

Analysis of microsatellite DNA resolves genetic structure and diversity of chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley

Michael A. Banks, Vanessa K. Rashbrook, Marco J. Calavetta, Cheryl A. Dean, and Dennis Hedgecock

Abstract: We use 10 microsatellite DNA markers to assess genetic diversity within and among the four runs (winter, spring, fall, and late fall) of chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. Forty-one population samples are studied, comprising naturally spawning and hatchery stocks collected from 1991 through 1997. Maximum likelihood methods are used to correct for kinship in juvenile samples and run admixture in adult samples. Through simulation, we determine the relationship between sample size and number of alleles observed at polymorphic microsatellite markers. Most samples have random-mating equilibrium proportions of single and multilocus genotypes. Temporal and spatial genetic heterogeneity is minimal among samples within subpopulations. An F_{ST} of 0.082 among subpopulations, however, indicates substantial divergence among runs. Thus, with the exception of our discovery of two distinct lineages of spring run, genetic structure accords with the diverse chinook life histories seen in the Central Valley and provides a means for discrimination of protected populations.

Résumé : Nous nous sommes servis de dix marqueurs microsatellites de l'ADN pour estimer la diversité génétique dans et entre les quatre remontes (hiver, printemps, automne et fin de l'automne) de quinnat (*Oncorhynchus tshawytscha*) dans la vallée centrale de la Californie. Nous avons étudié 41 échantillons de populations, représentant des poissons à reproduction naturelle et des stocks d'écloserie, prélevés entre 1991 et 1997. Les méthodes du maximum de vraisemblance ont servi à faire la correction pour la parenté génétique dans les échantillons de juvéniles et pour le mélange des remontes dans les échantillons d'adultes. Par la simulation, nous déterminons la relation entre la taille de l'échantillon et le nombre d'allèles observés aux marqueurs polymorphes microsatellites. La plupart des échantillons présentent des proportions équilibrées d'appariement aléatoire de génotypes à un et à plusieurs locus. L'hétérogénéité génétique temporelle et spatiale est minime entre les échantillons au sein des sous-populations. Un F_{ST} de 0,082 entre les sous-populations indique toutefois une divergence substantielle entre les remontes. Ainsi, à l'exception de notre découverte de deux lignées distinctes dans la remonte du printemps, la structure génétique concorde avec les divers cycles biologiques du quinnat observés dans la vallée centrale et fournit un moyen de distinguer les populations protégées.

[Traduit par la Rédaction]

Introduction

Many salmon populations throughout the Pacific Northwest have been extirpated and, of those remaining, the majority is at risk of extinction (Nehlsen 1994). At the southern limit of its range in North America, the Pacific salmon *Oncorhynchus* is particularly vulnerable, owing primarily to dry climate and human competition for water but also to in-

tensive exploitation and habitat disturbance. In California, three salmonid stocks are already protected, and listing of all remaining stocks has been proposed (National Marine Fisheries Service (NMFS) 1998, 1999). Salmon conservation concerns are particularly acute in California's Central Valley, habitat for four spawning runs of the chinook salmon (*Oncorhynchus tshawytscha*) and the source of water for two thirds of the state's inhabitants and its enormous agricultural industry. The stock at greatest risk, the Sacramento River winter-run chinook salmon, was listed as threatened by the state of California in 1989, when run size fell below 200, and as endangered by the federal government in 1994 (NMFS 1994).

Population genetic data are increasingly important in managing salmonid populations on the brink of extinction (Waples 1995). There is a need to identify protected stocks in mixed ocean harvests and in rivers or estuaries where dams or water diversions imperil out-migrating juveniles. Identification of broodstock in propagation programs is

Received March 8, 1999. Accepted February 17, 2000.
J15051

M.A. Banks,¹ V.K. Rashbrook, M.J. Calavetta,²
C.A. Dean, and D. Hedgecock. Bodega Marine Laboratory,
University of California at Davis, P.O. Box 247, Bodega Bay,
CA 94923, U.S.A.

¹Author to whom all correspondence should be addressed.
e-mail: mabanks@ucdavis.edu

²Current address: Hitachi Genetic Systems, 1201 Harbor Bay
Parkway, Suite 150, Alameda, CA 94502, U.S.A.

required to avoid admixture and hybridization among spawning runs. Genetic markers are useful for confirming parentage and relatedness in hatchery-bred fish and for verifying models of hatchery impacts on genetic diversity of naturally spawning stocks. In the case of Central Valley chinook salmon, however, a study of 39 allozyme loci revealed very slight divergence among fall- and winter-run chinook salmon, with Wright's standardized allele frequency variance $F_{ST} = 0.01$ (Bartley et al. 1992). Nielsen et al. (1994) reported substantially more divergence in frequencies of six mtDNA haplotypes, $F_{ST} = 0.24$, but the probability of any two Central Valley chinook haplotypes being identical is 0.7, precluding use of this marker for individual identification. Highly variable nuclear DNA markers, such as microsatellites, make possible genealogical analyses or genetic discrimination among closely related fish populations (Wright and Bentzen 1996). Thus, Banks et al. (1999) cloned microsatellite DNA markers from winter-run chinook salmon, developed polymerase chain reaction (PCR) methods for rapid genotyping, and verified Mendelian inheritance in families produced in a captive breeding program.

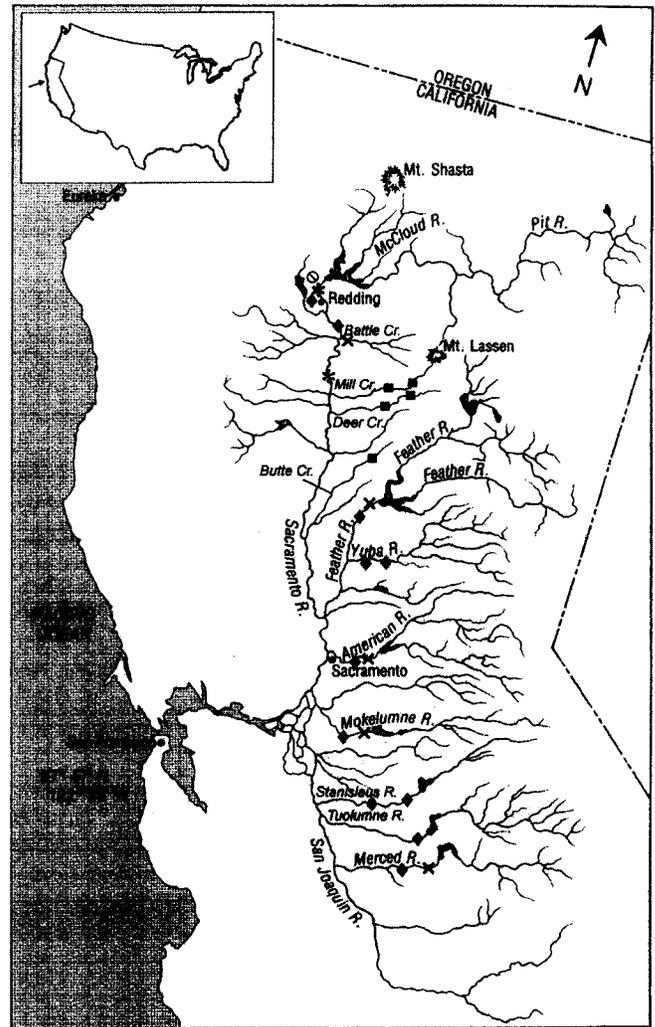
Here, we report the results of a genetic analysis of 41 naturally spawning and hatchery populations of California's Central Valley chinook salmon using 10 microsatellites. We use maximum likelihood methods to correct for family structure among samples comprising juveniles as well as to correct for run admixture in adult samples and demonstrate increased accuracy and precision as a result of these corrections. Our objective is to provide a baseline genetic profile of chinook salmon in the Central Valley that will allow the distinction of endangered stocks from their less threatened close relatives. The study reveals substantial diversity of chinook salmon at the southern limit of their North American range despite 150 years of major ecological disturbances (Yoshiyama et al. 1998). It also illustrates the power of microsatellite DNA markers to resolve the complex genetic structure of salmon populations within a major river basin.

Materials and methods

Sample collections

Between 1991 and 1997, tissue samples from 3032 chinook salmon were collected from 20 Central Valley sites by the California Department of Fish and Game (CDFG) and the U.S. Fish and Wildlife Service (Table 1). Samples comprised all four runs of chinook salmon and, for 11 sites, multiple year-classes. A complete list of all samples used in this study, their accession numbers if taken from the CDFG tissue archive (The Resource Agency, CDFG, 1701 Nimbus Rd, Rancho Cordova, CA 95670), and their genotypes are available at <http://www-bml.ucdavis.edu/cvdataset.htm>. Winter-run samples comprised both spawned carcasses (1995–1997) collected from the Sacramento River and broodstock collections (1991–1995) trapped at Keswick and Red Bluff diversion dams near Redding (Fig. 1). Owing to the inaccessibility of spring-run spawning habitat and consequent difficulties of collecting a sufficient number of spawned carcasses, out-migrating juveniles were sampled in addition to carcasses from Butte, Deer, and Mill creeks (1994–1997). The fall run, presently the most abundant chinook run in the Central Valley, is the only run found in the lower Sacramento – San Joaquin watershed (Yoshiyama et al. 1998). Samples from eight naturally spawning and five hatchery fall-run populations were collected from 1993 to 1996. Like the winter run, late-fall-run stocks are confined to the Sacramento

Fig. 1. The Sacramento and San Joaquin rivers and their tributaries. Sampling sites per run are indicated as follows: *, winter; ■, spring; ◆, fall; open circle with diagonal line, late fall; ×, hatcheries. See Table 1 for temporal and life history stage details for each population sample.



River where both hatchery (1993 and 1995) and naturally spawning (1995) adults were sampled.

Microsatellite typing

Tissue samples (caudal fin clips) were placed in a storage buffer (10 mM Tris-HCl, pH 8.0, 25 mM EDTA, 100 mM NaCl) and frozen until processed. Genomic DNA extraction varied according to whether tissue was obtained from live fish or degraded carcasses. With fresh tissue, 0.5 mm² was placed in 200 µL of 5% Chelex[®]100 (BioRad) in 96-well PCR trays (0.2 mL) (Corning-Costar). Tissue extracts were covered with a silicone-sealing mat (Corning-Costar), heated for 30 min at 60°C, and then boiled for 30 min at 103°C in a thermocycler (M.J. Research). Trays were centrifuged for 5 min at 3000 rpm to pellet fin debris and Chelex[®] beads. One hundred microlitres of supernatant was removed and transferred to a new 96-well tray to be used as template for PCR, leaving the rest of the DNA for long-term storage. DNA was extracted from carcass fin clips using the Puregene[®] DNA isolation kit (Gentra Systems).

Individuals were genotyped at up to 10 previously described unlinked microsatellite loci: *Ots-1*, -2, -3, -5, -9, and -10 (Banks et al.

Table 1. California's Central Valley chinook salmon scored at up to 10 microsatellite loci: collection data and within-population genetic parameters for adjusted samples (see text).

Run	Year	Collection site	Life stage	<i>L</i>	<i>N</i>	<i>H_e</i>	<i>H_o</i>	<i>A</i>	<i>F_{IS}</i>	H-W-C
Winter										
W1	1991	Keswick and Red Bluff dams	Adult	10	18	0.48	0.48	3.7	0.05	0.84
W2	1992	Keswick Dam	Adult	10	29	0.51	0.50	4.7	-0.01	0.99
W3	1993	Keswick and Red Bluff dams	Adult	10	11	0.52	0.66	3.6	-0.25	0.78
W4	1994	Keswick Dam	Adult	10	26	0.53	0.54	4.3	0.00	0.14
W5	1995	Keswick Dam	Adult	10	37	0.52	0.57	5.4	-0.11	0.61
W6	1995	Sacramento River	Spawned carcass	10	32	0.50	0.52	4.5	-0.01	0.79
W7	1996	Sacramento River	Spawned carcass	10	36	0.49	0.55	4.1	-0.11	0.95
W8	1997	Sacramento River	Spawned carcass	10	103	0.50	0.51	4.7	-0.02	0.86
Mean					37	0.51	0.54	4.4		
Spring										
S1	1994	Butte Creek	Spawned carcass	10	67	0.60	0.59	7.3	0.03	0.12
S2	1996	Butte Creek	Spawned carcass	5	41	0.66	0.65	4.4	0.06	0.81
S3	1996	Butte Creek	Juvenile	5	35	0.68	0.71	5.8	-0.02	0.68
S4	1997	Butte Creek	Spawned carcass	10	117	0.59	0.54	7.2	0.13*	0.00*
Mean					65	0.63	0.62	6.18		
S5	1994	Deer Creek	Juvenile	10	30	0.52	0.52	5.8	0.03	0.64
S6	1995	Deer Creek	Spawned carcass	10	25	0.56	0.57	6.9	0.04	0.20
S7	1996	Deer Creek	Juvenile	6	73	0.73	0.76	9.7	-0.03	0.89
S8	1997	Deer Creek	Spawned carcass	10	49	0.60	0.57	8.3	0.12*	0.04*
S9	1995	Mill Creek	Spawned carcass	10	15	0.54	0.65	4.7	-0.14	1.00
S10	1996	Mill Creek	Juvenile	10	40	0.54	0.57	6.6	0.00	0.73
Mean					51	0.61	0.61	6.6		
Fall										
Hatchery										
F1	1993	Coleman Hatchery	Adult	10	144	0.62	0.62	9.8	-0.04	0.16
F2	1994	Coleman Hatchery	Adult	8	95	0.53	0.56	6.5	-0.06	0.07
F3	1995	Coleman Hatchery	Adult	10	95	0.63	0.65	9.9	0.01	0.56
F4	1995	Feather River Hatchery	Adult	10	95	0.62	0.61	10.3	0.00	0.19
F5	1996	Feather River Hatchery	Adult	7	94	0.72	0.72	10.8	0.01	0.76
F6	1995	Merced River Hatchery	Adult	10	95	0.63	0.59	9.5	0.067*	0.02*
F7	1995	Mokelumne River Hatchery	Adult	10	95	0.61	0.61	9.6	0.01	0.79
F8	1995	Nimbus Hatchery	Adult	10	95	0.60	0.61	10.2	0.01	0.75
Mean					101	0.62	0.62	9.58		
Fall										
Naturally spawning										
F9	1995	American River	Spawned carcass	8	90	0.57	0.60	5.9	-0.07	0.89
F10	1995	Feather River	Spawned carcass	7	75	0.56	0.53	5.4	0.07	0.14
F11	1996	Feather River	Spawned carcass	5	53	0.68	0.72	6.6	-0.04	0.48
F12	1995	Merced River	Spawned carcass	8	88	0.54	0.54	6.3	0.00	0.04*
F13	1995	Mokelumne River	Spawned carcass	10	94	0.60	0.60	9.7	0.00	0.12
F14	1994	Stanislaus River	Spawned carcass	8	26	0.50	0.46	5.0	0.12*	0.27
F15	1995	Stanislaus River	Spawned carcass	8	27	0.54	0.54	4.9	0.05	0.09
F16	1994	Toulumne River	Spawned carcass	8	15	0.52	0.55	4.5	-0.07	0.65
F17	1995	Toulumne River	Spawned carcass	8	29	0.54	0.52	5.1	0.08	0.49
F18	1996	Toulumne River	Spawned carcass	5	78	0.63	0.65	6.2	-0.02	0.24
F19	1995	Sacramento River	Spawned carcass	10	94	0.61	0.60	9.3	0.02	0.08
F20	1996	Yuba River	Spawned carcass	5	54	0.64	0.63	7.4	-0.03	0.77
Mean					60	0.58	0.58	5.94		
Fall grand mean					77	0.59	0.59	7.33		
Late fall										
LF1	1993	Coleman Hatchery	Adult	10	143	0.55	0.55	8.9	0.02	0.25
LF2	1995	Keswick Dam	Adult	10	90	0.56	0.53	8.9	0.07*	0.00*
LF3	1995	Coleman Hatchery	Adult	10	90	0.55	0.58	8.6	-0.04	0.17
Mean					108	0.56	0.55	8.8		
Grand mean					64	0.58	0.59	6.7		
Total					2638					

Note: *L*, number of loci; *N*, adjusted sample size; *H_e*, expected heterozygosity; *H_o*, observed heterozygosity; *A*, average number of alleles per locus; *significant ($\alpha = 0.05$).

Table 2. Classification and examples of hypotheses for observed arrays of progeny phenotypes and their associated probabilities.

Classes of hypotheses	Genotype		F ₁ phenotype ^a				Probability ^b
	Parent 1	Parent 2	I	II	III	IV	
Two alleles	AA	AB	<i>n</i> ₁	<i>n</i> ₂			$[n_T!/(n_1!n_2!)](0.5)^{n_1}(0.5)^{n_2}$
	AB	AB	<i>n</i> ₁	<i>n</i> ₂	<i>n</i> ₃		$[n_T!/(n_1!n_2!n_3!)](0.25)^{n_1}(0.25)^{n_2}(0.5)^{n_3}$
Three alleles	AA	BC	<i>n</i> ₁	<i>n</i> ₂			$[n_T!/(n_1!n_2!)](0.5)^{n_1}(0.5)^{n_2}$
	AB	AC	<i>n</i> ₁	<i>n</i> ₂	<i>n</i> ₃	<i>n</i> ₄	$[n_T!/(n_1!n_2!n_3!n_4!)](0.25)^{n_1}(0.25)^{n_2}(0.25)^{n_3}(0.25)^{n_4}$
Four alleles	AB	CD	<i>n</i> ₁	<i>n</i> ₂	<i>n</i> ₃	<i>n</i> ₄	$[n_T!/(n_1!n_2!n_3!n_4!)](0.25)^{n_1}(0.25)^{n_2}(0.25)^{n_3}(0.25)^{n_4}$

^aFull-sib progeny can have from one to four phenotypes, depending on parental genotypes; the *n_i* are the numbers of progeny observed in each phenotypic category.

^bTerms of the relevant bi-, tri-, or tetranomial probability distributions; *n_T* is the sum of the *n_i*.

1999), *Ots-104* and *-107* (Nelson and Beacham 1999), *Oneμ13* (Scribner et al. 1996), and *Omy-77* (Morris et al. 1996). The forward PCR primer was labeled with a fluorescent phosphoamidite (HEX or fluorescein). PCR products were electrophoresed, 96 at a time with allelic controls, on a 45.0 cm wide by 22.5 cm high 8% denaturing polyacrylamide gel at 50 W for 60–150 min. DNA fragments were visualized on the FMBIO[®] fluorescent imaging system (Hitachi Software Engineering America Ltd.). Gels were manually scored and independently verified by at least one other researcher. The data were entered into a relational database (Paradox[®], Borland, version 7) and double-checked for accuracy. Individuals that did not produce repeatable genotypes were not included in the analyses. Variation in the quality of DNA extracted from carcass tissues made genotypes difficult to score; consequently, fewer than 10 loci were assayed in many carcass samples (Table 1).

Statistical analyses

We developed and applied methods to address three sampling issues: relatedness among juveniles, run admixture, and limits to precision of allele frequency estimates for highly polymorphic loci. Each of the methods is described below.

Correcting for kinship in samples of juveniles

Four samples of the spring run (Deer Creek, 1994 and 1996, Mill Creek, 1996, and Butte Creek, 1996) comprised juveniles only. We investigated and attempted to correct for kinship in these samples because the presence of full- or half-sibs could bias allele frequency estimates (Allendorf and Phelps 1981).

The odds of two individuals being full-siblings rather than unrelated can be calculated using the program Kinship 1.2 (Goodnight and Queller 1999). We first explored the power of this approach using 53 winter-run offspring comprising seven full-sib and three half-sib families in a captive breeding program (Banks et al. 1999). These fish afforded 393 pairwise comparisons among known full-sibs, 307 comparisons among known half-sibs, and 678 comparisons among unrelated individuals. Odds for a full-sib relationship were calculated with both five and nine loci to mimic the information available in the spring-run juvenile samples. Significance thresholds for the full-sib LOD score (the log of the odds ratio) were taken from simulation results at an empirical level of $\alpha = 0.01$. Threshold LODs were 1.04 and 1.14 for five and nine loci, respectively, corresponding to Type II errors of 0.72 and 0.50. Percentages of known full-sib, half-sib, and unrelated pairs yielding significant full-sib LOD scores when scored at five loci are 42.2, 5.6, and 0.4%, respectively. The comparable percentages using nine loci are 57.2, 6.8, and 0.6%, respectively, showing the expected improvement in power with more loci. This exercise suggests that LOD score classification of a full-sib relationship between two winter-run individuals is conservative. The test has low power, detecting a little more than half of true full-sibs when nine loci are used, but suitable protection against Type I error, classifying very few truly unrelated pairs as full-sibs.

To analyze the spring-run juvenile samples, we first deleted individuals missing more than one or two loci (depending on the number of loci sampled) to avoid potentially spurious results in evaluating kinship. Adult spring-run samples for Mill and Deer creeks and from Butte Creek were pooled into two reference populations for calculation of genotypic probabilities under the two kinship hypotheses (full-sib versus unrelated). We selected and grouped pairs of individuals for which full-sib LOD scores were significant ($p < 0.01$), as above, removing them from their sample. The Type II errors were approximately 0.6 for the two samples studied at five loci (Deer Creek, 1996, and Butte Creek, 1996) and 0.4 for the two samples studied at eight or nine loci (Deer Creek, 1994, and Mill Creek, 1996).

We next examined the genotypes of these kin groups for evidence against a simple full-sib hypothesis, i.e., more than four alleles per locus or impossible genotypic combinations, and identified the largest subset of potential full-sibs for further analysis. The remaining individuals in a kinship group could be explained as half-sibs, although this sometimes constrained the genotype of the common parent. For each putative full-sib group, we calculated the probability for the observed array of progeny genotypes under one or more mating-type hypotheses. For example, if there were one AA and two AB in a full-sib group with three members, we would calculate the relevant terms of the binomial probability distribution for {0.5, AA : 0.5, AB}, implying a mating type AA × AB, the trinomial probability distribution for {0.25, AA : 0.5, AB : 0.25, BB}, implying an AB × AB mating, and the tetranomial probability distribution for {0.25, AA : 0.25, AB : 0.25, AC : 0.25, BC}, implying an AB × AC mating (Table 2). The probabilities for each mating type, given the observed progeny, were then multiplied by the frequency of the implied mating type, given baseline allelic frequencies and assuming Hardy–Weinberg–Castle (H–W–C) equilibrium, to yield the likelihood for each possible mating type. The mating type with the maximum likelihood was selected, and the inferred parental genotypes were then substituted into the data set for the juveniles.

Correction for run admixture

Because different spawning runs of chinook salmon can overlap in space and time, a given collection of adults can comprise a mixture of two or more runs. Such an admixture of spring run in broodstock collected for hatchery propagation of winter run (samples W1–W5, Table 1) was first suspected on phenotypic grounds in 1995. Extensive genetic analyses subsequently confirmed this admixture based on linkage disequilibrium (D. Hedgecock et al., unpublished data), a well-known feature of admixed populations (Waples and Smouse 1990). Linkage disequilibrium was measured using the program GENETIX version 3.3 (available at <http://www.univ-montp2.fr/~genetix/genetix.htm>) and found to be significant for 60% of the dilocus combinations. Nonwinter contaminants were identified and removed from the winter-run broodstock sample by the following procedure: (i) individuals with uncommon gametic types contributing to the significance of multiple pairwise

Fig. 2. UPGMA phenogram derived from CSE using raw data at five microsatellite loci (*Ots-2, -3, -9, and -10* and *Oneμ13*) for 41 chinook salmon population samples (see Table 1) from California's Central Valley. The scale indicates genetic distance; numbers at nodes indicate clusters with bootstrap results greater than 50%.



disequilibria were identified, (ii) these individuals were removed from the broodstock sample and the linkage equilibrium of the remainder of the broodstock was verified, (iii) the culled individuals were reassigned to run on the basis of their multilocus genotypes by a maximum likelihood method implemented in the program WHICHRUN (Banks and Eichert 2000), and (iv) individuals with better than even odds of belonging to the winter-run broodstock were returned to that sample, and linkage equilibrium was again verified. This procedure resulted in the removal of 19 of the 140 winter-run broodstock before the present study (i.e., W1–W5 sam-

ples comprise 121 fish in Table 1). A similar procedure was applied here to each of the remaining samples for which more than 10% of all pairwise disequilibrium tests were significant at the 5% level. Smaller genetic distances among the spring, fall, and late-fall runs, however, reduce the power of assignment of individual genotypes to run.

Bootstrap analyses of number of alleles per locus

To compare the numbers of microsatellite alleles observed in different populations, we used the program BOOTSTRAP (M.

Table 3. Sizes of spring-run juvenile chinook salmon samples before and after adjustment for missing data and kinship.

Spring-run sample	No. original sample	No. loci typed	Individuals removed for missing data ^a	No. kin groups ^b	No. parents inferred ^c	Net adjustment for kinship ^d	Size of adjusted sample
Deer Creek, 1994	53	9	-17	4	13	-6	30
Deer Creek, 1996	81	5	-3	14	28	-5	73
Mill Creek, 1996	64	8	-14	11	26	-10	40
Butte Creek, 1996	74	5	-32	7	19	-7	35

^aIndividuals were removed if they were missing more than one of five loci or more than two of eight or nine loci.

^bGroups of individuals among which there was significant probability of being full-sibs, as determined by kinship (Goodnight and Queller 1999).

^cEach kin (full-sib) group has at least two parents; in some kin groups, the presence of more than four alleles or impossible genotype combinations suggested half-sib relationships requiring additional parents to be inferred.

^dNet adjustment is removal of juveniles belonging to kin group followed by substitution with their inferred parents.

Table 4. Heterogeneity testing within four subpopulations of chinook salmon from California's Central Valley with stepwise removal of samples to determine the largest subset of samples with homogeneous allelic frequencies.

Before adjustment ^a			After adjustment ^a		
Pool	Samples ^b	F_{ST}	Pool	Samples ^b	F_{ST}
Central Valley chinook	41 samples	0.0637*	Central Valley chinook ^c	38 samples	0.0745*
Winter	W1, 2, 3, 4, 5, 6, 7, 8 W4, 5	0.0172* 0.0047	Winter	W1, 2, 3, 4, 5, 6, 7, 8 W1, 2, 3, 4, 5, 6, 8	0.0087* 0.0035
Spring	S1, 2, 3, 4, 5, 6, 7, 8, 9, 10	0.0175*	Spring	S1, 2, 3, 4, 5, 6, 7, 8, 9, 10	0.0171*
Spring Deer and Mill creeks	S5, 6, 7, 8, 9, 10 S6, 7, 8, 9, 10	0.0038* 0.0009	Spring Deer and Mill creeks	S5, 6, 7, 8, 9, 10	0.0017
Spring Butte Creek	S1, 2, 3, 4	0.0065*	Spring Butte Creek	S1, 2, 3, 4	0.0045*
Fall and late fall	S1, 4	0.0053*		S2, 3, 4	0.0018
	F1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, LF1, 2, 3	0.0072*	Fall and late fall ^c	F2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, LF1, 2, 3	0.0078*
	F4, 5, 6, 7, 8, 10, 13, 15, 17, 18, 19	0.0011		F4, 5, 6, 7, 8, 10, 11, 13, 15, 16, 17, 18, 20	0.001
Late fall	LF1, 2, 3	0.0032*	Late fall	LF1, 2, 3	0.0034*
	LF1, 2	0.0011		LF1, 2	0.0015
			Central Valley chinook (homogeneous pools) ^c	—	0.0819*

Note: *results significant ($\alpha = 0.05$).

^aAdjustment refers to correction for kinship or run admixture (see text).

^bSee Table 1 for sample names.

^cExcluding fall naturally spawning sample from the Sacramento River, 1995, and Coleman Hatchery, 1993 and 1995.

Barré, Laboratoire Génétique et Pathologie, IFREMER BP 133, 17390 La Tremblade, France, personal communication) to simulate the relationship between sample size and number of alleles by bootstrap sampling for each of 10 loci. For this analysis, population samples were combined within each of the winter, spring from Butte Creek, spring from Deer and Mill creeks, fall naturally spawning, fall hatchery, and late-fall runs. At each sample size (from $n = 1$ to the total number of fish in a given run at a given locus), the program performed 1000 randomized samplings, counting the number of alleles and calculating the mean and standard error. Sign tests were then used to compare allelic diversity among runs at a common sample size of 50.

Population genetic analyses

Allelic frequencies and population genetic statistics — observed and expected heterozygosities, mean number of alleles per locus, pairwise linkage disequilibrium, Wright's F statistics, Nei's (1972) distance (D), and Nei's (1987) minimum distance (D_m) — were calculated using the program GENETIX version 3.3. Tests for ran-

dom mating proportions of genotypes within samples (H-W-C equilibrium) were calculated using the program GENEPOP version 3.1 (Raymond and Rousset 1995). The significance of F_{IS} and F_{ST} ($\alpha = 0.05$) was determined by performing 500 permutations in GENETIX at five loci (*Ots-2*, *-3*, *-9*, and *-10* and *Oney13*). Significance of F_{ST} was also determined for a subset of 26 samples that were scored at all 10 loci (*Ots-1*, *-2*, *-3*, *-5*, *-9*, *-10*, *-104*, and *-107*, *Oney13*, and *Omy77*, Table 1). We did not calculate any distance measures that assume the stepwise mutation model, as recent studies have demonstrated that this model is often falsified by complex microsatellite mutational events and does not perform any better than more traditional infinite allele measures (e.g., Paetkau et al. 1997; Estoup et al. 1998; Colson and Goldstein 1999). Four major subpopulations were evident from the unweighted paired group method using arithmetic averages (UPGMA) phenogram derived from raw data (see below and Fig. 2): all winter, spring from Butte Creek, and spring from Deer and Mill creeks and, lastly, all fall and late-fall population samples as a single clade. Heterogeneity among samples within subpopulations was assessed by signifi-

cance of F_{ST} . When a subpopulation was heterogeneous, sets of homogeneous samples were determined by stepwise removal of samples until F_{ST} for the remaining subpopulation pool was nonsignificant ($\alpha = 0.05$). The order of sample removal was determined by rank of average F_{ST} for within-pool pairwise comparisons. The above tests were repeated using Nei's D_m to explore whether an alternative distance measure resulted in the same findings.

Nei's D and the Cavalli-Sforza and Edwards (1967) chord measure (CSE) were calculated using GENDIST in the program PHYLIP (Felsenstein 1993) for data from five loci and repeated for the 26 population samples that were characterized at 10 loci. Neighbor-joining (Saitou and Nei 1987) and UPGMA trees (Sneath and Sokal 1973) were calculated using NEIGHBOR in PHYLIP. Maximum likelihood trees (Cavalli-Sforza and Edwards 1967) were calculated using CONTML in PHYLIP. Bootstrap results for assessing the frequency of occurrence, and thus significance, of each tree cluster were attained using SEQBOOT and CONSENSE in PHYLIP with 1000 replicates. Trees were visualized using TREEVIEW (Page 1996). We investigated phenetic distances using four alternative distance measures: F_{ST} , D , D_m , and CSE.

Results

Sample adjustments

Adjustments for relatedness were made to four spring-run juvenile samples (Deer Creek, 1994 and 1996, Mill Creek, 1996, and Butte Creek, 1996). Sixty-six of the original 272 individuals in these samples were removed before kinship analysis because of missing genotypic data. Of the remaining 206 individuals, 114 were involved in pairwise comparisons for which the hypothesis of a full-sib relationship was significantly more likely ($p < 0.01$) than the hypothesis that they were unrelated. These related individuals were removed from their samples and replaced with 86 parents whose genotypes were inferred by maximum likelihood methods, resulting in the adjusted sample sizes shown in Tables 1 and 3. Before adjustment for kinship, Butte Creek, 1996, and Mill Creek, 1996, had significant F_{IS} and Mill Creek, 1996, was not in random-mating (H-W-C) equilibrium; after adjustment, these populations have single-locus genotypic proportions that conform to random-mating expectations. Kinship adjustments also reduced the number of significant pairwise linkage disequilibria from 11 to four, over all four samples. More importantly, kinship adjustments changed the allelic frequencies in these samples. In the most extreme case, the frequency of the *Oneμ13¹⁵⁰* allele increased from 0.333 to 0.456 for the spring Butte Creek, 1996, sample after kinship adjustment. The cumulative effects of these allele frequency shifts on F_{ST} estimates are discussed below.

Previously, we found that 60% of pairwise linkage disequilibria were significant in winter-run samples W1–W5, owing to admixture of spring run in broodstock collections (see Materials and methods section). Likewise, 35% of pairwise linkage disequilibria were significant in two of the winter-run carcass samples, W7 and W8. Removing 17 individuals determined to be likely of nonwinter origin (odds ratios ranging from 1.03 to 9.68) from these samples restored equilibrium conditions. Of the 33 nonwinter samples examined in this study, six had more than 10% of pairwise linkage disequilibria significant at the 5% level, indicating potential admixture although at a lower level than winter run. These included spring Butte Creek, 1994 and 1997, fall Coleman Hatchery, 1993 and 1995, Sacramento River, 1995,

and late-fall Coleman Hatchery, 1993. Removing three individuals likely to have been of nonspring origin (odds ratios ranging from 1.50 to 1.55×10^5) from spring Butte Creek, 1994 and 1997, samples brought them into linkage equilibrium. However, significant linkage disequilibrium remained in two hatchery (Coleman Hatchery, 1993 and 1995) and one naturally spawning fall-run sample (Sacramento River, 1995) despite removal of 10, 24, and 10 individuals determined to be of nonfall origin, respectively. These three fall-run samples were thus dropped from further analyses. Two individuals determined to be non-late-fall (odds ratios of 3.50 and 4.10) were removed from the Coleman Hatchery, 1993, sample.

Adjustments for relatedness and run admixture increased the F_{ST} estimated among all samples from 0.0637 to 0.0745 and decreased the F_{ST} estimated within three of the four subpopulations adjusted (0.0172 versus 0.0087 for winter run, 0.0038 versus 0.0017 for spring run from Deer and Mill creeks, 0.0065 versus 0.0045 for spring run from Butte Creek, Table 4). The number of samples having homogeneous allele frequencies also increased. Only two of eight winter-run samples (W4 and W5) were homogeneous before adjustment, whereas seven were homogeneous after adjustment. The juvenile spring-run sample from Deer Creek, 1994, joined a homogenous pool of seven samples from Deer and Mill creeks after correction for kinship. Finally, none of the four Butte Creek samples were homogeneous prior to corrections for kinship and admixture, whereas three were homogeneous after adjustment (Table 4).

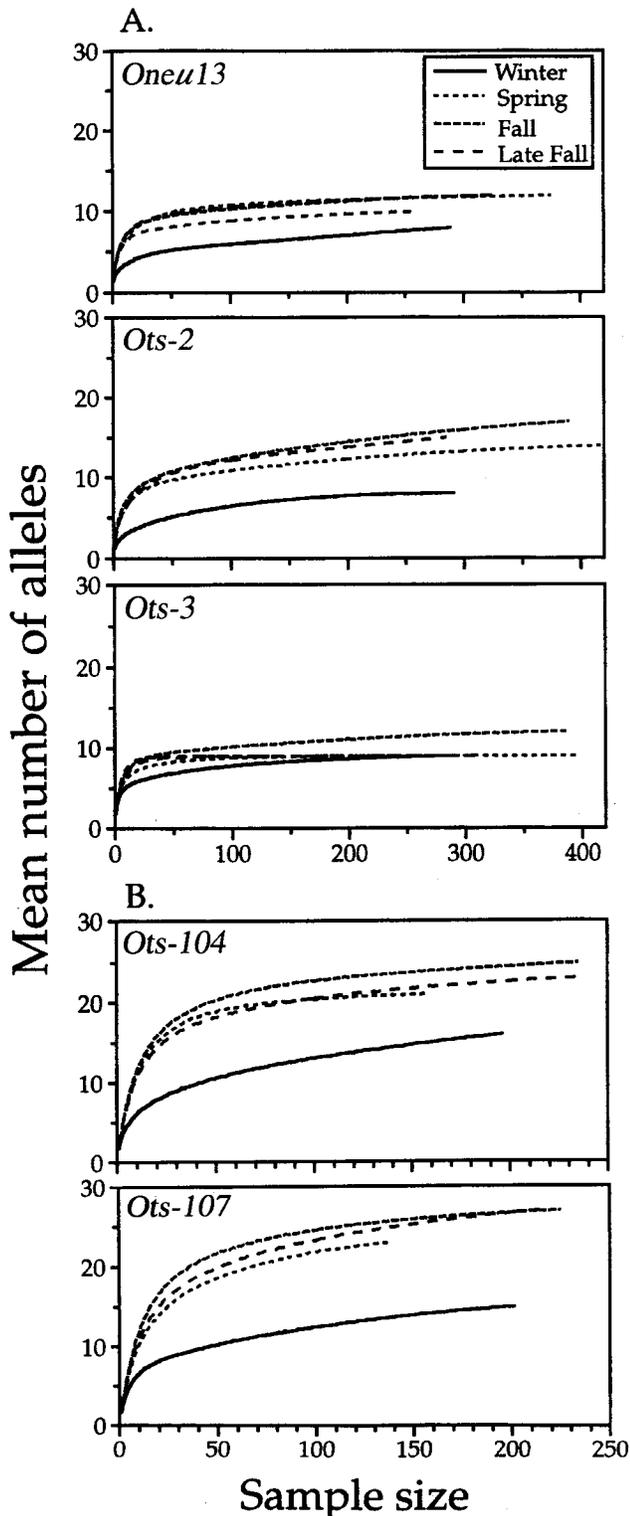
Genetic diversity within population samples

We now focus exclusively on results from adjusted samples, excluding the three fall samples with persistent linkage disequilibrium. Expected and observed heterozygosities within population samples are in close agreement, with means ranging from 0.51 to 0.63 and from 0.54 to 0.62, respectively (Table 1). Thirty-three of 38 tests for H-W-C equilibrium and significance of F_{IS} values within populations conform to random-mating expectations (Table 1). We note the few exceptions: significant F_{IS} values for two of 10 spring-run samples (Butte and Deer creeks, 1997), two of 20 fall-run samples (Stanislaus River, 1994, and Merced Hatchery, 1995), and one of three late-fall samples (Keswick, 1995). These results are supported by exact probability tests for H-W-C equilibrium, except that the test for the Stanislaus, 1994, fall sample is not significant and the test for the naturally spawning Merced River, 1995, fall-run sample is significant. Only one of 38 population samples had more than 10% of pairwise loci combinations in linkage disequilibrium (late-fall Coleman Hatchery, 1993) despite adjustment for admixture. The sample was retained because it was homogeneous with the late-fall Keswick, 1995, sample and the pooled samples were in linkage equilibrium.

Genetic heterogeneity within subpopulations

Samples within the four subpopulations determined from the major nodes of Fig. 2 come from different geographical sites and collection years and from naturally spawning or hatchery populations. Tests of homogeneity within these subpopulations resulted in the following population groupings (Table 4). Seven winter-run samples are homogeneous

Fig. 3. Results from bootstrap simulations for (A) dinucleotide and (B) tetranucleotide microsatellites illustrating the number of alleles detected in chinook salmon population samples as a function of sample size. Each curve was generated using 1000 iterations for each successive sample size. Homogeneous subpopulations were used except for spring run, where Butte, Deer, and Mill Creek samples were combined.



if the carcass sample of 1996 is removed. Three spring-run samples from Butte Creek are homogenous if the 1994 sample is removed. All six spring-run samples from Deer and Mill creeks are homogeneous. Thirteen fall samples are homogeneous after American River, 1995, Stanislaus River, 1994, Merced River, 1995, Coleman Hatchery, 1994, and all three late-fall samples are removed. No subset of the fall anomalies is homogeneous, but two late-fall population samples are homogeneous (Coleman Hatchery, 1993, and Keswick, 1995). The anomalous late-fall sample (Coleman Hatchery, 1995) is more closely related to the homogeneous late-fall subpopulation ($F_{ST} = 0.0072$) than to the homogeneous fall subpopulation ($F_{ST} = 0.0174$). Similar tests using D_m rather than F_{ST} resulted in the same homogeneous pools as above with one exception; it was not necessary to remove the fall Coleman Hatchery, 1994, sample before attaining a homogeneous subpopulation for fall run. Likewise, F_{ST} homogeneity tests performed for the 26 population samples scored at 10 loci affirm the same sample components among the five subpopulations except that two additional anomalous fall populations were identified; hatchery samples from the Feather and Merced rivers, 1995, were dropped to allow homogeneity among the remaining fall populations. This series of within-subpopulation tests of homogeneity define five major, more or less homogeneous subpopulations of chinook salmon in the Central Valley: winter, spring from Butte Creek, spring from Deer and Mill creeks, fall, and late fall. We note that temporal samples from the same geographic location are generally contained within the homogeneous pool for their respective subpopulation. The exceptions are distributed over different subpopulations (winter Sacramento River, 1996, spring Butte Creek, 1994, fall Stanislaus, 1994, and late-fall Coleman Hatchery, 1995). Thus, temporal variation within geographic site is not significant relative to the greater differences among subpopulations.

Comparisons among subpopulations

Bootstrapping results illustrating the number of alleles as a function of sample size are shown for the five most polymorphic loci (*Ots-2*, *-3*, *-104*, and *-107* and *Oneu13*, Fig. 3). Both the locus concerned and run sampled are important with respect to the number of alleles observed. For example, a sample size of 100 is sufficient to encounter 90% of the alleles at *Ots-3* in spring, fall, or late-fall runs, but in winter run, the minimum sample size required is 124. Likewise, a sample size of 137 is required for nonwinter runs at *Oneu13*, but winter requires 207. For the three more polymorphic loci, even larger samples sizes are necessary.

The simulated allele richness curves (Fig. 3) suggest that winter run consistently has fewer alleles per locus than do the other three runs. Winter run also has lower observed and expected heterozygosity and lower mean number of alleles per locus (4.4) than any of the other runs (6.7, 7.7, and 8.8 for spring, fall, and late fall, respectively, Table 1). Table 5 compares the number of alleles per locus for different populations at a common sample size of 50. In comparing winter with the other populations, fewer alleles are observed at eight or nine of 10 loci, which is significant by sign test. Likewise, within spring run, the pooled Butte Creek sample

has fewer alleles at seven of 10 loci than the pooled sample from Mill and Deer creeks.

Pooling of homogeneous samples within the five major subpopulations discussed in the preceding section increases the F_{ST} estimate for Central Valley chinook from 0.075 (calculated from 38 independent samples) to 0.082 ($p = 0.00$). Diversity among these five major subpopulations is illustrated by a UPGMA phenogram of relative genetic distance (Fig. 4). All four distance measures (F_{ST} , CSE, D , and D_m) resulted in topographically equivalent trees. Moreover, three methods of phenetic tree construction (neighbor-joining, UPGMA, and maximum likelihood estimation) yielded the same branching topology. Bootstrap values for the tree based on CSE were often higher than for the other trees (Fig. 4). Bootstrap values greater than 70% are obtained for all branch clusters, indicating statistically defined distinction of all five major subpopulations (Hillis and Bull 1993). Greater genetic distinctiveness of winter run is illustrated by the markedly larger genetic distance values in pairwise comparisons with all other runs. Spring Butte Creek samples are next most distinct, followed by spring Mill Creek and Deer Creek samples. Fall and late-fall runs have the lowest genetic distance estimate indicated by their inner cluster (Fig. 4).

Individual loci make markedly different contributions to average genetic distance among subpopulations, and different measures of genetic distance rank loci differently (Table 6). Spearman rank correlations indicate that F_{ST} is negatively, although not significantly, correlated with heterozygosity. Nei's D , on the other hand, is positively and significantly ($p < 0.05$) correlated with heterozygosity. Nei's D and D_m are both positively and significantly ($p < 0.05$) correlated with CSE, but their positive correlation with each other is not significant.

Discussion

The most important finding of this study is that chinook salmon of the Central Valley in California have substantial genetic diversity and structure. Analysis of microsatellite DNA variation reveals five distinct subpopulations that, with the exception of our discovery of two distinct lineages of spring run, are congruent with the winter, spring, fall, and late-fall spawning runs that have long been recognized (Fisher 1994). That such biological diversity has survived more than 100 years of massive habitat destruction, exploitation, and artificial propagation (Yoshiyama et al. 1998) is perhaps surprising but at the same time encouraging of efforts to protect these populations. Our data retrospectively support the designation of winter run and spring run as evolutionary significant units protected under the U.S. Endangered Species Act (NMFS 1994, 1999; Waples 1995). Winter run, whose blend of ocean- and stream-type life history characteristics is unique among all chinook salmon (Healey 1991), is the most distinctive of the subpopulations in the Central Valley. The next most distinctive subpopulations are the spring runs, particularly those in Butte Creek, which have unique life history adaptations (Yoshiyama et al. 1996). Formerly the most abundant chinook salmon throughout the Central Valley, spring runs are presently found in only a few tributaries of the Sacramento River, primarily those considered in this study (Fisher 1994;

Yoshiyama et al. 1996, 1998). Finally, fall and late-fall runs, although closely related, are significantly different at 10 microsatellite markers and differ in geographic range, run timing, and size at maturity (Fisher 1994).

The two most divergent runs in the Central Valley, winter run and spring run from Butte Creek, show signs of having experienced past reductions in size (bottleneck). Winter run has significantly fewer alleles at most loci and lower average heterozygosity than all other runs. Spring run from Butte Creek also has fewer alleles than spring run from Mill and Deer creeks. Thus, bottlenecks and genetic drift, which can affect highly polymorphic microsatellite markers in particular (Hedrick 1999), may have accelerated the genetic divergence measured for these spawning runs (Nei 1987). Whether the bottlenecks and reductions in genetic diversity coincided with precipitous declines of these stocks within the last 100 years (Yoshiyama et al. 1998) or resulted from earlier events is a matter for speculation. We note, however, that drift might also explain heterogeneity among samples from Butte Creek, as the abundance of spring run in this creek ranged from a high of 8700 adults in 1960 to a low of 10 fish in 1979 (CDFG 1998). Because mean age at reproduction is 3 years in spring-run chinook (Fisher 1994), the 1994 and 1997 samples were drawn from a weak year-class tracing back to the 1979 low.

Despite spatial and temporal overlap of chinook salmon spawning runs in the Central Valley, we have no evidence for natural hybridization among runs. A commonly held view is that most spring-run populations have hybridized with fall run and that Butte Creek spring run in particular has hybridized with the Feather River fall hatchery stock (Yoshiyama et al. 1998). Genotypic proportions in the Butte Creek spring run conform to random-mating expectations, however, and this subpopulation clusters farther from the fall run than does the spring run from Deer and Mill creeks, not closer as expected under the hybridization hypothesis. Admixture of runs can nevertheless occur in samples and appears to be a likely cause for significant linkage disequilibrium in broodstock collected for a winter-run hatchery supplementation program as well as in eight other samples from naturally spawning populations. Our ability to detect run origin among mixed samples is well defined for winter run owing to the genetic distinctiveness of this run. Identifying the provenance of individuals from runs other than the winter run is more difficult, however, owing to their greater similarity. Tests of linkage equilibrium nevertheless remain useful for detecting admixture in population samples (Waples and Smouse 1990).

Artificial hybridization in hatcheries, where mating choices are not made by the fish, could pose a risk to conservation of chinook salmon diversity in the Central Valley. Spring run were unwittingly hybridized with winter run in the early years of a hatchery supplementation program for the latter (D. Hedgecock et al., unpublished data). Broodstock are now typed for microsatellites upon collection, and only those with significant genotypic odds of being winter run are used in the hatchery. Here, we present evidence for significant linkage disequilibrium in two samples from the Coleman Hatchery fall stock (1993 and 1995), suggesting admixture and possible hybridization between fall and spring runs. Temporal overlap of both runs in Battle Creek

makes simultaneous capture of adults from both runs highly likely, and morphology does not provide reliable discrimination between them. This problem was first identified in the early operation of the Coleman Hatchery (Cope and Slater 1957) and remains a challenging problem, given limited power for individual genetic identification of non-winter-run fish. Admixture or hybridization of fall into late fall might also explain the significant linkage disequilibrium and failure of random-mating equilibrium in samples from the Coleman Hatchery late-fall stock. However, we have even less power for identifying non-late-fall individuals, given the smaller genetic distance apparent between late fall and fall. Finally, the Feather River Hatchery has long been suspected of hybridizing spring and fall runs on the basis of coded-wire tag returns (Yoshiyama et al. 1998). Yet, we observe neither linkage disequilibrium nor failure of random-mating equilibrium in samples from the hatchery or naturally spawning populations of fall chinook in the Feather River. In addition, the Feather River Hatchery samples considered in this study are not significantly different from the fall run, providing no evidence for past hybridization with Central Valley spring run.

Evaluating the extent of spatial and temporal heterogeneity within runs is critical to the successful development and implementation of conservation and management plans (Small et al. 1998; Nielsen et al. 1999; Tessier and Bernatchez 1999). Neither spatial nor temporal variation within runs appears to be important, however, compared with the much larger differences among the runs. Fall run, which was sampled extensively in both the Sacramento River and San Joaquin River drainages of the Central Valley, including both hatchery and naturally spawning populations, is largely homogeneous. Spring run from Deer and Mill creeks is homogeneous (although distinct from the second spring-run subpopulation in Butte Creek). Spatial variation is not an issue for the remaining runs, which presently have very restricted geographic ranges. On a temporal scale, overall homogeneity within the major subpopulations of chinook salmon in the Central Valley affirms a general lack of significant year-to-year variation within runs. Samples of winter run from 1991 to 1997 were homogeneous, with the exception of the 1996 carcass sample. Samples of spring run from Deer and Mill creeks were likewise homogeneous over a 4-year period, from 1994 through 1997. Fall run and late-fall run were each homogeneous over a 3-year period.

Finally, we find that local populations of chinook salmon in the Central Valley have proportions of microsatellite genotypes that conform to those expected from random mating among individuals. Random mating has, of course, been confirmed by innumerable studies in the vast literature on salmonid population genetics. The importance of confirming random mating for microsatellite markers in this study lies in the contrast that it provides for those exceptional failures of equilibrium genotypic distribution that we have attributed to kinship among juveniles or run admixture in samples.

Although substantial genetic structure of chinook salmon in the Central Valley was apparent from a cluster analysis of raw population data, the picture was brought into sharper focus by adjusting samples for kinship and run admixture. Relatedness of individuals within juvenile samples has long been recognized as a problem for estimation of population

allelic frequencies (Allendorf and Phelps 1981). In the past, geneticists have simply tried to avoid such samples, but today, we have the problem of sampling threatened populations whose juvenile life stages may be more abundant and more readily collected than adults. In the case of spring-run chinook in the Central Valley, for example, the inaccessibility of much of the spawning habitat greatly limits collection of spawning adults or postspawned carcasses (Yoshiyama et al. 1996, 1998). Fortunately, highly polymorphic microsatellite DNA markers now provide more statistical power for detecting kinship than was previously afforded by allozymes (O'Reilly et al. 1998; Fontaine and Dodson 1999). Thus, to maximize population genetic information for the threatened spring run, we corrected for relatedness in juvenile samples by replacing the genotypes of likely siblings with the most likely genotypes of their parents.

The second type of sample adjustment was for admixture of different spawning populations, which has also long been recognized as an issue in delineating spawning stocks that are incompletely isolated in space and time (Campton and Johnston 1985; Waples and Smouse 1990; Estoup et al. 1998). Again, microsatellite DNA markers now provide more statistical power than was previously afforded by allozymes, both for detecting linkage disequilibrium, a signal of sample admixture (Waples and Smouse 1990), and for assigning individuals to populations of origin (Banks and Eichert 2000). For samples with unusual levels of linkage disequilibrium, we first identified and removed individuals responsible for the disequilibrium and then reassigned them to their sample only if they were more likely to belong to the remainder of their sample than to any other run.

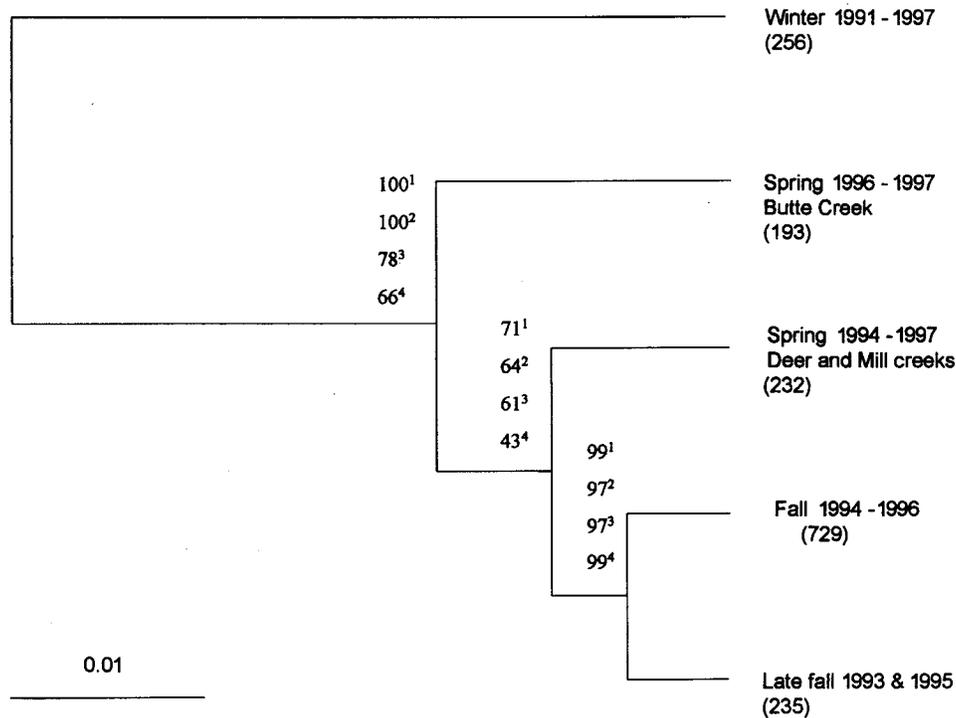
Although the difference in population structure before and after these adjustments appears small (cf. Figs. 2 and 4), the increase in the precision and accuracy of the population genetic analysis is significant and important for future applications of microsatellites to conservation. The two types of adjustments resulted in the net removal of 116 individuals, less than 5% of individuals genotyped in the study, with the following consequences. The number of populations displaying departure from single and multilocus equilibrium at a nominal significance level of 5% was reduced from 12 to five and from eight to one, respectively, further increasing confidence in the application of the H-W-C principle in mixed-stock analyses and individual assignment. Variation among samples within runs was also reduced, as evidenced by reductions in numbers of alleles and within-run F_{ST} values, resulting in an increase in the sizes of homogeneous pools of samples and the precision of allele frequency estimates for runs. Finally, genetic distance among runs was dramatically increased, as evidenced by an increase in F_{ST} from 0.064 to 0.082. These changes all favor the further application of these markers in conservation efforts.

Average allele frequency variance among the runs of chinook salmon in the Central Valley ($F_{ST} = 0.082$ for five loci or 0.078 for 10 loci) is substantial compared with the genetic divergence detected by microsatellite studies of other anadromous salmonids. Based on five microsatellite loci, Olsen et al. (1998) reported F_{ST} values of 0.026 and 0.032 for comparisons within and between odd- and even-year populations of pink salmon (*Oncorhynchus gorbuscha*), respectively. Small et al. (1998) estimated an F_{ST} of 0.058, based

Table 5. Means and standard errors obtained from bootstrap simulations of the number of alleles observed at a common sample size of 50.

	Winter		Spring				Fall					
	Mean	SE	Butte Creek		Mill and Deer creeks		Naturally spawning		Hatchery		Late fall	
			Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
<i>Ots-1</i>	2.37	0.02	3.35	0.02	3.22	0.02	2.95	0.01	3.71	0.02	3.60	0.02
<i>Ots-2</i>	5.20	0.03	8.08	0.03	10.49	0.03	10.74	0.04	12.76	0.04	10.53	0.04
<i>Ots-3</i>	6.86	0.03	7.97	0.01	8.06	0.02	9.50	0.02	8.85	0.01	8.97	0.01
<i>Ots-5</i>	2.00	0.00	1.32	0.01	1.99	0.00	2.78	0.02	2.59	0.02	2.37	0.02
<i>Ots-9</i>	3.00	0.00	3.00	0.00	3.00	0.00	3.00	0.00	3.00	0.00	3.00	0.00
<i>Ots-10</i>	2.42	0.02	4.59	0.02	4.32	0.02	4.74	0.01	5.00	0.02	3.69	0.02
<i>Ots-13</i>	5.23	0.03	8.64	0.03	10.16	0.02	9.47	0.03	9.72	0.03	8.11	0.02
<i>Omy-77</i>	1.63	0.02	3.37	0.02	3.51	0.02	4.83	0.01	4.44	0.02	3.27	0.03
<i>Ots-104</i>	10.67	0.04	16.18	0.02	17.69	0.02	20.17	0.03	26.16	0.07	18.10	0.04
<i>Ots-107</i>	10.33	0.04	15.88	0.04	20.42	0.02	23.48	0.03	21.76	0.05	19.95	0.05

Fig. 4. UPGMA phenogram derived from CSE using adjusted data from 10 microsatellite loci (*Ots-1*, -2, -3, -5, -9, -10, -104, and -107, *Omy-13*, and *Omy-77*) for 26 chinook salmon population samples from California's Central Valley. Populations were grouped in homogeneous subpopulations as defined in the Results section for 10 loci. Numbers in parentheses indicate mean sample sizes; numbers at nodes indicate the percentage of bootstrap simulations supporting a given cluster (1, CSE-UPGMA; 2, *D*-UPGMA; 3, CSE-neighbor joining; 4, *D*-neighbor joining). The scale indicates genetic distance.



on three microsatellite loci, for coho salmon (*Oncorhynchus kisutch*) in the Fraser River. McConnell et al. (1997), in a survey of eight microsatellite loci of Atlantic salmon (*Salmo salar*), reported an F_{ST} of 0.072 among populations in the Inner Bay of Fundy having different life histories.

Divergence among Central Valley chinook salmon runs has also been estimated using allozymes, mtDNA, and sequence variation in a major histocompatibility (MHC) gene. In a study of variation at 39 allozyme-coding loci, Bartley et al. (1992) estimated F_{ST} to be only 0.01 among five samples of chinook salmon from the Central Valley. However, the allozyme-based estimate includes more shared monomorphic loci than were allowed in our selection of polymorphic microsatellite markers. In addition, the identity of the winter-

run sample in the Bartley et al. (1992) study has been questioned, and a recent allozyme study has detected substantial divergence among runs (D. Teel and G. Winans, NMFS, Seattle, Wash., personal communication). Thus, protein and microsatellite DNA markers may paint more similar pictures of divergence of salmon runs in the Central Valley than presently appreciated. On the other hand, mtDNA and MHC markers appear, at first glance, to show greater divergence among runs than do microsatellite markers. Our average F_{ST} estimate for 10 microsatellite loci (0.078) is notably less than the F_{ST} of 0.24 calculated from the mtDNA data of Nielsen et al. (1994) or the 0.129 estimate from the MHC class II $\beta 1$ exon (Kim et al. 1999). However, these last two marker types have substantially fewer alleles than is typical

Table 6. Single-locus average genetic distance and heterozygosity estimates for chinook salmon from California's Central Valley.

Locus	H_e	F_{ST}	CSE	D	D_m
<i>Ots-104</i>	0.87884	0.0473	0.3308	0.5735	0.0258
<i>Ots-107</i>	0.86563	0.0573	0.4355	0.7009	0.0348
<i>Ots-3</i>	0.80355	0.0299	0.2086	0.2592	0.0156
<i>Oncu-13</i>	0.71252	0.1172	0.3720	0.3895	0.0481
<i>Ots-2</i>	0.66378	0.1687	0.4753	0.4872	0.0808
<i>Ots-10</i>	0.52825	0.0512	0.1428	0.1955	0.0204
<i>Ots-1</i>	0.4732	0.0227	0.1276	0.1429	0.0074
<i>Ots-9</i>	0.46908	0.1053	0.1932	0.2368	0.0229
<i>Omy-77</i>	0.19023	0.0531	0.1222	0.1309	0.0066
<i>Ots-5</i>	0.13883	0.1611	0.1415	0.1393	0.0139

Note: Population samples were combined into heterogeneous pools as defined for 10 loci. Loci were ranked according to the magnitude of heterozygosity.

of microsatellites. A number of researchers (Hedrick 1999 and references therein) have shown that, for highly variable loci such as microsatellites, F_{ST} is constrained by high within-population diversity. As gene diversities within subpopulations and the total population approach 1.0, the difference between them, which should represent the diversity among subpopulations, approaches zero. Thus, F_{ST} estimates for polymorphic marker types may be lower than those obtained from less variable marker types, as we observe when comparing our average estimate from microsatellites with those from mtDNA and MHC data. F_{ST} ranges from 0.023 to 0.169 for our microsatellite loci (Table 6) and is negatively, although not significantly, correlated with heterozygosity. In contrast, Nei's D is positively and significantly correlated with heterozygosity, which suggests that this measure may not be as affected by high heterozygosity as F_{ST} . Nei's D values of 0.395 for mtDNA (from data in Nielsen et al. 1994) and 0.648 for MHC (from data in Kim et al. 1999) are within the range that we observe for microsatellite loci (0.131–0.701). All markers suggest that winter run is the most distinctive subpopulation, followed by spring run and then fall and late fall, so there is general concordance among different types of markers concerning divergence among chinook salmon runs in the Central Valley.

In conclusion, we find substantial genetic diversity revealed by microsatellite DNA markers among California's Central Valley chinook salmon populations. Despite well-justified concern for the irreversible loss of genetic heterogeneity, owing to substantial habitat loss, water diversion, harvest pressure, and significant hatchery influence (Yoshiyama et al. 1998), we are impressed with the genetic diversity and structure that remain. Evidence for run admixture and resultant hybridization in hatcheries remains a pressing concern, and there is need for increased genetic power for verifying the run identity of hatchery broodstock. However, the potential for discrimination among the five chinook subpopulations with microsatellite loci has obvious importance for management and protection of specific stocks. With increasing numbers of salmon being listed as endangered species, we have urgent need for the best possible resolution among populations. Water policy, fishing quotas, and other regulatory decisions require precise information on specific stocks at risk. We report sufficient

genetic diversity among Central Valley chinook spawning runs to enable estimation of contributions to mixed samples and individual identification of winter run. The increase in precision offered by microsatellite loci, together with the sample adjustment methods developed in this study, demonstrates great promise for resolving closely related populations in a conservation context.

Acknowledgements

We thank S. Hamelberg and associates of the U.S. Fish and Wildlife Service and J. Navicky and associates of the California Department of Fish and Game for providing samples used in this study. We also thank R. Brown, S. Greene and associates of the California Department of Water Resources for supporting this research, and members of the salmon project at the Bodega Marine Laboratory for their assistance. We thank K. Belkhir (Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France) for various modified versions of GENETIX. We also thank two anonymous reviewers whose comments substantially improved this paper. This study was supported by funds from the Department of Water Resources, CALFED, and the U.S. Fish and Wildlife Service. This paper is dedicated to the memory of Nat Bingham, who recruited us to the cause of salmon conservation and fostered this work.

References

- Allendorf, F.W., and Phelps, S.R. 1981. Use of allelic frequencies to describe population structure. *Can. J. Fish. Aquat. Sci.* **38**: 1507–1514.
- Banks, M.A., and Eichert, W. 2000. WHICHRUN (version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. *J. Hered.* **91**: 87–89.
- Banks, M.A., Blouin, M.S., Baldwin, B.A., Rashbrook, V.K., Fitzgerald, H.A., Blankenship, S.M., and Hedgecock, D. 1999. Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tshawytscha*). *J. Hered.* **90**: 281–288.
- Bartley, D., Gall, G.A.E., Bentley, B., Brodziak, J., Gomulkiewicz, R., and Mangel, M. 1992. Geographic variation in population genetic structure of chinook salmon from California and Oregon. *Fish. Bull. U.S.* **90**: 77–100.
- California Department of Fish and Game (CDFG). 1998. A status review of the spring-run chinook salmon (*Oncorhynchus tshawytscha*) in the Sacramento River drainage. Candidate Species Status Rep. 98-01.
- Campton, D.E., and Johnston, J.M. 1985. Electrophoretic evidence for admixture of native and nonnative rainbow trout (*Salmo gairdneri*) in the Yakima River, Washington, (USA). *Trans. Am. Fish. Soc.* **114**: 782–793.
- Cavalli-Sforza, L.L., and Edwards, A.W.F. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution*, **32**: 550–570.
- Colson, I., and Goldstein, D.B. 1999. Evidence for complex mutations at microsatellite loci in *Drosophila*. **152**: 617–627.
- Cope, O.B., and Slater, D.W. 1957. Role of Coleman Hatchery in maintaining a king salmon run. U.S. Dep. Inter. Fish Wildl. Serv. Res. Rep. 47. pp. 1–22.
- 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. 1993. PHYLIP (Phylogeny Inference Package), version 3.57c. Department of Genetics, University of Washington, Box 357360, Seattle, WA 98105, U.S.A.
- Fisher, F.W. 1994. Past and present status of Central Valley chinook salmon. *Conserv. Biol.* **8**: 870–873.
- Fontaine, P.M., and Dodson, J.J. 1999. An analysis of the distribution of juvenile Atlantic salmon (*Salmo salar*) in nature as a function of relatedness using microsatellites. *Mol. Ecol.* **8**: 189–198.
- Goodnight, K.F., and Queller, D.C. 1999. Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Mol. Ecol.* **8**: 1231–1234.
- Healey, M.C. 1991. The life history of chinook salmon. *In Pacific life histories. Edited by C. Groot and L. Margolis.* University of British Columbia Press, Vancouver, B.C. pp. 39–52.
- Hedrick, P.W. 1999. Perspective: highly variable genetic loci and their interpretation in evolution and conservation. *Evolution*, **53**: 313–318.
- Hillis, D.M., and Bull, J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**: 182–192.
- Kim, J.T., Parker, K.M., and Hedrick, P.W. 1999. MHC differentiation in Sacramento River chinook salmon. *Genetics*, **151**: 1115–1122.
- McConnell, S.K.J., Ruzzante, E.D., O'Reilly, P.T., Hamilton, L., and Wright, J.M. 1997. Microsatellite loci reveal highly significant genetic differentiation among Atlantic salmon (*Salmo salar* L.) stocks from the east coast of Canada. *Mol. Ecol.* **6**: 1075–1089.
- 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 (NMFS). 1994. Endangered and threatened species; status of Sacramento River winter-run chinook salmon. *Fed. Register* **59**(No. 2): 440–450.
- National Marine Fisheries Service (NMFS). 1998. Endangered and threatened species; west coast chinook salmon; listing status change; proposed rule. *Fed. Register* **63**(No. 45): 11482.
- National Marine Fisheries Service (NMFS). 1999. Endangered and threatened species; threatened status for two chinook salmon evolutionary significant units (ESUs) in California. *Fed. Register* **64**(No. 179): 50 394 – 50 415.
- Nehlsen, W. 1994. Salmon stocks at risk: beyond 214. *Conserv. Biol.* **8**: 867–869.
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* **106**: 283–292.
- Nei, M. 1987. *Molecular evolutionary genetics.* Columbia University Press, New York.
- Nelson, R.J., and Beacham, T.D. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. *Anim. Genet.* **30**: 228–229.
- Nielsen, J.L., Tupper, D., and Thomas, W.K. 1994. Mitochondrial DNA polymorphism in unique runs of chinook salmon (*Oncorhynchus tshawytscha*) from the Sacramento – San Joaquin River basin. *Conserv. Biol.* **8**: 882–884.
- Nielsen, E.E., Hansen, M.M., and Loeschcke, V. 1999. Genetic variation in time and space: microsatellite analysis of extinct and extant populations of Atlantic salmon. *Evolution*, **53**: 261–268.
- Olsen, J.B., Seeb, L.W., Bentzen, P., and Seeb, J. 1998. Genetic interpretation of broad-scale microsatellite polymorphism in odd-year pink salmon. *Trans. Am. Fish. Soc.* **127**: 535–550.
- O'Reilly, P.T., Herbinger, C., and Wright, J.M. 1998. Analysis of parentage determination in Atlantic salmon. *Anim. Genet.* **29**: 363–370.
- Paetkau, D., Waits, L., Clarkson, P., Craighead, L., and Strobeck, C. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (*Ursidae*) populations. *Genetics*, **147**: 1943–1957.
- Page, R.D.M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.* **12**: 357–358.
- Raymond, M., and Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* **86**: 248–249.
- Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- 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. *Can. J. Fish. Aquat. Sci.* **53**: 833–841.
- Small, M.P., Beacham, T.D., Withler, R.E., and Nelson, R.J. 1998. Discriminating coho salmon (*Oncorhynchus kisutch*) populations within the Fraser River, British Columbia, using microsatellite DNA markers. *Mol. Ecol.* **7**: 141–155.
- Sneath, P.H.A., and Sokal, R.R. 1973. *Numerical taxonomy.* Freeman Press, San Francisco, Calif.
- Tessier, N., and Bernatchez, L. 1999. Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Mol. Ecol.* **8**: 169–179.
- Waples, R.S. 1995. Evolutionary significant units and the conservation of biological diversity under the Endangered Species Act. *Am. Fish. Soc. Symp.* **17**: 8–27.
- Waples, R.S., and Smouse, P.E. 1990. Gametic disequilibrium analysis as a means of identifying mixtures of salmon populations. *Am. Fish. Soc. Symp.* **7**: 439–458.
- Wright, J.M., and Bentzen, P. 1996. Microsatellites — genetic markers of the future. *Rev. Fish Biol. Fish.* **4**: 384–388.
- Yoshiyama, R.M., Gerstung, E.R., Fisher, F.W., and Moyle, P.B. 1996. Historical and present distribution of chinook salmon in the Central Valley drainage of California. *Sierra Nevada Ecosystem Project: final report to Congress. In Assessments, commissioned reports, and background information.* Center for Water and Wildlife Resources, University of California at Davis, Davis, Calif. pp. 309–362.
- Yoshiyama, R.M., Fisher, F.W., and Moyle, P.B. 1998. Historical abundance and decline of chinook salmon in the Central Valley region of California. *N. Am. J. Fish. Manage.* **18**: 487–521.