Summary of genetic data collected for rapid response Chinook salmon assignment, and evaluation of new markers and assignment tools.

Report for FY2010

Updated August 10, 2011

By:
Christian Smith and Denise Hawkins
Applied Program in Conservation Genetics
Abernathy Fish Technology Center
1440 Abernathy Creek Road
Longview, Washington 98632
(360) 425-6072

And
John Carlos Garza
Southwest Fisheries Science Center
110 Shaffer Road
Santa Cruz, California 95060
(831) 420-3903

Submitted to:
Scott Hamelberg
Coleman National Fish Hatchery
24411 Coleman Fish Hatchery Road
Anderson, California 96007
(530) 365-8622

INTRODUCTION

The Sacramento River system supports four distinct "runs" of Chinook salmon (*Oncorhynchus tshawytscha*): fall run, late-fall run, spring run, and winter run. Winter run Chinook salmon leave the ocean and enter the Sacramento River from November through June in an immature reproductive state. They migrate into the upper reaches of the Sacramento River, hold in cool waters released from Shasta Dam, and spawn from May through August between the city of Red Bluff (river mile [RM] 245) and Keswick Dam (RM 302), the upstream limit of migration. Most winter-run Chinook salmon spawn at age three, with the remainder spawning at ages two and four (Hallock and Fisher 1985).

Winter run Chinook salmon were listed as "threatened" under the Endangered Species Act in 1989 and their status was changed to "endangered" in 1994 (59 Federal Register 440). In 1989, the U.S. Fish and Wildlife Service (USFWS) began propagating winter run Chinook salmon to supplement natural production. The winter-run Chinook salmon supplementation program was initially located at the Coleman National Fish Hatchery (NFH) on Battle Creek, a tributary of the Sacramento River. In 1998, the program was moved to the newly constructed Livingston Stone NFH located at the base of Shasta Dam, to increase returns to the main stem Sacramento River.

Potential broodstock fish for the program at Livingston Stone NFH are trapped, primarily at Keswick Dam, and subjected to genetic analysis to confirm winter run identity. The protocol presently used to distinguish winter run from non-winter run Chinook salmon among individuals trapped at Keswick Dam was developed by researchers at University of California at Davis (UCD) in collaboration with USFWS (Greig and Banks 1999; Banks et al. 2000). Genetic screening was performed at UCD's Bodega Marine Laboratory for several years prior to being moved to the USFWS' Abernathy Fish Technology Center (AFTC) in 2003-2004. Since 2004, broodstock selection for the winter run Chinook salmon program has been carried out as a joint project between AFTC, the Red Bluff FWO (RBFWO) and Coleman NFH.

In the time since the genetic screening protocol was developed, new genetic markers and statistical tools for population assignment have been developed. Nearly all agency and university genetic laboratories working on Chinook salmon are sharing a standardized set of microsatellite markers (Seeb et al. 2007) that allow them to share data and thus perform genetic analyses much more efficiently. The microsatellite loci in the shared panel (commonly referred to as the GAPS microsatellites) are different from and more variable than the microsatellite loci used in the established protocol for winter run Chinook salmon identification. Single nucleotide polymorphisms (SNPs) are another type of genetic marker that is now being broadly used in Chinook salmon. SNPs possess many technical advantages over microsatellites (Smith et al. 2005), and NOAA Fisheries Southwest Fisheries Science Center (SWFSC) has recently developed a high-resolution SNP baseline for Central Valley Chinook salmon (Clemento et al. in press; J.C. Garza, manuscript in prep). New genetic assignment models (Rannala and Mountain 1997; Piry et al. 2004; Anderson et al. 2008) have also been developed and implemented in software packages (e.g.: GENECLASS2 (Piry et al. 2004) and ONCOR (Steven Kalinowski; available at http://www.montana.edu/kalinowski/Software/ONCOR.htm)). Differences between the models implemented in these newer programs and the model implemented in the established protocol, which uses WHICHRUN (Banks and Eichert 2000), may or may not result in improved accuracy for assigning Central Valley Chinook salmon as winter run or non-winter run.

The goals of this report are:

1) To summarize population genetic results for winter run Chinook salmon based on all data collected at AFTC (2004-2009) and also data collected at UCD where available. Specifically, we evaluated trend data for heterozygosity, allelic richness, genotypic disequilibrium, and effective population size.
2) To assess new genetic markers and assignment techniques and compare assignments made with these to assignments made using the established markers and protocol.

METHODS & MATERIALS

Genotype data

Available genotypes for three marker sets (rapid response microsatellites-7 loci, GAPS microsatellites-13 loci, and SNPs-94 loci) were analyzed. For mixture samples (samples assigned to run type) our initial quality control target was no more than 15% data missing from genotypes (i.e. individuals with genotypes <85% complete were to be excluded), however the actual numbers of failed loci allowed varied by marker set (see below). The following three paragraphs describe the sources of data for each marker set.

Genotypes for rapid response microsatellites for samples collected from 1996-2003 were provided by Robert Null, RBFWO. Genotypes for samples collected from 2004-2010 were taken from the AFTC database. The samples were collected either during boat carcass surveys (coordinated by Robert Null and Kevin Niemela, RBFWO), or at Keswick Dam as part of the broodstock collection program for Livingstone Stone NFH (Table 1). A total of 6,346 multi-locus genotypes were available. For some collections, only six of the seven loci were genotyped, so for this marker set we tolerated up to 2 missing loci (29% missing data) per individual. Correspondence analysis was also used to identify outlier individuals (i.e. individuals which appeared vastly different from all other Central Valley Chinook salmon). Such individuals may represent very rare genetic lineages, but more often represent genotyping errors or migrants from other basins. Individual fish sampled during boat surveys and at Keswick Dam were categorized as “mixture”, and a subset of these were subsequently categorized as “winter run” based on genetic assignment results (Table 1). The baseline used to evaluate these samples was the one used in the established rapid response protocol, which consisted of 915 Central Valley Chinook salmon samples collected between 1991 and 2000.

Genotypes for GAPS microsatellites for samples collected in 1992-1995, 1997, 1998, 2001, 2003 and 2004 were taken from the GAPS database. These genotypes were generated by the SWFSC lab in Santa Cruz, CA. In addition, samples collected at Keswick Dam in 2002, 2007 and 2008 were analyzed using the GAPS microsatellites at AFTC. Individuals genotyped at AFTC and missing more that 2 loci (15% missing data) were excluded from the analysis. As described

above, all individuals sampled at Keswick Dam were categorized as “mixture”, and a subset of these were subsequently categorized as “winter run” based on genetic assignment results. Individuals for which genotypes were taken from the GAPS baseline were all classified as “winter run” (Table 1). The baseline used to evaluate these samples was GAPS (downloaded March 2010), which consisted of 857 Central Valley Chinook salmon samples collected between 1992 and 2004.

Genotypes for 94 of the 96 SNPs presently used by NOAA Fisheries for analyzing mixture samples and reconstructing pedigree relationships of Chinook salmon were collected for samples from 2002, 2003, 2009, and 2010 (chemicals for the SNPs Ots_102867-609 and Ots_111312-435 were not available). All analyzed individuals had previously been assigned as winter run using the established protocol. Genotypes for these fish were collected in a collaboration between the AFTC and the SWFSC lab. Individuals missing more that 14 loci (15% missing data) were excluded from the analysis. The baseline used to evaluate these samples was the one presently in use for mixture analysis by the SWFSC, and consisted of 1,338 Central Valley Chinook salmon samples collected between 1992 and 2007

In addition to the full set of 94 SNPs we also wanted to test a smaller set that might be processed at lower cost. For this we ranked the 94 SNPs in order of decreasing allele frequency differences between winter run and the other runs. The 23 SNPs with the largest allele frequency differences were then selected as a subset. Finally, because we did not have a mixture genotyped with the SNPs, we simulated one for the mixture analyses by combining the winter run individuals we did have (Table 1) with several individuals that were removed from the baseline (25 spring run each from Mill Creek, Deer Creek, and Butte Creek; and 25 fall run from Battle Creek). Spring run are the most similar lineage to winter run (Banks et al. 2000; Garza et al. 2007), and our goal in including mostly spring run in our simulated mixture was to produce conservative estimates of the resolution which might be provided by SNPs. Within this simulated mixture, baseline fish were subject to the same missing data filter (15% missing data) as all other fish genotyped at AFTC.

Genetic profile of winter run Chinook salmon

Average heterozygosity across loci was calculated for each collection using GDA (Lewis and Zaykin 2001). Genetic diversity within each collection was also measured as average allelic richness across loci, the number of alleles observed in a collection corrected via rarefaction for unequal numbers of individuals per collection. Allelic richness was calculated using the program HPRARE (Kalinowski 2005). The log likelihood ratio statistic (G test) was used to test for Genotypic Disequilibrium (GD) between each pair of loci in each collection using GENEPOP (Rousset 2008).

Two measures of effective population size (N_e) were calculated based on each marker set. Variance N_e (related to the amount of variance in allele frequencies) was estimated using the temporal method (Waples 2005) as implemented in the program SALMONNB (Waples et al. 2007). For this analysis we assumed that, for each collection listed in Table 1, 5.7% of individuals were 2 years old, 91.7%were 3 years old, and 2.6% were 4 years old (USFWS 2009). Inbreeding N_e (related to the rate of increase in inbreeding) was estimated using the linkage disequilibrium method (Waples 2006) implemented in the program LDNE (Waples and Do 2008). In both cases, alleles with frequencies < 1% were omitted from the calculations.

Comparison of assignment methods

The established rapid response protocol (AFTC Genetics SOP # 21) involves calculation of the likelihood of assignment of an individual to each population in the baseline. The ratio of likelihoods of assignment to winter run versus the next most likely population is then evaluated. If the log of odds (LOD) ratio is greater than 2.0, then it is suggested that the probability that the individual is non-winter run is < 1/100 (Banks and Eichert 2000) and the individual is classified as winter run.

We compared results obtained using WHICHRUN to those obtained using other methods. The genotype assignment probability calculations described by Rannala and Mountain (1997) have proven very useful in fisheries applications and are implemented in the programs GENECLASS2 and ONCOR. As a further refinement, ONCOR uses the conditional maximum likelihood to estimate mixture proportions (Millar 1987) and incorporates these as priors for assignment tests in an iterative fashion. Instead of treating each unknown fish independently, ONCOR thus uses information contained in the mixture to assign individual fish.

All mixed samples (Table 1) were assigned as either winter run or non-winter run using WHICHRUN, GENECLASS2 and ONCOR. For WHICHRUN we followed the established rapid response protocol. In order to set comparable confidence criteria for the other programs, we set probability cut-offs for both GENECLASS2 and ONCOR at 99% (i.e. any individual with <99% probability of being assigned to winter run was classified as non-winter run).

Comparison of assignment using different marker sets

For optimal comparison of marker sets, one must have a single set of baseline samples, several representative mixture samples, and some samples of known origin genotyped for all marker sets. Since the baselines we are considering here were genotyped previously as parts of other projects this optimal situation was not realized. In comparing results for the marker sets considered here, it is important to recognize this limitation.

First, in order to gain a sense of the separation of winter run from non-winter run by each baseline we performed correspondence analysis using the program GENETIX (Belkhir et al. 2004), and evaluated the results of this analysis in light of the assignment tests using WHICHRUN.

Second, we performed correspondence analysis on the samples for which we had both rapid response and GAPS markers genotyped. This allowed us to evaluate relative divergence between winter run and non-winter run based on these two marker sets without the complication of having different individuals sampled for each marker set.

Third, we plotted LOD scores obtained from WHICHRUN using both microsatellite data sets, the 23 best SNPs (based on allele frequency differences) and the full set of 94 SNPs. Since we did not have a mixture sample genotyped for SNPs, we simulated one as described above. This artificial mixture was then assigned using WHICHRUN and the established protocol, and LOD frequency histograms were plotted.

RESULTS & DISCUSSION

Genotype data

Of 6,346 mixture individuals for which rapid response genotypes were available, 544 exhibited more than two failed loci and were thus excluded from further analysis. Another 38 individuals appeared to be extreme outliers based on correspondence analysis (results not shown). Examination of the sample identities revealed that most (31/38) of the outlier individuals were analyzed in the year when UCD and AFTC were in the process of standardizing allele calls. Standardization between the laboratories was accomplished by running sets of common samples for some alleles at each locus, and extrapolating the remaining allele conversions linearly. Since many of the baseline alleles had not been run at AFTC during the first several years, it is likely these outlier samples are the results of allele-calling discrepancies between mixture and baseline individuals.

Of 721 mixture individuals for which GAPS genotypes were produced at AFTC, 3 exhibited more than two failed loci, and were thus excluded from analysis.

Of 188 mixture individuals for which SNP genotypes were available, none exhibited more than fourteen failed loci, so none were excluded from analysis. Further, for the analysis of the 23 best SNPs (see below) no samples exhibited more than three failed loci, so none were excluded from analysis.

The number of alleles observed per data set was greatest for the GAPS markers (320 alleles), intermediate for the 94 SNP markers (188 alleles) and least for the rapid response markers (103 alleles; Appendix 1).

Genetic profile of winter run Chinook salmon

The winter run Chinook salmon program at Livingston Stone NFH was designed to enhance the natural population of winter run salmon that spawns below Keswick Dam. Several measures were taken in designing protocols for the hatchery specifically for the purpose of avoiding negative genetic impacts to the natural population. Our present analysis revealed no clear trends in either heterozygosity or allelic richness over time (Figure 1). While values for both statistics have varied over time, no overall increase or decrease was apparent.

In contrast with genetic diversity measures, GD did appear to change in recent years, exhibiting spikes in 2004 and 2010. One potential cause of high GD could be inclusion of a few non-winter run individuals (e.g. if a small number of spring run fish collected at Keswick Dam were incorrectly classified as winter run). Another possible explanation is a reduction in the number of families represented in the winter run Chinook salmon population. For example, such an increase in GD might be expected if a few families made disproportionately large contributions to the sample of adults collected at Keswick Dam. Additional analyses including correspondence analyses of individuals from each return year, and sibship analysis might provide further insight regarding the cause of high GD in some years.

Estimates of N_e varied based on the markers considered and on the method of estimation (variance N_e versus inbreeding N_e; Table 2). In interpreting the results of the two different methods, it is important to recall that variance N_e is used to make inferences about the present generation, whereas inbreeding N_e is used to make inferences about the parental generation (Waples 2005). Another consideration is that estimates based on the rapid response markers were based on much larger, and presumably more representative samples in most cases (calculations were not performed for samples containing <30 individuals). The estimates of variance N_e reveal an increase in N_e at approximately 6-year intervals, but the estimates of inbreeding N_e do not indicate increases in recent years. This makes sense given the increase in genotypic disequilibrium observed during this time (Figure 2), and could also indicate a greater relative contribution by the hatchery population during recent years. As with the genotypic disequilibrium results reported above, comparison of this result with the proportions of hatchery and wild individuals captured at Keswick Dam over time would be helpful.

Comparison of assignment methods

Assignment of individuals as winter run or non-winter run was highly consistent regardless of which assignment method was employed (Table 3). Concordance was highest among the methods implemented in ONCOR and GENECLASS2 (r^2 = 0.995-0.999, depending on the marker set), and lowest between WHICHRUN and GENECLASS2 when using the rapid response markers (r^2 = 0.964). Examination of the individuals for which there were discrepancies revealed that the most common conflict was over fish which both ONCOR and GENECLASS2 assigned as winter run, but WHICHRUN assigned as non-winter run (Figure 3a). This result indicates that the established protocol is relatively conservative in assigning fish to winter run.

Comparison of assignment using different marker sets

Correspondence analysis, combined with the results of assignment using WHICHRUN, allowed us to evaluate the distinction between winter run and non-winter run individuals using the different marker sets. For all marker sets examined, the primary axis of the correspondence analysis (x-axis in Figures 4-5) distinguished winter run from non-winter run individuals. The degree of separation between the winter and non-winter run individuals varied based on the markers examined. However, much greater overlap was apparent in the analysis employing the rapid response markers (Figure 4a) than in analyses employing other markers (Figure 4b, Figure 5). Figure 4a indicates that many individuals which are part of the winter run cluster (and thus likely are winter run) are assigned as non-winter run. Again, this suggests that the established protocol is conservative in assigning individuals to winter run.

It is important to note that many more individuals are represented in Figure 4a than in the others (4b, 5a, 5b). Therefore, a direct comparison of specific individuals is not possible and a better comparison of baselines would optimally include identical individuals genotyped with different markers. When we repeated the correspondence analysis including only individuals for which we had both rapid response and GAPS genotypes, however, the GAPS markers still appear to do a better job of separating the samples into two clusters along the x-axis (Figure 6).

While correspondence analysis plots are a good way to visualize groups in multi-dimensional data, LODs are the most direct way to measure how well alternative marker sets would perform for assigning individuals as winter run or non-winter run. We plotted histograms of LOD scores for all mixture fish from Table 1, and for individuals in the simulated mixture for the SNP data (Figure 7). The histograms show a clear increase in the distance between mean LOD values for winter run versus non-winter run populations when 94 SNPs or the GAPS markers are used. A result which is not apparent from these figures is that, when the 94 SNPs were used, 66 non-winter run individuals from the simulated mixture had 0 likelihood of being winter run and are thus missing from the bottom panel of Figure 7. The greater separation in LOD scores for winter run versus non-winter run provides strong evidence that accuracy in rapid response identification could be improved by switching to either the GAPS markers or the 94 SNPs. Use of the 23

SNPs also resulted in increased separation of LODs between winter run and non-winter run; however, the difference was not as pronounced as for the GAPS markers or 94 SNPs.

One other factor to consider regarding markers is the age of the samples included in each of the baselines. In order to account for changes in baseline population allele frequencies over time, it is recommended that the baselines should be updated approximately every 5-10 years (Waples 1990). In order to be confident that our baseline allele frequencies are representative of those in the populations of Central Valley Chinook salmon, we recommend that additional samples be added to the rapid response baseline immediately and to the GAPS and SNP baselines within the next five years or so.

Conclusions

We found no evidence that genetic diversity in the winter run Chinook salmon population has increased or decreased in recent generations. An increase in GD in 2004 suggested that variance in survival among families may have been particularly high sometime around 2001. Effective population size estimates were highly variable. However, differences in patterns indicated by variance and inbreeding N_e in recent years merit further investigation incorporating other types of data.

Comparison of assignment methods revealed that the two newer methods (implemented in GENECLASS2 and ONCOR) were slightly more consistent with one another than either was with WHICHRUN. For the application of distinguishing winter run from non-winter run we do not expect that differences among the programs would lead to major differences, especially if a marker set which provided stronger distinction among winter run and non-winter run populations was used (Figure 3b). GENECLASS2 is much more widely used and cited than the other two programs are combined; however, incorporation of the mixture proportion estimates may provide an advantage to ONCOR. Winter run, spring run and fall run Chinook salmon have different temporal migration distributions, and the run composition of individuals collected in the upper

Sacramento River is expected to change throughout the winter run broodstock collection period. By incorporating information about the mixture of fish collected each week at Keswick into the individual assignments to run type, we expect that ONCOR may provide increased assignment accuracy. Further, the model implemented in ONCOR for predicting assignment accuracy eliminates a source of bias which made earlier models overestimate accuracy (Anderson et al. 2008). ONCOR is thus expected to yield more realistic estimates of assignment accuracy than GENECLASS2.

Although there are several other marker sets available (Banks and Jacobson 2004; Garza et al. 2007), we chose to examine only the two that are standardized among all agencies working on Chinook salmon in California's Central Valley. The GAPS microsatellites were standardized among all labs working on Chinook salmon in order to provide information to the Pacific Salmon Commission more efficiently. The SNP panel examined here was developed by the SWFSC as an outgrowth of the GAPS dataset development process and for the benefit of all parties working on Chinook salmon in California, and SNP alleles are standardized across laboratories by definition. Use of a standardized marker set for the Livingston Stone broodstock program would 1) allow that program to benefit from updated baseline data contributed by other projects, 2) allow our partners to benefit from the data we collect, 3) provide some power to assign non-winter run individuals to spring run or fall run.

Based on the results presented here we expect that the accuracy and efficiency of the winter run Chinook salmon genetic identification program could be improved by switching to either GAPS microsatellites or SNPs. Although the established protocol is unlikely to lead to inclusion of non-winter run fish in the broodstock, it seems very likely that divergent families within winter run (i.e. winter run families which were not represented in the original baseline) could be excluded, which would have the effect of decreasing genetic diversity of the hatchery broodstock, relative to the natural winter run population. Since both the GAPS and SNP markers provide increased distinction between winter run and non-winter run individuals, the number of ambiguous individuals (i.e., individuals with LOD scores ~ -1 to 1) should be reduced. We also

expect that using ONCOR, rather than WHICHRUN, to perform assignments could be advantageous.

The cost for processing the GAPS microsatellites is comparable to that for processing the rapid response microsatellites. The annual chemical costs for processing 96 SNPs in rapid response mode using equipment available at AFTC would be approximately $5,000 higher than for either of the microsatellite sets. This is because SNP genotyping platform available at AFTC requires simultaneous processing of 94 samples (47 rapid response individuals) to achieve maximum cost efficiency. Rapid response events often include fewer than 47 individuals, so maximum cost efficiency would be unlikely for this application. The additional chemical cost, however, would allow use of the largest and fastest growing genetic baseline for Chinook salmon in California, and would thus reduce the need for the rapid response program to fund baseline updates. Use of the SNPs for the rapid response program would also mean that mixture analyses performed by NOAA Fisheries could assign winter run Chinook salmon to individual crosses performed at Livingston Stone NFH. Finally, use of the SNP markers would allow the rapid response data to be contributed to the shared SNP baseline and would thus greatly increase the value of the data generated by this program to conservation of Chinook salmon in California.

ACKNOWLEDGEMENTS

The genetic evaluation of Chinook salmon as potential broodstock at Livingston Stone NFH has been conducted since 2004 as a collaborative project between several USFWS offices including the Red Bluff Fish and Wildlife Office, Coleman NFH, Livingston Stone NFH, and Abernathy Fish Technology Center. Work by staff at all of these offices contributed to the data presented in this report. The SNP markers and baseline used in the analyses presented here were developed by the SWFSC Molecular Ecology Team. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

LITERATURE CITED

Anderson, E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 65(7):1475-1486.

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.

Banks, M. A., and D. P. Jacobson. 2004. Which genetic markers and GSI methods are more appropriate for defining marine distribution and migration of salmon? North Pacific Anadromous Fish Commission Technical Note 5:39-42.

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 & Aquatic Sciences 57(5):915-927.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Clemento, A. J., A. Abadía-Cardoso, H. A. Starks, and J. C. Garza. In press. Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, *Oncorhynchus tshawytscha*. Molecular Ecology Resources.

Garza, J. C., S. M. Blankenship, C. Lemaire, and G. Charrier. 2007. Genetic population structure of Chinook salmon (Oncorhynchus tshawytscha) in California's Central Valley. Draft Final Report for CalFed Project "Comprehensive Evaluation of Population Structure and Diversity for Central Valley Chinook Salmon":82pp.

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(4):318-320.

Hallock, R. J. and F. W. Fisher. 1985. Status of winter-run Chinook salmon, Oncorhynchus tshawytscha, in the Sacramento River. California Department of Fish and Game, Anadromous Fisheries Branch, January 25, 1985 office report, 28 pp.

Kalinowski, S. T. 2005. HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. Molecular Ecology Notes 5:187-189.

Lewis, P. O., and D. Zaykin. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html.

Millar, R. B. 1987. Maximum likelihood estimation of mixed stock fishery composition. Can. J. Fish. Aquat. Sci. 44:583-590.

Piry, S., and coauthors. 2004. GENECLASS22: A software for genetic assignment and first-generation migrant detection. Journal of Heredity 95(6):536-539.

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Science 94:9197-9201.

Rousset, F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8(1):103-106.

Seeb, L. W., and coauthors. 2007. Development of a standardized DNA database for Chinook salmon. Fisheries 32(11):540-552.

Smith, C. T., C. M. Elfstrom, J. E. Seeb, and L. W. Seeb. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14:4193-4203.

Waples, R. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics 7:167-184.

Waples, R. S. 1990. Temporal changes of allele frequency in Pacific salmon - Implications for mixed-stock fishery analysis. Can. J. Fish. Aquat. Sci. 47(5):968-976.

Waples, R. S. 2005. Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? Molecular Ecology 14(11):3335-3352.

Waples, R. S., and C. Do. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8:753–756.

Waples, R. S., M. Masuda, and J. Pella. 2007. SALMONNb: a program for computing cohort-specific effective population sizes (N-b) in Pacific salmon and other semelparous species using the temporal method. Molecular Ecology Notes 7(1):21-24.

Table 1.Number of samples for which genetic data are available for 7 rapid response microsatellites (RR), 13 GAPS microsatellites (GAPS), and 94 SNPs. The header row indicates whether each sample contained a mixture of run types (mix) or only the subset of fish that assigned as winter run (WR).

				Genetic Data		
Year	Site / Source	RR mix	RR WR	GAPS mix	GAPS WR	SNP WR
1992	GAPS baseline				18	
1993	GAPS baseline				4	
1994	GAPS baseline				16	
1995	GAPS baseline				17	
1996	Boat survey	33	33			
1997	Boat survey	110	98			
1997	GAPS baseline				3	
1998	Boat survey	484	436			
1998	Keswick Dam	152	124			
1998	GAPS baseline				17	
1999	Boat survey	238	234			
1999	Keswick Dam	41	23			
2000	Keswick Dam	150	94			
2000	Red Bluff Diversion Dam	8	6			
2001	Boat survey	397	366			
2001	Keswick Dam	230	202			
2001	Red Bluff Diversion Dam	6	2			
2001	GAPS baseline				35	
2002	Boat survey	239	227			
2002	Keswick Dam	251	193	119	101	47
2003	Boat survey	343	313			
2003	Keswick Dam	144	107			47
2003	GAPS baseline				10	
2004	Keswick Dam	496	345			
2004	GAPS baseline				15	
2005	Boat survey	203	182			
2005	Keswick Dam	468	378			
2006	Keswick Dam	365	268			
2007	Keswick Dam	212	154	219	156	
2007	Red Bluff Diversion Dam	5	0			
2008	Keswick Dam	377	191	380	198	
2009	Keswick Dam	282	240			47
2010	Keswick Dam	530	408			47
Total		5,764	4,624	718	590	188

Table 2. Estimates of effective population size of the winter run Chinook population based on variance in allele frequencies (temporal method) and increase in inbreeding (linkage disequilibrium method). Cell colors indicate low (red) to high (green) values. Samples used for these calculations are in the "WR" columns from Table 1.

	Variance N_e			Inbreeding N_e		
Year	RR	GAPS	SNP	RR	GAPS	SNP
1996	123.3			infinite		
1997	infinite			infinite		
1998	476.8			399.9		
1999	2566.2			infinite		
2000	148.2			infinite		
2001	442.1	58.8		438.7	308.2	
2002	infinite	infinite	365.8	671.1	283.8	93.3
2003	4527.7		infinite	infinite		182.4
2004	313.1			105.6		
2005	1669.4			252.6		
2006	527.8			123.5		
2007	498.4	107		234.1	135.7	
2008	infinite	infinite		249.7	800.3	
2009	10643.9		2089	312.4		278.7
2010	290.4		148.5	57		76.9

Table 3. Concordance between three computer programs in assigning the samples from Table 1 in columns labeled "mix" as either winter run or non-winter run. Results are shown based on the rapid response baseline and the GAPS baseline.

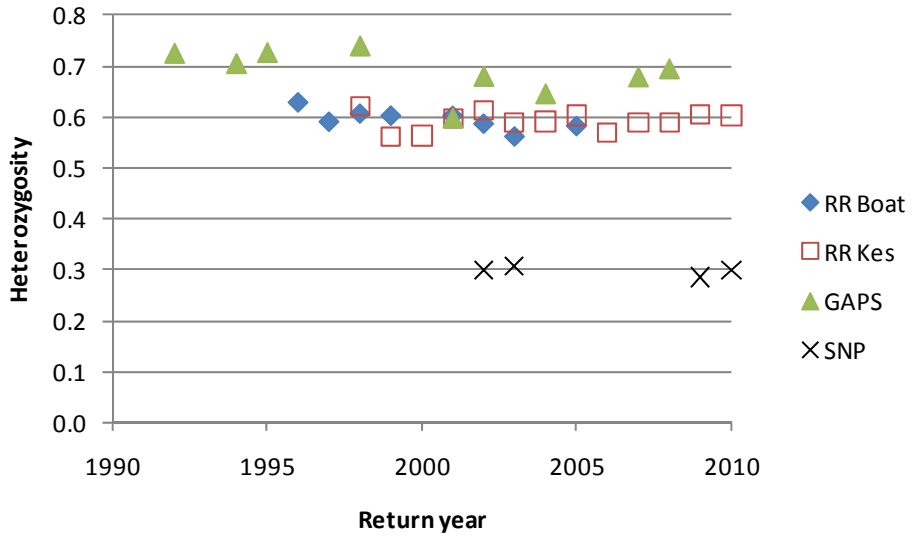
Rapid Response

	ONCOR	GENECLASS2	WHICHRUN
ONCOR	-		
GENECLASS2	0.995	-	
WHICHRUN	0.967	0.964	-

GAPS

	ONCOR	GENECLASS2	WHICHRUN
ONCOR	-		
GENECLASS2	0.999	-	
WHICHRUN	0.997	0.996	-

a)



b)

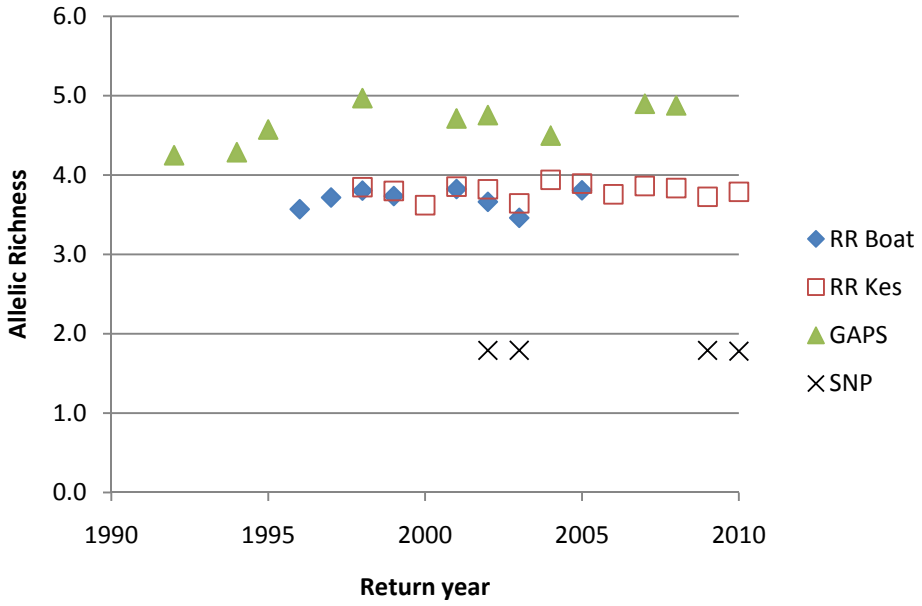


Figure 1. Genetic diversity in collections of winter run Chinook salmon as measured by observed heterozygosity (a) and allelic richness (b). Results are shown for rapid response (RR) markers for samples from boat surveys and Keswick Dam, and for GAPS and SNP marker sets. Note that SNPs are lower than the other marker sets on the y-axis in both figures due to the fact that SNPs only exhibit two alleles per locus.

Figure 2

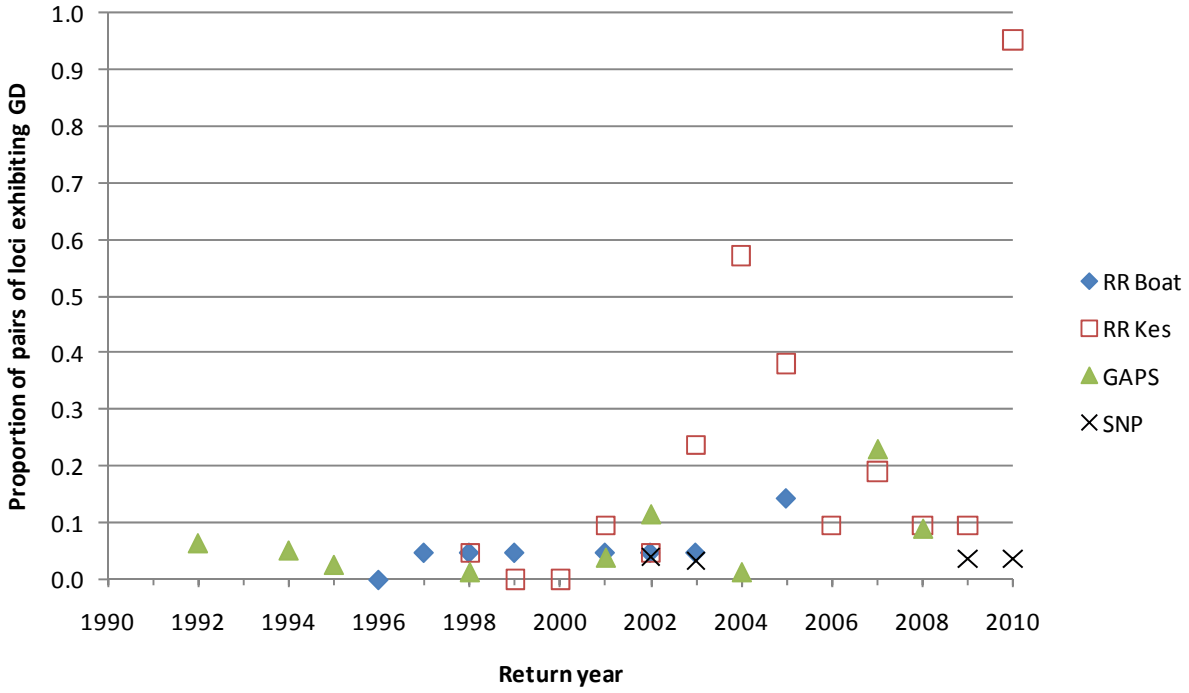
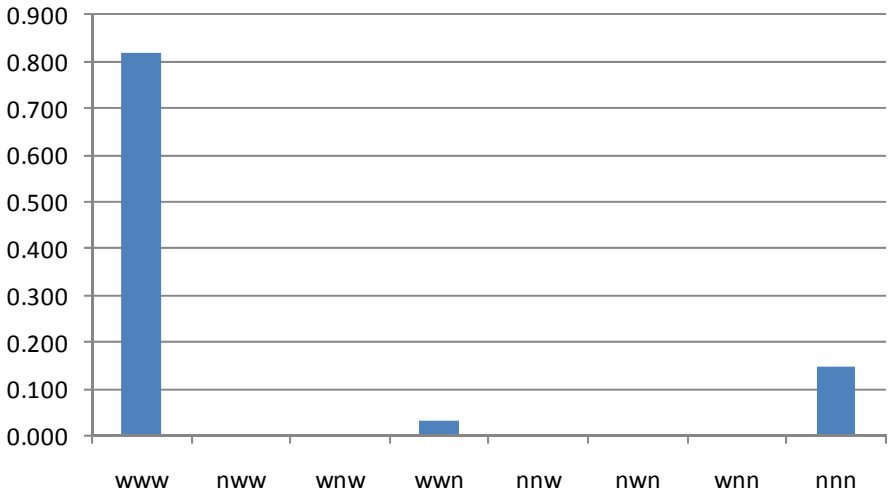


Figure 2. Genotypic disequilibrium (GD) observed in collections of winter run Chinook salmon. The y-axis indicates the proportion of pairwise locus comparisons exhibiting significant (α=0.05) GD within each collection. Results are shown for rapid response (RR) markers for samples from boat surveys and Keswick Dam, and for GAPS and SNP marker sets.

a)



b)

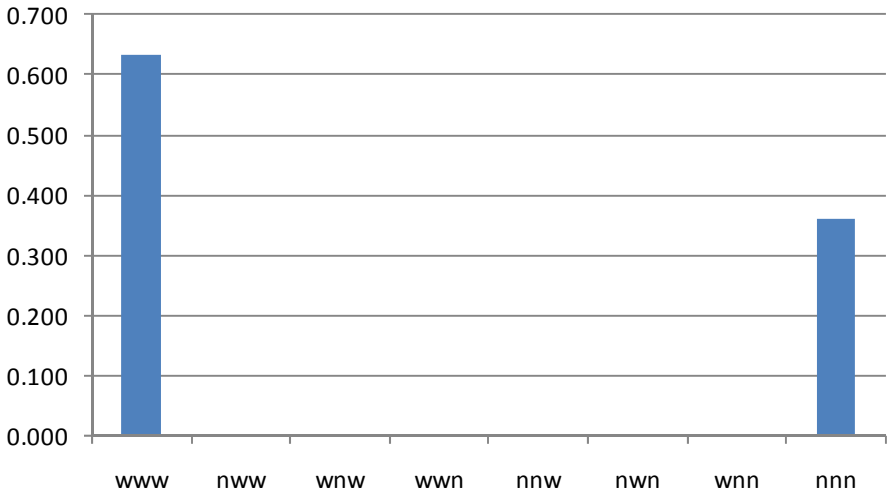


Figure 3. Proportions of mixture samples (Table 1) assigned as winter run (w) and non-winter run (n) by genetic assignment methods implemented in the programs ONCOR, GENECLASS2, and WHICHRUN (in that order). Results are shown for rapid response markers (a) and GAPS markers (b). Each category along the x-axis corresponds to a specific combination of assignment results. For example "wwn" indicates fish that were assigned as winter run by ONCOR, winter run by GENECLASS2, and non-winter run by WHICHRUN.

a)

b)

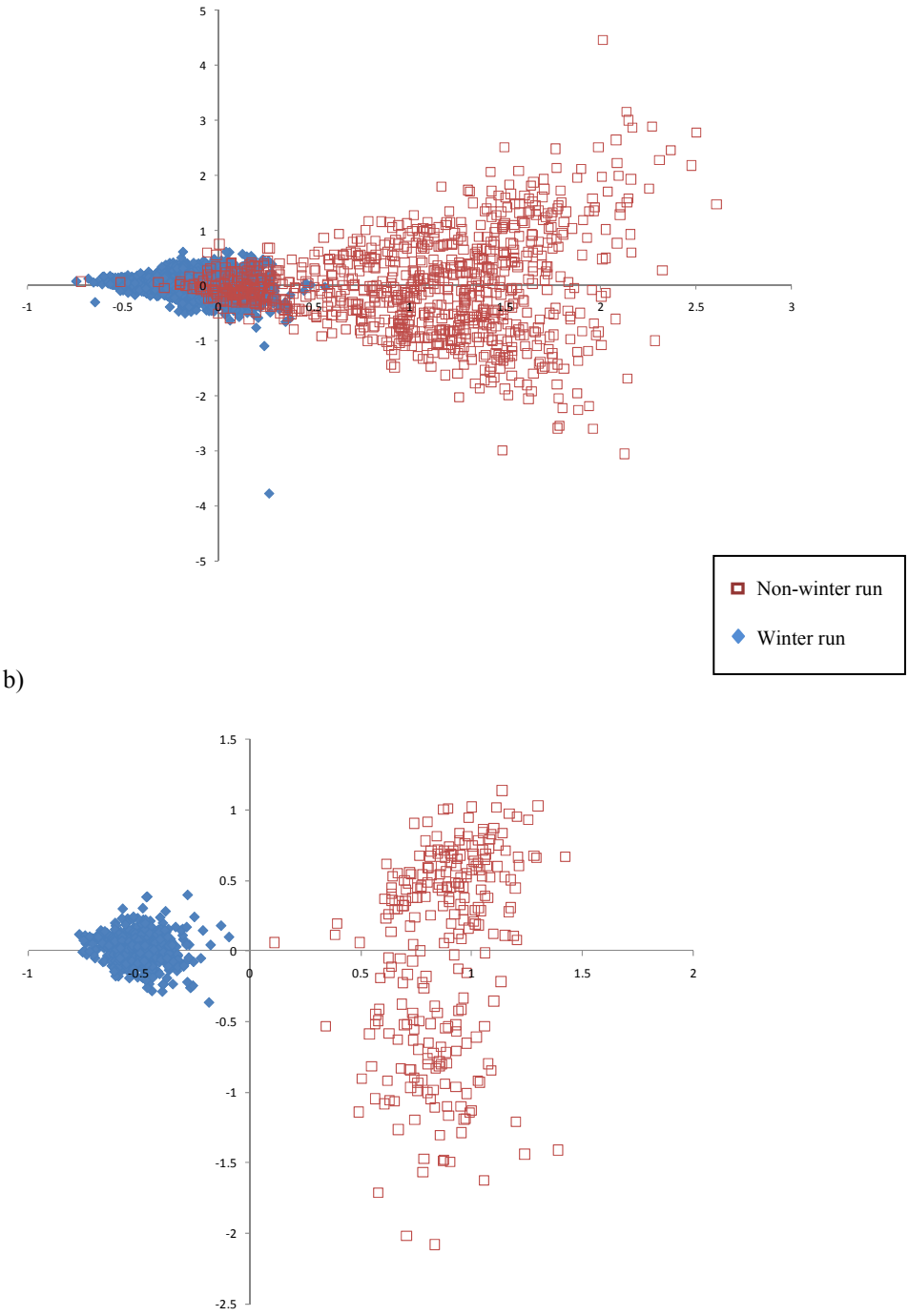


Figure 4. Correspondence analysis of mixture samples (Table 1) based on rapid response (a) and GAPS (b) marker sets. The shape and color of each point indicate genetic assignment using the established rapid response protocol (substituting the GAPS markers and baseline in b). These figures were intended to highlight differences between markers, but the different individuals represented in each figure may also have contributed to differences.

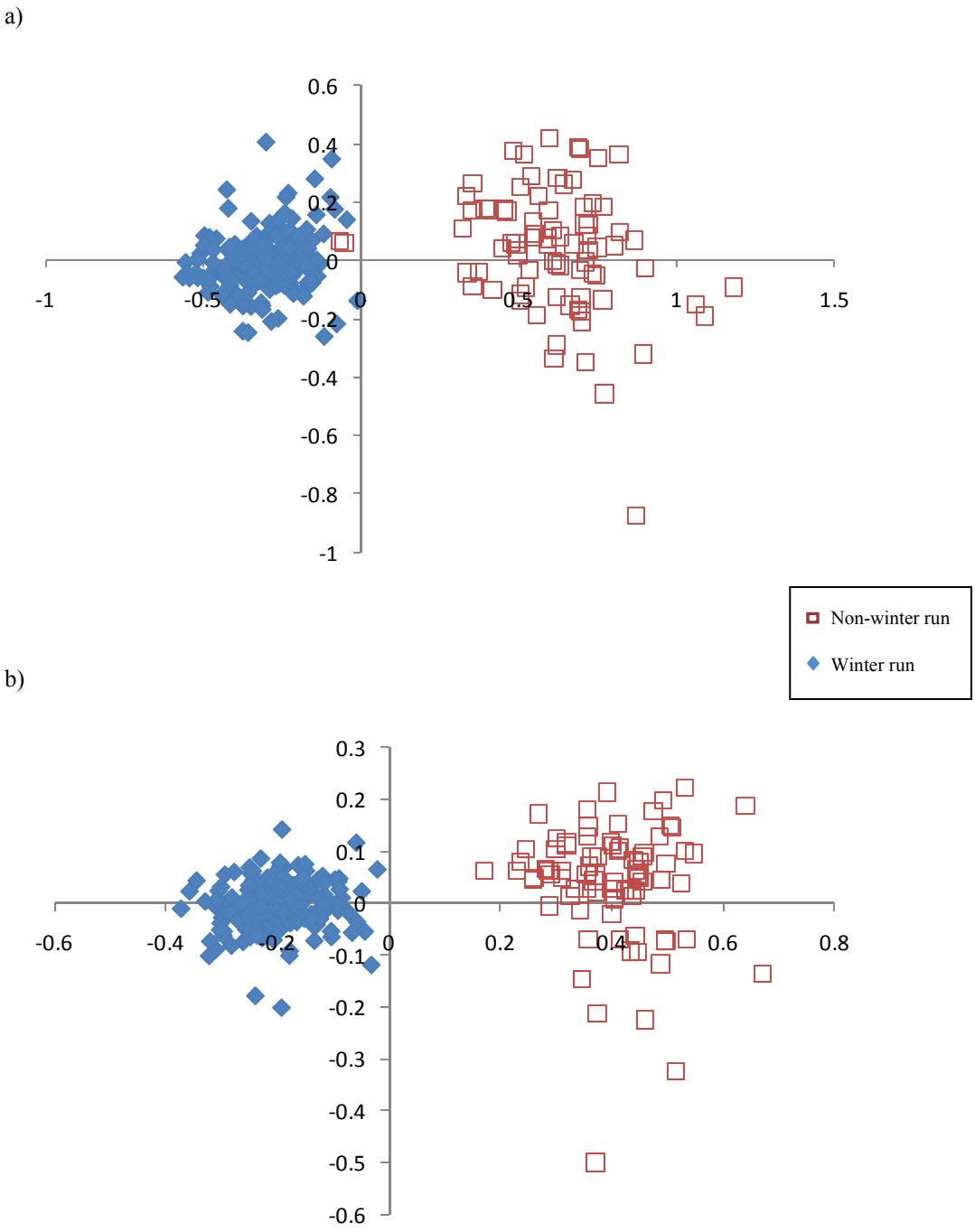
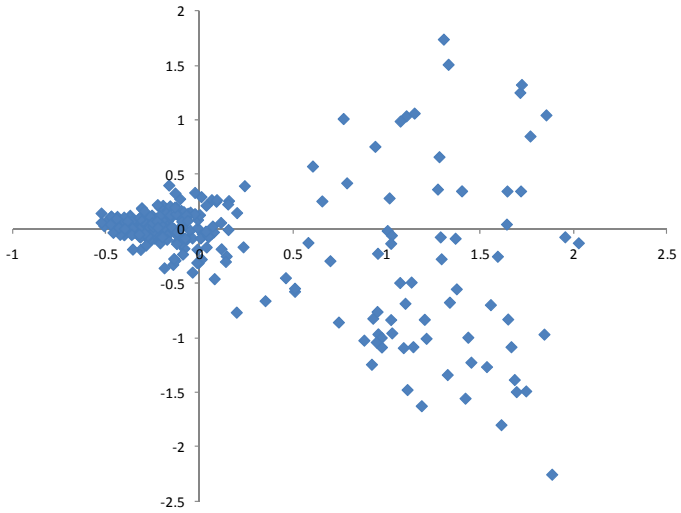


Figure 5. Correspondence analysis of simulated mixtures (see methods) based on 23 SNPs (a) and 94 SNPs (b). The shape and color of each point indicate genetic assignment using the established rapid response protocol (substituting the SNP markers and baseline).

a)



b)

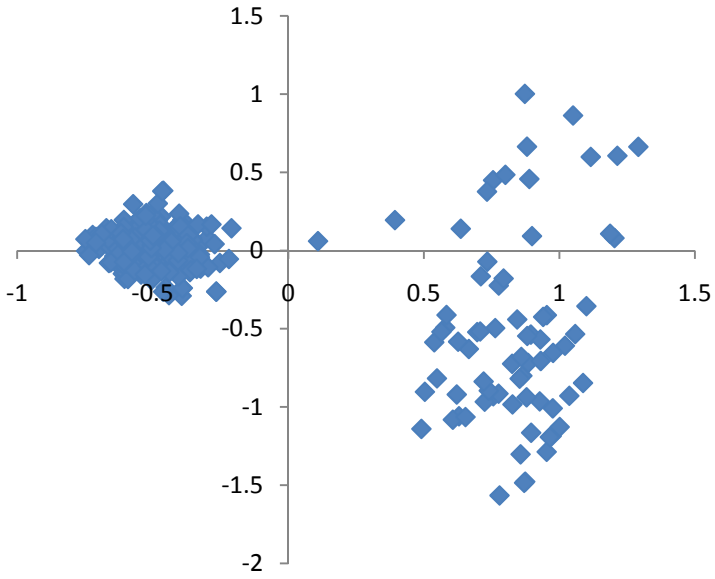


Figure 6. Correspondence analysis of individual fish for which both rapid response (a) and GAPS (b) genotypes were available. The degree of separation of the points into two clusters indicates the relative power of the marker sets to distinguish between winter run and non-winter run.

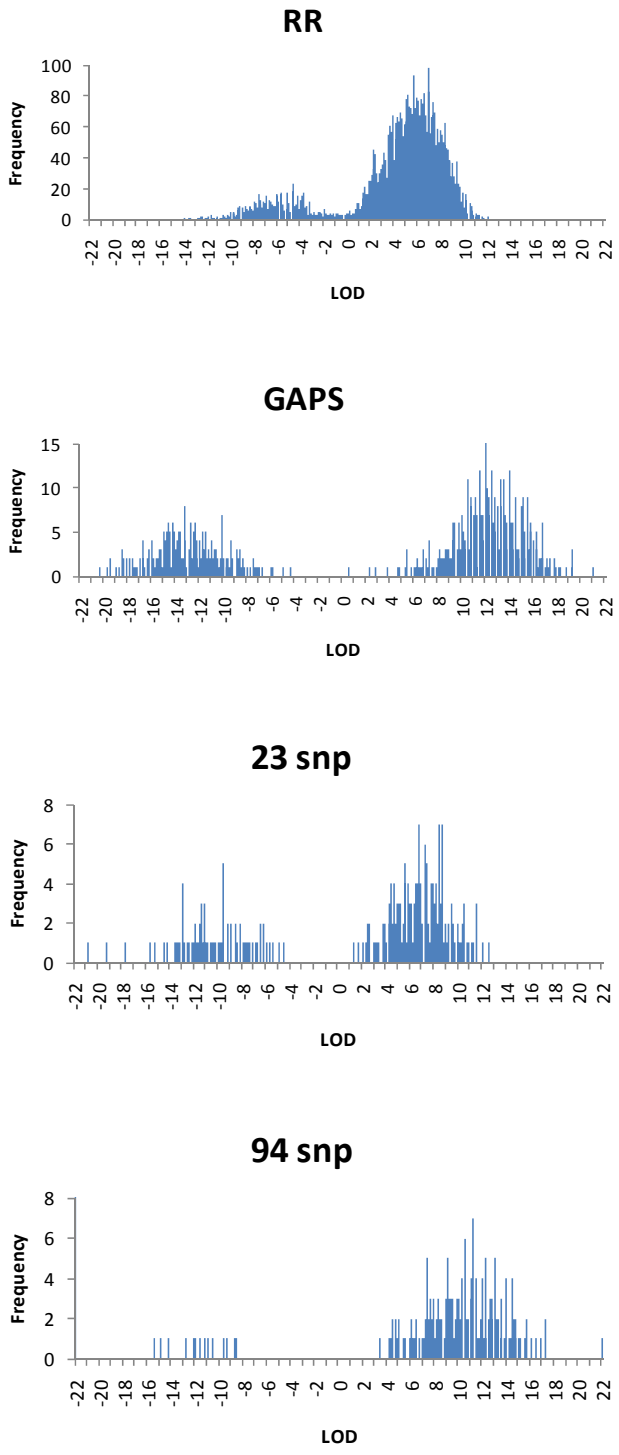


Figure 7. Distributions of LOD ratios resulting from analysis of 4 marker sets following the established rapid response protocol. Results for GAPS and rapid response (RR) were based on actual mixtures, whereas results for both SNP sets were based on simulated mixtures.

Appendix 1. Marker names and numbers of alleles in three sets of markers used to analyze Chinook salmon from California's Central Valley.

Baseline	Marker name	Number of alleles	Baseline	Marker name	Number of alleles
SNP			Rapid Response		
	Ots_94857-232	2		Ots2	18
	Ots_102213-210	2		Ots3	10
	Ots_104569-86	2		Ots9	3
	Ots_107285-93	2		Ots10	5
	Ots_110495-380	2		One13	12
	Ots_112419-131	2		Ots104	25
	Ots_118175-479	2		Ots107	30
	Ots_128302-57	2		Total	103
	Ots_131906-141	2			
	Ots_AsnRs-60	2	GAPS		
	Ots_mybp-85	2		Ogo2	15
	Ots_TAPBP	2		Ogo4	14
	Ots_96222-525	2		Oki100	36
	Ots_102414-395	2		Omm1080	44
	Ots_105105-613	2		Ots201b	31
	Ots_107806-821	2		Ots208b	45
	Ots_110551-64	2		Ots211	26
	Ots_112820-284	2		Ots212	29
	Ots_118205-61	2		Ots213	32
	Ots_128693-461	2		Ots3M	11
	AldB1-122	2		Ots9	4
	Ots_aspat-196	2		OtsG474	14
	Ots_myoD-364	2		Ssa408	19
	Ots_u07-07.161	2		Total	320
	Ots_96500-180	2			
	Ots_102420-494	2			
	Ots_105132-200	2			
	Ots_108007-208	2			
	Ots_110689-218	2			
	Ots_112876-371	2			
	Ots_118938-325	2			
	Ots_128757-61	2			
	AldoB4-183	2			
	Ots_CD59-2	2			
	Ots_Ots311-101x	2			
	Ots_u07-49.290	2			
	Ots_97077-179	2			

Baseline	Marker name	Number of alleles
SNP		
	Ots_102457-132	2
	Ots_105401-325	2
	Ots_108390-329	2
	Ots_113242-216	2
	Ots_122414-56	2
	Ots_129144-472	2
	Myc-366	2
	Ots_CD63	2
	Ots_PGK-54	2
	Ots_u4-92	2
	Ots_99550-204	2
	Ots_102801-308	2
	Ots_105407-117	2
	Ots_108735-302	2
	Ots_111666-408	2
	Ots_113457-40	2
	Ots_123048-521	2
	Ots_129170-683	2
	OTALDBINT1-SNP1	2
	Ots_EP-529	2
	Ots_Prl2	2
	OTSBMP-2-SNP1	2
	Ots_100884-287	2
	Ots_106499-70	2
	Ots_109693-392	2
	Ots_111681-657	2
	Ots_117043-255	2
	Ots_123921-111	2
	Ots_129458-451	2
	OTNAML12_1-SNP1	2
	Ots_GDH-81x	2
	Ots_RFC2-558	2
	OTSTF1-SNP1	2
	Ots_101119-381	2
	Ots_103041-52	2
	Ots_106747-239	2
	Ots_110064-383	2
	Ots_112208-722	2
	Ots_117242-136	2

Baseline	Marker name	Number of alleles
SNP		
	Ots_124774-477	2
	Ots_130720-99	2
	Ots_ARNT-195	2
	Ots_HSP90B-385	2
	Ots_SClkF2R2-135	2
	S71-336	2
	Ots_101704-143	2
	Ots_104063-132	2
	Ots_107074-284	2
	Ots_110201-363	2
	Ots_112301-43	2
	Ots_117432-409	2
	Ots_127236-62	2
	Ots_131460-584	2
	Ots_RAG3	2
	Ots_MHC1	2
	Ots_SWS1op-182	2
	unk_526	2
	Total	188