# Genetic structure of kokanee (*Oncorhynchus nerka*) spawning in tributaries of Lake Sammamish, Washington

Kenneth I. Warheit* and Cherril Bowman

Report Submitted to King County Department of Natural Resources and Parks, Water and Land Resources Division, and Trout Unlimited – Bellevue/Issaquah as partial fulfillment for Contracts 07-2047 (King County) and 07-2098 (Trout Unlimited)

*Corresponding author: Molecular Genetics Laboratory, Washington Department of Fish and Wildlife, 600 Capitol Way N. Olympia, WA 98501, phone 360.902.2595, fax 360.902.2943, email: warhekiw@dfw.wa.gov

Final report, June 2008

**Abstract**

The goal of this project is to determine the genetic structure of kokanee populations within the Lake Sammamish watershed, and to use these data to help determine the appropriate management plan for these populations. We genotyped 664 samples from 17 collections, primarily from Ebright (years 2000, 2001, and 2003), Laughing Jacobs (2000 and 2003), and Lewis (2000, 2001, and 2003) creeks, using two different genetic markers: 17 microsatellite loci and a 417-nucleotide base fragment of the mitochondrial DNA (mtDNA) cytochrome b gene. Mitochondrial DNA is maternally inherited and therefore reflects maternal lineages. All Lake Sammamish kokanee collections appear to possess relatively equal molecular diversity as reflected by the microsatellite markers, but the 2001 collections from Ebright and Lewis creeks showed greater diversity at the cytochrome b locus. Many of the Lake Sammamish collections are in molecular disequilibrium, resulting from family groups, relatively recent immigrants, and different maternal lineages present within each collection. Furthermore, these data also suggest that co-mingling with spawning late-run kokanee are fish from either an undiscovered kokanee population within Lake Sammamish, kokanee/sockeye from Sammamish River, Lake Sammamish sockeye, or sockeye or kokanee transplanted from out of basin stocks. Despite the movement of fish among the Lake Sammamish tributaries, Ebright, Laughing Jacobs, and Lewis creeks show significant among tributary population structure. That is, fish from these three tributaries are genetically distinct; however, fish from Ebright and Laughing Jacobs creeks appear more genetically similar to each other than either is to Lewis Creek. Finally, any hatchery-based supplementation program that obtains broodstock directly from spawning tributaries must address the fact that fish drawn from these sources may include: closely related individuals (i.e., family groups), and fish not native to that tributary, either individuals from neighboring tributaries, or from out-of-basin populations.

**Introduction**

Kokanee are the non-anadromous freshwater form of *Oncorhynchus nerka*, and are native to the Lake Washington/Lake Sammamish watershed (Berge and Higgins, 2003). Historically, this watershed supported three kokanee run-types: early-run from Issaquah Creek, now extinct; middle-run from the Sammamish River tributaries; and late-run from the Lake Sammamish

tributaries (Berge and Higgins, 2003; Young et al., 2004), although this conclusion is not universally accepted (J. Mattila, pers. comm., 2008). The late-run stocks were initially considered to be descendants from non-native transplants from Lake Whatcom Hatchery, but Young et al. (2004) have shown that the Lake Sammamish kokanee are genetically distinct from the Lake Whatcom Hatchery strain, and are assumed to be native to the Lake Sammamish basin. Furthermore, Young et al. (2004) have also shown that the middle-run stocks from Bear and Little Bears creeks, tributaries to the Sammamish River, are residualized sockeye, which in turn are most-likely native to the Lake Washington watershed (Spies et al., 2007). Since available information suggests middle-run kokanee are either extinct or highly introgressed with sockeye, the only remaining native kokanee in the Sammamish River/Lake Sammamish watershed are the late-run fish from the tributaries of Lake Sammamish.

Kokanee in the Lake Sammamish tributaries spawn primarily in Lewis, Ebright, and Laughing Jacobs creeks (Figure 1), where escapement since 1996, the first year of continuous counts, has been variable, ranging from less than 100 to over 4,000 (Jackson, WDFW Fish and Wildlife Biologist, pers. comm 2008), with a median count of just under 600 and a harmonic mean of 247 fish. This relatively low escapement and the fact that these populations appear to represent the only remaining native kokanee populations in the Lake Washington/Lake Sammamish watershed have prompted resource agencies to design management strategies to stabilize or increase the total number of kokanee spawning in Lake Sammamish tributaries. A hatchery-based augmentation program has been proposed for Lake Sammamish kokanee (A. Hoffmann, WDFW Region 4 Fish Program Manager, per comm., 2008); however, there is no consensus as to the appropriate composition of the broodstock. There appears to be two alternatives: (1) broodstock would be composed primarily of fish from the most abundant population (Lewis Creek), and juveniles from the hatchery program would be dispersed throughout the Lake Sammamish basin, including Ebright and Laughing Jacobs creeks; or, (2) tributary-specific broodstocks would be established, with an attempt to maintain separate hatchery populations for Lewis, Ebright, and Laughing Jacobs creeks. The historical relationships and the current level of natural gene flow among these three tributaries are relevant to the discussion as to which of these two alternative management strategies is most desirable.

The primary purpose of this report is to determine the genetic relationships among populations from the three primary kokanee-producing tributaries in Lake Sammamish: Lewis, Ebright, and Laughing Jacobs creeks. Here, we use mitochondrial DNA (mtDNA) sequence (cytochrome b) data to define maternal lineages within each population, and microsatellite data to assess differences within and among populations and for each maternal lineage. Our goal was first to determine if spawning aggregations (i.e., temporal collections) in each tributary are genetically distinct and relatively independent populations, and if genetic differences do exist, are they a function of historical separation and lineage sorting, or more recent differentiation, possibly driven by small population sizes and genetic drift.

This report expands upon previous work conducted by Washington Department of Fish and Wildlife (WDFW) on kokanee within the Lake Washington/Lake Sammamish watershed. Young et al. (2004) examined kokanee and sockeye collections from the Lake Washington/Lake Sammamish watershed and from potential out-of-basin sources populations. Kassler (2005) added to the samples analyzed by Young et al. (2004) individuals collected in 2002 from Webster Creek (Walsh Lake) within the Cedar River watershed. Although Kassler (2005) was unable to conclude definitive genetic relationships between the Webster Creek and other populations, Kassler did show significant genetic differentiation between Webster Creek and all other populations in the analysis. Young et al. (2004) and Kassler (2005) used the same suite of microsatellite loci, which in total was a subset of the loci used in this present study (Table 1). Our study here differs from Young et al. (2004) and Kassler (2005) in several components: (1) increased total number of microsatellite loci from nine to 17 (Table 1); (2) added mtDNA sequence data; (3) increased collection years sampled for the Lake Sammamish kokanee populations from 2000 to also include 2001, 2003, and 2004 (Table 2); and (4) included small collections from Vasa and Pine creeks (Table 2). We accepted the conclusions from Young et al. (2004) that the Lake Sammamish kokanee populations are native and most closely related to each other compared with the other collections in their data set. Furthermore, in this study we made no attempt to address the origin of the middle-run kokanee from Sammamish River tributaries (see Young et al., 2004, Spies et al., 2007), or the genetic affinities of the Webster Creek kokanee, although we included the Kassler (2005) Webster Creek samples in this study.

## Materials and Methods

*Samples*

All samples used in this study were collected as fin clips preserved in 100% ethanol and archived in the WDFW Molecular Genetics Laboratory collection. Except for Meadow_99, all samples were collected from adults, either live (year 2001, and 2003, in part) or from carcasses, by King County (Water and Land Resources Division [WLRD]), WDFW, and Seattle Public Utility (Table 2). The Meadow_99 collection was composed of juveniles taken from the Spokane Hatchery. The brood stock for these juveniles was from Meadow Creek (Lake Kootenay), British Columbia, which drains into the upper Columbia River (Table 2, Figure 1). A total of 664 samples from 17 collections were included in this analysis (Table 2). Fourteen of the 17 collections were from six spawning tributaries within the Lake Sammamish basin (Lewis, Ebright, Laughing Jacobs, Pine, Vasa, and Issaquah creeks, Figure 1). We used temporally replicated collections from Ebright, Laughing Jacobs, and Lewis creeks to assess temporal stability in allele frequencies within a spawning locality. One collection was taken from research gill nets sampled within Lake Sammamish (collection composed of *Oncorhynchus nerka* of various ages; Lake Samm_03, Table 2) and represents an aggregate sample. The remaining two collections are from outside the Lake Sammamish basin: Webster and Meadow creeks, as described above. The year 2000 collections from Lewis, Ebright, and Laughing Jacobs creeks, were included in Young et al. (2004), as were the 1993 Issaquah Creek and 1999 Meadow Creek collections. For the purpose of this report, an individual fish is referred to as a sample, samples obtained from a locality during one year of sampling is referred to as a collection, and collections from a single locality are jointly referred to as a population. DNA was extracted from each sample using Machery-Nagel silica membrane based kits following the manufacturer's standard protocol and eluting with a final volume of 100 µl.

*Microsatellites*

We used 17 microsatellite loci (Table 1); PCR amplified in seven multiplexes using fluorescently end-labeled primers and AB 9700 thermal cyclers. All reverse primers included a seven-nucleotide base extension (GTTTCTT) to the 5' end to promote the incorporation of a nontemplated adenosine to the 3' end of the PCR product (Brownstein et al, 1996). Each PCR

reaction was conducted within a total volume of 10 µl, of which 1 µl was unquantitated DNA. Also included in the PCR reaction were 0.2 mM dNTPs, 1.5 mM $MgCl_2$, and 0.05 units of GoTaq (Promega Corporation). Primer concentration varied among loci, as did the PCR annealing temperature and numbers of cycles (Table 1). For all PCR reactions, the thermal profile was as follows: 2 minutes of initial DNA denature at 94°C, followed by cycles of denature at 94°C (15 seconds), annealing at temperatures listed in Table 1 (30 seconds), and DNA extension at 72°C (60 seconds). Following the last cycle, the PCR is completed with a final extension at 72°C (10 minutes) and then held at 10°C until placed at 4°C refrigeration. PCR products were visualized using an ABI-3730 DNA Analyzer with internal size standards (GS500LIZ 3730) and GeneMapper 3.7 software. Allele binning and naming were accomplished using a modification of MicrosatelliteBinner-v1h (S. Young, WDFW; available from the author). MicrosatelliteBinner creates groups (bins) of alleles with similar mobilities (alleles with the same number of repeat units). The upper and lower bounds of the bins are determined by identifying clusters of alleles separated by gaps (nominally 0.4 base pairs in size, can vary among loci) in the distribution of allele sizes. The bins are then named as the mean allele size for the cluster rounded to an integer.

*Cytochrome b Sequences*

We used two WDFW-developed primers (SsaL14437: GCTAATGACGCACTAGTCG; SsaH14885: CTCAAATTCATTGTACAAGGG; S. Young, unpublished data) to sequence a 417-nucleotide base fragment of the mtDNA cytochrome b gene. This fragment begins 59 bases past the start of the cytochrome b gene (base number 15,435) as indicated on *Oncorhynchus nerka* sequence EF055889 in GenBank (Benson et al., 2007). We used a 15 µl PCR reaction to amplify the sequence fragment. Included in this reaction were 1.5 µl of extracted DNA, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, and 0.06 units of GoTaq (Promega Corporation), and 0.4mM of each primer. The thermal profile for this reaction was 3 minutes of initial DNA denature at 95°C, followed by 29 cycles of denature at 94°C (60 seconds), annealing at 55°C (60 seconds), and DNA extension at 72°C (60 seconds). Following the last cycle, the PCR is completed with a final extension at 72°C (10 minutes) and then held at 10°C until placed at 4°C refrigeration. The PCR product was purified using Shrimp Alkaline Phosphatase (SAP; 2 units), which degrades

the excess unused dNTPs by removing the 3'-phosphate group, and exonuclease I (2 units), which degrades the single stranded primers. We applied direct sequencing, in both directions using each primer in a separate reaction. We used BigDye version 3.1 (Applied Biosystems) at one-quarter strength (2 μl, with 3 μl of 5x buffer), 1 μl of the purified PCR product, and 0.16 mM of primer in a total of 20 μl per reaction. The thermal profile for these sequencing reactions differed slightly from the manufacturer's recommendations: 1 minute of initial DNA denature at 96°C, followed by 37 cycles of denature at 96°C (10 seconds), annealing at 50°C (7 seconds), and DNA extension at 60°C (4 minutes). Following the last cycle, the PCR is held at 10°C until placed at 4° C refrigeration. DNA sequences were visualized using an ABI-3100 Genetic Analyzer, and aligned and edited using Sequencher version 4.7 (Gene Codes Corporation).

*Data Analyses*

Genetic diversity and equilibrium

Molecular diversity for the microsatellite data was assessed using a variety of programs. The allelic range, total number of alleles, and observed heterozygosity for each locus was calculated using PowerMarker version 3.25 (Liu and Muse, 2005). We tested for significant heterozygote deficit at each locus, pooled across all collections with GENEPOP (Raymond and Rousset, 1995). We also used PowerMarker to ascertain the average number of samples and alleles per locus, and observed heterozygosity for each of the 13 collections, and FSTAT (Goudet, 2001) to calculate allelic richness. We used a program written in Matlab (version 7; The MathWorks; Warheit, unpubl. program) to calculate the mean unbiased expected heterozygosity (Nei, 1987) for each collection. Expected heterozygosity is the average expected heterozygosity across all loci and is based on within locus allele frequencies. It not only provides a measure of molecular diversity within populations, it also presents an estimate for molecular equilibrium within a population. Expected heterozygosity is the percent of expected heterozygotes within a population, given the current allele frequency and assuming Hardy-Weinberg equilibrium. If the observed heterozygosity deviates significantly from the expected heterozygosity, the population may not meet Hardy-Weinberg expectations. To assess the relative differences in expected heterozygosity among the collections, we resampled each collection by randomly selecting with replacement the appropriate number of individuals per collection (each collection with constant sample size equal to the sample size in the original dataset) and then calculated mean unbiased

expected heterozygosity for each collection using the new dataset. We repeated this bootstrap procedure 100 times and for each collection we constructed a box plot showing the 0.05, 0.10, 0.25, 0.50 (median), 0.75, 0.90, and 0.95 percentile for the expected heterozygosity over the 100 bootstrap runs. For cytochrome b data we calculated haplotype frequencies, haplotype diversity, and nucleotide diversity using Arlequin (Version 3.1; Excoffier et al. 2005). Haplotype diversity is equivalent to the expected heterozygosity in diploid data and is the probability that two randomly chosen haplotypes in a population are different (Nei, 1987). Nucleotide diversity is the mean number of nucleotide differences among all pairs of haplotypes within a collection. While haplotype diversity provides a measure of how different individuals are within a collection, nucleotide diversity provides a measure of how different haplotypes are within a collection. Both statistics provide a measure of molecular diversity.

We used GENEPOP and FSTAT to calculate measures of genetic equilibrium. We tested for Hardy-Weinberg equilibrium using two procedures. First, for each locus in each collection we calculated the probability of disequilibrium using GENEPOP, and then calculated the percentage of loci in each collection that are in disequilibrium at both the nominal P = 0.05, and P = 0.05, adjusted for multiple comparisons (Bonferroni correction)[1]. Second, we used FSTAT to calculate $F_{IS}$, a measure of heterozygote deficit (deficit of observed compared with expected heterozygosity) within collections, and tested the significance of the $F_{IS}$ statistic by permuting alleles among individuals within each collection. Finally, we used GENEPOP to test for linkage disequilibrium for each pair of loci in each collection, and calculated the percentage of locus-pairs in each collection that are in disequilibrium at both the nominal and adjusted P-values.

Geographic and temporal structure of populations

We assessed the geographic and temporal structure of populations (e.g., Ebright, Laughing Jacobs, and Lewis creeks) using three statistics. We used GENETIX (Belkhir, 2004) to calculate $F_{ST}$ for each pairwise combination of collections, and tested the significance of the $F_{ST}$ by permuting individual genotypes between pairwise collections. This procedure was repeated 100 times, and following each iteration, GENETIX calculated a new $F_{ST}$ for the permuted dataset.

[1] For all statistical tests we used both the nominal P = 0.05 and P = 0.05, adjusted for multiple comparisons. The adjustment varied depending on the number of simultaneous tests and was calculated as nominal alpha (0.05) divided by the number of simultaneous tests.

The null hypothesis was no genetic structure, as measured by $F_{ST}$, and the probability to reject the null hypothesis was determined by the number of permutations that produced an $F_{ST}$ greater than the $F_{ST}$ for the original data. We tested for genotypic differentiation for each pair of collections using GENEPOP, with a null hypothesis of no difference in genotypic distribution between pairwise collections. This procedure uses a contingency table of genotypes and tests for differences between pairs of collections using log-likelihood (G) test (Goudet et al., 1996). We used Arlequin to calculate $\Phi_{ST}$ values (haplotype equivalent of $F_{ST}$) and to conduct a differentiation test for cytochrome b data.

We used an analysis of molecular variance (AMOVA; Excoffier et al., 1992), implemented in Arlequin, to test for differences in geographic and temporal structure of populations, for both the microsatellite and cytochrome b data. For the microsatellite analysis we included only samples from Ebright, Laughing Jacobs, and Lewis creeks collected in 2000, 2001, and 2003, defined populations as collections (i.e., years; Table 2), and grouped populations by tributary. With this procedure we evaluated the percent of molecular variation that is attributed to differences among tributaries versus differences among years (i.e., collections) within a tributary. We conducted tests using one model where samples from 2001 were included, and another model where 2001 samples were excluded. We used a similar set of models for the cytochrome b data, except for two models we included 2004 collections from Ebright and Lewis creeks, and for one model we reversed the roles of collection and tributary, by defining populations based on tributaries and grouped populations based on the year samples were collected.

To visualize differences in the microsatellite genetic structure of collections, we conducted factorial correspondence analyses (FCA) using GENETIX. Factorial correspondence analysis is a technique similar to principal component analysis (PCA), except where in a PCA there is an eigenvalue decomposition of a covariance (or correlation) matrix, in FCA there is a decomposition of a chi-square statistic associated with a contingency table. The microsatellite data are converted to an allele frequency contingency table, with individual samples as rows, and alleles (each allele from each locus) as columns, and allele frequencies represented as a 0, 1, and 2, indicating the absence (0), heterozygote presence (1), or homozygotes presence (2) of an allele. The values in the contingency tables are weighted by the total number of alleles presented

in the data set[2], and the table is reduced using singular value decomposition. We tested for differences in the spatial distribution of collections along two dimensions in a FCA plot using Mahalanobis distances (differentiation defined as a Mahalanobis distance significantly greater than zero). Two collections are considered to have different spatial distributions in a FCA (and therefore difference sets of allele frequencies) if the Mahalanobis distance that separates their centroids (bivariate means) is significantly greater than zero. We calculated Mahalanobis distances using a program written by Warheit in Matlab (version 7; The MathWorks).

Genetic assignments

We measured contemporary gene flow among populations from Ebright, Laughing Jacobs, and Lewis creeks using a partial Bayesian assignment procedure for the microsatellite data. The baseline data for this analysis consisted of samples from the 2000, 2001, and 2003 collections from Ebright and Lewis creeks, 2000 and 2003 collections from Laughing Jacobs, 2003 collection from Pine Creek, 1993 collection from Issaquah Creek, 2002 collection from Webster Creek, and the 1999 collection from Meadow Lake. We employed a jackknife (leave-one-out) procedure to assign individual samples from all Ebright, Laughing Jacobs, and Lewis collections, as well as the 2003 Lake Sammamish aggregate collection, to a population. Here, an individual sample is removed from the baseline. Allele frequencies for each baseline collection are calculated without the inclusion of the test animal, and the Rannala and Mountain (1997) likelihood statistics are then calculated for the test fish using its microsatellite genotype and the new baseline collection allele frequencies. This likelihood statistic is the probability that an individual microsatellite genotype (i.e., the genotype from the test fish) is drawn from one of the baseline collections (technically, the probability of a genotype conditioned on a baseline collection). For each individual being assigned, likelihood is calculated for each baseline collection (ten likelihood values for each fish). We used Bayes' Theorem (Equation 1) to calculate the parental source collection for each individual (posterior probabilities; technically, the probability of the source collection conditioned on the genotype) for each individual, with

$$P(stock \mid genotype) = \left( \frac{P(genotype \mid stock) \cdot P(stock)}{P(genotype)} \right) \qquad \text{Equation 1}$$

[2] For data with no missing alleles, this would equal to the number of individuals times the numbers of loci times two. However, some samples had missing data for one or more loci, so for each sample the total number of alleles is equal to the total number of individual loci with data times two. This value is then summed across all samples.

*P(genotype|stock)* equal to the Rannala and Mountain (1997) likelihood statistic, *P(stock)* equal to the prior probabilities for each collection. *P(genotype)* is the overall probability of the data, and was calculated as the numerator from equation 1, summed across all collections. We assumed equal prior probabilities for all collections in the baseline. Although we calculated the collection-source probabilities for all ten baseline collections, as defined above, we aggregated the probabilities for each of the three main tributaries (Ebright, Laughing Jacobs, and Lewis) by adding together the collection-source probabilities for each tributary. This way, for each individual sample, we report the population-source probabilities for an aggregate Ebright, Laughing Jacobs, and Lewis Creek, in addition to the probabilities for Pine, Issaquah, and Webster creeks, and Meadow Lake. We used a program written in Matlab (version 7; The MathWorks; Warheit, unpubl. program) to calculate all probabilities.

Finally, we used Matlab (version 7; The MathWorks) to perform a principal component analysis (PCA) of the log-transformed Rannala and Mountain (1997) likelihood scores to search for outlier likelihoods that may indicate that individual samples are from a source population not present in the current data set (i.e., a source population that is not Ebright, Laughing Jacobs, Lewis, Pine, Issaquah, or Webster creeks, or Meadow Lake).

Relatedness

We calculated the pairwise relatedness (*sensu* Queller and Goodnight, 1989) of individuals within populations using the program IDENTIX (Belkhir et al., 2002). For this relatedness statistic, full siblings (or parent-offspring) have relatedness at 0.50, and half-siblings and first cousins are related at 0.25 and 0.125, respectively. Pairs of individuals with relatedness of zero or less are considered unrelated. We tabulated all pairwise relatedness values for each population and calculated frequency distributions and variances. Variances increase as the number of high relatedness values increase, and therefore, populations whose variances are statistically greater than random have a greater portion of closely related individuals than would be expected based on random mating. We tested for the significance of the relatedness variance using the permutation procedure in IDENTIX. Here, for each individual and for each locus, genotypes are randomly selected without replacement. Once all genotypes for all loci are selected, population pairwise relatedness values and their variances are calculated. We repeated

this procedure 100 times. The actual relatedness variance is considered significant if it is greater than the 95% level for the distribution of variances from the resampling procedure.

Phylogenetic Structure

We ascertained the historical relationships of the cytochrome b haplotypes by constructing a minimum evolution tree in Mega 4 (Tamura et al., 2007), using the program's default options. Here, for each plausible tree topology the sum of all branch lengths is computed and the tree with the smallest sum is chosen as the minimum evolution tree. These trees are computed as unrooted although a root can be placed with ancillary data. The tree presented here is unrooted, which indicates that we have not determined evolutionary direction or which haplotype is most ancestral.

## Results

*Microsatellites*

Molecular diversity and within-population structure

The number of alleles per locus across all 17 loci ranges from six (One-105 +a) to 40 (One-101 +a), with a median 17 alleles per locus. Heterozygosity range from 0.42 (Ots-3M +a) to 0.92 (Omm-1130 +a), with a median of 0.84, and although there is a relatively good correspondence between the number of alleles and heterozygosity among the loci, One-100 +a, and to a lesser degree One-114 +a, One-2 +a, and Omm-1130 +a show a significant deficit in heterozygotes (Table 1). There are no loci that show a significant excess of heterozygotes.

We examined molecular diversity within collections using average number of alleles per locus, allelic richness, and heterozygosity (Table 3). Although the number of alleles per locus within collections varied from roughly seven to 15, this statistic is greatly affected by the collection's sample size, which varied from nine to 97. Allelic richness measures allelic diversity within a collection by standardizing all collections within the data set to a collection sample size equal to the size for the collection with the lowest number of samples (Pine_03, Table 3). Therefore allelic richness provides a more useful statistic than number of alleles per locus to compare molecular diversity among collections. Here, with the exception of Webster_05, there is a

relatively constant allelic richness among collections; all late run populations from Lake Sammamish have allelic richness rounding to seven alleles per locus (Table 3).

Except for Webster_05, expected heterozygosity ranged from 0.73 to 0.83. Both Webster_05 and IssaEarly_93 show significantly lower heterozygosities than most other collections, indicating lower molecular diversity, while Meadow_99, LakeSam_03, Lewis_03, and Ebright_01 show higher expected heterozygosities (Figure 2). Among the three primary Lake Sammamish tributaries expected heterozygosities do not appear significantly different from each other, with the possible exception of the higher heterozygosities for Lewis_03 and Ebright_01 (Figure 2). All collections, except Pine_03, show some deviations from Hardy-Weinberg equilibrium, at least at one locus, and four collections (Laugh_00, Laugh_03, Lewis_03, and Meadow_99) show a significant deficit in observed heterozygosity, as measured by $F_{IS}$ (Table 3). Webster_05, Lewis_01, and especially Lewis_00 show a large number of loci that are in Hardy-Weinberg disequilibrium, but do not show a significant deficit in observed heterozygosity, as measured by $F_{IS}$ (Table 3). The average observed heterozygosity for these three collections is larger than the expected heterozygosity. For Lewis_00 ten of the 17 loci show an excess of observed heterozygotes, and four of these ten loci are in Hardy-Weinberg disequilibrium. Overall, nine and one of the 17 loci (0.53 and 0.06) in Lewis_00 are in Hardy-Weinberg disequilibrium at the P=0.05 and P=0.05 adjusted level, respectively (Table 3). The pattern for Lewis_01 is similar to that of Lewis_00; 14 of the 17 loci show an excess of heterozygotes. Of the three primary Lake Sammamish tributaries, all three Ebright collections are in Hardy-Weinberg equilibrium, Laughing Jacobs is in disequilibrium with significant heterozygote deficit, and Lewis shows a significant heterozygote deficit in 2003 and heterozygote excess in 2000 and 2001.

While Hardy-Weinberg disequilibrium signifies a non-random association of alleles within a locus, linkage disequilibrium indicates a non-random association of alleles among loci. This could result from the physical linkage of loci on the same chromosome, or statistical linkage as a result of population processes, such as a population being composed of an aggregate of samples from genetically different sources. We examined all pairwise combinations of loci within collections to determine if there is significant linkage disequilibrium. No pair of loci showed

significant linkage across all populations indicating that no two loci are physically linked on the same chromosome. However, in Ebright_01, Lewis_00, Lewis_01, Pine_03, and to a lesser extent Laugh_00, a relatively high percentage of locus-pairs show lineage disequilibrium (Table 3).

Geographic structure of populations

We demonstrate clear separation of IssaEarly_93, Meadow_99, and Webster_05 from each other and from the Lake Sammamish populations along three factorial correspondence analysis (FCA) axes (Figure 3). At this scale, the Lake Sammamish tributary populations form a tight and distinct cluster, except for a scatter of individuals that are drawn toward the middle of the plot. When the FCA includes only samples from the Lake Sammamish tributaries, we note significant differences among the Lake Sammamish tributary populations (Figure 4). The within tributary Mahalanobis distances for the 2000 and 2003 collections are not significant for Ebright ($D^2$=0.14, P=0.19), Laughing Jacobs ($D^2$=0.07, P=0.64), and Lewis creeks ($D^2$=0.03, P=0.59), but are significant among tributaries (12 pairwise comparisons; average $D^2$=2.4, P=2.2 x $10^{-6}$; median $D^2$=2.4, P=3.0 x $10^{-9}$). The 2001 collections for both Ebright and Lewis creeks are distinct and significantly different from each other and all other collections (7 pairwise comparisons each; average $D^2$=2.9, P=3.5 x $10^{-6}$ for Ebright_01 and $D^2$=3.6, P=7.6 x $10^{-10}$ for Lewis_01; median $D^2$=2.9, P=9.8 x $10^{-15}$ for Ebright_01, $D^2$=3.3 and P=5.6 x $10^{-17}$ for Lewis_01). Although each of these 2001 collections differ from their co-tributary collections in the same direction in Figure 4, there are no alleles at a high frequency that are shared only between these 2001 collections. The Pine Creek centroid clusters closely with the Ebright centroids (Figure 4), and is not significantly different from either Ebright_00 ($D^2$=0.20, P=0.46) or Ebright_03 ($D^2$=0.49, P=0.19). However, Pine is also not significantly different from Laugh_00 ($D^2$=1.8, P=0.0026) or Laught_03 ($D^2$=1.0, P=0.0091) at the Bonferroni-adjusted P-value, but is different from Lewis_00 ($D^2$=3.75, P=1.8 x $10^{-6}$) and Lewis 03 ($D^2$=2.6, P=0.0002).

The pattern of population structure, as measured by $F_{ST}$ is very similar to that measured by explicit tests for genotypic differentiation (Table 4), and reveals patterns similar to the FCA (Figure 4). The 2000 and 2003 collections from within Ebright, Laughing Jacobs, and Lewis creeks are not significantly different from each other in either allelic ($F_{ST}$) or genotypic structure,

although differences between Lewis_00 and Lewis_03 are significant at the un-adjusted P-value (Table 4). All three collections from Ebright are significantly different from collections from both Laughing Jacobs and Lewis creeks, and the Laugh_00 collection differs from all three collections from Lewis, but the 2003 collections from Laughing Jacobs and Lewis are not significant in their $F_{ST}$ or genotypic differentiation. Pine Creek shows a large number of non-significant values, but since the sample size for Pine is small (N=9), there is relatively little power to differentiate Pine Creek from the other populations. IssaEarly_93, Webster_05, and Meadow_99 are significantly different from each other and all the late run Lake Sammamish populations.

To examine the interaction between collection year and tributary, we conducted two analyses of molecular variance (AMOVA), both including and excluding the 2001 collections from Ebright and Lewis creeks (Table 5). For both models there is significant difference among groups (tributaries), but only for the model that included the 2001 collections was there significant differences among years (collections) within tributaries (Table 5). These results are consistent with both the $F_{ST}$ and genotypic differentiation tests, and the FCA. That is, there is no significant difference between the 2000 and 2003 collections within a tributary, and the tributaries themselves are significantly different from each other. In addition, the 2001 collections from both Ebright and Lewis are well differentiated from the other collections from each tributary, respectively, but their inclusion within their respective group (tributary) does not alter the significance of differentiation among the tributaries.

Genetic assignments

Results from the genetic assignment tests are presented as ternary plots (Figures 5 and 6). For Ebright, 0.70 and 0.69 of the samples collected in 2000 and 2003, respectively, are assigned back to Ebright, while only 0.07 and 0.06 are assigned to either Laughing Jacobs or Lewis creeks in 2000 and 2003, respectively (Figure 5). One sample is assigned to Pine creek in 2003, while the remaining samples (0.23 and 0.22 in 2000 and 2003, respectively) are unassigned. That is, 23% and 22% of the samples from 2000 and 2003 have posterior probabilities less that 0.80 and based on this criterion, are not assigned to a natal creek. The assignment results for 2000 samples from Lewis creek are similar to those from Ebright creek (Figure 5). Of the 92 samples genotyped,

0.74 of the individuals are assigned to Lewis, while 0.06 are assigned to either Laughing Jacobs or Ebright creeks, and 0.14 are unassigned. However, in 2003 fewer individuals are assigned to Lewis creek (0.60), while 0.11 are assigned to Laughing Jacobs, 0.05 are assigned to Ebright, and 0.20 are unassigned. Two samples from Lewis03 (4%) are assigned to Meadow_99 with high confidence (posterior probability equal 1.00). In general, individuals are not assigned back to the collection locality at as high of a frequency for Laughing Jacobs as they are for either Ebright or Lewis (Figure 5). One-quarter (0.25) of the samples from Laughing Jacobs are assigned to a population other than Laughing Jacobs in 2000, and this non-collection locality assignment rate rises to 0.35 in 2003. One sample from Laughing Jacobs in 2000 is assigned to Meadow_99 with high confidence (posterior probability equal 1.00).

Samples assign back to their collection locality at a higher rate for the 2001 set of samples (both Ebright and Lewis creeks) than for any other collection year (Figure 6). For Ebright in 2001, 0.77 are assigned to Ebright, 0.11 assigned to Lewis, and 0.12 unassigned. The majority of these unassigned individuals have posterior probabilities for Ebright greater than 0.50. For Lewis in 2001, 0.80 are assigned to Lewis, 0.06 assigned to Ebright, 0.01 assigned to Laughing Jacobs, and 0.13 unassigned. More samples collected from within Lake Sammamish in 2003 assign to Lewis (0.43) than Ebright (0.26), and, four of the 35 individuals sampled (0.11) are assigned to Meadow_99 (Figure 6).

Seven samples from Figure 5 and 6 assigned with high confidence to Meadow_99, but it seems unlikely that kokanee from Meadow_99 (or their direct descendents) would be present in Lake Sammamish, unless fish were transported from Lake Kootenay, British Columbia or Spokane Hatchery and stocked in Lake Sammamish. The methods used here to assign individual fish to a population assume that the population of origin for each fish being tested is present in the baseline data. If the population of origin is not present, a fish will be assigned to some other population and it is possible that the posterior probability for that assignment will be high. That is, the wrong assignment will have strong statistical support. However, if the population of origin is not present in the baseline data, the Rannala and Mountain (1997) log-likelihoods (see Material and Methods) for all populations should be unusually small. We examined all log-likelihoods associated with each collection for all individuals simultaneously by conducting a

principal component analysis (PCA). The first two axes from the PCA accounted for 76% of the log-likelihood variance and produced a plot showing a relatively tight cluster of points and a series of outlier points (Figure 7). Samples with low log-likelihood scores across all populations are shown to the right of the plot and individuals with low log-likelihood scores from the Lake Sammamish populations are shown at the bottom of the plot. In other words, samples with low log-likelihoods overall, but particularly low log-likelihoods for the Lake Sammamish populations will appear in the lower right quadrant of Figure 7. Genotypes from populations not included in the baseline will have low log-likelihoods overall, and particularly low log-likelihoods from the Lake Sammamish populations.

Thirteen individuals are outliers extending beyond the 99.5% confidence ellipse into the lower right quadrant of Figure 7. All seven individuals assigned as Meadow_99 are among these 13 outliers. It is highly unlikely that any of these 13 individuals are from Meadow_99 or any of the other baseline populations, in particular any of the known Lake Sammamish populations.

Relatedness

We calculated relatedness statistics for the Lake Sammamish tributary collections only (Table 6). Pairwise relatedness values varies greatly among these collections, ranging from 0.80 of the individuals from the Laugh_03 collection with no relatedness values greater than or equal to 0.40, to Lewis_00 where 0.59 of the individuals have at least one pairwise relatedness value greater than or equal to 0.40. For Lewis_01, for example, 54 individuals (0.56) have at least one pairwise relatedness value greater than or equal to 0.40, 28 individuals (0.29) have at least two pairwise relatedness values greater than or equal to 0.40, and 13, seven, seven, and one individual(s) have at least three, four, five, and six pairwise relatedness values greater than or equal to 0.40, respectively (Table 6).

All Lake Sammamish collections show significant variances in the their relatedness values. That is, the actual variances for each collection exceeded the 95% level for the collection's simulated data set, and in all cases the actual variances were greater than the maximum variance for the 100 simulated runs (Table 6). This indicates that compared with random mating of parents all collections show a significantly high number of closely related individuals (i.e., pairs of

individuals with high relatedness values). Among the Lake Sammamish tributary populations, Laughing Jacobs Creek had the highest variance, while Ebright Creek had the lowest variance (Table 6).

*Cytochrome b*

Haplotype identity and molecular diversity

We recorded complete sequences for a total of 603 individuals (Table 7). Most of the 417 base fragment was invariant among these samples, but at five positions there were mutations that resulted in a total of six haplotypes (Table 7). Haplotype B is the most common haplotype occurring in 72% of the samples shown in Table 7. Haplotype E is the least common, occurring in only one individual in Lewis_03. Haplotypes A, C – F are each one mutation different from Haplotype B, and two mutations different from each other (Figure 8). Mutations associated with Haplotypes A, C – E do not result in a change in amino acid structure of cytochrome b, compared with Haplotype B; however, Haplotype F is unique among the six haplotypes in having a different amino acid sequence (Figure 8). Haplotype F is a rare haplotype occurring in only eight individuals, three each in Ebright_01 and Lewis_01 (Table 7, Figure 9), and two in LakeSam_03 (not shown). In fact, the two 2001 populations show the greatest amount of both haplotype and nucleotide diversity among all collections in the data set. Besides Meadow_99, which was fixed at Haplotype B, Lewis_00 showed the least amount of molecular diversity (Table 7, Figure 9). Pine_08, with only eight individuals, was relatively diverse, showing three haplotypes, including Haplotype A, which occurred in Lewis_03 and LakeSam_03, and is the primary haplotype in Webster_05. All individuals from the 2004 collections and IssaEarly_03 (not shown Table 7, see Table 1) have Haplotype B.

Geographic structure of populations

With the exception of Ebright_01, Ebright and Laughing Jacobs creeks show no differentiation in terms of haplotype distribution or $\Phi_{ST}$ (Table 8). Both Ebright_01 and Lewis_01 show large $\Phi_{ST}$ values and are significantly differentiated from all populations, although they are not significantly different from each other or Laugh_03 at the P=0.05, adjusted. No differentiation from Laugh_03 may be due to a lack of power associated with small sample size for Laugh_03. The pattern of differentiation for Lewis Creek is complicated. Lewis_00 has a low level of both

haplotype and nucleotide diversity and is significantly differentiated from all other populations at P=0.05, but is not differentiated from Ebright_00 and Lewis_03 at P=0.05, adjusted. A similar pattern exists for Lewis_03; however, the only two populations significantly different from Lewis_03 at P=0.05, adjusted are Ebright_01 and Lewis_01. It appears that Ebright and Laughing Jacobs creeks have a relatively strong signal of sharing haplotypes, while the signal between Ebright and Lewis is weaker. As with the microsatellite data, Ebright_01 and Lewis_01 are divergent from other collection years from all other populations, but show similarities between themselves.

To test for the statistical effects associated with the interaction between tributary and collection year, we designed four AMOVA models (Table 9). When all collection years are included and collections are grouped by tributary, there is no statistical difference among tributaries in haplotype frequencies, but there is a significant year effect within tributaries. When all collection years are included and collections are grouped by year, there are significant differences among years and among tributaries within years, although the tributary effect (4.08% of the total variance) is weaker than the year effect (14.84%). When collection years 2001 and 2004 are removed, all the significant molecular variance resides within collections, with no significant among tributary, or among year-within tributary effects (Table 10). Since the 2004 collections included only five samples each from Ebright and Lewis creeks, and all ten samples are of Haplotype B, it appears that it is the removal of collection year 2001 that affects the level of significance among groups and among populations-within groups. When we restrict the analysis to collection years 2000 and 2003 from Ebright and Lewis creeks, there is a significant among tributary effect, but no significant among year within tributary effect. These data suggest that even if collection year 2001 is removed from the analysis, there is still a significant difference in haplotype frequencies between Ebright and Lewis creeks.

*Cytochrome b, relatedness, immigration, and microsatellite structure within collections*

Cytochrome b haplotypes define maternal lineages and from the results discussed above each collection has two or more maternal lineages (Table 7, Figure 9). For most collections a single maternal lineage (Haplotype B) is numerically dominant, but in the 2001 collections, Haplotype C is either the most common (Ebright_01) or in large numbers (Lewis_01). We used AMOVA

and FCA to determine if maternal lineages, family groups, and/or immigration (defined using individual assignments) produce microsatellite structure within collections. In all three tributaries different maternal lineages have significantly different microsatellite allele frequencies within collection year, although in each case the total amount of molecular variance explained by these differences ranges from 1.08 to 2.44% (Table 10).

To visualize the relative effects of maternal lineages, family groups, and/or immigration on microsatellite allele frequencies, we conducted a series of factorial correspondence analyses, one each for each Ebright, Lewis, Laughing Jacobs, and Lake Sammamish collection (Figures 10 – 12, respectively). In Ebright Creek each collection year showed different factors affecting microsatellite allele frequencies (Figure 10). In 2000, there appears to be very little microsatellite structure, as revealed by the FCA, with the possible exception of a group of three closely related individuals with a slightly divergent set of allele frequencies, which are also shared by unrelated individuals. However, in 2001, family groups (and correlated maternal lineages within the family groups) dominate microsatellite structure, and in 2003 the one individual with Haplotype C shows diverged allele frequencies.

A slightly different pattern exists for Lewis Creek (Figure 11). Here, in 2000 one, possibly two family groups affect microsatellite structure, while in 2001, a family group, maternal lineages (note different FCA distributions for Haplotypes B and C), and immigration (single individual assigned to Laughing Jacobs Creek) show divergent microsatellite allele frequencies. In 2003, three immigrants, one with the only Haplotype E in the entire data set, show extremely divergent allele frequencies. In additional the microsatellite allele frequencies for individuals with Haplotype D appear slight different from that of Haplotype B.

In Laughing Jacobs Creek, immigrants appear to be the most dominant factor affecting microsatellite allele frequency differentiation (Figure 12), although in 2000 one family group, with Haplotype D, also appears to be differentiated. Finally, microsatellite structure within the Lake Sammamish aggregate sample in 2003 appears to be entirely a function of immigrants with the six fish whose origins are not from within Ebright, Lewis, or Laughing Jacobs creeks showing the most divergent allele frequencies (Figure 12).

**Discussion**

The primary objective of this research project was to ascertain the genetic relationships among kokanee populations spawning in the three primary tributaries to Lake Sammamish (Ebright, Laughing Jacobs, and Lewis). As part of the process to achieve this objective, we examined not only among tributary genetic differentiation, but also the within collection and within tributary genetic properties. All Lake Sammamish kokanee collections appear to possess relatively equal microsatellite molecular diversity, although some populations (Lewis_03 and Ebright_01) are slightly more diverse. The 2001 collections from Ebright and Lewis creeks appear to have greater cytochrome b nucleotide and haplotype diversity, while Lewis_00 appears relatively devoid of cytochrome b diversity. There also appears to be considerable molecular disequilibrium within the Lake Sammamish kokanee collections.

One cause of molecular disequilibrium is hidden genetic structure within a population (either through natural processes or as a sampling artifact). That is, if populations are aggregates of smaller populations with little or no reproductive interaction among these smaller units, molecular disequilibrium can occur. This process is known as the Wahlund effect (Frankham et al., 2002). We have shown above that for each tributary, maternal (haplotype) lineages may have different allele frequencies and may have contributed to genetic disequilibrium (Figure 10-12). Immigration may also produce a Wahlund effect, especially if the immigration is recent and genetic introgression is minor or non-existence. Microsatellite structure in Ebright_03, Lewis_01, and especially Lewis_03, Laugh_00, and Laugh_03 appear to be affected by the presence of fish identified as being from populations other than the one being sampled. Eleven of the 13 individuals identified as coming from populations not present in this data set (see Figure 7) showed divergent allele frequencies in each of the collections in which they occurred. It is not know if these individuals were part of the breeding population within each collection, but the fact that they were present indicates that at least there is a potential to interbreed and therefore a potential genetically homogenize populations. There also appears to be some mixing among fish from within Ebright, Lewis, and Laughing Jacobs creeks (Figures 5 and 6), and in many collections these immigrant fish did not show greatly differentiated microsatellite allele

frequencies than the resident fish collected from natal tributaries (bottom panels in Figures 10-12).

Another source of molecular disequilibrium is the presence of family units (and therefore the possibility of inbreeding) within collections. The high variance of relatedness values for all collections suggests that these collections are composed of family groups. This is especially true for Ebright_01, Laugh_00, Lewis_00, and Lewis_01, where family groups show divergent microsatellite allele frequencies (Figures 10-12). These four collections also showed the greatest number of individuals with at least one pairwise relatedness value greater than or equal to 0.40 (Table 6). Despite the presence of family groups within collections, there are no apparent effects from inbreeding or genetic drift, in that the numbers of alleles per locus and heterozygosity do not appear to be diminished, compared with Meadow_99, for example (Table 3, Figure 2).

In summary, the populations used in this analysis may be influenced by several factors affecting their genetic equilibrium. In many respects, these populations appear as aggregates in themselves, composed of family units, immigrants, and/or different maternal lineages. The fact that maternal lineages may have different allele frequencies within some populations suggests that these maternal lineages are not interacting reproductively. This can result from slight differences in spawn timing or location, or from assortative mating. Also, these maternal lineages may represent different life histories within each tributary, and these collections may be aggregates of kokanee, sockeye, and/or residualized sockeye.

Genetic assignment tests suggest that there can be mixing among fish from each tributary (i.e., inter-tributary immigration), and this appears to occur at a higher rate in Laughing Jacobs Creek than the other two tributaries (Figures 5 and 6). These assignment tests also indicated that there are fish within the Lake Sammamish watershed that did not originate from any of the tributaries within Lake Sammamish (that is, tributaries from which we have collected baseline data). There are at least four possibilities as to the sources of these fish: (1) Sammamish River kokanee or sockeye, (2) Lake Sammamish sockeye, (3) Lake Sammamish kokanee from one or more unsampled population(s) (e.g., beach spawning kokanee; Berge, King County WLRD, pers.

comm., 2008), and (4) sockeye or kokanee fish transplanted from out of basin stock. Genetic analyses of additional data may be able to resolve this question.

Finally, Ebright, Laughing Jacobs, and Lewis creeks are genetically differentiated, and these differences seem to be maintained across all collections years. Ebright and Laughing Jacobs appear more similar to each other than either is to Lewis in both their microsatellite and cytochrome b structure, but this is most apparent with cytochrome b, where there are no statistical differences in the haplotype frequencies for both 2000 and 2003. In general, within tributary collections for 2000 and 2003 are similar for all three tributaries, although for Lewis, there appears to be slightly more differences between 2000 and 2003 than in the other two tributaries (Tables 4 and 8).

**Conclusions**

1. If the genetic samples present here appropriately represent their source population (i.e., collections are not biased), many populations are in genetic disequilibrium.
2. Depending on the collection analyzed, the genetic disequilibrium appears to be a function of the family groups, immigration, and difference in the genetic structure among maternal lineages.
3. If immigration is affecting genetic equilibrium, the source population for the immigrants must have different allele frequencies than the population receiving the immigrants. If this is the case, this suggests that immigration is a relatively recent phenomenon (i.e., insufficient time for population to reach equilibrium).
4. The presence of genetic structure associated with differences in maternal lineages suggests that within a tributary there may be structure associated with spawn timing or location, or, these maternal lineages may represent different life histories (kokanee, sockeye, or residualized sockeye). Hans Berge (King County, WLRD, pers. comm., 2008), who collected many of these samples, states that it is very doubtful that samples taken from within Ebright, Laughing Jacobs, or Lewis creeks in 2000, 2001, and 2003 are

sockeye. Nevertheless, the "populations" analyzed here might actually represent aggregate samples.

5. Fish from collection year 2001 show very different genetic characteristics than fish from either 2000 or 2003.
6. There is evidence that fish whose origins are not from any of the Lake Sammamish kokanee baseline populations (Ebright, Laughing Jacobs, Lewis creeks) are present within Lake Sammamish. These fish may be Sammamish River kokanee/sockeye, Lake Sammamish sockeye, beach spawning kokanee from within Lake Sammamish, or sockeye or kokanee fish transplanted from out of basin stocks.
7. Ebright, Laughing Jacobs, and Lewis creeks show significant among tributary population structure. Fish from Ebright and Laughing Jacobs creeks appear more genetically similar to each other than either is to Lewis Creek.

**Recommendations**

The purpose of this project was to determine the genetic structure of kokanee population within the Lake Sammamish watershed, and to make use of these data to help determine the appropriate management plan for these populations. The conclusions from this project can be generalized into four observations that should have bearing on kokanee management activities. First, based on the genetic samples used in this study, kokanee populations cannot be clearly defined. That is, many of the collections from Ebright, Laughing Jacobs, Lewis creeks appear to be aggregates, composed primarily of fish from that tributary (with the exception of Laughing Jacobs), but from different maternal lineages and different family groups. One possible concern is that there are two different forms of *Oncorhynchus nerka* present in the tributaries, and these forms are not easily discernable from external characteristics. If this is true, we will need to design a genetic screening process to determine if we have the appropriate brood stock from the tributary. We would need to analyze additional samples to design this genetic screen. Second, a small number of fish from outside the Lake Sammamish basin appear to be present in the Lake Sammamish system. Alternatively, there is an unknown (and therefore uncollected) kokanee population within Lake Sammamish. As with the maternal lineage issue discussed above, the presence of

these alien fish can complicate a hatchery program if they become incorporated as broodstock. We should be able to design a genetic screen for these fish as well, but again, this would require the analysis of additional data. Third, there does appear to be movement of fish from one tributary to another, but because many of the kokanee populations analyzed here are in disequilibrium, it is possible that this movement is fairly recent. Regardless of the degree to which fish may move about, from these data, it seems that immigration has not removed genetic differences among the tributaries. Fourth, the three tributaries populations are distinct genetically, although there does appear to be a greater similarity between the Ebright and Laughing Jacobs creeks, than either has with Lewis Creek. This differentiation appears to be long term in that the some maternal lineages (e.g., Haplotype B) present in different tributaries have significantly different allele frequencies. A hatchery program designed to preserve the genetic diversity present in the Lake Sammamish watershed would need to recognize that the significant genetic differences among the tributary populations represents unique genetic diversity. However, based on the current structure of Lake Sammamish kokanee populations, if the data presented here is indicative of current and future collections from Ebright, Lewis, and Laughing Jacobs creeks, any random sample of adults from each of these tributaries (as you would take for a hatchery supplementation program) may include: (1) closely related individuals, (2) individuals from other tributaries, and (3) out-of-basin individuals.

**Acknowledgments**

Funding for this project was generously provided by Trout Unlimited, King County Department of Natural Resources and Parks, WDFW, and the Washington State General Fund. This project benefited greatly from discussions with Denise Hawkins, Craig Busack, Sewall Young, and Scott Blankenship. Hans Berge (King County, WLRD), and Scott Blankenship and Anne Marshall (WDFW) provided comments of this report.

**Literature Cited**


Banks, M.A., M.S. Blouin, B.A. Baldwin, V.K. Rashbrook, H.A. Fitzgerald, S.M. Blankenship, and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tshawytscha*). J. Hered. 90:281-288.

Belkhir K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Belkhir, K., V. Castric, and F. Bonhomme. 2002. IDENTIX, a software to test for relatedness in a population using permutation methods. Mol. Ecol. Notes 2:611-614.

Benson, D.A., I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and D.L. Wheeler. 2007. GenBank. Nucl. Acids Res. 35 (Database Issue): D21-D25.

Berge, H.B., and K. Higgins. 2003. The current status of kokanee in the greater Lake Washington Watershed. King County Department of Natural Resources and Parks, Water and Land Resources Division. Seattle, Washington. 50pp.

Brownstein M.J., J.D. Carpten, and J.R. Smith. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20:1004–1010

Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:149-491.

Excoffier, L., G. Laval, S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Frankham, R., J.D. Ballou, and D.A. Briscoe. 2002. Introduction to Conservation Genetics. Cambridge University Press, Cambridge. 617 + xxi pp.

Goudet J, M. Raymond, T. De Meeüs, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144:1933-1940.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

Kalinowski, S.T. 2003. Genetic Mixture Analysis 1.0. Department of Ecology, Montana State University, Bozeman MT 59717. Available for download from http://www.montana.edu/kalinowski

Kassler, T.P. 2005. A microsatellite DNA analysis of Walsh Lake kokanee. Unpublished report, WDFW. 11 pp.

Liu, K., and S.V. Muse. 2005. PowerMarker: Integrated analysis environment for genetic marker data. Bioinformatics 21:2128-2129.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Olsen, J.B., S.L. Wilson, E.J. Kretschmer, K.C. Jones, and J.E. Seeb. 2000. Characterization of 14 tetranucleotide microsatellite loci derived from sockeye salmon. Mol. Ecol. 9:2185-2187.

Queller, D.C., and K.F. Goodnight. 1989. Estimating relatedness using genetic markers. Evolution 43:258-275.

Ranalla B., J.L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. USA 94: 9197-9201

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

Rexroad III, C.E., R.L. Coleman, A.M. Martin, W.K. Hershberger, and J. Killefer. 2001. Thirty-five polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). Anim. Gen. 32:316.331.

Scribner, K.T., J. Gust, and R.L. Fields. 1996. Isolation and characterization of novel salmon microsatellite loci: Cross –species amplification and population genetic applications. Can. J. Fish. Aquat. Sci. 53:833.841.

Small, M.P., T.D. Beacham, R.E. Withler, and R.J. Nelson. 1998. Discriminating coho salmon (*Oncorhynchus kisutch*) populations within the Fraser River, British Columbia. Mol. Ecol. 7:141-155.

Spies, I.B., E.C. Anderson, K. Naish, and P. Bentzen. 2007. Evidence for the existence of native population of sockeye salmon (*Oncorhynchus nerka*) and subsequent introgression

with introduced populations in a Pacific Northwest watershed. Can J. Fish. Aquat. Sci. 64:1209-1221.

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

Young, S.F., M.R. Downen, and J.B. Shaklee. 2004. Microsatellite DNA data indicate distinct native populations of kokanee, *Oncorhynchus nerka*, persist in the Lake Sammamish Basin, Washington. Environ. Biol. Fishes 69:63-79

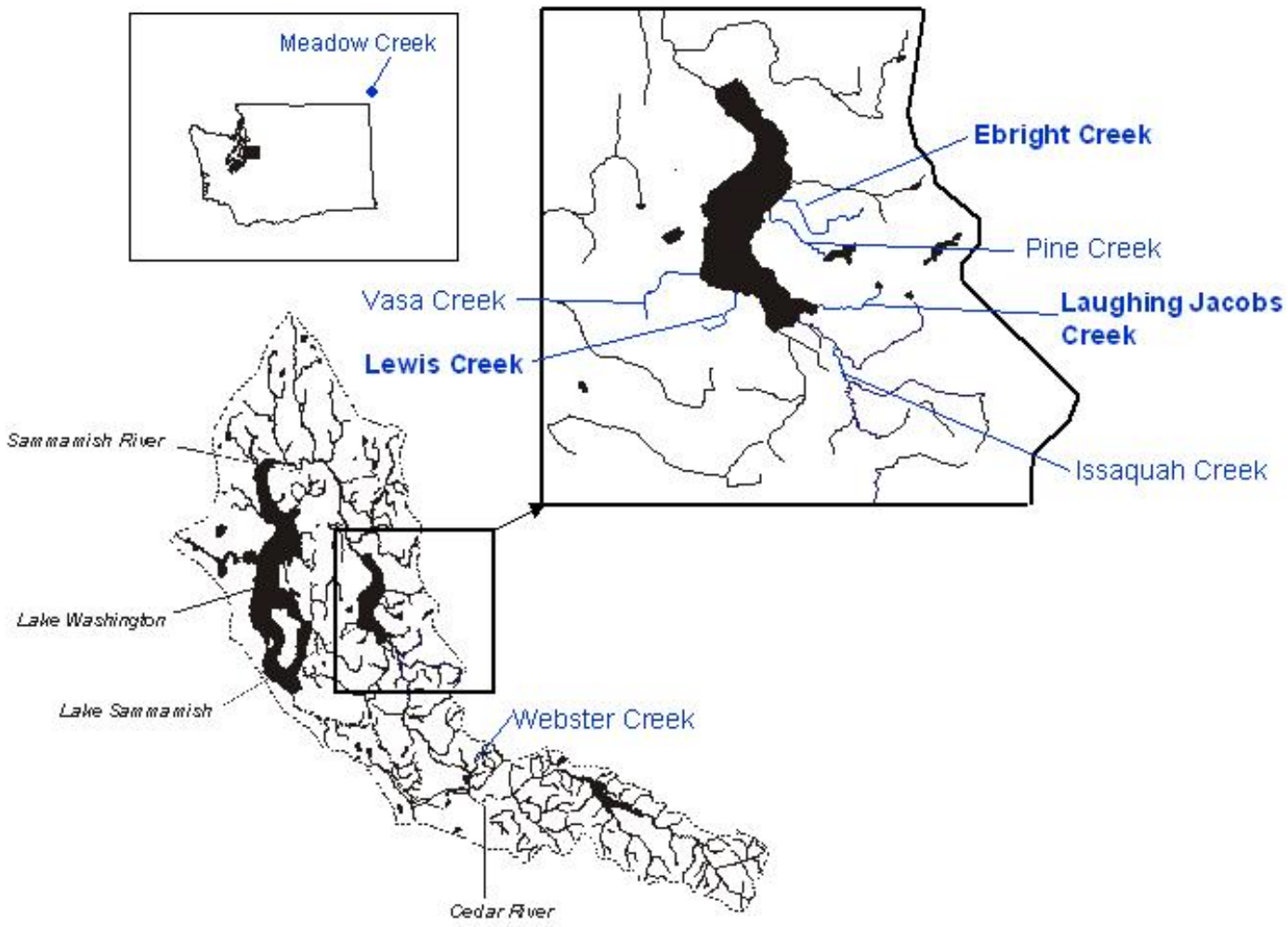


**Figure 1.** Collections localities, with the lower left plot detailing WRIA 8 and the Lake Washington/Lake Sammamish watershed. The upper right corner plot is a portion of the Lake Sammamish watershed, with the primary Lake Sammamish tributary collections in bold.

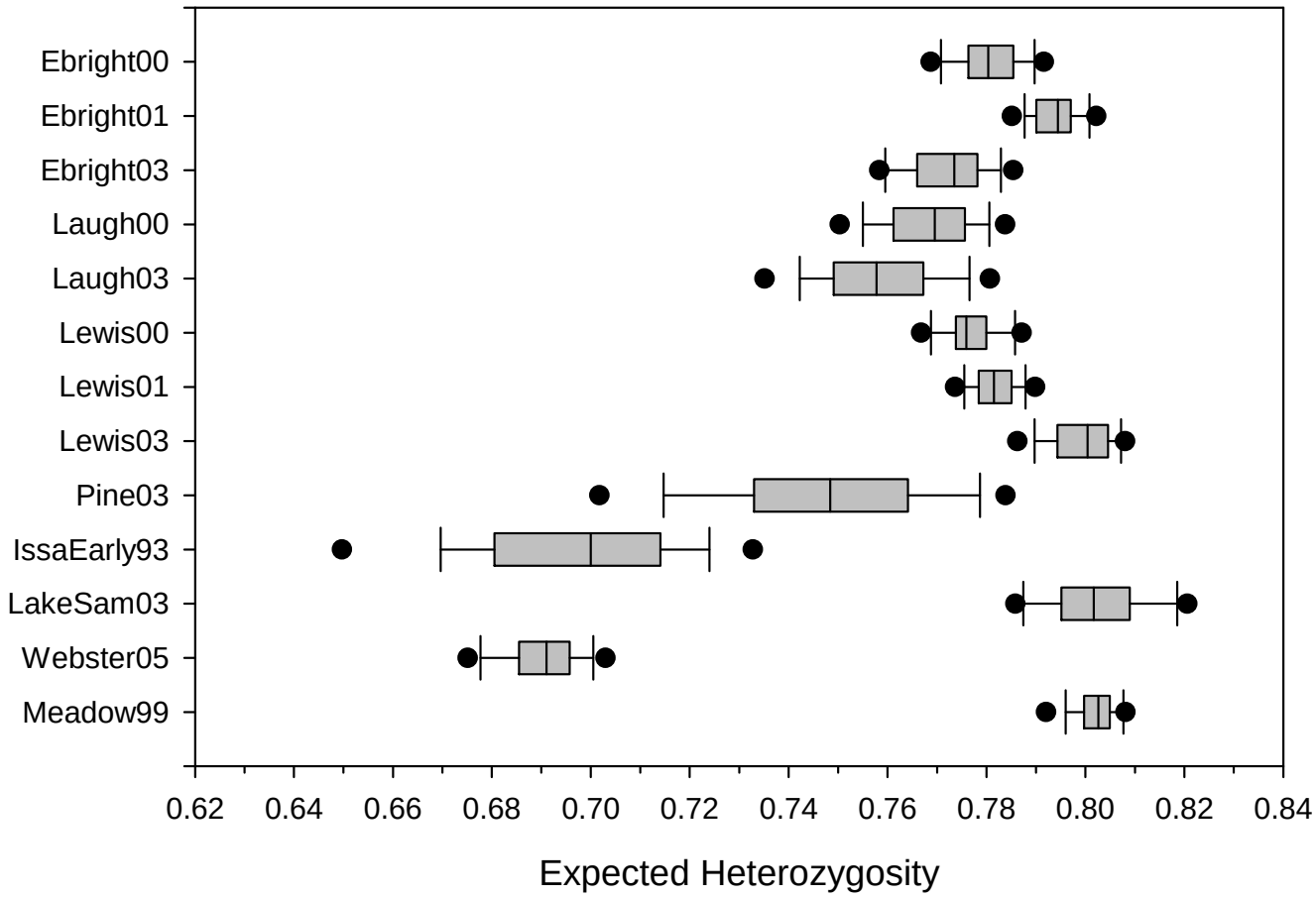


**Figure 2.** Difference in expected heterozygosity among collections as measured by a bootstrap resampling procedure (see Materials and Methods). For each collection, the box represents the range for 50% of the bootstrap runs (25$^{th}$ [left edge of box] to 75$^{th}$ [right edge] percentile, and the median value is the vertical line within the box. The bars that extend out from the box denote the 10$^{th}$ and 90$^{th}$ percentiles, while the dots indicate the 5$^{th}$ and the 95$^{th}$ percentile.

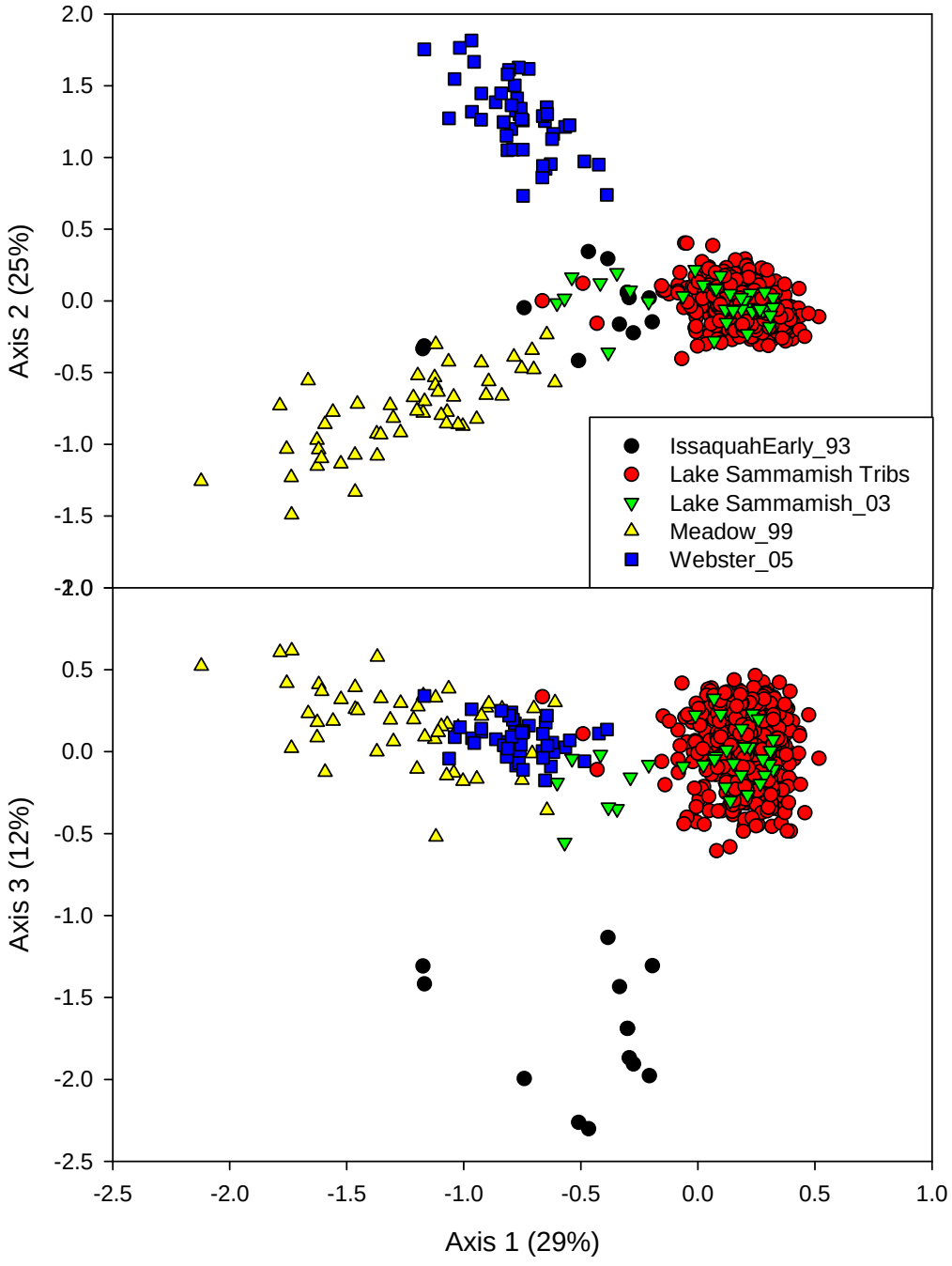


**Figure 3.** Factorial correspondence analysis Axes 1 versus 2 (top) and Axes 1 versus 3 (bottom) for 13 collections in the data set. All Lake Sammamish tributary collections are represented as one symbol.

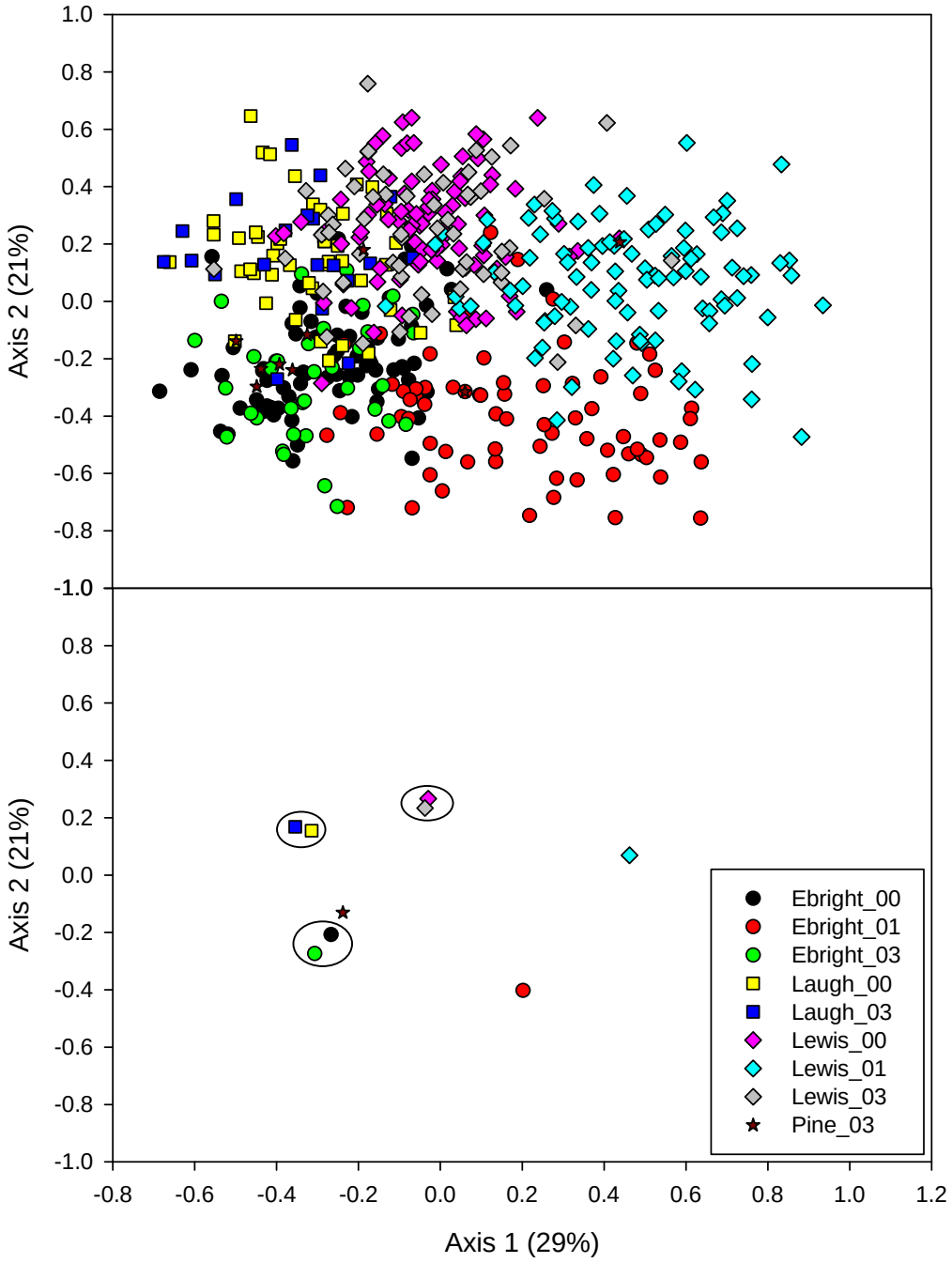


**Figure 4.** Factorial correspondence analysis of Lake Sammamish tributary collections only. Upper plot show the ordination of individual samples, while the bottom plot shows the centroids for each collection. Collection centroids enclosed within a circle are not significantly different, as measured by Mahalanobis distances (see Materials and Methods).

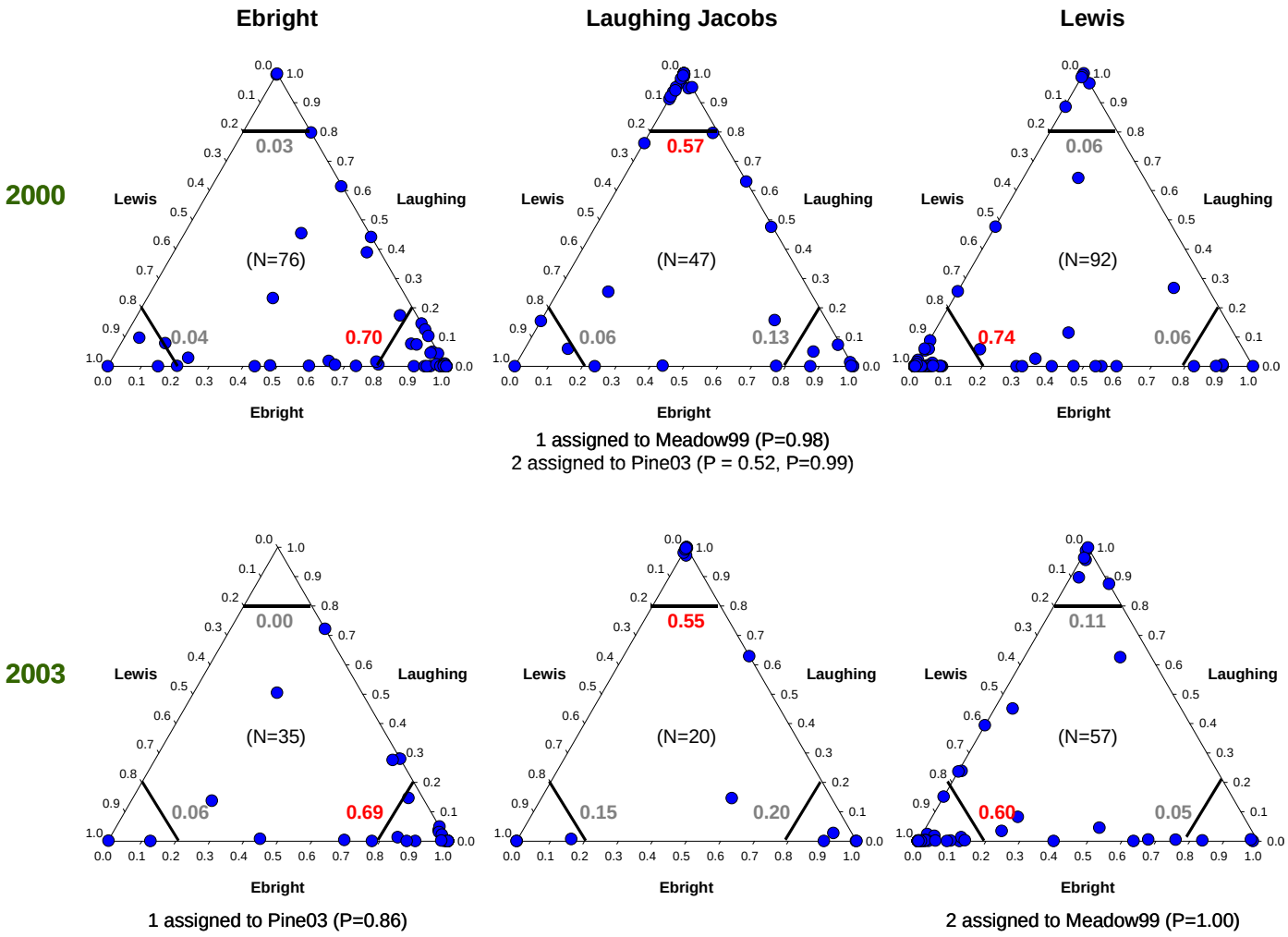


**Figure 5.** Ternary plots showing assignment probabilities for all individuals from Ebright (left), Laughing Jacobs (center), and Lewis (right) sampled in 2000 (top) and 2003 (bottom). Individuals with assignment probabilities equal to or greater than 0.80 to Ebright, Laughing Jacobs, and Lewis creeks are shown in the lower right, top, and lower left of each plot, respectively, and are denoted by proportions (red proportions are correct assignments). For example, 0.70 of the individuals from Ebright00 were assigned to Ebright, and 0.07 were assigned to either Laughing Jacobs or Lewis. Twenty-three percent were unassigned.

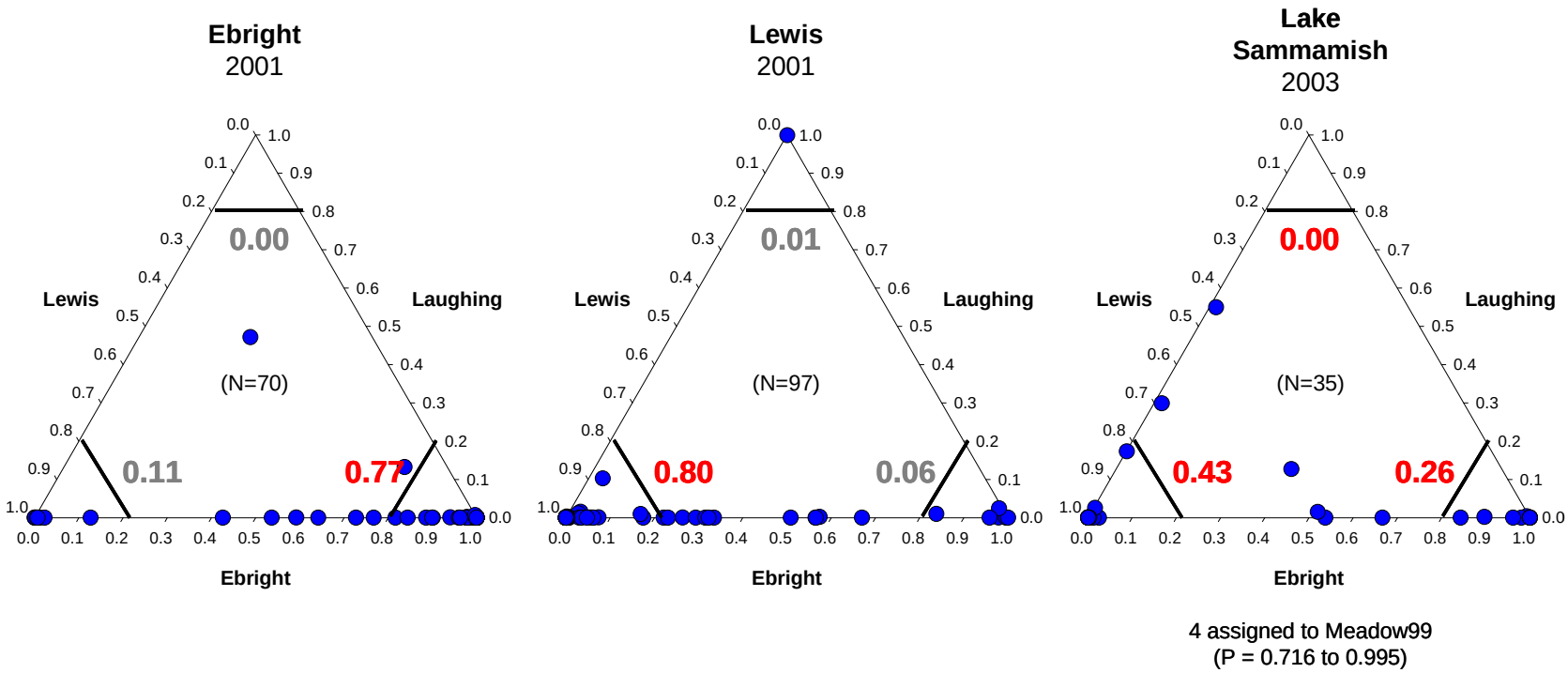


**Figure 6.** Ternary plots showing assignment probabilities for all individuals from Ebright (left), Lewis (center), and Lake Sammamish (right) sampled in 2001 (Ebright and Lewis) and 2003 (Lake Sammamish). See Figure 5 for details.

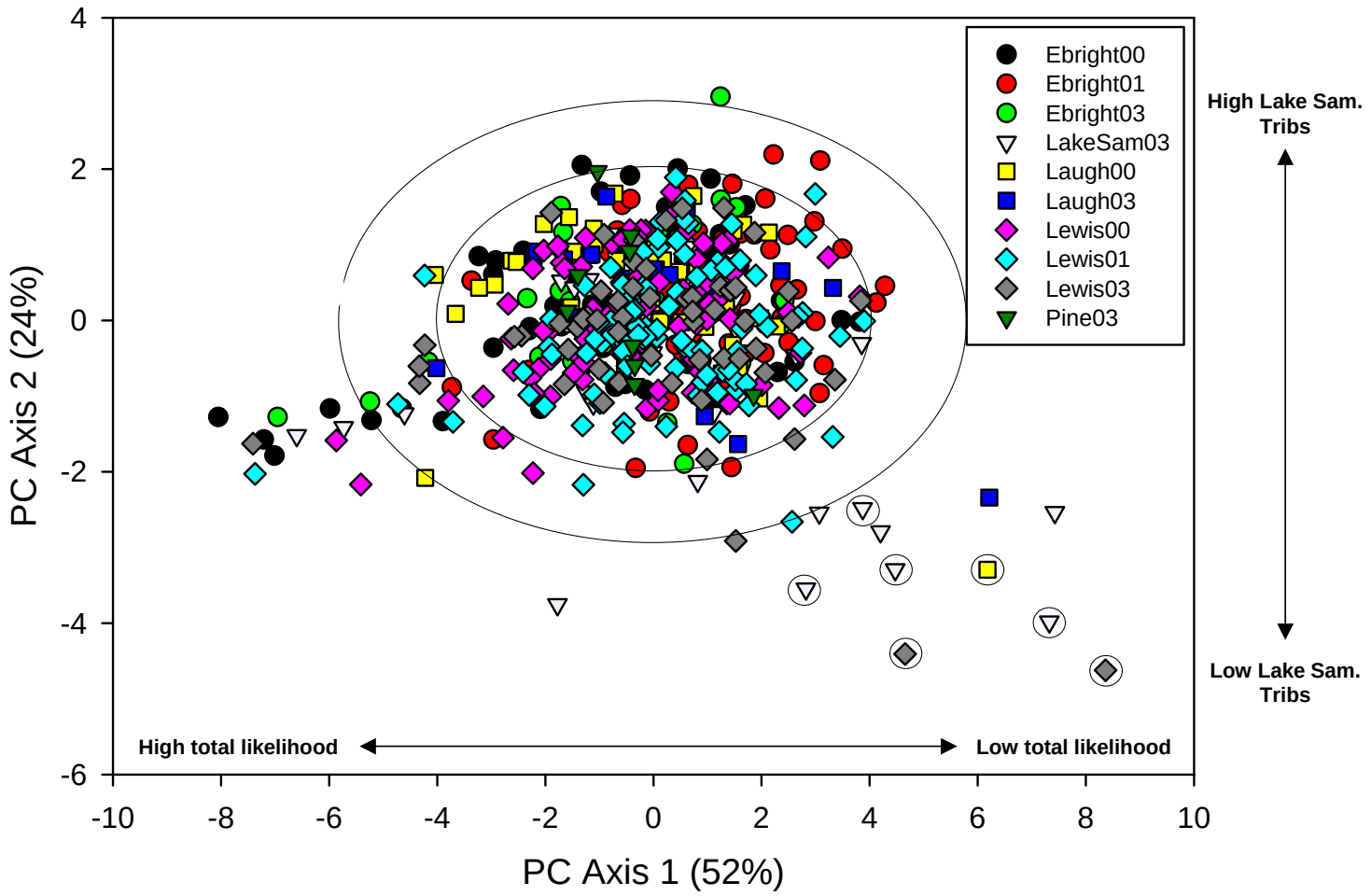


**Figure 7.** Principal component analysis (PCA) of the log-likelihood scores for all individuals sampled from within Lake Sammamish watershed. Legend refers to the source population of samples. The circles around seven of the outliers denote individuals that were assigned to Meadow_99 (see Figures 5 and 6). The 95% and 99.5% confidence ellipses are approximated by the inner (±1.96 standard deviations) and outer ellipses (±2.81 standard deviations), respectively

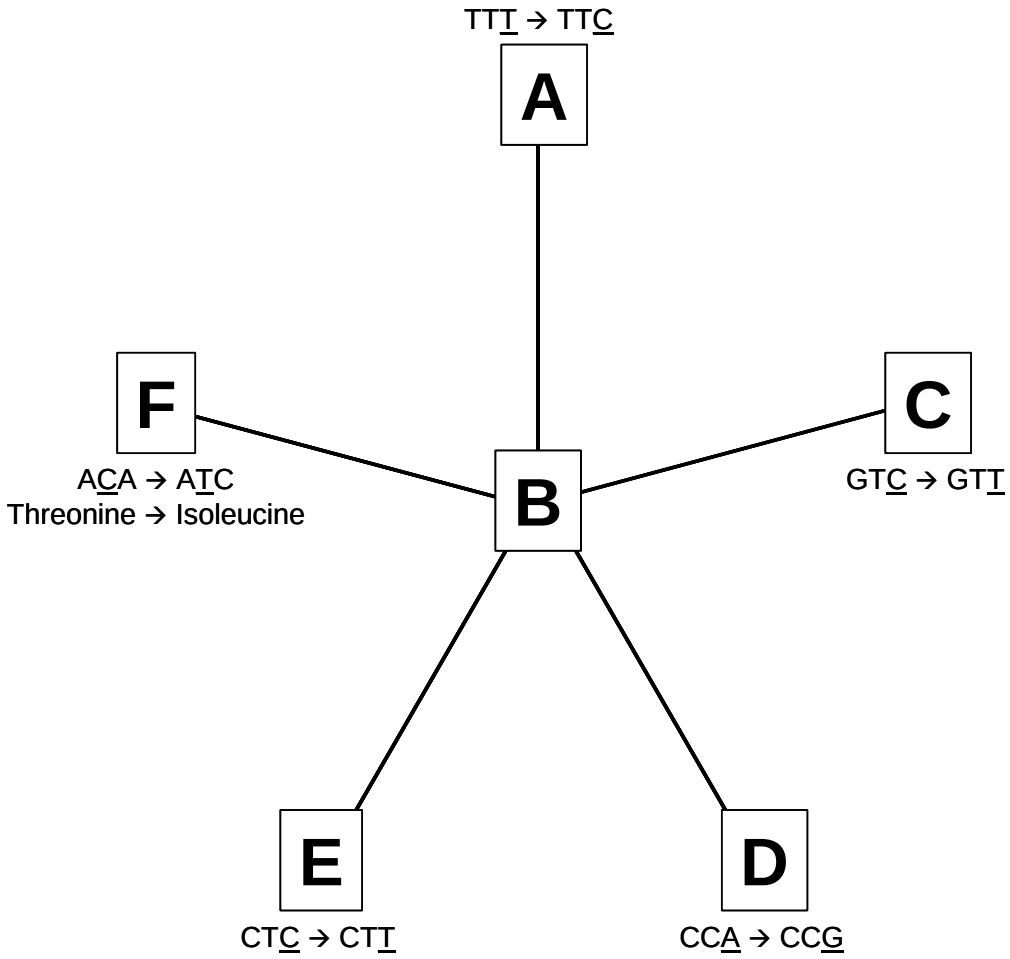


**Figure 8.** Unrooted minimum evolution tree for the six cytochrome b haplotypes. Each line connecting the haplotypes represents one mutation. Under each haplotype we show the mutation relative to Haplotype B. For example Haplotype A differs from Haplotype B by having a C at position 15,649 (see Table 8) rather than a T. Cytochrome b is a protein coding gene and mutations associated with Haplotypes A, C – E occur in the third position of the triplet coding for a particular amino acid. Mutations in the third position are neutral and do not result in a change in amino acid. Haplotype F is defined by a mutation in the second position in the triplet and resulted in the change in the amino acid structure of the protein.

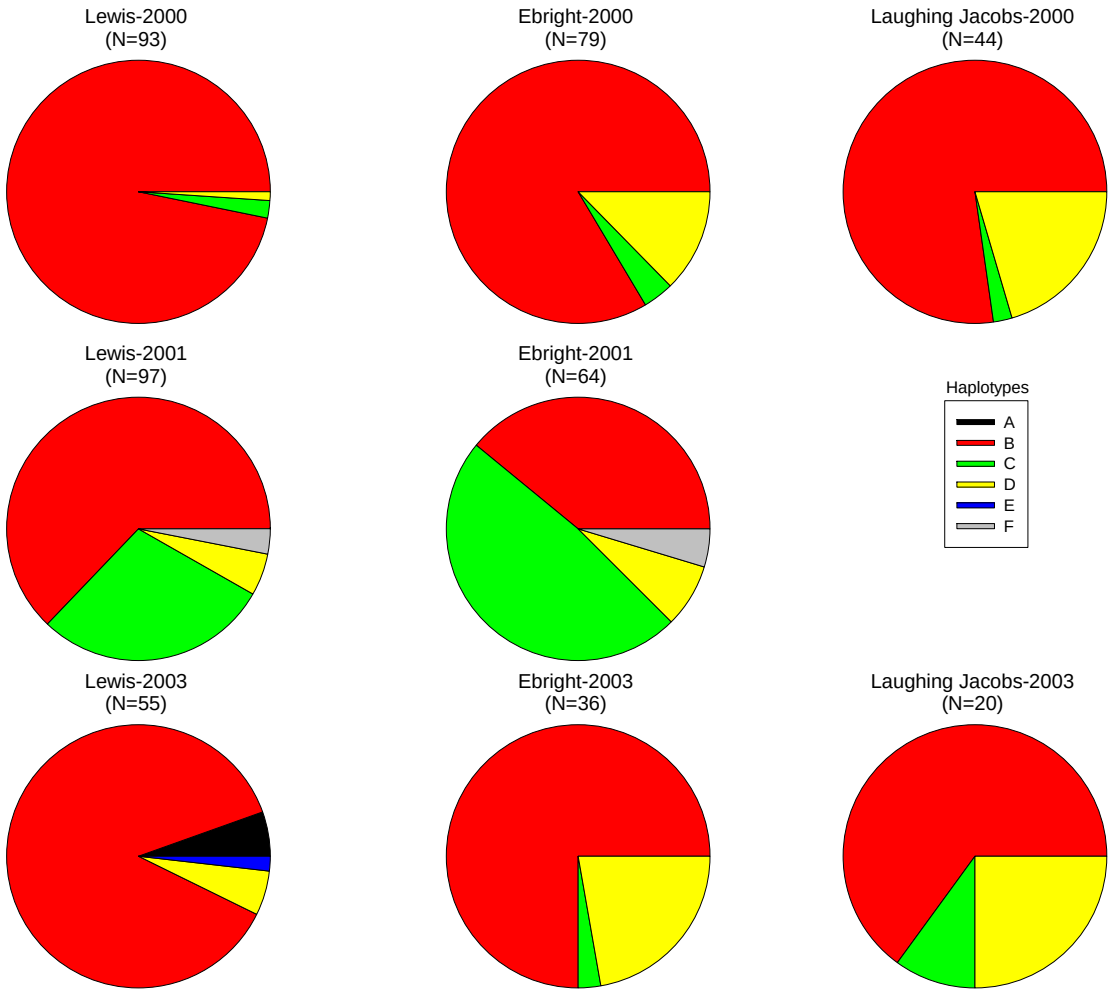


**Figure 9.** Cytochrome b haplotype frequencies for the collections from the primary populations of kokanee from Lake Sammamish

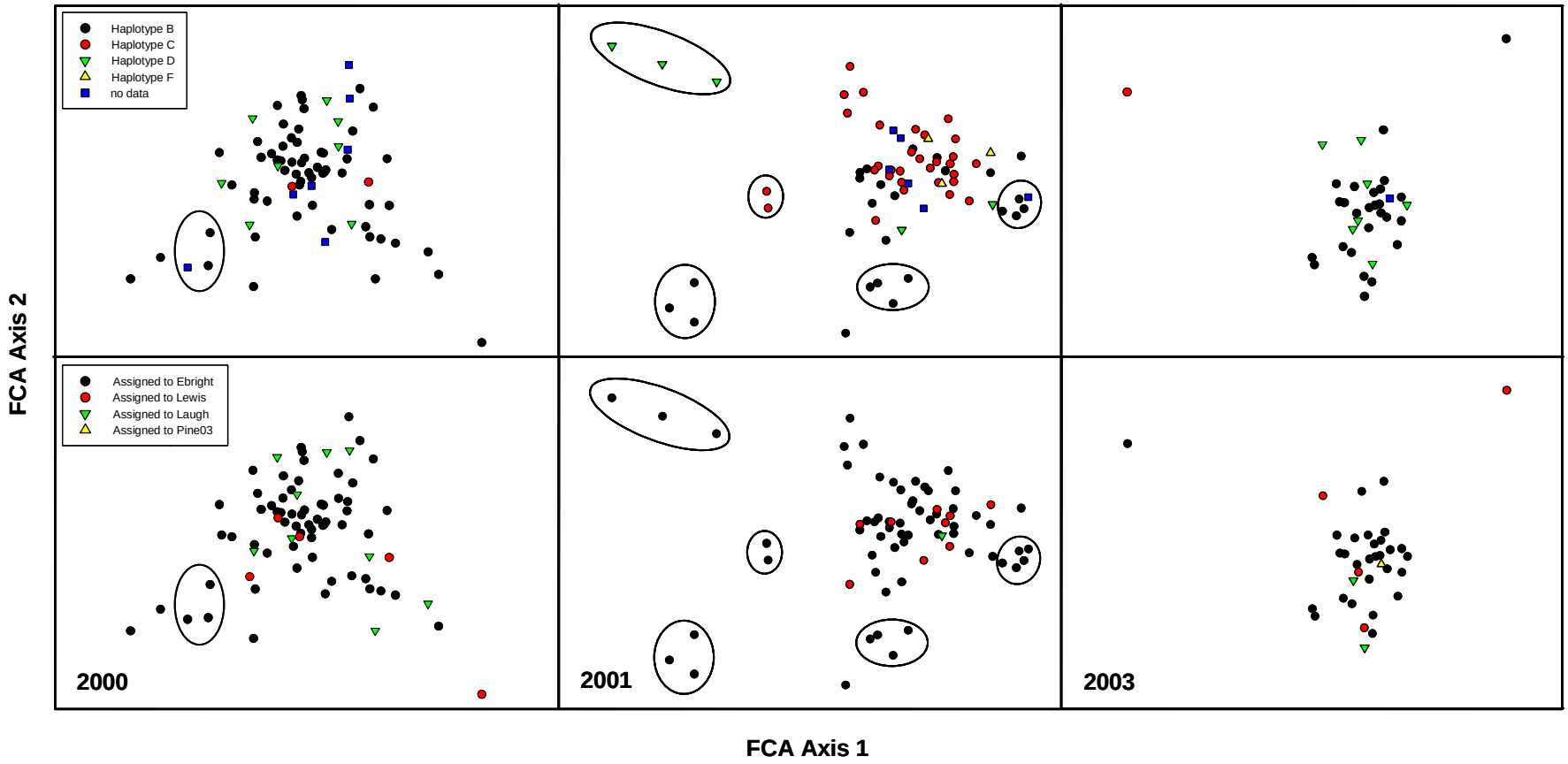


**Figure 10.** Factorial correspondence analysis for the three temporal collections from Ebright Creek. Top and bottom panels for each collection year show the same analysis, except in the top panels the individuals are shown as cytochrome b haplotypes and in the bottom panels the individuals are shown as their genetic assignments (as in Figures 5 and 6, except here the assignment criterion is 0.50). The legends in the top and bottom left panels apply to each of the top and bottom panels, respectively. Ellipses enclose groups of individuals with relatedness values greater than or equal to 0.40 (see Table 6). Not all groups of related individuals are denoted with an ellipse.

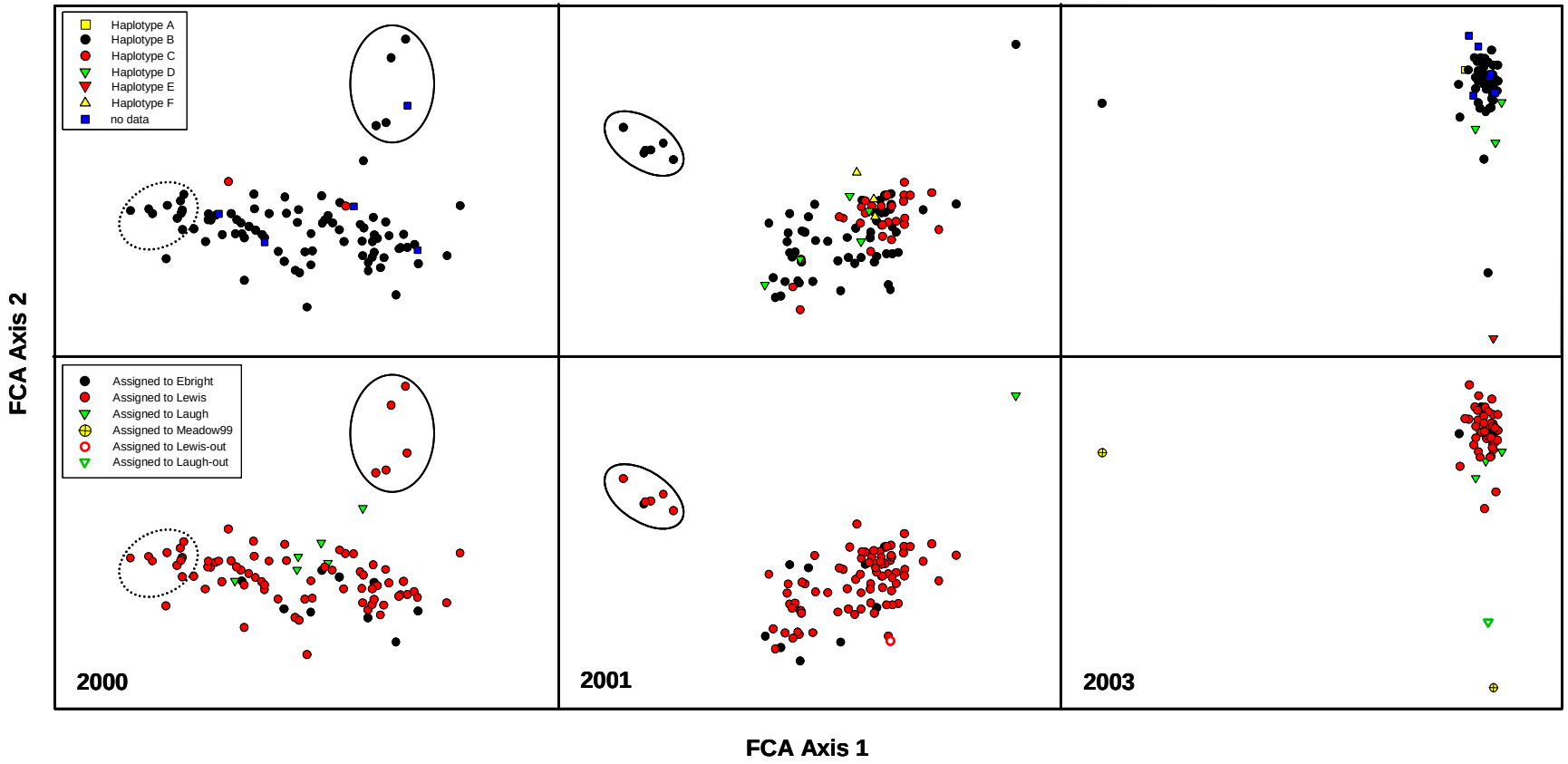


**Figure 11.** As in Figure 10, except factorial correspondence analysis for the three temporal collections from Lewis Creek. Dotted ellipse shown in the 2000 collection indicates that not all individuals within the ellipse are closely related; there are six individuals within the dotted ellipse with relatedness values greater than or equal to 0.40. Lewis- and Laugh-out refer to samples assigned to Lewis or Laughing Jacobs creeks, respectively, but were outside the 99.6% confidence ellipse in Figure 7 and are most likely from a populations not included in this data set.

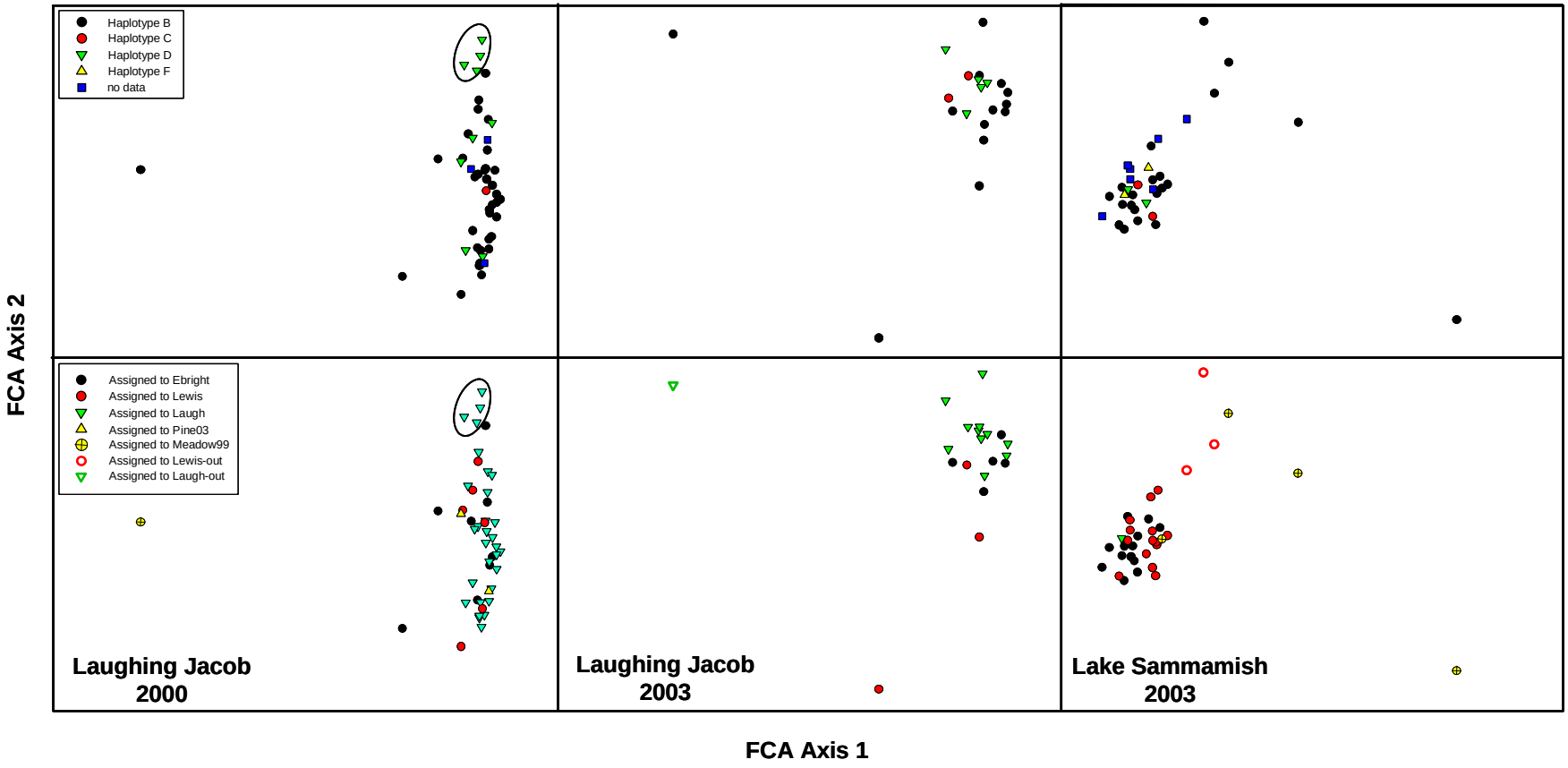


**Figure 12.** As in Figure 10, except factorial correspondence analysis for the two temporal collections from Laughing Jacobs Creek, and the 2003 collection from within Lake Sammamish. Lewis- and Laugh-out refer to samples assigned to Lewis or Laughing Jacobs creeks, respectively, but were outside the 99.6% confidence ellipse in Figure 7 and are most likely from a populations not included in this data set.

**Table 1.** Names, descriptions, and PCR conditions for the microsatellite loci used in this analysis. All loci within a multiplex were amplified together in a single reaction, except for Multiplex C, where only One105 and Ots103 where amplified together. One101 was pooled with One105 and Ots103 for electrophoresis with the ABI 3730. $H_o$ refers to the observed heterozygosity, with shading and bold equal to significant heterozygote deficit at P= 0.05 and P=0.05, adjusted for multiple comparisons, respectively.

| Locus Name | Reference[1] | Dye | Primer (mM) | Multiplex | Annealing Temp | # PCR Cycles | Total # Alleles | Allelic Range (bp) | | $H_o$ | Young et al.[2] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Min | Max | | |
| One-108 +a | a | 6fam | 0.09 | One-A | | | 18 | 190 | 266 | 0.83 | ✓ |
| One-110 +a | a | hex | 0.06 | One-A | 55° | 35 | 25 | 216 | 297 | 0.88 | ✓ |
| One-100 +a | a | ned | 0.11 | One-A | | | 36 | 304 | 480 | **0.88** | ✓ |
| One-102 +a | a | 6fam | 0.08 | One-B | | | 16 | 203 | 279 | 0.80 | ✓ |
| One-114 +a | a | vic | 0.10 | One-B | 53° | 40 | 25 | 209 | 321 | 0.86 | ✓ |
| One-115 +a | a | ned | 0.05 | One-B | | | 17 | 182 | 263 | 0.84 | ✓ |
| One-105 +a | a | 6fam | 0.08 | One-C1 | | | 6 | 132 | 153 | 0.64 | ✓ |
| Ots-103 +a | b | vic | 0.07 | One-C1 | 55° | 35 | 21 | 146 | 229 | 0.88 | ✓ |
| One-101 +a | a | ned | 0.06 | One-C2 | | | 40 | 177 | 408 | 0.90 | ✓ |
| Omm-1135 +a | c | hex | 0.09 | One-F | 53° | 40 | 13 | 226 | 256 | 0.70 | |
| Omm-1139 +a | c | ned | 0.08 | One-F | | | 8 | 130 | 145 | 0.51 | |
| Omm-1085 +a | c | 6fam | 0.27 | One-G | | | 16 | 100 | 167 | 0.86 | |
| Omm-1070 +a | c | vic | 0.09 | One-G | 60° | 40 | 30 | 203 | 347 | 0.89 | |
| Ots-3M +a | d | ned | 0.05 | One-G | | | 11 | 135 | 173 | 0.42 | |
| One-2 +a | e | 6fam | 0.05 | One-H | | | 16 | 222 | 291 | 0.80 | |
| Omm-1142 +a | c | hex | 0.08 | One-H | 48° | 35 | 12 | 120 | 154 | 0.75 | |
| Omm-1130 +a | c | ned | 0.08 | One-H | | | 36 | 225 | 427 | 0.92 | |

[1] a = Olson et al. (2000), b = Small et al. (1998), c = Rexroad et al. (2001), d = Banks et al. (1999), e = Scribner et al. (1996)
[2] Young et al. (2004) used a subset of the markers applied in this current study. Kassler (2005) used the same set of loci as Young et al.

**Table 2.** Collection information for samples used for this analysis. Collection Code refers to the Washington Department of Fish and Wildlife, Molecular Genetics Laboratory's accession code for the collection. See Figure 1 for localities for each collection.

| Collection | Location | Collection Year | Collection Code | Collecting Agency[1] | CytoB N | Microsat N |
|---|---|---|---|---|---|---|
| Ebright_00 | Ebright | 2000 | 00HA[2] | WDFW, KC | 79 | 76 |
| Ebright_01 | Ebright | 2001 | 01JE | KC | 64 | 70 |
| Ebright_03 | Ebright | 2003 | 03ML | KC | 36 | 35 |
| Ebright_04 | Ebright | 2004 | 04IF | KC | 5 | 1 |
| Laugh_00 | Laughing Jacobs | 2000 | 00DY[2] | WDFW, KC | 44 | 47 |
| Laugh_03 | Laughing Jacobs | 2003 | 03MM | KC | 20 | 20 |
| Lewis_00 | Lewis | 2000 | 00DX[2] | WDFW, KC | 93 | 92 |
| Lewis_01 | Lewis | 2001 | 01JD | KC | 97 | 97 |
| Lewis_03 | Lewis | 2003 | 03MK | WDFW, KC | 55 | 57 |
| Lewis_04 | Lewis | 2004 | 04IC | KC | 5 | 4 |
| Pine_03 | Pine Creek | 2003 | 03MN | KC | 8 | 9 |
| Vasa_04 | Vasa | 2004 | 04ID | KC | 1 | - |
| IssaEarly_93 | Issaquah (Early) | 1993 | 93WA[2] | WDFW, KC | 13 | 13 |
| IssaEarly_03 | Issaquah (Early) | 2003 | 03MP | KC | 2 | 1 |
| LakeSam_03 | Lake Sammamish | 2003 | 03MQ | KC | 31 | 35 |
| Webster_05 | Webster | 2002 | 05MX[3] | SPU | 46 | 45 |
| Meadow_99 | Meadow | 1999 | 99OE[2] | CT | 48 | 49 |
| Total | | | | | 647 | 651 |

[1] CT = Colville Tribe, KC = King County, WLRD, SPU = Seattle Public Utilities, WDFW = Washington Department of Fish and Wildlife
[2] Collections included in Young et al. (2004) and Kassler (2005)
[3] Collection included in Kassler (2005)

**Table 3.** Measures of microsatellite molecular diversity for each of the 13 collections. Ho = observed heterozygosity, He = expected heterozygosity. For both Hardy-Weinberg and linkage disequilibrium the table shows the proportion of loci within that collection that were in disequilibrium at the P=0.05 and P=0.05 adjusted levels. Bold typeface denotes those $F_{IS}$ values that are significantly greater than zero, indicating significant deficit of observed heterozygosity, compared with expected heterozygosity.

| Collection | N | N per locus | # alleles per locus | Allelic Richness | Ho | He | $F_{IS}$ | Disequilibrium | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Hardy-Weinberg | | Linkage | |
| | | | | | | | | P=0.05 | P=0.05 adjusted | P=0.05 | P=0.05 adjusted |
| Ebright_00 | 76 | 72.82 | 12.06 | 7.16 | 0.7867 | 0.7846 | -0.003 | 0.06 | 0.00 | 0.18 | 0.01 |
| Ebright_01 | 70 | 69.29 | 12.00 | 7.30 | 0.8132 | 0.8015 | -0.015 | 0.12 | 0.00 | 0.46 | 0.15 |
| Ebright_03 | 35 | 34.24 | 10.41 | 6.94 | 0.7855 | 0.7821 | -0.004 | 0.12 | 0.00 | 0.04 | 0.00 |
| Laugh_00 | 47 | 46.35 | 11.65 | 7.03 | 0.7482 | 0.7699 | **0.029** | 0.35 | 0.00 | 0.35 | 0.07 |
| Laugh_03 | 20 | 19.76 | 9.88 | 7.24 | 0.7418 | 0.7776 | **0.047** | 0.06 | 0.00 | 0.10 | 0.00 |
| Lewis_00 | 92 | 89.18 | 12.18 | 7.13 | 0.7925 | 0.7898 | -0.003 | 0.53 | 0.06 | 0.51 | 0.15 |
| Lewis_01 | 97 | 95.82 | 12.47 | 6.97 | 0.8005 | 0.7858 | -0.019 | 0.24 | 0.00 | 0.51 | 0.24 |
| Lewis_03 | 57 | 55.65 | 13.35 | 7.39 | 0.7661 | 0.8041 | **0.046** | 0.24 | 0.00 | 0.10 | 0.02 |
| Pine_03 | 9 | 9.00 | 6.94 | 6.94 | 0.7974 | 0.7885 | -0.012 | 0.00 | 0.00 | 0.55 | 0.49 |
| IssaEarly_93 | 13 | 12.71 | 7.06 | 6.28 | 0.7524 | 0.7293 | -0.033 | 0.12 | 0.00 | 0.20 | 0.00 |
| LakeSam_03 | 35 | 33.53 | 12.82 | 8.14 | 0.8105 | 0.8269 | 0.02 | 0.12 | 0.06 | 0.04 | 0.00 |
| Webster_05 | 45 | 44.47 | 6.24 | 4.85 | 0.7430 | 0.7007 | -0.061 | 0.24 | 0.00 | 0.26 | 0.01 |
| Meadow_99 | 49 | 48.18 | 15.47 | 8.31 | 0.7930 | 0.8128 | **0.025** | 0.12 | 0.06 | 0.02 | 0.01 |

**Table 4.** $F_{ST}$ values (above diagonal) and genotypic differentiation probabilities (below diagonal) for all pairwise combinations of collections. Bold and shading typeface indicates significance at the P=0.05 and P=0.05 adjusted, respectively. For $F_{ST}$ -values significance is measured against the null hypothesis of random distribution of individuals between collections. For genotypic differentiation significance is measured against the null hypothesis of random distribution of genotypes between collections.

| | Ebright_00 | Ebright_01 | Ebright_03 | Laugh_00 | Laugh_03 | Lewis_00 | Lewis_01 | Lewis_03 | Pine_03 | IssaEarly_93 | LakeSam_03 | Webster_05 | Meadow_99 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ebright_00 | - | **0.0146** | -0.0003 | **0.0201** | **0.0126** | **0.0121** | **0.0238** | **0.0096** | 0.0065 | **0.0819** | **0.0072** | **0.1159** | **0.0748** |
| Ebright_01 | **0.0000** | - | **0.0141** | **0.0289** | **0.0236** | **0.0174** | **0.0161** | **0.0154** | **0.0233** | **0.0678** | **0.0098** | **0.1220** | **0.0712** |
| Ebright_03 | 0.1600 | **0.0000** | - | **0.0197** | **0.0146** | **0.0156** | **0.0282** | **0.0098** | 0.0004 | **0.0895** | **0.0108** | **0.1163** | **0.0699** |
| Laugh_00 | **0.0000** | **0.0000** | **0.0000** | - | -0.0029 | **0.0203** | **0.0297** | **0.0153** | **0.0128** | **0.0877** | **0.0154** | **0.1287** | **0.0689** |
| Laugh_03 | **0.0000** | **0.0000** | **0.0000** | 0.7764 | - | **0.0129** | **0.0256** | **0.0069** | 0.0099 | **0.0835** | **0.0111** | **0.1332** | **0.0645** |
| Lewis_00 | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | - | **0.0140** | **0.0027** | **0.0189** | **0.0723** | **0.0069** | **0.1187** | **0.0620** |
| Lewis_01 | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | - | **0.0149** | **0.0236** | **0.0646** | **0.0101** | **0.1179** | **0.0684** |
| Lewis_03 | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0009** | **0.0112** | **0.0000** | - | 0.0072 | **0.0670** | **0.0044** | **0.1119** | **0.0590** |
| Pine_03 | 0.1826 | **0.0000** | 0.3705 | **0.0084** | 0.0655 | **0.0000** | **0.0000** | 0.2270 | - | **0.0899** | 0.0063 | **0.1152** | **0.0584** |
| IssaEarly_93 | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | - | **0.0595** | **0.1442** | **0.0919** |
| LakeSam_03 | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0004** | 0.0871 | **0.0000** | - | **0.0971** | **0.0479** |
| Webster_05 | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | - | **0.1146** |
| Meadow_99 | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | - |

**Table 5.** Analysis of molecular variance (AMOVA) including only the Ebright, Laughing Jacobs, and Lewis Creek collections. Two models are presented. The first model (left) includes the 2001 collections from Ebright and Lewis creeks, and the second model (right) excludes the 2001 collections.

| | Percent Variation | Percent Variation |
|---|---|---|
| Among Groups | 0.97 | 1.41 |
| P (rand. value > observed) | 0.0097 | 0.0000 |
| Among Populations within Groups | 0.86 | -0.02 |
| P (rand. value > observed) | 0.0000 | 0.5415 |
| Within Populations | 98.18 | 98.61 |
| MODEL | | |
| Groups: | Tributary | Tributary |
| Populations: | Collection Year | Collection Year |
| Years Included: | 00, 01, 03 | 00, 03 |

**Table 6.** For each of the Ebright, Laughing Jacobs, and Lewis collections, the proportion of individuals from that collection with zero to six pairwise relatedness values greater than or equal to 0.40. For example, 0.78 of the individuals from Ebright00 have no relatedness values greater than or equal to 0.40, 0.22 of the individuals have at least one pairwise relatedness value greater than or equal to 0.40, 0.11 have at least two pairwise relatedness values greater than or equal to 0.40, and upwards to 0.01 have at least four pairwise relatedness values greater than or equal to 0.40. The right side of the table shows the actual variance of pairwise relatedness values for each collection, and statistics for the variance of the simulated collections, as described in the Material and Methods section.

| | Number of Pairwise Relatedness > = 0.40 | | | | | | | Variance | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Greater than or equal to | | | | | | Empirical | Simulated | | |
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | | Median | 5% | 95% |
| Ebright00 | 0.78 | 0.22 | 0.11 | 0.08 | 0.01 | 0.00 | 0.00 | 0.017 | 0.014 | 0.013 | 0.015 |
| Ebright01 | 0.57 | 0.43 | 0.33 | 0.14 | 0.04 | 0.00 | 0.00 | 0.017 | 0.012 | 0.011 | 0.012 |
| Ebright03 | 0.71 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.016 | 0.013 | 0.012 | 0.014 |
| Laugh00 | 0.49 | 0.51 | 0.30 | 0.13 | 0.04 | 0.00 | 0.00 | 0.024 | 0.014 | 0.014 | 0.015 |
| Laugh03 | 0.80 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.024 | 0.012 | 0.01 | 0.014 |
| Lewis00 | 0.41 | 0.59 | 0.33 | 0.16 | 0.09 | 0.05 | 0.00 | 0.020 | 0.013 | 0.013 | 0.014 |
| Lewis01 | 0.44 | 0.56 | 0.29 | 0.13 | 0.07 | 0.07 | 0.01 | 0.020 | 0.013 | 0.013 | 0.014 |
| Lewis03 | 0.70 | 0.30 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.015 | 0.013 | 0.012 | 0.014 |

**Table 7.** Cytochrome b molecular diversity and haplotype frequency. The last line of the table indicates what base number along the mtDNA sequence with the mutation that defines each haplotype. Relative to the other five haplotypes, Haplotype B had no mutations.

| Population | Total # of Individuals | Haplotype Diversity | Nucleotide Diversity | Haplotype Frequency | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | A | B | C | D | E | F |
| Ebright_00 | 79 | 0.30 | 0.00072 | 0.00 | 0.84 | 0.04 | 0.13 | 0.00 | 0.00 |
| Ebright_01 | 64 | 0.74 | 0.00179 | 0.00 | 0.39 | 0.48 | 0.08 | 0.00 | 0.05 |
| Ebright_03 | 36 | 0.41 | 0.00099 | 0.00 | 0.75 | 0.03 | 0.22 | 0.00 | 0.00 |
| Laugh_00 | 44 | 0.38 | 0.00091 | 0.00 | 0.77 | 0.02 | 0.20 | 0.00 | 0.00 |
| Laugh_03 | 20 | 0.58 | 0.00140 | 0.00 | 0.65 | 0.10 | 0.25 | 0.00 | 0.00 |
| Lewis_00 | 93 | 0.06 | 0.00015 | 0.00 | 0.97 | 0.02 | 0.01 | 0.00 | 0.00 |
| Lewis_01 | 97 | 0.57 | 0.00138 | 0.00 | 0.63 | 0.29 | 0.05 | 0.00 | 0.03 |
| Lewis_03 | 55 | 0.25 | 0.00059 | 0.05 | 0.87 | 0.00 | 0.05 | 0.02 | 0.00 |
| Pine_03 | 8 | 0.50 | 0.00120 | 0.13 | 0.75 | 0.00 | 0.13 | 0.00 | 0.00 |
| IssaEarly_93 | 13 | 0.28 | 0.00068 | 0.00 | 0.85 | 0.00 | 0.15 | 0.00 | 0.00 |
| Webster_05 | 46 | 0.23 | 0.00056 | 0.87 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 |
| Meadow_99 | 48 | 0.00 | 0.00000 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | | | 15,649 | | 15,844 | 15,778 | 15,523 | 15,504 |

**Table 8.** $\Phi_{ST}$ values (above diagonal) and genotypic differentiation probabilities (below diagonal) for all pairwise combinations of Lake Sammamish tributary populations. Bold and shading typeface indicates significance at the P=0.05 and P=0.05 adjusted, respectively.

| | Ebright00 | Ebright01 | Ebright03 | Laugh00 | Laugh03 | Lewis00 | Lewis01 | Lewis03 |
|---|---|---|---|---|---|---|---|---|
| Ebright00 | - | **0.2834** | 0.0075 | 0.0016 | 0.0221 | 0.0638 | **0.1249** | 0.0201 |
| Ebright01 | **0.0000** | - | **0.2583** | **0.2695** | **0.1809** | **0.3886** | **0.0460** | **0.3107** |
| Ebright03 | 0.4152 | **0.0000** | - | -0.0250 | -0.0269 | **0.2070** | **0.1423** | **0.0745** |
| Laugh00 | 0.5145 | **0.0000** | 1.0000 | - | -0.0185 | **0.1749** | **0.1443** | **0.0618** |
| Laugh03 | 0.1441 | **0.0044** | 0.4978 | 0.3078 | - | **0.2773** | **0.0923** | **0.1071** |
| Lewis00 | **0.0029** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | - | **0.1786** | **0.0280** |
| Lewis01 | **0.0000** | **0.0329** | **0.0001** | **0.0000** | **0.0226** | **0.0000** | - | **0.1509** |
| Lewis03 | **0.0474** | **0.0000** | **0.0234** | **0.0209** | **0.0062** | **0.0115** | **0.0000** | - |

**Table 9.** Analysis of molecular variance (AMOVA) including only the Ebright, Laughing Jacobs, and Lewis Creek populations. Four models are presented. The first and second models include all collection years from all tributaries. The third model includes only collection years 2000 and 2003, the two years from which all three tributaries have collections. The fourth model includes collection years 2000 and 2003 for Ebright and Lewis only

| | Percent Variation | Percent Variation | Percent Variation | Percent Variation |
|---|---|---|---|---|
| Among Groups | -2.77 | 14.84 | 6.00 | 5.94 |
| P (rand. value > observed) | 0.6002 | 0.0098 | 0.1300 | 0.0000 |
| Among Populations within Groups | 18.05 | 4.08 | 0.69 | 1.47 |
| P (rand. value > observed) | 0.0000 | 0.0000 | 0.3079 | 0.1017 |
| Within Populations | 84.72 | 81.08 | 93.31 | 92.60 |
| Model | | | | |
| Groups: | Tributary | Collection Year | Tributary | Ebright v. Lewis |
| Populations | Collection Year | Tributary | Collection Year | Collection Year |
| Years Included | 00, 01, 03, 04 | 00, 01, 03, 04 | 00, 03 | 00, 03 |

**Table 10.** Analysis of molecular variance (AMOVA) testing for interaction between collection year and maternal (haplotype) lineage in terms of microsatellite allele frequency differences

| | Percent Variation | Percent Variation | Percent Variation |
|---|---|---|---|
| Among Groups | 0.24 | -1.20 | -0.81 |
| P (rand. value > observed) | 0.3724 | 0.7087 | 0.6011 |
| Among Populations within Groups | 1.08 | 1.75 | 2.44 |
| P (rand. value > observed) | 0.0010 | 0.0215 | 0.0000 |
| Within Populations | 98.68 | 99.44 | 98.37 |
| Model | | | |
| Groups: | Collection Year | Collection Year | Collection Year |
| Populations | Haplotype | Haplotype | Haplotype |
| Tributary | Ebright | Laughing Jacob | Lewis |