Development of a SNP genetic baseline for Chinook populations in Puget Sound

Developed for

PSC Southern Boundary Restoration and Enhancement Fund

Contract Number – Pacific Salmon Commission # SF-2008-I-15

WDFW Contract No. CAPS 08-1245

Report by

Kenneth I. Warheit[1], Todd Kassler, Jennifer Von Bargen, and Sewall Young

Washington Department of Fish and Wildlife, Fish Science Division, 600 Capitol Way N., Olympia, WA 98501

July 2011

[1] Kenneth.warheit@dfw.wa.gov

Introduction

Within the past decade Washington Department of Fish and Wildlife (WDFW) worked with a core set of other fish genetics laboratories to compile a microsatellite baseline for Chinook populations ranging from California to Alaska (GAPS Chinook baseline; Seeb *et al.* 2007). This baseline is used to identify genetically differentiated populations or population aggregates in many PSC and other managed fisheries in the eastern Pacific. Overall, this baseline performs well, but provides greater discriminatory power to populations aggregated into reporting groups than for the individual populations themselves. This is particularly true for late run (fall) Chinook stocks in Puget Sound where there has been a history of broodstock sharing among many hatcheries, resulting in a widespread distribution within Puget Sound of fish with Soos Creek (Green River) hatchery ancestry. The consequence of this management and other activities is relatively low power to differentiate many Chinook populations in Puget Sound using the GAPS microsatellite baseline.

During the past several years the GAPS consortium of labs has developed hundreds of single nucleotide polymorphism (SNP) loci for Chinook[2], and despite the existence of a functional microsatellite baseline, the GAPS consortium initiated this development of SNP markers for several reasons. First, the group recognized that there is a collection of populations within the coastwide database, similar to Puget Sound, where the microsatellites provide insufficient resolution to differentiate important fishery-related populations. SNPs developed to target these "problem" populations may provide the necessary genetic resolution to differentiate these stocks

[2] Laboratory work for this project was conducted 2009-early 2010 when there was approximately 100 SNP loci from the GAPS consortium. This report is based on these 96 of these SNPs. However, since early 2010, with advances in SNP discover methods, the number of validated Chinook SNPs is now several hundred, and continues to increase.

in mixed-stock fishery analyses. Second, SNP markers provide higher laboratory throughput at the same or lower cost than microsatellites. Third, the genotyping error rate for SNPs is lower than that for microsatellites, reducing laboratory error and technician time for allele scoring. Finally, SNPs require little standardization among contributing laboratories, essentially eliminating the time consuming and expensive practice of standardizing microsatellite alleles. As such, SNP markers are more amenable than microsatellites for large-scale fishery studies, especially management activities that require in-season analyses. Therefore, a SNP Chinook baseline for Puget Sound would enable quicker and less expensive genetic analyses of fisheries, and has the potential to be used to differentiate populations that cannot be distinguished using the existing microsatellite baseline.

The goal of this PSC-funded project is to build a SNP baseline for 27 Chinook populations in Puget Sound. These specific populations constitute most, perhaps all populations that we would expect to contribute to PSC or local fisheries. The purpose in developing this baseline is two-fold. First, targeted SNP data may help differentiate Puget Sound stocks that cannot be differentiated using the existing microsatellite baseline. Second, the escapement targets for Chinook populations in Puget Sound affect many PSC managed fisheries, and a comprehensive SNP baseline will provide fishery managers a more immediate and less expensive means to manage these fisheries than the use of a microsatellite-based baseline.

In this report we present a series of analyses on the Puget Sound Chinook SNP baseline, including tests of molecular equilibrium and phylogenetic hypotheses. However, the main purpose of this project was to generate the SNP baseline and to determine if this baseline is an

improvement over the microsatellite baseline in its power to differentiate populations in a mixed-fishery analysis. Although the baseline data itself is not included as part of this report, it is available upon request and will be included in the GAPS Chinook SNP baseline.

Methods

Samples

We analyzed 3551 samples from 81 collections representing 31 populations of Chinook salmon from Washington State (Table 1, Figure 1). Twenty-one collections are from Puget Sound proper; four populations are from Hood Canal, including three populations from the Skokomish Basin; and two populations each are from the Strait of Juan de Fuca (Elwha_H and Dungeness_R), Washington outer coast (NF_Makah_H and MathenyCr/Queets), and upper Columbia River (WenatcheeR_SU and Chiwawa_Sp). The N.F. Nooksack, N.F. Stillaguamish, and the Skykomish populations are composed of both hatchery- and natural-origin collections (the hatcheries in these systems use both hatchery- and natural-origin fish as broodstock). We maintained the Clear Creek Hatchery and Nisqually River as separate populations because the Clear Creek Hatchery broodstock is segregated from the natural run within the Nisqually River. Although the George Adams Hatchery uses both hatchery- and natural-origin fish for broodstock, we kept the hatchery collection as a separate population from the two natural populations in the Skokomish Basin because it was unclear to us at the start of this analysis if the N.F. and S.F. Skokomish populations are distinct.

The Puget Sound Chinook evolutionary significant unit (ESU) is defined as all naturally spawning populations of Chinook in Puget Sound proper, Hood Canal, and in the Strait of Juan

de Fuca from the Elwha River east (Ruckelshaus *et al*. 2006). Therefore, for this analysis, we considered as Puget Sound all hatchery- and natural-origin populations occurring within the boundaries of this ESU. We also included as outgroups four populations from outside this ESU: NF_Makah_H and MathenyCr from the Washington Coast, and WenatcheeR_Su and Chiwawa_Sp from the upper Columbia River.

Laboratory Procedures

DNA Extraction. We used two procedures to extract DNA from samples. For most samples, we extracted genomic DNA from a small piece of fin tissue (approximately 2 mm) using Macherey-Nagel nucleospin tissue kits following the manufacturer's recommended procedures, and eluted to a final volume of 100 µL. Alternatively, we extracted DNA by digesting a small piece of fin tissue in a 5% chelex (BioRad Chelex 100 resin) solution containing 0.4 mg proteinase K (Sigma). Following digestion at 65 degrees C for 180 min, the samples were heated for 10 minutes at 95 degrees C to denature proteins.

SNP amplification and visualization. We genotyped 96 single nucleotide polymorphisms (SNPs), 75 of which were part of the set standardized by the GAPS consortium of labs (Narum *et al.* 2008, Seeb *et al.* 2007) (Tables 2 and 3), using either an Applied Biosystem (AB) 7900HT Fast Real-Time PCR System or a Fluidigm EP1 System. For those samples genotyped using the AB7900HT, we dried within a 384 well plate 5µL of DNA extract diluted 1:50 with water yielding approximately 2.5 ng of template DNA for each sample. The dried template was re-hydrated using 0.25X Taqman Universal PCR Master Mix (Applied Biosystems), and 1X of custom Taqman assay (Applied Biosystems) to a final reaction volume of 5 µL. Each of the 96

SNPs were PCR amplified separately using either MJ Research PTC-200 or Applied Biosystems 9700 thermocyclers, with the following thermal profile: initial denature step of 10 minutes at 95°C; 40 cycles of 15 seconds at 94°C, 1 minute at 60°C, followed by a final indefinite holding step at 10°C. Following amplification, SNPs were visualized and analyzed using the AB 7900HT and AB SDS 2.3 software.

Most samples were genotyped using Fluidigm EP1 System. Here, we produced two mixtures. First, for the assay loading mixture we combined 1X Assay Loading Reagent (Fluidigm), 2.5X ROX Reference Dye (Invitrogen), and 10X custom Taqman assay (Applied Biosystems) to a final reaction volume of 5 µL. Second, for the sample loading mixture we combined 1X Taqman Universal PCR Master Mix (Applied Biosystems), 0.05X AmpliTaq Gold DNA Polymerase (Applied Biosystems), 1X GT Sampling Loading Reagent (Fluidigm) and 2.1 µL template DNA to a final reaction volume of 5 µL. We used 96.96 dynamic array integrated fluidic circuits (chips; Fluidigm), and pipetted 4 µL of assay loading mix into each of the 96 wells on the assay side of a chip and 5 µL of the sample loading mix into each of the 96 wells on the sample side of a chip. We then placed each chip into the Fluidigm IFC Controller to combine the assay and sample loading mixtures, followed by PCR amplification using Fluidigm's Eppendorf thermal cycler. The PCR reaction required an initial set of steps of 70°C for 30min, 25°C for 5 min, 52.3° for 10 sec, 50.1°C for 1 min 50sec, 98°C for 5 sec, 96°C for 9 min 55 sec, 96°C for 15 sec, 58.6°C for 8 sec, and 60.1°C for 43 sec, followed by 40 cycles of 58.6°C for 10 sec, 96°C for 5 sec, 58.6°C for 8 sec and 60.1°C for 43 sec with a final hold at 20°C. Following amplification, SNPs were visualized using the Fluidigm EP1 and the BioMark Data Collection software (Fluidigm), and analyzed using Fluidigm SNP Genotyping Analysis software.

We used a “pre-amp” step, prior to the PCR amplification described above, for samples that produced weak or no initial amplification following the Fluidigm EP1 protocol. For this pre-amp protocol, established by Fluidigm (Fluidigm Specific Target Amplification Quick Reference), we combined 1X Buffer (Qiagen) and a primer mix containing all 96 Taqman assays at a final 0.2µM, with 1.25µL of DNA template, to a final volume of 5 µL per reaction. This mixture was then PCR amplified using either MJ Research PTC-200 or Applied Biosystems 9700 thermocyclers at 95.0°C for 15 min followed by 10 cycles of 94.0°C for 30 sec and 60.0°C for 3 min, ending with a 10°C hold. Products were then diluted 1:5 with 1X TE, and used as template in the sample loading mixture.

Statistical Analyses

We calculated allele frequencies for all SNPs and populations using the program PowerMarker (Liu and Muse 2005). We used Genepop (v 4.0.9; Raymond and Rousset 1995, Rousset 2008) to calculate Hardy-Weinberg equilibrium (probability test) and linkage disequilibrium for each locus (or pair of loci for the linkage test) in each population with 10000 dememorization steps, 20 batches, and 5000 iterations per batch. P-values for each test were combined across all loci, groups of loci, populations, or groups of populations (depending on the test) using Fisher’s method ($X^2 = -2\sum_{i=1}^{k} \log_e(p_i)$, where p^i is the p-value for the ith test, and X^2 has a chi-square distribution with 2k degrees of freedom). We used MatLab (v. 2010a; The MathWorks) to implement Fisher’s method.

We conducted phylogenetic analyses of all populations using chord (Cavalli-Sforza and Edwards 1967) and allele-sharing (Jin and Chakraborty 1994) pairwise distances. We bootstrap resampled the dataset 1000 times by randomly selecting, with replacement, SNP loci while maintaining the original number of loci in each analysis. Following each bootstrap sample, both distance measures were calculated, and a phylogenetic tree was constructed based on each distance matrix using the neighbor-joining method (Saitou and Nei 1987). Therefore, for each pairwise distance method we produced 1000 trees corresponding to the 1000 bootstrap resampled datasets. Finally, we constructed 50% consensus neighbor-joining trees for each distance from the 1000 trees. We used PowerMarker (Liu and Muse 2005) to calculate distance measures and to conduct the bootstrap resampling, and MEGA version 4 (Tamura *et al*. 2007) to construct trees.

The two alleles associated with each SNP are complementary. That is, the frequency of one allele is 1 minus the frequency of the other allele, and therefore, only one allele per SNP is needed to define that SNP within each population. We constructed a 27 by 68 Puget Sound population by SNP allele frequency matrix using only the major allele (i.e., the allele from each locus with the higher frequency calculated across all populations) for each SNP and converted this matrix to a mean-centered allele frequency matrix. Mean-centered allele frequencies are computed by subtracting from the allele frequency for each SNP in each population the mean frequency for each SNP calculated across all 27 Puget Sound populations. We then constructed a variance-covariance matrix from the mean-centered matrix, and subjected this matrix to a principal component analysis (PCA). We produced three sets of data from the PCA; for each principal component we computed (1) scores, which are weighted linear combination of the

variables for each individual (here, weighted combination of SNP major allele frequencies for each population), (2) coefficients (or eigenvectors) for each variable (SNP), which are the weights associated with the PC scores, and (3) eigenvalue, which is the variance associated with that principal component; a sum of all eigenvalues across all principal components equals the total variance associated with the original variables. To more easily interpret the coefficients, we transformed the coefficients into product-moment correlations ("loadings") between each SNP and the respective principal component by multiplying each coefficient by the square-root of the eigenvalue for that component (e.g., PC axis 1) and then dividing that product by the standard deviation of the original SNP calculated across all populations (see Dillon and Goldstein 1984).

We plotted scores for each of the Puget Sound populations for the first four principal components. To determine where the outgroup populations would be plotted with respect to the ordination of the Puget Sound populations, given the principal component coefficients calculated using only the Puget Sound populations, we first subtracted from each of the outgroup populations' major allele frequency the mean Puget Sound allele frequency for each SNP, as we had done for each Puget Sound population. We then multiplied this mean-centered outgroup allele frequency matrix by the matrix of coefficients. This product produced principal component scores for the outgroup populations, based on the SNP covariance structure of the Puget Sound populations, enabling us to project the outgroup populations' scores onto Puget Sound Chinook principal component space.

We conducted F_{ST} outlier tests, using two different models, to evaluate if SNP loci showed higher population differentiation than expected based on neutral variation, given a specific level

of heterozygosity. Loci with outlier F_{ST} values are candidates for directional selection (Lewontin and Krakauer 1973, Excoffier *et al* 2009). We used both the finite island model, which assumes that all populations are independent (Beaumont and Nichols 1996) and the hierarchical island model (Slatkin and Voelm 1991), which will reduce the number of false outlier loci compared with the finite island model, if the populations are structured hierarchically (Excoffier *et al* 2009). Both models require diploid data, so for these analyses we removed AOts015, which is a mitochondrial SNP. The hierarchical island model also requires that you *a priori* aggregate populations into more than one group. We designated population groups based on the results of phylogenetic and principal component analyses (see Results and Discussion section). We implemented all F_{ST} outlier tests using *ARLEQUIN* (v. 3.5; Excoffier and Lischer 2010). For the finite island model we used 20,000 simulations and 100 simulated demes, while for the hierarchical model we used 50,000 simulations, 10 simulated groups, and 100 simulated demes per simulated group.

Finally, we used ONCOR (S. Kalinowski, unpublished) to calculated the power of the 68-SNP dataset to accurately identify each Puget Sound population in a mixed-stock fishery analysis (MSA). We compared the results from the SNP power analysis with a similar analysis we conducted with the GAPS microsatellite baseline for the same set of populations. For both the SNP and microsatellite analyses we used the Anderson *et al.* (2008) 100% simulation procedure in the mixture analysis module in ONCOR, which simulates a fishery sample composed entirely (100%) of individuals from a single population, and a new baseline from the original baseline, maintaining the same sample size per population as the original baseline.. The composition (mixture proportions) of the fishery sample is estimated using a conditional maximum likelihood

procedure (Millar 1987) and the expectation-maximization (EM) algorithm. The process is iterative in that the mixture proportion for each run is used as priors for each successive run, and the process is repeated until the proportions stabilize. The EM algorithm starts with uniform priors. The 100% simulation process is repeated for each Puget Sound population, with each fishery sample composed of 200 individuals. To assess variability in our power estimate we repeated the entire process (100% simulations for each population) 100 and 50 times, for the SNP and microsatellite database, respectively.

Results and Discussion

SNP Loci

Although our initial analysis included 96 SNP loci, we used only 68 of the 96 SNPs for final analyses (Table 2). Of the 28 SNPs eliminated from the analysis, 19 are part of the set standardized by the GAPS consortium of labs. We eliminated a SNP if it was invariant throughout the dataset, if its minor allele frequency was less than 0.01, if it was missing (not scored) from at least one entire population, or if it was missing from greater than 50% of all samples (Table 3). All SNPs are of nuclear-origin, except AOts015 (Ots_C3N3), which is of mitochondrial-origin. One SNP locus was inadvertently scored twice: AOts032 (Ots_HSP90B-100) was incorrectly logged into the WDFW database twice, correctly as Ots_HSP90B-100, and incorrectly as AOts082 (Ots_HSP90B.100). This error was not discovered until most analyses were completed, and as a result, this SNP was included twice in all analyses.

Nineteen of the 68 SNPs are invariant in at least one Puget Sound population, six SNPs are invariant in at least 12 Puget Sound populations, three SNPs are invariant in 19 Puget Sound

populations, and one SNP is in variant in 26 of the 27 Puget Sound populations (Table 2). The minor alleles are defined based on the 27 Puget Sound populations and are specified in Table 2. Within Puget Sound populations and across all SNPs, minor allele frequencies range from 0.0% to 86.2%, but if the four outgroup populations are included, minor allele frequencies range from 0.0% to 100.0% (Table 2). The median minor allele frequencies across all 31 populations range from 0.0% to 48.1%.

Thirty-three of the 67 nuclear SNPs show Hardy-Weinberg equilibrium $P < 0.05$. However, across 32 of these 33 loci (excluding AOts086), the percentage of Puget Sound populations with Hardy-Weinberg equilibrium $P < 0.05$ and $P < 0.05$-adjusted[3] range 14.8 – 29.6% and 3.7 – 8.7%, respectively (Table 2). Although these results do suggest a degree of disequilibrium with some loci, that degree of disequilibrium is considerably less than that in AOts086 (Ots_MetA), where all 27 Puget Sound populations are in Hardy-Weinberg disequilibrium. The median observed heterozygosity for this locus is 0.10, compared with a median expected heterozygosity of 0.39. This locus is not part of the set that was standardized by the GAPS consortium of labs.

Pairwise linkage (gametic) disequilibrium among these SNPs is relatively rare, with only 6% of the 61,506 pairwise combinations of 68 loci, across 27 Puget Sound populations, showing linkage disequilibrium at $P < 0.05$. For each specific locus-pair, except AOts032 and AOts082, which are the same locus and therefore linkage $P = 0.00$, the median (and maximum) percentage

[3] We used Bonferroni adjusted P-values (0.05/number of independent tests) to account for experiment-wise error rate. Because some loci are fixed in some populations, the number of independent tests is not constant across all loci or populations. For each locus, the P-value was adjusted using the number of populations for which a test statistic was calculated; this ranged from one (i.e., no adjustment; AOts046) to 27, with a median of 27 populations (out of 27 populations). For each population, the P-value was adjusted using the number of loci for which a test statistic was calculated; this ranged from 54 to 67, with a median of 61 loci (out of 67 nuclear loci).

of populations with linkage P < 0.05 and P < 0.05-adjusted, across all locus-pairs, is 4.0% (33.3%) and 0.0% (33.3%), respectively. For each individual locus, the probability of pairwise linkage disequilibrium, across all populations, rangedfrom 0.00 (AOts001 and AOts007) to 1.00 (20 loci), with a median probability of P = 0.559 (Table 2).

Populations

The observed and expected heterozygosities median (and range) across all populations is 0.26 (0.19 – 0.30) and 0.27 (0.25 – 0.29) respectively. Although 17 and 11 populations show Hardy-Weinberg equilibrium probabilities of P < 0.05 and P < 0.05-adjusted, respectively, only NF_Stillaguamish population stands out as having an exceptionally high chi-square statistic, and a large number of loci with Hardy-Weinberg equilibrium probabilities of P < 0.05 and P < 0.05-adjusted (Table 4). Following NF_Stillaguamish, GeorgeAdams_H has the next highest number of loci with Hardy-Weinberg equilibrium probabilities of P < 0.05 and P < 0.05-adjusted. NF_Stillaguamish has the lowest, and GeorgeAdams_H has the highest observed heterozygosities, while NF_Stillaguamish and Soos_H has the largest differences between observed and expected heterozygosities (Table 4).

Six populations have pairwise linkage disequilibrium P < 0.0000003 (SF_Nooksack, L_Skagit_R_Fa, NF_Stillaguamish, UW_H_SuFa, White_H, Elwha_H), while the remaining populations have pairwise linkage disequilibrium P > 0.50 (Table 4). The median (and maximum) percentage of locus-pairs with linkage P < 0.05 and P < 0.05-adjusted for these six populations are 8.5% (10.3%) and 0.1% (0.1%), compared with 4.7% (6.4%) and 0.0% (0.1%) for the remaining 21 populations (Table 4). Except AOts032 and AOts082 (see above), no pair

of loci are in linkage disequilibrium for each of the six populations; however, in AOts025 and AOts067 linkage $P < 0.05$ for NF_Stillaguamish, UW_H_SuFa, White_H, and Elwha_H, but $P > 0.93$ for SF_Nooksack and L_Skagit_R_Fa. With the exception of AOts032 and AOts082, no two pairs of SNP loci have significant linkage disequilibrium across a broad range of populations, and the higher (but minor) overall linkage disequilibrium in SF_Nooksack, L_Skagit_R_Fa, NF_Stillaguamish, UW_H_SuFa, White_H, Elwha_H suggests more a potential problem with the collections themselves (e.g., presence of admixture or family structure) than with the SNP loci.

Phylogenetic Analysis

We rooted the two consensus trees using the four Washington outer coast and upper Columbia River populations, assuming that the Puget Sound populations are monophyletic with respect to these four outgroup populations. Although we did not attempt to test explicitly the hypothesis of Puget Sound monophyly there is evidence that the Washington coastal and upper Columbia Chinook lineages are not derived from within Puget Sound and that rooting the Puget Sound trees with these outgroup populations is consistent with previously published phylogenetic hypotheses (Waples *et al.* 2004; Beacham *et al.* 2006). The structure of the 50% consensus trees derived from chord and allele-sharing distances are nearly identical, with only two differences (U_Skagit_Su/NF_Stillaguamish and SamishFall/Issaquah_Cr_SuFa nodes), and are presented as a single tree (Figure 2). Three monophyletic groups within Puget Sound, with at least 50% bootstrap support, are evident: north Puget Sound spring/summer, south Puget Sound/Hood Canal mostly fall, but also includes White River spring, and Strait of Juan de Fuca populations. This overall structure is generally consistent with that in Warheit *et al.* (2007) and Ruckelshaus

et al. (2006), who used microsatellites and allozymes, respectively, and suggests that the phylogenetic hypothesis presented in Figure 2 is consistent across a range of molecular markers and should be considered the primary hypothesis to be tested by further analyses. The north Puget Sound spring/summer group is further divided into a Nooksack group and a group consisting of spring/summer-run populations from the Skagit and NF Stillaguamish River. This later group consists, in part, of a collection of populations from rivers draining into the Whidbey Island basin. However, three populations from this same area are not included in this group: the fall-run populations from the lower Skagit and SF Stillaguamish Rivers, and the summer-run hatchery populations form the Skykomish River. All three of these populations remain unresolved in our consensus tree (Figure 2). The collection of mostly hatchery fall-run populations from south Puget Sound and Hood Canal form a well defined, but poorly resolved group. Here, the only two nodes that are resolved differentiate the spring-run White River hatchery population from the fall-run populations, and the George Adams hatchery and Hamma Hamma lineages, both from Hood Canal. Finally, the two populations the from the Strait of Juan de Fuca form a lineage with 99% bootstrap support, although the Elwha_H population is fall-run and the Dungeness_R population is spring/summer-run (Table 1, Figure 2).

Principal Component Analysis

The first two principal components from an analysis of the 27 Puget Sound populations account for a 64% of the total variance across all 68 SNP loci and clearly separate spring/summer north Puget Sound (NPSSpSu), fall south Puget Sound (SPSFall), and Strait of Juan de Fuca (SJF) Chinook populations (Figure 3; see Figure 2 for populations included in each of these groups). The L_Skagit_R_Fa (fall), SF_Stillaguamish (fall), and Skykomish_H_SU (summer) populations

are intermediate between the NPSSpSu and SPSFall clusters, and the White_H (spring) population plots near but is not within the SPSFall cluster. This is consistent with the phylogenetic analysis discussed above (Figure 2). K-means clustering (Dillon and Goldstein 1984) with K = 3 partitions the populations into three groups, corresponding to NPSSpSu, SPSFall, and SJF clusters, but places L_Skagit_R_Fa, SF_Stillaguamish, and Skykomish_H_SU within the NPSSpSu cluster. K-means clustering with K = 4 maintains the NPSSpSu, SPSFall, and SJF clusters, and places L_Skagit_R_Fa, SF_Stillaguamish, and Skykomish_H_SU into a fourth independent cluster. The centroids for each of the NPSSpSu, SPSFall, and SJF clusters occur at the ends of the vectors that originate at the origin (0,0; Figure 3). The NPSSpSu and SPSFall vectors are nearly complementary (159 degrees), highly correlated (r^2 = -0.94), and are coincident with PC Axis 1. The SNPs with the largest positive or smallest negative coefficients and $-0.80 \geq r^2 \geq 0.80$ with PC Axis 1 are AOts010, AOts012, AOts015, AOts040, AOts050, AOts053, AOts055, and AOts060 (Table 5). These SNPs have a large influence of the ordination of populations along PC Axis 1 and are important in differentiating NPSSpSu from SPSFall populations. The SJF vector is nearly orthogonal with both the NPSSpSu (89 degrees, r^2 = 0.02) and SPSFall (111 degrees, r^2 = -0.36) vectors, and is nearly coincident with PC Axis 2. The SNPs with the largest positive or negative coefficients and $-0.80 \geq r^2 \geq 0.80$ with PC Axis 2 are AOts003, AOts016, and AOts031 (Table 5). These SNPs have a large influence of the ordination of populations along PC Axis 2 and are important in differentiating SJF from NPSSpSu and SPSFall populations.

Principal components three and four account for 8% and 5% of the total variance, respectively, across all 68 SNP loci and separate the Nooksack and White River populations from the other

Puget Sound populations (Figure 4). The vector leading to the White River population is nearly orthogonal with the vector leading to the centroid between the two Nooksack populations (96 degrees, $r^2 = -0.11$) and these two vectors are coincident with the PC Axis 3 and 4, respectively. Overall, the correlation coefficients for PC Axes 3 and 4 are lower than those for Axes 1 and 2, reflecting the decreasing variance accounted for by each principal component moving from Axis 1 to 4, and none of the SNPs show significant correlation ($-0.80 \geq r^2 \geq 0.80$) with either Axes 3 or 4 (Table 5).

The four outgroup populations projected onto the first two Puget Sound principal components are plotted on the NPSSpSu and SJF side of PC Axis 1. The ordinations of two of these populations (NF_Makah _H and WenatcheeR_SU) are correlated with the NPSSpSu vector ($r^2 =$ 0.99 and 0.87, respectively) and negatively correlated with the SPSFall vector ($r^2 = -0.88$ and -0.99, respectively). This suggests that the allelic covariance structure that generated Figure 3 and separates the SPSFall populations from both the NPSSpSu and SJF populations along PC Axis 1 is derived in the SPSFall populations. That is, the difference between SPSFall, and NPSSpSu and SJF populations along PC Axis 1 is a function of an evolutionary change in the covariance among the SNP loci in the SPSFall populations from a state that exists in the NPSSpSu, SJF, and outgroup populations.

All four outgroup populations projected onto Puget Sound principal components 3 and 4 cluster with the two Nooksack populations, with the Matheny Creek population plotted almost on top of the Nooksack centroid (end of the Nooksack vector). This suggest that the allelic covariance structure that generated Figure 4 and separates Nooksack populations from the other Puget

Sound populations is derived, compared with that of the outgroup populations, in all but the Nooksack populations. This hypothesis is inconsistent with the phylogenetic hypothesis presented in Figure 2 and indicates that the covariance structure in the Nooksack populations is convergent with the outgroup populations, or that the phylogeny in Figure 2 is incorrect, and that the rest of the NPSSpSu populations, SPSFall and White_H populations, and SJF populations are more closely related to each other than any are to the Nooksack populations. One of the differences between Fig 2 and the phylogeny presented in Ruckelshaus *et al.* (2006) is the placement of the Nooksack populations. In Ruckelshaus *et al.* (2006), the Nooksack populations are the sister group to the NPSSpSu, SPSFall, and White _H monophyletic group, and as with Figure 2, the SJF populations are the sister group to the rest of the Puget Sound populations.

Detecting loci under selection

For the hierarchical island model (Slatkin and Voelm 1991) we assembled the Puget Sound populations into three groups, corresponding to the NPSSpSu, SPSFall + White_H., and SJF groups in Figures 2 and 3, leaving out of the analysis the outgroups, and L_Skagit_R_Fa, SF_Stillaguamish, and Skykomish_H_SU, the three populations intermediate between NPSSpSu, SPSFall. The finite island model (Beaumont and Nichols 1996) does not assume hierarchical structure, and for this model we included all 27 Puget Sound populations. In addition, although *ARLEQUIN* (v. 3.5; Excoffier and Lischer 2010) reports both the F_{ST} and F_{CT} statistics for the hierarchical island model, we describe here only the F_{CT} results. F_{ST} measures the proportion of total variance that can be attributed to differences among populations, without regard to the hierarchical structure, while F_{CT} measures the proportion of total variance than can be attributed to differences among groups only (e.g., among NPSSpSu, SPSFall + White_H., and SJF).

Overall, there is little difference between the results from the hierarchical and finite island models (Figure 5), both suggest that AOts050 and AOts055 are the only SNPs that are candidates for being influenced by natural selection. These two SNPs show F_{ST} and F_{CT} values that are significantly greater ($p > 0.99$ and $P > 0.95$) than would be expected from a neutral model and heterozygosities, in the finite and hierarchical island models, respectively. AOts050 and AOts055 are among the eight SNPs that are highly correlated with PC Axis 1, and suggests that some of the differentiation among the populations (and between NPSSpSu and SPSFall groups) along PC Axis 1 may be a function of current or past selective pressures.

Differentiating populations in a mixed-stock fishery analysis (MSA)

Both the SNP and microsatellite data performed poorly in simulated MSA, with only seven of the 27 populations, for both datasets, showing mean estimated proportions equal to or greater than 0.90 (Figure 6). Since each simulated fishery was composed of only a single population, if either the microsatellite or SNP data were able to differentiate all Puget Sound populations, without error, the results from the simulated MSA for each population would have been 1.0. The median (and range) mean estimated proportions for the Puget Sound populations are 0.65 (0.06 – 0.99) and 0.66 (0.17 – 0.99) for the SNPs and microsatellites, respectively. The performance for the Issaquah_Cr_SuFa, NF_Skokomish_R_Fa, and SF_Skokomish_R populations was particularly poor for both the microsatellite and SNP data (Figure 6). When we aggregated the populations into four groups (NPSSpSu, SPSFall, SJF, and a group composed of Skagit_R_Fa, SF_Stillaguamish, and Skykomish_H_SU) 25 of the 27 populations, for the SNP dataset, showed mean estimated proportions equal to or greater than 0.90. The median (and range) estimated proportions for the Puget Sound populations in their aggregates was 0.99 (0.77 – 0.998). The

three unresolved populations from the phylogenetic analysis (Figure 2) had the three lowest mean estimated proportions at 0.77, 0.88, and 0.90, for SF_Stillaguamish, Skykomish_H_SU, and Skagit_R_Fa, respectively. This would be expected given their ambiguous phylogenetic relationships (Figure 2), and intermediate and ungroup position in the principal component analysis (Figure 3).

Overall, of the 26 populations with both SNP and microsatellite data[4] the SNP data show increased power for 12 populations to be differentiated within a mixed-stock fishery analysis, decreased power for seven populations, and no change for seven populations (Figure 6). However, the distribution of these results is not uniform throughout Puget Sound. The SNP data provide greater population resolution than the microsatellite data among the SPSFall with increase power for eight of the 14 populations, decreased power for three populations, and no change for four populations (Figure 6). However, despite the increase in resolution, the median estimated proportions for the SPSFall populations is only 0.48, indicating that on average for these populations their estimated proportion within a mixed-stock fishery analysis will be half of its true value. These populations are mostly hatchery populations with a very recent shared ancestry as a result of extensive broodstock sharing, primarily from the Soos_H populations. These populations are also tightly clustered in the principal component analysis (Figure 3) and are unresolved in the phylogenetic analysis (Figure 2). Despite its origin from Soos_H, the mean estimated proportion for the UW_H_SuFa population is 0.96. This hatchery population is an inbreed experimental population from the University of Washington and as a result, has a distinct allele frequency among the SPSFall populations.

[4] At the time of this analysis, the microsatellite genotyping of the SF_Stillaguamish population had not been completed and therefore was not included here.

The SNP data do not provide greater population resolution than the microsatellite data among the NPSSpSu populations, with only two of the seven populations (Marblemount_H_Sp and NF_Stillaguamish) showing increased power to be differentiated within a mixed-stock fishery analysis (Figure 6). However, these seven populations can be more readily differentiated in a mixed-stock fishery analysis than the SPSFall populations, with median estimated proportions equal to 0.77 and 0.89 for the SNP and microsatellite datasets, respectively.

Conclusions

The Puget Sound Chinook SNP baseline is a work in progress. We met the goals of this project in that we generated a SNP baseline for 27 populations from Puget Sound, extensively analyzed this baseline using several molecular genetic metrices, and tested and compared the power of the baseline to differentiate the Puget Sound populations in mix-stock fishery analyses. However, these 68 SNPs did not provide us with an improved ability to differentiate populations beyond that in the microsatellite baseline, and in some respects, at least for some of the NPSSpSu populations, the microsatellite baseline data provided better resolution than the SNP data. The bulk of the 68 SNPs discussed in this report were developed by GAPS laboratories without the intended purpose of discovering SNPs specific for Puget Sound populations, so there is no expectation that these SNPs would differentiate these closely related populations. For this reason, WDFW is now collaborating with the University of Washington (Drs. James and Lisa Seeb laboratory) in several projects to screen other existing SNPs and to develop SNPs with the explicit design of better differentiating populations within both the NPSSpSu and SPSFall

groups. If we are successful in developing these SNP markers, they will be added to the existing 68 SNPs developed here and to the GAPS Chinook SNP baseline.

Acknowledgements

The Pacific Salmon Commission's Southern Boundary Restoration and Enhancement Fund and the Washington State General Fund provided the financial support for all elements of this project. We thank Angus Mackay and Victor Keong for managing this contract for the Pacific Salmon Commission, Scott Blankenship (WDFW) for discussion, reviewing this report, and for his overall knowledge of SNPs and Chinook genetics, which he gladly shared with us, Lisa and Jim Seeb for sharing their knowledge of Chinook SNPs, and the GAPS consortium of laboratories for developing the 96 SNPs screen during this project.

Figure legends:

Figure 1. Map of northwest portion of Washington State showing rivers from which samples were collected and used for this project. Circles identify general location from where samples were collected but are not intended to be spatially accurate in terms of river mile of collection. Open circles just north and east of Seattle are the locations of the Grovers Creek Hatchery. (Miller Bay, Kitsap Peninsula), and University of Washington Hatchery collections (see Table 1).

Figure 2. Fifty percent consensus neighbor joining tree constructed from both chord (Cavalli-Sforza and Edwards 1967) and allele-sharing (Jin and Chakraborty 1994) pairwise distances matrices. Tree is rooted using the four populations from outside of Puget Sound. Numbers

beside each node indicate the percentage of 1000 bootstrap runs in which that node was present using the allele sharing (left) and chord (right) distance matrices. Populations are group based on run-timing, location, and phylogenetic structure (except SamishFall, which is located in north Puget Sound, but clusters with south Puget Sound populations).

Figure 3. The first two axes from a principal component analysis of 27 Puget Sound Chinook populations and 68 SNPs. H and N in the legend refer to hatchery- and natural-origin populations, respectively. The three symbols in the plot that show both white and black filling indicate populations with both hatchery- and natural-origin populations. NPSSpSu equal spring/summer-run populations from north Puget Sound; SPSFall equals fall-run populations from south Puget Sound; and SJF equals populations from the Strait of Juan de Fuca (see also Figure 2). The three vectors indicate the position of centroids for the three clusters of populations. The actual ordination of the Chiwawa_Sp populations is (1.23, -1.04).

Figure 4. Axes 3 and 4 from a principal component analysis of 27 Puget Sound Chinook populations and 68 SNPs. H and N in the legend refer to hatchery- and natural-origin populations, respectively. The three symbols in the plot that show both white and black filling indicate populations with both hatchery- and natural-origin populations. The two vectors indicate the position of the White River populations, and the centroids for the two Nooksack populations.

Figure 5. F_{ST} outlier tests using finite island model (above) and hierarchical island model (below). Symbols represent SNP loci and lines represent the null distribution quantiles from the

coalescent simulations, ranging from 1%, 5%, 50% (median; solid line), 95%, and 99% from the bottom to the top line.

Figure 6. Results from the 100% simulations to test the power to differentiate populations in a mixed-stock fishery analysis. Bars are for the 68-SNP dataset, while the filled circles are for the microsatellite dataset, and both represent the mean estimated proportion from the 100 (SNP) and 50 (microsatellite) 100% simulated fisheries analyses for each population. The error bars for both the SNPs and microsatellites indicate the 95% confidence interval for these means. The +, -, and o symbols at the bottom of each bar indicate if the SNP dataset provided better, worse, or the same level of population differentiation, respectively, compared with the microsatellite dataset. The datasets were considered the same if their 95% confidence intervals overlapped. The three vertical lines within the plot separate North Puget Sound, South Puget Sound, Hood Canal, and Strait of Juan de Fuca populations, as you move from left to right. The 90% estimated proportion is indicated with a horizontal solid line.

References

Anderson EC, Waples RS, Kalinowski ST (2008) An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, 65, 1475–1486

Beacham TD, Jonson KL, Supernault J, Wetklo M, Deng L, Varrnavskaya N (2006) Pacific Rim population structure of Chinook salmon as determined from microsatellite analysis. *Transactions of the American Fisheries Society* **135**, 1604–1621.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond.* B **263**, 1619-1626.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19**, 233-257.

Dillon WR, Goldstein M (1984) *Multivariate Analysis. Methods and Applications.* John Wiley and Sons, New York. 387 + xii pp.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285–298.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Jin L, Chakraborty R (1994) Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. *Molecular Biology and Evolution* **11**, 120-127.

Kalinowski S (unpublished) ONCOR. A computer program for genetic stock identification. Available at http://www.montana.edu/kalinowski/Software/ONCOR.htm.

Lewontin RC, Krakauer, J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175-195.

Liu K, Muse SV (2005) PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics*, 21, 2128-2129.

Millar, R.B. 1987. Maximum likelihood estimation of mixed stock fishery composition. *Canadian Journal of Fisheries and Aquatic Sciences*, 44, 583–590.

Narum SR, Banks M, Beacham TD, Bellinger MR, Campbell MR, Dekoning J, Elz A, Guthrie III CM, Kozfkay C, Miller KM, Moran P, Phillips R, Seeb LW, Smith CT, Warheit K, Young SF, Garza JC (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, 17, 3464–3477.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**,248-249.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Ruckelshaus MH, Currens KP, Graeber WH, Fuerstenberg RR, Rawson K, Sands NJ, Scott JB (2006) *Independent populations of Chinook salmon in Puget Sound*. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-78, 125 p.

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Slatkin M, Voelm L (1991) F_{ST} in a hierarchical island model. *Genetics* 127, 627–629.

Seeb LW, Antonovich A, Banks MA, Beacham TD, Bellinger MR, Blankenship SM, Campbell MR, Decovich NA, Garza JC, Guthrie III CM, Lundrigan TA, Moran P, Narum SR, Stephenson JJ, Supernault KJ, Teel DJ, Templin WD, Wenburg JK, Young SF,Smith CT (2007) Development of a standardized DNA database for Chinook salmon. *Fisheries*, 32,540–552.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Waples RS, Teel DJ, Myers JM, Marshall AR (2004) Life-history divergence in Chinook salmon: historic contingency and parallel evolution. *Evolution* **58**, 386-403.

Warheit KI, Narum S, Kostow K (2007) Aggregating stocks for harvest management and an improved genetic stock identification. In: *Genetic Stock Identification Workshop Final Genetics Workgroup Report* (ed. Pacific Salmon Commission), pp. 105-121. Pacific Salmon Commission http://www.psc.org/GSIWorkshop/2nd_Workshop_Sept2007/Day3/Genetics_Workgroup_Recommendations.pdf

Table 1. Collections, populations, and localities of Chinook samples used in this analysis. N = sample size per population; for Production, N and H equal natural- and hatchery- origin samples, respectively; for Run Time, Sp, Su, and Fa equal spring, summer, and fall, respectively; and the first two digits from the WDFW Collection Code correspond to the year in which the samples were collected.

Population Code	Location	N	Production	Run Time	WDFW Collection code(s)
NF_M_Nooksack	N.F. Nooksack R. (Kendall Hatchery)	48	N/H	Sp	86AB, 98BA
SF_Nooksack	S.F. Nooksack R.	100	N	Sp	86AC, 93EI
SamishFall	Samish R.	95	H	Fa	98HJ
U_Skagit_Su	upper Skagit R.	129	N	Su	94DV, 95DN, 98FJ
Marblemount_H_Sp	Marblemount Hatchery (upper Skagit R.)	99	H	Sp	06EO
Suiattle_R	Suiattle R.	146	N	Sp	89AE, 98DL, 99DJ
U_Sauk_R_SpSu	upper Sauk R.	91	N	Sp/Su	94EL, 98DN, 99ED, 06FG
L_Skagit_R_Fa	lower Skagit R.	141	N	Fa	98EC, 06EN
NF_Stillaguamish	N.F. Stillaguamish R. (Whitehorse Hatchery)	231	N/H	Su	87AK, 96EB, 01DY, 01DZ
SF_Stillaguamish	S.F. Stillaguamish R.	78	N	Fa	92FP, 93EM, 94DO, 95DP, 96BT, 97OE
Skykomish_H_SU	Skykomish R. (Wallace Hatchery)	152	N/H	Su	96BU, 96BZ, 00FT
Grovers_Cr_H	Grovers Cr. Hat. (Miller Bay, Kitsap Pen)	143	H	Fa	04AT, 04HT
UW_H_SuFa	Univ. WA Hatchery (Lake Washington)	146	H	Su/Fa	04JW
Issaquah_Cr_SuFa	Issaquah Creek	133	N	Su/Fa	99FS, 04HV
Bear_Cr_SuFa	Bear Creek	107	N	Su/Fa	98GC, 99FR, 03NU, 04IP, 04IQ,
Cedar_R_SuFa	Cedar R.	169	N	Su/Fa	94EE, 02EA, 03NT, 03NV, 04HS,
Soos_H	Soos Creek Hatchery (Green R.)	95	H	Fa	98HB
White_H	White R./Hupps Springs Hatchery	287	H	Sp	98CO, 98CP, 02HR
S_Prairie_Cr	South Prairie Creek (Puyallup R.)	93	H	Fa	98CK, 99FI, 02KI
Clear_Cr_H	Clear Creek Hatchery (Nisqually R.)	143	H	Fa	05KB
Nisqually_R_SuFa	Nisqually R.	102	N	Su/Fa	98ED, 99EH, 99FB, 00FO, 00FP, 06EL
GeorgeAdams_H	George Adams Hatchery (Skokomish R.)	143	H	Fa	05JZ
NF_Skokomish_R_Fa	N.F. Skokomish R.	96	N	Fa	98FH, 99FG, 00GL, 04HH, 05IT, 06DP
SF_Skokomish_R	S.F. Skokomish R.	98	N	Fa	05IS, 06DO
Hamma_Hamma_R	Hamma Hamma R.	138	N	Fa	99EP, 00HJ, 01GY, 02IU

Table 1. (con't)

Population Code	Location	N	Production	Run Time	WDFW Collection code(s)
Elwha_H	Elwha R.	143	H	Fa	96AF, 96AG
Dungeness_R	Dungeness R.	91	N	Sp/Su	04FI, 04HP
NF_Makah_H	N.F. Makah R. (Coast)	33	H	Fa	03OC
MathenyCr	Matheny Creek (Coast)	15	N	Sp/Su	95EI
WenatcheeR_SU	Wenatchee R. (Columbia R. Basin)	31	N	Su	93DD
Chiwawa_Sp	Chiwawa R. (Columbia R. Basin)	35	N	Sp	93DH

Table 2. Description and statistics for 68 SNPs included in this analysis. Locus is the standardized GAPS name for the SNP. Minor allele is the allele for each locus whose Puget Sound frequency is less than 0.50. Puget Sound Minor is the frequency range of the minor allele across all Puget Sound populations and Total Minor is the allele frequency for Puget Sound minor allele calculated using the Puget Sound and outgroup populations. Hardy-Weinberg and Linkage Disequilibrium (LD) statistics were calculated for Puget Sound populations only. "All Pops P-value" is the Hardy Weinberg equilibrium probability calculated across all populations using Fisher's method. $P \leq 0.05$ and $P \leq 0.05$-adj are the percentages of all Puget Sound populations with Hardy-Weinberg probabilities less than or equal to 0.05 and 0.05 with Bonferroni adjustments, respectively. "LD All Pops" is the linkage disequilibrium probability per locus based on Fisher's method summed across all pairwise combinations of that locus in all populations.

WDFW_Code	Locus	Minor Allele	Major Allele	# Pops Invariant	Puget Sound Minor	Total Minor	Hardy-Weinberg			LD All Pops
							All Pops P-value	Pops P ≤ 0.05	Pops ≤ 0.05-adj	
AOts001	Ots_113242-216	C	T	0	(0.330 - 0.679)	0.859	0.001	0.222	0.037	0.000
AOts002	Ots_113457-40R	C	T	0	(0.211 - 0.565)	0.964	0.033	0.148	0.074	0.273
AOts003	Ots_123048-521	C	A	0	(0.021 - 0.423)	0.500	0.160	0.111	0.037	0.730
AOts004	Ots_128757-61R	A	deletion	0	(0.216 - 0.504)	0.534	0.042	0.074	0.037	0.020
AOts005	Ots_94857-232R	C	T	0	(0.163 - 0.443)	0.703	0.227	0.074	0.037	0.273
AOts006	Ots_94903-99R	T	G	0	(0.058 - 0.436)	0.704	0.811	0.037	0.037	0.022
AOts007	Ots_96222-525	C	T	0	(0.381 - 0.741)	0.741	0.374	0.037	0.037	0.000
AOts008	Ots_96500-180	T	G	0	(0.106 - 0.361)	0.500	0.001	0.111	0.037	0.051
AOts009	Ots_96899-357R	A	T	5	(0.000 - 0.055)	0.125	1.000	0.000	0.000	1.000
AOts010	Ots_97077-179R	G	T	0	(0.020 - 0.345)	0.500	0.806	0.077	0.000	0.993
AOts011	Ots_AldB1-122	T	C	0	(0.015 - 0.319)	0.319	0.000	0.111	0.037	0.974
AOts012	Ots_aldb-177M	TTG	ATA	0	(0.246 - 0.688)	0.955	0.244	0.111	0.037	0.001
AOts014	Ots_aspat-196	C	G	0	(0.022 - 0.199)	0.400	0.002	0.074	0.037	0.546
AOts015	Ots_C3N3	G	T	4	(0.000 - 0.505)	1.000	-	-	-	1.000
AOts016	Ots_CD59-2	A	G	0	(0.263 - 0.862)	0.862	0.000	0.148	0.037	0.009
AOts018	Ots_cox1-241	C	T	0	(0.133 - 0.466)	0.917	0.001	0.074	0.037	0.006
AOts020	Ots_EP-529	A	G	0	(0.023 - 0.181)	0.500	0.924	0.074	0.037	0.969
AOts021	Ots_ETIF1A	A	C	0	(0.056 - 0.529)	0.879	0.025	0.185	0.074	0.572

Table 2. (con't)

WDFW_Code	Locus	Minor Allele	Major Allele	# Pops Invariant	Puget Sound Minor	Total Minor	Hardy-Weinberg All Pops P-value	Hardy-Weinberg Pops P ≤ 0.05	Hardy-Weinberg Pops ≤ 0.05-adj	LD All Pops
AOts022	Ots_FARSLA-220	G	A	0	(0.021 - 0.291)	1.000	0.000	0.185	0.074	1.000
AOts023	Ots_FGF6A	T	G	0	(0.020 - 0.172)	0.750	0.000	0.148	0.037	1.000
AOts025	Ots_GDH-81x	deletion	C	0	(0.191 - 0.626)	0.626	0.006	0.111	0.074	0.009
AOts028	Ots_GPDH-338	A	G	1	(0.000 - 0.121)	0.121	0.168	0.174	0.043	1.000
AOts029	Ots_GPH-318	T	C	1	(0.000 - 0.096)	0.250	0.982	0.087	0.043	1.000
AOts031	Ots_GTH2B-550	C	G	0	(0.076 - 0.549)	0.600	0.021	0.185	0.037	0.543
AOts032	Ots_HSP90B-100	C	T	4	(0.000 - 0.148)	0.900	0.839	0.111	0.000	1.000
AOts033	Ots_IGF-I.1-76	T	A	19	(0.000 - 0.071)	0.177	1.000	0.000	0.000	1.000
AOts036	Ots_IL8R_C8	T	C	0	(0.073 - 0.250)	0.296	0.659	0.111	0.000	0.573
AOts037	Ots_MHC1	A	G	0	(0.120 - 0.346)	0.922	0.407	0.111	0.037	0.003
AOts038	Ots_MHC2	T	G	0	(0.188 - 0.528)	0.967	0.177	0.037	0.037	0.003
AOts039	Ots_mybp-85	C	T	0	(0.304 - 0.676)	0.985	0.000	0.185	0.037	0.050
AOts040	Ots_Myc-366	T	C	0	(0.031 - 0.313)	0.313	0.232	0.148	0.000	0.057
AOts041	Ots_myo1a-384	C	A	2	(0.000 - 0.259)	0.259	0.349	0.120	0.000	0.979
AOts043	Ots_nkef-192	T	C	17	(0.000 - 0.022)	0.783	0.807	0.125	0.125	1.000
AOts044	Ots_NOD1	C	G	0	(0.091 - 0.324)	0.697	0.000	0.222	0.037	0.270
AOts046	Ots_LWSop-638	C	T	26	(0.000 - 0.021)	0.467	1.000	0.000	0.000	0.935
AOts047	Ots_Ots311-101x	deletion	AA	0	(0.037 - 0.218)	0.218	0.675	0.074	0.000	0.984
AOts048	Ots_P450	T	A	3	(0.000 - 0.168)	1.000	0.042	0.087	0.087	1.000
AOts049	Ots_P53	A	G	0	(0.141 - 0.508)	0.508	0.029	0.111	0.037	0.015
AOts050	Ots_PGK-54	T	A	0	(0.242 - 0.726)	1.000	0.000	0.222	0.037	0.029
AOts051	Ots_Prl2	G	A	0	(0.175 - 0.553)	0.817	0.000	0.111	0.074	0.112
AOts052	Ots_RAG3	T	C	0	(0.196 - 0.541)	0.871	0.002	0.222	0.037	0.005
AOts053	Ots_RFC2-558	deletion	A	0	(0.201 - 0.564)	1.000	0.004	0.185	0.074	0.014
AOts054	Ots_S7-1	C	T	0	(0.322 - 0.715)	0.900	0.000	0.222	0.074	0.001

Table 2. (con't)

WDFW_Code	Locus	Minor Allele	Major Allele	# Pops Invariant	Puget Sound Minor	Total Minor	Hardy-Weinberg All Pops P-value	Hardy-Weinberg Pops P ≤ 0.05	Hardy-Weinberg Pops ≤ 0.05-adj	LD All Pops
AOts055	Ots_SClkF2R2-135	T	A	0	(0.025 - 0.585)	0.585	0.000	0.185	0.074	0.016
AOts058	Ots_SWS1op-182	T	A	0	(0.065 - 0.404)	0.672	0.009	0.148	0.037	0.691
AOts060	Ots_TGFB	T	C	0	(0.142 - 0.455)	0.481	0.105	0.148	0.000	0.194
AOts061	Ots_TLR3	T	C	0	(0.088 - 0.443)	0.793	0.011	0.148	0.037	0.027
AOts063	Ots_u07-07.161	T	C	0	(0.129 - 0.356)	0.569	0.064	0.074	0.037	0.629
AOts064	Ots_u07-18.378	T	A	0	(0.021 - 0.255)	0.500	0.001	0.148	0.074	0.511
AOts065	Ots_u07-25.325	T	C	3	(0.000 - 0.218)	0.462	0.938	0.043	0.000	1.000
AOts066	Ots_u07-49.290	G	A	0	(0.112 - 0.440)	0.732	0.099	0.074	0.037	0.011
AOts067	Ots_u07-53.133	C	T	0	(0.178 - 0.513)	0.875	0.000	0.296	0.037	0.077
AOts070	Ots_u211-85	C	T	0	(0.110 - 0.471)	0.552	0.061	0.074	0.037	0.001
AOts072	Ots_u6-75	T	C	19	(0.000 - 0.074)	0.206	1.000	0.000	0.000	1.000
AOts073	Ots_unk526	A	G	0	(0.037 - 0.158)	0.296	0.039	0.111	0.037	0.995
AOts075	Ots_ZR-575	A	G	3	(0.000 - 0.117)	0.906	0.001	0.130	0.087	1.000
AOts079	Ots_GST.207	T	C	0	(0.095 - 0.286)	0.286	0.000	0.185	0.037	0.646
AOts081	Ots_hnRNPL.533	T	A	0	(0.026 - 0.199)	0.645	0.719	0.037	0.000	0.929
AOts082	Ots_HSP90B.100	C	T	4	(0.000 - 0.147)	0.891	0.800	0.167	0.000	1.000
AOts083	Ots_HSP90B.385	A	G	0	(0.032 - 0.343)	0.359	0.057	0.148	0.037	0.269
AOts085	Ots_LEI.292	A	G	19	(0.000 - 0.027)	0.027	1.000	0.000	0.000	1.000
AOts086	Ots_MetA	C	T	0	(0.138 - 0.567)	0.567	0.000	1.000	1.000	0.275
AOts088	Ots_P450-288	G	A	0	(0.207 - 0.439)	0.667	0.004	0.074	0.000	0.104
AOts091	Ots_Tf-3545	C	G	12	(0.000 - 0.107)	0.107	0.542	0.167	0.000	1.000
AOts093	Ots_u07-17.135	A	G	6	(0.000 - 0.056)	0.118	0.692	0.150	0.000	1.000
AOts094	Ots_u07-17.373	deletion	A	2	(0.000 - 0.074)	0.439	0.034	0.120	0.040	1.000
AOts095	Ots_u07-19.260	C	T	0	(0.010 - 0.242)	0.242	0.708	0.148	0.000	1.000
AOts096	Ots_u07-20.332	A	C	3	(0.000 - 0.215)	0.215	0.999	0.043	0.043	1.000

Table 3. SNPs genotyped but removed from dataset prior to detailed analyses.

WDFW_Code	Locus	GAPS_75	Reasons for Removal
AOts013	Ots_AsnRS-60	Yes	missing from entire populations
AOts017	Ots_CD63	Yes	missing in greater than 50% of all samples
AOts019	Ots_EndoRB1-486	Yes	invariant
AOts024	Ots_FGF6B_1	Yes	missing from entire populations
AOts026	Ots_GH2	Yes	invariant
AOts027	Ots_GnRH-271	Yes	missing in greater than 50% of all samples
AOts030	Ots_GST-375	Yes	minor allele frequency < 0.01
AOts034	Ots_Ikaros-250	Yes	missing from entire populations
AOts035	Ots_IL11	Yes	minor allele frequency < 0.01
AOts042	Ots_myoD-364	Yes	missing in greater than 50% of all samples
AOts045	Ots_nramp-321	Yes	missing in greater than 50% of all samples
AOts056	Ots_SERPC1-209	Yes	minor allele frequency < 0.01
AOts057	Ots_SL	Yes	missing from entire populations
AOts059	Ots_TAPBP	Yes	missing from entire populations
AOts062	Ots_Tnsf	Yes	missing from entire populations
AOts068	Ots_u07-57.120	Yes	missing in greater than 50% of all samples
AOts069	Ots_u202-161	Yes	missing from entire populations
AOts071	Ots_u4-92	Yes	missing in greater than 50% of all samples
AOts074	Ots_zP3b-215	Yes	minor allele frequency < 0.01
AOts078	Ots_BAC-E9	No	invariant
AOts084	Ots_IL-1RA.173	No	missing from entire populations
AOts087	Ots_Ots2	No	missing in greater than 50% of all samples
AOts089	Ots_PrpI.120	No	minor allele frequency < 0.01
AOts090	Ots_PSMB1.197	No	invariant
AOts092	Ots_TUBA-454	No	missing in greater than 50% of all samples
AOts097	Ots_u07-24.441	No	missing from entire populations
AOts098	Ots_u07-53.185	No	missing from entire populations
AOts101	Ots_U608.861	No	missing in greater than 50% of all samples

Table 4. Genetic equilibrium statistics for the 27 Puget Sound populations included in this report. For each population, Fisher's Method combines all 68 SNP loci, or all pairwise combinations of locus-pairs for Hardy-Weinberg and Linkage Disequilibrium, respectively. P-values ≤ 0.05-adj refers to Bonferroni-adjusted P-values. See Table 1 for the description of each population.

	Hardy-Weinberg					Heterozygosity		Linkage Disequilibrium				
Population Code	Fisher's Method			Loci w/ P-value				Fisher's Method			Locus Pair w/ P-value	
	Chi-sqr	df	P-value (all loci)	≤ 0.05	≤ 0.05-adj	Exp	Obs	Chi-sqr	df	P-value (all loci)	≤ 0.05	≤ 0.05-adj
NF_M_Nooksack	131	112	0.10	0.07	0.02	0.26	0.23	2917	3494	1.00	0.042	0.000
SF_Nooksack	173	120	0.00	0.12	0.02	0.27	0.25	4567	3882	0.00	0.084	0.000
SamishFall	123	116	0.30	0.02	0.02	0.27	0.27	3212	3420	0.99	0.047	0.000
U_Skagit_Su	182	124	0.00	0.11	0.02	0.27	0.29	3705	3970	1.00	0.046	0.000
Marblemount_H_Sp	154	116	0.01	0.14	0.03	0.26	0.26	3511	3540	0.63	0.064	0.000
Suiattle_R	201	124	0.00	0.15	0.03	0.26	0.23	3599	3910	1.00	0.053	0.001
U_Sauk_R_SpSu	175	122	0.00	0.07	0.02	0.27	0.26	3784	4156	1.00	0.049	0.000
L_Skagit_R_Fa	223	130	0.00	0.17	0.02	0.29	0.30	4968	4402	0.00	0.088	0.000
NF_Stillaguamish	1299	118	0.00	0.88	0.59	0.28	0.19	4092	3538	0.00	0.078	0.001
SF_Stillaguamish	199	124	0.00	0.11	0.03	0.28	0.25	3856	4018	0.97	0.055	0.000
Skykomish_H_SU	198	124	0.00	0.21	0.02	0.27	0.25	3997	4154	0.96	0.046	0.000
Grovers_Cr_H	141	120	0.10	0.08	0.02	0.28	0.28	3614	3618	0.52	0.059	0.000
UW_H_SuFa	216	112	0.00	0.20	0.04	0.27	0.26	4229	3080	0.00	0.103	0.001
Issaquah_Cr_SuFa	123	114	0.26	0.05	0.02	0.27	0.27	3307	3776	1.00	0.039	0.001
Bear_Cr_SuFa	158	124	0.02	0.06	0.02	0.27	0.25	3751	4030	1.00	0.049	0.000
Cedar_R_SuFa	139	124	0.17	0.10	0.02	0.26	0.26	3577	4030	1.00	0.046	0.000
Soos_H	262	108	0.00	0.20	0.06	0.26	0.22	3056	3380	1.00	0.042	0.000
White_H	218	116	0.00	0.14	0.03	0.25	0.24	3849	3304	0.00	0.085	0.001
S_Prairie_Cr	182	118	0.00	0.10	0.02	0.27	0.24	3319	4020	1.00	0.036	0.000
Clear_Cr_H	142	122	0.10	0.08	0.02	0.28	0.28	3420	3780	1.00	0.043	0.000

Table 4. (con't)

Population Code	Hardy-Weinberg					Heterozygosity		Linkage Disequilibrium				
	Fisher's Method			Loci w/ P-value				Fisher's Method			Loci Pair w/ P-value	
	Chi-sqr	df	P-value (all loci)	≤ 0.05	≤ 0.05-adj	Exp	Obs	Chi-sqr	df	P-value (all loci)	≤ 0.05	≤ 0.05-adj
Nisqually_R_SuFa	138	114	0.06	0.12	0.02	0.27	0.26	3321	3420	0.89	0.058	0.000
GeorgeAdams_H	222	122	0.00	0.25	0.05	0.27	0.30	3587	3760	0.98	0.053	0.000
NF_Skokomish_R_Fa	149	122	0.05	0.10	0.02	0.27	0.26	3452	3898	1.00	0.047	0.000
SF_Skokomish_R	148	122	0.05	0.07	0.02	0.28	0.27	3369	3904	1.00	0.037	0.000
Hamma_Hamma_R	126	118	0.29	0.03	0.02	0.27	0.26	3485	3896	1.00	0.049	0.000
Elwha_H	162	130	0.03	0.08	0.02	0.27	0.26	4727	4288	0.00	0.068	0.000
Dungeness_R	158	126	0.03	0.08	0.02	0.27	0.25	3713	4284	1.00	0.047	0.000

Table 5. Principal component coefficients for each SNP locus for the first four components (Axis 1 through Axis 4), and the correlation between each SNP locus and that component (r^2). Correlation coefficients greater than or equal to 0.80 and less than or equal to -0.80 are highlighted with bold and italic typeface.

Locus	Axis 1	r^2	Axis 2	r^2	Axis 3	r^2	Axis 4	r^2
AOts001	-0.0935	-0.48	0.0136	0.05	0.3186	0.70	0.0904	0.16
AOts002	-0.0857	-0.39	0.2360	0.76	-0.0089	-0.02	-0.1196	-0.19
AOts003	-0.0672	-0.30	-0.2802	***-0.87***	0.0782	0.15	-0.0283	-0.04
AOts004	0.0911	0.50	0.0080	0.03	-0.2139	-0.49	0.1157	0.22
AOts005	0.0537	0.30	0.1371	0.55	0.0902	0.22	0.2928	0.57
AOts006	-0.0883	-0.49	0.0275	0.11	-0.0116	-0.03	0.3385	0.65
AOts007	0.0235	0.11	0.1994	0.67	-0.0249	-0.05	0.0945	0.16
AOts008	-0.0064	-0.04	-0.0483	-0.23	0.0057	0.02	-0.1989	-0.46
AOts009	0.0137	0.31	0.0131	0.21	0.0050	0.05	0.0394	0.31
AOts010	0.2212	***0.90***	0.0113	0.03	0.1366	0.23	0.0569	0.08
AOts011	0.1415	0.79	0.0861	0.34	-0.0327	-0.08	0.1278	0.25
AOts012	-0.2653	***-0.87***	0.0190	0.04	0.2006	0.28	0.1003	0.11
AOts014	0.0609	0.64	-0.0358	-0.26	-0.0186	-0.08	0.0608	0.22
AOts015	-0.3478	***-0.91***	-0.1667	-0.31	-0.1632	-0.18	0.1357	0.12
AOts016	-0.0996	-0.35	0.3261	***0.81***	-0.0918	-0.14	-0.2773	-0.34
AOts018	0.0367	0.17	0.2112	0.70	0.2424	0.48	0.1222	0.20
AOts020	-0.1281	***-0.93***	0.0129	0.07	-0.0244	-0.07	-0.0483	-0.12
AOts021	0.1568	0.63	0.2204	0.63	0.0777	0.13	-0.0493	-0.07
AOts022	-0.1142	-0.75	-0.0807	-0.38	0.0092	0.03	-0.0178	-0.04
AOts023	0.0266	0.36	0.0099	0.10	-0.0689	-0.40	0.0107	0.05
AOts025	0.1175	0.51	-0.1927	-0.59	0.1409	0.26	0.1467	0.22
AOts028	0.0031	0.04	-0.0345	-0.30	0.0514	0.27	0.0013	0.01
AOts029	-0.0230	-0.38	0.0405	0.48	-0.0096	-0.07	-0.0486	-0.28
AOts031	0.1053	0.40	0.2983	***0.80***	-0.0899	-0.14	-0.0925	-0.12
AOts032	-0.0332	-0.37	0.1036	***0.81***	-0.0127	-0.06	0.0376	0.14
AOts033	-0.0128	-0.36	0.0306	0.62	0.0015	0.02	0.0008	0.01
AOts036	-0.0163	-0.15	-0.0551	-0.37	0.0425	0.17	-0.1182	-0.39
AOts037	0.1491	***0.81***	0.0131	0.05	-0.1022	-0.23	0.0695	0.13
AOts038	0.0981	0.52	0.1343	0.50	-0.0241	-0.05	0.0292	0.05
AOts039	-0.1012	-0.46	0.1160	0.38	-0.1026	-0.20	0.1443	0.23
AOts040	0.1764	***0.85***	-0.0431	-0.15	-0.1246	-0.25	-0.1213	-0.20
AOts041	0.1669	***0.88***	-0.0037	-0.01	0.0392	0.09	0.0990	0.18
AOts043	-0.0028	-0.19	0.0102	0.48	0.0073	0.21	-0.0022	-0.05
AOts044	0.0347	0.28	-0.1002	-0.57	-0.1071	-0.36	-0.0667	-0.19
AOts046	-0.0007	-0.07	-0.0007	-0.05	0.0001	0.01	0.0009	0.03
AOts047	-0.0493	-0.40	-0.0441	-0.25	-0.0787	-0.27	-0.1725	-0.48
AOts048	-0.0691	-0.70	-0.0405	-0.29	-0.0554	-0.24	0.1191	0.42
AOts049	-0.1464	-0.56	-0.2142	-0.58	-0.2376	-0.38	0.0012	0.00

Table 5. (con't)

Locus	Axis 1	r2	Axis 2	r2	Axis 3	r2	Axis 4	r2
AOts050	-0.3089	***-0.85***	0.2347	0.46	-0.0305	-0.04	0.0479	0.05
AOts051	-0.1684	-0.69	0.1701	0.50	0.0182	0.03	0.2225	0.32
AOts052	0.0341	0.17	0.0952	0.34	-0.1984	-0.42	0.0534	0.09
AOts053	-0.2427	***-0.88***	0.1150	0.30	0.0428	0.07	-0.1114	-0.14
AOts054	-0.1840	-0.75	0.1318	0.38	-0.2208	-0.38	0.2038	0.29
AOts055	0.3395	***0.90***	0.0743	0.14	-0.1519	-0.17	0.2273	0.21
AOts058	-0.0643	-0.36	-0.1886	-0.76	-0.0993	-0.24	-0.1097	-0.22
AOts060	0.2203	***0.91***	0.0126	0.04	-0.1701	-0.30	-0.0272	-0.04
AOts061	-0.0864	-0.50	0.1430	0.58	-0.1543	-0.38	-0.0243	-0.05
AOts063	0.0720	0.55	0.0229	0.12	0.1739	0.57	0.0752	0.20
AOts064	0.0736	0.52	-0.1067	-0.54	-0.0311	-0.09	-0.0479	-0.12
AOts065	-0.0506	-0.38	0.1195	0.64	0.0124	0.04	-0.1266	-0.33
AOts066	-0.1609	-0.73	0.0542	0.17	0.2549	0.49	0.0099	0.02
AOts067	0.0048	0.02	0.1021	0.35	-0.2780	-0.56	0.1099	0.18
AOts070	-0.0162	-0.09	-0.0648	-0.26	-0.0857	-0.20	-0.2978	-0.58
AOts072	-0.0145	-0.35	0.0498	***0.85***	0.0120	0.12	-0.0101	-0.09
AOts073	-0.0197	-0.27	0.0328	0.32	-0.0899	-0.52	-0.0328	-0.16
AOts075	-0.0426	-0.57	-0.0264	-0.25	-0.0467	-0.26	0.0638	0.30
AOts079	-0.0218	-0.21	0.0634	0.44	-0.0767	-0.32	-0.0232	-0.08
AOts081	-0.0575	-0.52	-0.0590	-0.38	-0.0183	-0.07	0.0070	0.02
AOts082	-0.0344	-0.41	0.0913	0.77	-0.0155	-0.08	0.0355	0.15
AOts083	-0.0130	-0.09	-0.1382	-0.66	0.2173	0.62	0.0827	0.19
AOts085	0.0048	0.30	-0.0020	-0.09	0.0040	0.11	-0.0083	-0.18
AOts086	-0.1386	-0.53	-0.1841	-0.50	-0.2794	-0.45	0.2562	0.34
AOts088	-0.0263	-0.18	-0.0559	-0.26	0.1051	0.30	0.1172	0.27
AOts091	-0.0593	***-0.81***	-0.0278	-0.27	-0.0461	-0.27	0.0117	0.06
AOts093	-0.0034	-0.08	0.0097	0.17	-0.0300	-0.31	-0.0450	-0.38
AOts094	-0.0162	-0.31	-0.0108	-0.15	0.0118	0.10	-0.0176	-0.12
AOts095	-0.0889	-0.57	0.1247	0.57	0.0654	0.18	-0.1429	-0.32
AOts096	-0.0102	-0.09	0.0425	0.28	-0.0715	-0.28	-0.0474	-0.15

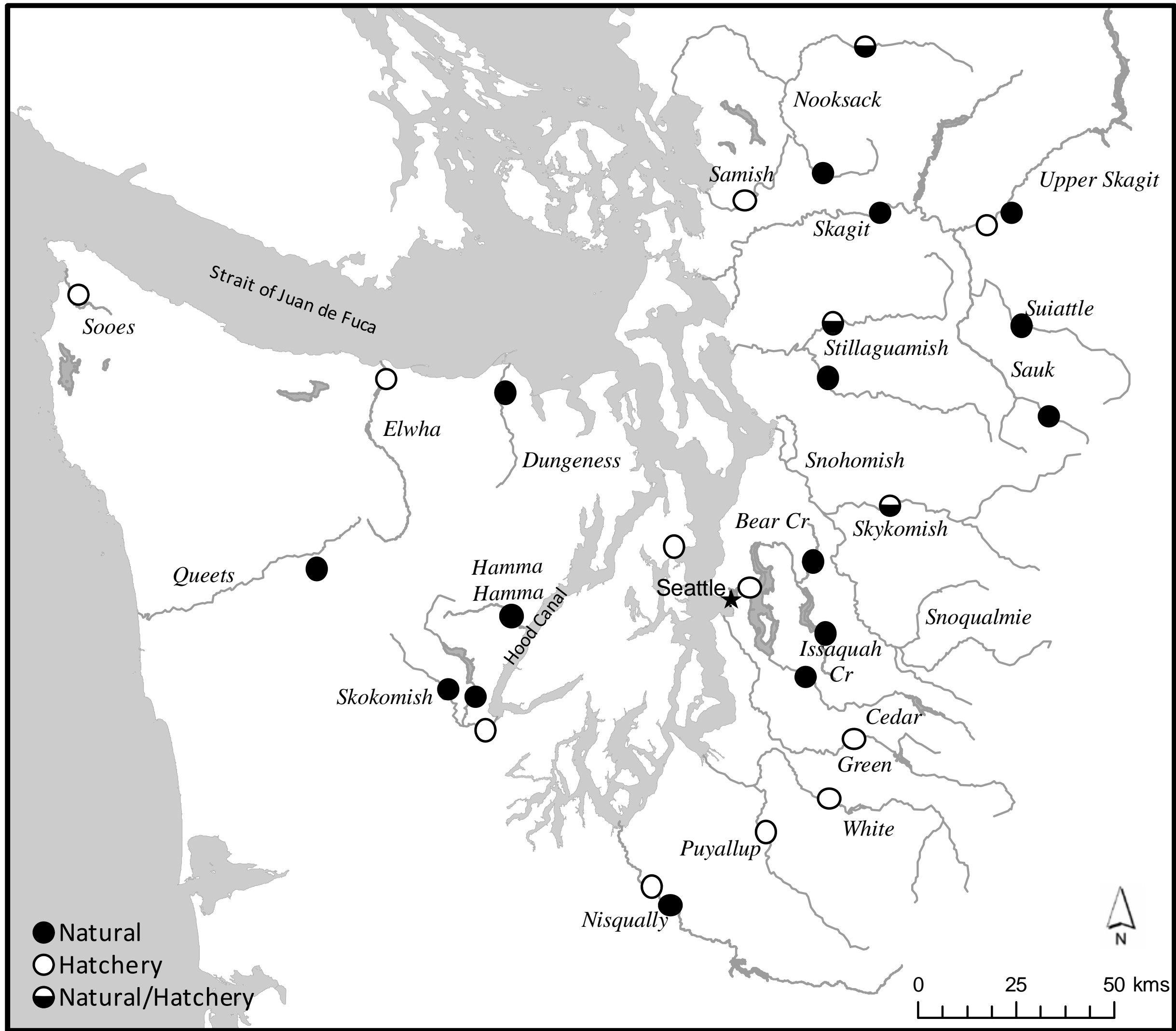

Figure 1.

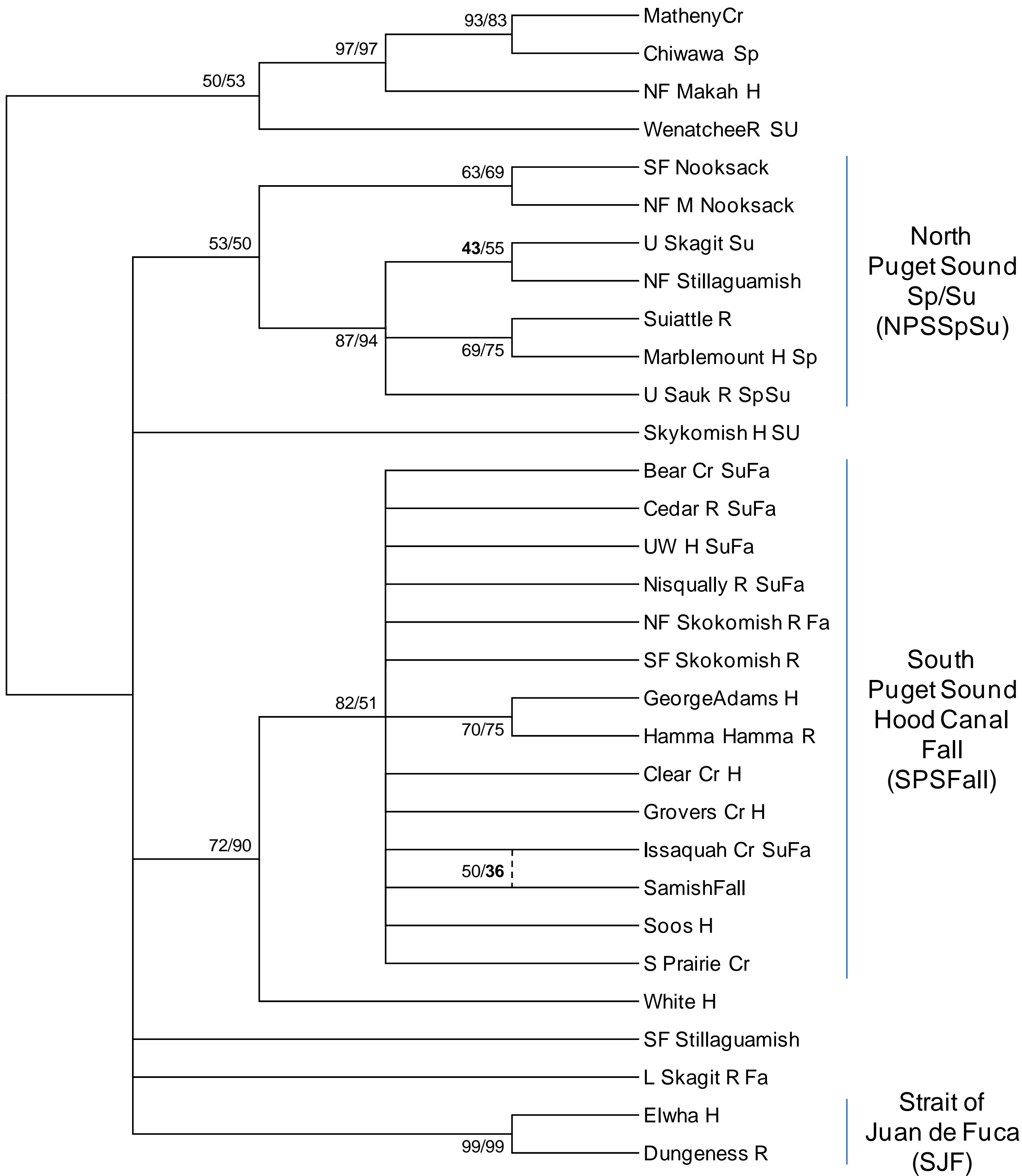

Figure 2.

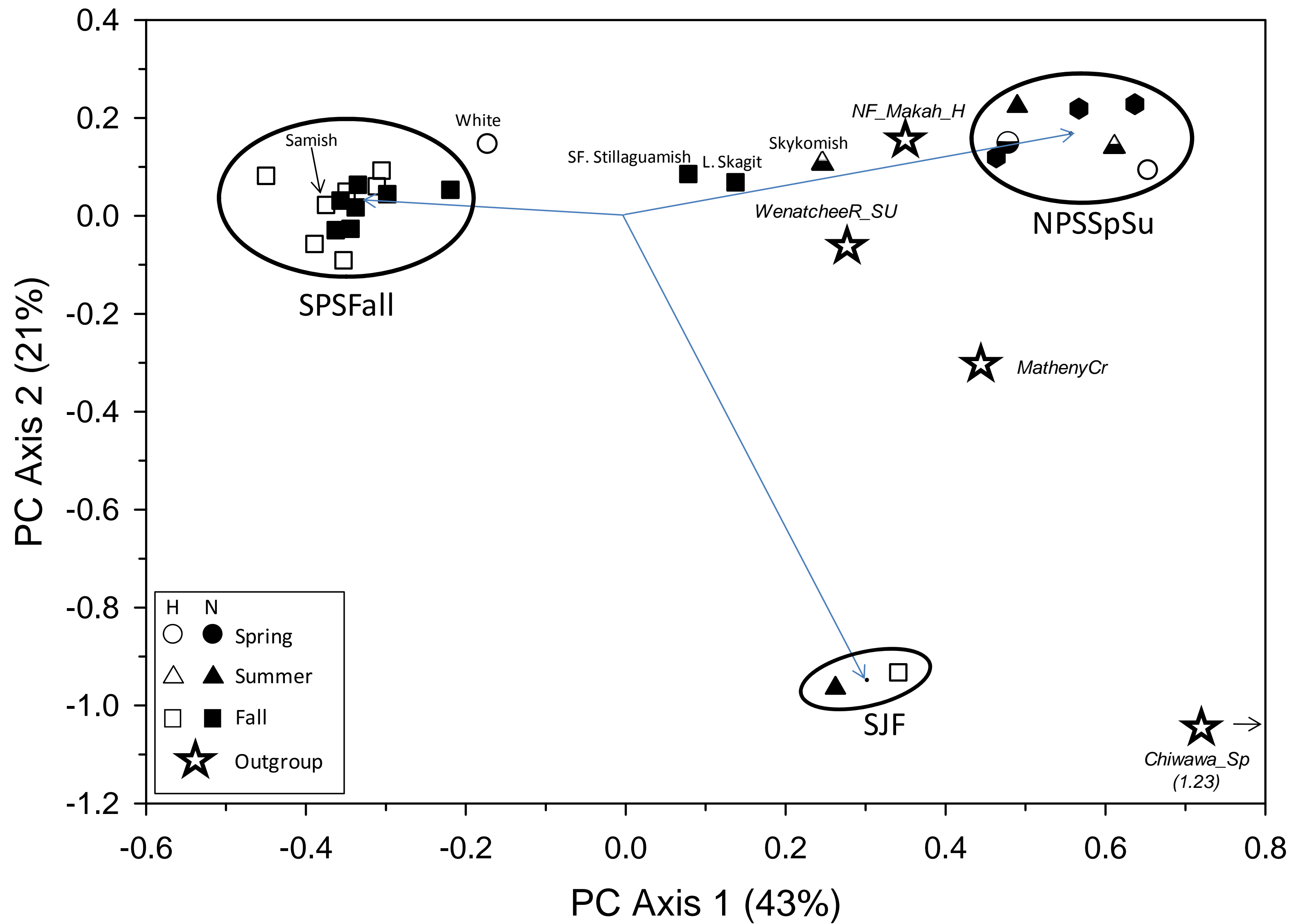

Figure 3.

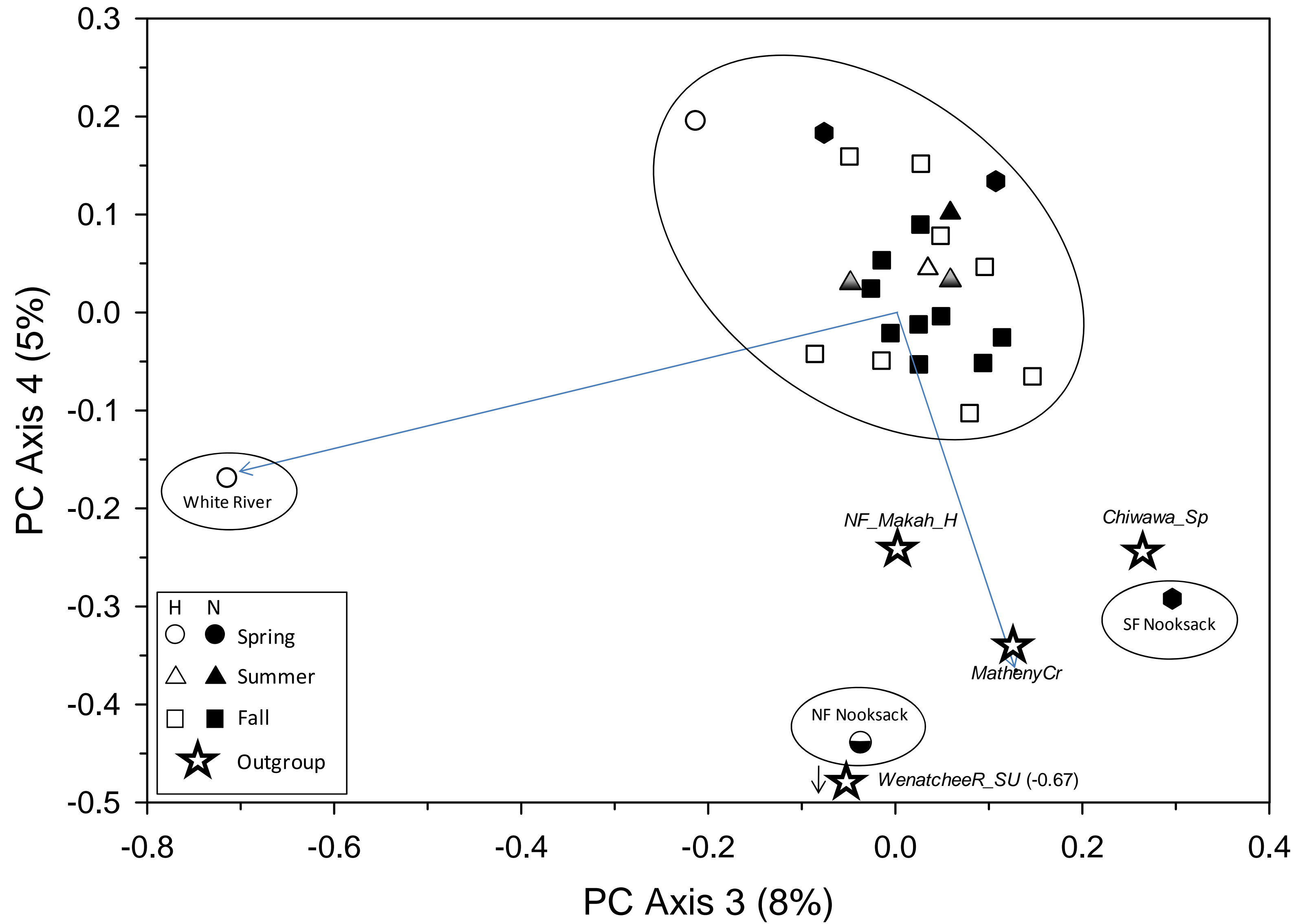

Figure 4.

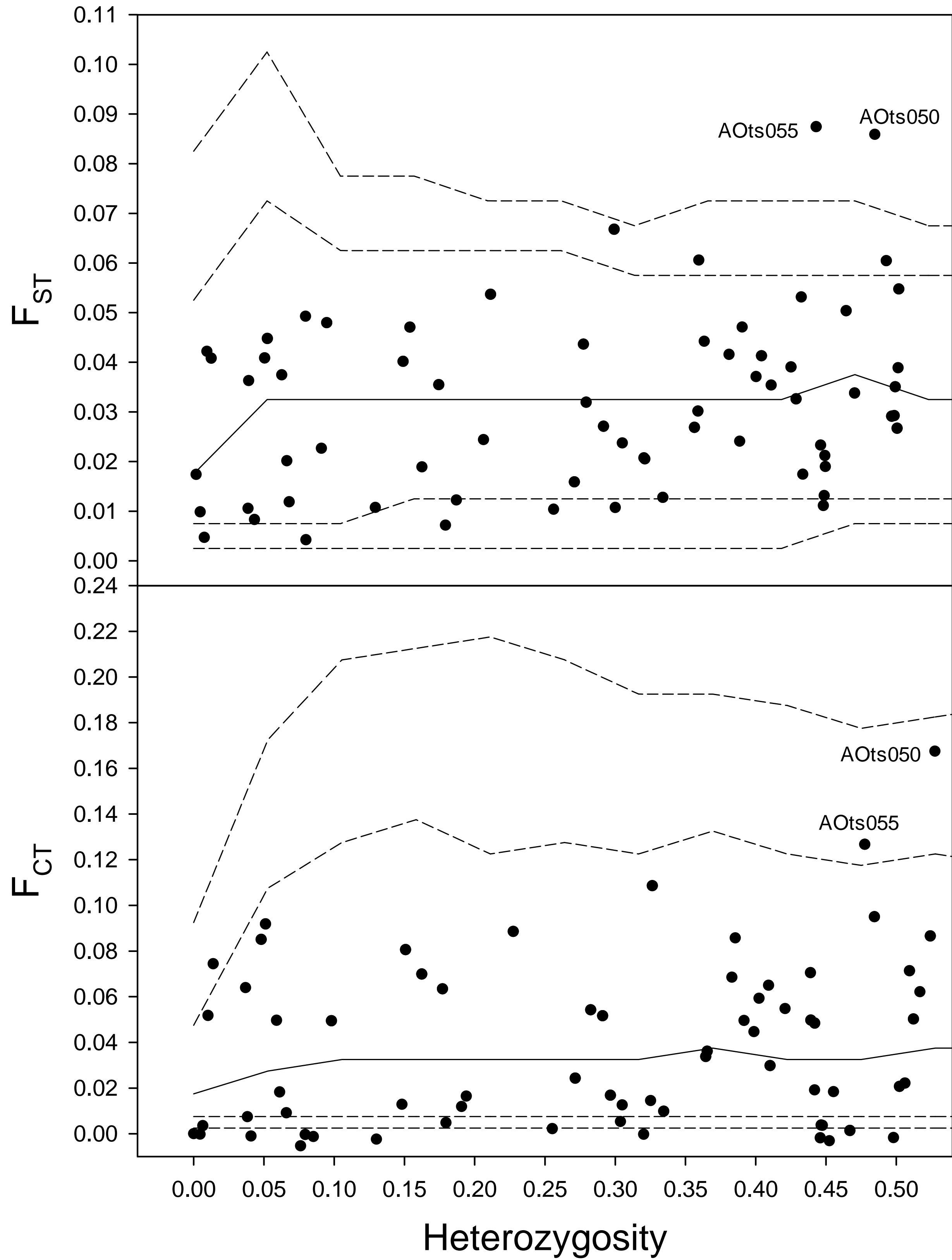

Figure 5.

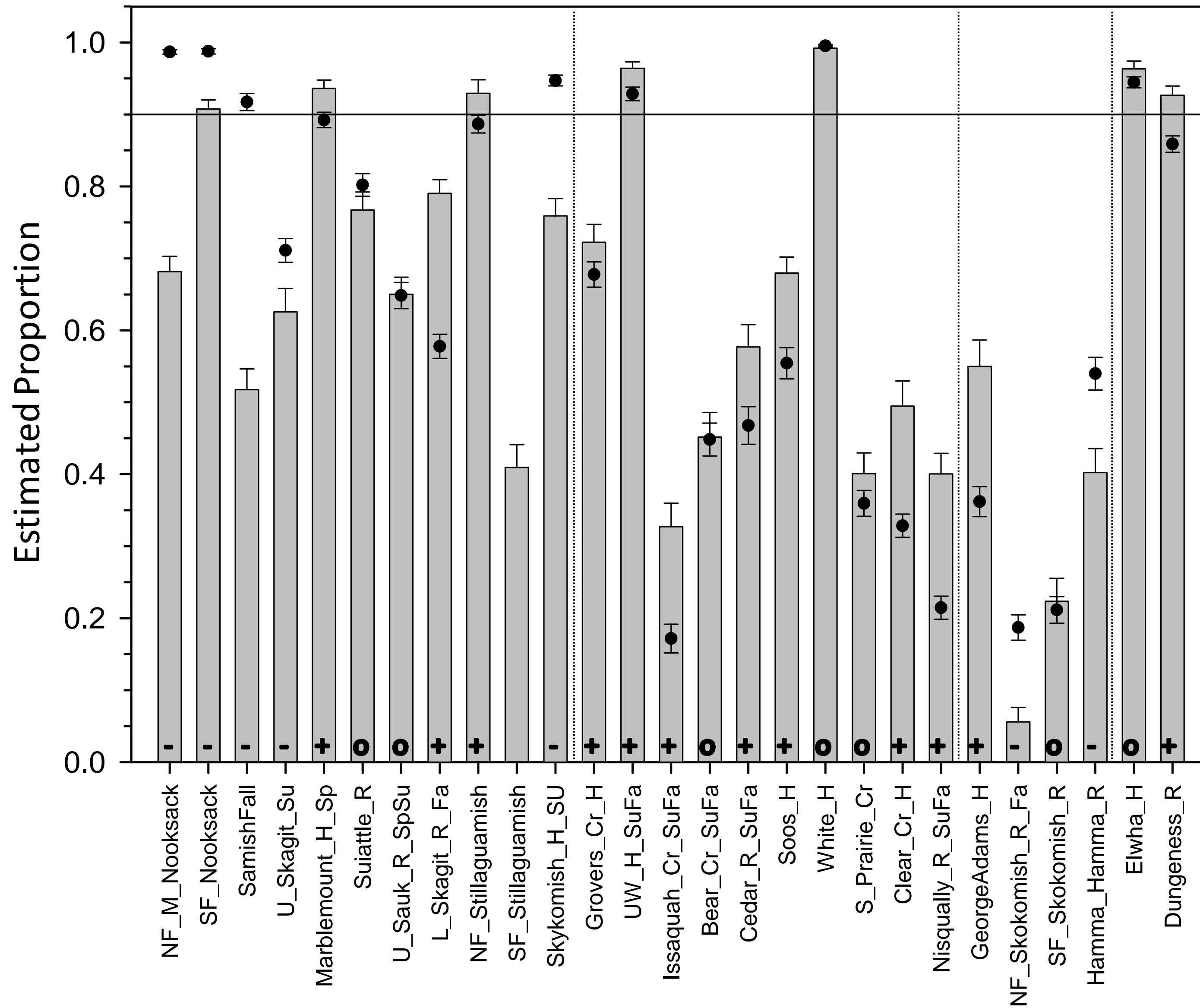

Figure 6.