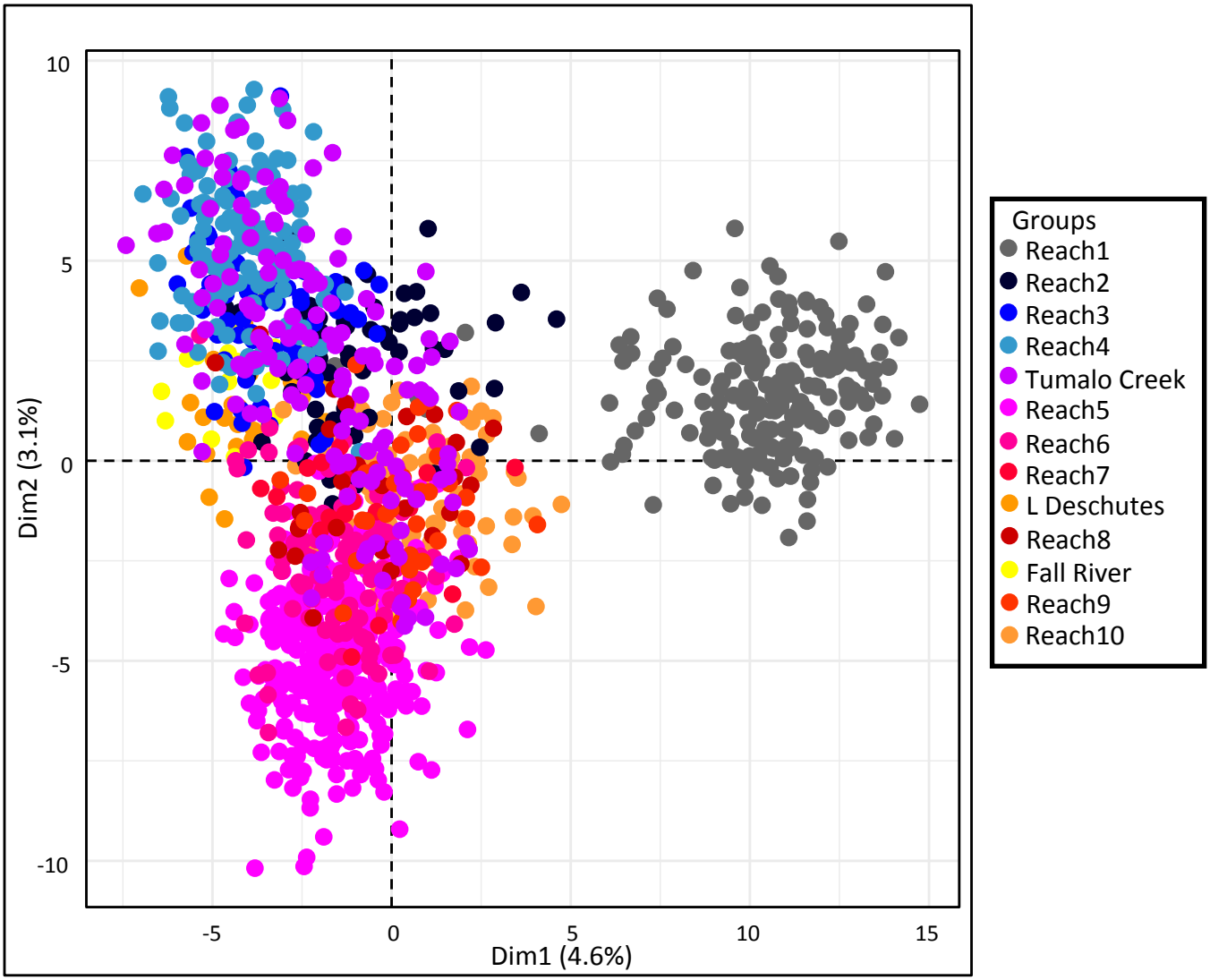


Figure 8: Plot of Bayesian Information Criterion (BIC) scores produced by *K*-means clustering with the number of estimated groups in the dataset (*K*) ranging from 1-30. All sampling groups of Deschutes redband were included in this analysis.

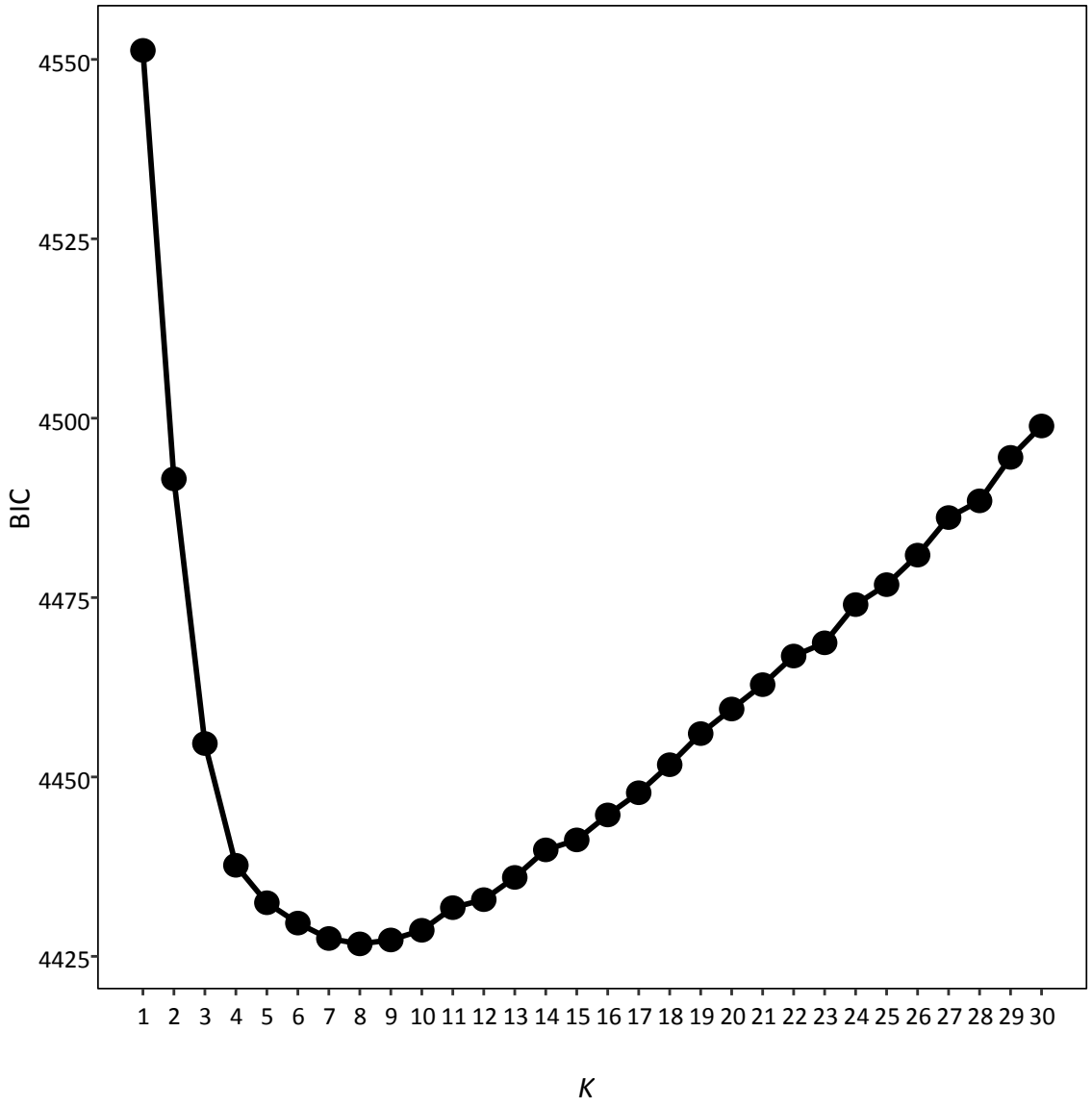


Figure 9: Barplots of membership probabilities for genetic groups identified via *K*-means clustering. The three plots reflect membership at *K*=6, 7, and 8. The genetic groups that were identified at each *K*-value are denoted by distinct colors. Each vertical bar represents an individual redband trout grouped according to sampling groups, which are separated by vertical dotted lines.

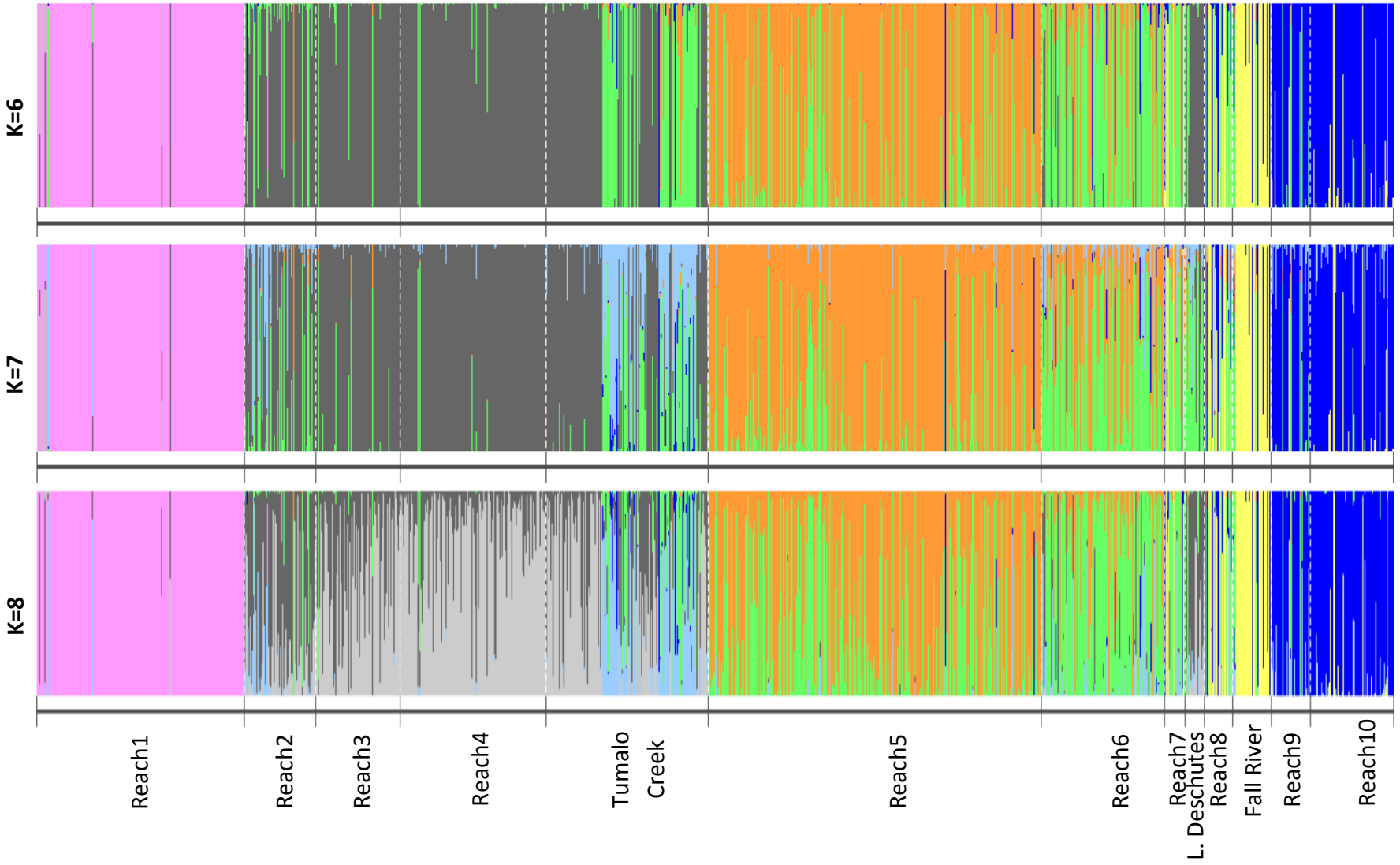


Figure 10: Plot of N_b values estimated for each sampling group. Results from both the linkage disequilibrium ("$N_{b\text{-}LD}$") and sibship ("$N_{b\text{-}Sibs}$") methods are presented. The y-axis is on a $\log_{15}$ scale. For the sampling units, "R"=Reach, "TC"=Tumalo Creek, "LD"=Little Deschutes, and 'FR"=Fall River.

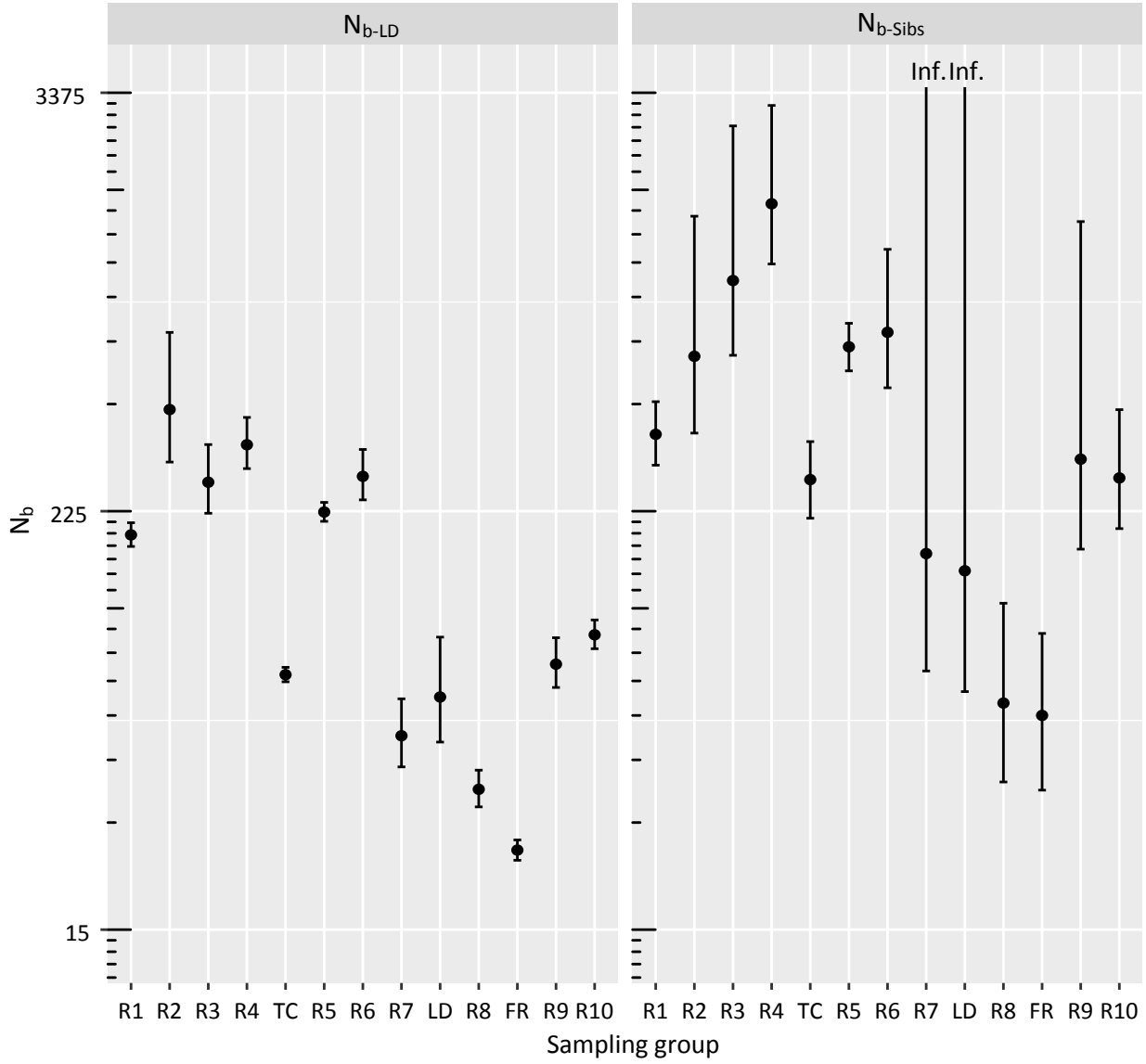


Figure 11: Plot of the relationship between sample size and number of family groups identified by COLONY2 for our sampling groups. The solid black line represents a fitted linear model based on the data and the dotted line represents an exact 1:1 relationship.

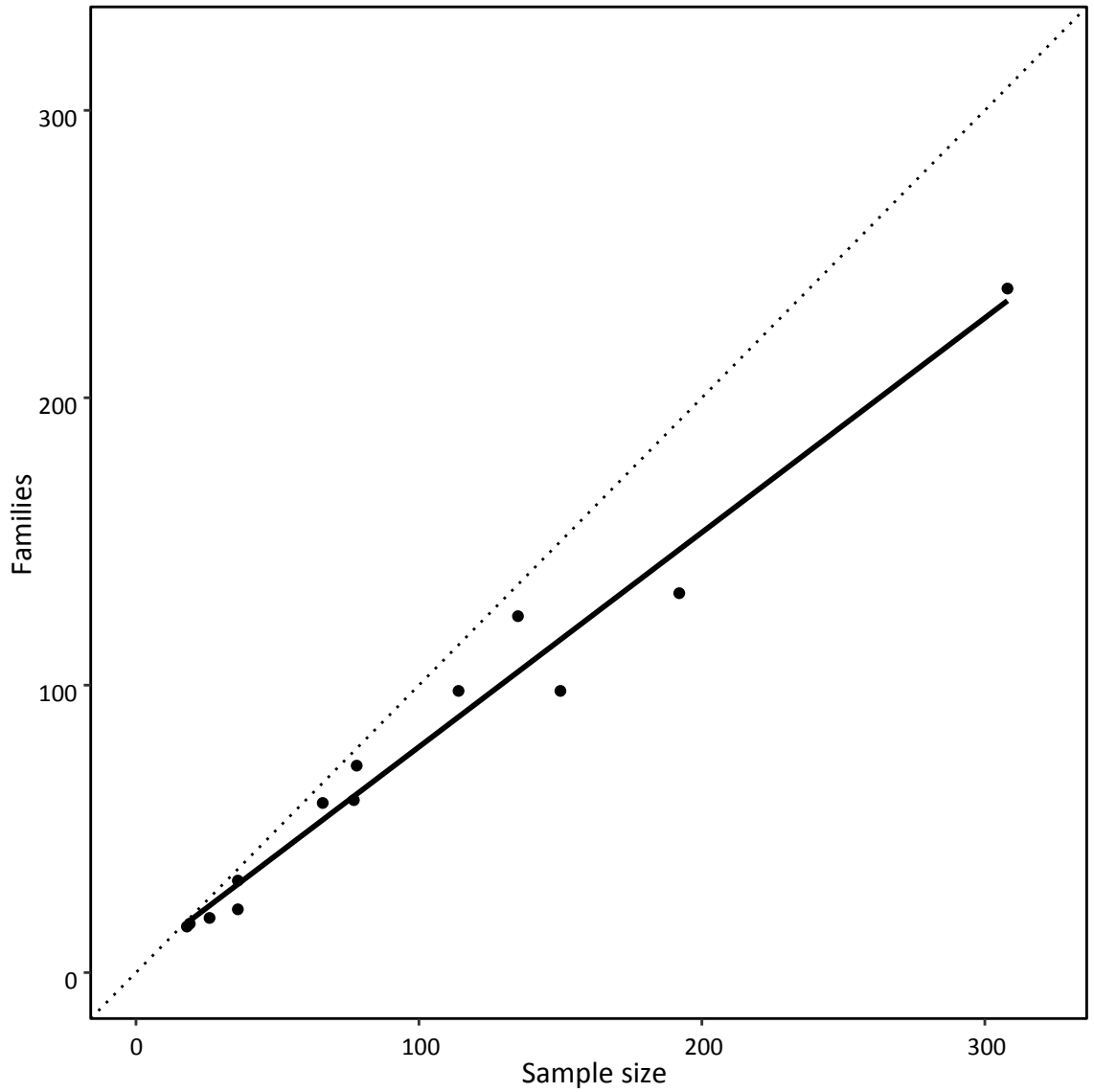
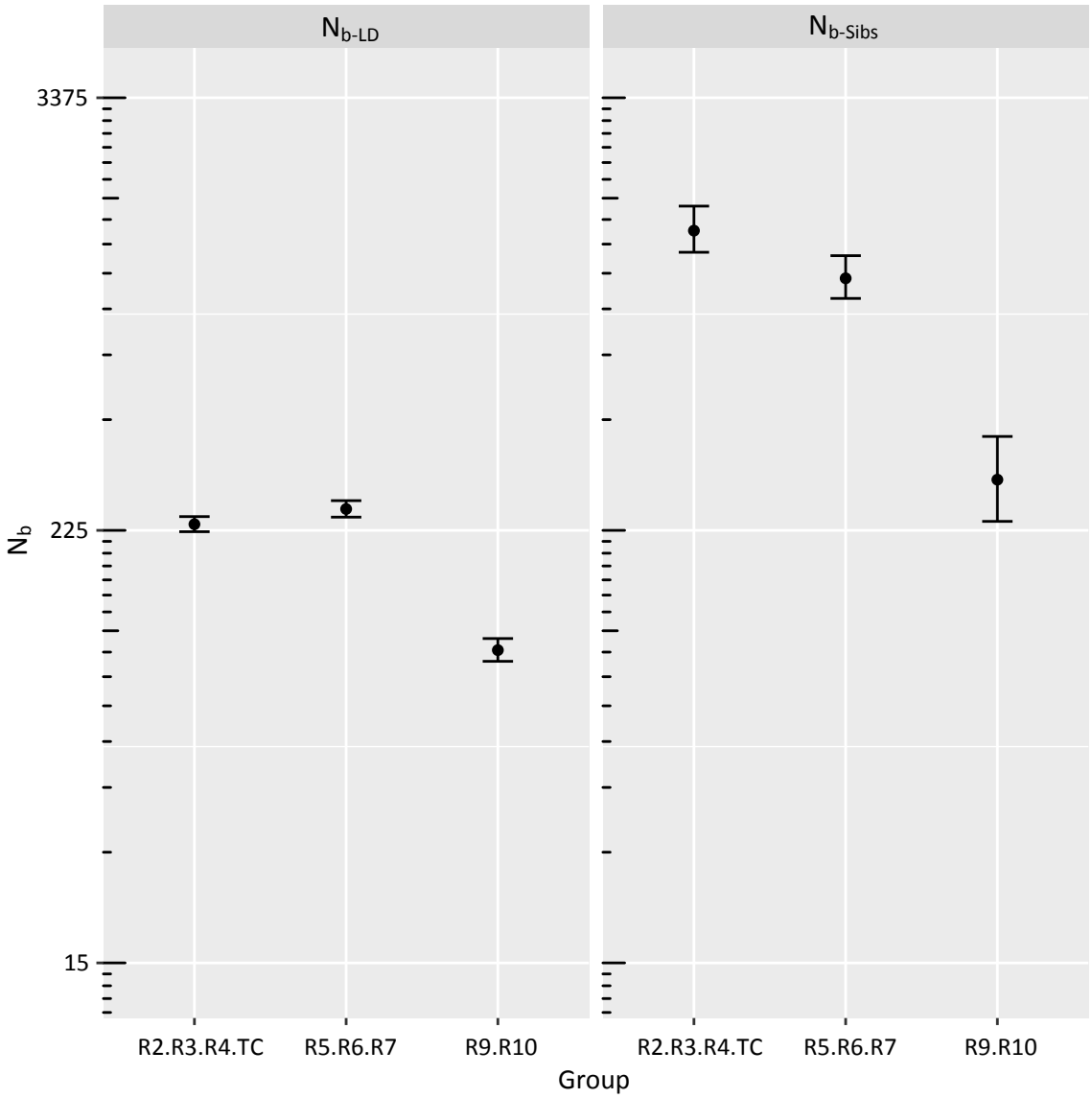


Figure 12: Plot of N_b values estimated for sampling groups that were combined based on patterns of genetic structure. Results from both the linkage disequilibrium ("$N_{b\text{-}LD}$") and sibship ("$N_{b\text{-}Sibs}$") methods are presented. The y-axis is on a $\log_{15}$ scale. For the sampling units, "R"=Reach.

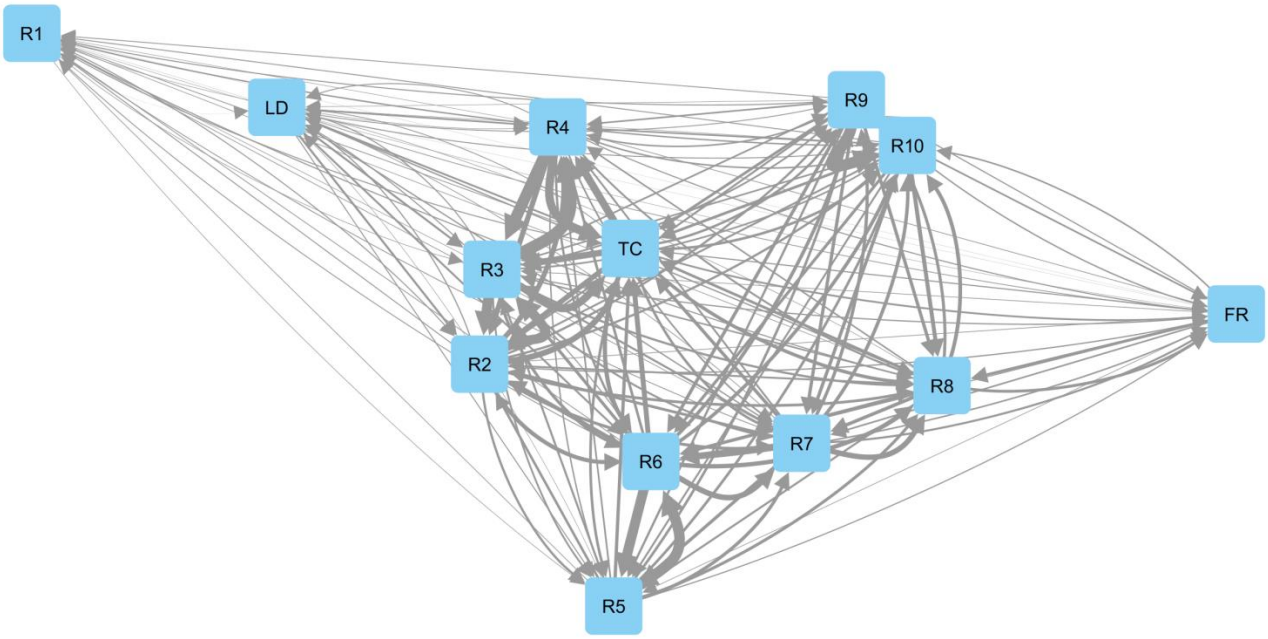


Appendix

Appendix 1: Matrix of pairwise G_{ST} values between sampling groups.

	R2	R3	R4	TC	R5	R6	R7	LD	R8	FR	R9	R10
R1	0.048	0.065	0.070	0.052	0.061	0.053	0.048	0.095	0.048	0.082	0.045	0.043
R2		0.005	0.012	0.011	0.025	0.015	0.014	0.027	0.022	0.044	0.024	0.027
R3			0.004	0.009	0.031	0.020	0.021	0.028	0.026	0.046	0.030	0.033
R4				0.008	0.037	0.025	0.026	0.033	0.029	0.047	0.033	0.035
TC					0.022	0.014	0.015	0.033	0.016	0.034	0.022	0.023
R5						0.006	0.014	0.052	0.022	0.051	0.023	0.027
R6							0.005	0.030	0.013	0.040	0.016	0.019
R7								0.033	0.004	0.023	0.011	0.012
LD									0.039	0.061	0.049	0.049
R8										0.016	0.013	0.012
FR											0.037	0.033
R9												0.002

Appendix 2: Network plot of relative migration levels between sampling groups. The arrows refer to the direction of gene flow from one group to another and the edge values reflect the level of gene flow. The width of the arrows corresponds to the estimate of migration: wider arrows indicate higher values and higher migration. The network is edge-weighted so that sampling groups with higher migration rates are closer in space. Sampling groups are coded with "R" standing for Reach, "TC" for Tumalo Creek, "LD" for Little Deschutes, and "FR" for Fall River.



Appendix 3: Pairwise matrix of relative migration values for all sampling groups. Values range from zero to one with higher values indicating higher gene flow. Values below the diagonal represent the amount of gene flow from the population in the vertical column to the population in the horizontal column. Values above the diagonal represent the amount of gene flow from the population in the horizontal column to the population in the vertical column.

	R1	R2	R3	R4	TC	R5	R6	R7	LD	R8	FR	R9	R10
R1	*	0.13	0.09	0.09	0.13	0.09	0.12	0.12	0.05	0.13	0.08	0.13	0.15
R2	0.12	*	0.81	0.45	0.52	0.23	0.37	0.28	0.14	0.26	0.14	0.22	0.22
R3	0.09	0.85	*	1.00	0.65	0.19	0.28	0.20	0.13	0.22	0.14	0.19	0.18
R4	0.08	0.46	0.97	*	0.71	0.16	0.23	0.18	0.12	0.20	0.13	0.17	0.16
TC	0.11	0.46	0.56	0.65	*	0.26	0.38	0.25	0.11	0.32	0.19	0.24	0.25
R5	0.10	0.25	0.21	0.18	0.31	*	0.90	0.31	0.09	0.28	0.14	0.25	0.26
R6	0.11	0.36	0.29	0.26	0.44	0.85	*	0.49	0.13	0.39	0.17	0.32	0.32
R7	0.10	0.26	0.21	0.19	0.31	0.30	0.52	*	0.10	0.47	0.21	0.30	0.32
LD	0.06	0.21	0.21	0.18	0.19	0.13	0.20	0.15	*	0.15	0.11	0.12	0.12
R8	0.10	0.19	0.17	0.17	0.26	0.21	0.31	0.32	0.09	*	0.28	0.29	0.34
FR	0.06	0.10	0.10	0.11	0.14	0.10	0.13	0.17	0.06	0.26	*	0.14	0.17
R9	0.12	0.21	0.18	0.17	0.26	0.23	0.31	0.31	0.09	0.31	0.15	*	0.94
R10	0.12	0.20	0.17	0.17	0.26	0.20	0.28	0.31	0.09	0.37	0.18	0.90	*

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