Microsatellite DNA for the management and protection of California's Central Valley chinook salmon (*Oncorhynchus tshawytscha*)

Final Report for the Amendment to Agreement No. B-59638

Dennis Hedgecock
Principal Investigator

University of California, Davis
Bodega Marine Laboratory
2099 Westside Road
Bodega Bay, CA 94923-0247
(707) 875-2075
dehedgecock@ucdavis.edu

Microsatellite DNA for the management and protection of California's Central Valley chinook salmon (*Oncorhynchus tshawytscha*)

Final Report for the Amended Contract

This report covers activities under an amendment to a contract that began in November 1997. The contract was extended, at no cost, to October 31, 2001, and subsequently amended and extended to June 30, 2002. As this amendment brings to conclusion a study that originated in May 1994, under a parent DWR contract (Agreement No. B-59638), we also provide an account of the evolution in approach and methods for genetically discriminating juvenile winter chinook during their migration through the Sacramento-San Joaquin Delta. Under the amendment, we were to complete three remaining tasks in this seven-year effort:

Task 1: Genetic analysis of chinook salvage and Delta monitoring samples in the 2000/1 season of juvenile migration.

Deliverables:

(1) Reports of salvage results for each set of tissues transmitted to us throughout the out-migration season (approximately weekly or bi-weekly); (2) a final report on winter run salvaged in the 2000/1 season; (3) a final summary report on the past six years of salvage and Delta sampling.

Under the amended contract, we completed analyses of genetic data for all salvage samples provided to us, from 21 August 2000 to 29 April 2001, following the protocols established in previous work and detailed below. Preliminary results for the 2000/1 salvage season were communicated to the DWR, in a series of e-mails from Cheryl Dean to Sheila Greene, from 22 Nov. 2000 to 30 October 2001, thus satisfying the first deliverable under this task. We also completed genetic analysis of samples from the IEP Delta monitoring program over the 1995/6 through 2000/01 seasons of juvenile migration. We compile here the results for six years of sampling from the salvage facilities and from Delta monitoring surveys; as this includes analysis of the 2000/1 salvage data, this report satisfies the last two deliverables.

Background

In the course of this research, we first developed or adapted microsatellite DNA markers for chinook salmon and carried out a baseline survey of diversity within and among chinook salmon stocks in the Central Valley (Banks et al. 1999; Banks et al. 2000). This study revealed substantial genetic divergence among the winter, spring, fall, and late-fall spawning runs that have long been recognized by biologists; in addition, we discovered that of spring chinook comprised two distinct lineages (one in Butte Creek, the other in Deer and Mill Creeks). We showed that winter chinook salmon could be identified, either as a fractional component of a mixed-stock population or as individuals, using microsatellite DNA markers. We then developed high-throughput genotyping methods and statistical tools for "real-time" genetic analysis and run-assignment of Delta salvage and monitoring samples (Hedgecock et al. 2001).

We further showed that genetically identified winter chinook in salvage samples from the 1995/6 through the 1999/0 seasons occurred outside the size-time criteria for winter juveniles migrating through the Delta (Fisher 1992). This final report adds considerable detail to earlier reports and substantiates that genetic analysis provides a more accurate and reliable means than size-time criteria for identifying winter chinook in the Delta.

The management questions motivating this research were how many winter run salmon are taken at the state and federal pumping facilities in the Sacramento-San Joaquin Delta each year and can take be assessed in real time so that pumping operations might be modified to protect winter run. Two challenges emerged in applying genetics to the management of water and salmon resources in the Delta. Given the genetic distinctiveness of winter chinook from other chinook salmon in the Central Valley and substantial baseline data on Central Valley spawning stocks, we initially proposed use of classic mixed stock analysis (MSA), wherein one statistically partitions large, mixed population samples into the fractional contributions of different spawning populations (Utter and Ryman 1993). MSA has been very successfully employed for several decades to manage ocean harvests of Pacific salmon for the protection of threatened or endangered stocks. Our plan was to use MSA to determine what proportion of a week's salvage at state and federal Delta pumping facilities was from the winter run *vs.* other genetically distinct runs of the Central Valley. From the fractional contribution of winter-run to salvage, we further hoped to infer the take of winter run, the quantity of interest to protection of the salmon.

A basic assumption of MSA is that allele frequencies in the mixed population can be accurately and precisely determined, so that they can then be partitioned into fractional contributions from baseline populations. However, we came eventually to understand that the salvage facilities do not randomly or uniformly sample the juvenile chinook salmon impacted by water-pumping operations. In the mixed salvage sample, each individual fish represents a larger and potentially unique number of fish impacted by pumping operations. Expansion of salvage numbers to estimate take and loss is based on formulae that account for variation in flows, pumping rates, screening efficiencies, predation rates, handling mortalities, and other factors, as set forth in the Delta Pumping Plant Fish Protection (4-Pumps) Agreement (1986) between the state Department of Fish and Game (DFG) and Department of Water Resources (DWR). Furthermore, the samples taken for genetic analysis are not a random sample of salvaged fish, as the proportion of fish sampled on any given day depends on the number of fish in the salvage facility. As a result, allele frequencies in the impacted juvenile chinook population and the fractional contributions of different runs to take are not likely to be estimated accurately. Quantifying winter run loss and "take" in the Delta remains a formidable challenge.

Prevented from using MSA, we turned to methods for assigning individuals to run of origin, using a log-likelihood ratio test that is based on probabilities for an individual genotype in baseline population samples for Central Valley chinook salmon (Banks and Eichert 2000; Banks et al. 2000). Individual assignment, while not solving the problem of quantifying take, does provide insight into the timing of winter run migration through the Delta, revealing that size-time criteria based on a growth model are not appropriate for quantifying loss and take. We show here that the individual assignment test provides a statistically robust procedure for determining whether a juvenile chinook salmon comes from the winter run or from any of the non-winter runs. We show further that the temporal pattern of winter occurrence, combined with the rather

2

constant sizes of winter run juveniles throughout the season of migration through the Delta, suggests that winter chinook are not growing in the Delta but are instead migrating rapidly through it.

The assignment test is based on the odds that an individual, multi-loci genotype belongs to one run rather than another. The odds for winter assignment are calculated as the ratio of that genotype's frequency in the winter-run to its frequency in other spawning runs. As originally implemented in WHICHRUN (Banks and Eichert 1999), only a single, alternative run is represented in the denominator of the odds ratio. If the frequency of the individual genotype is highest in winter, the alternative tested is the run with the second highest frequency; if the genotype frequency is not the highest in winter, the alternative tested is the run with the highest frequency. However, by maximizing the denominator, this test minimizes winter odds and may fail to assign true winter fish, when the most likely alternative is unlikely to have been present at the time and place of capture. For example, spring run is often the second most likely run for winter genotypes but is unlikely to be abundant at certain times when winters are captured. This conservative bias was ameliorated by using the average of the odds for each of the runs, which is equivalent to assuming that all five runs are present in equal proportions. On the other hand, false-positive assignments to winter run can result if a non-winter run predominates at the time an individual is captured, such as during the fall-run juvenile migration through the Delta. To ameliorate this problem, we developed a Bayesian approach to individual assignment, adjusting the fundamental log likelihood ratio of genotypic probabilities by prior probabilities for relative run abundance, which are based on MSA estimates of run-composition.

Materials and Methods

Sample collection
All samples used for this research were obtained under appropriate federal and state permits for taking and possessing tissues of endangered species.

Salvage. Caudal fin clips were collected non-lethally as part of a cooperative program by DFG, DWR, and the Tracy Fish Collection Facility, to monitor juvenile migration of juvenile salmon. Fin clips and collection data (site, date, time, fork length) were obtained from a subset of all salmon taken at the salvage facilities of the CVP (Tracy Fish Collection Facility) and SWP (Skinner Delta Fish Protection Facility) during six consecutive juvenile migration seasons (1995/6, 1996/7, 1997/8, 1998/9, 1999/0, 2000/1). These fin clips were logged into the DFG tissue archive at Rancho Cordova and subsequently transferred to the Bodega Marine Laboratory (BML), with chain-of-custody documentation. Sampling of fish at the salvage facilities is non-random and non-uniform. Based on preliminary genetic results, the sampling strategy was revised, during 1998/9, to be size-stratified, depending on date. Fin clips were taken from all fish captured from August through November; from all fish above 55 mm fork length (FL) and from sub-samples of fish smaller than 55 mm FL from December through February 15 (later amended to mid March); and from random sub-samples of all fish from February 15 (later mid-March) through May. The final date of sampling was subsequently extended into late June. Fin clips were stored frozen in buffer (10mM Tris-HCL, pH 8.0, 25mM EDTA, 100mM NaCl) until processed. Over this period, 7136 samples were received at BML for genetic analysis (Fig. 1).

Delta monitoring. Samples of juvenile chinook salmon were obtained from the IEP Delta monitoring program. Prior to 1999, samples were taken intensively four times per year at a variety of sites in the lower Sacramento River and the Delta, using beach seines, rotary screw traps, and mid-water trawls. Very few winter run were observed in these samples. Based on preliminary genetic results from the salvage sampling, the Genetics subcommittee of the IEP Salmon Project Work Team revised the sampling program on December 2, 1998, so that sampling would be continuous from Knight's Landing (KNL), the lower Sacramento River (LSR), and two sites in the Delta (DLC, DLS), using rotary screw traps and mid-water trawls only. Further, the sampling was size-stratified, depending on time of year, in the same manner as for the salvage sampling. From December 1995 to April 2001, 3876 samples were received at BML, of which 3011 fit the criteria for analysis established by the Genetics subcommittee.

Laboratory methods

DNA was extracted by placing 0.5 mm^2 of tissue in 200 µl of 5% Chelex 100 (BioRad), heating for 30 min. at 60˚C, then boiling for 30 min. at 103˚C. Samples were genotyped at five previously described microsatellite loci, *Ots-2*, *Ots-3*, *Ots-9*, *Ots-10* (Banks et al. 1999) and *Oneµ-13* (Scribner et al. 1996). PCR was performed using two multiplexed PCR amplifications with a HEX or fluorescein forward labeled PCR primer (Greig and Banks 1999). PCR products from 95 individuals plus controls were electrophoresed for two hours at 50 watts on a 45 cm-wide by 27 cm-tall, 8% denaturing polyacrylamide gel. DNA fragments were visualized on the FMBIO fluorescent imaging system (Hitachi Software Engineering Co.), then scored using BioImage IQ auto-banding software and verified by at least two researchers. In the case of salvage samples, when the five-loci genotype was a likely winter run (methods described below), the sample was genotyped at two additional markers, *Ots-104* and *Ots-107* (Nelson and Beacham 1999), to increase power for winter run assignment.

Statistical methods

We previously surveyed 41 samples of Central Valley chinook populations for variation at 10 microsatellite markers and identified five genetic subpopulations (winter, spring from Butte Creek and from Deer and Mill Creeks, fall, and late fall; Banks et al. 2000). Allele-frequency data for homogeneous subsets of samples representing the five runs – 212 winter, 113 Deer and Mill Creek spring, 87 Butte Creek spring, 300 Fall, and 219 Late Fall – provide baseline allele-frequency data for individual assignment tests and mixed stock analyses (see below). The baseline data set is available at http://www-bml.ucdavis.edu/imc.

A log-likelihood ratio test is used to assign individuals either to the winter run or to the composite, non-winter chinook population. For each individual, we calculate the likelihood or the probability of its multi-loci genotype in each of the five baseline populations, using the program WHICHRUN (Banks and Eichert 1999; available at http://www-bml.ucdavis.edu/ whichrun.htm). An odds ratio is then constructed in a spreadsheet by dividing the winter likelihood with the average likelihood of observing that genotype in the remaining populations. The log (to the base 10) of this odds ratio or LOD score is used to assign the individual, based on their genotypes at five loci (*Ots-2*, *Ots-3*, *Ots-9*, *Ots-10*, and *Oneµ-13*). A LOD score of zero corresponds to a likelihood ratio of 1:1 or even odds that the fish in question belongs either to the winter run or to a non-winter population. Individuals with a LOD score greater than or equal to zero are assigned to the winter population, while those with a LOD score less than zero are

assigned to the non-winter population. If the LOD score for an individual is greater than zero but less than 2.0, then the individual is genotyped at two additional loci, *Ots-104* and *Ots-107*, and the LOD score recalculated.

The accuracy of the assignment tests using these criteria and the two sets of loci (5 or 7) has been assessed through simulations, as discussed under the next task. Also investigated in the next section are the criteria for a minimum assignable genotype. Because of the diagnostic power of *Ots-2* and *Oneµ-13*, an individual can lack neither of these to be assigned; if it has those two loci, then it cannot be missing data for more than one of the other three loci. If an individual is provisionally assigned to the winter run based on LOD score at 5 loci greater than zero but less than 2.0, then it must be typed at both *Ots-104* and *Ots-107* to be assigned more definitively.

The assignment method described so far assumes that all components of the mixture are equally frequent. This assumption can result in false assignments when populations mix in highly unequal proportions. To compensate for unequal proportions of different chinook runs taken in the salvage operations, we employ a Bayesian method for incorporating prior information on relative run frequencies into the calculation of likelihood, using Bayes' rule:

$$P(\theta \mid X) = \frac{P(\theta)P(X \mid \theta)}{P(X)},$$

where $P(\theta \mid X)$ represents the conditional likelihood of a fish from a given run, $P(\theta)$ is the relative abundance of the run estimated by MSA, $P(X \mid \theta)$ is the frequency of the multi-loci genotype in that run, and $P(X)$ is a normalizing factor,

$$\sum_{i=1}^{5} P(\theta_i)P(X_i \mid \theta_i).$$

We correct the assignment LOD score for each winter, by calculating the ratio of the conditional probability of winter to the average conditional probability for the four non-winter runs.

For each individual assigned to the winter run on the basis of the genotypic LOD score, we construct a contextual mixed population by selecting the closest 100 samples to the assigned individual, based on fork length (mm) and date of capture (days). To insure the independence of prior probabilities from the test individual, we determine the Euclidean distance from the test individual to all samples collected in years other than the year in which the assigned individual was captured. The Euclidean distance for each individual is calculated, with an Excel spreadsheet macro, as:

$$\delta_i = \left[\left(FL_X - FL_{Yi} \right)^2 + \left(DoC_X - DoC_{Yi} \right)^2 \right]^{1/2},$$

where FL_X, FLYi, DoCX, and DoC_{Yi} are the fork lengths and dates of capture of the assigned individual, *X*, and contextual individual Y_i, respectively. The 100 individuals closest to the assigned individual in size and date of capture are then subjected to mixed stock analysis (MSA), which estimates the fractional representation of spawning runs in the contextual population. Of course, this MSA is not strictly valid, owing to the non-random and non-uniform nature of

salvage sampling discussed above. Adjustment of the assignment odds, however, does not require as great a precision as estimates of winter take; indeed, the correction is greatest when non-winter fish predominate, such as fall run smolts, at ~120 mm FL, after mid-May (Fig. 1). The final Bayesian assignment test, utilizing as prior probabilities the relative run proportions estimated for the closest 100 contextual individuals, is calculated by a spreadsheet macro.

Accuracy and precision of mixed stock analysis is calculated using Statistical Program for Analyzing Mixtures 3.2 (SPAM, version 3.6, available through the Gene Conservation Laboratory link at http://www.cf.adfg.state.ak.us). Two hundred individuals from each of five stocks are created through permutation of the Central Valley baseline dataset; 1000 permutated data sets are used to determine the mean and standard error of fractional run contributions to the mixture. Analysis was performed using five microsatellite loci, *Ots-2,-3,-9,-10*, and *Oneμ-13*, which are routinely typed on all fish with LOD<0.0 (non-winters).

We examine genetic statistics for winter juvenile populations in each of the six years, comparing them with the baseline adult winter population. We test whether single-locus genotypic proportions conform to those expected under random mating (the Hardy-Weinberg-Castle or HWC equilibrium), using GENEPOP (v. 3.2; Raymond and Rousset 1995). Within-population measures of genetic variability and equilibrium, observed and expected heterozygosity, mean number of alleles per locus, Wright's F_{IS} statistic, and gametic-phase disequilbrium for all pairs of loci, are calculated with the program GENETIX (v. 4.1) (Belkhir et al. 2001). GENETIX is also used to test the significance of genetic variation among adult and juvenile winter population samples, by calculating Wright's standardized measure of inter-population variance, F_{ST}, and comparing this to the distribution of F_{ST} in 5000 permutations in which alleles are shuffled randomly between adult and juvenile populations. The test is done sequentially, starting with juvenile data pooled over the six year-classes. If F_{ST} between the adult and pooled juvenile populations is significant, we calculate the pairwise F_{ST} among the adults and the six juvenile samples. The most divergent juvenile population is then removed, the remaining juvenile populations are pooled, and the adult-juvenile F_{ST} recalculated. The process is repeated until a homogeneous collection of adult and juvenile populations is obtained.

The relationship of size to date of capture for winter juveniles is examined with a linear model, FL = YR + day(YR), where FL is fork length, YR is the year of capture and day(YR) is the date of capture nested within year, using SAS (v. 6.1, Proc GLM, SAS Institute, Cary, Ind.).

Results

Data quantity and quality
Of the 7127 fin-clips from six years of salvage sampling received by BML for genetic analysis, 87 (1.22%) did not amplify, 276 (3.87%) do not meet the criteria for assignment, and 6752 (94.7%) have assignable multi-loci genotypes and (Table 1). For the last two years of the project, when better accounting was made, we observed eight instances of no fin-clip in the tube sent and three cases of steelhead samples, which were identified with molecular markers (Greig et al. 2002). The proportions of individuals genotyped that were successfully assigned, by year, are 0.967, 0.904, 0.986, 0.954, 0.976, and 0.975, for 1995/6, 1996/7, 1997/8, 1998/9, 1999/0, and 2000/1, respectively.

Table 1. Fate of salvage (A) and Delta monitoring (B) samples selected for genetic analysis by DWR.

A. Salvage	1995/6	1996/7	1997/8	1998/9	1999/0	2000/1
Received from archive	340	1031[*]	1112	1771	1214	1662
No sample / No data	4	–	–	–	0	8
No amplification	6	5	3	6	49	18
Steelhead	0	0	0	0	2	1
# of fish unassigned	11	99	16	81	28	41
# of fish assigned to run	319	927	1093	1684	1135	1594

[*]Does not include 18, non-winter, adipose-fin-clipped samples sent by mistake.

B. Delta monitoring	1995/6	1996/7	1997/8	1998/9	1999/0	2000/1
Received, met criteria	788	801	605	396	203	218
Not analyzed	–	–	–	–	2	10
# unassigned before 2002	93	72	64	16	–	–
# assigned before 2002	695	729	539	143	–	–
# unassigned in 2002	–	–	2	222	133	117
# assigned in 2002	–	–	0	15	68	91

Of the 3011 Delta monitoring samples submitted to BML, 2440 were successfully amplified. Scoring was completed on 2351 of these by Cheryl Dean, and of these, 2105 or 90.5% yielded assignable multi-loci genotypes. An additional 648 samples were genotyped, after Cheryl's departure in 2002, by a less experienced worker; because these gels were difficult to score, likely winter run were identified by the diagnostic *Ots-2*^{66} and *One-13*^{160} alleles. These samples were genotyped a second time for all seven markers, so that a LOD score could be accurately determined. We discuss below how many winters might have been missed with this criterion. Rates of DNA amplification and successful assignment by experienced personnel, in both the salvage and Delta monitoring portions of this project, compare favorably with those obtained for screw-trap sampling of winter juveniles at the Red Bluff Diversion Dam (Final Report, AFRP Cooperative Agreement, UC Davis and USFWS, January 1998 to September 2001).

Identification of winter run chinook in salvage samples
Individual assignment of salvage samples to winter run
Individual assignment tests, using five loci and assuming equal probability of sampling from the five runs, identify 761 winter-run out of the 6752 juveniles genotyped over six years of salvage sampling (Table 2). This sum is reduced by 5.4%, to 720, when the 5-loci assignment test is adjusted for run abundances, and by slightly less than that, to 727, using seven loci and assuming equal run-abundance. Adjusting the 7-loci assignments for unequal run abundances decreases winter assignments by 2.3%, from 727 to 711, suggesting a decreasing importance of run abundances as genotypic odds are improved by the addition of more loci. With enough loci, winter run could be reliably assigned no matter what the context. Most of the fish assigned to winter run over the course of the six-years were collected in 2000/1, 573 of 711 (80.6%). Of the 711 fish assigned to winter, 679 (95.5%) are found within the winter growth bounds; these 679 winter juveniles represent 45.1% of the 1506 fish with assignable genotypes that are within the winter growth bounds.

Table 2. Number of salvage samples assigned to winter run over six years. The number of fish assigned based on five or seven loci and assuming equal population sizes of all five runs are given in the "EPS" columns. Fish assigned based on five or seven loci and the prior probabilities of the various runs based on MSA of nearest 100 juveniles by size and date are given in "Bayes" columns.

Year	No. fish assigned	EPS at 5 loci	Bayes at 5 loci	EPS at 7 loci	Bayes at 7 loci
1995-6	319	17	13	11	9
1996-7	927	11	7	6	6
1997-8	1093	32	27	27	27
1998-9	1684	76	64	67	63
1999-0	1135	42	33	34	33
2000-1	1594	583	576	582	573
Totals	6752	761	720	727	711

Genetics of juveniles assigned to winter run

If these juveniles are correctly assigned to winter run, then they should have the same population genetic parameters as the baseline winter adult population. The 711 samples assigned to winter

Table 3. Measures of genetic variability at seven microsatellite loci for salvage samples assigned to winter run, in each of six years, using the Bayesian correction for unequal run abundances. H_{exp} and H_{nb} are estimates of average expected heterozygosity, the latter corrected for sampling bias. H_{obs} is the observed heterozygosity. $P_{(0.95)}$ is the proportion of loci at which the most common allele does not exceed a frequency of 0.95. Mean k is average number of alleles per locus. The same statistics are presented for the winter chinook baseline (Banks et al. 2000).

Year	N	H_{exp}	H_{nb}	H_{obs}	$P_{(0.95)}$	Mean k
1995/6	9	0.634	0.674	0.730	1	4.57
s.d.		0.165	0.211			
1996/7	6	0.595	0.649	0.691	1	3.29
s.d.		0.123	0.115			
1997/8	27	0.569	0.579	0.607	1	5.71
s.d.		0.154	0.145			
1998/9	63	0.606	0.611	0.624	1	6.71
s.d.		0.146	0.155			
1999/0	33	0.584	0.593	0.615	1	6.29
s.d.		0.149	0.169			
2000/1	572	0.588	0.589	0.597	1	9.43
s.d.		0.121	0.128			
Winter	268	0.608	0.610	0.624	1	9.00
s.d.		0.176	0.197			

run, using seven loci and adjusting for unequal run abundance, appear to have similar levels of genetic diversity as the winter adult population (Table 3). All loci are polymorphic in all samples, under the definition that the most common allele has a frequency no greater than 0.95. Expected heterozygosities (H_{exp} and the non-biased estimate, correcting for sampling variance, H_{nb}) and observed heterozygosities (H_{obs}) are in close agreement with the values in winter run. Mean values for the juvenile samples range from 0.579 to 0.674 for H_{nb} and from 0.597 to 0.730 for H_{obs}, compared to 0.610 (H_{nb}) and 0.624 (H_{obs}) for the adult winter baseline population. The large standard deviations on mean expected heterozygosities suggest that these values are not different from the observed value or from each other. Tests for agreement of genotypic proportions with those expected under random mating (HWC equilibrium) yield non-significant results, as do tests of the significance of F_{IS} within year-class. Although five juvenile samples have lower average numbers of alleles per locus than the 2000/1 juvenile sample or the adult baseline population, this is attributable to the smaller sample sizes in the former samples.

In randomly mating populations, the frequencies of composite genotypes for pairs of unlinked loci should be predictable from the frequencies of the alleles at each locus, a condition known as

Table 4. Proportion of 1000 permutations of the data resulting in a coefficient of linkage disequilibrium as large or larger than the one observed for each pairwise combination of seven loci. The winter adult population has one of 21 values less than 5% (bold), about what is expected by chance. The 1998/9 and 2000/1 salvage samples show significantly higher levels of disequilibrium.

Loci-pair	1995/6	1996/7	1997/8	1998/9	1999/0	2000/1	Winter
Ots-2 – Ots-3	0.369	0.473	0.441	0.275	**0.028**	**0.000**	0.290
Ots-2 – Ots-9	0.472	0.267	0.195	0.227	0.276	**0.020**	0.416
Ots-2 – Ots10	0.135	0.325	0.939	0.462	0.966	0.359	0.870
Ots-2 – One13	0.802	1.000	0.121	0.240	0.549	**0.000**	**0.022**
Ots-2 – Ots-104	0.748	0.183	0.105	**0.001**	0.622	**0.031**	0.276
Ots-2 – Ots-107	0.143	0.054	0.110	0.736	0.180	**0.004**	0.335
Ots-3 – Ots-9	**0.031**	0.210	0.486	**0.000**	0.297	**0.001**	0.304
Ots-3 – Ots10	0.670	0.649	0.897	0.434	0.748	0.260	0.442
Ots-3 – One13	0.745	0.073	0.666	0.076	0.311	**0.014**	0.122
Ots-3 – Ots-104	0.294	0.670	0.184	**0.030**	0.349	**0.047**	0.888
Ots-3 – Ots-107	0.209	0.375	**0.023**	**0.028**	0.075	**0.000**	0.692
Ots-9 – Ots10	0.194	0.167	0.986	0.481	0.297	0.289	0.175
Ots-9 – One13	0.579	1.000	0.076	0.810	0.406	**0.044**	0.558
Ots-9 – Ots-104	0.290	0.782	0.328	0.116	0.525	**0.001**	0.945
Ots-9 – Ots-107	0.509	0.244	0.138	0.361	0.199	**0.023**	0.192
Ots10 – One13	0.302	1.000	**0.041**	0.242	0.774	0.589	0.557
Ots10 – Ots-104	0.230	0.331	0.962	0.598	0.249	0.301	0.275
Ots10 – Ots-107	0.212	0.159	0.887	0.097	0.330	**0.003**	0.132
One13 – Ots-104	0.287	0.590	0.254	0.064	0.070	**0.004**	0.945
One13 – Ots-107	0.898	0.858	0.323	0.102	0.258	**0.000**	0.331
Ots-104 – Ots-107	0.618	0.719	0.611	0.055	0.831	**0.022**	0.557
No. < 0.05	1	0	2	4	1	16	1
Prop. < 0.05	0.048	0.000	0.095	0.190	0.048	0.762	0.048

linkage or gametic-phase equilibrium. For the adult winter baseline population, for example, a permutation test reveals only random deviation from linkage equilibrium (Table 4, last column). Significant linkage disequilibrium for the 1995/6 sample was eliminated after the Bayes' correction removed putative winters (see Table 2). Linkage equilibrium tests for the juvenile samples show four year-classes to be similar to the winter baseline with fewer than 5-10% of pairwise loci combinations in linkage disequilibrium. However, the samples for 1998/9 and 2000/1 have 19% and 76%, respectively, of pairwise loci combinations with linkage disequilibrium.

Linkage disequilibrium in these two samples may be caused by residual admixture of non-winter run fish, although this seems unlikely given our diagnostic power (see next section), or by relatedness among individuals. We investigate relatedness, using the program *Kinship* (Goodnight and Queller 1999), which performs an explicit test of the hypothesis that a pair of individuals is more likely to share alleles because they are full-sibs than if they are unrelated, given the allele frequencies in the winter baseline population. The proportion of pairs having significant full-sib relatedness within the 1998/9 sample is 0.078 (120/1540), while the proportion of full-sib pairs within the 2000/1 sample is 0.089 (14483/162165).

Finally, we test whether there is significant genetic divergence among the individuals assigned to winter run from salvage samples and the winter baseline population (Table 5). Here, genetic distance is measured by Wright's standardized measure of allele-frequency variance between samples, F_{ST}. The significance of F_{ST} is tested by randomly shuffling individuals among populations 1000 times, which produces a distribution of F_{ST} for randomized data. If 5% or fewer of the permuted values are as large or larger than the observed value, then we judge the observed F_{ST} to be significantly different from zero. Only the 2000/1 sample differs from the baseline population in pairwise tests, but 5 of 15 pairwise tests among salvage samples are significant. We assess the sources of heterogeneity among these samples by calculating, first, that global $F_{ST} = 0.0085$ (P=0.0) among all seven populations. After removing the 2000/1 sample, $F_{ST} = 0.0033$ (P=0.031) for the remaining six populations. Only after additionally removing the 1998/9 sample does F_{ST} for the remaining five samples, 0.0033, become non-significant (P=0.115). Whatever causes linkage disequilibrium in the 1998/9 and 2000/1 samples appears also to cause significant genetic distance from other winter populations.

Table 5. Pairwise tests of genetic divergence among the salvage samples assigned to winter run and the winter baseline population. The number of individuals, N, is given in the first row. In the remaining rows are Wright's standardized variance of allele frequency, F_{ST}, (above the diagnonal) and (below the diagonal) the percent of 1000 permutations of individuals among populations that yielded an F_{ST} as large or larger than the one observed. Significant values are in bold.

	1995/6	1996/7	1997/8	1998/9	1999/0	2000/1	W base
N	9	6	27	63	33	573	268
1995/6		-0.020	**0.024**	0.006	0.011	**0.023**	0.006
1996/7	81.9		-0.006	-0.007	-0.004	0.001	-0.013
1997/8	**4.6**	61.0		**0.009**	0.006	0.001	0.005
1998/9	27.6	64.6	**3.5**		**0.010**	0.010	**0.002**
1999/0	15.0	58.5	15.9	**2.3**		0.005	0.004
2000/1	**2.6**	41.1	33.3	**0.0**	6.9		**0.011**
W base	22.7	77.7	11.5	12.3	10.5	**0.0**	

Juvenile migration patterns at the pumps

The distribution of assigned winter run, relative to the 6752 juvenile salmon genotyped, differs substantially from that expected under the growth model (Figure 1). While many winters are within the predicted growth-curve bounds (95.5%), a fraction of winter juveniles (4.5%) arrives at the pumps earlier than predicted; conversely, about half (50.7%) of the juvenile salmon within the predicted winter growth lines are not assignable to the winter run. These proportions are highly variable across years, however. In 1998/9, only half of the winters were within the growth curves; except in 2000/1, most fish within the winter curves were non-winter (Table 6)
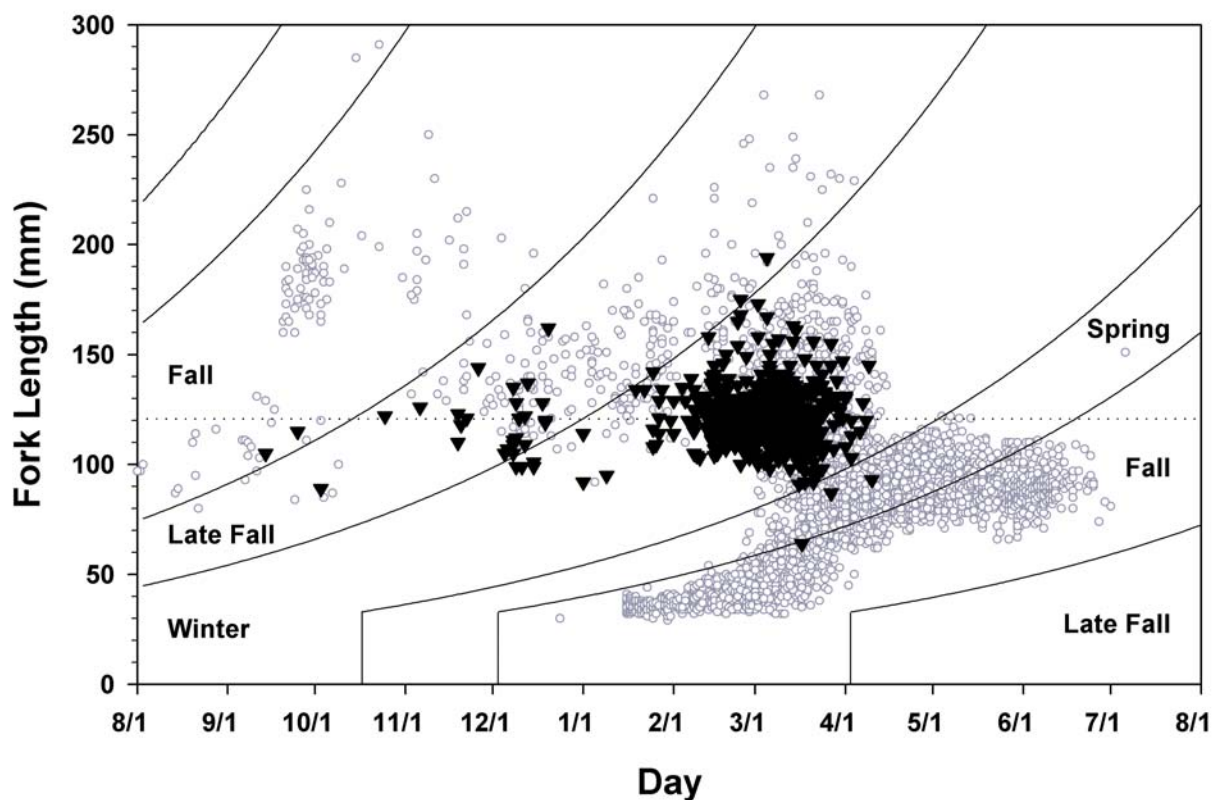

Figure 1. Distribution, by fork length and day of capture, of 6752 juvenile chinook salmon with assignable genotypes over a six-year period, from August 1995 through July 2001. The 711 juveniles assigned to winter run, based on genotypic odds adjusted for the abundance of other runs (LOD>0.0), are shown by dark inverted triangles. Mean fork length of winter run juveniles (121 mm) is shown by the dotted line. The predicted upper and lower growth curves for each of the four Central Valley runs are plotted for comparison (Fisher 1992).

Table 6. Proportions of winter juveniles within the growth curves and proportions of fish within the growth curves that are assigned to winter or not. In and out under category refers to the area between the winter growth curves.

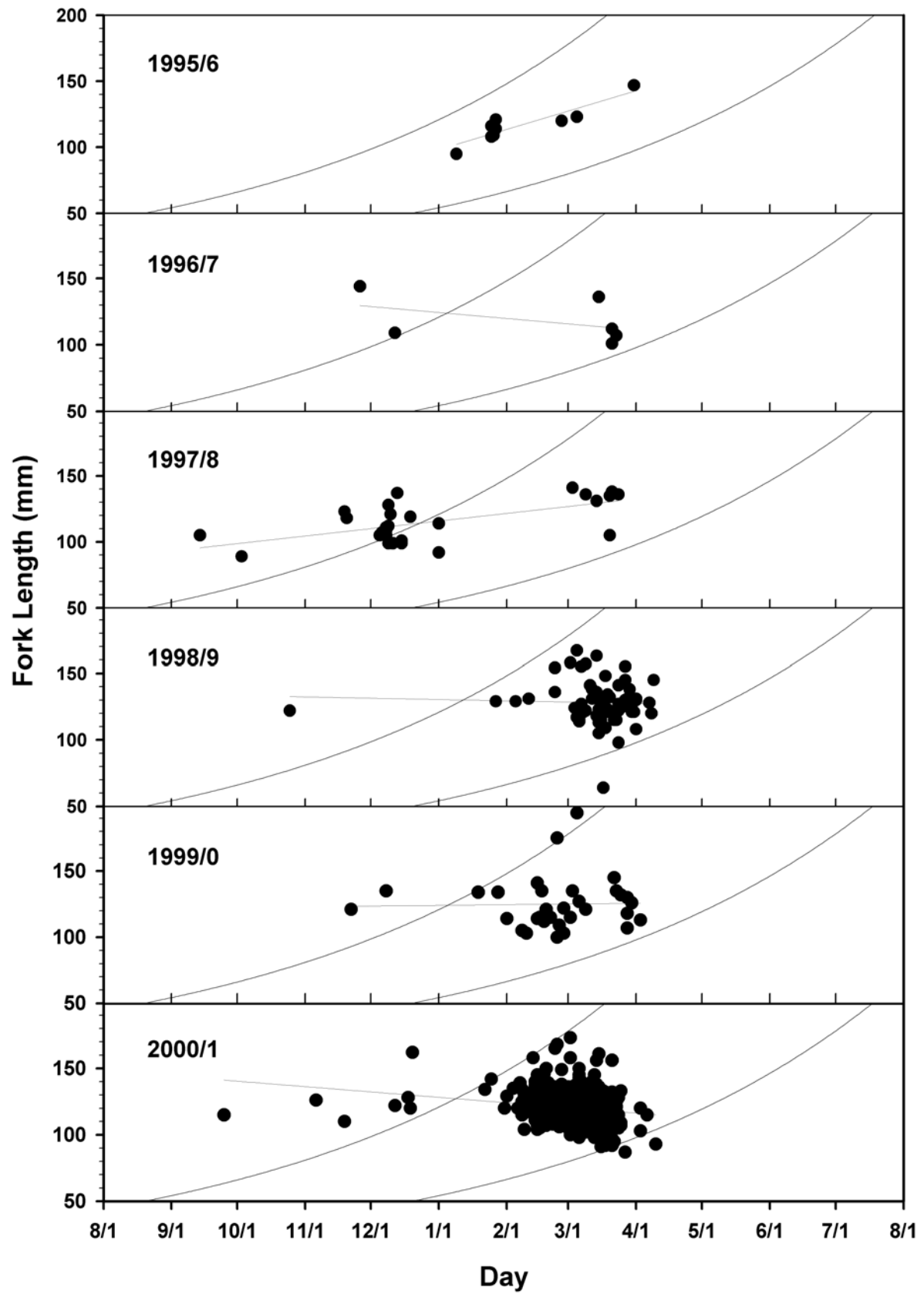| Category	1995/6	1996/7	1997/8	1998/9	1999/0	2000/1	Total
N, Winter	9	6	27	63	33	573	711
% Winter in	100.00%	66.67%	48.15%	96.83%	87.88%	98.25%	95.50%
% Winter out	0.00%	33.33%	51.85%	3.17%	12.12%	1.75%	4.50%
N, in	165	104	43	182	211	673	1378
% in, non-W	94.55%	96.15%	69.77%	66.48%	86.26%	16.34%	50.73%
% in, Winter	5.45%	3.85%	30.23%	33.52%	13.74%	83.66%	49.27%

Figure 2. Distribution of winter juveniles by fork length and date of capture for each year of sampling.

Table 7. Regression of fork length (FL) on date of capture for juveniles assigned to the winter run in six different years. Juveniles in different years differ significantly in mean FL. Only three of the regressions have a significantly non-zero slope; of these, two are positive (1995/6 and 1997/8) and one is negative (2000/1).

Year	N	Mean FL	Adj-r^2	Slope	t	P[t]
1995/6	9	117.0	0.787	0.496	3.03	0.0025
1996/7	6	118.2	0.010	-0.144	-1.50	0.1351
1997/8	27	115.4	0.359	0.186	4.10	<0.0001
1998/9	60	128.3	0.0	-0.028	-0.41	0.6802
1999/0	33	125.0	0.0	0.018	0.24	0.8097
2000/1	571	120.0	0.039	-0.136	-4.42	<0.0001

The linear model, FL = YR day(YR), is significant ($F = 7.77$, d.f. = 11/694, $P < 0.0001$), accounting, however, for only about 11% of the total variance in fork length among the 711 fish assigned to winter run. The dependent variables, YR and day(YR), are each significant, suggesting differences among years in average size and in the regression of fork length on date (Figure 2; Table 7). Only three years have significant slopes of FL on day, however, accounting for 79% of variance in FL in 1995/6, 36% in 1997/8, and 4% in 2000/1. The slope is positive for the 1995/6 and 1997/8 juveniles but negative for the 2000/1 juveniles, suggesting the absence of the positive trend expected if juvenile winter chinook were growing in the Delta.

Identification of winter chinook in Delta monitoring samples

For this portion of the project, we attempted to genotype 2999 fish at five loci (*Ots-2, -3, -9, -10*, and *One-13*), of which 2280 yielded assignable genotypes and 317 were assigned to the winter run (Table 8). Nearly 90% of samples yielded assignable genotypes for the first three years. Complete genotypes were not recorded for many of the 648 samples analyzed from the 1998/9, 1999/0, and 2000/1 seasons. For these collections, we identified potential winters based on

Table 8. Numbers of chinook salmon that were genotyped (Total) and that were assignable to the winter run (W) from four sites (KNL, Knight's Landing; LSR, Lower Sacramento River; DLC, Delta Central; DLS, Delta south) over six years of Delta monitoring.

	1995/6		1996/7		1997/8		1998/9		1999/0		2000/1	
Site	Total[a]	W	Total[a]	W	Total[a]	W	Total[b]	W	Total[b]	W	Total[b]	W
KNL	0	0	0	0	70	65	11	7	4	1	0	0
LSR	476	42	569	13	376	61	4	4	4	2	83	70
DLC	219	7	100	4	91	6	132	11	60	15	8	8
DLS	0	0	59	1	2	0	11	0	0	0	0	0
Totals	695	49	728	18	539	132	158	22	68	18	91	78

[a] totals for these years are the number of assignable genotypes.

[b] totals for these years include fish initially identified as likely winters by alleles at *Ots-2* or *One-13* or both.
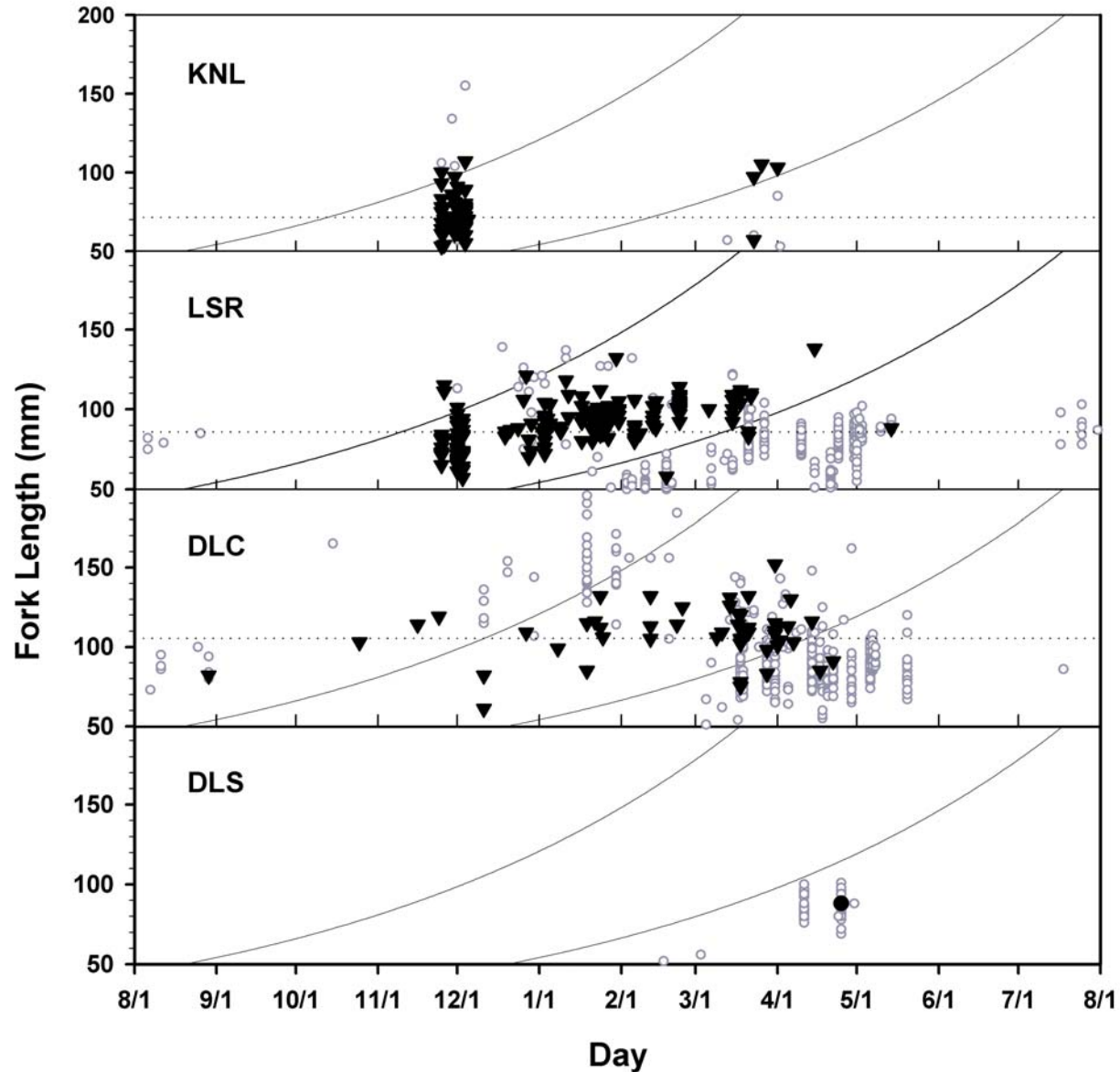
Figure 3. The distributions, by fork length and day of capture, for juvenile salmon captured at four sites in the IEP Delta monitoring program (Knight's landing, KNL; the lower Sacramento River, LSR; the central Delta, DLC; and the southern Delta, DLS) and subsequently genotyped. Juveniles assigned to winter run are depicted by solid inverted triangles. Mean fork lengths of winter juveniles are shown by the dotted lines in all but the last panel, which has only one winter fish (KNL, 71.3 mm; LSR, 85.8 mm; DLC, 105.4mm). The predicted upper and lower growth curves for winter juveniles are plotted for reference.

diagnostic alleles at *Ots-2* and *One-13*; if a fish had at least one *Ots-2*[66] allele or at least one *One-13*[160] allele, we re-typed the individual to obtain the complete five-loci genotype and LOD score. This method would correctly identify every one of the winter adults in the baseline population (n=187). For the larger sample of juveniles, however, we may have missed some true winters, using this criterion. We estimate the number of true winter juveniles missed by the allele criterion by, first, dividing the number assigned to winter run (107) by the proportion of

true winters meeting the allele criterion. In winter run, 93.95% of *Ots-2* genotypes contain at least one *66* allele, while 87.08% of *One-13* genotypes contain at least one *160* allele. Of the 648 samples genotyped, 285 were missing *One-13*. For the 285 fish that were scored only for *Ots-2*, we estimate $[(107/0.9395)\times(285/648)] = 50.1$ winters. For the remaining fish with both *Ots-2* and *One-13* scores, we estimate $[(107/(1-(1-0.9395)\times(1-0.8708))\times(363/648)] = 60.4$ winters. Thus, we estimate that we missed no more than $(50.1+60.4) – 107 =3.5$ winter fish, using this criterion

Juvenile migration patterns in the Delta
The distributions of winter run juveniles, by fork length and date of capture at four sites sampled by the Delta monitoring program, are given in Fig. 3. These distributions are similar to those observed in the salvage samples, in that the winter run are not confined to the area defined by the upper and lower growth curves and show a limited range in FL. The mean fork length of juvenile winter run does increase in the downstream direction: KNL (71.3 mm), LSR (85.8 mm), and DLC (105.4 mm). Mean FL in the salvage samples is 121 mm, although this may be biased upward by the size-selective nature of the salvage operation.

Task 2: Documentation of the laboratory protocol required for future Delta and salvage genetic monitoring.

Deliverables:

(1) critique of best markers to use for winter run identification; (2) analysis of the power of *Ots-2* flanking-sequence polymorphism in winter-run diagnosis; (3) assessment of extent and consequences of size homoplasy at *Ots-2*; (4) documentation of statistical procedures used for run-assignment.

A thorough description of genetic identification protocols for winter chinook would help define contract needs for future routine monitoring of chinook salmon in the Delta and in fish salvage operations. The microsatellite DNA markers that we use for identification of winter-run work very well but are not necessarily the most efficient or the minimum set of markers. In a related project on spring run identification, Dr. Michael Banks developed both additional microsatellite DNA markers and computer simulation methods (WHICHLOCI) for evaluating which markers provide the best discrimination of protected stocks. Using this program and our baseline data, we simulate the power of the seven loci that we have routinely used; we also identify a smaller set of markers that might provide more efficient typing with better power. We provide a critique of the minimum set of markers identified by these simulations, because some markers may be unsuitable for real time analyses or may be prone to scoring errors not accounted for in the simulation.

Included in this objective is the Ph.D. dissertation research of Scott Blankenship, who examined the association of alleles at the microsatellite *Ots-2* with nucleotide polymorphisms in the flanking sequence. These flanking markers may be as good as *Ots-2* for winter-run diagnosis and may be scored more efficiently than *Ots-2* itself. Blankenship's dissertation also provides new insights into the evolution of microsatellite DNA loci and the extent and consequences of size homoplasy (convergent evolution of similarly sized alleles) at these markers. This work is fundamental to the interpretation of microsatellite DNA data.

The statistical procedures for individual assignment are partially documented under the first task. That description pulls together components that have been published separately, including the calculation of genotypic frequencies and log-likelihood ratio tests (Banks and Eichert 2000) and the adjustment of genotype likelihood for relative stock abundance (Hedgecock et al. 2001 and in prep). Many of these methods assume complete multi-loci genotype information for each individual, which is not always achieved in real studies, particularly with degraded samples, such as carcasses or screw trap samples. Here, we discuss criteria that have been developed for assignable genotypes together with their statistical support.

Critique of markers for winter chinook assignment

Simulations with WHICHLOCI

WHICHLOCI is a program based on WHICHRUN, the program that we use to calculate the likelihood a given individual multi-loci genotype in each baseline population and the odds that that genotype belongs to one or another run, based on the log of the odds ratio or LOD-score. WHICHLOCI can be run in a simulation mode that allows one to assess the power of assigning individuals from a particular run to their correct population, locus by locus, with defined thresholds for correct assignment and miss-assignment. These simulations can determine the power of a given set of loci, such as the seven loci used to assign salvage samples to winter run, or find a minimum set of loci that can assign true winters at the defined thresholds. WHICHLOCI was run in both modes to produce the results presented in Table 9. Simulations are run for baseline populations of 10,000, 500, and 200 individuals. Greater accuracy in determining assignment or miss-assignment are achieved with larger simulated populations, but the program does not expand the number of alleles with sample size and may thus overestimate assignment.

Table 9. Simulated accuracy and power of the standard set of 7 loci[a] and a minimum set of 4 loci[b] for assigning true winter chinook salmon, at the predefined thresholds of 99% correct assignment and 1% miss-assignment.

	Correct Assignments of True Winters				Miss-assignments of non-Winters			
Loci	Mean %	Variance	SD	SE	Mean %	Variance	SD	SE
A. Simulated population of 10,000 for each baseline								
7	99.644	0.348	0.590	0.019	0.749	0.726	0.852	0.027
4	99.721	0.246	0.496	0.016	0.794	0.792	0.890	0.028
B. Simulated population of 500 for each baseline								
7	98.755	1.318	1.148	0.115	2.520	3.471	1.863	0.186
4	99.795	0.183	0.428	0.043	0.876	1.217	1.103	0.110
C. Simulated population of 200 for each baseline								
7	99.294	0.705	0.840	0.084	1.443	1.376	1.173	0.117
4	99.746	0.234	0.484	0.048	0.773	0.774	0.880	0.088

[a] In order of value: *One-13, Ots-2, Ots-107, Ots-104, Ots-3, Ots-9, Ots10.*

[b] In order of value: A) *One13, Ots-211, Ots-2, Ots-212*; B) *Ots-211, One13, Ots-2, Ots-212*; C) *Ots-2, Ots-211, Ots-212, One13.*

Note that the standard set of seven loci correctly assigns 99% of true winters, with a rate of miss-assignment that is just over 1%. At the same time, a minimum set of four loci, two of which are included in the standard set and two of which were developed for spring run assignment (Greig and Banks, in prep.), perform slightly better than the standard seven loci. This suggests that the genotyping of salvage samples could be made more efficient by developing multiplex PCRs for these four markers.

Other considerations in choosing markers

Although WHICHLOCI is a good tool for evaluating the statistical efficiency with which individuals might be assigned to run, other factors must be considered in choosing markers for high-throughput screening of salvage samples. For example, tetranucleotide loci, such as *Ots-107*, *-104*, *-211*, or *-212* have very large numbers of alleles, which can lead to drop-out of large allele in the PCR, owing to the competitive nature of the amplification reaction, or to mis-identification of alleles, unless appropriate and abundant molecular size standards are employed, which can increase cost. Nevertheless, with increases in the efficiency of genotyping tetranucleotide microsatellites or even single-nucleotide polymorphisms, such as discussed in the next section, it seems likely that a more efficient set of markers than what has been used to date could be developed in the near future.

Ots-2 and its flanking sequence

Microsatellites, which have rapidly become the preferred marker in population genetics, reliably assign individual fish to the Sacramento River winter *vs*. fall, late-fall, or spring chinook runs of California's Central Valley (Banks et al. 2000). A substantial proportion of this discriminatory power comes from O*ts-2*, a simple CA repeat, which is expected to evolve rapidly under the stepwise mutation model (SMM). We have sequenced a 300 bp region around this locus and typed 668 microsatellite-flanking sequence haplotypes, in order to explore further the basis of this microsatellite divergence. Three sites of nucleotide polymorphism in the *Ots-2* flanking sequence define five haplotypes that are shared by California and Canadian populations. The *Ots-2* microsatellite alleles are non-randomly distributed among these five haplotypes, in a pattern of gametic disequilibrium that is also shared among populations. Divergence between winter run and other Central Valley stocks appears to be caused by a combination of surprisingly static evolution at *Ots-2* within a context of more rapidly evolving haplotype frequencies. The single-nucleotide polymorphisms in the *Ots-2* flanking sequence provide the same diagnostic power as *Ots-2* itself, together with the advantages of automated, high-throughput SNP typing. A paper describing this work, drawn from Scott Blankenship's Ph.D. thesis (Blankenship 2001), will appear in the November 2002 issue of *Molecular Ecology*. A second MS, drawn from Dr. Blankenship's thesis, has been drafted. This paper uses computer simulations of microsatellite evolution to show that, were microsatellites evolving by rapid stepwise mutation, as commonly believed, they would be poor population markers. The implication from these papers is that the focus in the future should be on typing SNPs and understanding haplotype evolution.

Assignable genotypes

DNA extracted from carcasses (Feather River study) may often be degraded to the extent that PCR amplification is poor. Consequently, individuals with missing data are common in these data sets. Missing data can lead to miss-assignment, if information at particularly diagnostic loci

is absent. On the other hand, discarding individuals because of incomplete data may underestimate the proportion of winter run. Thus, we tested the effect of missing data on assignment. Given the goal of protecting endangered winter run, we choose to err on the side of including non-winter fish as winter (type-II error) than to err on the side of rejecting fish which are genuinely winter (type-I error).

We use WHICHLOCI to investigate this problem, simulating baseline populations (n=500) for the five Central Valley runs of chinook and calculating systematically the percentages of correct and incorrect assignment for each of the 127 combinations of 1 to 6 loci. The denominator in the assignment ratio is the average likelihood of all other runs, *i.e.* we assume equal run abundance in simulated sampling from a mixed population. We set parameters of LOD>0.0 and use 1000 iterations to determine assignment percentages. The maximum assignment is obtained with all five of the core loci (98.8% correct assignment, 1.7% miss-assignment), which are done first under our stratified genotyping strategy, or with all seven loci, adding *Ots-104* and *Ots-107* to the core five (99.7% correct assignment, 0.5% miss-assignment). These simulations show that loci are not equal in the assignment test. For example, *Ots-2* is the best single marker for correct winter-run assignment (95.3% correct assignment) but only the second best marker at minimizing incorrect assignments (22.97%). The marker yielding the least incorrect assignment, 2.6%, is *One-13*, which is second in correct assignment (88.2%). Together, these two loci produce correct assignment /miss-assignment percentages of 97.1% and 4.8%, respectively.

Rules for an assignable genotype

- An assignable genotype from among the core five loci must include both *One-13* and *Ots-2* and be missing no more than one of the remaining three loci (missing *Ots-3*, 98.2% correct, 2.6%; missing *Ots-9*, 98.5% correct, 2.3% incorrect; and missing *Ots-10*, 98.3% correct, 2.1% incorrect).

- For the extended set of seven loci, all combinations of five or six loci suffice. Mean percentages of correct and incorrect assignment for five of seven loci (excluding the combination of the five core loci) are 98.7% and 2.1%, respectively, and for six of seven loci, 99.4% and 1.0%, respectively.

Task 3: Analysis of the genetic affinities of Feather River spring run to other Central Valley chinook.

Deliverables:

(1) A comprehensive analysis of genetic affinity of Feather River unknowns with the five subpopulations of Central Valley chinook salmon at 7-10 highly informative loci.

Background

The status of spring chinook in the Feather River is contentious because spring run is listed under federal and state ESA laws and the impact of the Feather River Hatchery on this ESU is uncertain. CDFG identifies as "unknown" the run of chinook salmon entering the Feather River

Hatchery prior to 15 September. Tagging studies of hatchery stocks suggest that progeny of spring spawning adults can return in the fall and *vice versa* (Brown and Greene 1994).

As a part of our baseline study of genetic diversity of chinook salmon in Central Valley (Banks et al. 2000), we showed that the Feather River Hatchery fall stock belonged to the fall lineage with no evidence for spring-run affinities. DWR subsequently requested a similar analysis of Feather River spring or "unknown" samples (although this task was not stipulated in the parent contracts). Preliminary analyses, the first, using the five "core" loci (*Ots-2, -3, -9, -10* and *One-13*), and a second, adding *Ots-104* and *–107*, showed that "unknowns" were quite distinct from the spring chinook lineages found in Mill, Deer, and Butte Creeks but very closely related to the fall lineage. Indeed, we found that pools of both spring and fall samples were genetically homogeneous. As these results were controversial, we undertook a more comprehensive analysis under the amended contract, involving additional samples from 1999 and 2000 and more loci, with comparisons to the existing non-winter baseline populations of Central Valley chinook (Banks et al. 2000; Table 5).

Materials and Methods

Tissues of suspected spring run fish, from five different years, were provided from the CDFG archive (Table 10). These samples were either from fish that spawned at the Feather River Hatchery before 15 October, from carcasses collected in the river in early October, or from fish caught by anglers on the Feather River, during May or June. The hatchery gates were atypically opened in June 1994 to obtain a sample of spring chinook. CDFG labels all but the 1994 samples as unknown. In addition to typing the seven markers presently used for winter-run diagnosis and described in previous sections, we typed five of six new microsatellite markers developed for spring-run identification (*Ots-204, -208, -209, -211* and *-213*; Grieg and Banks, submitted).

Table 10. Samples of adult, potential spring chinook salmon provided to BML for genetic analysis. FRH is the Feather River Hatchery; FR is the Feather River.

Year	Race	Location	Date	Life stage	N
1994	Spring	FRH	6/6/94	Adult, spawning	25
1995	Unknown	FRH	10/2/95	Adult, spawning	95
1996	Unknown	FR Angler	6/3-21/96	Adult	17
1996	Unknown	FR Carcass	10/3-9/96	Adult, carcass	78
1996	Unknown	FRH	9/30/96	Adult, spawning	95
1999	Unknown	FRH	9/23-10/4/99	Adult, spawning	115
2000	Unknown	FR Angler	5/6-6/12/00	Adult	50

Results

Based on 12 microsatellite loci, the spring and "unknown" samples form a cohesive set of genetically similar populations that is distinct from the fall run (Table 11). Average F_{ST} among the Feather River spring samples is 0.002; 11 of the 15 pairwise comparisons among the six Feather spring samples are not significantly different from zero. The four non-zero distances are comparisons either between hatchery samples or between hatchery and wild samples. On the other hand, divergence between spring and fall runs in the Feather River (mean pairwise $F_{ST} = 0.01$) is significant, with 11 out of 12 pairwise comparisons being non-zero at the 5% level of

significance. Finally, the Feather River spring run, though distinct from Central Valley fall run (mean pairwise F_{ST} = 0.008, all values significantly non-zero), is genetically much closer to fall run than to either Deer-Mill Creek spring (mean pairwise F_{ST} = 0.016) or Butte Creek spring (mean pairwise F_{ST} = 0.034). Interestingly, of the two Feather River hatchery fall-run samples, the 1996 sample is similar to the fall baseline and different from all six Feather River spring samples, while the 1995 sample is different from fall baseline but is closer to the Feather River spring run.

Table 11. Genetic distance (above diagonal, Wright's standardized variance of allele frequencies, F_{ST}) among samples of Feather River chinook salmon (in bold). The significance of F_{ST} (below diagonal) is the percent of 5000 permutations yielding an F_{ST} as large or larger than that observed. Samples are coded by run (F, fall; S, spring; U, unknown but collected during the spring spawning season), year, and source (H, hatchery; W, in river; angler and carcass samples from 1996 combined). Included are comparisons to baseline populations (spring run from Butte Creek and from Deer and Mill Creeks, fall run from various locations given in Banks et al. 2000).

	F95H	F96H	S94H	U95H	U96H	U96W	U99H	U00W	B-Sp	D-Sp	Fall
F95H		0.002	0.006	0.002	0.005	0.005	0.004	0.002	0.032	0.017	0.007
F96H	7.84		0.013	0.006	0.006	0.008	0.007	0.004	0.030	0.015	0.001
S94H	2.50	0.00		**0.004**	**0.004**	**0.006**	**0.000**	**0.003**	0.042	0.020	0.015
U95H	4.38	0.00	**7.60**		**0.002**	**0.004**	**0.001**	**0.001**	0.032	0.015	0.005
U96H	0.20	0.00	**5.94**	**2.86**		**0.004**	**0.001**	**0.001**	0.027	0.014	0.005
U96W	0.22	0.00	**11.18**	**1.58**	**0.78**		**0.004**	**0.001**	0.037	0.017	0.008
U99H	0.50	0.00	**50.98**	**11.04**	**8.28**	**0.62**		**-0.001**	0.034	0.016	0.009
U00W	8.78	0.46	**14.32**	**16.92**	**21.00**	**27.04**	**64.48**		0.034	0.016	0.007
B-Sp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.019	0.025
D-Sp	0.00	0.02	0.60	0.00	0.00	0.00	0.00	0.06	0.00		0.016
Fall	0.00	11.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

Two Feather River spring chinook samples have significant levels of linkage disequilibrium: the 1999 hatchery sample (n=115; significant LD in 7 of 66 pairs of loci) and the smaller 2000 angler survey sample (n=50; significant LD in 8 of 66 pairs of loci). The former sample comprises fish collected on three different dates, 9/23/99 (n=23), 9/27/99 (n=38), and 10/4/99 (n=54). When these sub-samples are analyzed separately, the number of loci-pairs with significant LD drops from 7 in the composite population to 2, 4, and 4 (of 66), respectively, suggesting that there are significant genetic differences among the three temporal samples. Analysis of F_{ST} in pairwise comparisons among the 1999 sub-samples reveals that the 9/23 sample differs significantly from the other two, which are not different from each other. Splitting the U99H samples into three temporal samples makes eight populations for the next step in the analysis, to determine the largest homogeneous set of Feather River spring samples. We begin with an analysis of the global F_{ST} among all eight samples, 0.0028, which is exceeded in none of the 1000 permutations of the data (P = 0.0). We then remove U99H-9/23 and recalculate the global F_{ST} among the remaining seven samples, 0.0019, which is exceeded in only 4 of 1000 permutations (P = 0.004). Removal of the next most distant sample, U95H, yields F_{ST} = 0.0017 (P = 0.021) for the remaining six samples. Finally, by removing U96H, F_{ST} for the remaining five samples, 0.0016, is not significantly different from zero (P = 0.097). These five, genetically homogeneous samples are combined to form a Feather River pooled sample (n=262) for cladistic analysis. These hatchery samples, particularly U99H-9/23 may be mixtures of fall and spring chinook.

A tree depicting the genetic affinity of the Feather River spring pool with other chinook was constructed, using Cavalli-Sforza and Edwards (1967) measure of genetic distance and the unweighted pair-group method, in Phylip 3.0 (Felsenstein 1993). This tree shows that Feather River spring chinook is most closely related to the fall chinook and only distantly related to other spring runs in the Central Valley (Fig. 4).
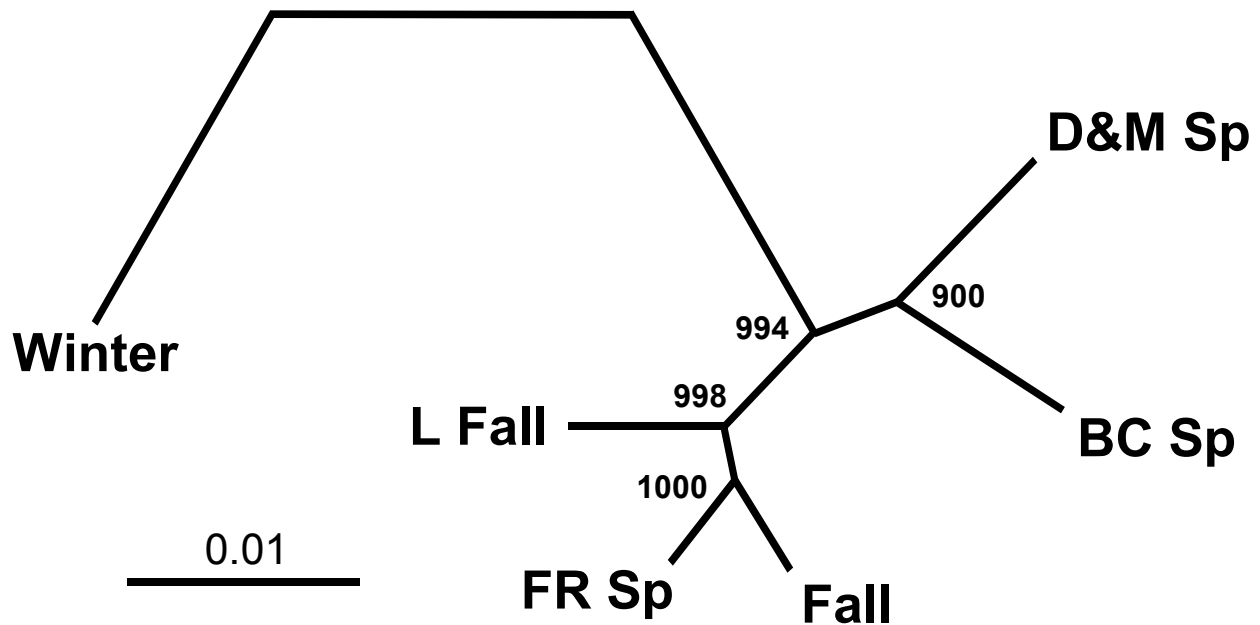

Fig. 4. Genetic distances among Central Valley chinook races (W, winter; D&M Sp, spring from Deer and Mill Creeks; BC Sp, spring from Butte Creek; L Fall, late fall from the upper Sacramento River; and Fall, from various Central Valley locations, including the Feather River; see Banks et al. 2000) and a genetically homogeneous pool of spring chinook from the Feather River (FR Sp). The tree is based on data for 12 microsatellite loci and was constructed in Phylip 3.0 (Felsenstein 1993), using Cavalli-Sforza and Edwards (1967) measure of genetic distance and the unweighted pair-group method arithmetic averaging (UPGMA).

References

Banks, M. A., M. S. Blouin, B. A. Baldwin, V. K. Rashbrook, H. A. Fitzgerald, S. M. Blankenship, and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in Chinook salmon (Oncorhynchus tschawytscha). Journal of Heredity. 90:281-288.

Banks, M. A., V. K. Rashbrook, M. J. Calavetta, C. A. Dean, and D. Hedgecock. 2000. Analysis of microsatellite DNA resolves genetic structure and diversity of chinook salmon (Oncorhynchus tshawytscha) in California's Central Valley. Canadian Journal of Fisheries and Aquatic Sciences 57: 915-927.

Banks, M. A., and W. Eichert. 2000. WHICHRUN (Version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. Journal of Heredity 91:87-89.

Belkhir K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2001. 1996-2001 GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Blankenship, S. M. 2001. Assessing the validity of microsatellite DNA markers for population analysis: patterns of neutral variation and the coalescent approach to modeling molecular evolution. PhD disseretation, University of California, Davis, December 2001., Davis.

Blankenship, S. M, D. Hedgecock, and B. P. May. 2002. Evolution of a perfect simple-sequence-repeat locus in the context of its flanking sequence. Molecular Biology and Evolution 19: 1943-1951.

Brown, RL and S Greene.1994. Evaluation of the Feather River Hatchery as mitigation for the construction of the State Water Project's Oroville Dam. In, Proceedings of the Annual Meeting of Irrigation and Drainage, Sacramento. US Committee on Irrigation and Drainage. Denver.

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

Felsenstein, J. 1993. PHYLIP (Phyologeny Inference Package), version 3.57c. Department of Genetics, University of Washington, Box 357360, Seattle, WA 98105, U.S.A.

Fisher. F. W. 1992. Chinook salmon, *Oncorhynchus tshawytscha*, growth and occurrence in the Sacramento River system. DFG, Inland Fisheries Division Office Report, June 1992.

Goodnight, K.F., and D. C. Queller. 1999. Computer software for performing likelihood tests of pedigree relationship using genetic markers. Mol. Ecol. 8: 1231–1234.

Greig, C., and M. A. Banks. 1999. Five multiplexed microsatellite loci for rapid response run identification of California's endangered winter chinook salmon. Animal Genetics 30:318-320.

Greig, C., J. M. Robertson, and M. A. Banks. 2002. Rapid PCR-based species tests for threatened sympatric salmonids. Conservation Genetics 3:83-86.

Hedgecock, D., M. A. Banks, V. K. Rashbrook, C. A. Dean, and S. M. Blankenship. 2001. Applications of population genetics to conservation of chinook salmon diversity in the Central Valley. In: Contributions to the Biology of Central Valley Salmonids, R. L. Brown, ed., Fish Bulletin 179: 45-70. California Department of Fish and Game.

Nelson, R.J., and Beacham, T.D. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. Animal Genetics 30: 228–229.

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

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 applications. Canadian Journal of Fisheries and Aquatic Science 53: 833–841.

Utter, F. M., and N. Ryman. 1993. Genetic markers and mixed stock fisheries. Fisheries 18:11-21.