Population genetics of *Oncorhynchus mykiss* in the Santa Clara Valley Region

John Carlos Garza & Devon Pearse

Final Report to the Santa Clara Valley Water District (SCVWD).

March 2008

Abstract

Steelhead/rainbow trout of the species *Oncorhynchus mykiss* are found in all of the major drainages of the Santa Clara Valley, which includes streams that drain into both San Francisco and Monterey Bays. Most of the basins in this area have dams or other impoundments and many of the resulting reservoirs have been stocked with hatchery rainbow trout. Genotype data was collected from 18 highly variable microsatellite molecular markers in more than one thousand trout from the Santa Clara Valley region sampled by Santa Clara Valley Water District biologists and a sample of adult steelhead from the San Lorenzo River in Santa Cruz County. The analyses examined population structure within the region, relationships between populations above and below barriers to anadromy and population genetic diversity. Analysis focused on 21 "population" samples, comprised of fish sampled in a specific location or year, that were used to represent fish collected from the Coyote, Guadalupe, Pajaro, Permanente/Stevens Creek, San Francisquito, San Lorenzo, and San Tomas Aquino basins. Additional analyses were conducted with data from the same microsatellite markers in rainbow trout hatchery stocks and steelhead from coastal and California Central Valley populations. These analyses looked at whether specific fish may have been produced by or descended from hatchery strains used in local stocking efforts, as well as providing biogeographic context for the Santa Clara Valley regional results.

In general, substantial structure was found, with populations within a basin most closely related to other populations from the same basin, regardless of whether they were sampled above or below a known barrier to anadromy. This is due to some combination of pre-impoundment historic shared ancestry, downstream migration and limited (possibly anthropogenic) upstream migration. However, lower genetic diversity in above-barrier populations indicates a lack of substantial genetic input upstream and highlights lower effective population sizes for above-barrier populations.

Several analyses found a clear signal of coastal steelhead ancestry in all population samples. Individual assignment tests indicated that less than 1% of all fish sampled were of recent hatchery strain origin. Examination of phylogeographic trees indicated that the Santa Clara Valley trout populations are generally most closely related to coastal steelhead populations from the two steelhead Distinct Population Segments (DPSs) that

include other San Francisco and Monterey Bay trout populations. In addition, these trees showed clear separation between all Santa Clara Valley populations, Central Valley steelhead and hatchery trout strains. The population samples from Coyote Creek in 1999 and 2000 were both of sufficient size to analyze temporal changes in genetic composition. The samples from the two years were not significantly differentiated, indicating temporal stability in genetic composition, although this should be evaluated in the future with temporal samples that span more than one generation.

As part of this contract, attempts were made to recover data for the nuclear microsatellite loci studied here from museum specimens collected in 1897 and 1909 from the Pajaro River and Coyote Creek basins, but these efforts were unsuccessful. However, subsequent analysis as part of a state funded project to analyze specimens from throughout central California was successful in extracting mitochondrial DNA data from some of these samples. Those results will be available shortly in a separate report.

Introduction

The Santa Clara Valley is the site of the third largest municipality in California, San Jose, and is at the southern end of the second largest urban area on the west coast of North America, the San Francisco Bay Area. It also sits at the edge of the largest estuarine system on the west coast of the coterminous United States, San Francisco Bay. Santa Clara Valley and County contain streams that drain mainly into southern San Francisco Bay, but also into Monterey Bay through the Pajaro River. In spite of the highly urbanized nature of the northern Santa Clara Valley and the lower reaches of some of the large streams in the basin, most of the streams contain assemblages of native fish.

Steelhead rainbow trout (*Oncorhynchus mykiss*) populations in the central California region and in the Santa Clara Valley are divided into two Distinct Population Segments (DPSs), formerly Evolutionarily Significant Units (ESUs). In the South Central California Coast (SCCC) DPS, which extends south from the Pajaro River in Monterey Bay to just north of the Santa Maria River in San Luis Obispo County, steelhead were listed as Threatened under the US Endangered Species Act (ESA) by the National Oceanic and Atmospheric Administration in 1997. Fish from the Russian River in Sonoma and Mendocino Counties to just north of the Pajaro River (Soquel Creek), including San Francisco Bay, are included in the Central California Coast Steelhead (CCC) DPS and were also ESA-listed as Threatened in 1997. A subsequent genetic analysis by Garza et al. (in review) indicated that the genetic division between San Francisco and Monterey Bay steelhead populations is actually just south of the Golden Gate. A primary limiting factor for steelhead populations in the central California region is access to freshwater habitat due to dams and water diversions, which are common in

the region. Most of these barriers lack fish passage structures that permit upstream migration. When fish from the species *O. mykiss* are currently found above such barriers they are considered to be resident rainbow trout, regardless of ancestry, and are not afforded protection under the state or federal ESAs.

In this project, trout from the species *Oncorhynchus mykiss*, commonly known as steelhead or rainbow trout, were studied in basins of the Santa Clara Valley using molecular genetic techniques and population genetic analysis of microsatellite DNA. To provide insight into questions of population structure in this geographic area, data were collected from 18 highly variable microsatellite genetic markers and variation analyzed to trace ancestry and evaluate genetic distinction among populations. Microsatellites, also known as simple tandem repeat loci, have been used in numerous studies of salmonids and have proven to be a valuable tool for elucidating population genetic structure (Carlsson & Nilsson, 2001; Castric et al. 2001; Spidle et al. 2001; Olsen et al. 2003; Poissant et al. 2005; Crispo et al. 2006; Garza et al. in review).

Previous genetic work on population structure of steelhead in California has relied primarily on mitochondrial DNA (e.g. Berg and Gall 1988; Nielsen et al. 1997), which is a single gene that is often not reflective of population history or true relationships (Chan and Levin 2005), or small numbers of microsatellite loci and inadequate population sampling, which can also lead to inaccurate inference regarding population structure, particularly on a relatively small geographic scale. However, recent work on *O. mykiss* in coastal California using a large number of microsatellite loci has demonstrated that genetic structure can be easily identified with such data both at larger scales (Aguilar and Garza, 2006; Garza et al. in review; Clemento et al. in prep) and at relatively fine ones

(Deiner et al. 2007; Pearse et al. 2007a). For example, *O. mykiss* populations in the Russian River separated by waterfalls were highly genetically distinct, whereas those found above and below the two major dams (Warm Springs and Coyote) were found to show little genetic distinction (Deiner et al. 2007). In the Klamath River, genetic relationships of trout populations above barriers with those below barriers do not vary with geographic distance, whereas genetic relationships between populations below barriers do (Pearse et al. 2007a), a pattern referred to as isolation by distance.

Genotypes were collected from 1073 individual fish at the same 18 polymorphic microsatellite loci employed in these other studies. Genetic variation of these fish was analyzed and the data were also combined with data from other populations of California trout to better understand the genetic results in a regional context. The goals of the study were to use population genetic analyses of the data to assess origins and ancestry of trout populations from Santa Clara Valley streams, better understand the relationship of these trout populations to others in California, and to provide information on genetic diversity and population structure within the Santa Clara Valley. Fish populations from rivers and creeks that flow to both San Francisco and Monterey Bays were evaluated. The basins for which trout populations were studied include Coyote Creek, the Guadalupe and Pajaro Rivers, and the smaller basins of San Tomas Aquino Creek, San Francisquito Creek, and Permanente/Stevens Creeks. Fish collected both above and below barriers to anadromy in most of the study basins were included in the analyses. Since offspring of anadromous steelhead from the San Lorenzo River were stocked in the Pajaro River basin for many years, up until 1997 (D. Strieg, Monterey Bay Salmon & Trout Project, pers. comm.), and its headwaters intercalate with those of many Santa Clara Valley basins (e.g. Guadalupe,

Permanente/Stevens), a population sample of anadromous adult steelhead sampled at a weir on the San Lorenzo River in Santa Cruz County was also genotyped.

There are a number of impoundments in the study basins and hatchery-raised trout of a variety of strain origins have been planted in nearly all of the reservoirs above them over the last 100+ years. Many of these trout were likely of diverse geographic and phylogenetic origin, as movements of salmonids from basin to basin and from state to state was common until recently. In addition, many of the strains used in stocking by California State and cooperative hatcheries, both currently and in the past, were developed from trout populations in both distant (e.g. Kamloops, Canada) and unknown locations. Some hatchery strains probably have contributions from populations of both anadromous steelhead in coastal and interior streams and isolated inland basins, although they are now phenotypically resident rainbow trout. Many of the basins in the Central Coast and San Francisco Bay Area region have been stocked recently with such rainbow trout from Fillmore Hatchery on the Santa Clara River in Ventura County. For example, reservoirs in the Pajaro, Coyote and Guadalupe basins were recently stocked with fish from Fillmore Hatchery. Microsatellite data from the five strains of trout currently raised at Fillmore Hatchery – Coleman, Virginia, Wyoming, Mt. Whitney Early, Mt. Whitney Late – were we also included in some of the genetic analyses to detect reproduction from hatchery fish and to determine if any of the sampled populations had a large degree of ancestry from these stocked fish. For some analyses, data from population samples of Central Valley anadromous trout in both the northern (Battle Creek) and southern (Stanislaus and Tuolomne Rivers) portion of the Sacramento/San Joaquin basin and a sample from a divergent hatchery rainbow trout strain (Junction Kamloops) from Hot

Creek Hatchery in eastern California were also included.

In addition to the hatchery trout and Central Valley steelhead strains, data from the same genes have been collected in almost 100 other populations of steelhead from California (Aguilar and Garza 2006; Pearse et al. 2007a; Garza et al. in review; Clemento et al. in prep), covering the entire range of steelhead in the state. Data from many of these populations, and for 14 of the 18 microsatellite genes, were combined with those from the Santa Clara County trout populations, to put local relationships in a geographic context and to identify relationships of Santa Clara County trout populations to those from other parts of California. This combined dataset was used to construct phylogeographic trees that depict summarized genetic relationships.

Methods

Sampling

Most of the fish analyzed in this study were sampled by Santa Clara Valley Water District biologists Samples were collected from 15 different streams and 30 separate sampling locations within Santa Clara County (Figure 1). Each sample consisted of small pieces of dried tissue (1-2 mm) collected from the upper portion of each fish's caudal fin and preserved through desiccation on blotter paper. A population sample of anadromous fish from the San Lorenzo River (Santa Cruz County; N=69) was also included in the study, as fish from this basin have been used extensively for stocking in the Pajaro River drainage and because these adult trout were collected at a weir and are known to be steelhead. Upon receipt, all samples were catalogued and transferred to tubes in 96 well microplates for DNA extraction.

DNA Extraction

Total nucleic acids were extracted from the tissue samples using Qiagen DNeasy Tissue Kits, following the manufacturer's recommended protocol for animal tissues and using a BioRobot 3000 (Qiagen, Inc.) for all liquid handling. Approximately $2mm^2$ of tissue was digested in 180µL of Qiagen buffer ATL and 20µL proteinase K and kept overnight in a shaking incubator at 55°C. The DNA was then bound to the DNeasy silica-gel membrane with the addition of 200µL Qiagen buffer AL and 200µL of ethanol, washed with 500µL each of Qiagen buffer AW1 and AW2, and finally eluted in 200µL buffer AE (Qiagen, 2000). Extracted DNA was kept frozen at 20°C until it was diluted (10:1 with autoclaved, distilled water) and distributed to 96 well plates for microsatellite

amplification via polymerase chain reaction (PCR).

Eleven of the original 1,073 fish were removed from the data set due to poor PCR amplification, in which more than half the loci failed. One of these individuals (M020937, COYO-545, Coyote Creek, Montague FUM3) also appeared to be an outlier relative to all other individuals examined based on Factorial Correspondence Analysis and may have been another salmonid species, such as Chinook salmon. With the removal of these 11 individuals, the total remaining number of individuals was 1,062. It is worth noting that this is an extremely low failure rate for such a study and reflects very high tissue quality. A further 16 individuals were not used in the population analyses because they came from sites with very few samples (e.g. Almaden Reservoir, Coyote Reservoir, Alameda Creek). The final number of fish included in the population genetic analyses was therefore 1,046 (Table 1).

The total data set was then divided into "population" samples for analysis. The primary division was between basins, and then by tributary or locality. All distinct localities were classified as populations and samples taken from above barriers were always separated from those taken from below-barrier populations (i.e. within the Guadalupe River, the Pajaro River, and Stevens Creek). Fish sampled in Coyote Creek downstream migrant trapping were further subdivided by sampling year. Throughout this report, fish from each of these groups are referred to as populations for convenience and without any assumptions about the biological details underlying this designation.

Genotyping

Genotypic data at 18 microsatellite loci was collected for fish in all population samples (Table 2). PCR was carried out in 15µL aliquots containing 4µL purified and diluted template DNA, 6.35µL H_2O, 1.5µL ABI 10X II PCR buffer, 0.9µL $MgCl_2$, 1.2µL dNTPs, 0.05µL DNA polymerase (Amplitaq, Applied Biosystems), and 1µL fluorescent-labeled oligonucleotide primers (Integrated DNA Technologies, Inc.). Variable cycling regimes were carried out on MJ Research (PTC 225) thermal cyclers to maximize PCR product. The typical profile consisted of a two minute pre-denaturation at 95 ˚C, then two amplification stages: (a) 10 cycles of denaturation at 95 ˚C for 15s, annealing at 53 ˚C for 15s, and extension at 72 ˚C for 45s; (b) 25 cycles at 89 ˚C for 15s, 55 ˚C for 15s, and 72 ˚C for 45s. The routine concluded with a final extension phase of 72 ˚C for 5 minutes and indefinite hold at 10 ˚C. PCR products were pooled to equalize peak heights and take advantage of multiple label colors and two non-overlapping ends of the measurable size range (50bp-500bp) within each lane. A mix of Formamide, loading dye and internal size standard was added to the pooled PCR product, denatured at 95 ˚C for 3 minutes and immediately transferred to ice. The samples were then electrophoresed on an ABI Prism 377 DNA sequencer. Gel imaging, lane tracking and allele size for loci run with the ABI 377 were scored with GENESCAN version 3.1.2 and GENOTYPER version 2.1 software (Applied Biosystems). At least two people performed all size scoring independently, discrepancies were identified and, if a resolution was not reached, the sample was rerun. If a discrepancy persisted through the second analysis, the fish was not scored at that locus. A representative fraction was re-genotyped as a control for data quality.

Data Analysis

Expected heterozygosity (Nei 1987), observed heterozygosity and number of alleles were calculated for each sample population. In order to compensate for variation in sample sizes, genetic diversity was also assessed using allelic richness as estimated with the rarefaction method in FSTAT version 2.9.3.2 (Goudet 2001). Deviations from Hardy-Weinberg equilibrium (HWE) were examined utilizing the Markov Chain Monte Carlo (MCMC) approximation of an exact test implemented in the GENEPOP program version 3.4 (Raymond and Rousset 1995). The alternative hypotheses of heterozygote deficiency and heterozygote excess were both tested with Markov chain parameters of 10,000 dememorization, 1000 batches and 1000 iterations per batch. Linkage (gametic phase) disequilibrium (LD) was also evaluated to examine segregation independence of the 18 microsatellite loci in each of the sample populations and using the same type of MCMC approximation of an exact test as implemented in GENEPOP. MCMC parameters were the same as those used for the heterozygosity exact tests. Disequilibrium results were summarized as the percentage of loci in a population out of equilibrium (HWE) or the percentage of locus pairs in a population that were in disequilibrium (LD).

Genetic differentiation between sample populations was examined with several methods. Using the test for genic differentiation in GENEPOP, a Fisher's exact test was employed to calculate the probability of the null hypothesis (H_O) that allele frequencies were identical across populations. Pairwise differentiation between all pairs of populations was also quantified using F_{ST}, as estimated by Weir and Cockerham's (1984) Θ estimator, and significance (> 0) assessed by the permutation algorithm in the Genetix software package (Belkhir et al. 2004) with 10,000 replicates.

Individual-based assignment tests were used to further evaluate the degree of recent gene flow between the sample populations of Santa Clara Valley trout, as well as the hatchery rainbow trout strains. This analysis assigns each individual fish to its most likely population of origin, using its genotype alone and through comparison to a collection of potential source populations. The semi-Bayesian allele frequency estimation algorithm of Rannala and Mountain (1997) and the leave-one-out procedure as implemented in GeneClass version 2.0.g (Piry et al. 2004) were utilized. Although application of assignment tests can be used to detect first generation migrants (Rannala and Mountain 1997), misassignment, or assignment of an individual to a population other than that of its sampled location, should not be interpreted as migration with juvenile and/or resident fish, but as a signal of recent common ancestry. Patterns of misassigned fish highlight similarities in genetic composition (allele frequencies) between sample populations/locations. Misassignments may also occur randomly if an individual has a genotype composed of alleles that are common to many groups, since there is no statistical power for assignment in such situations.

Phylogeographic trees were constructed using matrices of Cavalli-Sforza & Edwards' (1967) chord distance (CSE), using the software package PHYLIP version 3.57c (Felsenstein 1993). This genetic distance was chosen because of its statistical properties (Felsenstein 2003) and because it most reliably recovers the correct topology (branching pattern) for phylogeographic trees (Takezaki and Nei 1996). The neighbor-joining algorithm was used to determine tree topology and a consensus tree was assembled from 1,000 bootstraps of the distance matrix with the CONSENSE program of the PHYLIP software package. Internal branch lengths on the consensus tree are scaled by the number

of times that relationship was found in the neighbor-joining trees constructed with the bootstrap samples, and is a measure of confidence in that branch. Only bootstrap values above 50% are generally reported on such trees.

These phylogeographic tree-building analyses were carried out with several different datasets. First, all of the populations genotyped for this study (Santa Clara Valley & San Lorenzo River, Santa Cruz County) were analyzed and both the most probable tree and the bootstrap consensus tree reported. Several additional analyses of this dataset combined with data from trout that were analyzed in other studies were also carried out. These subsequent analyses utilized only the 14 microsatellite loci where the data could be confidently combined. The 4 additional loci could not be combined due to differences in the original data collection methods for at least one of the populations. The first such analysis combined the Santa Clara Valley data with that from the Fillmore Hatchery trout strains, a divergent rainbow trout strain derived from a Canadian (Kamloops) population, and several Central Valley steelhead populations, including one comprised of anadromous adults (Battle Creek). Both the most probable tree and the bootstrap consensus tree are reported. The second such analysis combined the Santa Clara Valley population samples described here with the 60 population samples from coastal steelhead populations from the Oregon border to Morro Bay (San Luis Obispo County) analyzed by Garza et al. (in review), and 20 population samples from the Monterey Bay region south to Los Angeles County analyzed by Clemento et al. (in prep), which includes pairs of populations from above and below major dams. The Fillmore Hatchery strains were also included in this analysis. The Fillmore Hatchery strains analyzed (Coleman, Virginia, Wyoming, Mt. Whitney early and late) represent the major strains used in stocking in the

southern and central portion of coastal California in the recent past. Only the most probable tree is reported for this analysis.

Factorial correspondence analysis (FCA), which is a canonical algorithm similar to principal components analysis, and as implemented in the Genetix software program (Belkhir et al. 2004), was also used to qualitatively explore the distribution of genotypes in the data. FCA uses frequencies of different alleles as the components or axes in a three dimensional visual representation of individual genotypes. This analysis helps to identify outlying individual fish and to visualize overlap in the distribution of individual genotypes from different populations. The FCA method was conducted on the full dataset from the Santa Clara Valley populations only, as well as the combined dataset with the Fillmore Hatchery strains and California Central Valley steelhead populations used for the phylogeographic trees.

Results and Discussion

Population structure

Phylogeographic trees were used to visually and quantitatively evaluate genetic relationships of Santa Clara Valley trout populations both with each other and with other California trout populations. This analysis first created matrices of genetic distances, using Cavalli-Sforza and Edwards (CSE; 1974) chord metric, and then employed the most commonly used (and accurate) tree building method, neighbor-joining (Saitou and Nei 1987), to construct unrooted branching networks (trees) of trout populations. The neighbor-joining tree describing the relationships of the Santa Clara Valley trout populations is found in Figure 2a. This tree is the most probable tree constructed with the data and provides information about both the branching relationships (topology) and the divergence of populations (terminal branch lengths). Bootstrap analysis was then used to evaluate the support across loci for individual internal branches and the majority rule bootstrap consensus tree is reported in Figure 2b. This consensus tree is constructed by bootstrap resampling with replacement from the original dataset to create multiple replicate datasets, in which some loci may be represented more than once and others not at all, and then calculating the CSE genetic distance matrix and building a neighbor joining tree for each one. The consensus tree is then constructed, with the length of the internal branches proportional to the number of trees in which the branch was observed. For example, when a bootstrap proportion and length of an internal branch that groups three populations is 85, that means that the grouping was found in 85% of the trees constructed with the replicate datasets of bootstrap resamples.

The phylogeographic tree analysis revealed a general clustering of populations by basin of origin. Most of the populations from the Guadalupe, Coyote and Pajaro drainages formed distinct, exclusive clusters in both the neighbor-joining and bootstrap consensus trees. The only exceptions to this are the intermingling of the Guadalupe-Los Gatos Creek and Guadalupe-Main populations with the Stevens and San Francisquito Creek populations and the intermingling of the San Lorenzo River population within the Pajaro River populations. There is no obvious explanation for the first pattern (nor is one really necessary, given the lack of strength of the association), other than greater migration between the two basins than between the lower and upper Guadalupe. The close relationship of the San Lorenzo River and Pajaro River trout populations is likely at least partly due to extensive stocking of juvenile trout in the Pajaro basin by the Monterey Bay Salmon and Trout Project in the 1990s using fish raised from San Lorenzo River steelhead broodstock. This stocking ceased in 1997 with the prohibition on interbasin transfers of anadromous fish by the California Department of Fish and Game in response to Endangered Species Act listings. A similar close genetic relationship between the San Lorenzo and Pajaro River trout populations has also been reported previously (Garza et al. in review; Sundermeyer 1999).

The FCA results for these populations were similar (Figure 3). This analysis produces a visual representation of individual genotypes arrayed by principal components of the allele frequency distributions of population samples. A close relationship of all Santa Clara Valley populations was found, with moderate differentiation in allele frequencies mainly associated with different basins, primarily Guadalupe, Coyote, San Tomas

Aquino and the Pajaro. They all overlap in the central cluster where the Permanente/Stevens and San Francisquito genotypes are also found.

The construction of phylogeographic trees that also include several California Central Valley steelhead populations and many of the rainbow trout strains commonly used in central California stocking activities found clustering of all of the Santa Clara Valley populations to the exclusion of Central Valley steelhead and hatchery trout strains (Figure 4a). These groupings were supported by long internal branch lengths (Figure 4a) and high bootstrap values (Figure 4b). The Coyote 1998 sample was intermediate between the Central Valley/hatchery and Santa Clara Valley trout populations, which is consistent with the high proportion of hatchery fish found in this sample (see assignment results in Table 4). These data clearly demonstrate very limited recent gene flow or migration between Santa Clara Valley and Central Valley trout populations.

The FCA results for the expanded dataset (Figure 5) are similar to those of the phylogeographic tree analysis (Figure 4 a & b). The Santa Clara Valley/coastal steelhead lineage is largely differentiated on its own axis, with most similarity to the Central Valley steelhead lineage, but little overlap (z axis differentiation not readily visible in Figure 5). There is no overlap between the genotypes of the Santa Clara Valley/coastal steelhead lineage and the hatchery trout strains, indicating a general lack of large scale introgression of hatchery strains into coastal trout populations. In addition, the Central Valley and hatchery trout strains are most similar, but the diversity in genotypes of the hatchery trout strains is much greater, reflecting the multiple phylogeographic origins of these hatchery trout strains.

There was broad concordance between the results of this and previous genetic studies in the relationships of Santa Clara Valley steelhead populations with other coastal California steelhead populations. In general, populations from the drainages that empty to San Francisco and Monterey Bays cluster with others from the same geographic/genetic group (Figure 6), indicating that these populations are part of the coastal steelhead group, with populations generally most closely related to those from other basins in close proximity. The one exception, San Tomas Aquino-Saratoga Creek, clustered with Monterey Bay (Pajaro River) populations and not with San Francisco Bay populations. There are several possible explanations for this pattern, including the possibility that Saratoga Creek or some nearby location was stocked with fish from a Monterey Bay region population. In fact the closest relationship of Saratoga Creek aside from the Pajaro-Uvas populations is the San Lorenzo River, which is also closely related to the Pajaro River populations. However, it is also possible that this association is simply due to the limited power of the dataset to simultaneously estimate so many pairwise genetic distances with perfect success. The analysis of a larger genetic dataset with these samples and/or the collection of data from fish sampled in Saratoga Creek in a different year could be used to try to resolve this question. However, it should be emphasized that the Saratoga Creek population is still very similar to other Santa Clara Valley/San Francisco Bay populations, but just slightly more similar to Monterey Bay populations. One other result that is worth noting is that the Pajaro River populations group in slightly different parts of the tree. The Uvas and Llagas Creek above-barrier populations cluster with the northern Monterey Bay group and the Uvas below-barrier population clusters with the southern Monterey Bay group. The significance of these associations is difficult to

determine, since recent gene flow and a lack of power in the combined dataset limit the ability to accurately estimate all such relationships with perfect accuracy.

Another area of concordance between the current and previous data sets is that when they are combined, the two population samples from the San Francisquito drainage, one from the current study and one from Garza et al. (in review), cluster together closely on the phylogeographic tree. These samples were taken in the same year, but still provide an internal control for combination of the datasets, indicating no discrepancies between the two datasets that might lead to erroneous inference when they are combined.

There were significant deviations from both Hardy Weinberg (within a locus) and linkage (between loci) disequilibria in most of the population samples, although generally not above the amount expected by chance alone, when not corrected for multiple tests. This is very common with salmonids when premigratory juveniles and/or resident fish from relatively small populations are sampled and is primarily due to family structure (siblings) in the data (Allendorf and Phelps 1981; Castric et al. 2002; Deiner et al. 2007). However, the Coyote Creek 2000 sample had such extensive disequilibrium that it may have other biological significance. A common source for such disequilibrium is admixture, or the sampling of two or more populations that are believed to be one. In the case of the Coyote Creek 2000 sample, it is possible that admixture of more than one genetically differentiated population from Coyote Creek is included in the sample. The fish in this population sample were from migratory smolts collected in a downstream migrant trap in the lower reach of Coyote Creek, below the confluence of most major tributaries. Upper Penitencia Creek, for example, may have contributed fish to this collection. This population is moderately differentiated from populations in other

tributaries that were sampled in downstream migrant trapping, and about 6% of the fish sampled in the Coyote 2000 collection assign to Upper Penitencia. More extensive sampling of trout populations in the Coyote Creek basin could help to resolve this and would also help to further elucidate population structure in this relatively large basin. The upper reaches of Coyote Creek may support other trout populations, and they may be present in these downstream migrant samples, but the upstream sites are not represented by direct collections in this study.

Matrices of pairwise values of F_{ST}, the standardized variance in allele frequencies, between populations were examined for patterns of population structure (Table 3). F_{ST} is a measure of how much of the total genetic variation is found between the populations. Since small populations lose genetic variation more quickly than larger ones, values of F_{ST} are dependent upon population size, with larger values in smaller populations resulting from recent loss of variants shared with other closely related populations. Since population size also determines how much genetic variation can be maintained in the population, measures of genetic diversity are generally correlated with F_{ST} in *O. mykiss* populations (Pearse et al. 2007a; Garza et al. in review). In the current data set, pairwise F_{ST} values were highly correlated with the number of alleles found in each population ($p<<0.01$; $r^2 = 0.89$). This means that absolute values of F_{ST} are not directly comparable for different populations without taking into account the levels of genetic diversity. However, the relative values of F_{ST} for a population with different populations still provide insight, as do distributions of F_{ST} values. These relative values of F_{ST} provided a similar signal to the assignment tests and genetic distance-based trees, with the mean intrabasin value (0.0715±0.0079) significantly smaller than that from comparisons of

population samples in different basins (0.1051±0.0067). In addition, the overall patterns of F_{ST} are consistent with other analyses of differentiation, including the lack of significant differentiation between temporal samples taken from Coyote Creek.

Individual assignment test analysis found high accuracy of assignment for Santa Clara Valley trout populations (Table 4). The overall accuracy of assignment to population sample of origin was 80.9%. However, more than half of the misassignments were to other population samples from the same basin (e.g. Uvas Creek to Bodfish Creek) and approximately one quarter of misassignments were between the Coyote 1999 and Coyote 2000 samples, which were not significantly differentiated. When misassignment to another population in the same basin is not considered an error, the assignment accuracy is 91.4%. When the full 18 locus dataset is used with Santa Clara Valley populations only, the assignment accuracy to population sample increases to 82.8%, but identification of hatchery fish is not possible. If probability calculations were used to apply an exclusion criterion, it would be possible to further increase accuracy by failing to assign fish with ambiguous genotypes or recent hybrid ancestry (Pearse et al. 2007b), thereby ensuring very high accuracy of assignment of individual fish to basin of origin, as well as identifying first generation migrants, when adult fish are studied.

This high accuracy of assignment indicates a substantial amount of population structure within the study area, which is typical of trout populations in the coastal California Distinct Population Segments (DPSs). In general, population structure in California trout populations is dependent upon geographic distance, with individual misassignments between locations primarily occurring between geographically proximate basins and between tributaries with basins (Pearse et al. 2007a; Garza et al. in review).

While some of this structure is likely associated with local adaptation, much of it is due simply to limited migration and family structure from the presence of siblings and higher order relatives in local populations, as well as reductions in population size.

The data from these 18 microsatellite loci (Table 2) and the high accuracy of individual assignment test analysis on even a small scale indicates that these molecular markers can be useful as a reference baseline for genetic stock identification techniques to determine basin and tributary of origin for individual trout in management or forensic applications. Care would be required to update reference databases frequently, to account for temporal shifts in allele frequencies due to changing population and family structure that would decrease assignment power.

Genetic diversity

Genetic diversity can be measured in a number of ways, but most of them are closely related to either the mean number of alleles per locus observed in the sample, or the heterozygosity of the sample. Heterozygosity is the proportion of individuals that have two distinct alleles on their two chromosomes. Heterozygosity for microsatellite loci in salmonids is typically in the range 0.5-0.7. The number of alleles is highly dependent upon the particular microsatellite loci evaluated, the number of loci genotyped and the population sample size, whereas heterozygosity is not. In addition, the mean number of alleles is generally a fairly sensitive measure of genetic diversity relative to differences or changes in population size, whereas heterozygosity is not (Amos and Harwood 1998).

Genetic diversity was generally moderate in Santa Clara Valley trout populations (Table 1). Most population samples had levels of genetic variation well within the range

observed in several studies of coastal steelhead populations with the same genetic markers (Aguilar & Garza 2006; Pearse et al. 2007a; Garza et al. in review; Clemento et al. in prep). However, comparison of both the allelic richness and heterozygosity genetic diversity measures indicates that the Permanente Creek and San Tomas Aquino-Saratoga Creek populations are substantially less genetically diverse than the other populations examined here. This is a result of smaller effective population size, indicating that the fish sampled in these two creeks are descended from a much smaller number of ancestors than are the population samples from other basins included in the study. Genetic diversity in Permanente Creek, in particular, was so low that the sample probably represents only a very small number of families and likely even a single family, since there were not more than 4 alleles (the maximum possible for a full sibling family) at any of the 18 microsatellite loci. However, the absence of linkage or Hardy-Weinberg disequilibria in the Permanente Creek population indicate that it is not a single sibship from this generation.

Heterozygosity for this population was also extremely low, indicating extreme inbreeding. The value of observed heterozygosity found in Permanente Creek, 0.248, indicates that less than one quarter of the microsatellite loci in each of these fish had a different allele on the paternal and maternal chromosomes. This is almost half the value for the population with the next lowest value of heterozygosity (San Tomas Aquino-Saratoga Creek; 0.488, Table 1) and the lowest value observed with these microsatellite markers in more than 100 California trout populations. It is important to note that measures of allelic richness and heterozygosity cannot be directly compared between studies that use different sets of genetic markers. This is true for nuclear genetic markers

(e.g. microsatellites) even when the same marker type is used, unless enough markers are used in both studies so that the values of the diversity measures have begun to asymptote to the parametric value for the population, which is a set of conditions rarely met with non-model organisms.

Measures of genetic diversity were compared between populations sampled above and below barriers. The two most appropriate measures for comparison are allelic richness, which scales the number of alleles by sample size, and observed heterozygosity, which is the proportion of chromosomes in the population with different microsatellite allele sizes. Allelic richness was higher in below-barrier populations than in above-barrier populations (5.38±0.73 vs. 4.66±1.18), even when the Permanente Creek outlier is excluded (5.38±0.73 vs. 5.04±0.58). Observed heterozygosity was also higher in below-barrier populations (0.616±0.054 vs. 0.570±1.36), but not when Permanente Creek is excluded (0.616±0.054 vs. 0.617±0.042).

These results indicate smaller effective size in above-barrier populations, which is consistent with the expectation of decreased upstream migration and, potentially, also less available spawning and rearing habitat for above-barrier populations, relative to ones with more migratory opportunities. These two forces can lead to gradual genetic erosion, which can contribute to eventual population extirpation (Srikwan and Woodruff 2000). Facilitating upstream migration might help to alleviate such eventual genetic effects, but may also counteract potential adaptation of above-barrier populations that is expected because of the strong selection against downstream migration in such populations.

The results are also consistent with the known insensitivity of heterozygosity compared with allelic diversity measures, relative to reductions in population size

(Cornuet and Luikart 1996, Garza and Williamson 2001). This difference is so great that a population will lose 80% of its alleles while losing only 20% of the heterozygosity in genomic variation (Amos and Harwood 1998).

Chinook salmon

San Francisco Bay and, in particular, Santa Clara Valley streams, have both steelhead/rainbow trout (*O. mykiss*) and Chinook salmon (*O. tshawytscha*) populations. Previous work by Garcia –Rossi and Hedgecock (2002) found that Chinook salmon from the Guadalupe River were most closely related to Central Valley Fall Chinook salmon, perhaps not surprisingly, and not to California Coastal Chinook, which range from the Russian River north to just south of the Klamath River. Although not formally part of this study, genotypes of 28 Chinook salmon juveniles collected in Guadalupe River traps by SCVWD biologists were analyzed. These fish were genotyped with the 13 microsatellite markers recently incorporated into a rangewide reference database of Chinook salmon constructed by the Genetic Analysis of Pacific Salmonids (GAPS) consortium of salmonid genetics labs from university, state, federal and tribal agencies (Seeb et al. 2007). Of the 28 fish analyzed, 25 were assigned to the California Central Valley Fall group. The other three, however, were assigned to lower Columbia River stocks, which is reflective of the close relationship seen between the Central Valley and lower Columbia River populations in coastwide phylogeographic analyses (Seeb et al. 2007). Analysis of Chinook salmon carcasses or other tissue from adult fish would enable evaluation of whether these fish truly represent recent migrants from the Columbia River or whether it

is simply a reflection of recent common ancestry of the Columbia and Sacramento River basin stocks of Fall Chinook.

Museum Specimens

An additional component of this project was an attempt to obtain and analyze genetic data from museum specimens of *O. mykiss* juveniles collected in 1897-1909 by John Otterbein Snyder and associates, and currently maintained at the National Museum of Natural History, which is part of the Smithsonian Institution. These specimens are part of a collection that includes population samples from 7 basins in Central California, including the Pajaro River and Coyote Creek. Sites sampled in these early collections include Uvas and Llagas Creek in the Pajaro River basin and Coyote Creek near Gilroy Hot Springs. Genetic analysis of museum specimens is a difficult endeavor and its success generally depends upon both the preservation method of the specimens and characteristics of the gene sequences targeted for analysis. In general, specimens that have been stored in or preserved with formalin are nearly impossible to extract genetic material from, whereas those preserved solely with ethanol yield genetic data in short fragments. In addition, nuclear DNA, such as microsatellites, is harder to extract than mitochondrial DNA (mtDNA).

The project Principal Investigators (PIs) sampled the Smithsonian collection in 2005 and DNA extraction was performed on tissue from these specimens using recognized techniques for "ancient" DNA, including physically-segregated laboratory space, UV irradiation, as well as sequestered laboratory reagents and filter tip-equipped pipettes.

PCR amplification of the microsatellite loci used in this study was attempted using several different laboratory protocols. None of these were successful in providing interpretable genetic data and it is not expected that data comparable to those reported here for the contemporary samples will be recoverable from the museum specimens. However, subsequent tests demonstrated that small segments of mtDNA are recoverable from the specimens. This is likely due to the high copy number of mitochondrial genes relative to nuclear genes, such as the microsatellites, as well as the substantially shorter DNA fragment targeted by the mtDNA primers. Such mtDNA data provides much less information than do the microsatellites analyzed here and is much more time and resource consuming to produce because of the short segments recovered and the necessary use of DNA sequencing. Nevertheless, the opportunity to directly observe population genetic change over a 100 year time span for California trout populations is so unique, even at the limited scale provided by mtDNA, that analysis of these specimens is still being pursued. State funding has been secured through UC Santa Cruz to analyze mtDNA in the entire Smithsonian collection from the 1897 and 1909 California collections, which includes population samples of trout from the Pajaro and Coyote basins in the Santa Clara Valley, as well as populations of trout from the Salinas, San Lorenzo and Eel Rivers. These results will be reported in a future publication at the end of the project and will acknowledge the SCVWD contribution to the preliminary analysis of these specimens.

Conclusions

Genetic diversity and population structure of trout from the species *O. mykiss* in the Santa Clara Valley were analyzed using 18 microsatellite loci. There was a clear signal of coastal steelhead ancestry in all populations examined, with populations from a particular basin generally most closely related to those from nearby basins. No substantial introgression of hatchery trout into Santa Clara Valley populations was found, although a small number of hatchery fish were captured in the Coyote Creek downstream migrant trap in 1998.

Planting of hatchery fish began in Santa Clara County more than 100 years ago (Thompson 1879). While the genetic analyses reveal minimal introgression of hatchery trout into native Santa Clara County *O. mykiss* populations, there are several potential ecological impacts associated with the practice of planting hatchery trout. The primary ecological issues associated with introducing hatchery fish are effects on carrying capacity, competition, predation, and disease transmission.

Carrying capacity of California coastal streams is often limited by food availability, riparian cover, and suitable spawning habitat. Santa Clara County streams have experienced substantial disturbance over the last 100 years, due to extensive urbanization and habitat alteration, and the carrying capacity of these streams for *O. mykiss* is also drastically affected by altered hydrography, non-point source pollution and introduced/invasive species. As such, many streams that support *O. mykiss* populations are likely at or near ecological carrying capacity and the introduction of additional fish may cause displacement of existing fish and/or density-dependent mortality (Brannon et al. 1999). There is also evidence that stocked fish can have deleterious effects on wild

populations of *O. mykiss* through competitive displacement (Fausch 1988). Predation is another potential negative consequence of introduction of hatchery trout, as they are potential predators of juvenile salmonids (Steward and Bjornn 1990) and other native aquatic species, such as amphibians (Pilliod and Peterson 2001). The stocking of hatchery trout can also lead to the introduction and transmission of novel diseases. Whirling disease (*Mysobolus cerebralis*), in particular, is a concern in coastal California streams (Modin 1998). Introduction of any new species or stock of fish can be a vector for disease that can potentially wipe out or drastically reduce entire populations of fish and amphibians. Given all of the potential detrimental consequences for native *O. mykiss* populations of planting hatchery trout, it would be unwise to conclude that trout planting operations in Santa Clara County have not had negative effects on the native aquatic fauna, solely because of a lack of direct hybridization/introgression. Future planting operations in Santa Clara County should proceed with caution and further research on potential ecological effects should be undertaken.

Both the tree-based analyses and the matrix of genetic differentiation statistics (F_{st} values) indicate that the relationships of Santa Clara Valley trout populations are most dependent upon their basin location and not whether they are found above or below dams and other barriers to anadromy. There was generally a close relationship between trout populations above and below dams within the same basin and, in tree analyses with other California trout populations, the position of above-barrier population samples was consistent with ancestry of coastal steelhead origin. Temporal stability of population structure was found in Coyote Creek, as the population samples taken in 1998, 1999, and 2000 were largely undifferentiated. However, this analysis obscures substantial

heterogeneity within each sample, which is due to population and family structure within the Coyote Creek basin. Further sampling and investigation of upstream trout populations should help to better elucidate this structure. In addition, further investigation of the ancestry of fish from Permanente and Saratoga Creeks should help to determine the reasons for the anomolous results found for these two populations.

Phylogeographic tree analyses that also included Central Valley steelhead and hatchery trout populations, found little evidence of gene flow between Santa Clara Valley and Central Valley trout populations. This is interesting, as Central Valley populations must pass through San Francisco Bay to get to their Sacramento/San Joaquin tributary spawning habitat, but consistent with current classification. In contrast, Chinook salmon populations from the San Francisco Bay are clearly related to Central Valley populations, and are classified as such, and not as coastal Chinook populations.

Garza et al. (in review) examined population structure of coastal steelhead in California and found a pattern of isolation by distance and relatively infrequent but consistent short distance migration (straying) as a major force maintaining population cohesion and genetic diversity in coastal steelhead populations. Analysis of the Santa Clara Valley dataset combined with this dataset found great consistency between the results of the two studies, with the Monterey Bay populations generally clustering with other Monterey Bay populations and the San Francisco Bay populations all in the same region of the tree, as well. This analysis also highlighted a distinction between the Llagas and Uvas Creek populations, which may be partially a result of past stocking with steelhead from the San Lorenzo River in the lower Pajaro River basin. However, the

majority of the inferred population genetic relationships implicated recent migration of limited distance as the dominant force influencing contemporary population structure in this region for anadromous species, which is perhaps not surprising given the importance of migration/straying for salmonid life history.

Two previous studies, Deiner et al. (2007) and Clemento et al. (in prep) examined population genetic structure of trout above and below barriers to anadromy in California coastal basins. Deiner et al. (2007) found that trout breeding above dams in the Russian River were closely related to steelhead trout returning and breeding below dams in the basin, whereas those above natural waterfalls were not. Clemento et al. (in prep) examined trout populations above and below major dams in 5 large river basins in central and southern California and also found a close relationship between trout populations above and below these recent dams; all of the trout populations evaluated in that study were found to be of recent coastal steelhead ancestry. The results of the current study of trout populations in the Santa Clara Valley region yield similar interpretation. The populations of trout in the Guadalupe, Pajaro, and Permanente/Stevens basin above dams that we studied are all of recent steelhead ancestry. It is difficult to pinpoint the timing and magnitude of recent gene flow. However, future individual-based genetic tagging studies could estimate both migration rates between geographically proximate basins, and between localities within a basin.

Acknowledgements

David Salsbery, Lisa Porcella, and Melissa Moore, biologists with the Santa Clara Valley Water District initiated this study and provided most of the trout tissue samples analyzed. They also contributed to the conclusions section on ecological effects of trout stocking and provided other significant contributions, including biological information and the map for Figure 1. They are expected to collaborate on the preparation of a manuscript on these results to be submitted to the e-journal San Francisco Estuary and Watershed Science. Laboratory contributors to the project include Scott Blankenship, Anthony Clemento, Cheryl Dean, Celeste Gallardo, Libby Gilbert, Andres Martinez and Edith Martinez. The samples for two of the Central Valley steelhead populations were provided by Katie Perry and George Edwards of the Native Anadromous Fish and Watershed Branch of the California Department of Fish and Game as part of an ongoing comprehensive study of Central Valley trout populations. The other, from Battle Creek, was provided by biologists from the US Fish and Wildlife Service.

References cited

Aguilar A, Garza JC (2006) A comparison of variability and population structure for major histocompatibility complex and microsatellite loci in California coastal steelhead (*Oncorhynchus mykiss*). Molecular Ecology 15:923-937.

Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences 38:1507-1514.

Amos W, Harwood J (1998). Factors affecting levels of genetic diversity in natural populations. Philosophical Transactions of the Royal Society of London B 353:177-186.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. (2004) GENETIX 4.05: logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000. Université de Montpellier II, Montpellier (France).

Berg WJ, Gall GAE (1988) Gene flow and genetic differentiation among California coastal rainbow trout populations. Canadian Journal of Fisheries and Aquatic Sciences 45:122-131.

Brannon EL, Currens KP, Goodman D, Lichatowich JA, McConnaha WE, Riddell BE, Williams RN (1999) Review of artificial production of anadromous and resident fish in the Columbia River Basin, Part I: A scientific basis for Columbia River production program, Northwest Power Planning Council. 139 pp.

Carlsson J, Nilsson J (2001) Effects of geomorphological structures on genetic differentiation among brown trout populations in a northern Boreal river drainage. Transactions of the American Fisheries Society 130:36-45.

Castric V, Bernatchez L, Belkhir K, Bonhomme F (2002) Heterozygote deficiencies in

small lacustrine populations of brook charr *Salvelinus fontinalis Mitchill* (Pisces, Salmonidae): a test of alternative hypotheses. Heredity, 89:27-35.

Castric V, Bonney F, Bernatchez L (2001) Landscape structure and hierarchical genetic diversity in the Brook Charr, *Salvelinus fontinalis*. Evolution, 55:1016-1028.

Cavalli-Sforza L, Edwards A (1967) Phylogenetic analysis: models and estimation procedures. Evolution, 32:550-570.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59:720-729.

Condrey MJ, Bentzen P (1998) Characterization of coastal cutthroat trout (O*ncorhynchus clarki*) microsatellites and their conservation in other salmonids. Molecular Ecology 7:783-792.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153:1989-2000.

Crispo E, Bentson P, Reznick DR, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. Molecular Ecology 15:49-62.

Deiner K, Garza JC, Coey R, Girman DJ (2007) Population structure and genetic diversity of trout (*Oncorhynchus mykiss*) above and below natural and manmade-barriers in the Russian River, California. Conservation Genetics 8:437-454.

Fausch KD (1988) Tests of competition between native and introduced salmonids in streams: what have we learned? Canadian Journal of Fisheries and Aquatic Sciences 45:2238-2246.

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package), version 3.57c. Department of Genetics, University of Washington, Box 357360, Seattle, Washington, 98105.

Felsenstein J (2003) Inferring Phylogenies. Sinauer Associates, Sunderland MA.

Garcia-Rossi D, Hedgecock D (2002) Provenance analysis of Chinook salmon (*Oncorhynchus tshawytscha*) in the Santa Clara Valley watershed. Report to the Santa Clara Valley Water District.

Garza JC, Gilbert-Horvath E, Spence B, Williams TH, Fish H, Gough S, Anderson JH, Hamm D (in review). Population structure of steelhead in California.

Garza JC, Williamson E (2001) Detection of reduction in population size using data from microsatellite loci. Molecular Ecology 10:305-318.

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

Nielsen JL, Fountain MC, Wright JM (1997) Biogeographic analysis of Pacific trout (*Oncorhynchus mykiss*) in California and Mexico based on mitochondrial DNA and nuclear microsatellites. In: *Molecular Systematics of Fishes* (ed. Kocher TD, and C.A. Stepien), pp. 53-73. Academic Press, London.

Olsen JB, Miller SJ, Spearman WJ, Wenburg JK (2003) Patterns of intra- and inter-population genetic diversity in Alaskan coho salmon: Implications for conservation. Conservation Genetics 4:557-569.

Pearse DP, Donohoe C, Garza JC (2007a) Population genetics of steelhead (*Oncorhynchus mykiss*) in the Klamath River. Environmental Biology of Fishes 80:377-387.

Pearse DP, Wooninck L, Dean C, Garza JC (2007b) Identification of Northeastern Pacific rockfish (*Sebastes* spp.) using multilocus nuclear DNA genotypes. Transactions of the American Fisheries Society 136:272–280.

Pilliod DS, Peterson CR (2001) Local and landscape effects of introduced trout on amphibians in historically fishless watersheds. Ecosystems 4:322-333.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004) GeneClass2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity 95:536-539.

Poissant J, Knight TW, Ferguson MM (2005) Nonequilibrium conditions following landscapes on the genetic structure of a stream-dwelling fish. Molecular Ecology 14:1321–1331.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences USA 94:9197-9221.

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280-1283.

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

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

Seeb LW, Antonovich A, Banks MA, Beacham TD, Bellinger MR, Campbell M, Decovich NA, Garza JC, Guthrie CM, 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.

Spidle AP, Schill WB, Lubinski BA, King TL (2001) Fine-scale population structure in Atlantic salmon from Maine's Penobscot River drainage. Conservation Genetics 2:11-24.

Srikwan S, Woodruff DS (2000) Genetic erosion in isolated small mammal populations following rain forest fragmentation. In: Genetics, Demography and Viability of Fragmented Populations. Young, A. & G. Clarke, eds. Cambridge University Press, Cambridge. Pp. 149-172.

Steward CR, Bjornn TC (1990) Supplementation of salmon and steelhead stocks with hatchery fish: A synthesis of published literature. U.S. Department of Energy, Bonneville Power Administration Project 88-100. Portland, OR.

Sundermeyer R (1999) Molecular genetics of steelhead in the Pajaro River basin. Master of Science Thesis, San Jose State University.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389-399.

Thompson FP (1879) Report of the Commissioners of Fisheries of the State of California for the Years 1878 and 1879. Sacramento, California.

Vincent ER (1987) Effects of stocking catchable-size hatchery rainbow trout on two wild trout species in the Madison River and O'Dell Creek, Montana. North American Journal of Fisheries Management 7:91-105.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. Evolution 38:1358-1370.

Figure legends:

Figure 1: GIS map of Santa Clara Valley area stream system with sampling localities indicated.

Figure 2: Phylogeographic tree of Santa Clara Valley trout populations a) neighbor joining tree constructed with CSE distances b) bootstrap consensus tree-majority rule from 1000 bootstrap replicates.

Figure 3: Factorial correspondence analysis of individual genotypes from 18 microsatellite loci for all Santa Clara Valley populations from this study and the San Lorenzo River. a & b are the same plot viewed from different angles.

Figure 4: Phylogeographic tree of Santa Clara Valley trout populations with Central Valley trout populations and Fillmore Hatchery trout strains included. Fourteen loci only used in analysis. a) neighbor joining tree constructed with CSE distances b) bootstrap consensus tree-majority rule from 1000 bootstrap replicates.

Figure 5: Factorial correspondence analysis of individual genotypes from 14 microsatellite loci for all populations from this study with Central Valley trout populations and Fillmore Hatchery trout strains included. The Central Valley and Fillmore Hatchery trout are the same as in Figure 4.

Figure 6: California-wide tree combining data from the Santa Clara Valley populations with all coastal steelhead trout populations analyzed by Garza et al. (in review) and Clemento et al. (in prep), which span the entire coastal range of the species in California. Data from 14 microsatellite loci are included in the neighbor-joining tree constructed with CSE genetic distances. Populations in blue are from the present study.

Table legends:

Table 1: Samples and summary statistics

Table 2: Microsatellite loci

Table 3: Matrix of all pairwise values of F_{ST}, as estimated by Weir and Cockerham's (1984) estimator.

Table 4: Assignment matrix.

Basin	Tributary/Location	N	N_a	A_r	Expected Hz	Observed Hz	% loci sig. out of HWE	% locus pairs. in LD
Coyote	Mainstem-1998 (B)	15	6.9	6.2	0.673	0.601	11.1	0
	Mainstem-1999 (B)	68	8.7	5.5	0.647	0.600	16.7	18.3
	Mainstem-2000 (B)	217	9.6	5.3	0.627	0.579	66.7	52.3
	Upper Penitencia (B)	20	7.4	6.1	0.701	0.660	5.6	0.7
Guadalupe	Mainstem (B)	15	5.7	-	0.661	0.603	5.6	0
	Guadalupe Ck (B)	126	6.8	4.6	0.596	0.590	0.0	4.6
	Guadalupe Reservoir (A)	20	5.2	4.6	0.651	0.625	5.6	0.7
	Herbert Ck (A)	51	6.8	5.1	0.659	0.670	0.0	1.3
	Los Gatos Ck (B)	62	8.2	5.7	0.662	0.668	16.7	15.7
	Lexington Reservoir (A)	30	6.6	5.0	0.651	0.609	0	0
	Austrian Gulch (A)	20	4.8	4.2	0.581	0.544	11.1	1.3
San Tomas Aquino	Saratoga Ck (B)	86	5.1	3.8	0.533	0.488	22.2	19.6
Permanente	Permanente Ck (B)	20	2.3	2.1	0.256	0.248	0	0
	Stevens Ck (B)	32	8.2	5.9	0.658	0.611	5.6	1.3
	Stevens Reservoir (A)	20	6.7	5.6	0.641	0.589	5.6	0.7
San Francisquito	Los Trancos (B)	29	7.7	5.9	0.691	0.670	11.1	0.7
Pajaro	Llagas Ck (A)	20	5.5	4.8	0.644	0.625	0	0
	Uvas Ck (A)	25	7.8	5.9	0.684	0.655	5.6	0.7
	Uvas Ck (B)	44	6.6	4.8	0.624	0.639	27.8	9.2
	Bodfish Ck (B)	57	5.5	5.4	0.664	0.686	33.3	19
San Lorenzo River	Mainstem (B)	69	10.2	6.1	0.676	0.659	16.7	7.8
Total/Mean		**1046**	**6.8**	**5.0**	**0.6**	**0.6**	**10.1**	**5.2**

Table 1: Sample data and summary statistics for Santa Clara Valley trout genotyped as part of this study and classified in 21 "population" groups. (A) and (B) refer to whether samples were taken above or below known barriers to anadromy. N is number of fish analyzed for that population sample. N_a is observed number of alleles. A_r is allelic richness. Hz is heterozygosity. HWE and LD are defined in the text.

Locus	Primer sequences (5'-3')	No. of Alleles	Range (bp)	Reference
Oki23	F-TGTGCTATAGGGTGAATGTGC	21	118-210	Spidle et al. unpublished,
	R-AACACAGGCATCCCCACTAA			GenBank AF272822
Omy1011	F-AACTTGCTATGTGAATGTGC	26	136-260	Spies et al. unpublished,
	R-GACAAAAGTGACTGGTTGGT			GenBank AY518334
Omy27	F-TTTATGGCTGGCAACTAATGT	7	97-109	McConnell et al. 1995
	R-TTTATGTCATGTCAGCCAGTG			
Omy77	F-CGTTCTCTACTGAGTCAT	21	80-140	Morris et al. 1996
	R- GTCTTTAAGGCTTCACTGCA			
One11	F-GTTTGGATGACTCAGATGGGACT	7	114-124	Scribner et al. 1996
	R-CCTGCTGCCAACACTGTCAA*			
One13	F-TCATACCCCATGCCTCTTCTGTT	20	206-248	Scribner et al. 1996
	R-GGGTGGAGAGACAGGTATCTTGTC*			
Ots1	F-TAGCGTTCACCTGGATTCCC	13	201-293	Banks et al. 1999
	R-CATGCTATTTCCAGACGGCA*			
OtsG3	F-GGACAGGACCGTCTGCTAAATGACTG	19	139-243	Williamson et al. 2002
	R-GGATGGATTGATGAATGGGTGGG			
OtsG43	F-AACTCCCGTTGACAATTTACTGTTG	15	145-209	Williamson et al. 2002
	R-TTTTGGCAAAGTTGGCTACTCTG			
OtsG85	F-CCATGTCAGCACTGACTTAAT	35	129-285	Williamson et al. 2002
	R-GGATGTTGTTCCTAATGTTTT			
Ots103	F-AGGCTCTGGGTCCGTG	6	58-92	Beacham et al. 1998
	R-TGATATGGTGTGATAGCTGG			
OtsG243	F-TTATTAAACTGCACTGTCTAACTACA	5	107-117	Williamson et al. 2002
	R-GTATGCAGCAAGCCAGGTG			
OtsG249	F-ATGGCAGTTAAGAGAACAAAAGTT*	22	147-243	Williamson et al. 2002
	R-GTACAACCCCTCTCACCTACCC			
OtsG253	F-CGCTGCAGAAACATTTTCGA*	25	165-269	Williamson et al. 2002
	R-AATTGGGTCATTAAGGCTCTGTGG			
OtsG401	F-CTGCCCTGAGAAGCTGGAGTGCTC	20	165-249	Williamson et al. 2002
	R-TTGCCCCACCCTTGCATCTATCCA			
OtsG409	F-GTAGCCATTTGTGTCACCATCATT	3	86-90	Williamson et al. 2002
	R-CATTCTCCTGCCTCACAGAGTTTA			
Ssa85	F-AGGTGGGTCCTCCAAGCTAC	21	96-157	O'Reilly et al. 1996
	R-ACCCGCTCCTCACTTAATC			
Ssa289	F-CTTTACAAATAGACAGACT	10	105-125	McConnell et al. 1995
	R-TCATACAGTCACTATCATC			

Table 2: Eighteen microsatellite lociused to genotype *Oncorhynchus mykiss* in this study. Primer sequences total number of alleles and range in allele size observed in the study populations is included, as is the reference for the original description. *Indicates primer was redesigned from original reference sequence for optimizatio purposes; Note that Banks et al. (1999) contains incorrect primer sequences.

	Coyote Ck-1999 (B)	Coyote Ck-2000 (B)	Coyote Ck-Up. Penitencia (B)	Guadalupe River (B)	Guadalupe Creek (B)	Guadalupe Reservoir (A)	Guadalupe-Herbert Ck (A)	Guadalupe-Los Gatos Ck (B)	Guadalupe-Lexington (A)	Guadalupe-Austrian Gulch (A)	San Tomas-Saratoga Ck (B)	Permanente Ck (A)	Perm-Stevens Ck (B)	Perm-Stevens Reservoir (A)	San Francisquito (B)	Pajaro-Llagas Ck (A)	Pajaro-Uvas Ck (A)	Pajaro-Uvas Ck (B)	Pajaro-Bodfish Ck (B)	San Lorenzo River (B)
Coyote Ck-1998 (B)	**0.0166**	0.0205	0.0273	0.0515	0.0760	0.0751	0.0501	0.0478	0.0692	0.1228	0.1301	0.3033	0.0412	0.0477	0.0345	0.0613	0.0440	0.0691	0.0596	0.0338
Coyote Ck-1999 (B)		**0.0054**	0.0228	0.0390	0.0876	0.0936	0.0578	0.0489	0.0778	0.1251	0.1346	0.2573	0.0452	0.0515	0.0408	0.0767	0.0527	0.0647	0.0650	0.0371
Coyote Ck-2000 (B)			0.0306	0.0463	0.1009	0.1083	0.0679	0.0647	0.0906	0.1349	0.1282	0.2461	0.0576	0.0648	0.0553	0.0822	0.0532	0.0669	0.0716	0.0414
Coyote Ck-Up. Penitencia (B)				0.0385	0.0822	0.0643	0.0584	0.0412	0.0777	0.1152	0.1289	0.3138	0.0371	0.0487	0.0353	0.0683	0.0409	0.0647	0.0629	0.0316
Guadalupe River (B)					0.0865	0.0816	0.0611	0.0466	0.0729	0.1290	0.1450	0.3479	0.0308	0.0656	0.0573	0.0877	0.0550	0.0624	0.0709	0.0455
Guadalupe Creek (B)						0.0732	0.0713	0.0763	0.0959	0.1542	0.1711	0.2731	0.0796	0.1003	0.0880	0.1197	0.1058	0.1104	0.0979	0.0849
Guadalupe Reservoir (A)							0.0757	0.0723	0.0621	0.1290	0.1450	0.3344	0.0758	0.1015	0.0745	0.1116	0.0855	0.1098	0.0887	0.0760
Guadalupe-Herbert Ck (A)								0.0581	0.0623	0.1018	0.1327	0.2487	0.0585	0.0611	0.0618	0.0999	0.0637	0.0712	0.0689	0.0520
Guadalupe-Los Gatos Ck (B)									0.0795	0.1272	0.1346	0.2671	0.0216	0.0400	0.0401	0.0763	0.0634	0.0758	0.0787	0.0547
Guadalupe-Lexington (A)										0.0853	0.1413	0.3247	0.0928	0.1036	0.0649	0.1090	0.0842	0.0870	0.0839	0.0760
Guadalupe-Austrian Gulch (A)											0.2256	0.3908	0.1403	0.1459	0.1291	0.1807	0.1267	0.1356	0.1493	0.1291
San Tomas Aquino-Saratoga Ck (B)												0.3480	0.1123	0.1551	0.1306	0.1392	0.1236	0.1172	0.1172	0.1118
Permanente Ck (A)													0.2572	0.2483	0.3179	0.3199	0.2937	0.3312	0.2567	0.2564
Permanente-Stevens Ck (B)														0.0225	0.0527	0.0659	0.0622	0.0737	0.0651	0.0496
Permanente-Stevens Reservoir (A)															0.0651	0.0800	0.0724	0.0891	0.0825	0.0615
San Francisquito (B)																0.0769	0.0599	0.0589	0.0689	0.0485
Pajaro-Llagas Ck (A)																	0.0477	0.0870	0.0693	0.0549
Pajaro-Uvas Ck (A)																		0.0658	0.0510	0.0270
Pajaro-Uvas Ck (B)																			0.0392	0.0432
Pajaro-Bodfish Ck (B)																				0.0392

Table 3: Pairwise values of Fst, the standardized variance in allele frequencies between populations, for all of the 20 "population" samples from this study. Values in bold are not significantly different from zero, indicating no differentiation.

TruePop	Coyote Ck-1998 (B)	Coyote Ck-1999 (B)	Coyote Ck-2000 (B)	Coyote Ck-Up. Penitencia (B)	Guadalupe River (B)	Guadalupe Creek (B)	Guadalupe Reservoir (A)	Guadalupe-Herbert Ck (A)	Guadalupe-Los Gatos Ck (B)	Guadalupe-Lexington (A)	Guadalupe-Austrian Gulch (A)	San Tomas-Saratoga Ck (B)	Permanente Ck (A)	Perm-Stevens Ck (B)	Perm-Stevens Reservoir (A)	San Francisquito (B)	Pajaro-Llagas Ck (A)	Pajaro-Uvas Ck (A)	Pajaro-Uvas Ck (B)	Pajaro-Bodfish Ck (B)	San Lorenzo River (B)	Filmore-Coleman	Fillmore-MtWhitE	Fillmore-MtWhitL	Fillmore-Virginia	Fillmore-Wyoming	N
Coyote Ck-1998 (B)		6	3																			6					15
Coyote Ck-1999 (B)	2	32	20	4				1	2					1		4					2						68
Coyote Ck-2000 (B)		35	149	13	3			3	4	1				1	1	2					5						217
Coyote Ck-Up. Penitencia (B)		1	6	11												1			1								20
Guadalupe River (B)		2	1	1	8	1								2													15
Guadalupe Creek (B)						126																					126
Guadalupe Reservoir (A)							19														1						20
Guadalupe-Herbert Ck (A)		1						47								2					1						51
Guadalupe-Los Gatos Ck (B)		1		1					58					1	1												62
Guadalupe-Lexington (A)								1		28	1																30
Guadalupe-Austrian Gulch (A)							1			1	18																20
San Tomas Aquino-Saratoga Ck (B)												86															86
Permanente Ck (B)													20														20
Permanente-Stevens Ck (B)			1		3				3					22	3												32
Permanente-Stevens Reservoir (A)			1						1						16		1					1					20
San Francisquito (B)		2	1	1	1									1		23											29
Pajaro-Llagas Ck (A)																	20										20
Pajaro-Uvas Ck (A)			3															22									25
Pajaro-Uvas Ck (B)																			37	4	3						44
Pajaro-Bodfish Ck (B)																			2	52	3						57
San Lorenzo River (B)	1	1	4	1						1					2	1	1	1	1	2	53						69
N correct/site	**0**	**31**	**149**	**11**	**8**	**126**	**19**	**47**	**58**	**28**	**18**	**86**	**20**	**22**	**16**	**23**	**20**	**22**	**37**	**52**	**53**	**80.9**	**% correct assign**				
% correct/site	**0**	**46**	**69**	**55**	**53**	**100**	**95**	**92**	**94**	**93**	**90**	**100**	**100**	**69**	**80**	**79**	**100**	**88**	**84**	**91**	**77**	**91.4**	**% correct to basin**				

Table 4: Matrix of individual genotypic assignments for all fish in the study, with 5 Fillmore Hatchery trout strains included as possible populations of origin. Rows represent the assigned population of origin for each fish from each populations and the columns represent all fish assigned to a given population. The most likely population of origin is always reported, even if the probability is low. Colors represent intrabasin assignments.

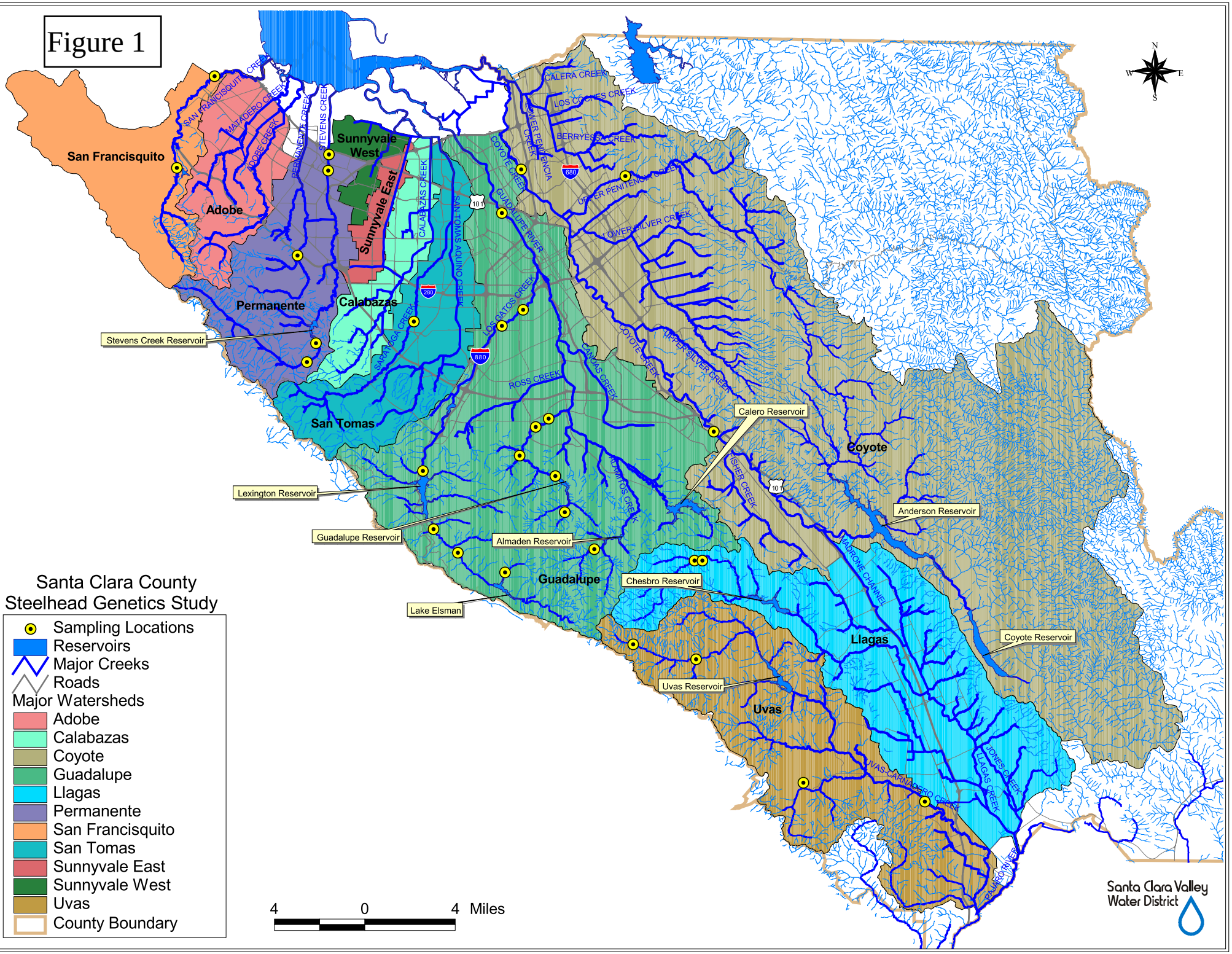


Figure 1
San Francisquito
Adobe
Sunnyvale West
Sunnyvale East
Permanente
Calabazas
San Tomas
Guadalupe
Llagas
Uvas
Coyote
Stevens Creek Reservoir
Lexington Reservoir
Guadalupe Reservoir
Almaden Reservoir
Lake Elsman
Chesbro Reservoir
Uvas Reservoir
Calero Reservoir
Anderson Reservoir
Coyote Reservoir
SAN FRANCISQUITO CREEK
MATADERO CREEK
MOORE CREEK
PERMANENTE CREEK
STEVENS CREEK
CALERA CREEK
LOS COCHES CREEK
UPPER PENITENCIA CREEK
LOWER PENITENCIA CREEK
BERRYESSA CREEK
COYOTE CREEK
GUADALUPE RIVER
UPPER SILVER CREEK
LOWER SILVER CREEK
CALABAZAS CREEK
SARATOGA CREEK
SANTA CLARA VALLEY CREEK
LOS GATOS CREEK
ROSS CREEK
CANOAS CREEK
GUADALUPE CREEK
ALMADEN CREEK
FISHER CREEK
ALMADEN CHANNEL
LLAGAS CREEK
JONES CREEK
UVAS CARNADERO CREEK
PAJARO RIVER
Santa Clara County
Steelhead Genetics Study
Sampling Locations
Reservoirs
Major Creeks
Roads
Major Watersheds
Adobe
Calabazas
Coyote
Guadalupe
Llagas
Permanente
San Francisquito
San Tomas
Sunnyvale East
Sunnyvale West
Uvas
County Boundary
4 0 4 Miles
N
W
E
S
Santa Clara Valley
Water District

Figure 2a

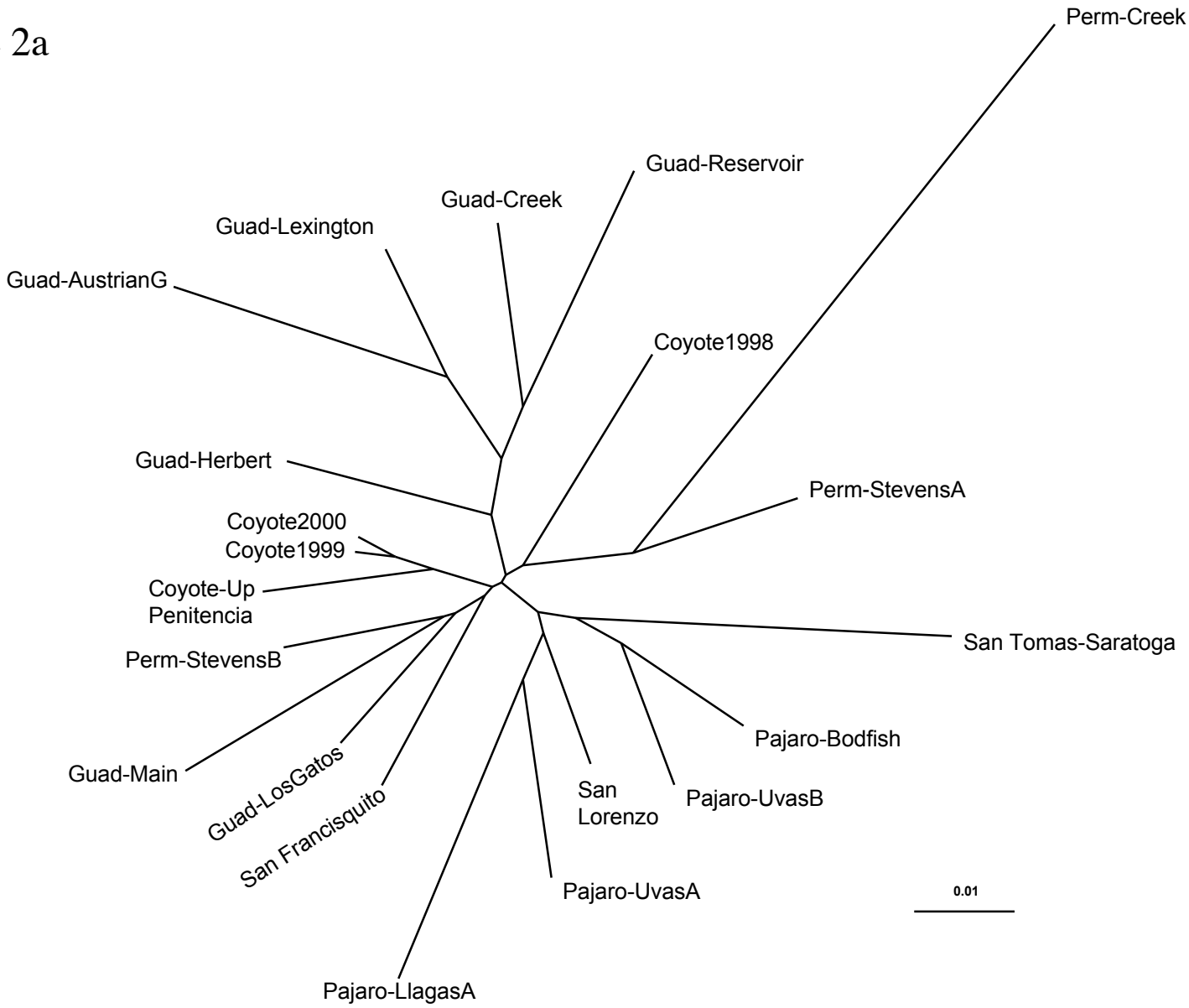

Perm-Creek
Guad-Reservoir
Guad-Creek
Guad-Lexington
Guad-AustrianG
Coyote1998
Guad-Herbert
Perm-StevensA
Coyote2000
Coyote1999
Coyote-Up
Penitencia
San Tomas-Saratoga
Perm-StevensB
Pajaro-Bodfish
Guad-Main
Guad-LosGatos
San Francisquito
San
Lorenzo
Pajaro-UvasB
Pajaro-UvasA
0.01
Pajaro-LlagasA

Figure 2b

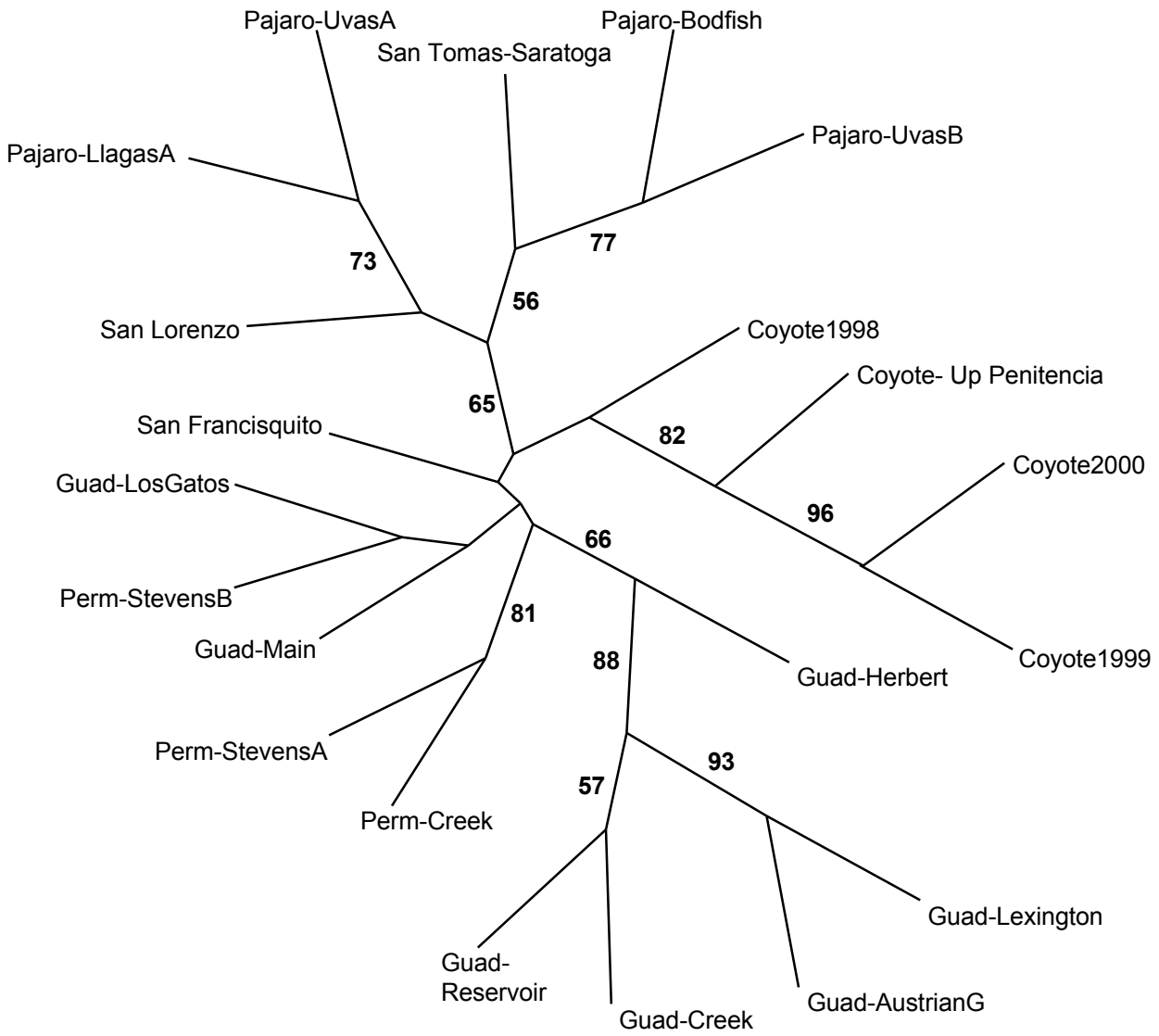

Pajaro-UvasA
Pajaro-Bodfish
San Tomas-Saratoga
Pajaro-UvasB
Pajaro-LlagasA
77
73
56
San Lorenzo
Coyote1998
Coyote- Up Penitencia
65
San Francisquito
82
Coyote2000
Guad-LosGatos
96
66
Perm-StevensB
81
Guad-Main
88
Coyote1999
Guad-Herbert
Perm-StevensA
93
57
Perm-Creek
Guad-Lexington
Guad-
Reservoir
Guad-AustrianG
Guad-Creek

Figure 3a

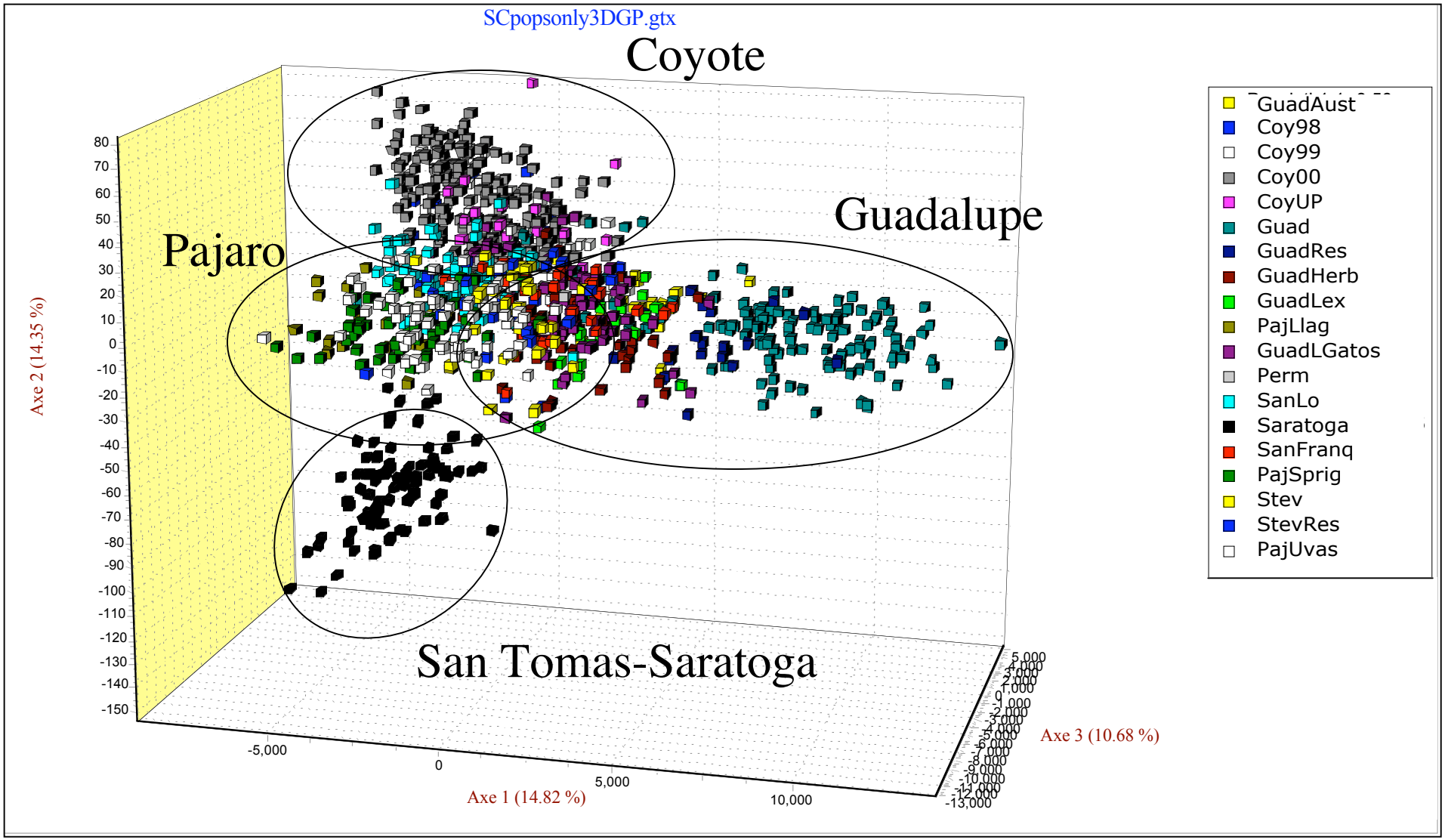

SCpopsonly3DGP.gtx
Coyote
Guadalupe
Pajaro
San Tomas-Saratoga
Axe 2 (14.35 %)
Axe 1 (14.82 %)
Axe 3 (10.68 %)
GuadAust
Coy98
Coy99
Coy00
CoyUP
Guad
GuadRes
GuadHerb
GuadLex
PajLlag
GuadLGatos
Perm
SanLo
Saratoga
SanFranq
PajSprig
Stev
StevRes
PajUvas

Figure 3b

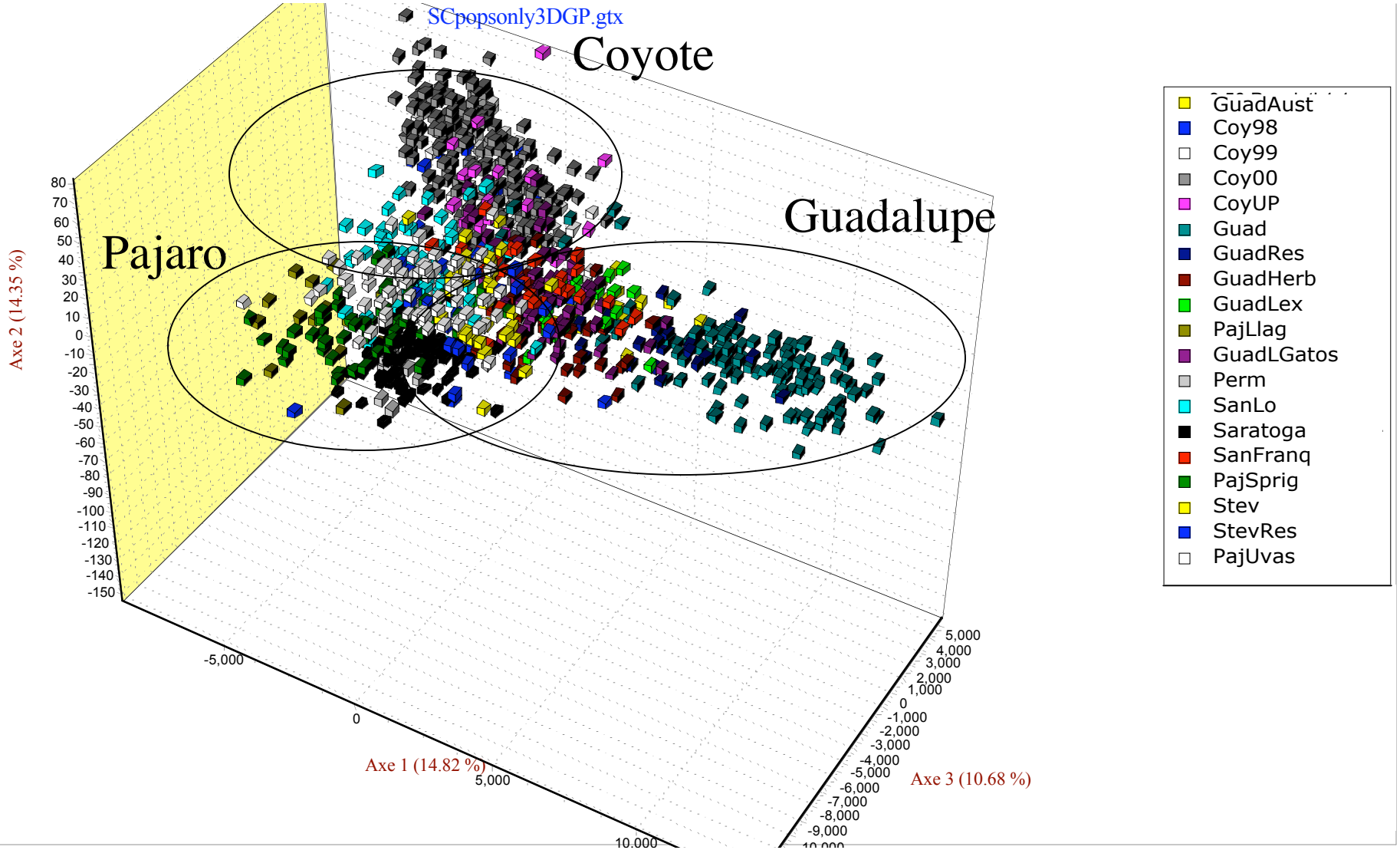

SCpopsonly3DGP.gtx
Coyote
Pajaro
Guadalupe
Axe 2 (14.35 %)
Axe 1 (14.82 %)
Axe 3 (10.68 %)
GuadAust
Coy98
Coy99
Coy00
CoyUP
Guad
GuadRes
GuadHerb
GuadLex
PajLlag
GuadLGatos
Perm
SanLo
Saratoga
SanFranq
PajSprig
Stev
StevRes
PajUvas

Figure 4a

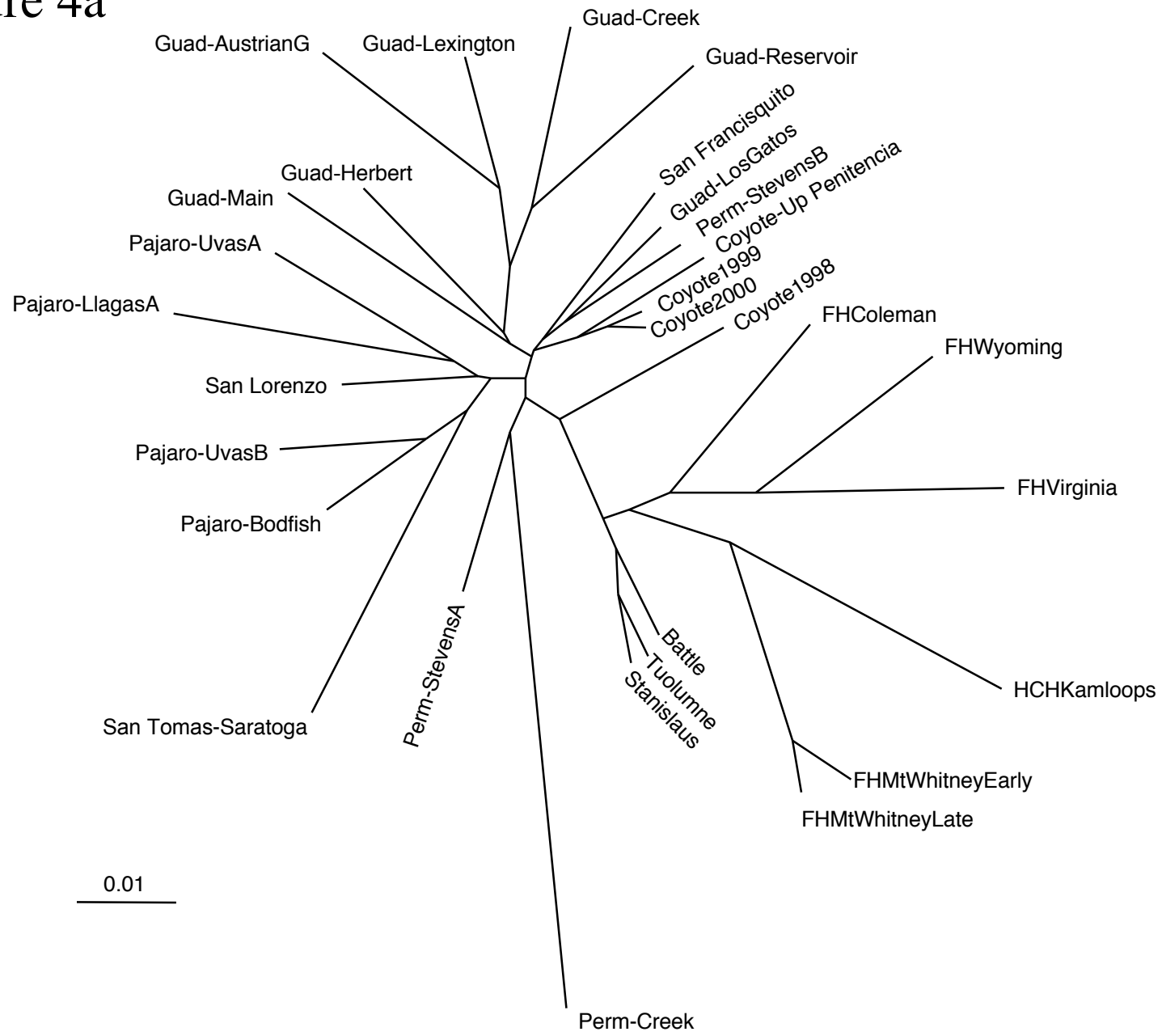

Guad-Creek
Guad-AustrianG
Guad-Lexington
Guad-Reservoir
San Francisquito
Guad-LosGatos
Perm-StevensB
Coyote-Up Penitencia
Guad-Herbert
Guad-Main
Pajaro-UvasA
Coyote1999
Coyote2000
Coyote1998
Pajaro-LlagasA
FHColeman
FHWyoming
San Lorenzo
Pajaro-UvasB
FHVirginia
Pajaro-Bodfish
Battle
Tuolumne
Stanislaus
Perm-StevensA
HCHKamloops
San Tomas-Saratoga
FHMtWhitneyEarly
FHMtWhitneyLate
0.01
Perm-Creek

Figure 4b

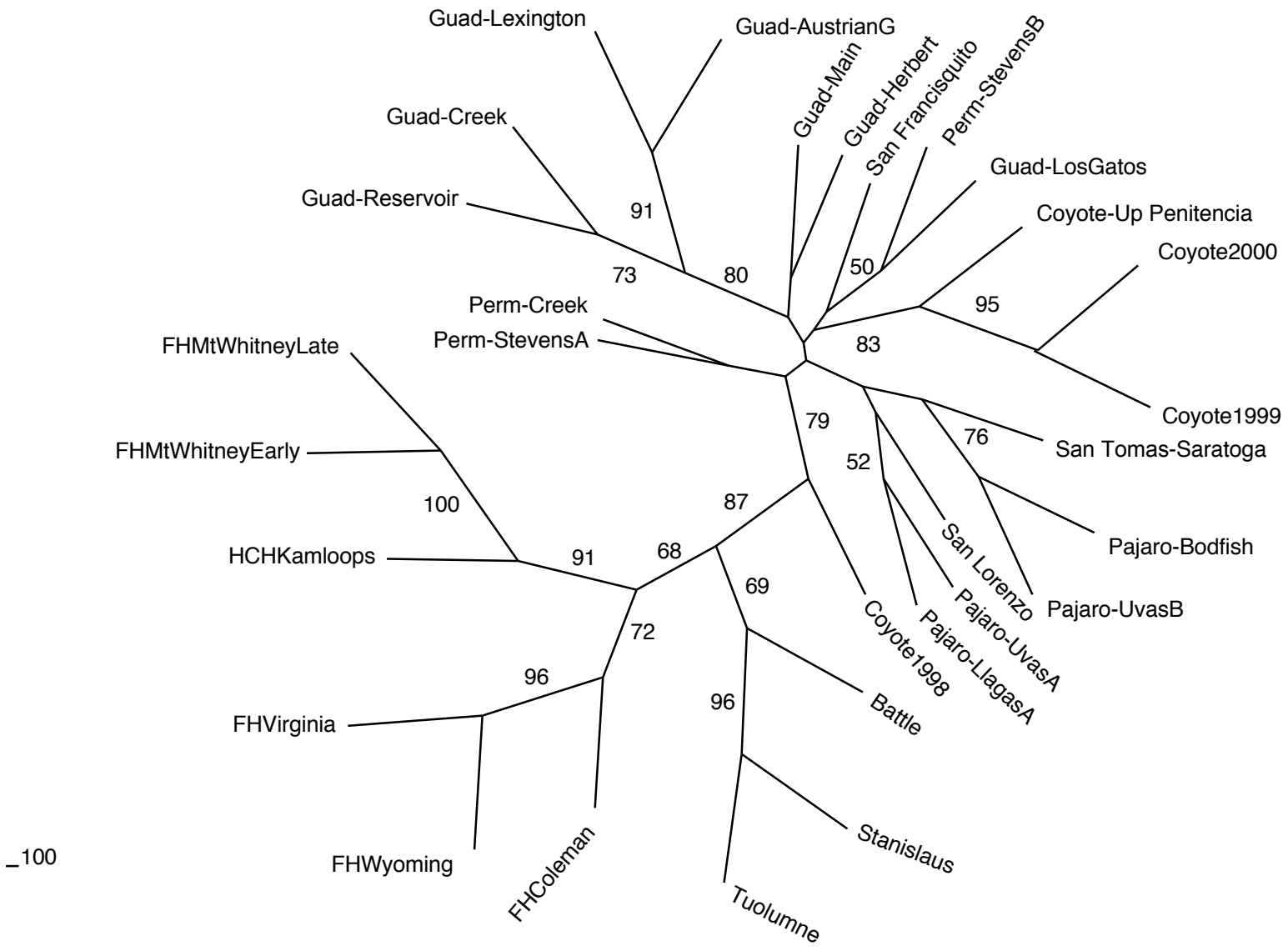

Guad-Lexington
Guad-AustrianG
Guad-Main
Guad-Herbert
San Francisquito
Perm-StevensB
Guad-Creek
Guad-LosGatos
Guad-Reservoir
Coyote-Up Penitencia
Coyote2000
Perm-Creek
Perm-StevensA
FHMtWhitneyLate
Coyote1999
San Tomas-Saratoga
FHMtWhitneyEarly
Pajaro-Bodfish
HCHKamloops
San Lorenzo
Pajaro-UvasB
Pajaro-UvasA
Pajaro-LlagasA
Coyote1998
Battle
FHVirginia
Stanislaus
FHWyoming
FHColeman
Tuolumne
91
73
80
50
95
83
79
76
52
100
87
91
68
69
72
96
96
100

Figure 5

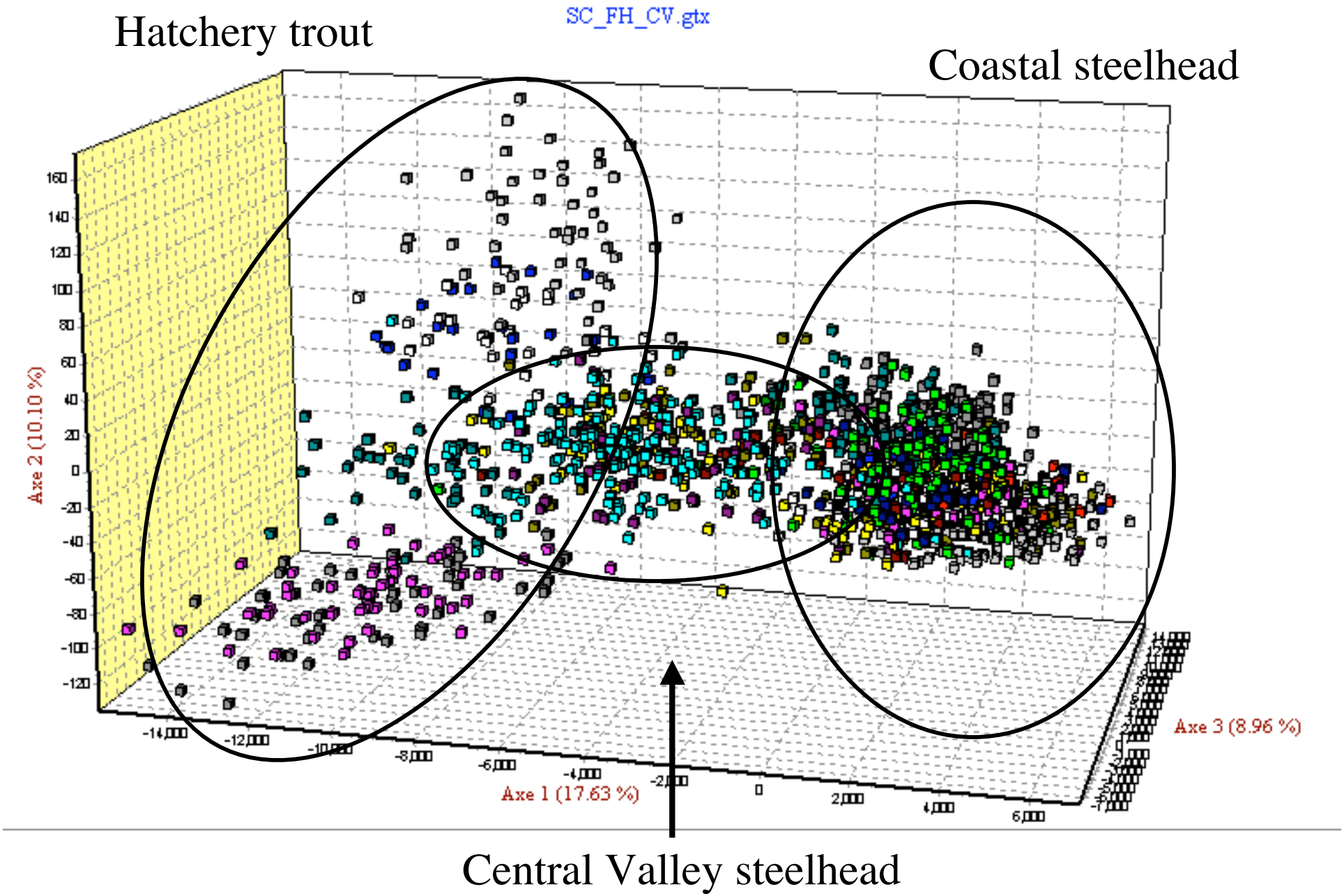
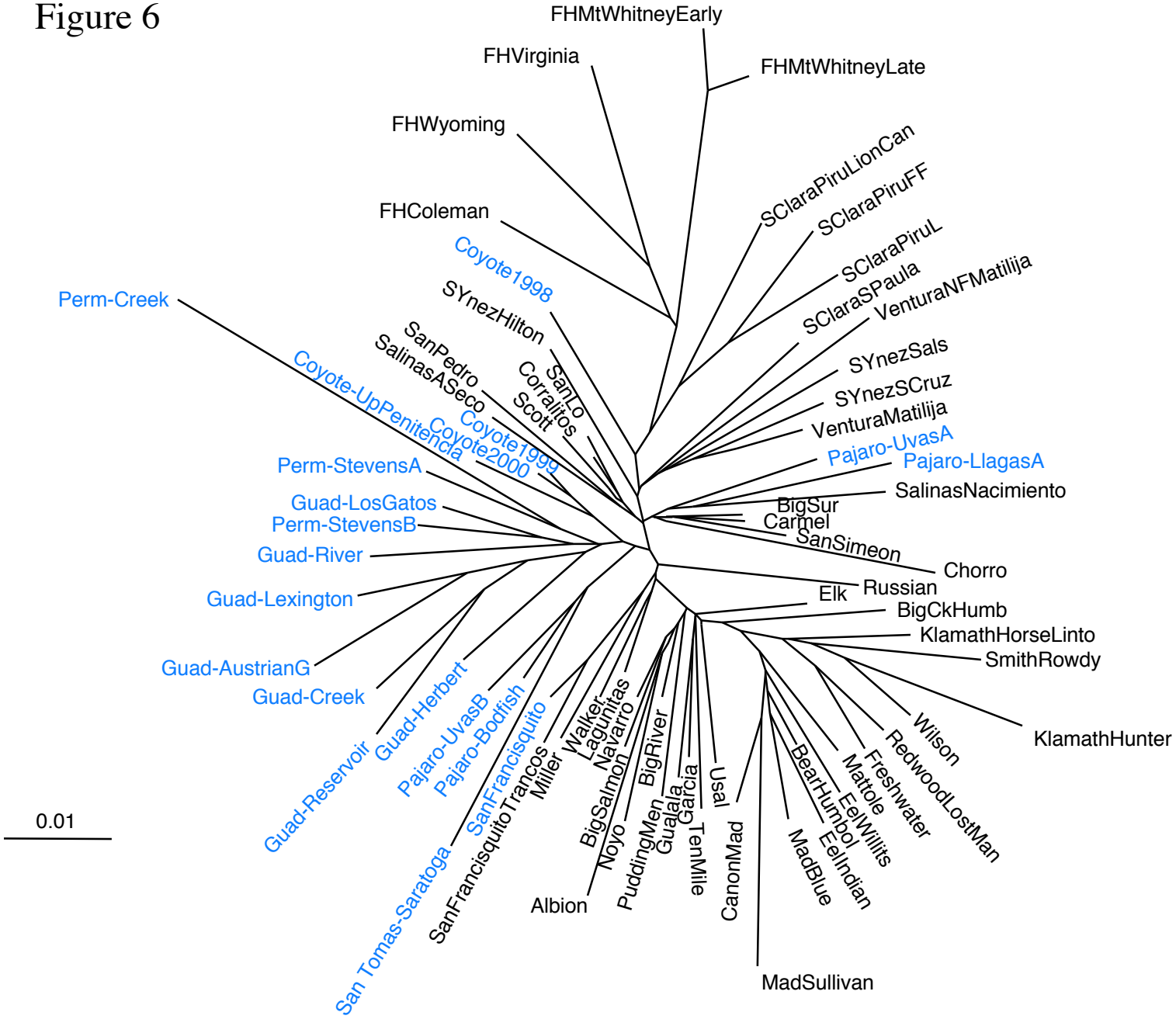

Hatchery trout
SC_FH_CV.gtx
Coastal steelhead
Axe 2 (10.10 %)
Axe 1 (17.63 %)
Axe 3 (8.96 %)
Central Valley steelhead

Figure 6


FHMtWhitneyEarly
FHVirginia
FHMtWhitneyLate
FHWyoming
SClaraPiruLionCan
FHColeman
SClaraPiruFF
SClaraPiruL
Coyote1998
SClaraSPaula
VenturaNFMatilija
Perm-Creek
SYnezHilton
SanPedro
SYnezSals
SalinasASeco
SanLo
Corralitos
SYnezSCruz
Coyote-UpPenitencia
Scott
VenturaMatilija
Coyote1999
Coyote2000
Pajaro-UvasA
Pajaro-LlagasA
Perm-StevensA
SalinasNacimiento
Guad-LosGatos
BigSur
Perm-StevensB
Carmel
SanSimeon
Guad-River
Chorro
Guad-Lexington
Elk
Russian
BigCkHumb
KlamathHorseLinto
Guad-AustrianG
SmithRowdy
Guad-Herbert
Guad-Creek
Pajaro-UvasB
Pajaro-Bodfish
KlamathHunter
Guad-Reservoir
SanFrancisquito
Walker
Wilson
Lagunitas
Navarro
RedwoodLostMan
Miller
BigRiver
Freshwater
BigSalmon
Mattole
BearHumbol
San Tomas-Saratoga
SanFrancisquitoTrancos
Albion
Noyo
PuddingMen
Gualala
Garcia
TenMile
Usal
CanonMad
MadBlue
EelWillits
EelIndian
MadSullivan
0.01