1999

Hatchery influence on Pajaro River Steelhead analyzed with microsatellite DNA

Dagmar Ruth Sundermeyer
San Jose State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

HATCHERY INFLUENCE ON PAJARO RIVER STEELHEAD ANALYZED WITH MICROSATELLITE DNA

A Thesis

Presented to

The Faculty of the Department of Biological Sciences

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Dagmar Ruth Sundermeyer

December 1999

APPROVED FOR THE DEPARTMENT OF BIOLOGICAL SCIENCES

Dr. Jerry Smith

Dr. Chris Brinegar

Dr. David Matthes

APPROVED FOR THE UNIVERSITY

ABSTRACT

HATCHERY INFLUENCE ON PAJARO RIVER STEELHEAD ANALYZED WITH MICROSATELLITE DNA

by D. Ruth Sundermeyer

Steelhead (*Oncorhynchus mykiss*) have been listed in California as threatened or endangered under the Endangered Species Act. Five microsatellite loci were examined in fourteen populations of *O. mykiss* in the Pajaro River watershed at the northern boundary of the south central coast Evolutionarily Significant Unit (ESU). Most of these populations were closely related to two populations from the San Lorenzo River from the adjacent central coast ESU, the source of most of the recent hatchery-reared steelhead planted in the Pajaro system. Some resident trout populations above barriers to migration were genetically different from steelhead in the San Lorenzo River, and probably represent relic native populations. Phylogenetic relationships given by Cavalli-Sforza's chord distance and Nei's D_A distance, which are based on random genetic drift, were more reasonable than those given by Goldstein's $(\delta\mu)^2$ and Slatkin's Rho, which assume stepwise mutation, based upon what is known about the ecology of these watersheds.

ACKNOWLEDGMENTS

Dr. Jerry Smith, Dr. Chris Brinegar and Dr. David Matthes reviewed this manuscript and provided valuable comments. Funding was provided by the Santa Clara Valley Water District, the Graduate Research Fellowship Program of the Department of Biological Sciences of San Jose State University, the Golden West Women's Flyfishing Association, and Dr. Jerry Smith. Phylogenetic results in this study were put into ecological context based upon Dr. Jerry Smith's two decades of experience in these two watersheds. Electrofishing assistance was provided by Dr. Jerry Smith. Dr. Chris Brinegar provided molecular technical and scientific assistance, as well as laboratory equipment and space. Laboratory protocols were based upon protocols developed by Dr. Jennifer Nielsen and Monique Fountain. Dr. Nielsen provided laboratory assistance as well. Dr. Robert Fowler provided electrophoresis equipment. Andrea Henke and Ingrid Udranszky provided valuable technical assistance. Figures were created by Angelique Benicio.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Introduction

As wild fish are locally adapted to their natal streams, hatchery fish introduced from other watersheds could make a population less fit. Hatchery-produced steelhead (*Oncorhynchus mykiss*) comprise a significant portion of naturally spawning populations in many areas, but insufficient data exist on both the relative abundance of hatchery and natural stocks and the ecological interactions between hatchery and native fish. However, hatcheries are considered a factor in the decline of many populations of Pacific salmonids (Nehlson 1997). Recovery of a species can be defined as the restoration of viable, self-sustaining populations in their natural habitats (Waples 1994). Information on genetic structure within and between adjacent populations is crucial to recovery decisions.

The National Marine Fisheries Service (NMFS) has listed as threatened or endangered, ten Evolutionarily Significant Units (ESUs) of west coast steelhead trout in Washington, Oregon, Idaho and California. Nonanadromous "resident rainbow trout" above permanent barriers to migration were not listed. NMFS policy provides that a population will be considered a distinct population segment, and hence a "species" under the Endangered Species Act, if it represents an ESU of the species. A population must satisfy two criteria to be considered an ESU (Waples 1991):

1) It must be substantially reproductively isolated from other conspecific population units, and
2) It must represent an important component of the evolutionary legacy of the species.

This study used microsatellite analysis to examine the genetic structure of populations in a key central California coastal watershed of an ESU of *O. mykiss*. The Pajaro River is the northern boundary of the south central California coast ESU (ESU 10) (Fig. 1), and preliminary studies indicated it may be an important biogeographic boundary (Nielsen, *et al.* 1994).

Several factors make the Pajaro watershed an important system to study:

1) Hatchery origin steelhead, especially from the San Lorenzo River system, may have altered the genetic composition of fish in some or all tributaries of the Pajaro.
2) Several reservoirs or other barriers on the tributaries created boundaries between resident and anadromous forms of *O. mykiss* which may or may not share a common gene pool. Fish above barriers may represent native relic stocks, while hatchery plants may have significantly affected fish below the dams.
3) Boundaries of the ESUs have been based on studies with minimal sampling in watersheds.

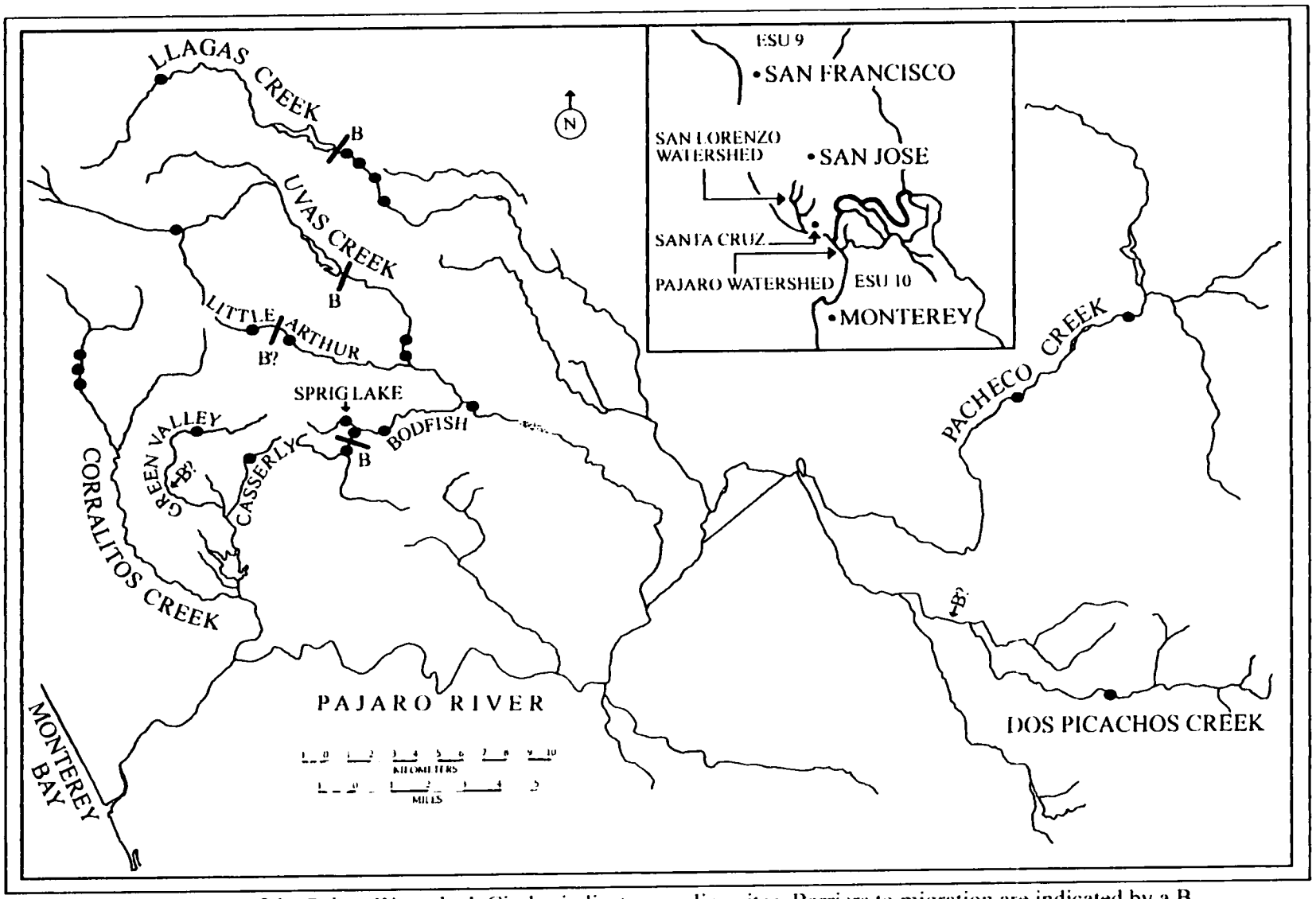


Fig. 1. Map of the Pajaro Watershed. Circles indicate sampling sites. Barriers to migration are indicated by a B.

Intensive sampling can elucidate the genetic structure within this watershed, and its relationship to that of the San Lorenzo River to the north.

This study investigated the genetic composition of *O. mykiss* in tributaries of the Pajaro River with analysis of microsatellite DNA (tandem repeat sequences). It compared Pajaro fish to fish from the San Lorenzo River in the adjacent ESU to the north to assess the extent to which tributaries may have been affected by hatchery plants and which tributaries appear to contain relatively native stocks. This study also assessed the accuracy of several different measures of genetic distances commonly used for microsatellites (Appendix A and B) by comparing phylogenetic relationships generated by these distances to what is known about the ecology of the Pajaro River watershed.

Genetic effects of hatchery stocking should be interpreted in the context of ecological data to aid in management and restoration decisions. The majority of anadromous steelhead spawn in these tributaries between January and April and some steelhead spawn in more than one year. The juvenile fish (young-of-the-year) rear in the stream. The fish either migrate to the ocean after the first year (as yearlings) or spend an additional year in the stream and emigrate at age 2+ (Shapovalov and Taft 1954). Since many tributaries of the Pajaro River have major water diversions, in and out migration is difficult, especially in drought years. In contrast, most streams in the central coast ESU immediately to the north tend to have smaller coastal plains and easier up and down migration between the ocean and freshwater spawning and rearing habitat (Jerry Smith, San Jose State

University, San Jose, CA, personal communication). Reservoirs have been built in the Pajaro watershed within the last century, creating barriers to migration and creating resident (nonanadromous) trout populations.

Out-of-basin hatchery fish have been planted in the Pajaro watershed. Mad River steelhead were planted by the California Department of Fish and Game (CDFG) in Uvas Creek in 1975-1981 (National Marine Fisheries Service 1996). CDFG has regularly planted central coast steelhead smolts, mostly of San Lorenzo River origin, in various tributaries of the Pajaro since the early 1980's. Sprig Lake, a seasonal impoundment on Bodfish Creek, is stocked with catchable (hatchery raised for fishing) rainbow trout (the nonanadromous form of *O. mykiss*) for children to fish for, and there is a possibility they may survive and interbreed with the steelhead in the creek.

Many microsatellite DNA loci have been developed for salmonids. The polymerase chain reaction (PCR) readily amplifies microsatellite DNA from minute amounts of tissue, and this enables the use of nonlethal sampling methods on endangered and threatened fish. Since variation in these VNTRs (variable number tandem repeats) is great and microsatellite DNA is thought to be selectively neutral, microsatellite analysis is a powerful tool for analyzing gene flow among populations. Genetic differences accumulate between populations that are strongly reproductively isolated from each other, as gene flow is substantially reduced. Microsatellite mutation and random genetic drift cause differences in allele sizes and allelic frequencies among these populations.

These differences are used to measure the degree of reproductive isolation between populations.

Several measures of genetic distance are commonly used to analyze differences in allelic frequencies of microsatellites between populations (Appendix B). Goldstein's $(\delta\mu)^2$ (Goldstein *et al.* 1995) and Slatkin's R_{ST} (Slatkin 1995) are based on the stepwise mutation model (SMM) of Ohta and Kimura (1973). These distances assume a mutation adds or subtracts a repeat unit in an allele, and the difference in the number of repeats gives information about the time between the new allele and ancestral alleles. There are also classical distance measures that rely on multidimensional geometry rather than a particular evolutionary model, such as Cavalli-Sforza's chord distance D_C (Cavalli-Sforza and Edwards 1967) and Nei's D_A (Nei *et al.* 1983). These two distances assume that drift is the main cause of differentiation, not the rapid stepwise mutation assumed in the SMM.

Materials and methods

Sample collection

Fish were captured with a backpack electroshocker (Smith Root Type 7) in tributaries of the Pajaro River from fourteen sites above and below barriers to migration (Fig. 1). Two sites were sampled in the San Lorenzo River watershed, one in the mainstem at Boulder Creek and one in a tributary, Zayante Creek. One to four locations were sampled in each "site". Most of these sites were represented by a single location, but five of the sites

contained two to four locations because of low fish densities. A nonintrusive (small, partial) caudal fin clip was taken from thirty to sixty fish at each site (only twelve from Sprig Lake) from young-of-the-year, yearling and age 2+ fish. Fin clips were put in a small piece of folded Whatman chromatography paper and air dried overnight. Samples were stored in labeled coin envelopes. Most fish were caught over the summer of 1997, but additional fish were sampled in the summer of 1998 from Pacheco and Llagas creeks because an insufficient number of fish were found in 1997. Since steelhead populations have overlapping age classes, sampling over two years should not affect genetic results. Thirty fish were randomly selected from each site for analysis and additional fish were selected when some of the DNA didn't amplify.

PCR and electrophoresis

For DNA extraction, a small piece of fin (approximately 2 mm square) was incubated in 200 ul of 5% Chelex (Biorad, dissolved in HPLC grade water) at 65^{O}C for 20 minutes, vortexed, incubated at 95^{O}C for 10 minutes, vortexed, and microcentrifuged at full speed for 3 minutes. The DNA was stored at -80^{O}C. PCR protocols were adapted from those used by Jennifer Nielsen (personal communication) at Hopkins Marine Laboratory in Pacific Grove, CA. A PCR master mix was made for each locus, and 1 µl of DNA was added to 6.15 µl of master mix for each reaction. Final concentrations of reagents in each PCR reaction were 0.87 mM dNTPs (Promega), 0.87 µM each primer (Operon), 3 mM $MgCl_2$, (4 mM $MgCl_2$ for Omy77), 0.15 units *Taq* DNA polymerase (Promega). Each

PCR reaction contained the following buffer concentrations: 0.0145 M NH_4S0_4, 0.00857 M β-mercaptoethanol, and 0.0586 M TRIS-HCl, with pH adjusted to 8.6.

DNA was amplified in a Model PTC-100 thermocycler (MJ Research) for five di-nucleotide microsatellite loci (Table 1). The thermocycler was programmed to denature at 94°C for 1 minute, followed by 30 or 32 cycles of denature at 94°C for 40 seconds, anneal at 50 or 52°C for 1 minute, and extend at 72°C for 1 minute.

Table 1. Microsatellite loci primer sequences, annealing temperatures and number of PCR cycles, and source of the primers.

Locus	primer pair (5'-3')	Anneal/cycles	Source
Ssa14	CCT TTT GAC AGA TTT AGG ATT TC CAA ACC AAA CAT ACC TAA AGC C	52°C/32	McConnell *et al.* 1995
Ssa289	CTT TAC AAA TAG ACA GAC T TCA TAC AGT CAC TAT CAT C	52°C/32	McConnell *et al.* 1995
Omy77	CGT TCT CTA CTG AGT CAT GGG TCT TTA AGG CTT CAC TGC A	52°C/30	Morris *et al.* 1996
Omy207	ACC CTA GTC ATT CAG TCA GG GAT CAC TGT GAT AGA CAT CG	50°C/30	O'Connell[1]
Oneμ2	GGT GCC AAG GTT CAG TTT ATG TT CAG GAA TTT ACA GGA CCC AGG TT	52°C/30	Scribner *et al.* 1996

[1]Communicated to Jennifer Nielsen, Hopkins Marine Laboratory, Pacific Grove, CA.

PCR products were run on a Hoefer Pokerface II electrophoresis system at 65 watts, on 6% denaturing polyacrylamide gels (USB). Gels were silver stained using the protocol

outlined in Promega's *Technical Manual, DNA Silver Staining System* (1993). Thirty-five fish selected from among all populations were run on a Perkin Elmer ABI automatic sequencer, model 672, at Hopkins Marine Laboratory (Pacific Grove, CA) to be used as size standards and positive controls on gels at San Jose State University. Alleles were scored with these size standards. Detailed protocols are included in Appendix C.

Statistical analysis

Allele frequencies were calculated in GENEPOP Version 3.1b (Raymond and Rousset, 1995). Tests for deviations from Hardy-Weinberg equilibrium and genotypic linkage disequilibrium were calculated with the exact tests of GENEPOP. A Holm's sequential Bonferroni correction for $\alpha = 0.05$ was used to correct for errors due to multiple testing when evaluating departure from Hardy-Weinberg equilibrium and genotypic linkage disequilibrium (Lessios 1992). Average heterozygosities for each population were calculated in DISPAN of Tatsuya Ota (http://med-humgen14.bu.edu/ota/program/dispan/dispan.html).

Cavalli-Sforza's chord distance D_C (Cavalli-Sforza and Edwards 1967) was calculated in the GENDIST program of PHYLIP Version 3.5c of Joseph Felsenstein (http://evolution.genetics.washington.edu/phylip.html). The CONTML program of PHYLIP (Felsenstein 1981), using global rearrangements and jumbling the data input ten times, generated a phylogenetic tree. CONTML estimates phylogenies by the restricted maximum likelihood method based on the Brownian motion model and is based on the

model of Cavalli-Sforza and Edwards (Felsenstein 1981). Nei's genetic distance D_A was generated with 2000 bootstrap replications, and a neighbor-joining tree was generated with NJBAFD2 of Naoko Takezaki (http://www.bio.psu.edu/people/faculty/Nei/lab/programs.html).

Distances and trees based on the stepwise mutation model (SMM) were generated for comparison to trees using D_C and D_A. $(\delta\mu)^2$ with 2000 bootstrap replications and a neighbor-joining tree were generated in NJBAFD2. Slatkin's Rho, an unbiased estimator of Slatkin's R_{ST} was calculated in RSTCALC 2.2 (Goodman 1997), then a neighbor-joining tree was generated in PHYLIP. RSTCALC globally standardized the data set so alleles were expressed in terms of standard deviations from the global mean rather than repeat unit number. This way, loci with low variances and a high degree of differentiation do not contribute less to the final value of R_{ST} than loci with high variances. This program bootstraps to obtain 95% confidence intervals for R_{ST}. All phylogenetic trees were displayed with TREEVIEW (Page 1996). A detailed description of how the software was used is in Appendix D.

Cavalli-Sforza and Edward's chord distance and Nei's distance D_A were used to describe the main clustering patterns in the Pajaro system because the assumptions of these distances probably fit this data better than distances based on a stepwise mutation model, Goldstein's $(\delta\mu)^2$ and Slatkin's R_{ST}. Many of the genetic differences between populations of steelhead and rainbow trout have probably occurred because of recent barriers to

migration and artificial hatchery stocking patterns rather than from a mutational process. A distance that assumes genetic drift may be more appropriate than a distance based on the magnitude of differences in allele sizes. Takezaki and Nei's (1996) computer simulation showed that Nei's D_A and Cavalli-Sforza's chord distance D_C generally performed better in topology construction under many different conditions than any of the other distances tested. Goldstein's $(\delta\mu)^2$ performed better for branch length estimation, and possibly for finding microsatellite differences for more distantly related populations, such as the distance between closely related species. Both Goldstein's $(\delta\mu)^2$ and Slatkin's average square have a higher coefficient of variation, which may help explain a poorer performance in topology reconstruction for closely related populations. Goldstein too has found that D_C and D_A reconstruct closely-related phylogenies better than the distances based on an SMM (Goldstein and Pollock 1997). However, the validity of any measure of genetic distance should be considered in context of what is known about the ecology and history of the area being studied.

Results

Description of loci

The number of alleles at these loci ranged from nine in Ssa289 to forty-four in Oneμ2. Average population heterozygosities were high, ranging from 0.7062 to 0.8623. Expected and observed values for heterozygosity and homozygosity at each locus were generally consistent, although there were many low allelic frequencies at the more

polymorphic loci, Omy77, Omy207 and Oneμ2 (Table 2). Allelic frequencies for each locus in each population are given in Appendix E.

P values for the Hardy-Weinberg equilibrium test were compared with Holm's sequential Bonferroni corrected values for $\alpha = 0.05$ (Table 3) (Lessios 1992). Twenty-three of eighty comparisons had uncorrected P values of less than 0.05, but because of the large number of comparisons, corrections were used and only eleven corrected values were significant. In lower Llagas Creek, four out of five loci showed significant departures from Hardy-Weinberg equilibrium (P values of 0.0000). Only seven out of the other fifteen sites had a departure in one locus and eight of the fifteen had none. With the data for lower Llagas excluded, three of the loci had only one or zero out of fifteen significant P values, but Omy77 had two, and Omy207 had three other significant departures from Hardy-Weinberg equilibrium.

Thirty-six of one hundred sixty tests showed P values of less than 0.05 for genotypic disequilibrium, or apparent linkage between loci. However, with Holm's sequential Bonferroni corrected values for $\alpha = 0.05$, to correct for the large number of comparisons, only fourteen of the comparisons were significant (Table 4). Lower Llagas had five of the fourteen significant corrected P values. Seven of the ten comparisons between pairs of loci showed at least one site with a significant departure for linkage disequilibrium. Three sites were found in Ssa289/Oneμ2 and in Omy77/Oneμ2.

Table 2. Average heterozygosity for Pajaro and San Lorenzo River sites. Expected and observed homozygosity and heterozygosity are given for each locus and each population. N is the sample size for each site.

		Ssa14				Ssa289				Omy77				Omy207				Oneμ2				
	Average	Homozyg.		Heterozyg.		Homozyg.		Heterozyg.		Homozyg.		Heterozyg.		Homozyg.		Heterozyg.		Homozyg.		Heterozyg.		
	heterozygosity	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	N
Green Valley	0.7509	7.4	6	22.6	24	9.1	8	20.9	22	9.5	12	20.5	18	5.9	7	24.1	23	5.4	5	24.6	25	30
Casserly	0.8107	6.7	6	23.3	24	9.2	5	20.8	25	6.2	12	23.8	18	3.0	2	27.0	28	3.4	2	26.6	28	30
Corralitos	0.8605	5.1	4	24.9	26	7.8	7	23.2	24	2.6	2	27.4	28	3.8	10	27.2	21	2.0	3	28.0	27	30-31
Llagas up	0.7062	13.9	11	21.1	24	11.5	10	23.5	25	13.4	19	19.6	14	6.2	13	28.8	22	5.7	5	29.3	30	33-35
Llagas low	0.6451	8.9	8	21.1	22	18.7	16	11.3	14	11.6	28	18.4	2	5.0	5	25.0	25	9.0	15	21.0	15	30
Uvas up	0.8473	6.5	8	22.5	21	6.1	7	22.9	22	3.9	10	26.1	20	2.9	9	26.1	20	3.0	5	27.0	25	29-30
Uvas low	0.8623	5.7	6	24.3	24	7.1	10	22.9	20	3.3	7	25.7	22	2.6	9	27.4	21	1.9	2	28.1	28	30
Bodfish up	0.8167	5.2	5	24.8	25	8.6	6	21.4	24	4.5	11	24.5	18	5.5	9	24.5	21	3.4	3	26.6	27	29-30
Bodfish low	0.8507	3.8	4	28.2	28	10.5	5	21.5	27	4.3	9	27.7	23	2.3	6	29.7	26	3.0	6	29.0	26	32
Little Arthur Redwood	0.7398	10.3	7	20.7	24	16.2	16	14.8	15	4.7	3	25.3	27	4.9	11	26.1	20	4.0	1	27.0	30	30-31
Little Arthur MtM	0.7450	18.9	19	12.1	12	8.3	8	22.7	23	5.1	4	24.9	26	4.0	4	27.0	27	3.0	1	28.0	30	30-31
Pacheco	0.8234	5.6	8	24.4	22	7.8	7	22.2	23	7.2	11	22.8	19	3.0	4	27.0	26	2.8	1	27.2	29	30
Dos Picachos	0.7633	11.7	10	21.3	23	8.4	13	23.6	19	10.0	13	22.0	19	4.0	5	28.0	27	3.8	5	26.2	25	30-33
Sprig Lake	0.8184	1.0	0	8.0	9	1.7	1	9.3	10	3.9	6	8.1	6	1.7	1	9.3	10	1.6	1	8.4	9	9-12
San Lorenzo Zayante	0.7619	13.3	13	16.7	17	8.8	9	21.2	21	6.5	16	23.5	14	4.0	3	26.0	27	3.2	1	26.8	29	30
San Lorenzo Boulder	0.8496	7.5	7	22.5	23	6.9	7	23.1	23	3.9	4	25.1	25	2.4	4	27.6	26	1.7	2	28.3	28	29-30

Table 3. Estimated exact P values and standard errors for Hardy-Weinberg equilibirium tests for each microsatellite locus at Pajaro and San Lorenzo river sites. Values significant at Holm's sequential Bonferroni corrected levels of $\alpha = 0.05$ are boxed.

	Ssa14		Ssa289		Omy77		Omy207		Oneμ2	
	P value	S.E.	P value	S.E.	P value	S.E.	P value	S.E.	P value	S.E.
Green Valley	0.5506	0.0092	0.2018	0.0116	0.1847		0.1090	0.0123	0.1770	0.0167
Casserly	0.0225	0.0064	**0.0028**	0.0015	0.0288	0.0080	0.1882	0.0220	0.9609	0.0092
Corralitos	0.3792	0.0245	0.7729	0.0131	0.3675	0.0193	**0.0000**	0.0000	0.3154	0.0389
Llagas up	0.0824	0.0093	0.8208	0.0067	0.0212	0.0060	**0.0000**	0.0000	0.3254	0.0198
Llagas low	**0.0000**	0.0000	0.4186		**0.0000**		**0.0000**	0.0000	**0.0000**	0.0000
Uvas up	0.0219	0.0054	0.7617	0.0105	0.0040	0.0023	0.0634	0.0124	0.2604	0.0336
Uvas low	0.0088	0.0041	0.1411	0.0103	0.1516	0.0168	**0.0013**	0.0013	0.0760	0.0229
Bodfish up	0.9614	0.0061	0.4296	0.0091	0.0127	0.0039	0.0599	0.0134	0.0130	0.0088
Bodfish low	0.7041	0.0250	0.1124	0.0113	0.1168	0.0167	0.0084	0.0060	0.0770	0.0180
Little Arthur Redwood	0.2945	0.0181	0.4716	0.0171	0.1572	0.0236	**0.0000**	0.0000	0.7289	0.0279
Little Arthur MtM	0.3061	0.0230	0.0878	0.0085	0.1139	0.0168	0.2263	0.0275	0.0854	0.0203
Pacheco	0.2116	0.0199	0.5887	0.0120	**0.0016**	0.0016	0.0154	0.0059	0.2462	0.0289
Dos Picachos	0.9589	0.0077	0.1687	0.0091	0.1464	0.0153	0.5480	0.0171	0.1527	0.0230
Sprig Lake	0.9157	0.0090	0.8740	0.0057	0.5485		0.1099	0.0120	0.0251	0.0049
San Lorenzo Zayante	0.1270	0.0205	0.9741	0.0025	**0.0000**	0.0000	0.7033	0.0304	0.0948	0.0139
San Lorenzo Boulder	0.3706	0.0244	0.1541	0.0106	0.3970	0.0218	0.0291	0.0100	0.6944	0.0370

Table 4. Genotypic disequilibrium P values and standard errors (S.E.) between pairs of loci. Values significant at Holm's corrected $\alpha = 0.05$ are boxed.

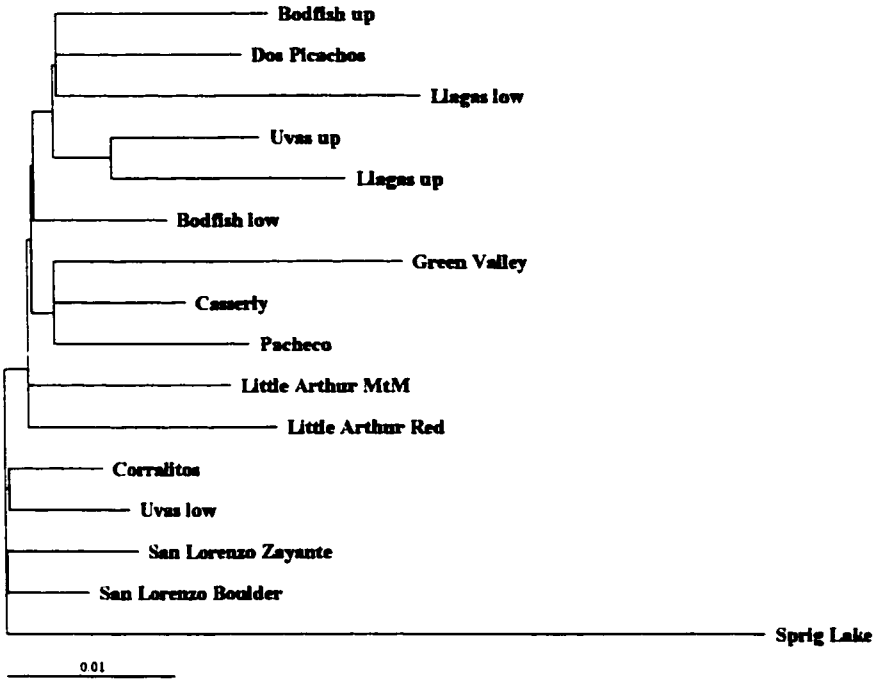
	Ssa14	Ssa289	S.E.	Ssa14	Omy77	S.E.	Ssa289	Omy77	S.E.	Ssa14	Omy207	S.E.	Ssa289	Omy207	S.E.
Green Valley	0.4702	0.0311		0.0711	0.0103		0.5252	0.0286		0.8255	0.0240		0.8760	0.0219	
Casserly	0.1757	0.0224		0.2622	0.0355		0.9129	0.0173		0.2504	0.0353		0.5497	0.0397	
Corralitos	0.1095	0.0227		0.0962	0.0255		1.0000	0.0000		1.0000	0.0000		0.8866	0.0254	
Llagas up	0.1066	0.0177		0.6175	0.0289		0.9091	0.0147		0.0599	0.0186		0.0046	0.0035	
Llagas low	0.0334	0.0036		0.0000	0.0000		0.1328	0.0070		0.0001	0.0001		0.0541	0.0046	
Uvas up	0.1376	0.0281		0.8218	0.0324		0.1520	0.0296		1.0000	0.0000		1.0000	0.0000	
Uvas low	0.2436	0.0316		0.1328	0.0300		0.5476	0.0450		1.0000	0.0000		0.0945	0.0231	
Bodfish up	0.2637	0.0301		0.1421	0.0326		0.0051	0.0033		1.0000	0.0000		0.3305	0.0311	
Bodfish low	0.4814	0.0432		0.1494	0.0343		0.5836	0.0375		1.0000	0.0000		0.3271	0.0399	
Little Arthur Red	0.0782	0.0113		0.2042	0.0219		0.0926	0.0162		0.1865	0.0255		0.4858	0.0297	
Little Arthur Mt M	0.2010	0.0223		0.0000	0.0000		0.0368	0.0135		0.0037	0.0037		0.2120	0.0314	
Pacheco	0.0981	0.0200		0.0084	0.0068		0.0512	0.0140		0.0920	0.0225		0.4324	0.0390	
Dos Picachos	0.3085	0.0380		0.7233	0.0368		0.9286	0.0165		0.4807	0.0451		0.7752	0.0354	
Sprig Lake	1.0000	0.0000		1.0000	0.0000		1.0000	0.0000		1.0000	0.0000		1.0000	0.0000	
San Lorenzo Zayante	0.7437	0.0232		0.2029	0.0299		0.0468	0.0153		0.2713	0.0365		0.3122	0.0373	
San Lorenzo Boulder	0.7726	0.0282		0.2722	0.0348		1.0000	0.0000		0.2298	0.0339		0.5561	0.0415	

	Omy77	Omy207	S.E.	Ssa14	Onep2	S.E.	Ssa289	Onep2	S.E.	Omy77	Onep2	S.E.	Omy207	Onep2	S.E.
Green Valley	0.9982	0.0014		0.0866	0.0195		0.4155	0.0381		0.5524	0.0276		0.6051	0.0387	
Casserly	0.6386	0.0433		0.0307	0.0129		0.0232	0.0130		1.0000	0.0000		0.0000	0.0000	
Corralitos	1.0000	0.0000		1.0000	0.0000		1.0000	0.0000		1.0000	0.0000		1.0000	0.0000	
Llagas up	0.1711	0.0291		0.2764	0.0331		0.0367	0.0140		0.3135	0.0360		0.1753	0.0308	
Llagas low	0.0044	0.0018		0.0000	0.0000		0.0007	0.0003		0.0048	0.0022		0.0015	0.0007	
Uvas up	1.0000	0.0000		0.4853	0.0480		0.3270	0.0422		0.0897	0.0274		0.0783	0.0257	
Uvas low	0.0881	0.0276		0.0618	0.0190		0.0000	0.0000		1.0000	0.0000		0.0276	0.0136	
Bodfish up	0.2141	0.0376		0.2068	0.0370		0.0352	0.0128		0.5895	0.0443		0.8383	0.0330	
Bodfish low	0.0000	0.0000		1.0000	0.0000		0.4709	0.0432		0.3763	0.0466		1.0000	0.0000	
Little Arthur Red	0.2183	0.0326		0.0793	0.0162		0.4639	0.0306		0.0000	0.0000		0.0366	0.0151	
Little Arthur Mt M	0.0688	0.0237		0.0000	0.0000		0.0335	0.0124		0.0000	0.0000		0.0582	0.0206	
Pacheco	0.6010	0.0382		0.0400	0.0172		0.0904	0.0232		0.0003	0.0003		0.0173	0.0110	
Dos Picachos	1.0000	0.0000		1.0000	0.0000		1.0000	0.0000		0.0317	0.0155		1.0000	0.0000	
Sprig Lake	0.4950	0.0253		1.0000	0.0000		1.0000	0.0000		0.1034	0.0103		0.0671	0.0096	
San Lorenzo Zayante	0.4114	0.0448		0.0099	0.0095		0.0022	0.0022		0.4126	0.0435		0.0334	0.0173	
San Lorenzo Boulder	0.1141	0.0292		1.0000	0.0000		1.0000	0.0000		1.0000	0.0000		1.0000	0.0000	

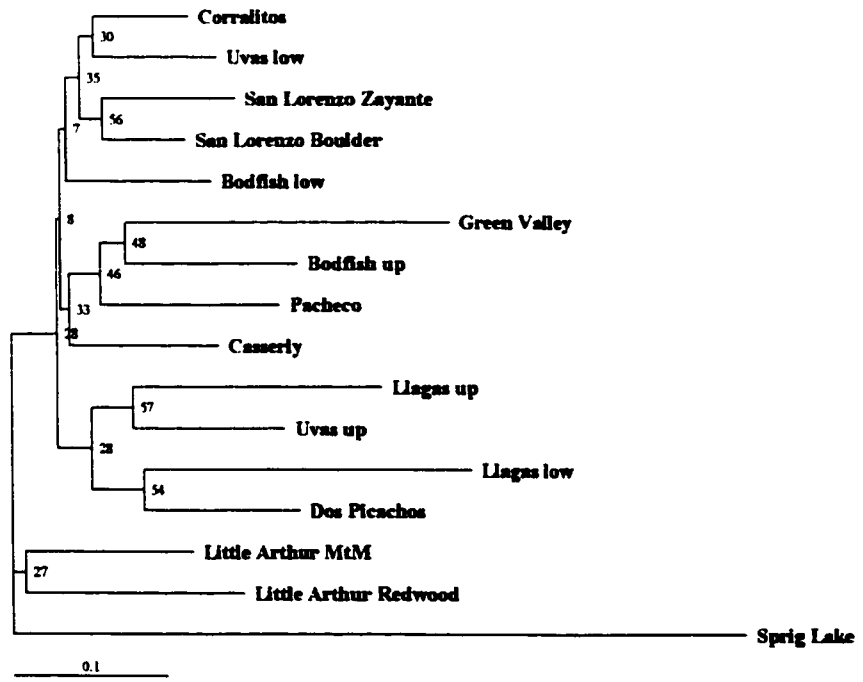
Distances and topologies

In the PHYLIP CONTML tree, by far the tightest clustering of populations is between Corralitos Creek, lower Uvas Creek and the two San Lorenzo sites, San Lorenzo River at Boulder Creek ("Boulder"), and Zayante Creek (Fig. 2). Branch lengths and confidence intervals are given in Appendix F. Cavalli-Sforza chord distances among these sites range from 0.0247 to 0.0358 (Table 5). As a point of reference, the chord distance between the two San Lorenzo sites, Boulder and Zayante Creek, is 0.0249. Boulder was more closely related to Corralitos Creek (0.0247) than it was to Zayante Creek, and nearly as close to lower Uvas Creek (0.0257). Lower Bodfish Creek and Casserly Creek, the most accessible tributaries of Uvas Creek and Corralitos Creek, also show relatively low chord distances to the two San Lorenzo river sites (0.0265 - 0.0439) and to each other.

Overall, there were no other tight geographical clusters (Fig.2). While other sites in the Pajaro watershed show clustering patterns in CONTML, branch lengths and chord distances are generally much longer than found in the San Lorenzo, Corralitos and lower Uvas group. In fact, branch lengths appeared to be more informative measures of genetic differentiation than the clustering patterns were. Upper Uvas and upper Llagas creeks for example, cluster together in CONTML but have a chord distance of 0.0458 between them. The clustering reflects the dissimilarity of the two sites to other sampled sites. Green Valley and lower Llagas have much longer branch lengths than other sites and are distant from even the sites with which they cluster, reflecting their genetic



(a)



(b)

Fig. 2. Unrooted phylogenetic trees for Pajaro and San Lorenzo River watersheds based upon data from five microsatellite loci. a) CONTML tree. b) D_A in NJBAFD. Numbers at branch points are bootstrap values.

Table 5. Cavalli-Sforza chord distances among Pajaro and San Lorenzo sites, based upon 5 microsatellite loci.

	Green Valley	Casserly	Corralitos	Llagas up	Llagas low	Uvas up	Uvas low	Bodfish up	Bodfish low	Lit. Arthur Red	Lit. Arthur MtM	Pacheco	Dos Picachos	Sprig Lake	San Lorenzo Z
Casserly	0.0588														
Corralitos	0.0619	0.0359													
Llagas up	0.0791	0.0580	0.0497												
Llagas low	0.0760	0.0652	0.0622	0.0758											
Uvas up	0.0731	0.0429	0.0353	0.0458	0.0717										
Uvas low	0.0729	0.0320	0.0252	0.0596	0.0706	0.0464									
Bodfish up	0.0567	0.0481	0.0428	0.0634	0.0746	0.0565	0.0586								
Bodfish low	0.0620	0.0347	0.0275	0.0562	0.0662	0.0442	0.0339	0.0372							
Lit. Arthur Red	0.0889	0.0439	0.0437	0.0704	0.0882	0.0604	0.0509	0.0562	0.0447						
Lit. Arthur MtM	0.0688	0.0451	0.0448	0.0695	0.0655	0.0560	0.0447	0.0495	0.0436	0.0443					
Pacheco	0.0603	0.0398	0.0448	0.0534	0.0781	0.0541	0.0437	0.0433	0.0456	0.0576	0.0511				
Dos Picachos	0.0712	0.0441	0.0404	0.0541	0.0550	0.0425	0.0462	0.0508	0.0418	0.0557	0.0479	0.0638			
Sprig Lake	0.1321	0.1133	0.1046	0.1282	0.1271	0.1049	0.0986	0.1196	0.1099	0.1072	0.1065	0.1060	0.1272		
San Lorenzo Z	0.0572	0.0352	0.0309	0.0545	0.0679	0.0463	0.0358	0.0488	0.0439	0.0534	0.0416	0.0427	0.0500	0.1053	
San Lorenzo B	0.0593	0.0349	0.0247	0.0554	0.0603	0.0442	0.0257	0.0407	0.0265	0.0476	0.0391	0.0395	0.0435	0.1097	0.0249

distinctiveness. For lower Llagas, the distinctiveness was due to very low allelic diversity.

The catchable trout population from Sprig Lake is the most widely separated in CONTML, and has chord distances from other sites ranging from 0.0986 to 0.1321. It also has a few alleles that are not present in other sites. The most common allele for Omy77 in this population is absent in other populations

Nei's distance D_A from NJBAFD2 produced generally the same tree as PHYLIP CONTML (Fig. 2). The tightest clustering was among Corralitos, lower Uvas creeks and the two San Lorenzo sites. Sprig Lake was again widely separated from all other sites. Green Valley and lower Llagas creeks again had the longest branch lengths. Both the CONTML and NJBAFD2 D_A trees cluster the two Little Arthur sites out from the others.

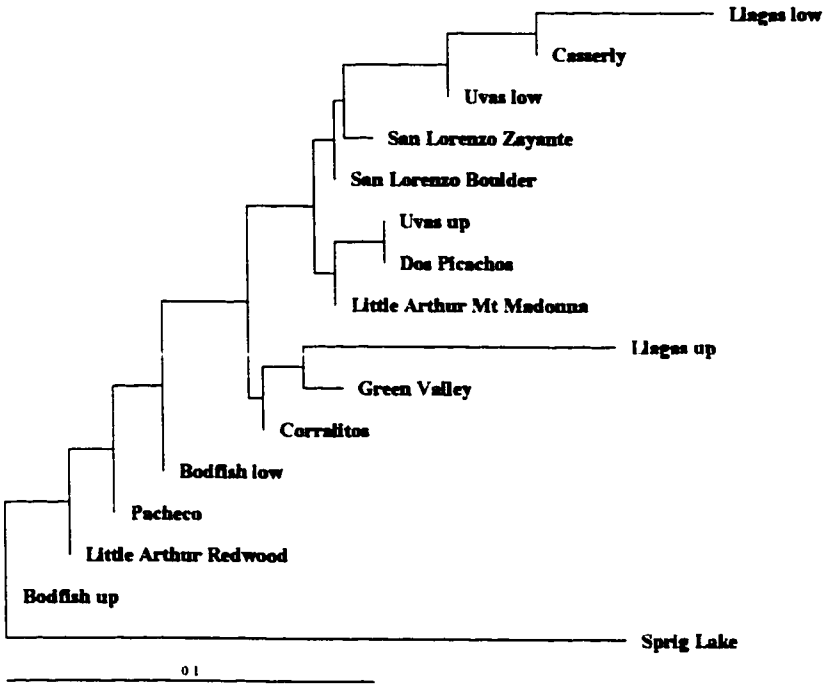
These two trees differ slightly. Casserly and lower Bodfish creeks show low distances from the San Lorenzo/Corralitos/lower Uvas cluster in both trees, but lower Bodfish creek clusters with them in NJBAFD2. In CONTML, upper Bodfish Creek clusters with lower Llagas and Dos Picachos creeks, while in NJBAFD2 upper Bodfish moves to the Green Valley/Pacheco/Casserly cluster. However, in both trees branch lengths are long. When Cavalli-Sforza's chord distance calculated in PHYLIP's GENDIST was used to generate a neighbor-joining tree, it also generated a tree with a clustering pattern like D_A in NJBAFD2.

The neighbor-joining trees generated with genetic distance measures based on the stepwise mutation model, Slatkin's Rho and Goldstein's $(\delta\mu)^2$, differ substantially from the CONTML tree and from each other in several ways (Fig. 3, Table 6). There was no clear San Lorenzo/Corralitos/lower Uvas cluster as there was in CONTML, and the relationships among these streams were very different. Slatkin's Rho from RstCalc shows the tightest relationship of Boulder with Little Arthur Creek at Mt. Madonna Road, Dos Picachos Creek and lower and upper Uvas Creek (Rho values from -0.0048 to 0.0113). Corralitos doesn't cluster with this group, but has a distance of -0.0033 from Boulder. Green Valley and lower Bodfish creeks are moderately related to Boulder (Rho values 0.0329 and 0.271) while Casserly, upper Bodfish and lower Llagas are much more distant (0.0419 to 0.0525). Goldstein's $(\delta\mu)^2$ in the neighbor-joining program of NJBAFD2 separates San Lorenzo's Zayante Creek from the San Lorenzo at Boulder and from all the other sites. Goldstein's $(\delta\mu)^2$ and Rho differ also in the relationships of upper Bodfish and lower Llagas.

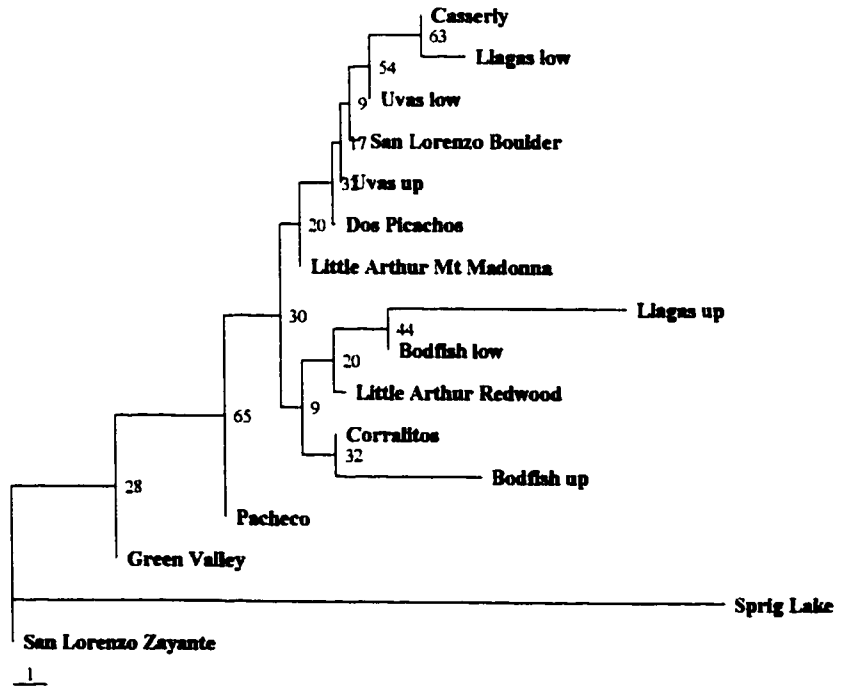
Discussion

Hardy-Weinberg equilibrium and linkage disequilibrium

Eleven out of eighty tests for departures from Hardy-Weinberg equilibrium were found to be significant in this study. Four of the eleven significant P values were found in lower Llagas Creek. These results could be caused by a population bottleneck, despite the fact that five sites were sampled over two years. Up and especially downstream passage on



(a)



(b)

Fig. 3. Neighbor-joining trees for Pajaro and San Lorenzo River watersheds based upon five microsatellite loci. a) Rho from RSTCALC. b) $(\delta\mu)^2$ from NJBAFD2. These are unrooted trees.

Table 6. Slatkin's Rho among Pajaro and San Lorenzo sites, based upon five microsatellite loci. Where Rho is negative, the within population variance in allele size is bigger than the among population variance, therefore there is no differentiation and the populations are essentially panmictic.

	Green Valley	Casserly	Corralitos	Llagas up	Llagas low	Uvas up	Uvas low	Bodfish up	Bodfish low	Lit. Arthur Red	Lit. Arthur MtM	Pacheco	Dos Picachos	Sprig Lake	San Lorenzo Z
Casserly	0.1324														
Corralitos	0.0163	0.0542													
Llagas up	0.0959	0.2333	0.0599												
Llagas low	0.1758	0.0320	0.1002	0.3408											
Uvas up	0.0733	0.0648	0.0155	0.1243	0.1297										
Uvas low	0.0694	-0.0062	0.0178	0.1297	0.0249	0.0171									
Bodfish up	0.0858	0.0505	0.0226	0.1524	0.1206	0.0860	0.0457								
Bodfish low	0.0303	0.0550	-0.0009	0.0617	0.1117	0.0583	0.0303	0.0057							
Lit. Arthur Red	0.0732	0.0631	0.0382	0.1292	0.1454	0.1120	0.0481	0.0208	-0.0020						
Lit. Arthur MtM	0.0165	0.0513	-0.0055	0.0824	0.0915	-0.0013	0.0093	0.0589	0.0221	0.0617					
Pacheco	0.0506	0.0681	0.0100	0.1351	0.1723	0.0790	0.0414	0.0106	-0.0084	0.0003	0.0329				
Dos Picachos	0.0522	0.0525	0.0092	0.1064	0.1035	-0.0133	0.0077	0.0805	0.0457	0.0915	-0.0096	0.0633			
Sprig Lake	0.2999	0.1977	0.2246	0.4365	0.3688	0.2696	0.1961	0.1464	0.1794	0.1582	0.2424	0.1818	0.2613		
San Lorenzo Z	0.0305	0.0494	0.0131	0.1866	0.0698	0.0353	0.0226	0.0535	0.0398	0.0784	0.0061	0.0477	0.0241	0.2384	
San Lorenzo B	0.0319	0.0419	-0.0033	0.1038	0.0525	0.0113	0.0064	0.0503	0.0271	0.0744	-0.0048	0.0505	0.0033	0.2733	0.0043

Llagas Creek is difficult because water spills only briefly from the reservoir in the winter, and because early spring dry back occurs in the long percolating valley floor portion of the stream. In the first year of sampling, only seven fish were caught at one site, even though three sites of useable habitat were sampled. In the second year, three sites had to be sampled to find the twenty-three additional analyzed fish. Allelic variation in this site was relatively low (Appendix E). The sampled fish could have represented the offspring of as few as one pair of spawning adults in each year.

Most of the remaining departures from Hardy-Weinberg equilibrium were found in Omy77 and Omy207, which had six of the seven departures at sites other than Llagas Creek. There are several possible explanations. One possibility is error in scoring alleles as the complex banding patterns in these two loci made them difficult to score. At all loci, for samples that were difficult to score, a new PCR was done and run next to size standards until alleles could be unambiguously scored. Omy77 and Ssa289 were scored independently by two people, and any fish for which alleles could not be agreed upon were run again next to size standards until alleles could be consistently and unambiguously scored. Another explanation could be that either homoplasy or null alleles skewed the data. Sequencing alleles was beyond the scope of the present project. Another possibility could be that the sample size was too small for such polymorphic loci with numerous rare alleles. However, it is interesting to note that for Oneμ2, the most polymorphic locus with forty-four alleles, only lower Llagas had a significant departure from Hardy-Weinberg equilibrium.

There were only a limited number of departures from linkage disequilibrium. It is possible some loci were linked. However, departures could have resulted from some of the factors discussed for Hardy-Weinberg equilibrium, such as error in scoring alleles or sample sizes that were too small for such polymorphic loci.

Two COMTML trees were generated with Omy77 and Omy207 removed from the data set to see if either locus changed the tree significantly (Appendix G). While the trees based on four loci differed slightly from the ones based upon all five loci, they did not change the general conclusions discussed below.

Distances and topologies

It is most useful to evaluate the genetic differences found in these watersheds in context of local ecological information. It is known that hatchery-reared smolts from the San Lorenzo River have been planted regularly in Uvas Creek and occasionally in Corralitos Creek. During drought years smolts have also been planted in the lower Pajaro River, which reduces the chance of smolt stranding but probably also increases widespread straying of adults within the Pajaro River system.

All of the trees clearly separate Sprig Lake out from all other populations. This small, seasonally dammed lake is stocked with catchable rainbow trout for "children-only" fishing. This population contained some alleles not found in the other sites. The most common Omy77 allele in Sprig Lake is absent from other sites and may serve as a marker

to distinguish these fish. Most of the catchable trout are caught (Jerry Smith, personal communication), but in any case there is no evidence of gene flow to steelhead in Bodfish or Uvas creeks.

In CONTML, the tight clustering of Corralitos and lower Uvas with both San Lorenzo River at Boulder and Zayante Creek suggest that these streams are closely related. The chord distances (D_C) between Boulder Creek especially and Corralitos and lower Uvas are small (0.0247 to 0.0358), about the same as the distances between the two San Lorenzo sites (0.0249) (Table 7). San Lorenzo River hatchery fish have been stocked in Corralitos and Uvas creeks in the past. In addition, the two streams are the most accessible of the Pajaro River tributaries and Corralitos Creek is the lowermost tributary.

Casserly and lower Bodfish creeks are the lowest and most accessible tributaries of Corralitos and lower Uvas creeks respectively. They are nearly as closely related to San Lorenzo River at Boulder (0.0265 - 0.0349) as are Corralitos and lower Uvas creeks, and appear to have heavy hatchery influence. Pacheco and Casserly creeks are grouped together in CONTML, but there is a moderate distance between them (0.0398). Pacheco Creek also is relatively close to Boulder. Although Pacheco Creek is in the upper watershed where straying is less likely, some of the fish sampled may have been hatchery fish planted as fingerlings in 1996 (Jerry Smith, personal communication).

Table 7. Genetic distances of Pajaro sites from San Lorenzo sites arranged in ascending order. The smallest distances represent the sites most closely related to San Lorenzo hatchery fish.

Stream	D_C from Boulder	Stream	D_C from Zayante	Stream	Rho from Boulder	Stream	Rho from Zayante
Corralitos	0.0247	Boulder	0.0249	Little Arthur MtM	-0.0048	Boulder	0.0043
Zayante	0.0249	Corralitos	0.0309	Corralitos	-0.0033	Little Arthur MtM	0.0061
Uvas low	0.0257	Casserly	0.0352	Dos Picachos	0.0033	Corralitos	0.0131
Bodfish low	0.0265	Uvas low	0.0358	Zayante	0.0043	Uvas low	0.0226
Casserly	0.0349	Little Arthur MtM	0.0416	Uvas low	0.0064	Dos Picachos	0.0241
Little Arthur MtM	0.0391	Pacheco	0.0427	Uvas up	0.0113	Green Valley	0.0305
Pacheco	0.0395	Bodfish low	0.0439	Bodfish low	0.0271	Uvas up	0.0353
Bodfish up	0.0407	Uvas up	0.0463	Green Valley	0.0319	Bodfish low	0.0398
Dos Picachos	0.0435	Bodfish up	0.0488	Casserly	0.0419	Pacheco	0.0477
Uvas up	0.0442	Dos Picachos	0.0500	Bodfish up	0.0503	Casserly	0.0494
Little Arthur Red	0.0476	Little Arthur Red	0.0534	Pacheco	0.0505	Bodfish up	0.0535
Llagas up	0.0554	Llagas up	0.0545	Llagas low	0.0525	Llagas low	0.0698
Green Valley	0.0593	Green Valley	0.0572	Little Arthur Red	0.0744	Little Arthur Red	0.0784
Llagas low	0.0603	Llagas low	0.0679	Llagas up	0.1038	Llagas up	0.1866
Sprig Lake	0.1097	Sprig Lake	0.1053	Sprig Lake	0.2733	Sprig Lake	0.2384

The two Little Arthur Creek sites are separated from each other by seasonal flashboard dams that might separate the populations. Although the two sites clustered together based upon some unique similarities, there is a chord distance between them of 0.0443. The lower site at Mt. Madonna Road was much more closely related to the San Lorenzo sites, which is probably due to hatchery fingerlings planted in 1996. The site was apparently a mixture of native and hatchery fish in 1997. A relatively new fish ladder could also have recently allowed a few hatchery steelhead access to the formerly resident trout population in the upper watershed. Little Arthur Creek is a tributary of Uvas, but the sites are separated from lower Uvas Creek with chord distances of 0.0447 and 0.0509. The upper site at Redwood Retreat may still have contained native fish in 1997, based upon large chord distances from San Lorenzo sites. The creek apparently is a combination of hatchery and wild fish and may become more similar to Uvas Creek with continued steelhead access to the upper watershed.

Llagas Creek fish were substantially different (0.0550 to 0.0882) because of low genetic diversity, apparently reflecting a population bottleneck. The difficult passage conditions and scarcity of fish during 1997 and 1998 sampling indicate that few steelhead presently utilize this stream.

Green Valley Creek segregated out from the other sites with long chord distances (0.0567 to 0.0791) and it was less polymorphic than most of the other sites. Fish passage up this stream is difficult, and the lack of smolt coloration among fish sampled in May 1997

indicates that this population may be resident trout. With distances of 0.0572 and 0.0593 from Zayante Creek and Boulder, Green Valley Creek could be one of the streams that has not been impacted by San Lorenzo hatchery fish.

Upper Uvas, upper Llagas, upper Bodfish and Dos Picachos creeks are clustered in CONTML and have high chord distances from the San Lorenzo sites (0.0407 to 0.0554). These streams may have predominantly native fish. Upper Uvas and upper Llagas creeks are separated from respective lower sites by reservoirs. Upper Bodfish Creek also is above waterfalls that are presently impassable. Dos Picachos Creek is separated by a long, seasonally dry, stretch of valley floor channel north of the town of Hollister. It is also one of the most upstream tributaries in the watershed, making straying less likely. Because they are above barriers, all four streams contain predominately resident fish, although a high density of apparently trapped yearling smolts above a diversion dam in June 1997 on Dos Picachos suggests that steelhead are also present in that stream in some years.

Nei's D_A is a modified version of D_C, so it is not surprising that a neighbor-joining tree using D_A, as the one generated in NJBAFD2, will have a tree that is similar to the one produced in CONTML. There is only one minor difference. Upper Bodfish is clustered with the Pacheco and Casserly group rather than with the above barriers group, but branch lengths are long and the clustering patterns may not be as meaningful as distances between sites.

PHYLIP's CONTML, Cavalli-Sforza and Edwards chord distance and Nei's D_A generate topologies that make sense based on what is known about barriers to migration in this watershed. They show that the tributaries of the Pajaro River that contain most of the steelhead have probably been substantially impacted by hatchery plants from the San Lorenzo. The boundary between south central and central ESUs is not distinct. Several streams above reservoirs and other substantial barriers to migration may contain native populations.

Stepwise mutation model distances compared to D_C and D_A

The programs that use R_{ST} and $(\delta\mu)^2$ give some different results from the CONTML program and from each other. They do not make as much sense based on what we know about migration barriers in these watersheds. $(\delta\mu)^2$ separates Zayante Creek from all other streams even though there are no barriers to migration to Zayante Creek. With Rho, Boulder Creek is still closely related to lower Uvas and Corralitos creeks, but also is closely related to upper Uvas and Dos Picachos creeks, and Little Arthur at Mt. Madonna Road. While it is very possible that Little Arthur at Mt. Madonna has received significant San Lorenzo hatchery influence, this is not likely for upper Uvas and Dos Picachos creeks because they have substantial barriers to migration.

The program RSTCALC standardizes allelic frequencies so that the more polymorphic loci do not influence the distances more than the less polymorphic loci. In this neighbor-

joining tree, the two San Lorenzo sites are not split up as they are with $(\delta\mu)^2$. Lower Llagas has a relatively long distance from other sites with Rho, which makes sense for a putative bottlenecked population. In contrast, the $(\delta\mu)^2$ value suggests that lower Llagas is closely related to Casserly, despite the fact that lower Llagas had very few alleles.

There are several possible explanations for why the distances based upon the stepwise mutation models do not fit with what is known about the ecology of this watershed. First, these distances are based upon a mutation model rather than drift or migration. With the kind of population bottlenecks and the amount of hatchery influence these streams are likely to see, this may not be appropriate. Second, there is a high degree of variance in these distances that may make it difficult to use them to separate closely related populations. Third, these loci may contain compound repeats that do not follow a stepwise mutation model. Enough is known about these two watersheds to say that the results from CONTML and D_A probably generate more accurate phylogenetic relationships.

Genetic distances are dependent on both time and effective population size, and they increase rapidly with population bottlenecks. Bottlenecks, therefore, can decrease the reliability of a given topology using any distance measure. Population bottlenecks are a very real possibility for steelhead populations during a drought year, especially in a system such as the Pajaro where dams and water diversions reduce flow at critical times

of the year. D_C and the program used in CONTML do not assume constant population sizes over time.

Both D_C and D_A assume gene frequencies change because of genetic drift, and are not based on an evolutionary model. Rho and $(\delta\mu)^2$ assume a complex mutation model that builds genetic differences over time. Some of the differences in streams in the Pajaro may largely be due to hatchery stocking of fish and to barriers to migration, all occurring within the last century. $(\delta\mu)^2$ may analyze well the genetic variation in populations that have had a true island model history (Goldstein *et al.* 1999). But it may not perform as well when used with steelhead populations, like those in the Pajaro, that have had so many recent perturbations.

Some caution should be used when evaluating the population structure inferred from this microsatellite data. It is not clear what the minimum number of loci or the minimum sample sizes should be for this kind of study. It depends in part on divergence level, how polymorphic the loci are, and population heterozygosity. We would expect lower divergence levels between populations of *O. mykiss* within a watershed than between populations more geographically distant.

Takezaki and Nei's computer simulation suggests that with increasing heterozygosity, sample size should increase, and that with a low divergence level sample sizes should also be larger (sample size $\approx$ 50) (Takezaki and Nei 1996). They say that in general it is

more important to use more loci than a larger sample size. Accuracy of topologies constructed with D_A and D_C is high if thirty or more loci are used, divergence is low and heterozygosities are high. Divergence between the Pajaro and San Lorenzo populations would be expected to be lower than between these streams and northern streams, such as the Mad River and other possible sources of hatchery influences.

Virtual microsatellites in computer simulations may behave very differently, however, from real microsatellites. Furthermore, studies based on mammalian data may not apply to fish, since microsatellites may behave differently in poikilotherms than in homeotherms (Brooker *et al.* 1994). In a comparative study between microsatellite and allozyme markers in European brown trout (*Salmo trutta*), seven microsatellite loci and samples sizes of twenty-five to thirty fish were enough to detect significant genetic differentiation between populations within two watersheds and to assign individuals correctly to their populations (Estoup *et al.* 1998). Clustering analysis of populations using Cavalli-Sforza's chord distance and a neighbor-joining dendogram, however, was less successful for distinguishing intradrainage nodes that are geographically close. It is interesting to note that in the Estoup *et al.* study, Cavalli-Sforza's chord distance and the Goldstein distance performed equally well in separating the two main drainages studied.

O'Connell and Wright (1997) suggest a minimum sample size of fifty fish per population for loci that show five to ten alleles, and that at least five loci be used for meaningful confidence limits and estimates of variance for F statatistics of Weir and Cockerham

(1984). Four of the five loci used in this study had more than ten alleles. They point out, however, that it is unrealistic to expect adequate sample sizes for highly variable loci, because that would involve very large sample sizes. In several streams of the Pajaro, it was difficult to get a minimum sample size of thirty fish. In these cases they recommend the use of numerical re-sampling techniques, e.g. bootstrapping. Bootstrapping can give some measure of precision (Hedges 1992), but the robustness of nodes in topologies when bootstrapping over only five loci may be calculated artificially high.

Management implications

Although the San Lorenzo and Pajaro Rivers are in two separate ESU's, the steelhead in many of the tributaries of the Pajaro are indistinguishable from steelhead in the San Lorenzo. Some straying could be expected to occur naturally, but hatchery plants have probably swamped the native stocks in Corralitos and lower Uvas creeks and their easily accessible tributaries. These streams have the majority of steelhead production in this watershed in most years. In general, it appears that San Lorenzo River stocks have affected those streams that do not have a substantial barrier to migration.

Remnant native populations are probably present in upper Llagas, upper Uvas, upper Bodfish, Dos Picachos, and Green Valley creeks. All of these tributaries have predominantly resident populations. Upper Llagas and upper Uvas creeks are above reservoirs, upper Bodfish is above a presently impassable fall, and Dos Picachos is separated by extensive, seasonally dry channels. These four sites are distinct from San

Lorenzo fish and from each other. They are the most distantly related from the San Lorenzo River sites. The upper site on Little Arthur Creek was apparently also a distinct native stock in 1997, but is now open to access by hatchery steelhead.

Pacheco Creek may have lost its steelhead runs because it often has difficult access, and little or no smolt outmigration in major drought years (1976-77, 1987-1991). It appears now to be a mixture of native and hatchery fish.

Population structures and genetic distances for this watershed based on Cavalli-Sforza's chord distance and Nei's D_A distance make more sense than those based on Slatkin's R_{ST} and Goldstein's $(\delta\mu)^2$. While it would be better to have larger sample sizes and more loci in this data set, it does appear that Cavalli-Sforza's chord distance D_C and Nei's D_A suggest an accurate population structure. At issue is how hatchery plants and relatively recent changes in the ecology of the watershed (dams, diversions) affect the assumptions of the computer programs that are available for microsatellite analysis. Future studies should examine real salmonid populations to determine what kind of genetic analysis is appropriate.

References

Allendorf, F.W., and Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmon fishes. *In* Evolutionary Genetics of Fishes. *Edited by* B. Turner. Plenum Press, New York. pp. 1-53.

Angers, B., and Bernatchez, L. 1997. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Mol. Biol. Evol. **14**(3): 230-238.

Ashley, M.V., and Dow, B.D. 1994. The use of microsatellite analysis in population biology: background, methods and potential applications. EXS **69**: 185-201.

Banks, M.A., Baldwin, B.A. and Hedgecock, D. 1996. Research on chinook salmon (*Onchorhynchus tshawytscha*) stock structure using microsatellite DNA. Bull. Natl. Res. Inst. Aquacult. **2**(Suppl.l): 5-9.

Brooker, A., Cook, D., Bentzen, P., Wright, J., and Doyle, R. 1994. Organization of microsatellites differs between mammals and cold-water teleost fishes. Can. J. Fish. Aquat. Sci. **51**: 1959-1965.

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

Cockerham, C. C., and Weir, B.S. 1993. Estimation of gene flow from F-statistics. Evolution, **47**(3): 855-863.

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

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin M., and Freimer, N.B. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA **91**: 3166-3170.

Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T., and Chakraborty, R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics **12**: 241-253.

Efron, B. 1982. The jacknife, the bootstrap, and other resampling plans. CBMS-NSF. Regional conference series in applied mathematics. No. 38. Society for Industrial and Applied Mathemetics. Philadelphia, PA.

Estoup, A., Garnery, L., Solignac, M., and Cornuet, J.M. 1995. Microsatellite variation in honey bee (*Apis melifera L.*) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics, **140**: 679-695.

Estoup, A., Rousset, F., Michalakis,Y., Cornuet, J.M., Adriamanga, M., and Guyomard, R. 1998. Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). Mol. Ecol. **7**: 339-353.

Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution, **35**: 1229-1241.

Felsenstein, J. 1984*a*. Distance methods for inferring phylogenies: a justification. Evolution, **38**(1): 16-25.

Felsenstein, J. 1984*b*. Phylogenies from gene frequencies: a statistical problem. Syst. Zool. **34**(3): 300-311.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, **39**: 783-791.

Ferguson, M.M., and Danzmann, R.G. 1998. Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? Can. J. Fish. Aquat. Sci. **55**: 1553-1563.

Fu, Y.X., and Chakraborty, R. 1988. Simultaneous estimation of all the parameters of a stepwise mutation model. Genetics, **150**: 487-497.

Goldstein, D.B., Linares, A.R. Cavalli-Sforza, L.L. and Feldman, M.W. 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics, **139**: 463-471.

Goldstein, D.B., and Pollock, D.D. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J. Hered. **88**: 335-342.

Goldstein, D.B., Roemer, G.W., Smith, D.A., Reich, D.E., Bergman, A., and Wayne, R.K. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics, **151**: 797-801.

Goodman, S.J. 1997. Rst Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. Molecular Ecology, **6**: 881-885.

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

Guo, S.W., and Thompson, E.A. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics, **48**: 361-372.

Haldane, J.B.S. 1954. An exact test for randomness of mating. J. Genet. **52**: 631-635.

Hedges, S.B. 1992. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. Mol. Biol. Evol. 9(2): 366-369.

Hillis, D.M. 1995. Approaches for assessing phylogenetic accuracy. Systematic Biology, **44**(1): 3-16.

Jarne, P., and Lagoda, P.J.L. 1996. Microsatellites, from molecules to populations and back. TREE **11**(10): 424-429.

Kimura, M., and Crow, J.F. 1964. The number of alleles that can be maintained in a finite population. Genetics, **49**: 725-738.

Lessios, H.A. 1992. Testing electrophoretic data for agreement with Hardy-Weinberg expectations. Mar. Biol. **112**:517-523.

Levinson, G., and Gutman, G.A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**: 203-221.

McConnell, S.K., O'Reilly, P., Hamilton,L., Wright, J.M., and Bentzen, P. 1995. Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar*): genetic differentiation of North American and European populations. Can. J. Fish Aquat. Sci. **52**: 1863-1872

Michalakis, Y., and Excoffier, L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics, **142**: 1061-1064.

Morris, D.B., Richard, K.R., and Wright, J.M. 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. Can. J. Fish. Aquat. Sci. **53**: 120-126.

National Marine Fisheries Service. 1996. Status review of west coast steelhead from Washington, Oregon, and California. NOAA Technical Memorandum NMFS-NWFSC-27.

Nehlsen, W. 1997. Pacific salmon status and trends - a coastwide perspective. *In* Pacific Salmon and their Ecosystems, *Edited by* D.J. Stouder, P.A. Bisson, and R.J. Naiman. Chapman and Hall, New York. pp. 41-50.

Nei, M. 1972. Genetic distance between populations. Am. Nat. **106**: 283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70**(12) Part I: 3321-3323.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, **89**: 583-590.

Nei, M. 1996. Phylogenetic analysis in molecular evolutionary genetics. Annu. Rev. Genet. **30**: 371-403.

Nei, M., Tajima, F., and Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. **19**: 153-170.

Nelson, R.J., Beacham, T.D., and Small, M.P. 1998. Microsatellite analysis of the population structure of a Vancouver Island sockey salmon (*Oncorhynchus nerka*) stock complex using nondenaturing gel electrophoresis. Mol. Mar. Biol. Biotechnol. **7**(4): 312-319.

Nielsen, J.L., Carpanzano, C., Fountain, M.C., and Gan, C.A. 1997. Mitochondrial DNA and nuclear microsatellite diversity in hatchery and wild *Oncorhynchus mykiss* from freshwater habitats in southern California. Trans. Am. Fish. Soc. **126**: 397-417.

Nielsen, J.L., Fountain, M.C., and Wright, J.M. 1997. Biogeographic analysis of Pacific trout (*Oncorhynchus mykiss*) in California and Mexico based on mitochondrial DNA and nuclear microsatellites. *In* Molecular Systematics of Fishes. *Edited by* T.D. Kocher and C.A. Stepien. Academic Press, New York. Chapter 5.

Nielsen, J.L., Gan, C.A., Wright, J.M., Morris, D.B., and Kelley, W.K. 1994. Biogeographic distributions of mitochondrial and nuclear markers for southern steelhead. Mol. Mar. Biol. Biotechnol. 3(5): 281-293.

O'Connell, M., and Wright, J.M. 1997. Microsatellite DNA in fishes. Reviews in Fish Biology and Fisheries **7**:331-363.

Ohta, T, and Kimura, M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. Camb. **22**: 201-204.

Olsen, J.B., Bentzen P., and Seeb, J.E. 1998. Characterization of seven microsatellite loci derived from pink salmon. Mol. Ecol. **7**(8): 1087-1089.

Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**: 357-358.

Promega Corporation. 1993. Technical manual, DNA silver staining system. Part # TMD005. Madison, WI.

Raymond, M., and Rousset, F. 1995. An exact test for population differentiation. Evolution, **49**(6): 1280-1283.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics, **145**: 1219-1228.

Ruzzante, D.E. 1998. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. Can. J. Fish. Aquat. Sci. **55**: 1-14.

Ruzzante, D.E., Taggart, C.T., Cook, D., and Goddard, S. 1996. Genetic differentiation between inshore and offshore cod (*Gadus morhua*) off Newfoundland: microsatellite DNA variation and antifreeze level. Can. J. Fish. Aquat. Sci. **53**: 634-645.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**(4): 406-425.

Schlötterer, C., and Tautz, D. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20**: 211-215.

Scribner, K.T., Arntzen, J.W., and Burke, T. 1994. Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo*, using allozyme, single-locus microsatellite, minisatellite, and multilocus minisatellite data. Mol. Biol. Evol. **11**(5): 737-748.

Scribner, K.T., Crane, P.A., Spearman, W.J., and Seeb, L.W. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Can. J. Fish. Aquat. Sci. **55**: 1748-1758.

Scribner, K.T., Gust, J.R., and Fields, R.L. 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic application. Can. J. Fish. Aquat. Sci. **53**: 833-841.

Shapovalov, L., and Taft, A.C. 1954. The life histories of the steelhead rainbow trout *(Salmo gairdneri gairdneri)* and silver salmon (*Oncorhynchus kisutch*) with special reference to Waddell Creek, California, and recommendations regarding their management. California Dept. of Fish and Game. Fish bulletin no. 98.

Shriver, M.D., Jin, L., Chakraborty, R., and Boerwinkle, E. 1993. VNTR allele frequency distribution under the stepwise mutation model. Genetics, **134**: 983-993.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics, **139**: 457-462.

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

Tessier, N., Bernatchez, L., Presa, P., and Angers, B. 1995. Gene diversity analysis of mitochondrial DNA, microsatellites and allozymes in landlocked Atlantic salmon. J. Fish Biol. **47**(Supplement A): 156-163.

Tessier, N., Bernatchez, L., and Wright, J.M. 1997. Population structure and impact of supportive breeding inferred from mitochondrial and microsatellite DNA analyses in land-locked Atlantic salmon *Salmo salar L.* Mol. Ecol. **6**(8): 735-750.

Valdes, A.M., Slatkin, M., and Freimer, N.B. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics, **133**: 737-749.

Valsecchi, E., Palsboll, P., Hale, P., Glockner-Ferrari, D., Ferrari, M., Clapham, P., Larsen, F., Mattila, D., Sears, R., Sigurjonsson, J., Brown, M., Corkeron, P., and Amos, B. 1997. Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera novaeangliae*). Mol. Biol. Evol. **14**(4): 355-362.

Waples, R.S. 1991. Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. Mar. Fish. Rev. **53**(3): 11-22.

Waples, R.S. 1994. Genetic considerations in recovery efforts for Pacific salmon. Conservation Biology, **8**(3): 884-888.

Weber, J.L., and Wong, C. 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. **2**: 1123-1128.

Weir, B.S. 1990. Genetic data analysis. Sinauer Publ., Sunderland, MA.

Weir, B.S., and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution, **38**(6): 1358-1370.

Wright, S. 1951. The genetical structure of populations. Ann. Eugen. **15**: 323-354.

Zhivotovsky, L.A., and Feldman, M.W. 1995. Microsatellite variability and genetic distances. Proc. Natl. Acad. Sci. USA **92**: 11549-11552.

Appendix A

Microsatellites Described

Microsatellite DNA is a form of variable number tandem repeats (VNTRs) composed of short tandem repeat segments of two to five base pairs per repeat unit. They are also known as simple sequence repeats (SSR). Di-, tri-, and tetra-nucleotide repeats are frequently used in studies of parentage and population structure because they have some of the highest mutation rates of any molecular tools used to date, about 2 5 X 10^{-5} to 10^{-2} in humans (Weber and Wong 1993). As noncoding regions of DNA, microsatellites are generally considered to be selectively neutral and inherited in a Mendelian fashion, although the latter assumption needs further testing with pedigree studies. They are ubiquitous throughout eukaryotic genomes and are easily amplified with PCR.

Three types of microsatellites have been defined, pure, compound and interrupted (Jarne and Lagoda 1996). Examples are:

Pure	$(GT)_n$
Compound	$(GT)_n(AT)_m$
Interrupted	$(GT)_nCATGC(GT)_m$

Pure repeats tend to be the most polymorphic, and may represent the ancestral state (Angers and Bernatchez 1997), while compound and interrupted repeats tend to be more stable (DiRienzo *et al.* 1994). Angers and Bernatchez sequenced seventy-eight alleles of

one locus among salmonids and found the allelic diversity within the microsatellite to be caused by changes in the number of repeats, base substitutions, and insertion of large repeat and nonrepeat sequences. At the intraspecific level, variation in the number of dinucleotide repeats was a more important source of diversity (39% of pairwise comparisons) than at the interspecific level. Homoplasy, alleles of the same length but different composition, was observed indicating that it may not always be accurate to infer mutational events from size information alone.

Microsatellites have not always succeeded in reconstructing phylogenies. This may be partly due to constraints on their mutational processes, including range constraints (large alleles greater than sixty repeats are rarely observed), irregularities and asymmetries in the direction of mutation, and microsatellite degradation (Goldstein and Pollock 1997). The high mutation rate of microsatellites is thought to be caused by polymerase slippage (Levinson and Gutman 1987, Schloetterer and Tautz 1992), and this mechanism may work differently in a cold water environment on cold-blooded organisms. Teleost fishes, cod, rainbow trout and Atlantic salmon, have microsatellites that are more than twice the length of the longest mammalian repeats, human, porcine and canine (Brooker *et al.* 1994). See Appendix B for a detailed description of mutational models.

Some loci are highly polymorphic in some populations or species but monomorphic in others, so a question can be raised about the long-term stability of microsatellite loci (Takezaki and Nei 1996). Much of the work on microsatelllites has been done on

mammalian genomes, and some caution should be used in extrapolating to fish. Salmonid diploidization from ancestral tetraploidy is not complete (Allendorf and Thorgaard 1984).

Appendix B

Mutational Models and Genetic Distances

Given the widespread use and importance of microsatellite data to population and taxonomic studies, it is critical to verify the assumptions made in phylogenetic analysis. There have been numerous studies testing the performance of various genetic distances. Some of these studies used computer simulations to test the performances of these distances with computer generated alleles (Takezaki and Nei 1996, Goldstein *et al.* 1995). Others tested the performance of distances with known populations (DiRienzo *et al.* 1994, Estoup *et al.* 1995, Goldstein *et al.* 1999), and others compared microsatellite data with other molecular methods, such as allozymes and mitochondrial DNA (Estoup *et al.* 1998). There is still some controversy over which methods work well under what circumstances. Goldstein reviewed the assumptions and methods commonly used for microsatellite analysis (Goldstein and Pollock 1997).

Two general mutation models have been developed. In the infinite alleles model (IAM) of Kimura and Crow (1964) each mutation creates a new allele with a low mutation rate μ. In the stepwise mutation model (SMM) of Ohta and Kimura (1973) a mutation adds or subtracts a repeat unit in an allele. Goldstein *et al.* (1995) and Slatkin (1995) developed genetic distance measures that utilize advantages of the stepwise mutation model ($(\delta\mu)^2$ and R_{ST}). The difference in the number of repeats gives information about the time between the new allele and ancestral alleles. The IAM does not use information of this sort, and it assumes the mutation process erases any memory of the previous

alleles. There are also classical distance measures that rely on multidimensional geometry rather than a particular evolutionary model (D_C and D_A).

For a distance to be effective, it is important that its expectation increases linearly with time. It is more precise if its coefficient of variance is low. There is often a tradeoff between these two considerations. In general, the distances based on the SMM have a higher coefficient of variance than the other distances, but they remain linear over a longer period of time.

ASD (Goldstein *et al.* 1995, Slatkin 1995) is based on the SMM. It is the average squared difference in number of repeat units for two alleles drawn from different populations isolated τ generations in the past. It is linear over time but has a high variance. The intra-population variance term makes it sensitive to fluctuations in population size (Goldstein and Pollock 1997). It has the form

$$\mathrm{ASD} = \sum_i \sum_j (i - j)^2 x_i y_j$$

where x_i and y_j are the frequencies in populations x and y with alleles having i and j repeats (Goldstein and Pollock 1997).

$(\delta\mu)^2$ (Goldstein *et al.* 1995) is based on the SSM model and is a variation of ASD. It assumes a strict stepwise mutation process of one repeat added or subtracted per

mutation, with no constraint of allele size. It increases linearly with time like ASD, but has a lower variance. It has been shown to be less sensitive to fluctuations in population size (Takezaki and Nei 1996). ASD and $(\delta\mu)$ assume that repeat motifs are pure, not interrupted. $(\delta\mu)^2$ has the form

$$(\delta\mu)^2 = (m_x - m_y)^2$$

where m_x and m_y are the means of allele sizes in populations x and y. It can be written as

$$(\delta\mu)^2 = \text{ASD} - V_x - V_y$$

where V_x and V_y are the variances in allelic score of populations x and y.

R_{ST} (Slatkin 1995) is based on the SSM model. It is closely related to $(\delta\mu)^2$ but mutations of more than one repeat unit are allowed. It is analogous to Wright's (1951) F_{ST} and can be used to estimate effective migration rates or times since population divergence. It assumes a high mutation rate μ per generation (in contrast to a low mutation rate for the infinite-allele mutation model) and an increment in allele size being a random variable with mean 0 and variance σ^2_m independent of allele size. A special case would be when $\sigma^2_m = 1$, the one-step model of Goldstein. It has the form

$$R_{ST} = (\bar{S} - S_W)/\bar{S}$$

Where S_W is the average sum of squares of the differences in allele size within each population (or twice the average of the estimated variances of allele size within each

population) and Sbar is twice the estimated variance in allele size in the collection of populations together.

D_C, Cavalli-Sforza and Edwards' (1967) chord distance, assumes that drift is the main cause of differentiation. It does not assume that population sizes have remained constant and equal in all populations. It has an expectation that rises linearly not with time, but with the sum over time of 1/N, where N is the effective population size. If two populations are located on the surface of a multidimensional hypersphere by allele frequencies at the jth locus, D_C for this locus gives the chord distance between these two populations, where the angle (θ_j) for the two populations is given by

$$\cos \theta_j = \sum_{i}^{m_j} \sqrt{x_{ij} y_{ij}}$$

Here x_{ij} and y_{ij} are the frequencies of the ith allele at the jth locus in populations x and y, m_j is the number of alleles at the jth locus, and r is the number of loci.

D_C takes the form

$$D_C = (2/\pi r) \sum_{j}^{r} \sqrt{2\left(1 - \sum_{i}^{m_j} \sqrt{x_{ij} y_{ij}}\right)}$$

(Takezaki and Nei 1996).

D_A (Nei *et al.* 1983) is also not based on a model of evolution. A modified version of D_C, it is based on the sum of the products of frequencies of alleles shared between populations. It has the form

$$D_A = 1 - (1/r) \sum_{j}^{r} \sum_{i}^{m_j} \sqrt{x_{ij} y_{ij}}$$

(Takezaki and Nei 1996).

D_C and D_A have been shown to reconstruct closely related topologies better than the SMM distances (Takezaki and Nei 1996, Goldstein *et al.* 1995) in computer simulations. Takezaki and Nei's (1996) computer simulations showed that Cavalli-Sforza and Edwards' chord distance (D_C) and Nei's D_A construct topologies more accurately than other distance measures under many different conditions. However, Goldstein's $(\delta\mu)^2$ was more effective for estimating evolutionary times. Takezaki and Nei (1996) also tested for the effect of a bottleneck on the accuracy of phylogenetic trees. D_A and D_C were the most accurate, although all distances are affected by bottlenecks. For populations that are recently separated, (less than 100 to 500 generations), D_A and D_C give topologies that are more accurate, but $(\delta\mu)^2$ may be more accurate for more distantly related populations. However, with range constraints, loci that have low mutation rates would be expected to be accurate over a longer period of time.

Di Rienzo *et al.* (1994) provide evidence that the single-step mutation model may be inadequate. In their two-phase mutation model, most mutations are single-step, but infrequent large jumps in repeat scores occur. A one-step mutation has a probability of p and a multistep mutation a probability of $1 - p$. The change in the number of repeat units comes from a specified distribution with a specified variance that allows for large changes. Computer simulations compared the expected values for frequency distributions of allele sizes for the two-phase model, the one-step model, and geometric models. The expected values were compared with observed values for ten SSR loci in the human Sardinian population whose genetic and demographic histories had already been reconstructed. The two-phase model provided the best fit for eight of the ten loci. Their data also suggested that as the variance in repeat number increased the frequency of magnitude of the second phase of the mutation increased.

Estoup *et al.* (1995) also compared the IAM and SMM in a phylogenetic study of African and European honeybees. For the SMM, the discrepancy between observed and expected numbers of alleles was consistently larger than the IAM. This was, they pointed out, in disagreement with the results of Edwards *et al.* (1992) and Shriver *et al.* (1993). The better fit of the IAM, they postulate, may in part be due to the compound repeat motif of the majority of their microsatellites, which included repeats of two or three different lengths. The presence of compound repeats would violate the assumption of SMM models that microsatellites evolve in a stepwise process of adding or subtracting perfect repeats.

Appendix C

Protocols

Collection Protocol

1. Using clean scissors cut a small piece of fin tissue from the caudal fin (tail) of the live fish. Tissue size should be approximately $2mm^2$. A wedge from the upper or lower lobe of the fin works fine. Adipose fins contain little DNA (mostly complex lipids) and do not provide an easy target for DNA extraction. Scissors should be well rinsed between samples to prevent cross-contamination of DNA. Do as much as possible to avoid contamination of the sample (touching it for example).
2. Tissues are collected and stored separately on chromatography paper or weighing paper in individual coin collecting envelopes. Envelopes are labeled with fish number, size class (standard length), location and date.
3. Immediately upon return from the field, samples are air-dried on the paper until all moisture has evaporated (about 24 hours at room temperature). Dried fin clips are repackaged separately.

Laboratory Protocol (Recipes at end)

DNA Extraction

1. Aliquot 200 µl of 5% Chelex from a stirring bottle into 1.5 ml microfuge tubes. Cut a small piece of fin (approx. 2mm square) on weighing paper with a razor blade and add it to a tube of Chelex. A toothpick may be used to facilitate the transfer of the sample from the paper to the tube. Gloves should be worn. Avoid contamination of DNA with DNA of previously handled fish.
2. Incubate the tubes at 65°C for 20 minutes.
3. Vortex for 5-10 seconds.
4. Incubate the tubes at 95°C for approximately 10 minutes.
5. Vortex for 5-10 seconds.
6. Microcentrifuge at full speed for 3 minutes.
7. Store in the freezer if not continuing directly on.

PCR

1. Thaw samples at 65°C, vortex, and spin for 3 minutes in a microcentrifuge. Label a set of 0.5 ml PCR tubes with sample name and primer. Label one for a negative control. Label as many as needed for positive controls/size standards
2. Make a master mix of the PCR reagents. $MgCl_2$ concentrations are:
   4 mM Omy77
   3 mM Omy207, Ssa14, Ssa289, Oneµ2

   For 10 vials mix:

	3mM Mg	4mM Mg
HPLC water	27.66 µl	24.80 µl
PCR Buffer (see recipes)	6.25 µl	6.25 µl
$MgCl_2$ (Promega, 25 mM)	8.59 µl	11.45 µl
dNTP mix (Promega, 10 mM each dNTP)	6.25 µl	6.25 µl
Primer Forward (Operon, 10 µM in HPLC water)	6.25 µl	6.25 µl
Primer Reverse (10 µM in water)	6.25 µl	6.25 µl
Taq (Promega, 5u/µl)	0.30 µl	0.30 µl

3. Aliquot 6.15 µl of master mix into each PCR tube. Add a drop of mineral oil. Pipet 1 µl of fin extract through the oil to the bottom of the tube.
4. Place the 0.5 ml tubes into the PCR machine. Run the PCR program appropriate for the locus.

   General thermocycler program:

Step 1	denature	94°C for 1 minute
Step 2	denature	94°C for 40 seconds
Step 3	anneal	50 or 52°C for 1 minute

Step 4	extend	72°C for 1 minute
Step 5	go to Step 2 for 30 or 32 cycles	
Step 6	4°C forever	
Step 7	End	

5. Refrigerate the PCR products.

Gel Preparation

The following protocol is for a 16.24 X 14.5 inch polyacrylamide gel for the Hoefer Poker Face system.

CAUTION: ACRYLAMIDE IS A NEUROTOXIN AND CANCER SUSPECT AGENT. AVOID INHALATION AND CONTACT WITH SKIN. ALWAYS DOUBLE GLOVE.
BIND SILANE IS TOXIC AND SHOULD BE USED IN A CHEMICAL FUME HOOD.

1. Mix the following reagents:

   - Acrylamide: Mix the acrylamide according to the directions on the bottle. Half a bottle is sufficient for one large gel. One bottle makes 125 ml. Unused acrylamide can be stored for up to two weeks in a dark bottle.
   - Bind Silane: In a chemical fume hood, prepare fresh binding solution by adding 3 µl Bind Silane to
     1 ml of 0.5% acetic acid in 95% ethanol
     (5 µl acetic acid + 995 µl 95% ethanol)
   - TBE buffer

2. Clean the glass plates with Alconox and water. The last rinse should be with distilled water. Care should be taken to avoid cross contamination of the plates; the surfaces of the plates should not touch each other, the same sponges, towels and gloves should not be used for both plates. Clean the plates twice with 95% ethanol and Kimwipe tissues.
3. Wearing gloves, spray Acrylease on to the large glass plate. With a dry Kimwipe, spread the Acrylease using a circular motion over the entire surface.
4. Wait five minutes for the Acrylease to dry. Remove the excess Acrylease with Kimwipe dampened with deionized H_2O. Dry the plate with another Kimwipe.
5. Wipe the small glass plate using a Kimwipe saturated with the freshly prepared Bind Silane. Be certain to wipe the entire plate surface with the saturated tissue.
6. Wait five minutes for the binding solution to dry. Wipe the small glass plate three to four times with 95% ethanol and Kimwipe tissues to remove the excess binding solution. Rub hard to remove all excess. Note: Failure to wipe excess binding solution from the small glass plate will cause the gel to stick to both plates, and the gel will be destroyed upon separation of the glass plates after electrophoresis.

7. Take care to prevent the treated surfaces from touching each other. Assemble the glass plates by placing 0.4 mm (or 0.35 mm) side and bottom spacer between the plates and use clamps to hold them in place. The foam tab on the side spacers should fit snugly against the short plate. Lean the assembled plates against a support at an angle that will allow the gel to be poured from one corner to the other.
8. Add TEMED and 10% ammonium persulfate to the acrylamide solution and stir.

62.5 ml	6% acrylamide
15.5 µl	TEMED
625 µl	10% ammonium persulfate (freshly prepared)

9. Draw the acrylamide solution up into a syringe. Carefully pour the acrylamide between the glass plates. Remove bubbles.
10. Position the gel horizontally. Insert a sharkstooth comb, straight side into the gel, until the teeth of the comb are at the edge of the large glass plate, or use a comb spacer. Secure the comb with clamps.
11. Pour some acrylamide into a microcentrifuge tube as a polymerization control. Allow polymerization for about an hour. Rinse the syringe with water before the residual acrylamide polymerizes. The gel may be stored overnight at room temperature if a paper towel saturated with dH_2O and plastic wrap are placed around the well end of the gel to prevent the gel from drying out. (Crystallization of the urea will destroy the gel.)

Electrophoresis

Note: Read the instruction manual for the electrophoresis equipment and the power supply before beginning. It contains important safety information.

Gel setup:

1. Remove the clamps from the polymerized acrylamide gel assembly and clean the outside of the glass plates with paper towels saturated with deionized water.
2. Remove the comb and bottom spacer. Clean excess acrylamide. Dry with towels.
3. Position the base so that the Poker Face II label is in front. Level the unit.
4. Install both sandwich supports. Place the support so that all six holes are engaged when the support is placed on the base.
5. Place the safety shield so that the long edge is flush with the front of the base. Align all four holes to the threaded holes in the base and tighten all four screws.
6. Apply a thin film of seal grease to both gaskets on the upper buffer chamber assembly. Place the assembly onto the short plate of the sandwich. Carefully align all bottom edges.
7. Install the clamps.
8. Pour enough 1X TBE buffer into the lower buffer chamber in the base so that the entire gel surface will be submerged. Lift the sandwich assembly and orient so that the upper buffer chamber is in the back. Insert the assembly through the sandwich supports into the lower chamber. Remove air bubbles between the plates with a pipet or syringe.

9. Plug the connectors (in the bottom shield) into the jacks in the base (behind the gel assembly). The two black plugs are on the left and the single red plug is on the right. Screw in both retaining thumbscrews until finger tight.
10. Add buffer to the upper buffer chamber.
11. With a pipet, squirt buffer into the wells to wash out any water or gel particles.
12. Install the safety lid, plugging the connectors into the clamps.
13. Prerun the gel at 65 watts for 1/2 hour.

Running the Gel:

1. Prepare the PCR samples by mixing
   5 µl sample with
   2.5 µl STR 3X loading solution
   Label the 0.6 ml tubes consecutively. If necessary, flash spin samples to bring contents to the bottom of the tube. Record which samples were loaded in each lane.
2. Denature the samples by heating at 95°C for 2 minutes. Place on ice.
3. Turn off and disconnect the electrophoresis equipment. Flush the urea from the well area and carefully insert the sharkstooth comb teeth into the gel approximately 1 - 2 mm.
4. Load 6 µl of each sample into a well. To prevent the gel from cooling, gel loading time should not exceed 20 minutes.
5. Run the gel at 65 watts for:
   Omy77 1 hour, 50 minutes
   Omy207 1 hour, 50 minutes
   Oneµ2 3 hours, 20 minutes
   Two sets may be run for the following loci:
   Ssa14 1 hour, 50 minutes, load second set 35 minutes into run
   Ssa289 1 hour, 50 minutes, load second set 30 minutes into run

Note: In a 6% gel, bromophenol blue migrates at approximately 25 bases and xylene cyanol migrates at approximately 105 bases.

Silver Staining (based on Promega's Technical Manual for Silver Staining)

1. While the gel is running, prepare the reagents, two liters of each per gel: developer, staining solution, fix/stop solution.
2. When the gel is done, remove the glass plates. Remove comb and side spacer, pry apart the glass plates with plastic wedge (not metal). Place the small plate with the gel in developing tray.
3. Follow the steps below:

	Solution	Time
a.	fix/stop solution	20 minutes
b.	dH_2O (millipore)	2 minutes
c.	repeat step b twice	2 X 2 minutes
d.	staining solution	30 minutes

e.	dH_2O (millipore)	10 seconds (critical step)
f.	developer (4 - 10^{o}C)	2 - 5 minutes (until alleles are visible)
g.	fix/stop solution (add to developer)	5 minutes
h.	dH_2O (millipore)	2 minutes

Note: The duration of step e is critical. Total time from immersion in dH_2O to immersion in developer should be less than 20 seconds. If that time is exceeded, repeat the staining procedure beginning with step d.

4. Allow gel to dry overnight. Cover it with a tray to keep off the dust.
5. Waste silver can be precipitated from the staining solution. Add 3 ml of 4 M NaCl to 2 liters of solution (2 g $AgNO_3$) and leave it to precipitate one or two days. Decant the water into the sink.

Photograph Gel

A photograph of the gel is made with automatic processor compatible film (Promega). It is like making a contact print of the gel.

1. In the darkroom with safelight, put the gel on the fluorescent light box, gel side up. Put a strip of film emulsion side (shiny side) in contact with the gel. Put a plexiglass plate on top to increase contact. Put a piece of black cardboard between gel and light box.
2. Turn on the light box, remove cardboard after light ceases flickering. Exposure times of 12 to 20 seconds are usually good.
3. Place film in GBX developer (Kodak) until bands appear, about 3 - 5 minutes. Rinse with water for one minute. Put paper in GBX fixer for three minutes. Rinse for one minute and hang to dry.

Composition of Solutions

Bind Silane

3 µl Bind Silane (Promega)
1 ml 0.5 % acetic acid in 95% ethanol

10% ammonium persulfate

0.125 g ammonium persulfate (Midwest Scientific)
1.25 ml dH_2O

developer solution

60 g sodium carbonate, anhydrous, certified ACS (Fisher Scientific)
3 ml 37% formaldehyde, certified ACS (Fisher Scientific)
20 µl sodium thiosulfate 0.2 g/ml (Sigma)
2000 ml millipore dH2O

Add sodium carbonate slowly to water. Chill solution to $10^{o}C$.
Prepare fresh for each use.

staining solution

2 g silver nitrate ($AgNO_3$)
3 ml 37% formaldehyde (Fisher Scientific)
2000 ml millipore dH_2O

fix/stop solution (10% acetic acid)

200 ml glacial acetic acid (Fisher Scientific)
1800 ml millipore dH_2O

10X TBE

107.8 g	Tris base	Dissolve the Tris base and EDTA in 800 ml millipore water. Slowly add the boric acid and monitor the pH until it is 8.3. Bring the volume to 1 liter with water.
7.44 g	EDTA (disoduim salt, dihydrate)	
55.0 g	boric acid	

STR 3X loading solution (Promega's Technical Manual for Silver Staining)

10 mM NaOH
95% formamide
0.05% bromophenol blue
0.05% xylene cyanol FF

Chelex

1.25 g Chelex (Biorad)
25 ml dH_2O (HPLC spectrograde)

10X PCR buffer (from Jennifer Nielsen's lab, Hopkins Marine Lab, Pacific Grove, CA)

For 1 ml:	670 µl	1 M Tris	Mix base with water and bring pH to 8.6.
	83 µl	2 M NH_4SO_4	Use HCl or NaOH to adjust pH.
	7 µl	14 M β-mercaptoethanol	
	240 µl	distilled water	

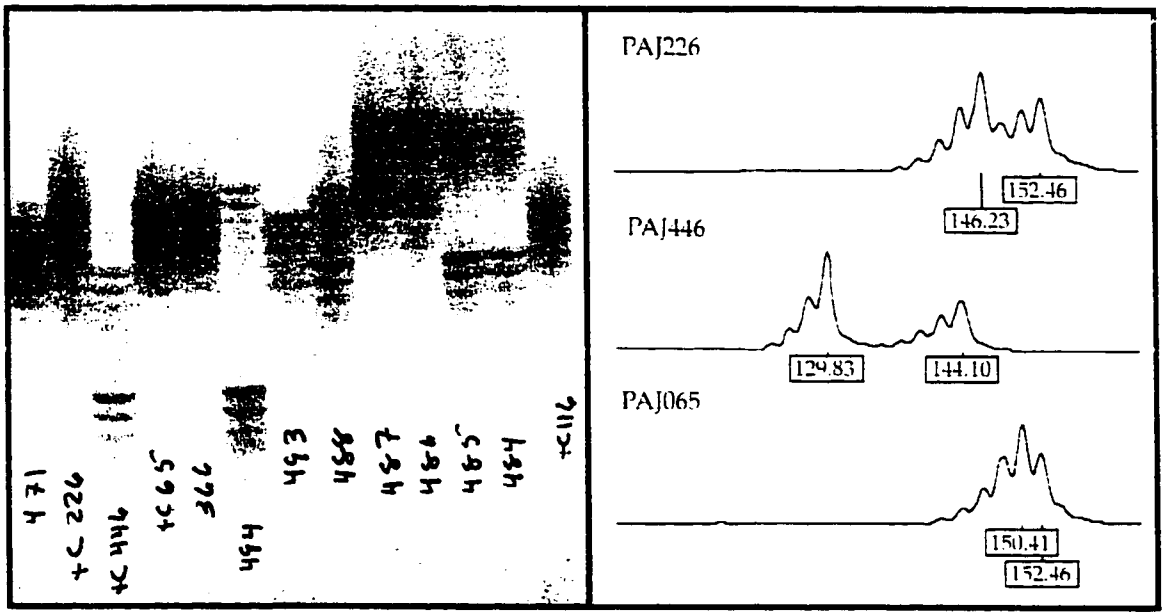
Scoring Alleles

PCR products from thirty-five fish were run on a Perkin Elmer ABI automatic sequencer, model 672, at Hopkins Marine Station in Pacific Grove, California, for Omy77, Ssa14, Oneμ2 and Ssa289. Twelve fish were run for Omy207. PCR conditions were the same as those used at San Jose State University except flourescently labeled primers were used. The largest peaks were used to define each allele size. A subset of these fish was used on the polyacrylamide gels as size standards and positive controls. (Fig C-1, C-2, C-3) Heterozygotes that had clearly separated alleles were used to determine allele shape, and that in turn helped to define heterozygote patterns that were composed of overlaying alleles (i.e. two or four base pairs apart).

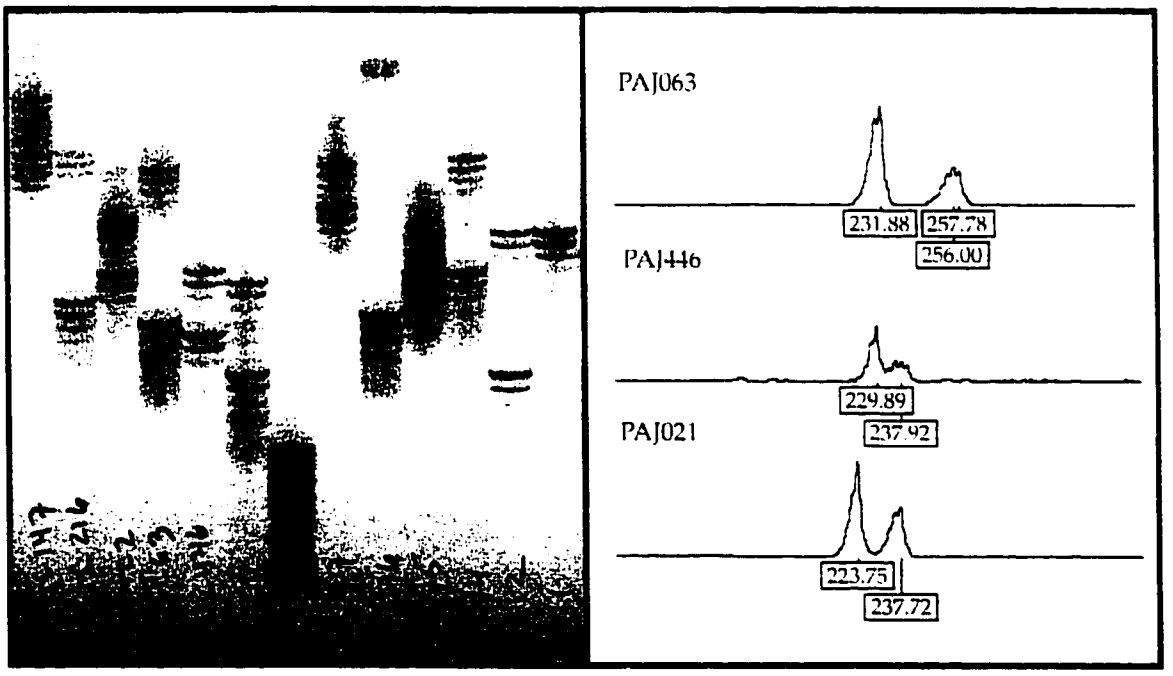
Gels were scored independently by two people for Omy77 and Ssa289. Any alleles that were not agreed upon were run again next to size standards and scored again. Final results had 100% agreement. The other three loci were scored by one person, and any alleles that were not clear were run again next to size standards until they could be unambiguously scored.

A scoring ruler was made by laying a small piece of paper on the gel and making tick marks along the edge where the darkest bands were. Ssa14, Oneμ2 and Ssa289 had stutter patterns that helped to extrapolate missing allele sizes. All of the loci had clearly defined patterns in the alleles. For those alleles that had a more complex band pattern

(Omy77 and Omy207), the top half or the bottom half of the pattern was used to discriminate alleles in heterozygotes with overlaying patterns.

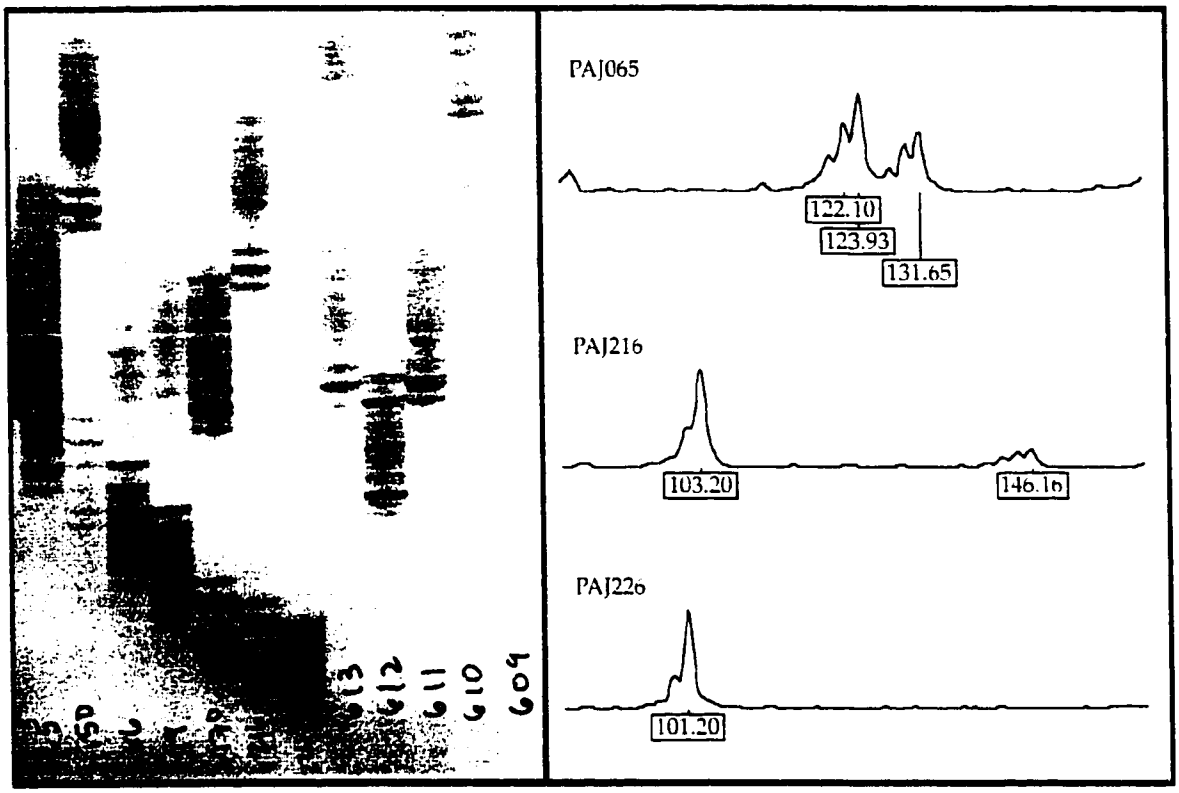


a

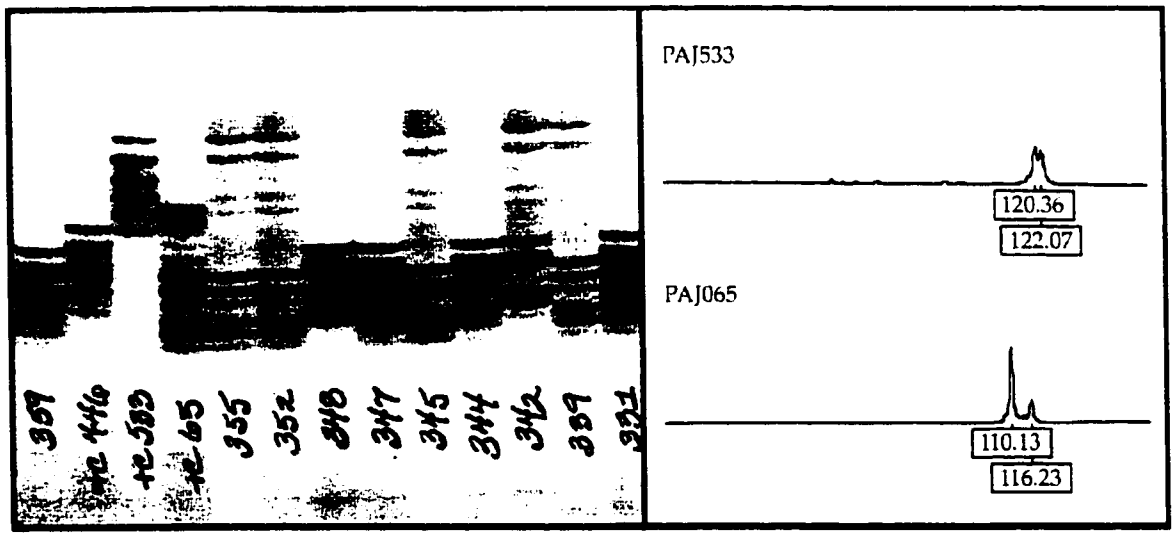


b

Fig. C-1. Alleles of a) Ssa 14 and b) Oneμ2. Results from an ABI automatic sequencer (right) were used as size standards (+c) on silver stained gels (left).



a



b

Fig. C-2. Alleles of a) Omy 207 and b) Ssa 289. Results from an ABI automatic sequencer (right) were used as size standards (+c) on silver stained gels (left).

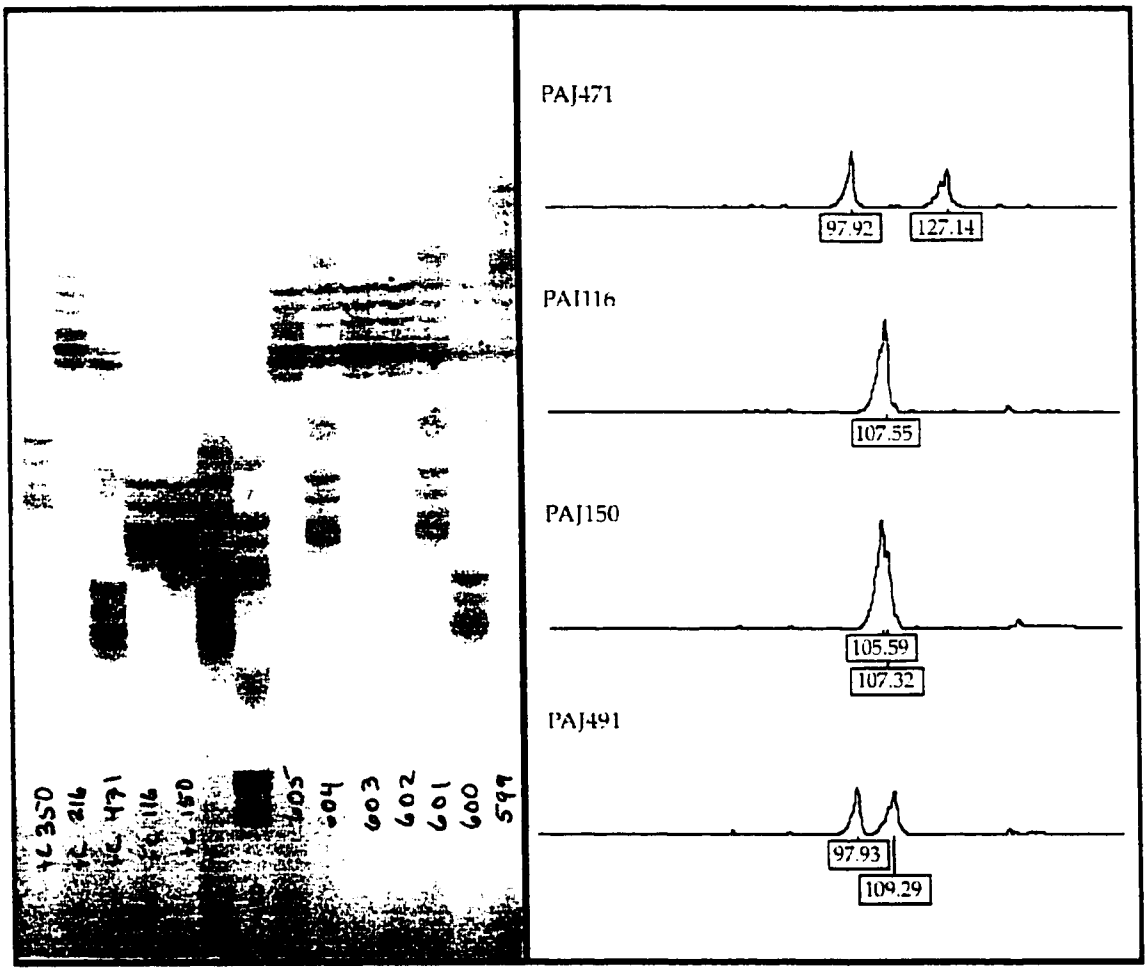


Fig. C-3. Alleles of a) Omy77. Results from an ABI automatic sequencer (right) were used as size standards (+c) on silver stained gels (left).

Appendix D

Genetic Analysis: Software Packages and Options Used

GENEPOP Version 3.1b (Raymond and Rousset 1995)

GENEPOP was used to compute basic information on the data set, including allelic frequencies, and exact tests for Hardy-Weinberg equilibrium and for genotypic disequilibrium among pairs of loci.

INPUT3D

A three-digit input file using the length, in base pairs, of microsatellite alleles was converted to a two digit coded input file in the subroutine INPUT3D.

Option 5 Basic information

This option was used to calculate allele frequencies, and observed and expected number of homozygotes and heterozygotes.

Option 1 Hardy-Weinberg exact tests

Suboption 3 Probability-test

The null hypothesis is:

Ho = "random union of gametes".

The "exact HW test" of Haldane (1954), Weir (1990), Guo and Thompson (1992) and others is used in this test. The probability of the observed sample is used to define the

rejection zone, and the P value of the test corresponds to the sum of the probabilities of all tables with the same or lower probability. A Markov chain method to estimate without bias the exact P value of this test (Guo and Thompson 1992) was used with the following settings:

Dememorization	1000
Batches	100
Iterations/batch	1000

Option 2 Genotypic linkage disequilibrium

Suboption 1 Test for each pair of loci in each population

The null hypothesis is:

Ho = "Genotypes at one locus are independent from geneotypes at the other locus."

Contingency tables for all pairs of loci in each population are created, and then a probability test (or Fisher exact test) is performed for each table using a Markov chain. For each locus pair within each population, the unbiased estimate of the P value is indicated, as well as the standard error. A global test (Fisher's method) for each pair of loci is performed across populations. The settings used were:

Dememorization	5000
Batches	100
Iterations	1000

PHYLIP (Phylogeny Inference Package) Version 3.5c

PHYLIP is a collection of software programs, most of them written and compiled by Joseph Felsenstein at the University of Washington. It has excellent documentation. GENDIST was used to calculate the Cavalli-Sforza chord distance D_C. CONTML was used to generate a neighbor-joining (NJ) tree based on the model of Edwards and Cavalli-Sforza. CONTML also generated trees with four loci, deleting Omy77 and Omy207 in turn. NEIGHBOR was used to generate a neighbor-joining tree using the distance Rho.

GENDIST

This program in PHYLIP was used to compute Cavalli-Sforza's chord distance (Cavalli-Sforza and Edwards 1967) from the set of allelic frequencies generated in GENEPOP. This distance assumes that differences between populations arise primarily from genetic drift. It does not assume population sizes have remained equal and constant in all populations. The setting used were:

A	Input file contains all alleles at each locus?	Yes
N	Use Nei's genetic distance?	No
C	Use Cavalli-Sforza chord measure?	Yes
R	Use Reynolds genetic distance?	No
L	Form of distance matrix?	Lower triangle
M	Analyze multiple data sets?	No
0	Terminal type?	ANSI
1	Print indications of progress of run?	Yes

CONTML (Gene Frequencies and Continuous Characters Maximum Likelihood Method) CONTML estimates phylogenies by the restricted maximum likelihood method based on the Brownian motion model (Felsenstein 1981). It gives an approximation of Cavalli-Sforza and Edwards' chord distance(Cavalli-Sforza and Edwards 1967) and takes that to give distance between particles undergoing pure Brownian motion. The input files were allelic frequencies from GENEPOP for five loci and for four loci. The settings used:

U	Search for best tree?	Yes
C	Gene frequencies or continuous characters?	Gene frequencies
A	Input file has all alleles at each locus?	Yes
O	Outgroup root?	Yes, use as outgroup species 14 (Sprig Lake)
G	Global rearrangements?	Yes
J	Randomize input order of species?	Yes
	Random seed	4273
	Number of times to jumble	10
M	Analyze multiple data sets?	No
0	Terminal type	ANSI
1	Print out the data at start of run	No
2	Print indications of progress of run	Yes
3	Print out tree	Yes
4	Write out trees onto tree file?	Yes

NEIGHBOR - Neighbor-Joining and UPGMA methods

This program generates neighbor-joining trees from distance matrices (Saitou and Nei 1987). It constructs a tree by successively clustering lineages, setting branch lengths as the lineages join. The input file was a distance matrix for Rho from RSTCALC. The settings used were:

N	Neighbor-joining or UPGMA tree?	Neighbor-joining
O	Outgroup root?	Yes, use as outgroup species 14
L	Lower-triangular data matrix?	Yes
R	Upper-triangular data matrix?	No
S	Subreplicates?	No
J	Randomize input order of species?	Yes, random seed 4273
		Jumble 10 times
M	Analyze multiple data sets?	No
0	Terminal type?	ANSI
1	Print out the data at start of run	No
2	Print indications of progress of run	Yes
3	Print out tree	Yes
4	Write out trees onto tree file?	Yes

DISPAN (Genetic Distance and Phylogenetic Analysis)

This program was developed by Tatsuya Ota at Pennsylvania State University.

DISPAN was used with GENEPOP allelic frequencies to compute average heterozygosity and its standard error for each population, D_A distances (Nei *et al.* 1983) and a neighbor-joining tree (Saitou and Nei 1987). Bootstrap tests (Efron 1982, Felsenstein 1985) were performed for this tree. The setting used were:

-da D_A distance

-g estimation of Hs, Ht, and Gst for each locus

-r2000 2000 bootstrap replications

-s426 random seed number

-tn NJ tree

NJBAFD (Neighbor-Joining Tree Construction from Allele Frequency Data)

NJBAFD was written by Naoko Takezaki at the National Institute of Genetics in Japan. NJBAFD2 is the win-32 bit version of the program for use in WINDOWS 95. It was used to compute D_A (Nei *et al.* 1983) and $(\delta\mu)^2$ (Goldstein *et al.* 1995) and to generate neighbor-joining trees from these distance matrices.

The alleles in the input file were an estimate of the number of nucleotide repeats. Primer lengths were subtracted from each allele, then the result divided by two to estimate the number of repeat units in each allele. The settings used in NJBAFD2 were:

-b2000 2000 bootstrap replications

-d0 or -d2 distance measures D_A or Dmyu

The cnvtre option was used to convert the treefile to a New Hampshire format to be used in TREEVIEW.

RSTCALC 2.2

Simon Goodman of the University of Edinburgh developed this program.

RSTCALC calculates Rho, an unbiased estimator of Slatkin's R_{ST} (Slatkin 1995). These values are unbiased with respect to differences in sample size between populations and difference in variance between loci. The data was transformed by expressing alleles in terms of standard deviations from the global mean rather than repeat unit number. Each locus then had a global mean allele size of zero and a standard deviation of one.

$$Ys = (Y - GM) / \text{std dev}$$

where Ys is the standardized value of allele Y, Y is allele (n) at locus (l), GM is the mean allele size in repeat units for locus (l) over the whole data set. The std dev is the standard deviation in allele size in terms of repeat units for locus (l) over the whole data set. To adjust for differences in sample size between populations, the variance components were obtained using conventional statistical approaches:

$$Rho = Sb/(Sb + SW)$$

where Sb is the component of variance that is between populations.

Bootstrapping was performed to obtain 95% confidence intervals. The setting used were:

1,000 iterations

2,000 bootstraps

Where negative values of Rho were generated, the within population variance in allele size was bigger than the among population variance in allele size. No differentiation in the populations could be determined.

TREEVIEW Version 1.5 (Page 1996)

TREEVIEW is a program for displaying and printing phylogenies.

Appendix E

Allelic Frequencies for Microsatellite Loci in *O. mykiss* Populations in the Pajaro and San Lorenzo River Watersheds.

Allelic frequencies for microsatellite loci in *O. mykiss* in the Pajaro and San Lorenzo rivers.

Ssa14																		
Fragment Length	126	128	130	134	136	138	140	142	144	146	148	150	152	154	156	158	160	162
Bodfish up		0.133	0.050	0.017		0.100			0.017	0.050	0.033	0.217	0.317					0.067
Dos Picachos			0.015			0.015		0.030	0.061	0.015	0.045	0.545	0.242	0.015				0.015
Llagas low									0.067		0.383	0.383	0.067	0.033				0.067
Uvas up	0.017	0.017		0.017						0.086	0.103	0.345	0.293	0.121				
Llagas up		0.014								0.043	0.029	0.557	0.300		0.014			0.043
Bodfish low		0.047	0.094	0.016	0.016	0.031			0.063	0.063	0.063	0.188	0.219	0.156	0.031			0.016
GreenValley										0.150	0.167	0.267	0.367					0.050
Casserly			0.183				0.067		0.017		0.033	0.400	0.167	0.100	0.017			0.017
Pacheco			0.067	0.017	0.133		0.033	0.067			0.083	0.350	0.200	0.050				
Little Arthur MtM			0.016		0.032				0 016		0.016	0.774	0.113	0.032				
Little Arthur Red			0.194	0.065	0.016				0.065			0.532	0.113		0.016			
Corralitos			0.067	0.017	0.017				0.033	0.133	0.017	0.317	0.200	0.133	0.033			0.033
Uvas low			0.100	0.017	0.083				0.017	0.017		0.350	0.183	0.167	0.033	0.033		
Zayante			0.017	0.050			0.017			0.050		0.650	0.150	0.017			0.017	0.033
Boulder			0.017						0.067	0.033	0.050	0.350	0.350	0.100	0.017			0.017
Sprig Lake	0.056			0.222	0.222		0.167		0.111	0.056		0.056				0.111		

Ssa289									
Fragment Length	110	112	114	116	118	120	122	124	128
Bodfish up	0.300	0.417	0.167			0.083			0.033
Dos Picachos	0.422	0.234		0.063		0.078	0.031		0.172
Llagas low	0.200	0.767	0.033						
Uvas up	0.328	0.241		0.017		0.172	0.138	0.017	0.086
Llagas up	0.200	0.471		0.014		0.217			0.043
Bodfish low	0.469	0.328	0.047	0.016		0.016	0.094		0.031
Green Valley	0.317	0.433	0.017	0.050		0.033	0.150		
Casserly	0.450	0.317		0.067	0.050	0.067	0.050		
Pacheco	0.233	0.433	0.117	0.100		0.083		0.033	
Little Arthur MtM	0.435	0.194	0.081	0.032		0.032	0.016	0.210	
Little Arthur Red	0.710	0.145		0.048		0.016	0.065	0.016	
Corralitos	0.387	0.274	0.048	0.016		0.097	0.161		0.016
Uvas low	0.333	0.333	0.100	0.067		0.100	0.050	0.017	
Zayante	0.450	0.283	0.017			0.083	0.100	0.067	
Boulder	0.367	0.283	0.067	0.083		0.100	0.033	0.067	
Sprig Lake	0.182	0.273	0.227	0.136		0.091		0.091	

Allelic frequencies for microsatellite loci in *O. mykiss* in the Pajaro and San Lorenzo rivers.

Omy77																				
Fragment Length	81	94	98	100	102	104	106	108	110	112	116	120	127	129	131	133	135	137	139	141
Bodfish up			0.190	0.103		0.172		0.017	0.138				0.034	0.259					0.086	
Dos Picachos			0.250	0.063		0.063		0.031	0.016	0.500				0.047	0.031					
Llagas low			0.450			0.433				0.083				0.033						
Uvas up			0.167	0.150			0.133	0.200	0.167	0.050		0.017	0.067	0.017				0.017	0.017	
Llagas up			0.045	0.015				0.242	0.015	0.045			0.591					0.015	0.030	
Bodfish low			0.219	0.031		0.125		0.031	0.031	0.109			0.078	0.234		0.063		0.047	0.031	
Green Valley						0.467			0.133					0.283		0.177				
Casserly	0.117		0.133	0.017		0.383	0.050	0.017	0.183	0.033			0.050			0.017				
Pacheco			0.133	0.017	0.017	0.100		0.017	0.317				0.350	0.033					0.017	
Little Arthur MtM			0.167	0.083		0.333		0.017	0.150	0.017			0.017	0.100			0.017	0.017	0.067	0.017
Little Arthur Red	0.033		0.250	0.017		0.083	0.017	0.167	0.050				0.017	0.050		0.067			0.250	
Corralitos		0.067	0.100	0.083		0.067		0.133	0.117	0.133			0.067	0.083		0.033			0.117	
Uvas low	0.085	0.017	0.153	0.068		0.068	0.017	0.136	0.203	0.153			0.034			0.034			0.034	
Zayante			0.133	0.017		0.217		0.100	0.383	0.033			0.033			0.083				
Boulder	0.017		0.207	0.034		0.155		0.086	0.241	0.086			0.034	0.052		0.034		0.034	0.017	
Sprig Lake				0.333			0.167		0.042		0.458									

Allelic Frequencies for Microsatellite Loci in *O. mykiss* in the Pajaro and San Lorenzo Rivers.

Omy207																			
Fragment Length	98	100	102	104	106	108	110	112	114	116	118	120	122	124	126	128	130	132	134
Bodfish up		0.367	0.033	0.033	0.017		0.067		0.017		0.017	0.067						0.167	
Dos Picachos				0.078		0.047	0.094		0.141				0.109					0.141	
Llagas low				0.233					0.233			0.017	0.133			0.067			
Uvas up			0.017	0.069			0.034		0.121		0.086	0.121					0.017	0.190	0.034
Llagas up				0.014					0.100			0.029			0.100			0.314	
Bodfish low		0.031	0.063	0.047			0.031	0.016	0.109	0.031	0.109	0.031	0.031		0.047	0.031		0.141	
Green Valley		0.133	0.050		0.017				0.050									0.367	
Casserly		0.033		0.033			0.050		0.167		0.150	0.100		0.017	0.033			0.167	
Pacheco	0.017	0.100	0.017		0.017				0.067		0.083	0.167				0.050		0.217	
Little Arthur MtM	0.032	0.032		0.113			0.048		0.113	0.032			0.129	0.016					
Little Arthur Red		0.048		0.065	0.032		0.113		0.129	0.048		0.016			0.194				
Corralitos	0.032			0.065				0.016	0.274		0.048	0.113	0.016		0.016	0.016		0.097	
Uvas low		0.017		0.050	0.017			0.050	0.200		0.067	0.017	0.067					0.117	
Zayante				0.050	0.050				0.283		0.033	0.183	0.050		0.017		0.050	0.117	
Boulder		0.017	0.017	0.050	0.067				0.183		0.083	0.033	0.083		0.017	0.067		0.117	
Sprig Lake				0.227					0.045	0.318	0.091	0.045			0.136		0.091		0.045

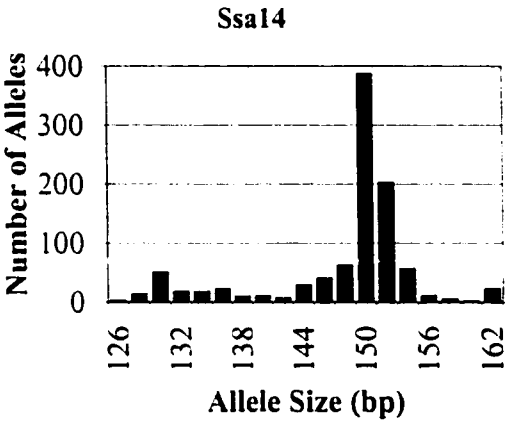
Fragment Length	136	138	140	146	148	150	154	156	158
Bodfish up	0.017	0.017		0.033			0.150		
Dos Picachos		0.016	0.047	0.078			0.250		
Llagas low			0.167					0.150	
Uvas up			0.052	0.172			0.052	0.034	
Llagas up	0.014	0.014		0.043	0.243		0.057	0.071	
Bodfish low	0.016			0.047	0.031		0.156	0.016	0.016
Green Valley	0.200	0.033	0.083	0.067					
Casserly	0.050		0.033	0.050			0.117		
Pacheco	0.017	0.050	0.067	0.083			0.050		
Little Arthur MtM	0.016		0.016	0.290	0.081		0.065		0.016
Little Arthur Red				0.032			0.306		0.016
Corralitos		0.048	0.048	0.032	0.016		0.161		
Uvas low		0.017	0.133	0.033	0.083		0.067		0.067
Zayante		0.050	0.017	0.017	0.017		0.067		
Boulder		0.067			0.050	0.017	0.117		0.017
Sprig Lake									

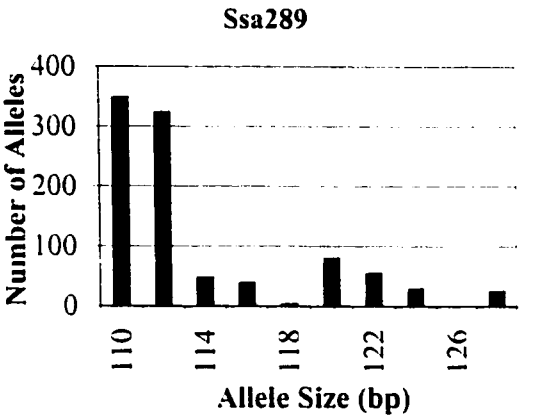
Allelic frequencies for microsatellite loci in *O. mykiss* in the Pajaro and San Lorenzo rivers

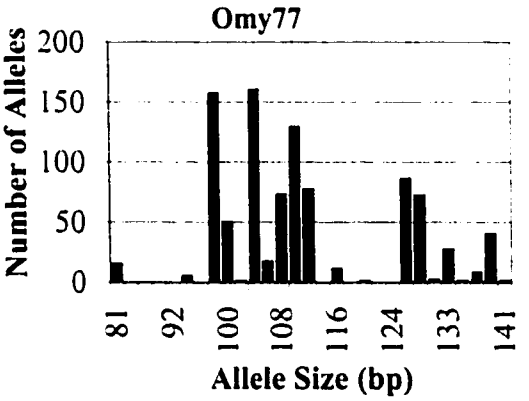
Oneμ2																					
Frag. Lengh	200	206	208	210	212	214	222	224	226	230	232	236	238	240	242	244	246	248	250	252	254	256
Bodfish up								0.050	0.017		0.100	0.017	0.217							0.017		0.150
Dos Picachos								0.017		0.100	0.117		0.300		0.017						0.067	0.033
Llagas low													0.433				0.050				0.333	0.033
Uvas up			0.033	0.017	0.017					0.050			0.183		0.117		0.050	0.033			0.033	0.217
Llagas up											0.143	0.014	0.257								0.200	0.029
Bodfish low			0.016					0.016		0.047	0.234	0.125	0.031	0.063	0.078		0.031		0.016	0.031	0.016	0.063
Green Valley								0.267	0.033				0.200	0.267				0.050			0.033	
Casserly								0.033		0.083	0.200		0.050	0.033	0.017					0.017	0.050	0.233
Pacheco								0.050			0.200	0.067	0.133	0.033	0.017	0.017	0.017		0.083	0.017		0.033
Lit.Art. MtM			0.032		0.016			0.145			0.145	0.032	0.048	0.016	0.032	0.016				0.016	0.016	0.194
Lit.Art. Red				0.016	0.016			0.032			0.242		0.032		0.210		0.065					0.048
Corralitos				0.017		0.017		0.067	0.017		0.100		0.183	0.017	0.067	0.033	0.050			0.017	0.050	0.100
Uvas low	0.016	0.016						0.016	0.098	0.115	0.180		0.033	0.033	0.033	0.016	0.016		0.016	0.016	0.049	0.033
Zayante								0.100	0.033		0.167	0.050	0.150	0.050		0.017		0.117		0.033		0.067
Boulder			0.033					0.083	0.033	0.017	0.133	0.033	0.177		0.017		0.067	0.033		0.067		0.033
Sprig Lake		0.150			0.150		0.250							0.050		0.300	0.050					

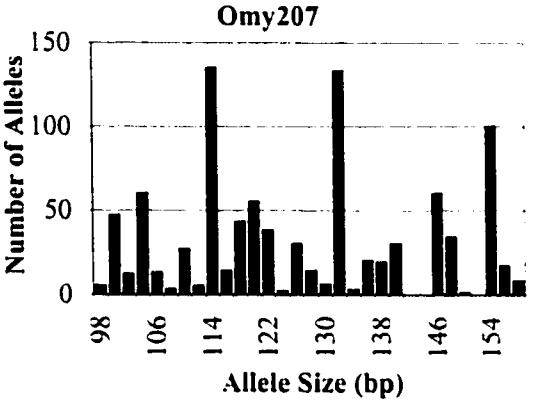
Frag. Lengh	258	260	262	264	266	268	270	272	274	276	278	284	286	290	292	308	312	354	364	372	378	382
Bodfish up	0.117		0.167	0.017	0.050	0.017										0.033			0.033			
Dos Picachos	0.033		0.083		0.067	0.067				0.050	0.050											
Llagas low		0.067	0.067	0.017																		
Uvas up		0.017	0.067	0.017		0.017			0.033	0.017	0.067		0.017									
Llagas up		0.043	0.057						0.043		0.200					0.014						
Bodfish low			0.109	0.016	0.016						0.016	0.016										0.063
Green Valley				0.033						0.050	0.067											
Casserly	0.017	0.017	0.083	0.083						0.017	0.050										0.017	
Pacheco	0.117	0.117				0.017					0.083											
Lit.Art. MtM	0.032		0.081						0.129	0.048												
Lit.Art. Red	0.145	0.032	0.016							0.016	0.081			0.048								
Corralitos	0.017		0.067	0.067				0.017				0.017		0.017		0.033		0.017	0.017			
Uvas low	0.016	0.033	0.066	0.049						0.016		0.049			0.016		0.016					0.049
Zayante		0.017	0.183			0.017																
Boulder	0.017	0.033	0.117	0.050			0.017			0.017		0.033							0.017	0.033		
Sprig Lake							0.050															

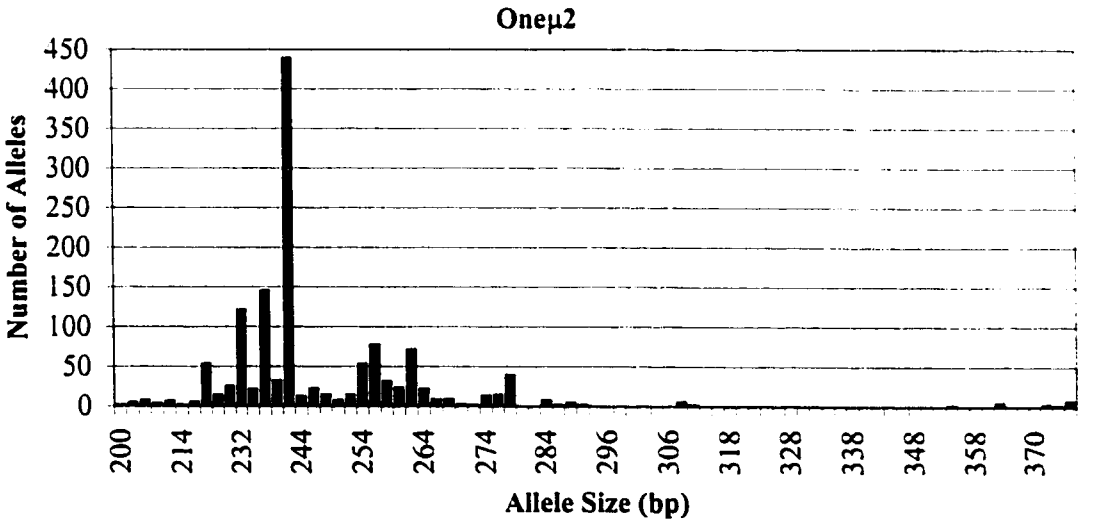
The number of alleles in each allele size (base pairs) for all populations



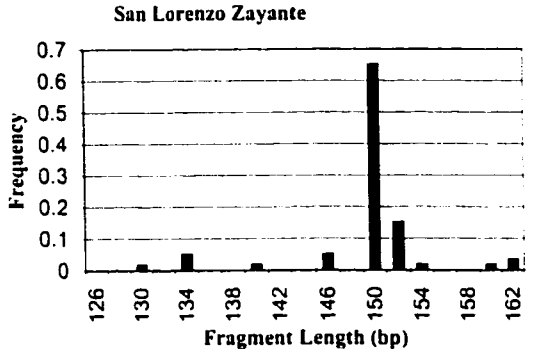


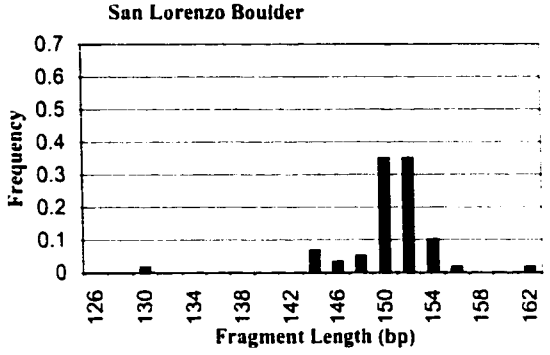


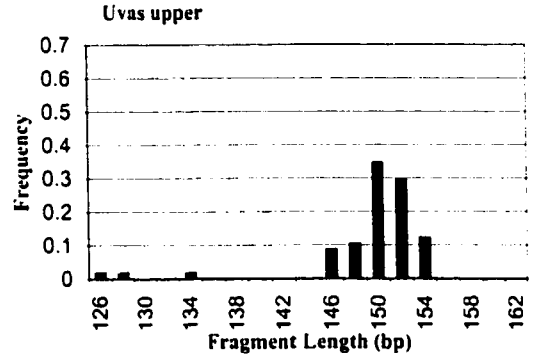


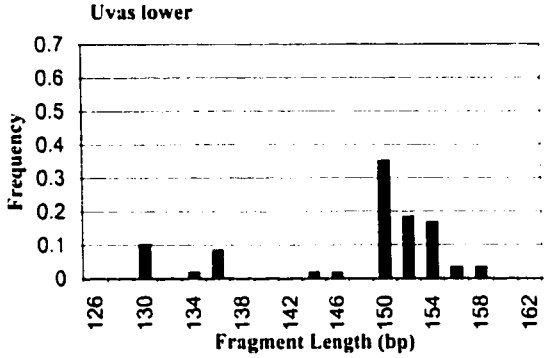


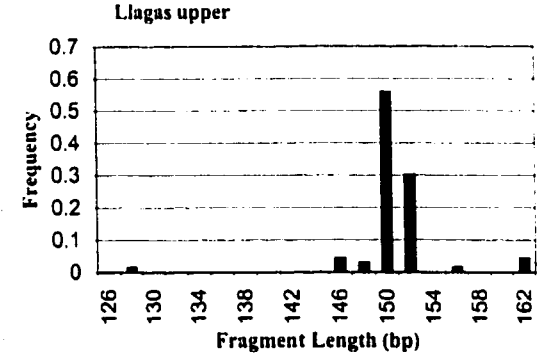
Ssa 14 Frequency Distributions for Each Population

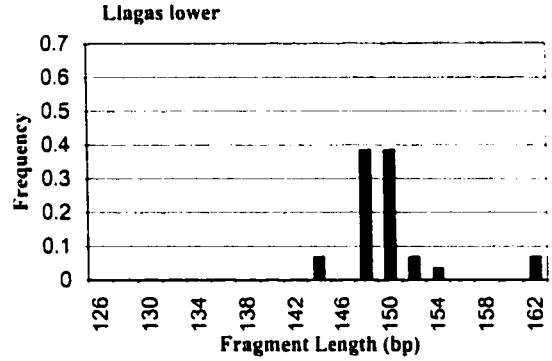


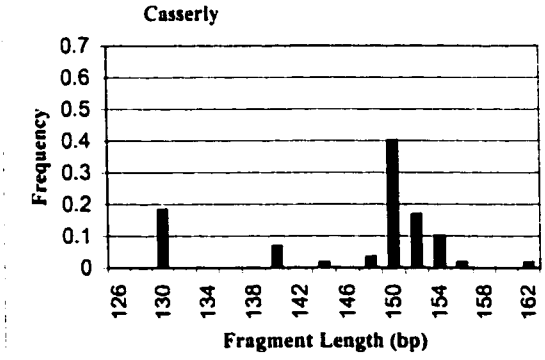


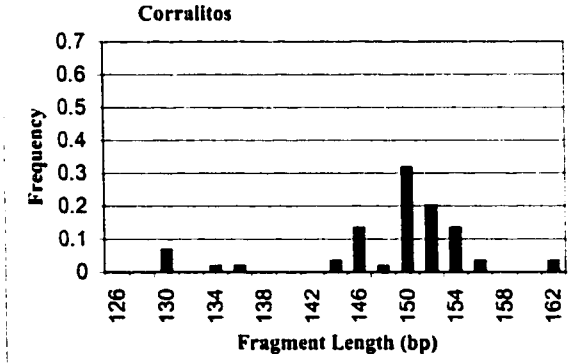




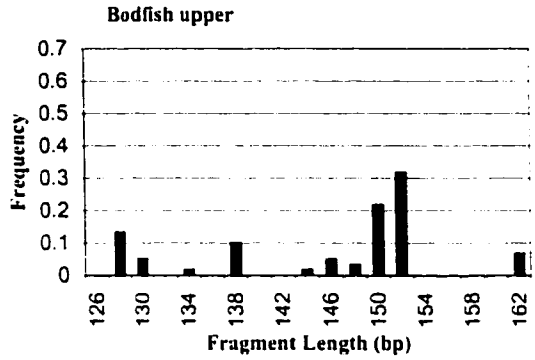


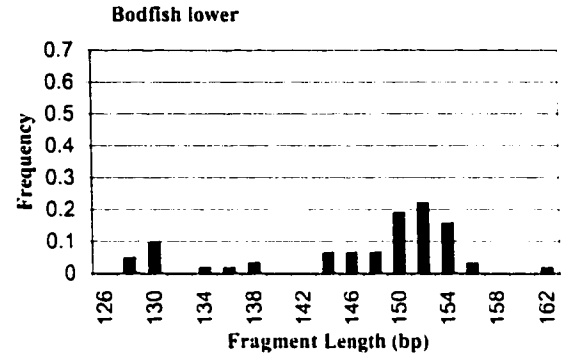


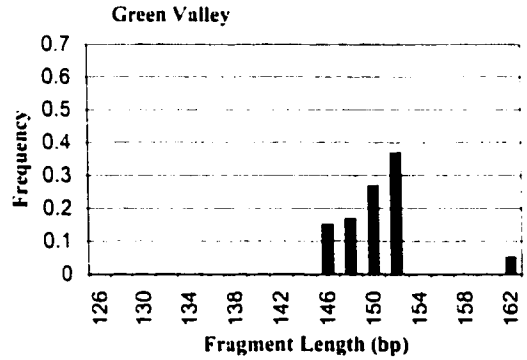


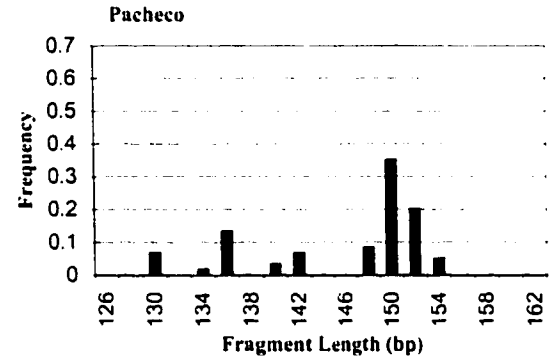


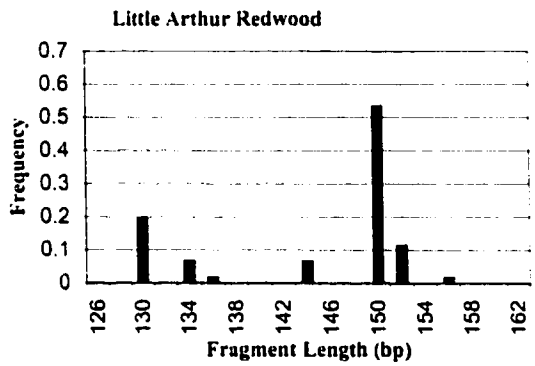
Ssa14 Frequency Distributions for Each Population

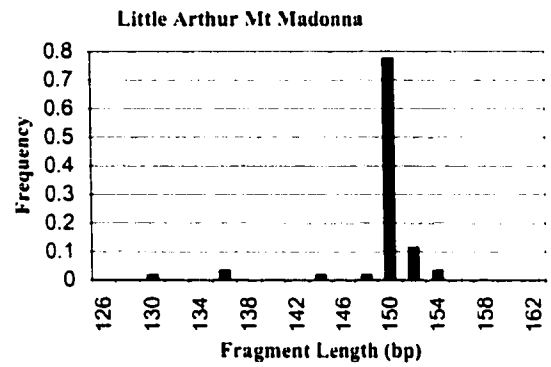


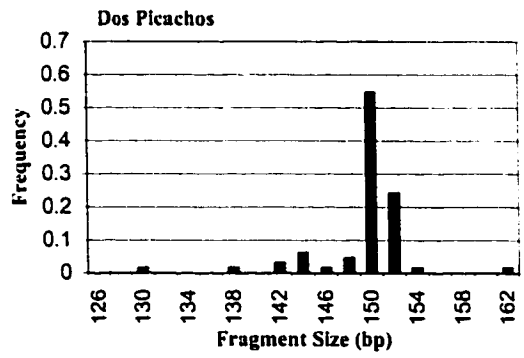


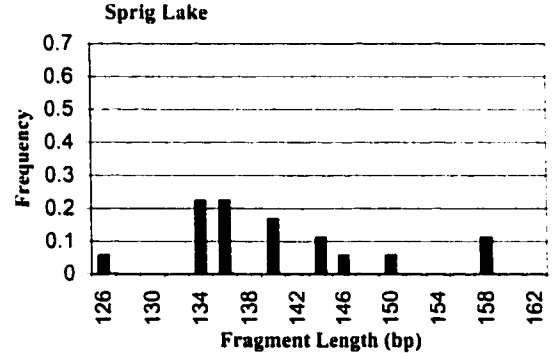




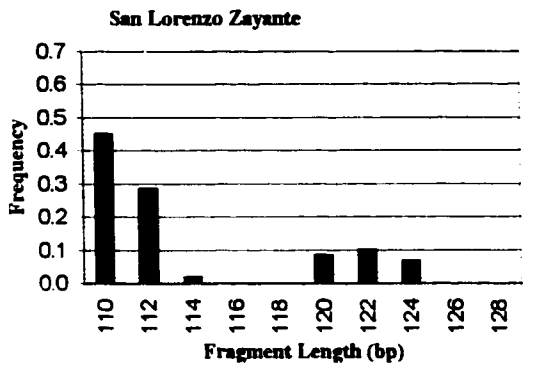


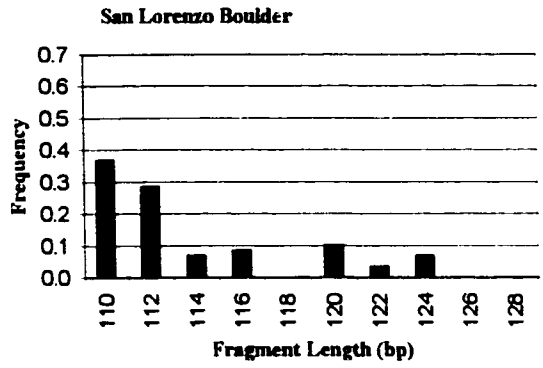


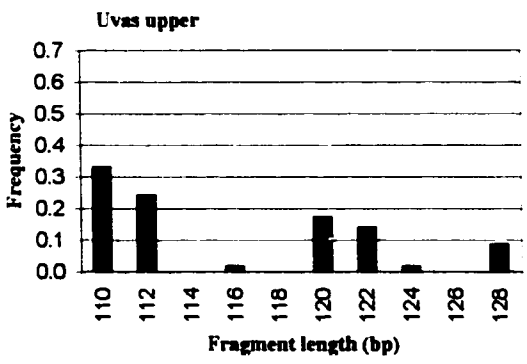


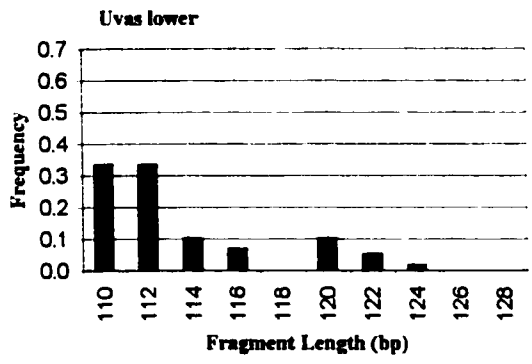


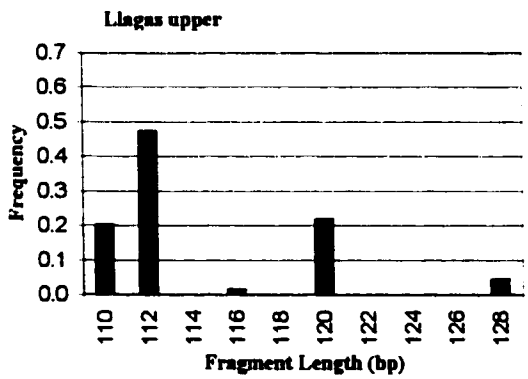
Ssa289 Frequency Distributions for Each Population

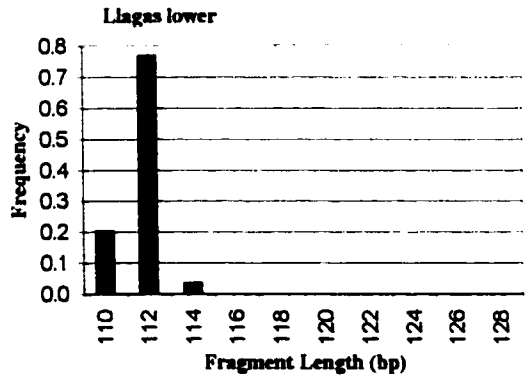


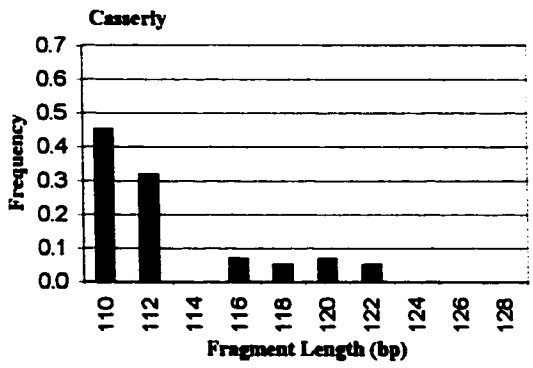


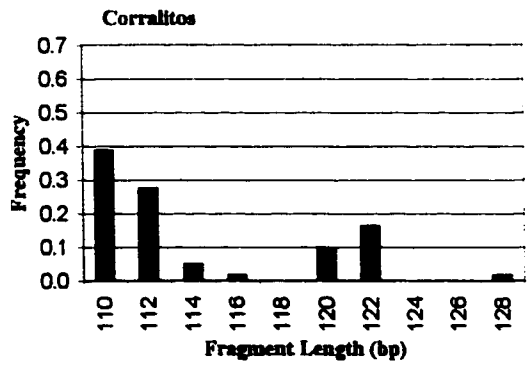




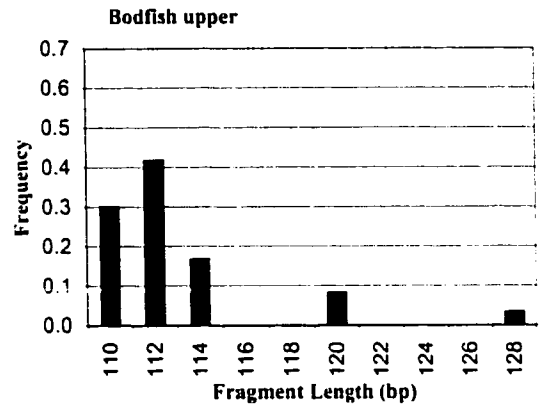


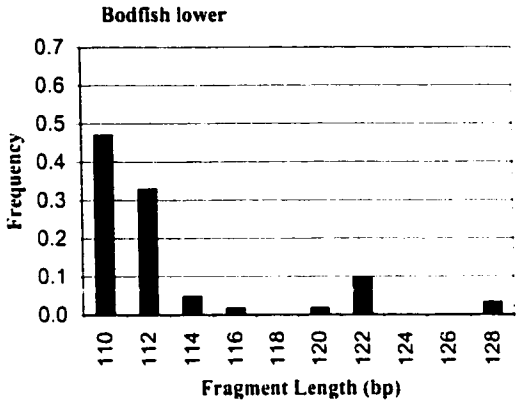


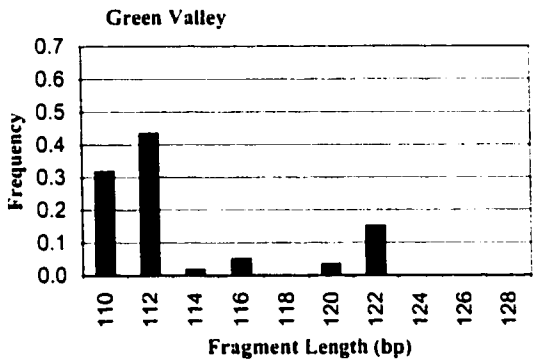


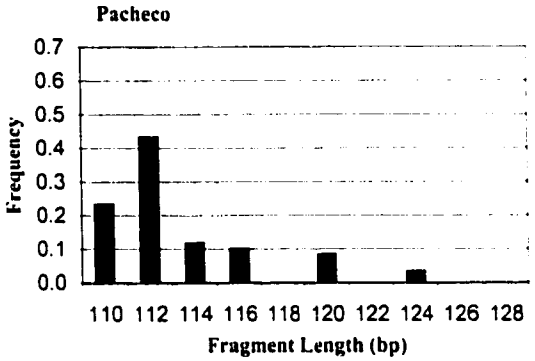


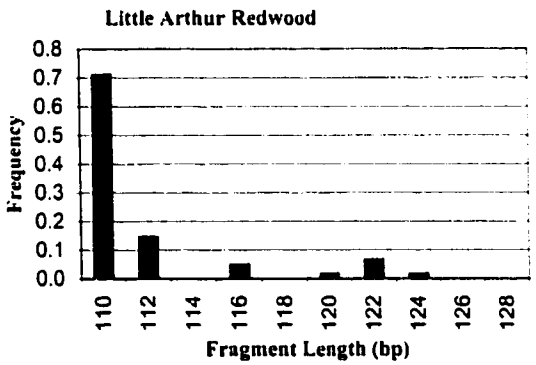
Ssa289 Frequency Distributions for Each Population

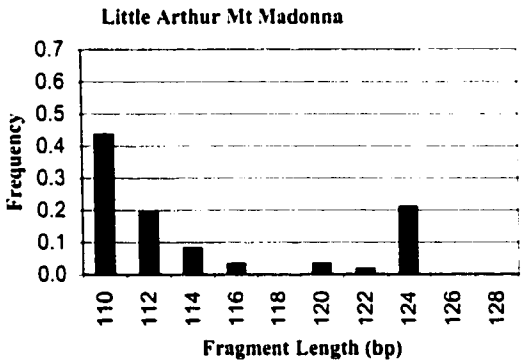


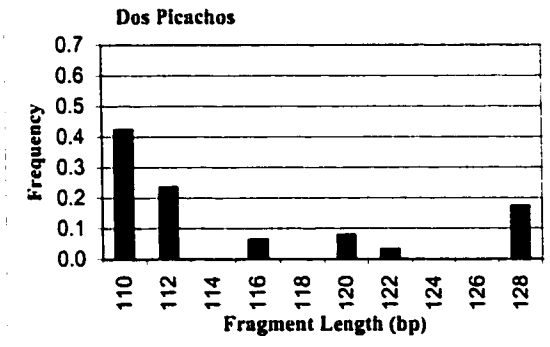


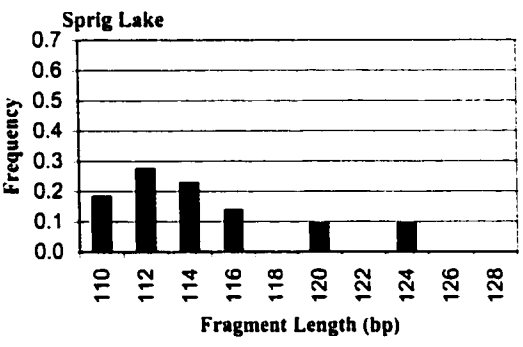




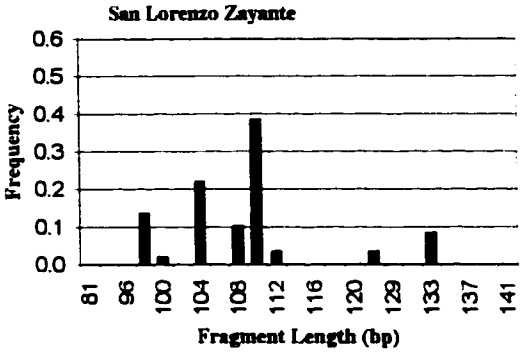


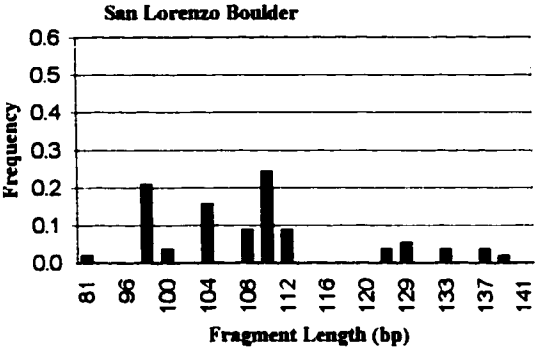


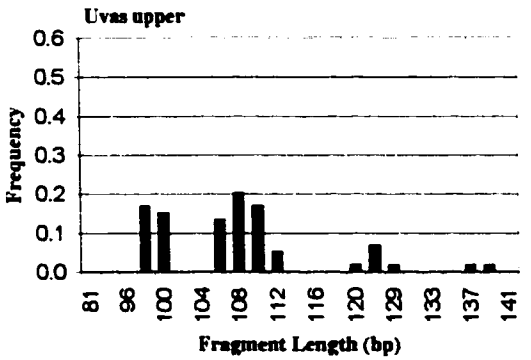


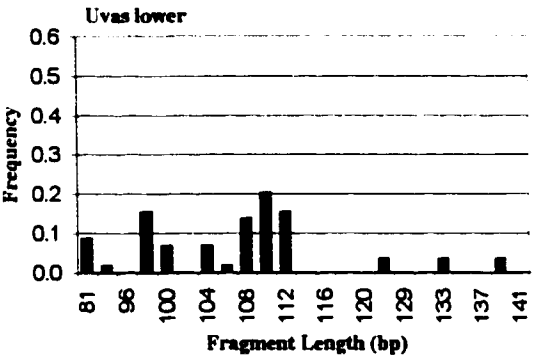


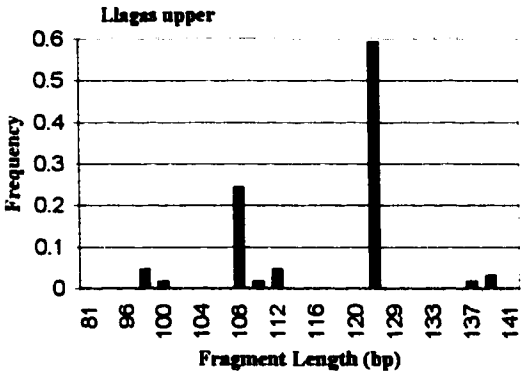
Omy77 Frequency Distributions for Each Population

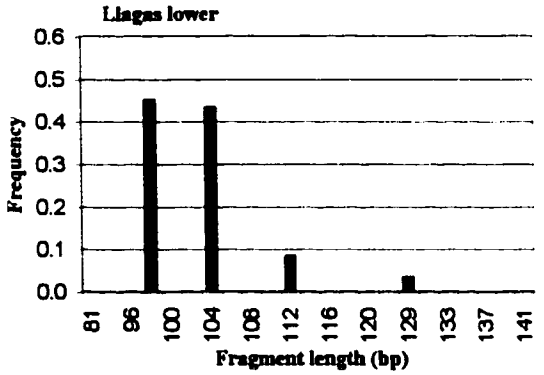


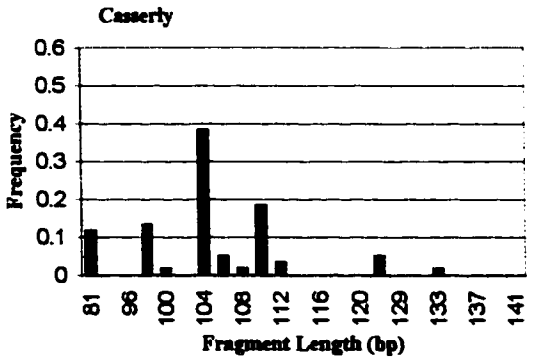


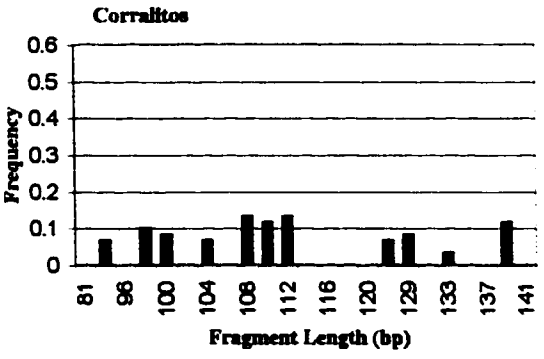




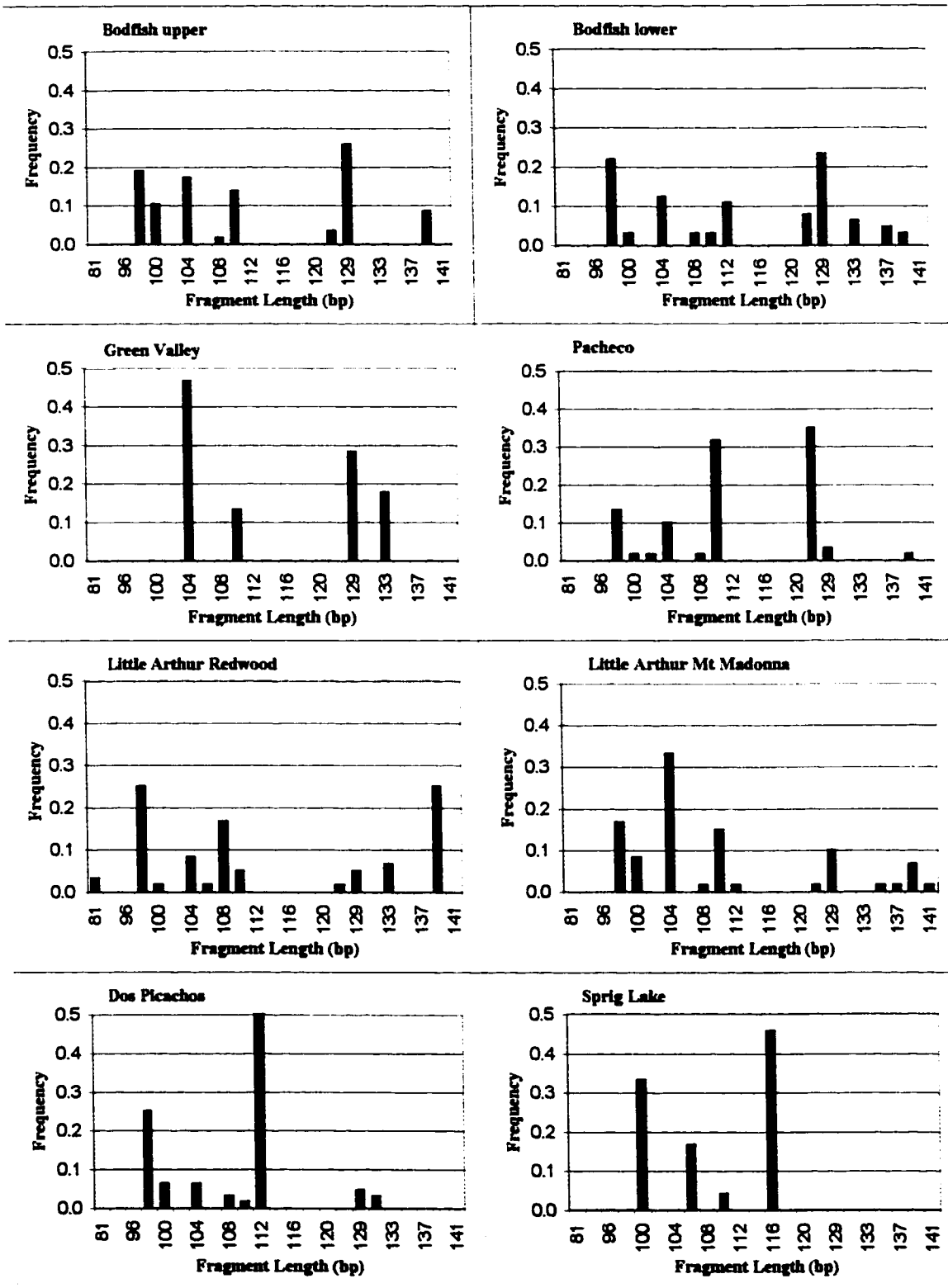




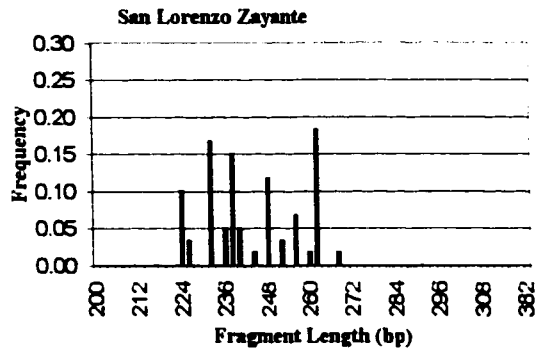


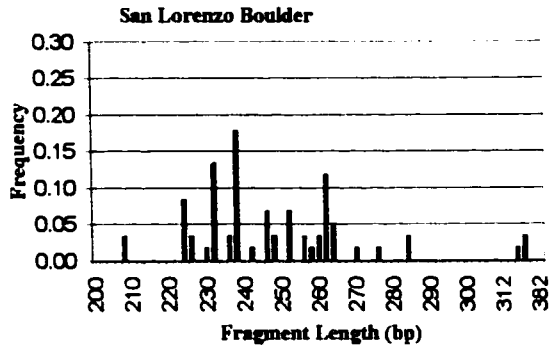


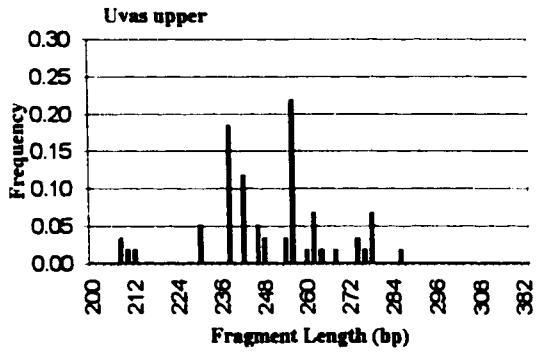
Omy77 Frequency Distributions for Each Population

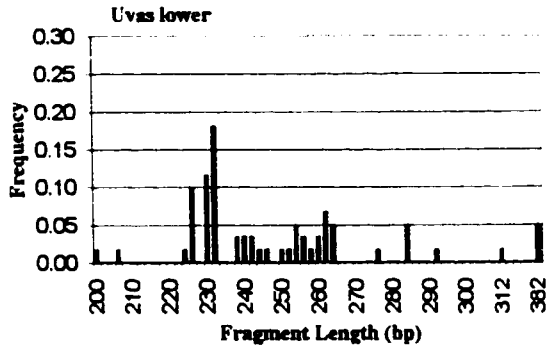


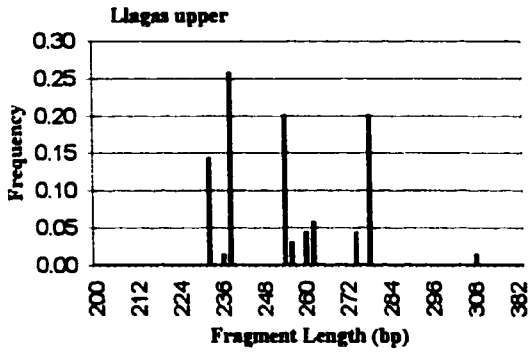
Oneμ2 Frequency Distributions for Each Population

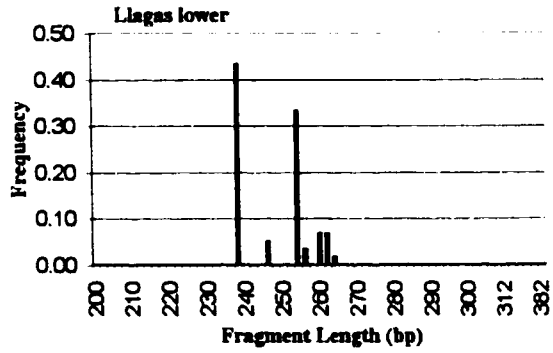


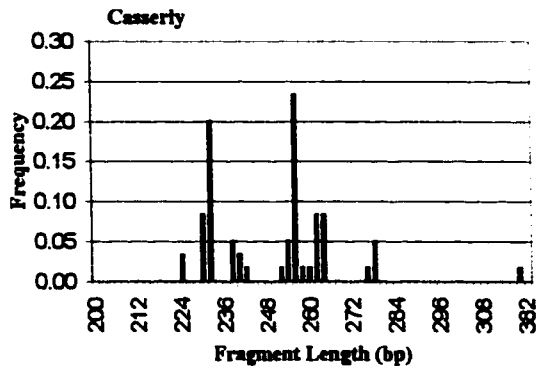


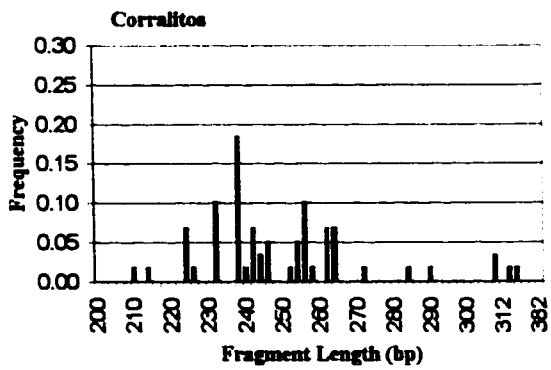




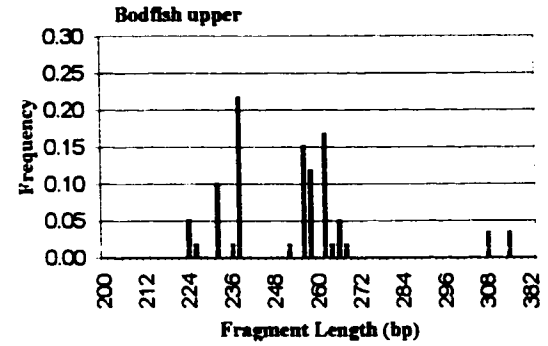


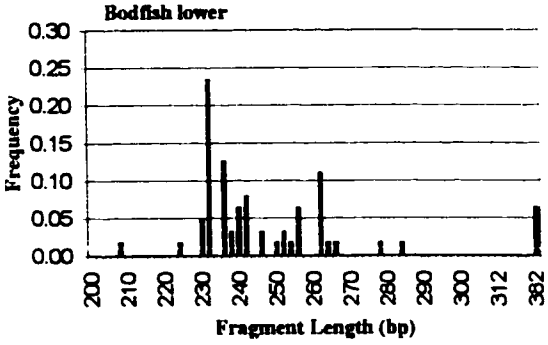


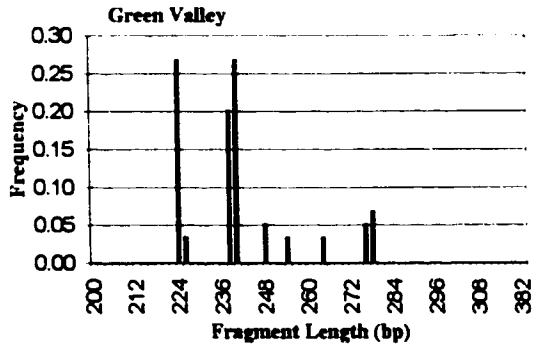


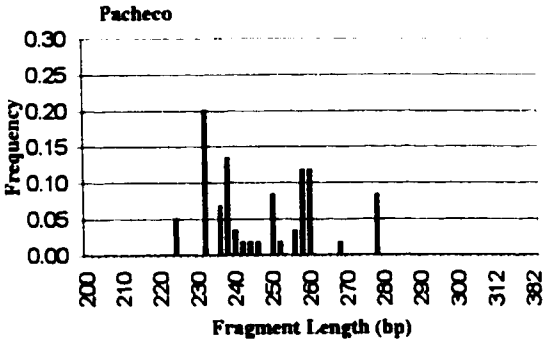


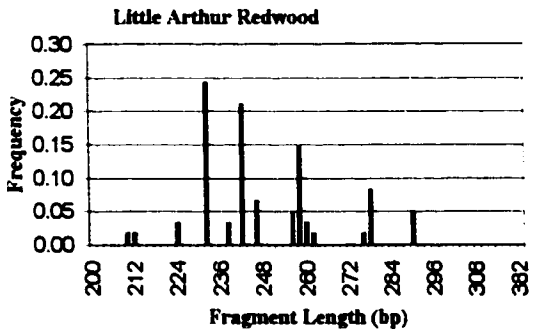
Oneμ2 Frequency Distributions for Each Population

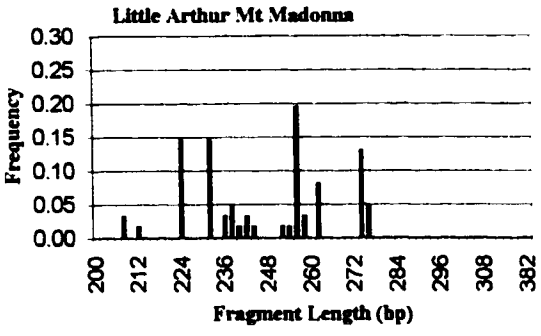


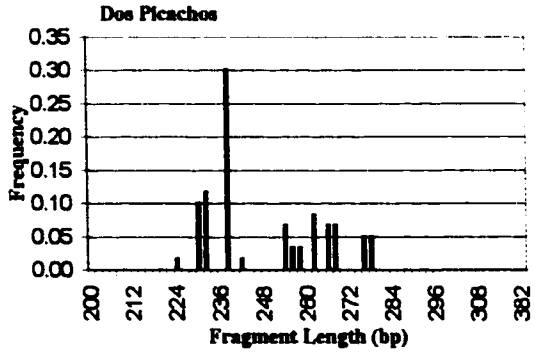


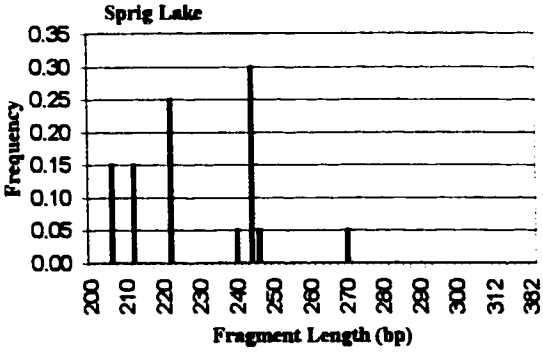




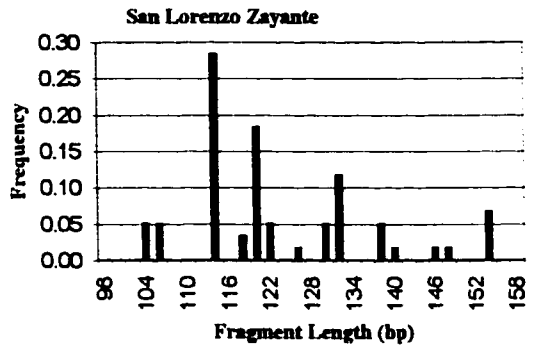


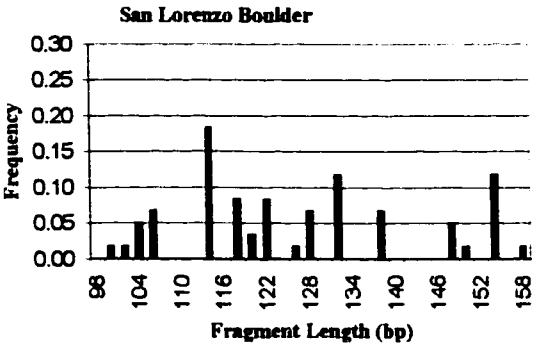


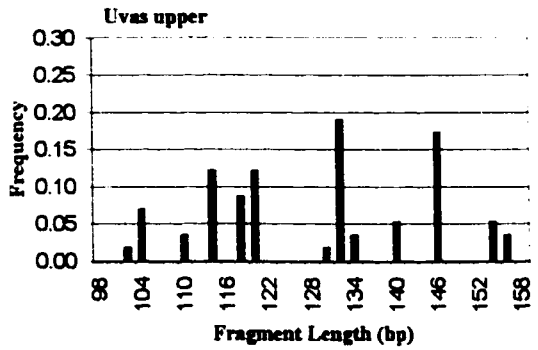


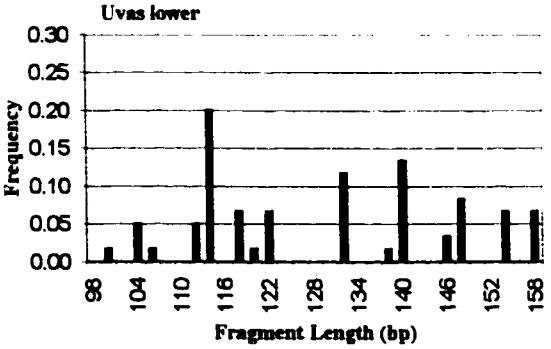


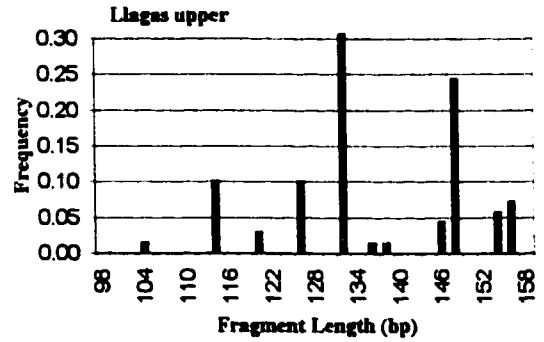
Omy207 Frequency Distributions for Each Population

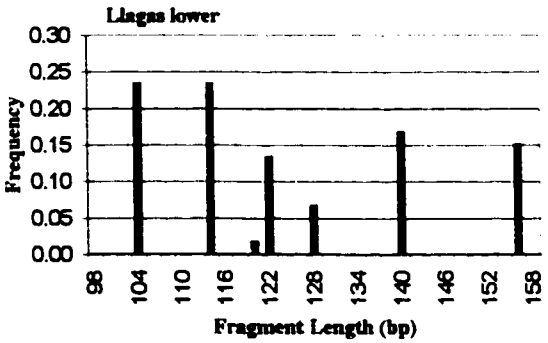


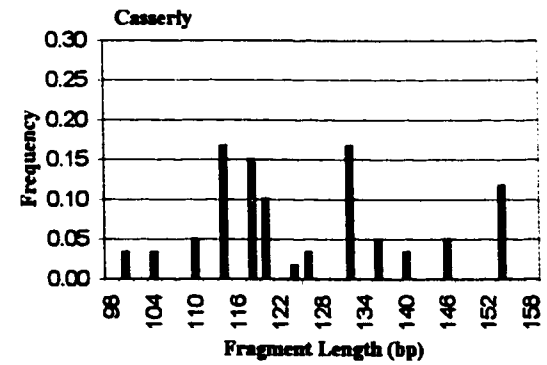


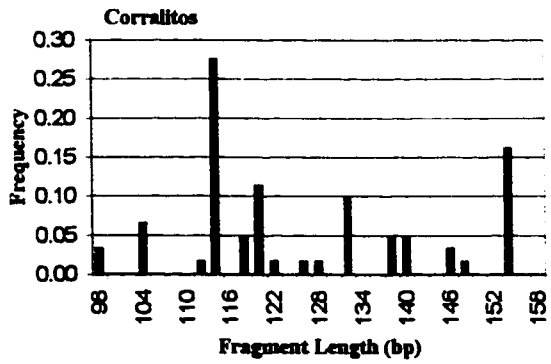




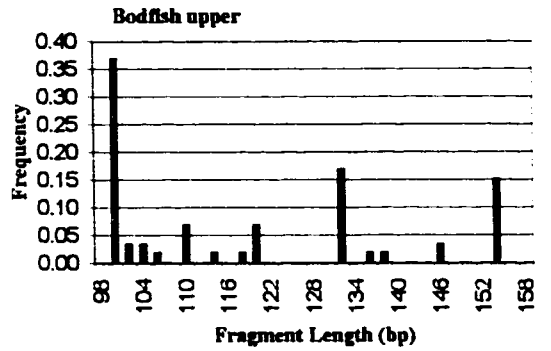


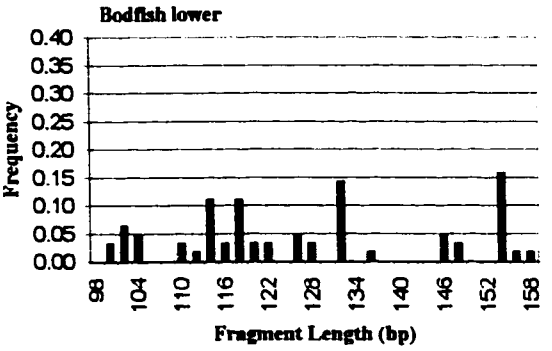


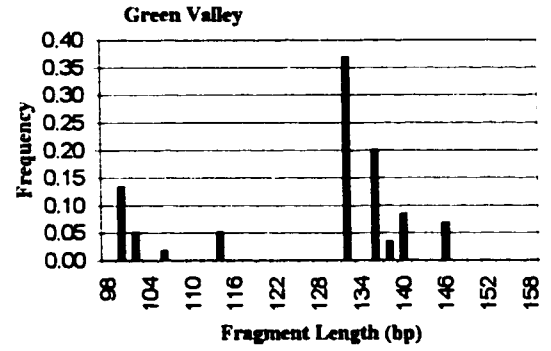


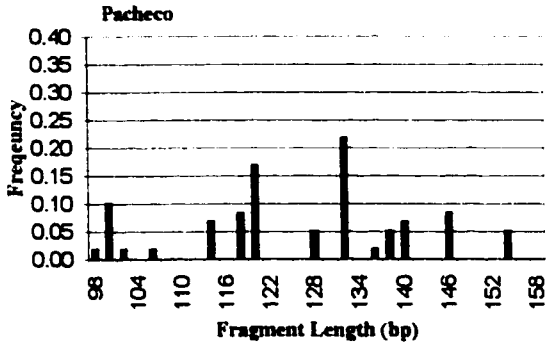


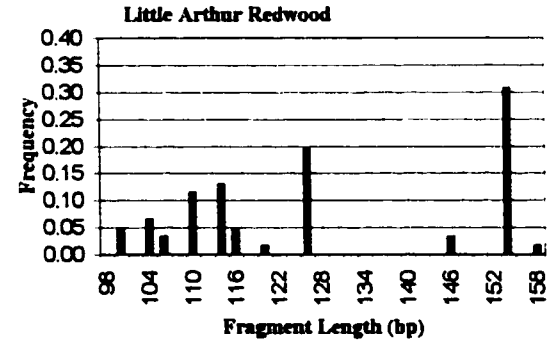
Omy207 Frequency Distributions for Each Population

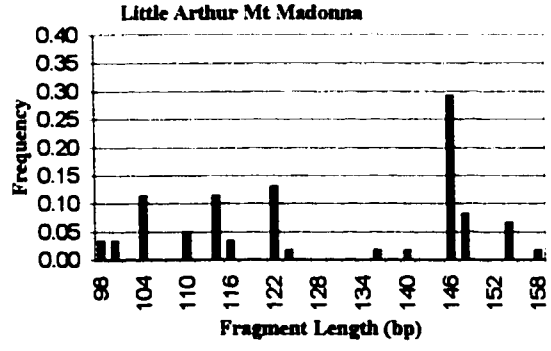


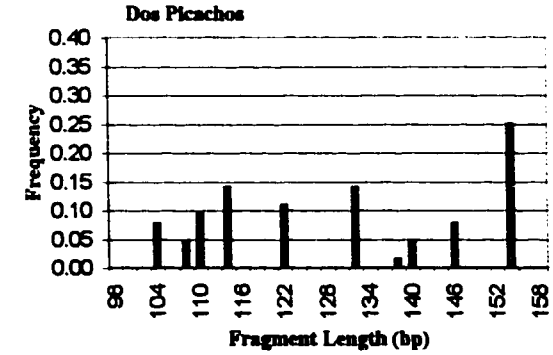


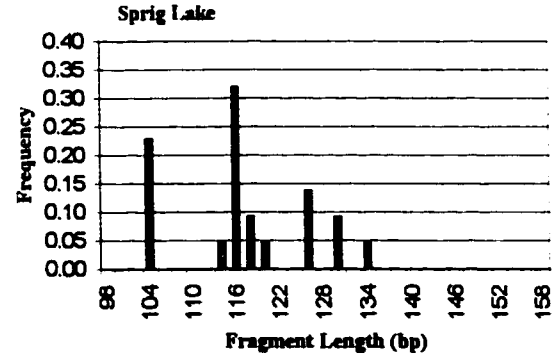








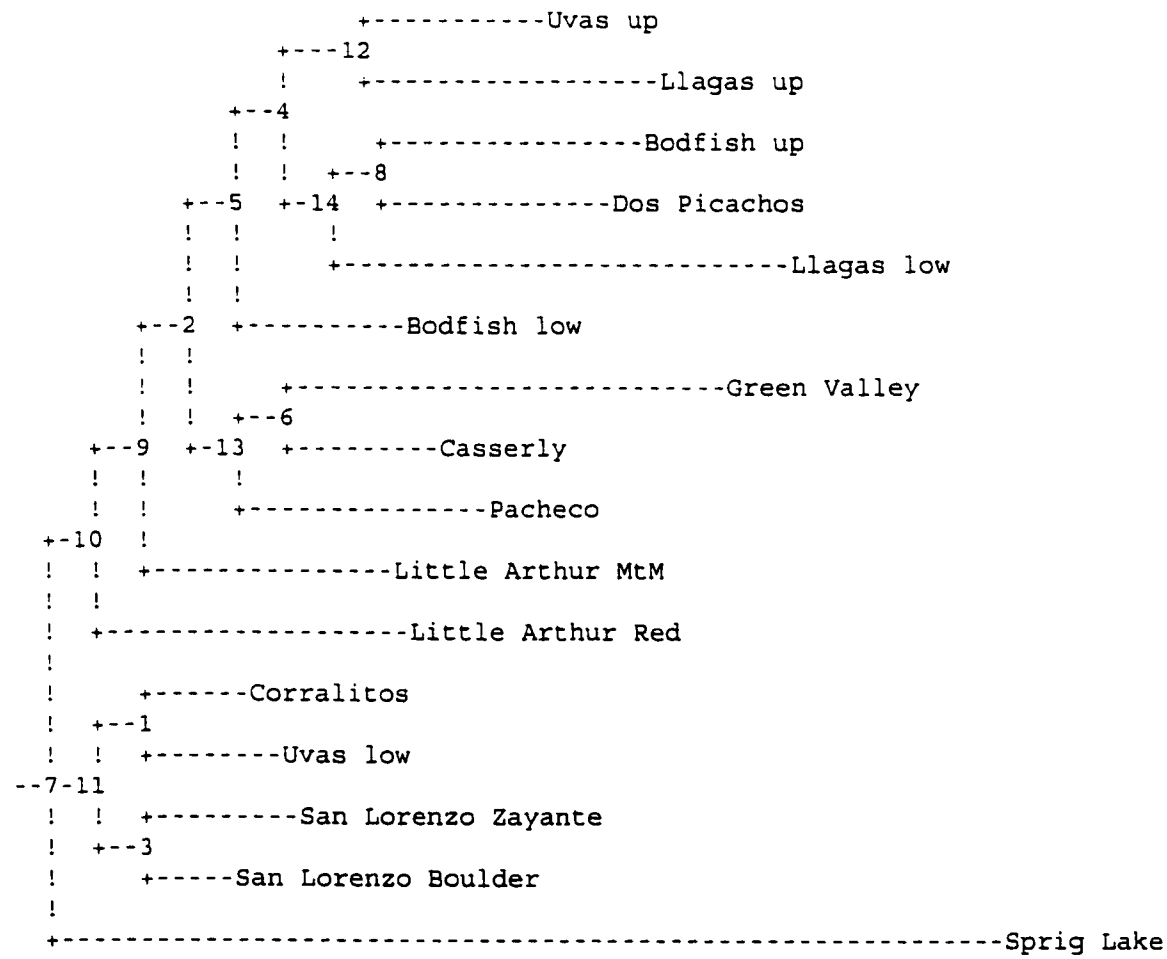




Appendix F

CONTML Tree and Confidence Limits from PHYLIP Version 3.5c for the San Lorenzo and Pajaro Watersheds

```
Continuous character Maximum Likelihood method version 3.572c
16 Populations,5 Loci
```



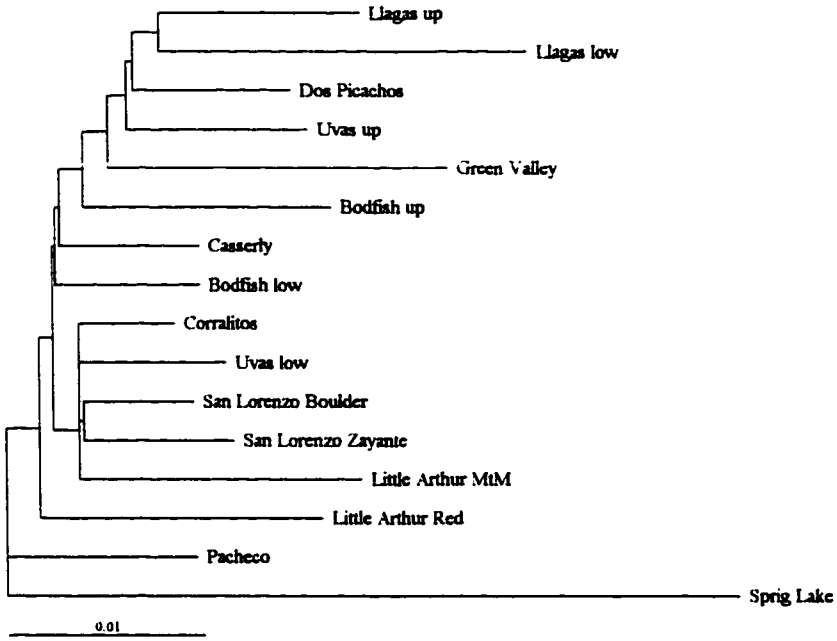

```
remember: (although rooted by outgroup) this is an unrooted tree!
Ln Likelihood =  2770.14300    examined 19299 trees
```

CONTML 16 populations, 5 loci

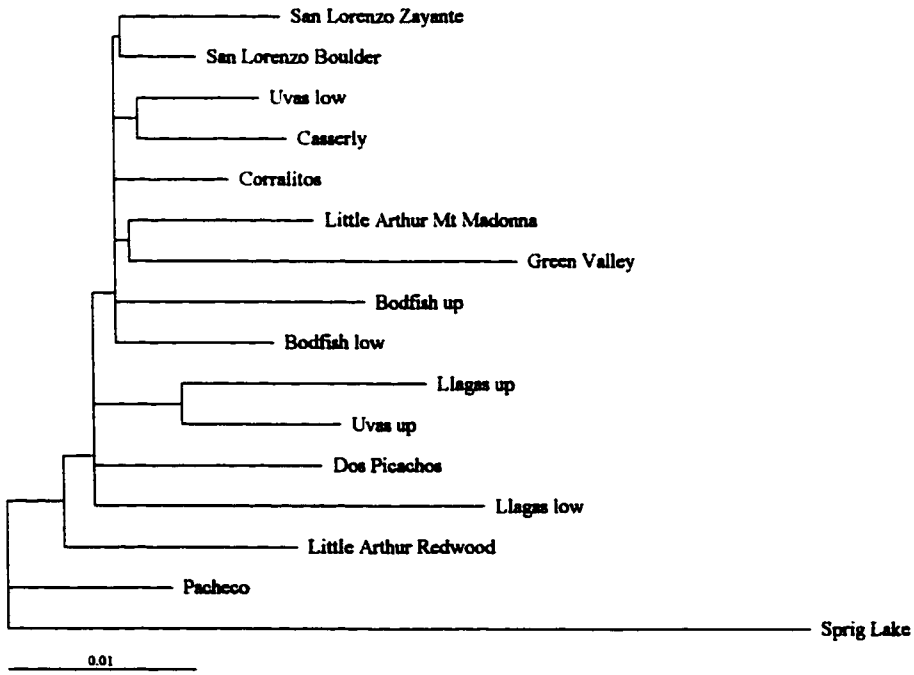
Between	And	Length	Approx. Confidence Limits	
7	10	0.00140	(0.00023,	0.00272)
10	9	0.00000	(-0.00106,	0.00121)
9	2	0.00027	(-0.00077,	0.00144)
2	5	0.00016	(-0.00096,	0.00144)
5	4	0.00118	(-0.00023,	0.00278)
4	12	0.00348	(0.00090,	0.00642)
12	Uvas up	0.00892	(0.00584,	0.01243)
12	Llagas up	0.01396	(0.00978,	0.01872)
4	14	0.00021	(-0.00147,	0.00212)
14	8	0.00000	(-0.00194,	0.00220)
8	Bodfish up	0.01273	(0.00923,	0.01672)
8	Dos Picach	0.01114	(0.00801,	0.01469)
14	Llagas low	0.02164	(0.01600,	0.02806)
5	Bodfish low	0.00795	(0.00572,	0.01050)
2	13	0.00131	(-0.00026,	0.00310)
13	6	0.00000	(-0.00190,	0.00217)
6	Green Valley	0.02071	(0.01528,	0.02689)
6	Casserly	0.00787	(0.00548,	0.01057)
13	Pacheco	0.01171	(0.00843,	0.01544)
9	LitArt MtM	0.01217	(0.00893,	0.01585)
10	LitArt Red	0.01486	(0.01097,	0.01929)
7	11	0.00000	(-0.00060,	0.00068)
11	1	0.00023	(-0.00074,	0.00134)
1	Corralitos	0.00550	(0.00397,	0.00724)
1	Uvas low	0.00711	(0.00520,	0.00929)
11	3	0.00010	(-0.00081,	0.00113)
3	San LorenZ	0.00774	(0.00570,	0.01007)
3	San LorenB	0.00475	(0.00342,	0.00626)
7	Sprig Lake	0.04497	(0.03391,	0.05753)

Appendix G

CONTML trees for Pajaro and San Lorenzo river watersheds based upon four microsatellite loci. a) Without Omy77 and b) Without Omy207.



(a)



(b)