Final Report
To

California Department of Fish and Game
Fisheries Restoration Grant Program

Grant #P0510530: Historical Baseline for Genetic Monitoring of Coastal California Steelhead, *Oncorhynchus mykiss*

Devon Pearse and John Carlos Garza

Institute of Marine Sciences
University of California, Santa Cruz

and

Fisheries Ecology Division, Southwest Fisheries Science Center
NOAA Fisheries Service

Author for correspondence:
Dr. Devon Pearse
NOAA/NMFS
110 Shaffer Road
Santa Cruz, CA 95060
devon.pearse@noaa.gov
Phone: 831-420-3906
Fax: 831-420-3977

Introduction

Direct information on long-term population trends is extremely difficult to obtain for most wild vertebrate species, particularly fish, yet this type of data is needed in order to accurately assess the impacts of past human activity, as well as to evaluate specific wildlife management actions and hatchery practices and understand the demographic history of a given population. From a population genetic perspective, patterns of genetic variation detected in modern populations may not accurately reflect the historical relationships among their ancestral populations. Recent anthropogenic impacts through habit modification or species management practices have the potential to significantly alter the genetic landscape of a given species. In such situations, inference based solely on data from extant populations may lead to inaccurate conclusions about population relationships and misguided management decisions.

The species *Oncorhynchus mykiss* encompasses anadromous and resident trout with diverse life-histories, and includes a number of named subspecies and races. Steelhead is the name given to members of *O. mykiss* that are anadromous, undertaking at least one sea migration. In California, six Distinct Population Segments (DPSs), formerly designated as Evolutionarily Significant Units (ESUs), of steelhead have been delineated on the basis of genetic, geographic and ecological variation: five coastal DPSs and one in the Central Valley (Busby *et al.*, 1996). With the exception of the Klamath Mountain Province DPS, all California steelhead are protected under the US Endangered Species Act (ESA), with the Southern California DPS classified as "Endangered" and all others as "Threatened" (NMFS 2006). Thus it is important to understand the population genetic

relationships among all California populations, as well as the impacts of human activities on steelhead migratory behavior.

Many fish species, particularly anadromous salmonids, have highly variable population dynamics, with large numbers of individuals reproducing in some years and few in others. Species or populations that have been reduced in size are vulnerable to an increased risk of further decline due to genetic effects caused by decreased genetic variation. The primary population parameters that influence this process are the effective population size and the frequency of migration among discrete breeding populations (Wright, 1931). Along with evolutionary genetic parameters such as mutation and natural selection, these factors determine the distribution of genetic variation among populations of a species, and are critical for understanding the relationships among populations in a given geographic area.

The availability of population samples taken more than 100 years ago from several California steelhead DPSs provided an almost unprecedented opportunity to evaluate the genetic composition of current populations relative to their ancestral stocks. We obtained tissue samples of *O. mykiss* (identified as *Salmo irideus* or *S. gairdneri*) from specimens housed at the Smithsonian Institution National Museum of Natural History and collected from several important coastal basins in 1897 and 1909 (Snyder, 1912). They were preserved in ethanol at least during most of this time, and preliminary tests indicated that amplifiable DNA could be recovered from many of the samples. Using these samples, we evaluated the historical population genetic composition and diversity of wild steelhead populations in the Northern California, Central California Coast, and South-Central California Coast Steelhead DPSs (Figure 1). We used these data

to provide a historical reference with which to compare patterns of genetic variation seen in modern *O. mykiss* populations and evaluate the effects of more than a century of human impacts on coastal steelhead populations in California. In doing so, we evaluate the following hypotheses:

1) Steelhead currently spawning in coastal California streams are derived from the steelhead populations historically present in those streams, and population genetic composition has not been significantly modified by hatchery or wild migrants from other streams: Samples of coastal steelhead from 1897 and 1909 provide a historical record of the native genetic composition of these runs prior to most hatchery and other anthropogenic impacts on the region.

2) The population genetic structure currently observed in coastal California steelhead is similar to historical population genetic structure: If hatchery practices and other anthropogenic causes have not greatly affected these steelhead populations, then historical genetic samples are expected to appear most genetically similar to each of their respective present-day stream populations, and patterns of population genetic structure estimated from historical and modern populations should be similar.

The results of this study thus have a direct application to population management and recovery by providing a historical baseline for monitoring, an indication of the long-term stability and historical uniqueness of central California steelhead populations, and data on the extent to which stocking practices have impacted the genetic composition of

each population. Data on the historical genetic diversity present in *O. mykiss* also allows evaluation of the success of current conservation measures and management plans in preserving the integrity of coastal California steelhead populations.

Methods

Samples and DNA extraction

Locations and number of samples from which DNA sequence data was successfully recovered are shown in Table 1. Although greater numbers of samples were available for some historical populations, the low quality of the extracted DNA created difficulties in amplification and sequencing that limited the number of samples that could be analyzed. In addition, we observed high variability in the success of PCR amplifications among samples from different locations and collection dates. In some cases, no usable amplification products were obtained from any sample in a given collection, while samples from other collections appeared to be relatively well preserved. We therefore chose to focus on obtaining a sufficient number of individual sequences from at least one location in each of the six sampled river basins.

All historical tissue samples were taken from ethanol-preserved specimens and dried on blotter paper for transport to the laboratory. It is not known if they were subjected to other preservation methods at any time in their curation. All further handling was done in an isolated room, separate from the main molecular biology laboratory to avoid contamination with modern *O. mykiss* DNA. Dedicated pipettes and laboratory

reagents were used for all historical DNA extractions, and all other equipment was cleaned with a bleach solution and dried prior to use. DNA extractions were conducted using the standard DNeasy 96 tissue protocol (Qiagen, Inc.), processed manually with centrifugation. Two blank extraction controls were included in each 96-well extraction plate, and these were used as negative controls in all PCR amplifications.

Sequences from modern *O. mykiss* were obtained for comparison with the distributions of haplotype frequencies observed in the historical samples. All modern samples were collected between 1997 and 2004 from the same watersheds that were sampled in the historical collections, in approximately the same locations, using standard electrofishing protocols (Garza et al. 2004, Table 1). DNA extractions of modern samples were conducted using a standard DNeasy 96 tissue protocol and processed on a Qiagen BioRobot 3000 (Qiagen, Inc.).

Microsatellite Loci

Amplification of microsatellite loci known to be informative for population genetic analysis in California *O. mykiss* was not successful when applied to DNA extracted from the historical collections. Although some individuals produced amplification products for some loci, these results were inconsistent and mixed with a high failure and allelic dropout rate. As a result, further attempts at microsatellite amplification were abandoned.

Mitochondrial DNA amplification

The mitochondrial (mtDNA) control region is one of the most commonly used molecular markers in a wide variety of taxa, and has proven useful in both population genetic and phylogeographic studies (Avise 2000). In *O. mykiss*, several studies have examined the distribution of genetic variation in the mtDNA control region in various parts of the species range (Bagley & Gall 1998; Nielsen et al. 1994, 1998; McCusker et al. 2000; Thrower et al. 2004). Recently, Graziano et al. (2005) described the 29 known 3' domain mtDNA control region haplotypes that have been observed in *O. mykiss* in order to standardize the nomenclature used by different researchers for this marker. We followed their nomenclature for the haplotypes observed in the present study.

Typically, the mtDNA control region is amplified using PCR primers that bind in conserved regions or in adjacent conserved tRNA genes. In *O. mykiss*, the primers P2 and S-Phe (Nielsen et al. 1994) amplify an approximately 193 base pair segment of the 3' domain of the control region, ending near the beginning of the phenylalanine tRNA gene. These primers were used successfully for all modern sample amplification and sequencing. However, the DNA extracted from the historical samples was degraded, making amplification of the full ~200 bp fragment difficult. We thus designed two sets of primers that amplify smaller overlapping fragments (Frag1 and Frag2, 86 and 79 base pairs, respectively. See Table 2). Together these fragments cover 139 bp in the middle of the P2/S-Phe fragment, and include all but two of the known variable sites (Table 2, Figures 4 and 5). For some historical individuals we were able to amplify a fragment by using a combination of the forward Frag2 primer in combination with the S-Phe primer,

which combined with Frag1 provided coverage of all known variable bases. To verify that all primer combinations were targeting the same mtDNA region, a single modern individual was used to amplify all possible forward and reverse primer combinations, and the resulting sequences were aligned and check for consistency.

All PCR amplifications included an extraction or water negative control to detect possible contamination, and amplification products were discarded when evidence of contamination was discovered. For all historical samples, laboratory work through the PCR amplification stage was conducted in an isolated laboratory using dedicated equipment free from contamination with modern salmonid tissues or extracts. In cases when initial amplification of a historical sample was weak, a second PCR amplification was sometimes performed following the same protocol, using a dilution of the first PCR as a template. Amplifications were performed in 30 or 40 ul reactions, and each reaction contained 1X ABI buffer, 5.0 mM $MgCl_2$, 0.1mM each dNTP, 1 uM each primer, 1 ul BSA buffer, 1 Unit Taq Gold DNA polymerase (Applied Biosystems), and 4 ul DNA extract. Historical sample DNA was used undiluted, while modern samples were diluted 1:20 prior to use in PCR.

Sequencing

Following PCR amplification, products were visually verified on 2% agarose gels stained with ethidium bromide (Figure 2). Successful amplifications were directly purified prior to sequencing using Qiagen PCR purification spin columns. Amplifications of Frag2 often contained a strong primer dimer band, smaller than the target fragment,

which prevented direct purification. To purify these samples, the entire amplification product was electrophoresed in a 3% agarose gel and the target band was excised from the gel and purified using the Qiagen Qiaquick gel purification kit, following the manufacturer's protocol. All sequencing reactions were carried out using BigDye sequencing chemistry (v1.1 and v3.1, Applied Biosystems). Following the sequencing reaction, products were purified using Sephadex, and electrophoresed on ABI 377 or 3730 automated sequencers. Sequences were checked and aligned by eye using Sequencher software (Figures 3 and 4, Gene Codes, Inc.).

Data analysis

All haplotype sequences were compiled and aligned in a single file and variable nucleotide sites were identified by eye and using the program PAUP* v4.0b10 (Swofford 2001). Ambiguous nucleotide sites critical to haplotype identification were rechecked in the original electropherograms, and a clean final sequence was determined for each individual retained in the final data set. Based on the observed nucleotide polymorphisms, each individual sequenced was assigned a haplotype designation (i.e. MYS01, MYS06, etc), and all population genetic analyses were based strictly on haplotype identity.

Standard population genetic statistics, including pairwise F_{ST} values and Markov chain approximations of exact tests of population differentiation (Raymond & Rousset 1995a), were calculated from the haplotype data using the program ARLEQUIN (Schneider *et al.*, 2000) to identify patterns of within- and among-population genetic

diversity. The program CONTRIB (Petit *et al.*, 1998) was used to measure gene diversity (Nei 1973) and haplotype richness (Hr), based on rarefaction given a limited sample size. The relationship between genetic differentiation (as measured by F_{ST}) and geographic distance between sampling locations was evaluated for both the historical and modern populations using the method of Slatkin (1993) in the ISOLDE option in Genepop (Raymond & Rousset 1995b).

Results

Haplotype identification

We first aligned the modern and historical sequences obtained here and compared them to published data on the 3' end of the mtDNA control region (Graziano et al. 2005). Unfortunately, for many of the historical individuals we were only able to sequence the shorter (Frag1+Frag2) consensus sequence. These sequences do not include two known variable sites (1,147 and 1,149), or two new variable sites (1,148 and 1,153) discovered in the present study in individuals with complete sequences. This did not affect our ability to uniquely identify some known haplotypes, because mutations at sequenced sites uniquely identify a single haplotype (e.g. MYS02 and MYS06, Table 3). However, a few pairs of published haplotypes are indistinguishable when the nucleotides present at these positions are unknown (e.g. MYS01 and MYS09, Table 3. See also Table 1 in Graziano et al. 2005). In order to compare haplotype frequencies across all individuals, including those with the shorter haplotype sequence, we synonymized known haplotypes that could

not be distinguished based on the shorter sequences and pooled their haplotype frequencies in all further analyses (Tables 3 and 4). For example, haplotypes MYS05 and MYS13 are differentiated by a single base change at a position (1,147) that is not present in the shorter sequences, so all individuals that produced either full MYS05 or MYS13 sequences or short sequences identified as one of these haplotypes were considered to possess haplotype MYS05or13 (Tables 3 and 4). In this case, haplotype MYS13 was not observed in any modern or full-length historical sequences, while haplotype MYS05 was observed in both time periods, so it is likely that all of the shorter sequences identified as MYS05or13 represent haplotype MYS05. Similarly, haplotype MYS09 was observed in only one historical and three modern individuals, and was pooled with haplotype MYS01 for the frequency analysis. This approach resulted in the loss of some information, but allowed direct comparison of observed haplotype frequencies across all individuals, populations, and time periods.

In aligning the mtDNA sequences, a discrepancy was apparent among published mtDNA haplotypes at positions 1036 and 1037 (highlighted in green in Figure 5). In Table 3 of Nielsen et al. (1998), these positions are reported as being G and T, respectively. However, in Nielsen et al. (1994, their sites 81 and 82 in Figure 2) and in the GenBank sequences which reference Nielsen et al. (1998), these sites appear in the order TG. All of our sequences are TG, so we will use this order for these positions. Neither site is polymorphic.

A second discrepancy among published reports and sequences deposited in GenBank concerns a polymorphism at either position 1042 or 1044 (Figure 5). Nielsen et al. 1998 report in Table 2 that position 1042 is variable, and that an “A” in this position

defines haplotype MYS11. Similarly, Nielsen et al. (1994) reported that a site equivalent to 1042 (their site 86) contains a G/A polymorphism and an "A" defines haplotype ST11. However, Table 1 in Graziano et al. (2005) shows the A/G polymorphism at position 1044, as does GenBank sequence accession number AY954383, reported by Graziano et al. (2005) as the same haplotype and cited in GenBank as coming from Nielsen et al. (1998). Bagley and Gall (1998) did not observe either polymorphism within *O. mykiss*. In all of our *O. mykiss* sequences, only the 1042 site was polymorphic, with an "A" observed in 13 individuals from the Hot Creek Hatchery Kamloops strain. All *O. mykiss* sequenced had a "G" at position 1044. Thus it appears the Nielsen et al. 1994 and 1998 reported the correct sequence, but that the incorrect sequence was submitted to GenBank, and was then used by Graziano et al. (2005).

Taking these two inconsistencies into account, the sequence from position 1030-1050 should read: 1029-CAGCGCTGTAAT**G**CGTACACT-1051, where the underlined bases are in the correct order, and the bold "G" is polymorphic, with "A" present only in MYS11.

The Lagunitas Sequence

An anomalous sequence was found in all but three individuals sequenced from the 1909 collection from Lagunitas Creek. This sequence was identified by a novel mutation at site 1,058 that was observed for Frag1 but was not reported in any published *O. mykiss* sequence (Figure 5). In addition, all individuals that possessed this mutation failed to produce any amplification for Frag2. However, we were able to sequence this region using the Frag2f primer paired with the published S-Phe primer. This sequence was

highly divergent from all other known *O. mykiss* sequences, but exactly matched a published sequence (Nielsen et al. 1994) from Coho salmon (*O. kisutch*). Our own sequencing of Coho salmon samples from the same location confirmed the identity. Thus, these samples were misidentified Coho salmon, not steelhead, and were removed from the present analysis.

New Haplotypes

Four haplotypes were observed in our samples that had not previously been described in the literature. Each one was observed at very low frequency (maximum of three individuals) in a single population. Two of the new haplotypes were found in modern population samples, and two were found only in historical samples (Table 3). We temporarily named the four new haplotypes based on the MYS numbering pattern as follows:

New30: Defined by new G>>A mutation at position 1,153, found in three modern individuals from Arroyo Seco 2001, and confirmed by PCR and resequencing. This haplotype is indistinguishable from MYS01or09 in the shorter sequences.

New31: Defined by a new combination of bases at previously known polymorphic sites. Found in at least three modern individuals in Coyote Creek, two in 1998 and one in 2000, and confirmed by PCR and resequencing. One nucleotide difference from MYS03. This haplotype is indistinguishable from MYS03 in shorter haplotype sequences.

New32: Defined by a new combination of bases at previously known polymorphic sites. Found in a single historical individual from Coyote Creek collected in 1909. This haplotype is uniquely identifiable in both shorter and full-length haplotype sequences, but was not confirmed by resequencing the individual.

New33: Defined by a C>T mutation at position 1,148. Found in a single individual from the Nacimiento River, 1909. Confirmed forward and reverse using primers Frag2 and S-phe, but not re-sequenced. This haplotype cannot be distinguished from MYS05or13 in shorter haplotype sequences.

Haplotype Frequencies

In total, we observed 15 MYS haplotypes, including the four newly described haplotypes, ten of which were uniquely identifiable based on the shorter Frag1/Frag2 consensus sequence (Table 3). The observed frequency of all haplotypes was determined for each population in the historical and modern collections (Table 4, Figure 6). Sample sizes were sufficient for all populations to consider each collection separately, with the exception of the three individuals in the historical sample from the main stem of the Pajaro River. These three individuals were combined with the ten sequenced individuals from Llagas Creek, a tributary of the Pajaro River (Tables 1 and 4). On average, the historical population samples had a larger number of haplotypes, higher gene diversity,

and greater haplotype richness than the modern populations, but none of these differences were statistically significant ($p>0.05$ for all paired t-tests, Table 4).

Population Structure

The distribution of haplotype frequencies was highly variable both among populations and between temporal samples from the same location (Figure 6). Pairwise F_{ST} values and exact tests of pairwise differentiation were calculated for all possible combinations of spatial and temporal collections, and these values are summarized in Table 5. The three modern hatchery samples were significantly different from all historical and modern populations except for the modern collection from the South Fork Eel River, which was not significantly differentiated from any hatchery sample (Table 5). Among the wild collections, significant pairwise F_{ST} values were observed among both temporal and spatial comparisons. After Bonferroni correction for multiple tests, 71 out of the 120 possible pairwise F_{ST} values were significantly different from zero with an alpha level of 0.05, with significant values evenly divided between temporal and spatial comparisons overall. Similarly, five out of the eight direct temporal comparisons had significantly different F_{ST} values between the historical and modern collections. Of the three locations with no significant difference between the two temporal samples, two, San Lorenzo and Pajaro-Uvas, were also not significantly different in the exact test of haplotype frequencies, while tests of all other temporal collection pairs were significant (Table 5).

To evaluate the correlation between genetic differentiation and geographic distance, transformed pairwise F_{ST} values were plotted against the geographic distance between each pair of collection locations (Figure 7), and the statistical significance of this correlation was evaluated using a Mantel test. Among the historical collections, we observed a strong, highly significant correlation between genetic and geographic isolation, with > 90% of the variance in genetic differentiation explained by distance (Figure 7a). In contrast, there was no significant correlation between genetic and geographic distance among the modern collections (Figure 7b).

Discussion

When populations exchange migrants in a frequency proportional to their geographic proximity, the genetic divergences among the populations develop a corresponding pattern such that the farther apart two populations are geographically, the greater the genetic isolation between them. This relationship between genetic divergence among populations and the geographic distances separating them was first described by Wright (1943) and termed "isolation by distance" (IBD). Such patterns are typical of natural populations, and have been documented in other studies of salmonids (Castric & Bernachez 2004), including California steelhead (Pearse et al. 2007; Garza et al. 2004), as well as many other taxa (e.g. fish, Pogson et al. 2001; plants, Cruse-Sanders & Hamrick 2004; but see Pearse et al. 2006 for an exception in turtles).

In the present study of historical population genetics of coastal California *O. mykiss*, we did not observe any significant changes in measures of average within-

population genetic diversity. Furthermore, while some populations displayed dramatic shifts in haplotype frequency between the two temporal collections, others remained almost unchanged (Figure 6). However, compared with the historical samples, we observed a dramatic reduction in the association between genetic divergence among populations and the geographic distance separating them in the modern collections. The finding of a highly significant pattern of isolation by distance in the historical samples, but not in the modern samples, is significant because it shows that the distribution of genetic variation in *O. mykiss* among river basins in central California has been significantly altered from its historical condition over the last one hundred years. The extremely strong correlation seen in the historical mtDNA data suggests that a much stronger biological pattern of isolation by distance was present prior to anthropogenic disturbances on *O. mykiss* populations.

The lack of a significant signal of isolation by distance based on the mtDNA data in the modern collections contrasts with the results of Garza et al. (2004), who used 18 highly polymorphic nuclear microsatellite markers and a broad sampling of coastal steelhead populations and found a significant correlation of genetic divergence and geographic distance (r^2=0.193, $p<0.0001$). This correlation remains when only data from the eight collection locations for which we also have historical samples are included (r^2=0.363, $p<0.05$, data from Garza et al. 2004, Clemento et al. unpublished data; Garza and Pearse unpublished data). However, although the correlations based on microsatellite data are statistically significant, the r^2 values are moderate compared with the historical mtDNA (r^2=0.909, $p<0.00001$), consistent with the lack of correlation found with the modern mtDNA data and a hypothesis of reduced isolation by distance in the modern

collections relative to the historical samples. In the modern mtDNA data, the reduced signal of isolation by distance can be attributed to both the presence of neighboring populations that are highly genetically divergent, and geographically distant populations that are genetically similar (Figure 7b). For example, while only one pairwise F_{ST} value was significant among the historical collections from the streams tributary to Monterey Bay, all of which are geographically close together, more than half of the comparable modern comparisons were significantly different from zero (Figure 1 and Table 5)

The above patterns could potentially be influenced by at least three distinct human impacts on *O. mykiss* populations. First, stocking of hatchery strains of *O. mykiss* could lead to changes in the haplotype frequency distribution in one native population, but have much weaker affects on nearby populations, leading to increased divergence between geographically proximate populations. Conversely, use of the same hatchery strain to stock many different river basins could potentially have the effect of genetically homogenizing geographically-distant populations, reducing genetic differentiation at larger scales and countering natural patterns of isolation by distance. A second major effect on patterns of isolation by distance could be caused by the creation of barriers to dispersal between geographically proximate populations. Like differential stocking of neighboring populations, reduced migration among geographically proximate populations would tend to increase their genetic differentiation. This effect could be greatly enhanced if populations also experienced population size reductions, increasing the effects of genetic drift on the distribution of haplotype frequencies. Finally, patterns of migration could be altered through increased long-distance straying of fish among watersheds due to habitat destruction. For example, if fish returning to spawn in their natal stream find

unsuitable habitat (e.g due to logging, agriculture, or development), they may attempt to spawn in another watershed, leading to increased gene flow among both geographically proximate and more distant populations. Thus, any of these potential impacts could contribute to the erosion of the strong, historically-present pattern of isolation by distance, leaving the much weaker signal seen in modern populations.

Acknowledgements

A number of people made this work possible. Primary among them is Edith Martinez, who performed much of the bench work and will be an author on the journal publication to come out of this work. Alicia Abadia, Anthony Clemento and Andy Martinez also contributed significantly. Jerry Finan, curatorial staff at the National Museum of Natural History, Division of Ichthyology, graciously assisted with sampling of the specimens.

References

Avise JC (2000) *Phylogeography* Harvard University Press, Cambridge, MA.

Bagley M, Gall A (1998) Mitochondrial and nuclear DNA sequence variability among populations of rainbow trout (*Onchorhynchus mykiss*). *Molecular Ecology* **7**, 945-961.

Castric V, Bernatchez L (2004) Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology* **13**, 1299-1312.

Cruse-Sanders JM, Hamrick JL (2004) Genetic diversity in harvested and protected populations of wild American ginseng, *Panax quinquefolius* L. (Araliaceae). *American Journal of Botany* **91**, 540-548.

Garza JC, Gilbert-Horvath L,Anderson J, Williams T, Spence B, Fish H (2004) Population structure and history of steelhead trout in California. In: J. Irvine, et al. (eds.), Workshop on Application of Stock Identification in Defining Marine Distribution and Migration of Salmon (Honolulu, HI, USA, November 1-2, 2003). North Pacific Anadromous Fish Commission, Technical Report **5**, 129-131.

Graziano SL, Brown KH, Nielsen JL (2005) Nomenclature of mitochondrial DNA haplotypes for *Oncorhynchus mykiss*. *Transactions of the American Fisheries Society* **134**, 1271-1273.

McCusker M, Parkinson E, Taylor EB (2000) Mitochondrial DNA variation in rainbow trout (*Oncorhynchus mykiss*) across its native range: testing biogeographical hypotheses and their relevance to conservation. *Molecular Ecology* **9**, 2089-2108.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA* **70**, 3321-3323.

Nielsen JL, Gan C, Thomas WK (1994) Differences in genetic diversity for mitochondrial DNA between hatchery and wild populations of *Oncorhynchus*. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 290-297.

Nielsen JL, Fountain MC, Favela JC, Cobble K, Jensen BL (1998) *Oncorhynchus* at the southern extent of their range: a study of mtDNA control-region sequence with special reference to an undescribed subspecies of *O. mykiss* from Mexico. *Environmental Biology of Fishes* **51**, 7-23.

NMFS (2006) Endangered and Threatened Species: Final Listing Determinations for 10 Distinct Population Segments of West Coast Steelhead, pp. 834-862. Federal Register v71, n3.

Pearse DE, Arndt AD, Valenzuela N, et al. (2006) Estimating population structure under nonequilibrium conditions in a conservation context: continent-wide population genetics of the giant Amazon river turtle, *Podocnemis expansa* (Chelonia; Podocnemididae). *Molecular Ecology* **15**, 985-1006.

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

Petit RJ, El Mousadki A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844-855.

Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution* **55**, 131-146.

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

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

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: A software for population genetic analysis. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264-279.

Thrower F, Guthrie CI, Nielsen J, Joyce J (2004) A comparison of genetic variation between an anadromous steelhead, *Oncorhynchus mykiss*, population and seven derived populations sequestered in freshwater for 70 years. *Environmental Biology of Fishes* **69**, 111-125.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Tables and Figures

Table 1: Samples analyzed in the present study from historical and modern time periods. Whenever possible, modern samples from the closest available location were used for comparison.

Watershed	Location	Historical		Modern	
		Year	N	Year	N
Eel River	South Fork	1897	30	2001, 03	24
Lagunitas Creek		1909	10	2001, 03	20
Coyote Creek		1909	15	1998-2000	40
San Lorenzo River		1897or1909	41	2002, 04	32
Pajaro River	Uvas Creek	1909	19	1997, 98	16
	Llagas Creek	1897	13	1997, 98	32
Salinas River	Arroyo Seco	1897, 1909	33	2001	32
	San Antonio River	1909	27	2003	22
	Nacimiento River	1909	24	2003	22
TOTAL:			212		240
American River	Hatchery		-	2002	24
Santa Clara River	Filmore Hatchery		-	2003	23
Hot Creek	Hatchery		-	2005	24

Table 2: Primers used for amplification and sequencing of the mtDNA control region.

Primer Name	Sequence	Reference
P2	TGTTAAACCCCTAAACCAG	Nielsen et al. 1994
S-Phe	GCTTTAGTTAAGCTACG	Nielsen et al. 1994
Frag1f	ACTTTTATGCACTTTAGCATT	This project
Frag1r	TCGAAAAGTTTATTAATGTAT	This project
Frag2f[1]	AAAGTATACATTAATAAACTTTTCG	This project
Frag2r[1]	CGTGGAAATGGCATTGATAA	This project

[1] The Frag2 amplicon is only 79 base pairs long, and proved difficult to sequence. To solve this, an M13 oligo extension was added to the 5' ends of the forward and reverse primers. This resulted in a 33 bp longer PCR product that produced much cleaner sequences, and these m13-tagged primers were used for most Frag2 amplifications.

Table 3: Polymorphic sites observed in the present study, with base numbers as defined by Bagley and Gall (1998) and haplotypes as defined by Graziano et al. (2005). See text for details.

	1,021	1,042	1,050	1,052	1,058	1,086	1,103	1,106	1,109	1,147	1,148	1,153
MYS01	T	G	T	T	A	T	A	A	G	G	C	G
MYS01or09	T	G	T	T	A	T	A	A	G	-	-	-
MYS02	C	G	T	T	A	T	A	A	G	G	C	G
MYS03	T	G	T	T	A	T	A	A	A	G	C	G
MYS04	T	G	T	T	A	C	G	A	G	G	C	G
MYS05	T	G	T	T	A	C	G	C	G	A	C	G
MYS05or13	T	G	T	T	A	C	G	C	G	-	-	-
MYS06	T	G	C	T	A	C	G	C	G	A	C	G
MYS08	T	G	T	T	A	C	A	C	G	A	C	G
MYS08or12	T	G	T	T	A	C	A	C	G	-	-	-
MYS09	T	G	T	T	A	T	A	A	G	A	C	G
MYS10	T	G	T	C	A	T	A	A	A	G	C	G
MYS11	T	A	T	T	A	T	A	A	G	G	C	G
MYS12	T	G	T	T	A	C	A	C	G	G	C	G
NEW30	T	G	T	T	A	T	A	A	G	G	C	A
NEW31	T	G	T	T	A	T	A	A	A	A	C	G
NEW32	T	G	T	T	A	C	A	A	A	G	C	G
NEW33	T	G	T	T	A	C	G	C	G	A	T	G
Coho	T	G	T	T	T	T	A	A	G	-	-	-

Table 4: Observed number of each haplotype in all historical and modern population sample, with the total number of observed haplotypes (# Haps), gene diversity (Hk), and haplotype richness (Hr). Mean values do not include hatchery populations. The population samples from Lagunitas Creek were excluded from the analysis because 7/10 individuals sequenced in the historical sample possessed a Coho-specific mtDNA haplotype.

	MYS 01or09	MYS 02	MYS 03	MYS 04	MYS 05or13	MYS 06	MYS 08or12	MYS 10	MYS 11	New 30	New 31	New 32	New 33	Coho	???	N	# Haps	Hk	Hr
Historical																			
SFE	27	0	2	0	0	0	0	0	0	0	0	0	0	0	1	30	2	0.133	0.70
LAG	1	0	1	0	1	0	0	0	0	0	0	0	0	7	0	10	3		
COY	5	0	8	0	1	0	0	0	0	0	0	1	0	0	0	15	4	0.638	2.733
SLO	11	0	13	0	16	0	0	0	0	0	0	0	0	0	1	41	3	0.676	1.992
PJU	2	0	3	0	5	2	6	0	0	0	0	0	0	0	1	19	5	0.804	3.857
PJL	2	0	0	0	3	2	6	0	0	0	0	0	0	0	0	13	4	0.744	3.00
SAS	2	0	5	0	16	6	2	0	0	0	0	0	0	0	2	33	5	0.684	3.266
SSA	4	0	4	1	7	0	0	0	0	0	0	0	0	0	11	27	4	0.725	2.813
SNC	0	2	4	0	14	0	0	0	0	0	0	0	1	0	3	24	4	0.533	2.474
Total	54	2	40	1	63	10	14	0	0	0	0	1	1	7	19	212			
Mean																22.8	3.78		2.60
Modern																			
SFE	15	5	4	0	0	0	0	0	0	0	0	0	0	0	0	24	3	0.562	1.958
LAG	8	1	6	0	4	0	0	0	0	0	0	0	0	0	1	20	4		
COY	12	0	15	0	4	0	0	0	0	0	3	0	0	0	1	35	4	0.561	1.816
SLO	15	0	7	0	8	0	2	0	0	0	0	0	0	0	0	32	4	0.688	2.633
PJU	3	0	7	0	3	1	2	0	0	0	0	0	0	0	0	16	5	0.767	3.784
PJL	0	12	9	0	11	0	0	0	0	0	0	0	0	0	0	32	3	0.683	2.00
SAS	8	0	0	0	18	0	0	0	0	3	0	0	0	0	0	29	3	0.605	1.963
SSA	0	14	0	0	8	0	0	0	0	0	0	0	0	0	0	22	2	0.485	1.00
SNC	0	16	1	0	0	0	4	1	0	0	0	0	0	0	0	22	4	0.455	2.165
ARH	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	1	0	0.00
HOT	10	0	0	0	0	0	0	0	13	0	0	0	0	0	0	23	2	0.514	1.00
FIL	19	0	0	0	0	0	5	0	0	0	0	0	0	0	0	24	2	0.344	0.989
Total	114	48	49	0	56	1	13	1	13	3	3	0	0	0	2	303			
Mean																26.5	3.44		2.16

Table 5: Estimates of population differentiation for all pairwise combinations. F_{ST} Values shown below diagonal, p-values for exact test of haplotype frequency differentiation above. Bold values are not significant at alpha of 0.05 after Bonferroni correction for multiple tests. Direct temporal comparisons are underlined.

	hSFE	**hCOY**	**hSLO**	**hPJU**	**hPJL**	**hSAS**	**hSNC**	**hSSA**	**mSFE**	**mCOY**	**mSLO**	**mPJU**	**mPJL**	**mSAS**	**mSNC**	**mSSA**	**mARH**	**mHOT**	**mFIL**
hSFE	*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hCOY	0.509	*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hSLO	0.560	0.322	*	**0.000**	0.000	**0.002**	**0.001**	**0.483**	0.000	**0.009**	**0.087**	**0.054**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hPJU	0.577	0.263	**0.085**	*	**0.658**	**0.112**	**0.002**	**0.055**	0.000	0.000	**0.008**	**0.361**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hPJL	0.645	0.300	**0.186**	**-0.035**	*	**0.010**	0.000	**0.005**	0.000	0.000	**0.002**	**0.031**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hSAS	0.581	0.312	**0.060**	**0.049**	**0.124**	*	**0.033**	**0.083**	0.000	0.000	0.000	**0.047**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
hSNC	0.686	0.391	**0.094**	**0.149**	0.256	**0.024**	*	**0.040**	0.000	0.000	0.000	**0.002**	**0.023**	0.000	0.000	**0.001**	0.000	0.000	0.000
hSSA	0.624	0.297	**-0.037**	**0.055**	**0.147**	**0.020**	**0.051**	*	0.000	**0.003**	**0.257**	**0.247**	**0.003**	**0.006**	0.000	0.000	0.000	0.000	0.000
mSFE	0.663	0.405	0.196	0.250	0.290	0.329	0.422	**0.204**	*	0.000	**0.002**	**0.000**	0.000	0.000	0.000	0.000	**0.003**	0.000	**0.000**
mCOY	0.616	0.404	**0.094**	0.201	0.317	0.254	0.330	**0.130**	**0.209**	*	**0.008**	**0.067**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mSLO	0.575	0.322	**0.026**	**0.095**	**0.153**	**0.144**	**0.223**	**0.018**	**0.067**	**0.111**	*	**0.125**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mPJU	0.599	0.288	**0.017**	**0.025**	**0.122**	**0.097**	**0.185**	**0.021**	**0.187**	**0.008**	**0.048**	*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mPJL	0.575	0.320	0.119	0.137	0.229	**0.120**	**0.103**	**0.098**	0.285	0.223	0.196	**0.111**	*	0.000	0.000	**0.008**	0.000	0.000	0.000
mSAS	0.623	0.357	**0.093**	**0.144**	**0.200**	**0.071**	**0.088**	**0.042**	0.307	0.327	**0.129**	0.200	0.201	*	0.000	0.000	0.000	0.000	0.000
mSNC	0.722	0.463	0.411	0.333	0.365	0.410	0.465	0.414	0.395	0.470	0.406	0.374	0.196	0.463	*	0.000	0.000	0.000	0.000
mSSA	0.709	0.432	0.309	0.290	0.346	0.274	**0.270**	0.289	0.396	0.451	0.347	0.339	**0.078**	0.310	**0.126**	*	0.000	0.000	0.000
mARH	0.927	0.741	0.473	0.590	0.657	0.604	0.748	0.580	**0.251**	0.538	0.318	0.591	0.625	0.561	0.781	0.766	*	0.000	**0.052**
mHOT	0.693	0.431	0.315	0.315	0.341	0.378	0.477	0.314	**0.261**	0.376	0.240	0.313	0.395	0.369	0.515	0.500	0.551	*	0.000
mFIL	0.770	0.530	0.328	0.335	0.338	0.439	0.565	0.354	**0.104**	0.386	0.154	0.350	0.473	0.398	0.587	0.587	**0.174**	0.347	*

Figure 1: Map of watersheds and O. mykiss DPSs included in the present study. Uvas and Llagas Creeks are tributaries of the Pajaro River, and the Arroyo Seco, San Antonio, and Nacimiento Rivers are all tributaries of the Salinas River.

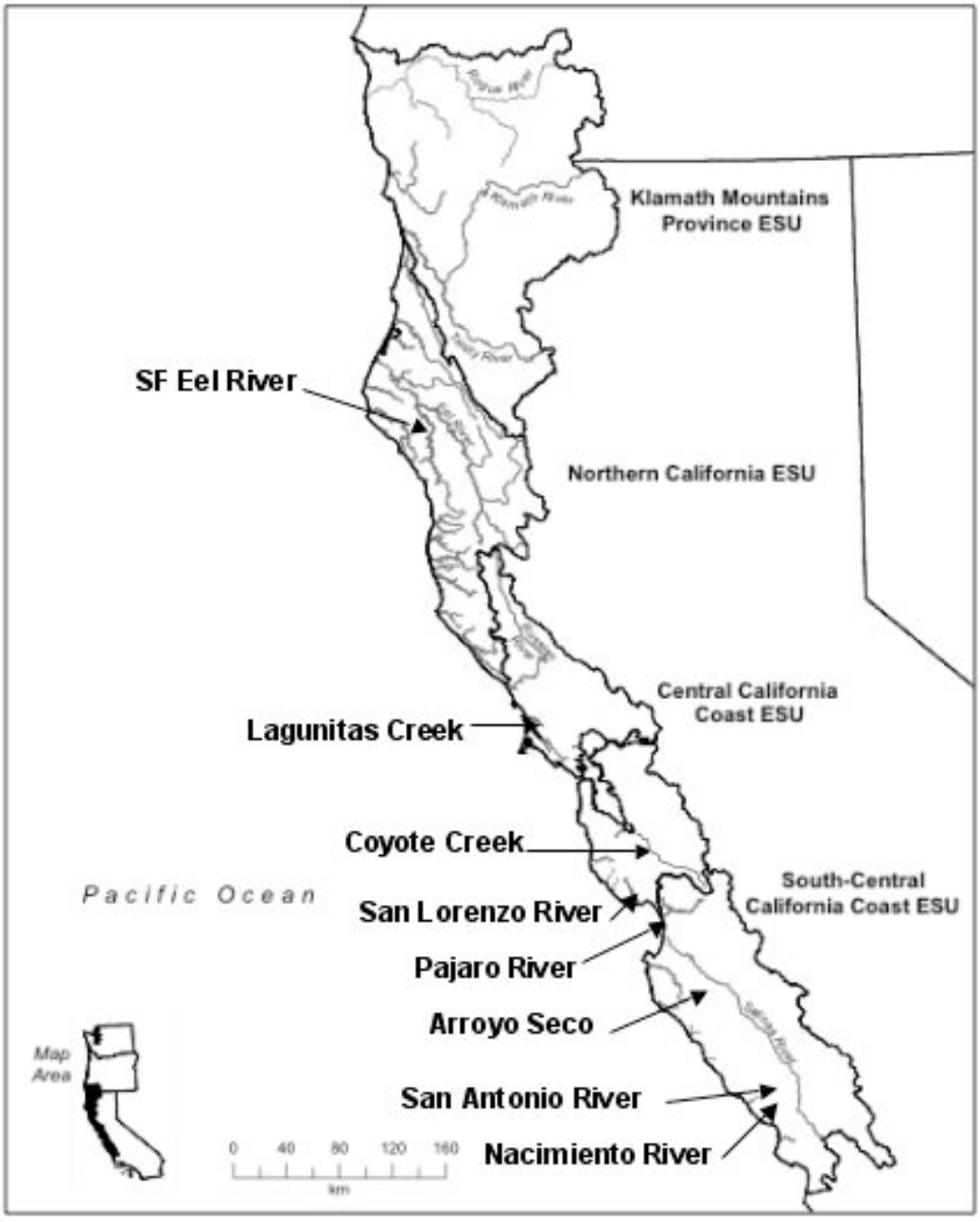
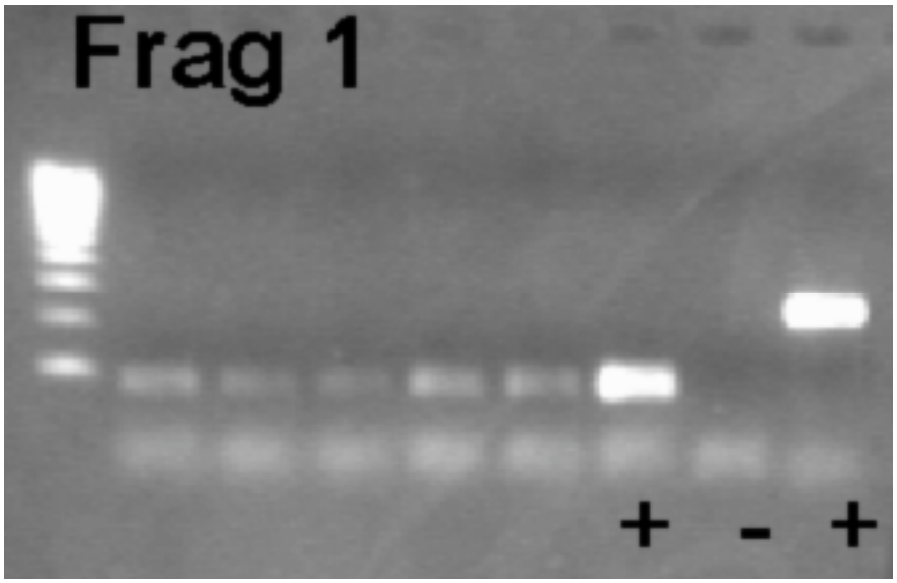


Figure 3: Agarose gel images of mtDNA amplification products.

a) Agarose gel image of Frag 1 PCR amplifications, including positive (modern sample) and negative (water extraction) controls. The far right lane shows a positive amplification using the primers P2 and S-phe, which produce a larger product.



b) Agarose gel showing amplification of Frag 2 using primers with the m13 nucleotide sequence added to the 5' end. This lengthens the final PCR product and produced cleaner sequence. The smaller second band present in many samples, including the negative control, is most likely a primer dimer. When present, affected samples required a gel purification step prior to sequencing the larger target band.

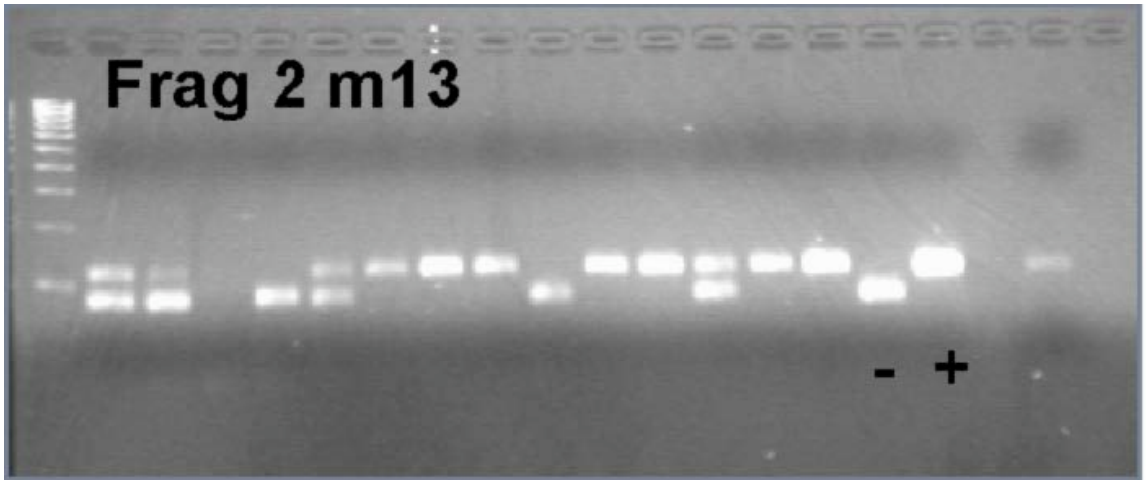


Figure 4: Image of Sequencher alignment window showing overlap of all mtDNA control region primer combinations used. Forward and reverse sequences for each fragment are indicated by “f” and “r” notation.

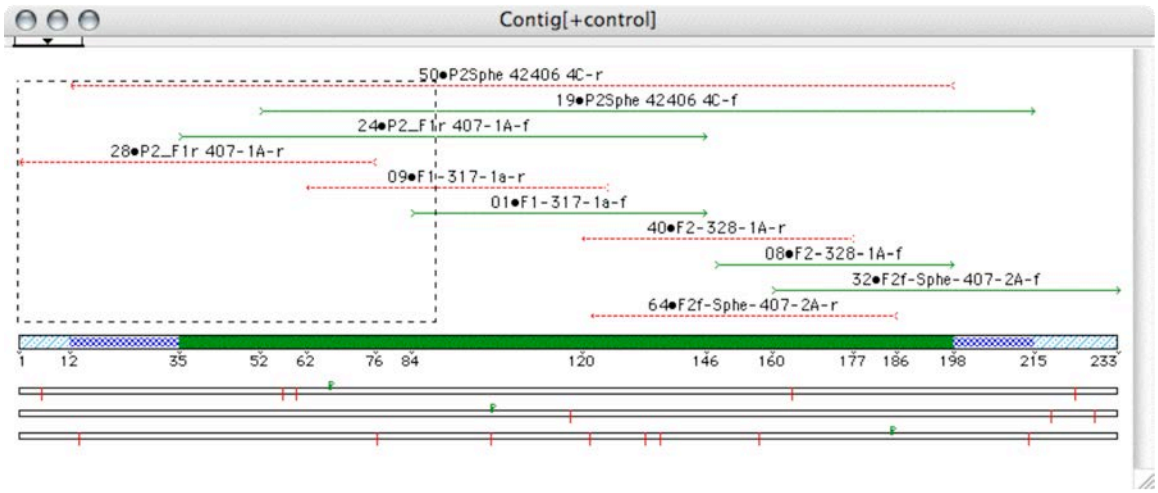


Figure 5: Sequence electropherograms of a) haplotype NEW30, with G>A mutation at postion 192 circled, and b) Typical sequence for other hapolotypes in this region. This is the second to last base in the full P2/S-Phe sequence, and is in the phenylalanine tRNA.

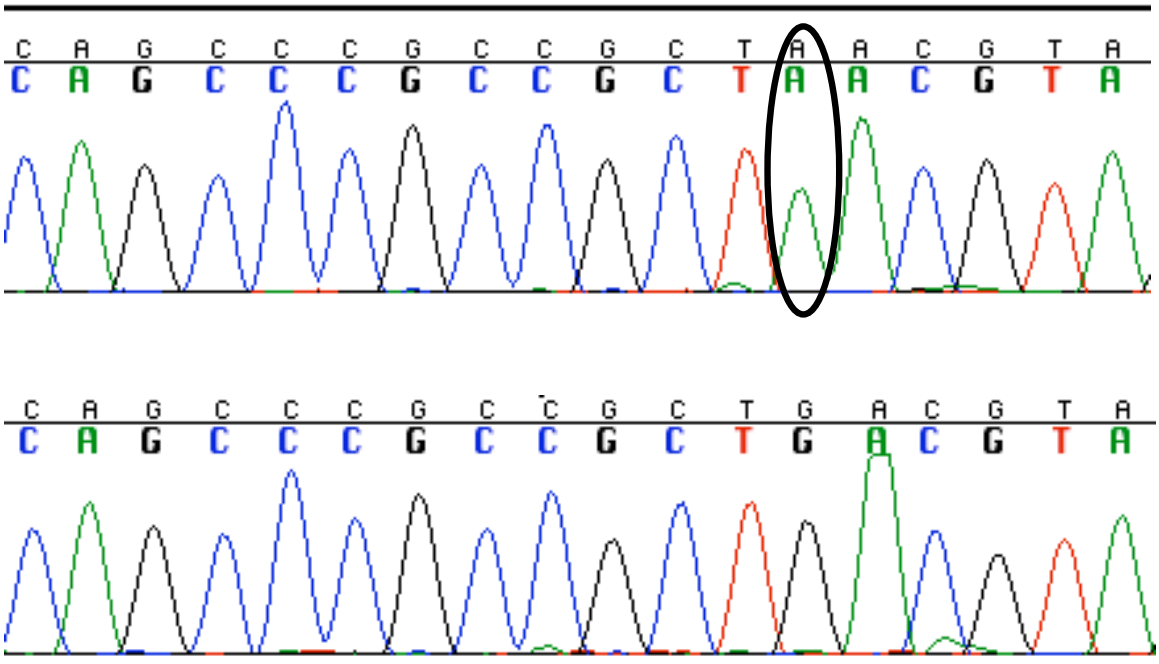


Figure 2:. Features of the *O. mykiss* MtDNA control region polymorphic right domain. Blue highlighted bases indicate sites misreported in the literature. Polymorphic sites indicated in bold and color coded: observed in this study (black), new polymorphic sites (pink), characteristic of *O. m. aguabonita* (orange), misreported in the literature (blue), characteristic of *O. kisutch* (red). See text for details.

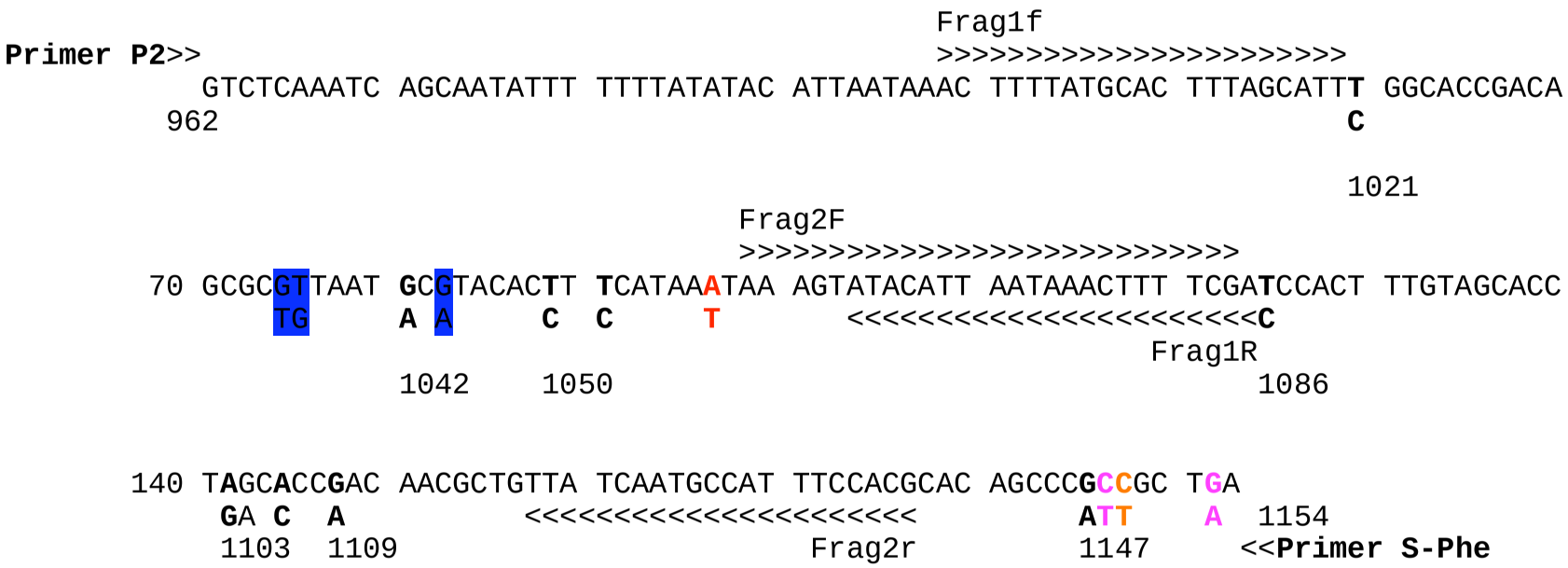


Figure 6: Pie charts showing frequencies of observed haplotypes in historical and modern collections for each location in the present study.

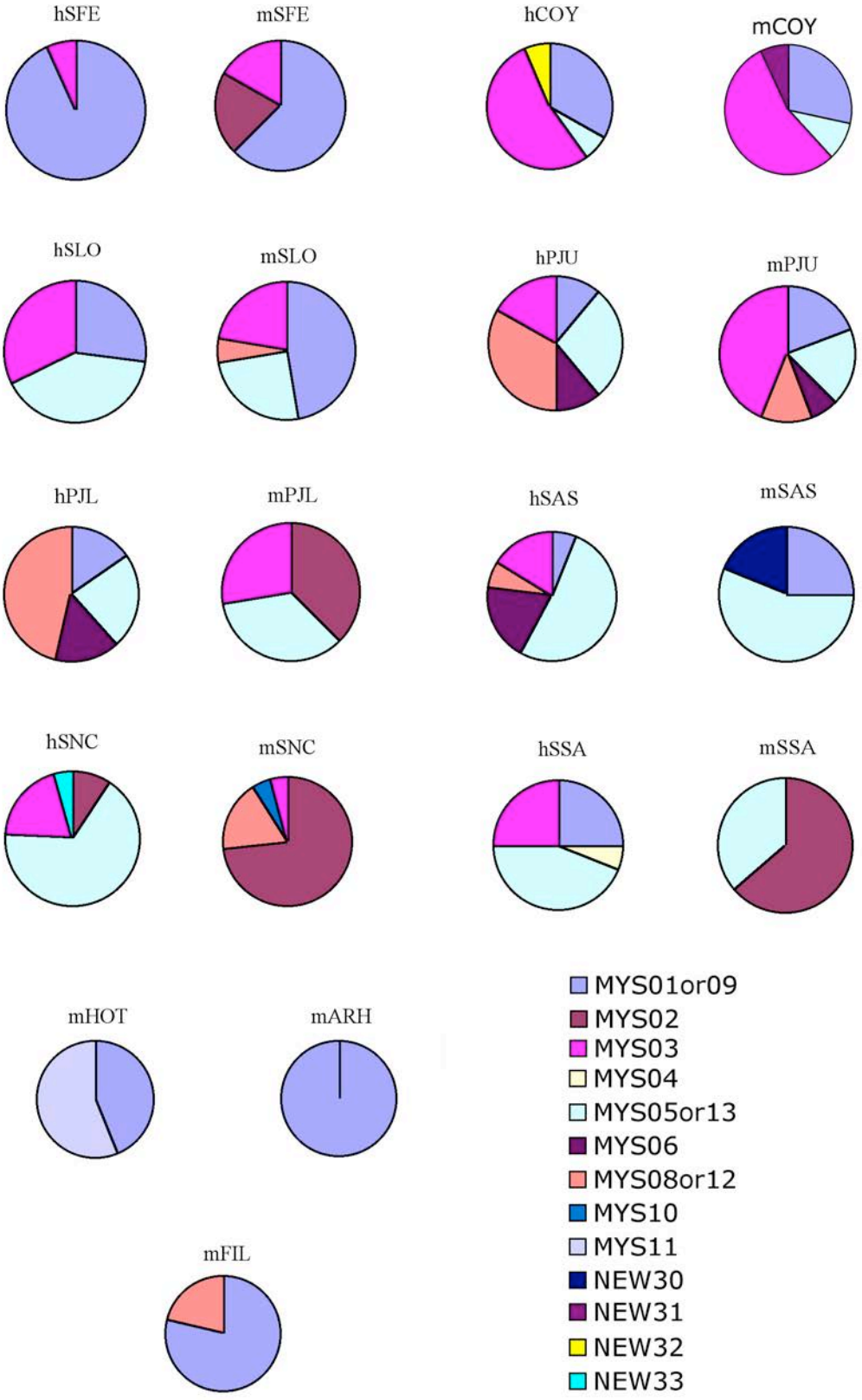
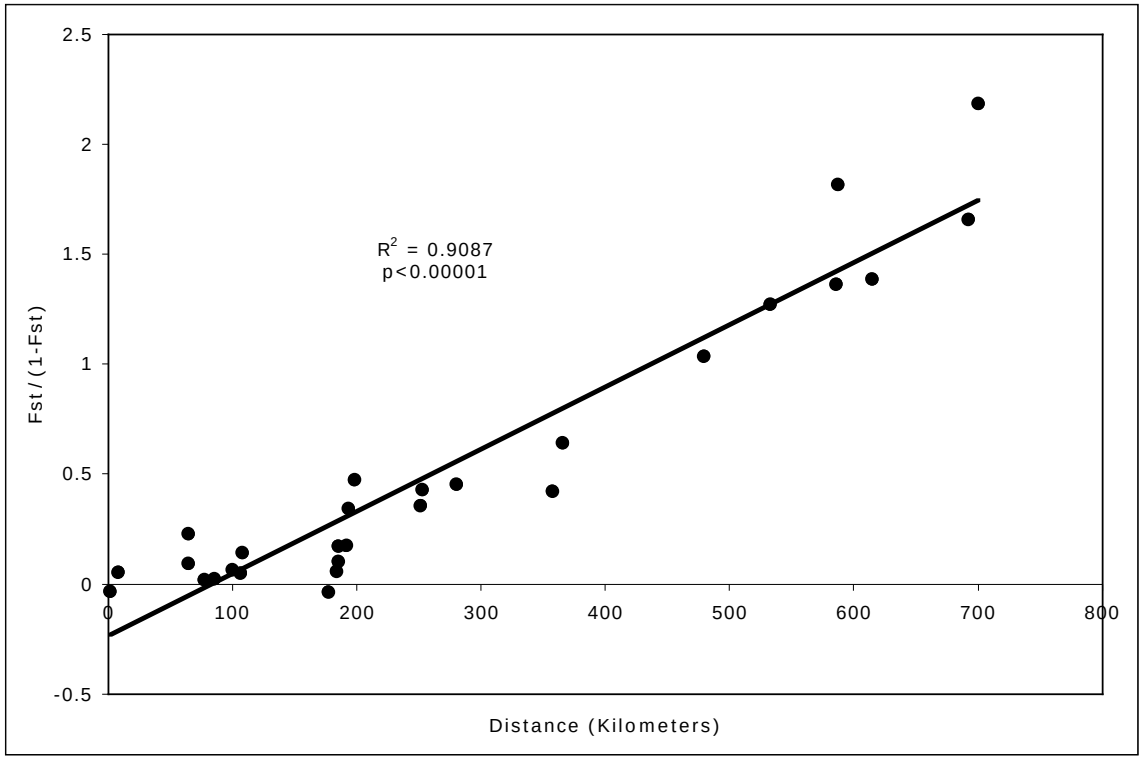


Figure 7. Correlation between genetic divergence and geographic distance for a) historical samples and b) modern samples from the same locations.

a)



b)