Western Alaska Salmon Stock Identification Program **Technical Document**:[1] 5

Title: Status of the SNP baseline for sockeye salmon **Version:** 1.0
Authors: T. Dann, A. Barclay, C. Habicht
Date: September 14, 2009

Introduction

The single nucleotide polymorphism (SNP) baseline for sockeye salmon that will be used for mixed stock analysis (MSA) to estimate stock contributions of catches sampled under the Western Alaska Salmon Stock Identification Program (WASSIP) is in a state of perpetual improvement. The collections that make up this baseline were collected over the past twenty years and were funded by many sources including the State of Alaska through general funds and disaster funds, the North Pacific Research Board, National Park Service, Federal Office of Subsistence Management, Pacific Salmon Commission, and the Exxon Valdez Oil Spill Trustee Council.

The suite of SNP markers screened for the baseline has also changed through time and will continue to grow or change as more markers become available. We currently screen for 42 nuclear and three mitochondrial markers, but the WASSIP Advisory Panel has requested that 96 SNP markers be incorporated into the baseline to improve the precision and accuracy of stock composition estimates. To meet this request, we are contracting the development of at least 50 SNP markers that are targeted to differentiate among sockeye salmon populations spawning within western Alaska and the Alaska Peninsula drainages (Technical Document 6). These new SNP markers will be assessed after screening a fraction of the baseline and the best-performing SNP markers will be added to the baseline during the winter of 2009/2010.

[1] This document serves as a record of communication between the Alaska Department of Fish and Game Commercial Fisheries Division and the Western Alaska Salmon Stock Identification Program Technical Committee. As such, these documents serve diverse ad hoc information purposes and may contain basic, uninterpreted data. The contents of this document have not been subjected to review and should not be cited or distributed without the permission of the authors or the Commercial Fisheries Division.

Here we present the current state of the baseline based on samples collected through the 2008 collection season and genotyped for the currently available 42 nuclear and three mitochondrial SNP markers.

Methods

Tissue Sampling

Baseline samples for SNP analyses were collected from spawning populations or obtained from existing agency archives from throughout the range of sockeye salmon in the Pacific Rim (Table 1). We used published genetic structure information (Beacham et al. 2006) to determine appropriate areas to sample outside the Bering Sea drainages. Target sample size for baseline collections was 95 individuals across all years for each population to achieve acceptable precision for the allele frequency estimates (Allendorf and Phelps 1981; Waples 1990a) and to accommodate our genotyping platform.

Laboratory Analysis

Assaying genotypes

Genomic DNA was extracted using a DNeasy® 96 Tissue Kit by QIAGEN® (Valencia, CA). Forty-five sockeye SNP markers were assayed (Table 2), three mitochondrial DNA (mtDNA) and 42 nuclear DNA (nDNA), using 5’ nuclease methods described in Seeb et al. (2009). Thirty-six assays originated from Smith et al. (2005) and Elfstrom et al. (2006). Nine new markers were developed using the methods of Smith et al. (2005) or Elfstrom et al. (2006) and sequencing fifty individuals, ten individuals collected at each of five geographic locations (Russia, Bristol Bay, Kodiak Island, Southcentral Alaska, and Southeast Alaska; Habicht et al. submitted). Individuals were sequenced in both directions, and sequences were aligned and screened for SNPs using Sequencher 4.5 software (Gene Codes Corporation).

Baseline population samples were genotyped using uniplex SNP genotyping performed in 384-well reaction plates and also by using the 48.48 array (Fluidigm Corporation) where 43 of the 45 markers were assayed in sets of 48 fish and *One_MHC2_190* and *One_STC-410* were assayed on the 384-well platform. With either platform, genotypes from generally 384 fish were visualized using the GeneMapper (uniplex platform; Applied Biosystems) and BioMark (array platform; Fluidigm Corporation) software programs and scored for each marker by two people simultaneously. Scores were entered and archived in the Gene Conservation Laboratory Oracle database, LOKI.

Quality control

Three measures were taken to ensure quality control of the baseline data:

1. Re-genotyping of samples – Eight percent of each collection was re-genotyped for all markers to ensure that genotypes were reproducible, to identify laboratory errors, and to measure rates of inconsistencies during repeated analyses on the uniplex and array platforms. We report here error rates for a representative baseline project which consisted of 87 baseline collections comprising 7,593 individuals (~ 15% of current baseline).

2. Exclusion of individuals with an excessive rate of drop-outs – A threshold of 80% scorable markers per individual was established and all individuals that did not meet this threshold were excluded from statistical analysis and use in the baseline. This threshold was set to exclude individuals with poor quality DNA. Poor quality DNA leads to lower reproducibility and therefore adds error to the allele frequency estimates. The value of 80% was chosen based upon the observation that many individuals with high quality DNA had some dropouts, but generally less than 20% of markers, while those with poor-quality DNA had higher drop-out rates. As a result, there was little difference in which individuals were excluded from analysis when picking the threshold as long as it was within the 70% to 90% range.

This rule (referred to as the "80% rule") will also be used for samples from fishery harvests to decrease errors and estimate variances caused by poor quality DNA and missing data. This approach is an attempt to balance the benefits from better data with the loss of power to accurately and precisely estimate stock proportions due to smaller sample sizes. One other potential disadvantage of this approach is the potential to introduce another form of bias if fish that are removed from analyses are not randomly distributed in the mixture. Heterogeneity in sample removal may introduce bias in subsequent estimates of stock proportions when samples with quality genotypic data are not representative of the entire harvest being sampled. We anticipate that bias will only be a concern if significant proportions of mixtures are excluded.

3. Exclusion of duplicate individuals – Finally, we searched for suspected duplicate fish within collections by identifying pairs of individuals that had identical multi-marker genotypes at 38 or more markers. If suspected duplicates were found, the second individual in each matching pair was removed from further analyses.

Statistical analysis

Heterozygosity and F_{ST}

Genotypic data were retrieved from LOKI and were used to calculate allele frequencies. Observed heterozygosity, expected heterozygosity, and F_{ST} (Weir and Cockerham 1984) were calculated for all markers using the program GDA (Lewis and Zaykin 2001).

Linkage disequilibrium

All pairs of nuclear markers were tested for gametic disequilibrium within each collection using GENEPOP (version 4.0; updated version of Raymond and Rousset 1995; Rousset 2008). We defined a pair of markers to be significantly out of gametic equilibrium if tests for gametic disequilibrium were significant ($P < 0.01$) for greater than half of all collections.

When gametic linkage was significant, we produced composite genotypes by ordering the alleles within each marker alphabetically and then stringing the alleles together by marker ordered alphanumerically. Markers that did not exhibit gametic disequilibrium with any other markers and markers that were combined were defined as loci for the remaining analyses. All mtDNA markers were combined into a single locus.

Pooling collections into populations

Collections taken at the same location at similar calendar days in different years were pooled as suggested by Waples et al. (1990). Technical Document 2 has a more detailed investigation of temporal variation among collections taken in different years at the same site and calendar time. Samples taken at the same location, but at substantially different calendar days, and samples taken from geographically proximate locations were tested for homogeneity using a chi-square test of allele frequency distributions across all loci. Groups of collections that demonstrated homogeneity (P > 0.01, not corrected for multiple tests) were pooled. The pooled and the remaining unpooled collections were defined as populations in further analyses. Our protocol was to drop populations from further analyses if they were represented by sample sizes of less than 80 fish.

Hardy-Weinberg equilibrium

Genotype distributions within collections were tested for deviation from Hardy-Weinberg expectation (H-W) using GENEPOP (version 4.0). These tests were repeated once collections were pooled into populations. For H-W, critical values ($\alpha = 0.05$) were adjusted for multiple tests within markers among collections and multiple tests across markers within collections (Rice 1989). The corrections for multiple tests resulted in low power to detect significant departures from H-W, so we also examined the number of departures from H-W by marker and by population prior to correcting for multiple tests to assess any patterns in departures from H-W.

Identifying markers under selection

LOSITAN (Antao et al. 2008), an implementation of the FDIST2 package of Beaumont and Nichols (1996), was used to identify markers that produce F_{ST} outliers. Markers with high outlier F_{ST} values are thought to be under dispersive selection. Due to limitations on the size of dataset used in this program and the geography of the application, we restricted this analysis to populations from the northern Alaska Peninsula, Bristol Bay, and the Kuskokwim River for the 42 nuclear markers. We chose running parameters based upon the following:

1. We chose to not use the "neutral" mean F_{ST} setting. This setting estimates a neutral F_{ST} from only markers that an initial run of LOSITAN reveals to not be under selection. A second and final run is computed incorporating all markers (giving each an estimated selection status) using the mean neutral F_{ST} obtained from the first run described above (Antao et al. 2008 page 3 # 6). We chose not to use this setting as this simulation analysis suggests that a majority of markers are candidates for balancing selection, more than we believe, and removing this many markers from the estimation of the mean F_{ST} results in a spuriously high mean F_{ST} estimate. However, we ran the analysis both using and not using the 'neutral' setting and found that results do not differ much (e.g., the same markers were identified as candidates for positive selection);
2. We chose to use the force mean F_{ST} setting because it approximates the desired average simulated F_{ST} to the average value observed in the dataset using a bisection algorithm (Antao et al. 2008 page 3 # 7);
3. We changed the sample size to more accurately represent the number of individuals we observed in most of the "islands" in our baseline (n=95);
4. We removed six populations from the Lake Clark and Upper Kuskokwim regions from the analysis because simulations based upon the island model may not be appropriate for a baseline with these populations included. There is evidence that Lake Clark sockeye salmon populations were recently founded and show signs of a bottleneck effect (Habicht et al. 2004), and there are probably high levels of isolation-by-distance for both of these groups of populations. We chose to remove

these specific populations as they were the most divergent on a Neighbor-Joining tree of pair-wise F_{ST}'s (data not shown);

5. We changed the expected number of populations to equal what we included in the simulations (i.e. 90 instead of 96);
6. We removed five markers from the analysis as they exhibit very low levels of heterozygosity. Beaumont and Nichols recommend discarding markers with heterozygosities less than 2/ (sample size), so we used 0.02 as our cut off for removal, which included: *One_ctgf-301*, *One_MARCKS-241*, *One_p53-534*, *One_RAG1-103*, and *One_RH2op-395*.

Population structure visualization

To visualize genetic population structure, Cavalli-Sforza and Edwards (1967) chord distances (CSE) were calculated from allele frequencies at the 42 SNP loci and plotted using the UPGMA method. We chose this measure of genetic distance because previous analyses have identified loci under positive selection and utilizing distance measures that assume neutral loci and are based upon genetic drift (i.e., pair-wise F_{ST}'s) may not be appropriate. While this measure is biased by unequal sample sizes, a substantial portion of the populations included in this baseline are of 95 individuals. CSE distances were used to produce two UPGMA trees: 1) all baseline populations and 2) restricted to populations from Western Alaska and the Alaska Peninsula (WAAP).

Hierarchical log-likelihood analysis

We examined the homogeneity of allele frequencies among populations within regions using a hierarchical log-likelihood ratio test (*G* test; Sokal and Rohlf, 1995). We included data from only nuclear loci and removed *One_MHC2_251* so as not to duplicate the divergence information provided by the two linked MHC loci. We examined *G*-statistics for each of 17 coastwide regions (Table 1), and summed *G*-statistics and degrees of freedom from 12 of these regions into three broad-scale regions (i.e., Western Bristol Bay YK, Eastern Bristol Bay, and Alaska Peninsula) for an examination of broad-scale population structure. These two levels of analysis correspond to the regional groupings used in the two UPGMA trees described above. We further summed test statistics across regions into Western Alaska (Norton Sound to South Alaska Peninsula) and Coastwide totals. Finally, we summed test statistics across loci for an overall measure of allele frequency homogeneity at the same hierarchical levels described above. As the number of populations within regions differed greatly (i.e., 3 populations in the Norton Sound region, 116 populations in the Western Gulf of Alaska region), we divided *G*-statistics by degrees of freedom to examine a measure of regional diversity less biased by sampling effort.

Baseline evaluation for MSA

Reporting groups were delineated based on geographic regions that were thought to be applicable for MSA analyses of mixtures captured under the WASSIP program. Within Norton Sound, Yukon and Kuskokwim Rivers, Bristol Bay and Alaska Peninsula, the reporting groups represent smaller geographic areas on the scale of commercial fishing districts. Outside of these areas, the reporting groups represent much larger geographic areas on the order of management regions or countries. During estimation of stock composition, populations were maintained separately within these reporting groups as recommended by Wood et al. (1987). Reporting group estimates were calculated by summing population estimates.

We then assessed the potential of the baseline to identify these reporting groups for MSA applications with simulations and proof tests. For the simulations, we generated 400 fish based on the population-specific allele frequencies from all the populations within each reporting group (i.e., 100% simulations). This process was repeated 1,000 times, and the mean and central 90% of the distribution of estimates were reported as the estimate and the 90% confidence interval. Simulated mixtures were analyzed using SPAM version 3.7b (Debevec et al. 2000; ADF&G 2001). For the proof tests, we created a test mixture by sampling approximately 200 fish from each reporting group; we rebuilt the baseline excluding the sampled fish. The test mixture was analyzed using BAYES (Pella and Masuda 2001) with a flat prior (with a weight of one fish). Estimates and 90% credibility intervals from three chains with different starting conditions were tabulated. We repeated this procedure for each reporting group. For both the simulations and proof tests, a critical level of 90% correct allocation was used to determine if the reporting group was acceptably identifiable (e.g., Seeb et al. 2000).

Results

Tissue Sampling

A total of 49,809 individuals from 562 collections representing 375 populations (Table 1; Figure 1) have been genotyped at the 45 SNP markers (Table 2). This baseline represents an increase of 120 populations to the 255 population baseline presented by the ADF&G Gene Conservation Laboratory (GCL) in its proposal to AYK SSI for WASSIP funding in 2007. Collection sites ranged from the western Kamchatka Peninsula (Russia) to Puget Sound, Washington. The most comprehensive collection was done in the densest portion of the species range, i.e., populations from rivers draining into the Bering Sea and areas adjacent to the Bering Sea (Figure 1). For some analyses we included a subset of collections from the Western Alaska/Alaska Peninsula region (WAAP). This subset was comprised of 20,856 individuals from 221 collections representing 137 populations ranging from the Norton Sound region in the north to the South Peninsula region to the south (Table 1; Figure 2).

Laboratory Analysis

The overall failure rate for successfully assaying genotypes at the 45 SNP markers in a representative project was 2.3%. The quality control process demonstrated a discrepancy rate of 0.58%. Assuming an equal error rate in the original and quality control genotyping process, our baseline collections were genotyped with a process that produced genotypes with an error rate of 0.29%. An average of 1.4 fish per collection was removed based upon the 80% rule for the collections that were included in this baseline (SD = 3.3). A majority of collections had no fish removed based upon the 80% rule (i.e., 317), and 102 collections had one fish removed while 12 collections each had greater than 10 fish removed.

Statistical Analysis

Heterozygosity and F_{ST}

Observed heterozygosity, expected heterozygosity, and F_{ST} for each of the nuclear markers, and only F_{ST} for each of the combined loci (see linkage disequilibrium results) are in shown in Table 3. Observed heterozygosity was lower than expected heterozygosity at every nuclear marker with the averages of 0.243 and 0.288, respectively. Observed heterozygosities ranged widely from 0.017 to 0.447.

The F_{ST} estimate over all markers was 0.149, but a few nuclear markers had considerably higher values. F_{ST} estimates for *One_MHC2_251* and *One_MHC2_190* were 0.303 and 0.356, respectively. Other markers with F_{ST} estimates greater than 0.2 included: *One_Tf_ex10-750*, *One_HpaI-99*, *One_STC-410*, *One_zP3b-49*, *One_Tf_ex3-182*, and *One_GHII-2465*. The remaining markers had F_{ST} values below 0.170 and only three markers had values below 0.050.

Linkage disequilibrium

Significant gametic disequilibrium was found between one pair of nuclear SNP markers (*One_MHC2_190* and *One_MHC2_251*; Table 4). Other pairs of markers that exhibited linkage disequilibrium within some collections, but below the threshold of 50% of the populations were: *One_GPDH* and *One_GPDH2* (34% of collections); *One_Tf_ex10-750* and *One_Tf_ex3-182* (19%); and *One_RF-112* and *One_RF-295* (7%). All of these pairs are known to be physically linked.

For the pair of linked nuclear SNP markers and the triplet of mitochondrial SNP markers *(One_CO1, One_Cytb_17,* and *One_Cytb_26)*, genotypes from each marker were pooled to form one haplotype locus*: One_MHC2_190_251* and *One_CO1_Cytb17_26*, respectively. After combining the pair of linked nuclear markers and the three mtDNA markers, the final analyses included 41 independent nuclear loci and 1 mitochondrial locus (described by three SNPs).

Pooling collections into populations

The 562 collections reduced to a total of 375 unique populations after pooling collections taken from similar locations over multiple years and from nearby sites that exhibited genetic homogeneity. Some tests for homogeneity between collections within the WAAP area were significant based upon our criterion. Of these, we pooled the following populations with temporal collections based upon the recommendations of Waples (1990): Goodnews River North Fork, Goodnews River Middle Fork, Tommy Creek, Upper Talarik Creek, and Idavain Creek. These represent 18% of the 28 pairs of collections taken from similar locations over multiple years within the WAAP area. The test for homogeneity between the two collections from the West Fork of the Black River (Chignik drainage) was also significant, but we have little metadata associated with the 1997 collection and so did not pool these collections for this baseline analysis. Technical Document 2 provides a more detailed investigation of this temporal diversity.

The average sample size per population was 133 fish, although a few populations outside the Western Alaska/Alaska Peninsula (WAAP) area were small with as few as 10 fish. Within the WAAP, the smallest population sample size was 47 fish. These populations with sample sizes below 80 fish were mistakenly included in subsequent analyses and are indicated by an asterisk in the population column of Table 1; they will be excluded in the final baseline. A substantial portion of the populations included in this baseline are of 95 individuals (i.e., 115), and 175 populations have a sample size greater than 95 individuals.

Hardy-Weinberg equilibrium

Significant departures from H-W were not found in any populations for the 42 nuclear SNP markers after correcting for multiple tests. However, before correcting for multiple tests, we did find some patterns in the distribution of departures from H-W. *One_MHC2_190* and *One_MHC2_251* were out of H-W in 29 and 30 populations, respectively, while no other marker was out of H-W equilibrium at more than 23 populations (Table 2; Figure 3). Nineteen populations were expected to be out of H-W equilibrium for each marker by chance at $\alpha = 0.05$.

We also detected eight populations with greater than twice as many markers out of H-W equilibrium than would be expected by chance (before correcting for multiple tests; Table 1; Figure 4). Two markers were expected to be out of H-W equilibrium for each population by chance at $\alpha = 0.05$. These included Avacha Bay, Dvu 'Yurta River, and Belaia River in Russia, the middle fork of the Goodnews River in western Alaska, Fish Creek and English Bay in Cook Inlet, Mill Creek in southeast Alaska, and Baker Lake in Washington. In all but one of the 61 cases, the significant departure from H-W at markers for these populations was due to an excess of homozygotes (i.e., positive F_{IS} values).

Identifying markers under selection

The results of the LOSITAN analysis clearly suggest that the two major histocompatibility complex markers (*One_MHC2_190* and *One_MHC2_251*; MHC) are very different from other markers and that statistically they are candidates for positive selection using these simulation parameters (Figure 5). LOSITAN also suggests *One_STC-410* and *One_ZNF-61* as candidates for positive selection, although the F_{ST} estimate for *One_ZNF-61* is not much greater than the upper bound of the mean F_{ST} estimate. We would expect 37 (total markers analyzed) minus 2 (MHC markers) = 35 X 0.05 (alpha) = 2 markers to be outside the bounds by chance, so excluding candidates for balancing selection, having two markers above the upper bounds is not unreasonable.

The LOSITAN output shows a lower bound that defines many markers as candidates for balancing selection. After removal of the two MHC markers, the F_{ST} mean and confidence interval bounds decreased and nine fewer markers are considered candidates for balancing selection. This also then includes two more markers (*One_STR07* and *One_Prl2*) as candidates for positive selection, but these were just above the upper bound (data not shown).

Population structure visualization

Genetic relationships among baseline populations are shown schematically in the UPGMA trees (Figures 6 and 7). On the tree with the whole Pacific Rim baseline, the deepest structure was found within the Eastern and Western Gulf of Alaska (Figure 6). A regional structuring of populations was the most common pattern with populations clustered by lakes and drainages. These patterns can most easily be visualized in the WAAP UPGMA (Figure 7), where most of the populations within some of the drainages or nursery lakes cluster together including the Naknek River, Alagnak River, and Chignik River.

Population relationships within some drainages are more complicated than others, which may be the result of a more complicated geography and other factors. The populations within the Wood River, which is made up of five large lakes, beach and tributary spawners, and early- and late-run timing, divide into four clusters. The populations within the Nushagak River, which is a long river with one branch that drains large lakes and other branches that are devoid of lakes, are divided into two clusters and an outlying population. The populations within the Kvichak River, which is made up of one large lake and one smaller lake, are in three clusters with one outlying population. These clusters are made of populations from Lake Clark (highly divergent), northeastern and southwestern Iliamna Lake, and a population spawning between the two lakes. Many of the populations within the North and South Peninsula, which contain many short rivers that drain directly into the ocean, are highly divergent from each other and may reflect the stronger influence of genetic drift on these smaller populations. The populations within the Egegik River cluster into one group representing tributary spawners from the eastern and north side of the nursery lake and a divergent population representing the south side of the nursery lake.

Finally, the Kuskokwim River and Norton Sound contained some of the most divergent collections. These included the Necons River and Telaquana Lake from the Kuskokwim River and Salmon and Glacial lakes that drain into Norton Sound. These Kuskokwim populations and the highly divergent Lake Clark populations were the populations removed from the LOSITAN analysis for markers under selection and are the most divergent in the WAAP area (top nodes; Figure 7).

Hierarchical log-likelihood analysis

Substantial heterogeneity in allele frequencies existed among populations within all fine- and broad-scale regions (Table 5). Each test for homogeneity of allele frequencies among populations within regions was highly significant (P < 0.01). The measure of regional diversity corrected for number of populations (i.e., G / df) highlights substantial diversity within particular regions, notably Norton Sound, Yukon Kuskokwim and Kvichak in the WAAP area (G / df = 17.27, 18.74, and 21.74, respectively; Figure 8), and Western Gulf of Alaska and Eastern Gulf of Alaska in the coastwide analysis (G / df = 37.74, and 26.16, respectively; Figure 9). Also notable is the relatively low within-region diversity for the WAAP area, especially within the Igushik, Wood, Naknek and Ugashik regions.

Different markers exhibit varying degrees of allele frequency divergence across regions. The *One_MHC2_251* marker is the most powerful included in this analysis at describing differences among populations for both the coastwide and WAAP regional scales, and exhibits similar discriminatory power in both regional areas (i.e., G / df = 82.14 and 78.88, respectively). Other markers are very useful at describing coastwide genetic diversity but not as useful within the WAAP study area (e.g., *One_E2* G / df = 25.79 and 9.79, respectively; Figure 10). Similarly, some markers show no differences among populations within some regions (e.g., *One_p53-576* G / df = 0.00 for Western Kamchatka through Yukon Kuskokwim, data not shown), but very high levels of diversity among populations for other regions (*One_p53-576* G / df= 26.36 for Western Gulf of Alaska).

Baseline evaluation for MSA

Three reporting groups failed to meet the critical level of 90% correct allocation in the 100% simulations (Igushik, Ugashik, and North Peninsula; 86%, 86% and 89%, respectively; Figure 11; Table 6). When fish were misallocated in the Igushik simulations, 10% were allocated to the

Wood River reporting group and 2% to the Nushagak reporting group. When fish were misallocated in the Ugashik simulations, 4% were allocated to the Egegik reporting group, 3% to the North Peninsula reporting group, and 2% to the Western Gulf of Alaska reporting group. When fish were misallocated in the North Peninsula simulations, 4% were allocated to the Western Gulf of Alaska reporting group and 2% to the South Peninsula reporting group. In general, the simulations indicated that most reporting groups can be distinguished from one another with a high degree of accuracy (mean = 93%).

Proof tests using the current baseline indicate that the 17 coastwide reporting groups can be distinguished from each other with a high degree of accuracy (mean = 97%; Figure 12; Table 7). Only one of the reporting groups (Western Gulf of Alaska; 89%) did not meet the critical level of 90% correct allocation. When fish were misallocated in the Western Gulf of Alaska proof test, 9% were allocated to the Eastern Gulf of Alaska reporting group.

Discussion

This sockeye salmon baseline is the most comprehensive SNP database available for any Pacific salmonid. It is also the most comprehensive genetic baseline of any marker type that includes high representation from all areas that are most likely to contribute to mixtures sampled under the WASSIP, with 127 populations from the WAAP areas. The WAAP is also the area where the majority of sockeye salmon are produced. Almost 50% of all of the sockeye salmon production in the world originate from Bristol Bay drainages alone (Eggers and Irvine 2007; Bugaev et al. 2008). The baseline is least complete for the US/Canada trans-boundary rivers that drain into Southeast Alaska and spawning areas in British Columbia. Major ancestral lineages from those regions that were identified in Beacham et al. (2006) are represented by one or more collections. Thus, despite some gaps in the baseline in this area, adequate samples exist so that fish originating from Eastern Gulf of Alaska populations not included in the baseline will most likely allocate to the large-scale Eastern Gulf of Alaska reporting group.

Population structure for sockeye salmon spanning the Pacific Rim was first described by Beacham et al. (2006). The baseline data for these studies are least complete in the densest portion of the species range. Such a baseline bias may impact MSA allocations. Their data, for example, indicated that 7% of a test sample of 62 fish from the western Bering Sea originated from the Alaska Peninsula and none originated from Bristol Bay. Data presented by Habicht et al. (submitted) suggest that Bristol Bay is the dominant regional stock of North American sockeye salmon migrating through the western Bering Sea, and Alaska Peninsula stocks are rarely present. This observation is supported by that of Bugaev et al. (2008), who used scale pattern analysis to report a dominant role for Bristol Bay stocks (55% of immature sockeye salmon) in summer 2006 BASIS surveys in the REEZ. Nevertheless, Beacham et al. (2006) provide a framework for future studies. The patterns of genetic relationships identified in this study are similar to those reported in Beacham et al. (2006) and provide a template to insure that samples used in this study adequately represent the major lineages of sockeye salmon at the extremes of the species range.

Marker F_{ST} and resolving power

Beacham et al. (2001) point out that the MHC markers provide a significant portion of the resolving power of the MHC/microsatellite data bases; merging of the MHC portions of the two data sets needs further evaluation given the different analysis methods between the studies. The two MHC markers in our study had the highest F_{ST} values among all the markers (Table 3) and the one MHC included in the log-likelihood ratio test had the highest *G* statistics in both the overall and the WAAP baseline (Figure 10), indicative of the resolving power of this locus for GSI. Among the other markers with high F_{ST} values, six others were above 0.2 and included: *One_Tf_ex10-750* (0.206); *One_HpaI-99*(0.218); *One_STC-410* (0.220); *One_zP3b-49* (0.266); *One_Tf_ex3-182* (0.268); and *One_GHII-2465* (0.275). Not surprisingly, these six were also identified in the log likelihood ratio test analysis as the only loci with degree-of-freedom-adjusted *G* statistics higher than 30 for the full baseline (Figure 10).

The log-likelihood ratio test analysis also showed that the loci with the highest *G* statistics for the full baseline were not identical to those for the WAAP area. For the WAAP area, the *G* statistics were generally lower with only five loci showing degree-of-freedom-adjusted *G* statistics above 20. Of these, four of the markers were identified as powerful for discriminating among populations within regions for the full baseline (the MHC marker, *One_Tf_ex10-750*; *One_HpaI-99; and One_zP3b-49*), while *One_ALDOB-135* was relatively powerful within the WAAP area but intermediate for the full baseline. *One_STC_410*, *One_TFex3-182* and *One_GHII-2465* had *G* statistics below 20 in the WAAP baseline, but higher than 30 in the full baseline. The log-likelihood ratio test might be a good test to identify the most useful markers by region as additional markers become available.

Markers under selection

Both MHC markers also appeared to be the markers under the strongest positive selection within WAAP (Figure 5). MHC is known to be under selection in salmonids (e.g. Atlantic salmon, Dionne et al. 2007). *One_STC-410* was also identified as a candidate locus under selection (Figure 5). *One_STC-410* is a SNP for the target locus stanniocalcin, which is a calcium- and phosphate-regulating hormone (Wagner 1994). Some loci with high F_{ST} values across the species range were not identified as candidates for positive selection within the WAAP area, but may be under selection outside of this area. These differences in selection and resolving power are indicated as large differences between the measure of within-area diversity (*G* / df) for the coastwide and WAAP areas in Figure 10 (e.g., *One_GHII-2465*). *One_Zp3b-49* is associated with the zona pellucida, an extracellular matrix that surrounds growing oocytes in mammals and fish and plays a role in gamete recognition, and therefore may be under selection (Epifano et al. 1995). *One_Tf_ex10-750* and *One_Tf_ex3-182* code for transferrin, which is an iron-binding protein that plays an important role in iron metabolism and resistance to bacterial infection in a variety of organisms. Positive selection for transferrin was detected in an analysis across salmonids (Ford et. al 1999).

The LOSITAN analysis also suggested a large number of markers as candidates for balancing selection. The expected relationships between H_e and F_{ST} were highly affected by the parameters used and the markers included the program. Given the large number of markers that were identified as candidates for balancing selection, more work needs to be done to determine if they are indeed under balancing selection or if some of the model assumptions have been violated. In that effort we are investigating an analysis of these markers in a Bayesian framework (i.e., BayeScan; Foll and Gaggiotti 2008) that may help better identify candidate markers under selection.

Deviations from H-W

We identified some factors that may explain why some populations were out of H-W equilibrium at more than twice the expected number of markers (5 at *P* = 0.05, not adjusted for multiple tests). Two of the populations that met this criterion were from places where samples taken early and late within calendar years were pooled (English Bay and Mill Creek). When chi-square tests were performed to test for homogeneity among these collections, English Bay had a *P*-value of 0.02 and Mill Creek had a value above 0.05. These *P*-values were above our critical value of 0.01 for pooling collections into populations. One possibility that either the early or late collections were mixtures of two run timings which resulted in the large number of markers out of H-W while producing relatively high *P*-values in the chi-square tests.

Three of the populations out of H-W equilibrium were taken in Russia and we have little metadata to determine which factors may contribute to departures from H-W (Avacha Bay, Dvu 'Yurta River, and Belaia River). The large number of deviant markers for Avacha Bay (12) indicates that this collection may be made up from a combination of populations, separated either temporally or spatially, but we have little information for this collection. The Dvu ’Yurta and Belaia river populations are each combinations of two collections taken in consecutive years. Again we do not have calendar day for these collections or any other metadata, but the *P*-values for the chi-square tests were below 0.01 for both of these tests, indicating that the collections differed between the two years. In future baseline analyses we may want to exclude the 1995 collections because they contain only 11 fish each.

The Middle Fork Goodnews River population was made up of three collections (1991, 2001, and 2007) and the chi-square test was highly significant (*P* < 0.01). The 2007 collection was made throughout June and July, while the other collections were made in mid July and early August indicating that there may be multiple populations in these samples that are temporally segregated.

The two Fish Creek collections were taken at similar calendar dates 16 years apart and had a highly significant chi-square test result ($P < 0.01$). These collections are of fish captured at the Fish Creek weir and may be a mixture of populations that segregate spatially within the Fish Creek drainage. These collections could not be pooled with the Fish Creek samples taken at the Big Lake Hatchery, which is in the Fish Creek drainage. This year we collected fish in Meadow Creek, another tributary to Fish Creek, with the hope that this collection can substitute for the weir collection in future baselines.

Finally, the collection from Baker Lake had more than five markers out of H-W equilibrium. We have no metadata from this location, but spatially segregated natural and artificial spawning areas that are used in Baker Lake to mitigate for dams (http://wdfw.wa.gov/fish/sockeye/bakerriver.htm) might be becoming reproductively isolated (i.e. Hendry et al. 2000). All but one of these departures from H-W expectations are the result of an excess of homozygotes, indicative of a Wahlund effect and consistent with observing an admixture of populations.

Population structure

The hierarchical analysis of allele frequency homogeneity highlighted high levels of diversity observed for some regions (e.g., Kvichak, Western Gulf of Alaska and Eastern Gulf of Alaska; Figures 8 and 9), although the range of many of the defined regions was large. These observations are often driven by large differences in allele frequencies observed between large groups of populations or for few outlier populations. Within the Kvichak region, this is the result of a strong divergence between populations within the Lake Clark and Iliamna nursery lakes that has been previously described (Habicht et al. 2004). The Western Gulf of Alaska region encompasses a geographically broad region with high levels of divergence among populations within the region. This divergence is largely driven by the clustering of populations within the Kodiak Archipelago, Kenai, Susitna and Copper rivers (Figure 6). In contrast, the large diversity observed within the Eastern Gulf of Alaska region results from a few highly deviant outlier populations (i.e., Kanalku Lake, Mahoney Creek, Tahltan Lake, Little Tahltan Lake, and Kah

Sheets Lake) with allele frequencies very discordant from two large, loosely clustered groups of the remaining populations. There is relatively little genetic diversity observed within the WAAP study area compared to the Gulf of Alaska regions, which may be the result of a more recent common ancestral population in the Beringia Refugium and many populations with large population sizes that likely retards the influence of genetic drift on genetic divergence.

Aside from some notable exceptions such as Norton Sound, Upper Kuskokwim and Lake Clark, the WAAP study area shows lower levels of genetic differentiation than areas in the Eastern and Western Gulf of Alaska (Figure 9, Table 5).

Baseline evaluation

Simulation and proof test results indicate that the 17 coastwide reporting groups can be distinguished from each other with a reasonable degree of accuracy. The two methods differ in that simulations generate hypothetical individuals from baseline allele frequencies, whereas proof tests remove known individuals from the baseline to be treated as mixture individuals. As such the proof tests provide a more realistic and robust methodology for testing the utility of the baseline at discriminating among reporting groups for GSI purposes. When fish were misallocated they were most often allocated to neighboring reporting groups and/or reporting groups with populations with very similar allele frequencies. For example, Pick Creek in the Wood River reporting group has allele frequencies similar to all of the Igushik populations, groups together with Igushik populations on trees, and can cause misallocation between these two adjacent reporting groups.

There are a number of potential sources of improvement in our baseline evaluation tests. The proof tests, for example, included only 200 individuals yet the WASSIP mixtures will generally be made up of 400 fish. The small sample sizes in the proof tests were necessitated by the small sample size of one reporting group (Norton Sound; 335 fish). The inclusion of additional SNPs will also likely increase resolving power due to an increase in the number of independent markers as well as the potential that some of the new SNPs are under selection and may represent adaptive differences among populations in the WASSIP area. Baseline evaluations that are comprised of more heterogeneous mixture compositions (i.e., not 100%) will provide a measure of baseline utility at discriminating among reporting groups in a more realistic fashion. There are statistical improvements that may improve our GSI resolving power and the results of baseline evaluation tests. Two such examples are the use of informative priors when using Bayesian methods for GSI and the use of a stratified estimate protocol (Technical Document 3).

Future analyses

1. Increase sample sizes for collections for which we have existing tissues to be genotyped.
2. Incorporate collections gathered in the 2009 field collection season into baseline analyses.
3. Remove populations with samples sizes of less than 80 fish (denoted with an asterisk in Table 1) for which we do not have existing tissues to be genotyped from the baseline.
4. Investigate temporal variation in allele frequencies for collections from similar locations in multiple years. Is this variation driven by loci under selection? Does this variation represent problems with our genotyping process? We foresee resampling populations to ensure that the baseline data are still valid and to help address these concerns.
5. Assess the suite of developing SNPs (see Technical Document 6) for utility in describing genetic variation within the WASSIP study area and for accurately and precisely estimating stock proportions in mixture samples from area fisheries.
6. Perform proof tests with 400 fish in reporting groups where adequate numbers of fish exist.
7. Perform simulations and proof tests using more heterogeneous mixture compositions (i.e., not 100%) to assess baseline utility at discriminating among reporting groups in a more realistic fashion.
8. Investigate why we saw a consistent pattern of lower observed heterozygosities than expected (Table 3).
9. Further investigate the utility of the loci identified in LOSITAN as loci under balancing selection. Loci under balancing selection may be good candidates to be replaced with loci under positive selection for MSA as new markers become available.

10. Conduct further analyses of genetic diversity, including AMOVA and Nei's gene diversity analysis, and examine *G* statistics for hierarchical levels within the WAAP area that may have more biologic meaning (e.g., populations within nursery lakes).

11. For these other levels of hierarchy, compare levels of heterogeneity using Fisher's *F*-test to better understand how diversity is distributed in the baseline.

12. Examine the distribution of allelic richness by region and ascertainment region to assess ascertainment bias.

13. Utilize statistical methods developed for estimating small proportions to increase the performance of MSA through decreased bias and increased precision. These methods might include the use of informative priors when using Bayesian methods for GSI and the use of a stratified estimate protocol (Technical Document 3)

14. Investigate the utility of reducing the range of the baseline to include only those populations that are likely to be present in WASSIP mixtures.

Literature Cited

ADF&G (Alaska Department of Fish and Game). 2001. SPAM Version 3.5: Statistics Program for Analyzing Mixtures. Alaska Department of Fish and Game, Commercial Fisheries Division, Gene Conservation Lab. Available for download from http://www.cf.adfg.state.ak.us/geninfo/research/genetics/software/spampage.php.

Allendorf, F. W., and S. R. Phelps. 1981. Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences. 38:1507-1514.Anderson, T. J. C., S. Nair, D. Sudimack, J. T. Williams, M. Mayxay, P. N. Newton, J.-P. Guthmann, F. M. Smithuis, T. T. Hien, I. V. F. van den Broek, N. J. White, and F. Nosten. 2005. Geographical distribution of selected and putatively neutral SNPs in Southeast Asian malaria parasites. Molecular Biology and Evolution 22(12):2362-2374.

Antao, T.,A. Lopes,R. Lopes, B.-P. Albano, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323 1471-2105. Available at: http://www.biomedcentral.com/1471-2105/9/323

Beacham, T. D., J. R. Candy, K. J. Supernault, T. Ming, B. Deagle, A. Schulze, D. Tuck, K. H. Kaukinen, J. R. Irvine, K. M. Miller, and R. E. Withler. 2001. Evaluation and application of microsatellite and major histocompatibility complex variation for stock identification of coho salmon in British Columbia. Transactions of the American Fisheries Society 130(6):1116-1149.

Beacham, T. D., B. McIntosh, C. MacConnachie, K. M. Miller, and R. E. Withler. 2006. Pacific rim population structure of sockeye salmon as determined from microsatellite analysis. Transactions of the American Fisheries Society 135(1):174-187.

Beaumont, M.A. and R.A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London B 263: 1619-1626.

Bugaev, A. V., I. I. Glevov, E. V. Golub, K. W. Myers, J. E. Seeb, and M. Foster 2008. Origin and distribution of sockeye salmon *Oncorhynchus nerka* local stocks in the western Bering Sea in August-October 2006. Izv. TINRO 153:88-108.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Evolution 21:550-570.

Debevec, E. M., R. B. Gates, M. Masuda, J. Pella, J. Reynolds, and L. W. Seeb. 2000. SPAM (version 3.2): Statistics Program for Analyzing Mixtures. Journal of Heredity 91: 509–510.

Dionne, M., K. M. Miller, J. J. Dodson, F. Caron, and L. Bernatchez. 2007. Clinal variation in MHC diversity with temperature: Evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution 61(9):2154-2164.

Eggers, D. M., and J. R. Irvine. 2007. Trends in abundance and biological characteristics for North Pacific sockeye salmon. North Pacific Anadromous Fish Commission Bulletin 4:53-75.

Ford, M. J., P. J. Thornton and L. K. Park. 1999. Natural selection promotes divergence of transferrin among salmonid species. Molecular Ecology 8: 1055–1061.

Elfstrom, C. M., C. T. Smith, and J. E. Seeb. 2006. Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon. Molecular Ecology Notes 6(4):1255-1259.

Epifano, O., L. Liang, and J. Dean. 1995. Mouse Zp1 encodes a zona pellucida protein homologous to egg envelope proteins in mammals and fish. Journal of Biological Chemistry 270(45):27254-27258.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Habicht, C., L. W. Seeb, K. W. Myers, E. Farley, J. E. Seeb. Submitted. Summer-fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single nucleotide polymorphisms (SNPs). Transactions of the American Fisheries Society. XX:XXX-XXX.

Habicht, C., J. B. Olsen, L. Fair, and J. E. Seeb. 2004. Smaller effective population sizes evidenced by loss of microsatellite alleles in tributary-spawning populations of sockeye salmon from the Kvichak River, Alaska drainage. Environmental Biology of Fishes 69(1-4):51-62.

Hendry, A. P., J. K. Wenburg, P. Bentzen, E. C. Volk, T P. Quinn. 2000. Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. Science 290:5491 516 – 518.

Lewis, P.O. and D. Zaykin. 2001. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0. URL http://lewis.eeb.uconn.edu/lewishome/software.html.

Pella, J., and M. Masuda. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. Fishery Bulletin 99(1):151-167.

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

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223-225.

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, J. E., C. E. Pascal, R. Ramakrishnan, and L. W. Seeb. 2009. SNP genotyping by the 5'-nuclease reaction: advances in high throughput genotyping with non-model organisms. . A. Komar, editor Methods in Molecular Biology, Single Nucleotide Polymorphisms, 2d Edition. Humana Press.

Seeb, L. W., M. A. Banks, T. D. Beacham, M. R. Bellinger, S. M. Blankenship, M. R. Campbell, N. A. Decovich, J. C. Garza, C.M. Guthrie III, T. A. Lundrigan, P. Moran, S. R. Narum, J. J. Stephenson, K. J. Supernault, D. J. Teel, W. D. Templin, J. K.Wenburg, S. F. Young, and C. T. Smith. 2007. Development of a standardized DNA database for Chinook salmon. Fisheries 32(11):540-552.

Seeb, L. W., C. Habicht, W. D. Templin, K. E. Tarbox, R. Z. Davis, L. K. Brannian, and J. E. Seeb. 2000. Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the *Exxon Valdez* oil spill. Transactions of the American Fisheries Society 129(6):1223-1249.

Smith, C. T., C. M. Elfstrom, L. W. Seeb, and J. E. Seeb. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14(13):4193-4203.

Sokal, R.R. and F.J. Rohlf. 1995. Biometry. 3rd Edition. Freeman, San Francisco, CA.

Wagner, G. F. 1994. The molecular biology of the corpuscles of Stannius and regulation of stanniocalcin gene expression. In: Fish Physiology, edited by N. Sherwood, and C. Hew. New York: Academic, vol. XIII, chapt. 9, p. 273-306.Waples, R. S. 1990a. Conservation genetics of Pacific salmon III. Estimating effective population size. Journal of Heredity 81(4):277-289.

Waples, R. S. 1990b. Temporal changes of allele frequency in Pacific salmon - implications for mixed-stock fishery analysis. Canadian Journal of Fisheries and Aquatic Sciences 47(5):968-976.

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

Wood, C. C., S. McKinnell, T. J. Mulligan, and D. A. Fournier. 1987. Stock identification with the maximum-likelihood mixture model: sensitivity analysis and application to complex problems. Canadian Journal of Fisheries and Aquatic Sciences. 44(4):866-881.

Technical Committee review and comments

Document 5: Status of the SNP baseline for sockeye salmon

Figure 3. It is worth noting that the null expectation for no linkage disequilibrium implicitly assumes an infinite parental population. One actually expects more than the nominal alpha fraction of significant tests simply due to drift. The fact that no general elevation of significant LD was found, despite rather large samples, suggests that most populations do not have small Ne.

Tests for selection. We also are suspicious of results of programs that suggest large numbers of loci apparently under selection. Evidence is accumulating that methods currently in use to identify 'outlier' loci do not fully account for variance in Fst due to historical population demography and population structure. See in particular the two references below:

Excoffier L, Hofer T, Foll M (2009). Detecting loci under selection in a hierarchically structured population. Heredity 103: 285–298.

Hermisson J (2009) Who believes in whole-genome scans for selection? Heredity 103, 283–284; doi:10.1038/hdy.2009.101; published online 5 August 2009

[*Unedited comments from "Panel comments October 2009.doc" related to Technical Document 5*.

Tables

Table 1. Baseline collection information organized geographically by reporting group and subdivided by population. Each line contains an individual collection with associated collection name, collection date (only year is provided for collections where calendar day was not known), and sample size. Some collections were pooled based on geographic proximity and tests of homogeneity (see text for methods). Collections that were pooled fall under the same number under the "Pop #" column. Populations that were out of H-W at more than twice the number of loci than expected by chance (5 loci @ $P = 0.05$) are noted with the number of loci out of H-W equilibrium under the H-W column. Populations with an asterisk (*) were represented by collections with total sample sizes of less than 80 fish. These populations will either have sample sizes increased in subsequent genotyping efforts or be dropped from future analyses.

Reporting group	Pop #	Population	H-W	Collection	Date	N
Western Kamchatka	1	Palana River		Palana River	6/27/2002	48
				Palana River	2002	50
	2	Tigil River		Tigil River	6/18/2002	100
	3	Bistraya River*		Bistraya River	8/16/1998	56
	4	Bolshaya River*		Bolshaya River	8/16/1999	29
				Bolshaya River	2003	40
	5	Kuril Lake		Etamink River Early	8/21/1990	29
				Etamink River Late	9/28/1990	48
				Kirushutk River	2000	49
				Etamink River	8/12/2002	46
				Khakizun Bay	8/25/2002	49
				North Far Bay	8/26/2002	50
	6	Gabruschka Bay*		Gabruschka Bay	8/25/2002	49
	7	Vichenkiya River		Vichenkiya River	2000	96
	8	Olada Bay*		Olada Bay	2000	50
	9	Ozernaya Bay		Ozernaya Bay	2000	50
				Ozernaya River	2000	49
				Ozernaya River	8/5/2003	50

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Ozernaya River	8/14/2002	50
						988
Eastern Kamchatka	10	Avacha Bay*	12	Avacha Bay	2002	60
	11	Kitilgina River*		Kitilgina River	6/29/1998	28
	12	Kozireuka River*		Kozireuka River	1994	40
	13	Dvu 'Yurta River	9	Dvu 'Yurta River	1994	77
				Dvu 'Yurta River	1995	11
	14	Belaia River	7	Belaia River	1994	69
				Belaia River	1995	11
	15	Hapiza River		Hapiza River Early	7/17/1998	96
				Hapiza River Late	9/2/1998	79
	16	Elovka River		Elovka River	1994	69
				Elovka River	1995	40
	17	Azabachje Lake*		Azabachje Lake	2004	30
	18	Kamchatka River Early*		Kamchatka River Early	6/1/1998	79
	19	Kamchatka River Late		Kamchatka River Late	7/21/1998	97
	20	Lake Potat*		Lake Potat	7/29/2001	49
	21	Lake Vati*		Lake Vati	8/7/2002	48
	22	Anana Lagoon*		Anana Lagoon Early	6/24/2002	30
				Anana Lagoon Late	7/4/2002	48
	23	Severnaya Lagoon		Severnaya Lagoon	6/26/2002	97
						1,058
Norton Sound	24	Salmon Lake		Salmon Lake	8/3/2001	96
	25	Glacial Lake		Glacial Lake	8/15/2004	144
	26	Unalakleet River		Unalakleet River	8/22/2007	95
						335

Reporting group	Pop #	Population	H-W	Collection	Date	N
Yukon Kuskokwim	27	Gisasa River*		Gisasa River	7/16/2005	47
				Gisasa River	6/28/2006	18
	28	Andreafsky River		Andreafsky River	6/28/2006	48
				Andreafsky River	7/19/2008	46
	29	Necons River		Necons River	8/1/2006	55
				Necons River	7/28/2007	95
	30	Telaquana Lake Outlet		Telaquana Lake Outlet	8/14/2003	96
	31	Telaquana Lake Beach*		Telaquana Lake Beach	10/4/2005	47
	32	Kogrukluk River		Kogrukluk River	7/6/2001	96
				Kogrukluk River	7/24/2007	48
	33	Salmon River		Salmon River	8/2/2006	142
	34	Kwethluk River		Kwethluk River	2007	141
	35	Kanektok River		Kanektok River	7/16/2002	95
				Kanektok River	7/10/2007	48
	36	Goodnews River North Fork		Goodnews River North Fork	7/23/2002	95
				Goodnews River North Fork	7/20/2006	47
	37	Goodnews River Middle Fork	6	Goodnews River Middle Fork	8/1/1991	48
				Goodnews River Middle Fork	7/15/2001	96
				Goodnews River Middle Fork	6&7/2007	47
						1,355
Togiak	38	Togiak River		Togiak Lake, Sunday Creek	8/21/2000	94
				Togiak Lake, Outlet	7/27/2006	95
	39	Ongivinuk Lake		Ongivinuk Lake	8/24/2006	142
	40	Nenevok Lake		Nenevok Lake	8/24/2006	142
	41	Gechiak Lake		Gechiak Lake	8/21/2000	96
	42	Kulukak Lake		Kulukak Lake	8/24/2006	142

Reporting group	Pop #	Population	H-W	Collection	Date	N
						711
Igushik	43	Ualik Lake		Ualik Lake	8/14/2003	95
	44	Ongoke Lake Lower		Ongoke Lake Lower	8/28/2007	143
	45	Ongoke Lake Upper		Ongoke Lake Upper	8/27/2007	94
	46	Amanka Lake		Amanka Lake	8/14/2003	94
						426
Wood	47	Lake Kulik beaches		Lake Kulik beaches	9/10/2007	95
				Lake Kulik beaches	9/10/2007	78
				Lake Kulik beaches	7/27/2008	8
	48	Grant River		Grant River	8/22/2007	92
	49	Lake Kulik		Lake Kulik	8/1/2001	96
	50	Silver Horn Beaches		Silver Horn Beaches	9/10/2007	95
				Silver Horn Beaches	9/10/2007	94
				Silver Horn Beaches	7/27/2008	124
	51	Hardluck Bay		Hardluck Bay Beaches	9/10/2007	95
				Hardluck Bay	9/1/2008	156
	52	Agulukpak River		Agulukpak River	8/21/2001	96
	53	Anvil Bay Beach		Anvil Bay Beach	8/20/2006	94
				N4 Beach	8/11/2006	94
	54	Little Togiak Lake		A Beach	8/8/2004	65
				A Beach	8/10/2005	30
	55	Pick Creek		Pick Creek	8/3/2001	93
				Pick Creek	7/22/2008	90
	56	Sixth Creek		Sixth Creek	8/1/2008	94
	57	Agulowok River		Agulowok River	8/22/2001	95
	58	Lynx Beach		Lynx Beach	8/11/2006	95

Reporting group	Pop #	Population	H-W	Collection	Date	N
	59	Lynx Creek		Lynx Creek	8/22/2001	96
	60	Ice Creek Upper*		Ice Creek Upper	8/10/2007	67
	61	Aleknagik Lake Creeks		Happy Creek	7/30/2001	95
				Bear Creek	8/2/2001	96
				Hansen Creek	8/4/2004	95
				Ice Creek Lower	8/9/2007	95
	62	Yako Creek*		Yako Creek	8/1/2008	68
	63	Yako Beach		Yako Beach	8/19/2006	95
	64	Eagle Creek		Eagle Creek	8/12/2007	93
	65	Mission Creek		Mission Creek	1998	93
						2,672
Nushagak	66	Mulchatna River Upper		Mulchatna River	8/27/2001	96
				Mulchatna River	8/27/2001	65
	67	Mulchatna River Lower		Koktuli River	8/13/2000	96
				Stuyahok River	8/14/2000	96
	68	Nushagak River Upper		Klutapuk Creek	8/18/2001	95
				King Salmon River	8/18/2001	96
				Upper Nushagak Sloughs	8/19/2001	96
	69	Chauekuktuli Lake beach		Chauekuktuli Lake Beach	8/22/2001	96
	70	Allen River		Allen River	8/22/2001	95
	71	Allen River Beach		Allen River Beach	8/17/2000	95
	72	Nuyakuk Lake		Nuyakuk Lake	8/16/2000	99
				Nuyakuk Lake South Beach	8/23/2001	94
	73	Tikchik Lake Creek		Tikchik Lake Creek	8/18/2000	95
	74	Tikchik River		Tikchik River	8/18/2001	96
						1,310

Reporting group	Pop #	Population	H-W	Collection	Date	N
Kvichak	75	Tlikakila River		Tlikakila River Glacier Fork	10/6/1999	47
				Tlikakila River Upper	9/24/2001	96
	76	Little Lake Clark		Little Lake Clark	10/9/1999	95
	77	Kijik River Lower		Kijik River Lower	9/18/2001	96
	78	Kijik River		Kijik River	9/19/2001	96
	79	Chulitna Lodge Beach		Chulitna Lodge Beach	10/5/1999	100
				Chulitna Lodge Ponds	10/1/1999	47
	80	Sucker Bay Lake		Sucker Bay Lake	9/14/2007	95
	81	Newhalen River		Tazimina River	8/29/2001	96
				Newhalen River	9/3/2002	96
	82	Tomkok Creek		Tomkok Creek	8/24/2000	95
				Tomkok Creek	8/28/2002	48
	83	Northeast Iliamna Lake		Knutson Bay Late	10/16/1999	95
				Bear Pond Late	10/17/1999	47
				Grass Pond Late	10/15/1999	44
				Pedro Ponds	1999	47
				Knutson Bay	8/27/2000	96
	84	East Iliamna Lake		Chinkelyes Creek	8/28/2000	97
				Finger Beach 1	8/24/2000	84
				Iliamna River	8/21/2004	46
	85	Iliamna River Late		Iliamna River Late	10/17/1999	96
	86	Iliamna Lake Islands		Fuel Dump Island	8/28/2000	99
				Triangle Island	8/16/2000	96
				Woody Island West Beach	8/19/2001	100
	87	Tommy Creek		Tommy Creek	8/24/2000	96
				Tommy Creek	8/19/2002	48
	88	Copper River		Copper River	8/23/1999	47
				Copper River	8/28/2000	96

Reporting group	Pop #	Population	H-W	Collection	Date	N
	89	South Iliamna Lake		Gibralter River	8/23/1999	47
				Belinda Creek	8/25/2000	95
				Dennis Creek	8/23/2000	96
				Gibralter River	8/25/2000	100
				Nick N Creek	8/25/2000	96
	90	Gibraltar Lake		Southeast Creek	8/26/2000	96
				Dream Creek	8/22/2001	96
	91	Upper Talarik Creek		Upper Talarik Creek	8/15/2004	94
				Upper Talarik Creek	8/10/2006	94
	92	Lower Talarik Creek		Lower Talarik Creek	8/26/2000	96
				Lower Talarik Creek	8/23/2001	70
						3,221
Alagnak	93	Moraine Creek		Moraine Creek	9/4/2001	96
				Funnel Creek Early	8/8/2004	171
				Moraine Creek	9/9/2004	96
				Moraine Creek Early	8/8/2004	190
	94	Battle Lake		Battle Creek	9/4/2001	96
				Battle Creek	9/8/2004	96
				Battle Lake Beach	9/11/2004	190
				Battle Lake Tributary	9/11/2004	192
	95	Nanuktuk Creek		Nanuktuk Creek	9/9/2004	191
				Nanuktuk Creek Early	8/9/2004	190
	96	Kulik River		Kulik River	9/5/2001	96
				Kulik River	9/8/2004	96
						1,700
Naknek	97	American River		American River	8/22/2000	95

Reporting group	Pop #	Population	H-W	Collection	Date	N
				American River	8/17/2001	95
	98	Grosvenor Lake		Grosvenor Lake	8/12/2003	96
	99	Hardscrabble Creek		Hardscrabble Creek	8/12/2003	95
	100	Iliuk Arm		Katolinat Creek #1	9/17/2006	48
				Margot Creek	8/15/2001	95
	101	East La Gorce Creek*		East La Gorce Creek	8/27/2006	47
	102	Headwater Creek		Headwater Creek	7/22/2001	132
	103	Brooks Lake		Brooks Lake	8/22/2000	100
	104	Dumpling Creek #1*		Dumpling Creek #1	8/26/2006	48
	105	Dumpling Creek #3		Dumpling Creek #3	9/17/2006	83
	106	Charlene Creek*		Charlene Creek	9/11/2006	47
	107	Lower Q-Tip Lake		Lower Q-Tip Lake	9/12/2006	86
	108	North La Gorce Creek*		North La Gorce Creek	9/10/2006	47
	109	Idavain Creek		Idavain Creek	8/23/2000	96
				Idavain Creek	8/29/2006	48
						1,258
Egegik	110	East Becharof Lake		Becharof Creek	8/11/2000	96
				Cabin Creek	8/15/2000	96
				Ruth Lake Outlet	8/12/2000	95
				Cleo Creek	8/16/2001	95
				Featherly Creek	8/16/2001	95
				Burls Creek	8/16/2006	93
				Salmon Creek	8/16/2006	190
	111	Kejulik River		Kejulik River Upper	8/8/2000	47
				Kejulik River	8/17/2001	96
	112	Becharof Lake North		Becharof Lake North Tributary	8/11/2008	189
	113	Becharof Lake South		Becharof Lake South Beach	8/11/2008	189

Reporting group	Pop #	Population	H-W	Collection	Date	N
						1,281
Ugashik	114	Ugashik Creek		Ugashik Creek	7/21/2001	96
	115	Ugashik Lake		Ugashik Narrows	8/24/2000	97
				Deer Creek	7/20/2001	96
				East Creek Mouth	8/8/2005	95
				Black Creek	8/24/2005	95
	116	Outlet Stream		Outlet Stream	8/26/2000	96
	117	Figure 8 Creek		Figure 8 Creek	8/22/2005	94
	118	Old Ham Creek		Old Ham Creek	8/22/2005	95
						764
North Peninsula	119	Cinder River		Mainstem Cinder River	7/29/2005	95
				Wiggly Creek	7/29/2005	80
	120	Lava Creek		Lava Creek	7/23/2004	92
				Mud Creek A	7/30/2005	95
	121	Meshik Lake		Meshik Lake Shoals	7/30/2005	95
				Meshik Lake Outlet	7/30/2005	95
	122	Meshik River		Blue Violet Creek	7/29/2002	92
				Landlock Creek	7/29/2002	96
				L Creek	7/30/2005	95
	123	Red Bluff Creek		Red Bluff Creek	7/30/2005	95
	124	Willie Creek		Willie Creek	8/27/2001	81
	125	Wildman Lake		Wildman Lake	7/30/2005	94
	126	Ocean River		Ocean River	2001	96
	127	Sandy Lake		Sandy Lake	6/30/2000	96
				Sandy Lake	7/8/2007	95
	128	Bear River Early		Bear River Early	6/30/2000	96

Reporting group	Pop #	Population	H-W	Collection	Date	N
	129	Bear River Late		Bear River Late	8/18/2000	96
	130	Hoodoo Lake		Hoodoo Lake	7/31/2001	95
				Hoodoo Lake Shoals	7/31/2005	95
				Nelson River	2007	47
	131	Nelson River		Nelson River	7/5/2000	96
	132	Davids River		Davids River	7/31/2005	95
	133	North Creek		North Creek	7/25/2007	91
	134	Paul Hansen Tributary		Paul Hansen Tributary	7/30/2002	95
	135	Outer Marker Lake		Outer Marker Lake	9/9/2004	95
	136	Swanson’s Lagoon		Swanson’s Lagoon	8/25/2008	95
	137	Peterson Lagoon		Peterson Lagoon	8/2/2005	95
	138	Whaleback Mountain Creek		Whaleback Mountain Creek	7/30/2002	96
	139	Summer Bay Lake		Summer Bay Lake	8/25/1999	96
	140	McLees Lake		McLees Lake	6/4/2004	142
						2,817
South Peninsula	141	Hansen Lake		Hansen Lake	8/2/2005	95
	142	Middle Lagoon		Middle Lagoon	7/28/2004	142
	143	Thin Point Lagoon		Thin Point Lagoon	8/1/2005	95
	144	Mortensen's Lagoon		Mortensen's Lagoon	8/2/2004	142
	145	Long John Lagoon		Long John Lagoon	8/1/2005	95
	146	Archeredin Lake		Archeredin Lake	8/3/2005	95
	147	Sanak Island		Sanak Island	8/24/2008	86
	148	Canoe Bay River		Canoe Bay River	8/26/2008	95
	149	Orzinski		Orzinski	7/1/2000	95
	150	Black Lake		Big Spring	1997	95
				Broad Creek	9/1/1997	94
				Boulevard Creek	9/1/1997	95

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Alec River	9/1/1997	96
				Fan Creek	1997	95
	151	Chiaktuak Creek Early		Chiaktuak Creek Middle	9/18/1997	94
				Chiaktuak Creek Early	1997	94
				Chiaktuak Creek Early	8/29/2008	174
	152	Chiaktuak Creek Late*		Chiaktuak Creek Late	10/23/1996	50
	153	West Fork Black River Upper		West Fork Black River Upper	8/28/2008	179
	154	West Fork Black River		West Fork Black River	1997	95
	155	Hatchery Beach Early		Hatchery Beach	9/15/1997	95
				Hatchery Creek Early	8/29/2008	94
				Cucumber Creek	8/29/2008	119
	156	Hatchery Beach Late		Hatchery Beach Late	10/18/1996	95
	157	Clark River Early		Clark River Early	8/28/2008	121
				Clark River Early	9/16/1997	96
	158	Clark River Late		Clark River Late	10/19/1996	95
	159	Chignik River		Chignik River	8/22/1998	95
	160	Surprise Lake		Surprise Lake	8/22/2008	95
						3,006
Western GOA	161	Upper Station Lower		Upper Station Lower	1993	95
	162	Upper Station Upper		Upper Station Upper	9/1/1993	95
	163	Upper Station Early		Upper Station Early	6/15/2000	95
	164	Akalura Lagoon Late		Akalura Lagoon Late	9/2/2005	95
	165	Frazer Lake Upper		Pinnell Creek Mouth	8/21/2008	78
				Stumble Creek Mouth	8/21/2008	95
				Courts Beach	8/21/2008	95
				Midway Creek Mouth	8/21/2008	93
				Midway Beach	8/21/2008	95

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Linda Creek Mouth	8/22/2008	95
	166	Hollow Fox Beach		Hollow Fox Beach	8/22/2008	95
	167	Frazer Lake Lower		Outlet Beach	8/20/2008	95
				Valarian Creek	8/21/2008	95
	168	Dog Salmon Creek		Dog Salmon Creek	8/22/2008	95
	169	Horse Marine Lake		Horse Marine Lake	9/2/2005	95
	170	Ayakulik River		Ayakulik River	7/26/2000	94
				Ayakulik River Late	8/14/2008	94
	171	Karluk Lake		O'Malley River	9/30/1999	95
				Lower Thumb River	9/30/1999	95
	172	Upper Thumb Lake		Upper Thumb Lake	7/24/2000	95
	173	Little River Lake		Little River Lake	7/15/1997	96
	174	Uganik Lake		Uganik Lake	7/15/1997	95
	175	Buskin Lake		Buskin Lake	6/26/2005	95
	176	Lake Louise		Lake Louise	8/3/2005	95
	177	Pasagshak Lake		Pasagshak Lake	7/15/2005	95
	178	Lake Miam		Lake Miam	9/2/2005	94
	179	Saltery Lake		Saltery Lake	1994	95
				Saltery Lake	8/26/1999	93
	180	Ocean Beach		Ocean Beach	8/29/2006	95
	181	Afognak Lake*		Afognak Lake	8/15/1993	79
	182	Malina Creek		Malina Creek	8/15/1993	80
	183	Thorsheim Lake		Thorsheim Lake	8/23/2006	83
	184	Portage Lake		Portage Lake	1998	96
	185	Paul's Lake*		Paul's Lake	1994	70
	186	Little Kitoi		Little Kitoi	9/10/1993	95
	187	Kaflia Lake		Kaflia Lake	8/27/2008	95
	188	Crescent Lake Upper		Crescent Lake Site 1	1994	48

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Crescent River	1995	95
	189	Crescent Lake Lower		Crescent River	7/1/1992	95
				Cresent Lake Site 2	1994	47
				Crescent River	7/7/2005	95
	190	Little Jack Creek		Little Jack Creek	9/6/2006	142
	191	South Fork Big River		South Fork Big River	8/14/2007	218
	192	Wolverine Creek		Wolverine Creek	7/5/1993	95
	193	Black Sand Creek		Black Sand Creek	8/13/2007	124
	194	Farro Lake Creek		Farro Lake Creek	8/13/2007	155
	195	McArthur River		McArthur River	1993	95
	196	Chilligan River		Chilligan River	1992	95
				Chilligan River	1994	48
	197	Chakachatna Slough		Chakachatna Slough	8/27/2008	95
	198	Beluga River		West Fork Coal Creek	1993	95
				Lone King Creek	9/4/2006	30
				Lone King Creek	8/27/2008	30
	199	Packers Lake		Packers Lake	7/1/1992	95
				Packers Lake	1993	48
	200	Moose Creek Yentna		Moose Creek Yentna	8/27/2007	106
	201	Puntilla Lake		Puntilla Lake	9/6/2006	143
	202	Red Salmon Lake		Red Salmon Lake	9/7/2006	131
	203	Trimble River		Trimble River Site 1	9/17/2007	61
				Trimble River Site 2	9/17/2007	47
	204	Canyon Creek		Skwentna River	9/20/2007	108
				Canyon Creek	9/20/2007	65
	205	Judd Lake		Judd Lake	8/23/1993	95
				Judd Lake	7/26/2006	94
	206	Trinity Lake		Trinity Lake	8/1/1992	94

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Trinity/Movie Lakes	9/2/1993	95
	207	Shell Lake		Shell Lake	9/3/1993	95
				Shell Lake	7/24/2006	95
	208	Hewitt Lake		Hewitt Lake	8/1/1992	49
				Hewitt Lake	8/2/2006	65
	209	Kichatna River		Kichatna River Site 1	8/27/2007	103
				Kichatna River Site 2	8/27/2007	19
	210	Yentna River West Fork		West Fork Unnamed Slough	9/1/1992	96
				West Fork Yentna River	9/10/1993	99
	211	Chelatna Lake		Chelatna Lake	8/28/1993	95
				Chelatna Lake	7/27/2006	95
	212	Swan Lake		Swan Lake	9/2/2006	95
				Swan Lake	8/15/2007	47
	213	Byers Lake		Byers Lake	1993	95
				Byers Lake	8/13/2007	95
	214	Spink Creek		Spink Creek	8/27/2007	30
				Spink Creek	8/30/2008	95
	215	Susitna River Sloughs		Susitna River Slough # 11	1995	50
				Susitna River Slough # 11	9/5/1996	6
				Susitna River Sloughs 8, 11, 21	9/5/1997	95
	216	Stephan Lake		Stephan Lake	9/2/1993	95
				Stephan Lake	7/28/2007	95
	217	Sheep River		Sheep River	8/30/2008	189
	218	Larson Lake		Larson Lake	9/1/1993	95
				Larson Lake	7/23/2006	95
	219	Mama and Papa Bear Lakes		Mama and Papa Bear Lakes	9/3/1997	50
				Talkeetna River Sloughs	9/4/1997	79

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Papa Bear Lake	8/28/2007	53
	220	Birch Creek		Birch Creek	1993	67
				Birch Creek	8/28/2007	133
	221	Nancy Lake		Nancy Lake	8/27/1993	95
	222	Big Lake		Big Lake	8/1/1992	95
	223	Fish Creek	6	Fish Creek	1993	95
				Fish Creek	8/15/1994	94
				Fish Creek	8/1/1992	95
				Fish Creek	8/5/2008	190
	224	Cottonwood Wasilla Creeks		Cottonwood Creek	1993	95
				Wasilla Creek	1998	71
	225	Eska Creek		Eska Creek	9/5/2006	95
	226	Jim Creek		Jim Creek	9/2/1997	95
	227	Bodenburg Creek		Bodenburg Creek	8/30/2006	143
	228	Sixmile Creek		Sixmile Creek	7/30/2008	94
	229	Williwaw Creek		Williwaw Creek	9/7/2006	39
				Williwaw Creek	8/23/2007	69
	230	Swanson River		Swanson River	8/21/1997	95
	231	Bishop Creek		Bishop Creek	1993	95
	232	Daniels Lake		Daniels Lake	1993	95
	233	Trail Lake Creeks		Railroad Creek	8/13/1997	95
				Johnson Creek	8/12/1997	88
	234	Moose Creek		Moose Creek	7/27/1993	47
				Moose Creek	1994	95
	235	Ptarmigan Creek		Ptarmigan Creek	8/1/1992	47
				Ptarmigan Creek	1993	95
	236	Tern Lake		Tern Lake	9/1/1992	47

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Tern Lake	1993	95
	237	Quartz Creek		Quartz Creek	8/6/1993	95
	238	Between Skilak and Kenai Lakes		Russian River below falls	8/2/1993	93
				Kenai River Late	9/11/1993	47
				Kenai River Early	8/18/1993	48
				Kenai River Site 1	8/22/1994	47
				Kenai River Site 2	8/22/1994	48
				Kenai River Site 4	8/22/1994	48
				Kenai River Early	1994	96
				Kenai River Site 3	8/22/1994	47
				Kenai River Site 5	9/9/1994	95
	239	Upper Russian Lake Late Bear Creek		Upper Russian Lake Late Bear Creek	8/29/1997	94
	240	Upper Russian Lake Early		Upper Russian River Early, Weir	7/1/1992	96
				Goat Creek	8/19/1997	95
	241	Upper Russian Lake Late South		Upper Russian Lake Late South	9/16/1999	95
	242	Upper Russian Lake Late North		Upper Russian Lake Late North	9/17/1999	95
	243	Lower Russian Lake Late Outlet		Lower Russian Lake Late Outlet	8/2/1993	95
	244	Hidden Lake		Hidden Creek	7/29/1993	95
				Hidden Lake North Shore	9/23/2008	95
	245	Skilak Lake Outlet		Skilak Lake	8/1/1992	96
				Skilak Lake Outlet Early	1994	140
				Skilak Lake Outlet Late	1994	140
				Skilak Lake	1995	48
	246	Tustumena Lake		Moose Creek	8/1/1992	96
				Nikolai Creek	7/1/1992	95
				Bear Creek	8/10/1993	95
				Glacier Flats Creek	8/4/1994	95
				Seepage Creek	1994	95

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Tustumena Lake Site A	1994	48
				Tustumena Lake Site B	1994	48
	247	English Bay	8	English Bay Early	6/1/1992	95
				English Bay Late	10/1/1992	95
	248	Delight River*		Delight River	1993	71
	249	Erb Creek		Erb Creek	8/1/1991	94
	250	Eshamy Creek		Eshamy Lake	10/1/1991	95
				Eshamy Creek	8/3/2008	95
	251	Main Bay		Main Bay	7/13/1991	94
	252	Coghill Lake		Coghill Lake	9/1/1991	96
				Coghill Lake North	8/27/1992	91
				Coghill Lake East	8/27/1992	94
	253	Miners Lake		Miners Lake	8/20/1991	93
	254	Eyak Lake Middle Arm		Eyak Lake Middle Arm	8/2/2007	95
	255	Eyak Lake South Beaches		Eyak Lake South Beaches	8/22/2007	94
	256	McKinley Lake		McKinley Lake	8/20/2007	95
	257	McKinley Lake Salmon Creek		McKinley Lake Salmon Creek	7/25/2007	95
	258	Mentasta Lake		Mentasta Lake	7/15/2008	197
	259	Tanada Creek		Tanada Creek	8/21/2005	94
	260	East Fork Gulkana River Fish Creek		East Fork Gulkana River Fish Creek	8/1/2008	211
	261	East Fork Gulkana River*		East Fork Gulkana River	8/1/2008	75
	262	Swede Lake		Swede Lake	8/13/2008	201
	263	Mendeltna Creek		Mendeltna Creek	8/22/2008	108
	264	Banana Lake		Banana Lake	8/18/2008	81
	265	Bear Hole		Bear Hole	8/14/2008	144
	266	St. Anne Creek		St. Anne Creek	7/15/2005	94
				St. Anne Creek	7/22/2008	205
	267	Mahlo River		Mahlo River	7/22/2008	191

Reporting group	Pop #	Population	H-W	Collection	Date	N
	268	Klutina River		Klutina River	8/21/2008	156
	269	Long Lake		Long Lake	9/7/2005	95
	270	Tebay River		Tebay River	8/18/2008	197
	271	Bremner River Salmon Creek		Bremner River Salmon Creek	8/17/2008	99
	272	Bremner River Steamboat Lake		Bremner River Steamboat Lake	8/17/2008	177
	273	Clear Creek		Clear Creek	8/24/2007	94
	274	Martin Lake		Martin Lake	7/26/2007	95
	275	Kushtaka Lake		Kushtaka Lake	8/9/2007	95
	276	Bering Lake		Bering Lake	7/12/1991	95
						17,259
Eastern GOA	277	East Alsek River		East Alsek River	10/15/2000	96
	278	Klukshu River		Klukshu River	8/23/2006	95
	279	Upper Tatshenshini		Upper Tatshenshini	2003	95
	280	Neva Lake		Neva Lake	7/11/2008	94
	281	Chilkat River Bear Flats		Chilkat River Bear Flats	8/9/2007	95
	282	Chilkat River Mule Meadows		Chilkat River Mule Meadows	8/1/2003	95
	283	Chilkat River Mosquito Lake		Chilkat River Mosquito Lake	8/4/2007	95
	284	Chilkat Lake Early		Chilkat Lake Early	7/29/2007	95
	285	Chilkat Lake Late		Chilkat Lake Late	8/12/2007	95
	286	Chilkoot River		Chilkoot River	10/3/2003	95
	287	Chilkoot Lake Beaches		Chilkoot Lake Beaches	7/21/2007	95
	288	Berners Bay		Berners Bay	8/18/2003	95
	289	Windfall Lake		Windfall Lake	7/31/2003	48
				Windfall Lake	8/2/2007	48
	290	Steep Creek		Steep Creek	8/20/2003	95
	291	Nahlin River		Nahlin River	7/31/2003	50
				Nahlin River	7/31/2007	34

Reporting group	Pop #	Population	H-W	Collection	Date	N
	292	Tatsamenie Lake		Tatsamenie Lake	1992	95
	293	Tatsamenie Lake		Tatsamenie Lake	2005	95
	294	Little Tatsamenie Lake		Little Tatsamenie Lake	9/21/1990	64
				Little Tatsamenie Lake	9/11/1991	25
	295	Little Trapper Lake		Little Trapper Lake	9/21/1990	95
	296	Kuthai Lake		Kuthai Lake	2006	95
	297	Taku River Mainstem		Taku River Mainstem	9/24/2007	95
	298	Snettisham Hatchery		Speel Lake	9/17/2003	95
				Snettisham Hatchery	11/27/2006	95
	299	Crescent Lake		Crescent Lake	9/10/2003	94
	300	Kook Lake		Kook Lake	7/30/2007	95
	301	Sitkoh Lake		Sitkoh Lake	9/26/2003	95
	302	Kanalku Lake		Kanalku Lake	7/7/2007	95
	303	Falls Lake		Falls Lake	9/2/2003	95
	304	Salmon Lake		Salmon Lake	7/21/2007	91
	305	Redfish Lake Beaches		Redfish Lake Beaches	8/10/1993	95
	306	Kutlaku Lake		Kutlaku Lake	9/17/2003	95
	307	Petersburg Lake		Petersburg Lake	8/23/2004	95
	308	Kah Sheets Lake		Kah Sheets Lake	8/25/2003	96
	309	Tahltan Lake		Tahltan Lake	2006	95
	310	Little Tahltan Lake		Little Tahltan Lake	9/24/1990	95
	311	Stikine Devil's Elbow*		Stikine Devil's Elbow	9/7/2007	55
	312	Scud River		Scud River	9/13/2007	88
	313	Porcupine River*		Porcupine River	9/13/2007	36
	314	Stikine Andy Smith Slough*		Stikine Andy Smith Slough	9/15/2007	10
	315	Stikine Fowler Slough*		Stikine Fowler Slough	9/15/2007	11
	316	Craig River*		Craig River	2006	12
				Craigson Slough	9/14/2007	43

Reporting group	Pop #	Population	H-W	Collection	Date	N
				Craig River	2007	5
	317	Iskut River		Iskut River	1985	30
				Iskut River	1986	24
				Iskut River	2002	29
	318	Shakes Slough Creek*		Shakes Slough Creek	8/22/2006	41
				Shakes Slough Creek	8/16/2007	13
	319	Mill Creek	7	Mill Creek Early	7/24/2007	95
				Mill Creek Late	8/12/2007	95
	320	Kunk Lake		Kunk Lake	9/14/2003	96
	321	Thoms Lake		Thoms Lake	9/2/2004	94
	322	Neck Lake		Neck Lake	4/23/2007	95
	323	McDonald Lake Hatchery Creek		McDonald Lake	9/15/1992	96
				McDonald Lake	9/5/2003	93
				Hatchery Creek	9/1/2007	95
	324	McDonald Lake Outlet		McDonald Lake Outlet	2007	95
	325	Gene's Lake		Gene's Lake	8/17/2007	95
	326	Helm Lake		Helm Lake	9/21/2005	95
	327	Heckman Lake		Heckman Lake	9/25/2004	95
				Heckman Lake	9/21/2007	95
	328	Mahoney Creek*		Mahoney Creek	8/15/2003	58
	329	Hugh Smith Lake Cobb Creek*		Hugh Smith Lake Cobb Creek	9/6/2007	62
	330	Hugh Smith Lake Bushmann Creek		Hugh Smith Lake Bushmann Creek	9/8/2004	95
	331	Salmon Bay Lake		Salmon Bay Lake	9/10/2004	95
	332	Red Bay Lake		Red Bay Lake	1992	50
				Red Bay Lake	9/13/2004	95
	333	Shipley Lake		Shipley Lake	9/8/2003	94
	334	Sarkar Lakes		Sarkar Lakes	2000	45
				Five Finger Creek	9/8/2005	50

Reporting group	Pop #	Population	H-W	Collection	Date	N
	335	Three Mile Creek		Three Mile Creek	9/30/2004	95
	336	Hetta Lake		Hetta Lake	10/1/2003	94
	337	Klakas Lake		Klakas Lake	9/12/2004	95
	338	Kegan Lake		Kegan Lake	9/10/2004	95
	339	Karta River		Karta River	8/25/1992	93
				McGilvery Creek	9/4/2003	96
	340	Luck Lake		Luck Lake	9/10/2004	94
	341	Sweetwater Lake		Sweetwater Lake	6/7/2003	47
				Sweetwater Lake	6/23/2007	95
	342	Essowah Lake		Essowah Lake	9/5/2004	96
	343	Bowser Lake		Bowser Lake	9/13/2001	95
	344	Damdochax Creek		Damdochax Creek	9/18/2001	94
	345	Tintina Creek		Tintina Creek	9/12/2006	94
	346	Meziadin Lake		Meziadin Lake	9/19/2001	91
				Meziadin Beach	9/26/2006	95
	347	Hanna Creek		Hanna Creek	9/3/2006	93
	348	Kitlope Lake		Kitlope Lake	8/3/2006	95
	349	Four Mile Creek		Four Mile Creek	8/29/2006	85
	350	Pinkut Creek		Pinkut Creek	8/25/2006	95
	351	Pierre Creek		Pierre Creek	8/30/2006	95
	352	Fulton River		Fulton River	2006	95
	353	Morrison Arm		Morrison Arm	9/7/2007	92
	354	Lower Tahlo River		Lower Tahlo River	1988	10
				Lower Tahlo River	1994	85
	355	Upper Babine River		Upper Babine River	2006	95
	356	Sustut River		Sustut River	2006	95
	357	Slamgeesh River		Slamgeesh River	8/7/2006	95
	358	Swan Lake		Swan Lake	10/15/2006	94

Reporting group	Pop #	Population	H-W	Collection	Date	N
	359	Nangeese River*		Nangeese River	9/19/2006	42
	360	Zymoetz River*		Zymoetz River	9/3/2006	64
	361	Nanika River		Nanika River	9/21/2007	94
	362	Kitsumkalum Lake*		Kitsumkalum Lake	11/6/2006	56
	363	Lakelse Lake		Lakelse Lake	8/22/2006	93
	364	Alastair Lake		Alastair Lake	9/14/2006	85
	365	Naden River		Naden River	1995	95
	366	Stellako River		Stellako River	9/28/2007	94
	367	Horsefly River		Upper Horsefly River	9/2/2001	95
				Lower Horsefly River	9/12/2001	95
	368	Chilko Lake		Chilko Lake	1/1/2001	95
	369	Raft River		Raft River	9/4/2001	95
	370	Adams River		Adams River	10/3/2007	95
	371	Birkenhead River		Birkenhead River	10/18/2007	95
	372	Weaver Creek		Weaver Creek	1/1/2001	94
	373	Harrison River		Harrison River	10/17/2007	95
	374	Baker Lake	6	Baker Lake	5/16/1996	97
	375	Cedar River		Cedar River	10/26/1994	96
						9,648

Table 2. Forty-five sockeye SNP markers assayed for this project; three mitochondrial DNA and 42 nuclear DNA. Forward and reverse primers and probes are given for each new Taqman assay. Loci that were out of H-W equilibrium at more than the number of populations expected by chance (19 populations @ P = 0.05) are noted with the number of populations out of H-W equilibrium (P = 0.05) under the H-W column.

Marker	Reference[1]	H-W
One_ACBP-79	A	
One_ALDOB-135	A	
One_ctgf-301	A	
One_CO1 [2]	A	
One_Cytb_17 [2]	A	
One_Cytb_26 [2]	A	
One_E2-65	B	
One_GHII-2165	A	21
One_GPDH-201	B	20
One_GPDH2-187	B	
One_GPH-414	A	
One_hsc71-220	A	
One_HGFA-49	B	21
One_HpaI-71	A	
One_HpaI-99	A	
One_IL8r-362		
F: TTGCTAGAAGCGTTGGTTATGATGA		
R: CAGCAAAATTGAGAAGTCACTAGGAAAA		
VIC- CAGCCAAAGAAGAGTC		
FAM- AGCCAAAAAAGAGTC		
One_KPNA-422	A	
One_LEI-87	A	
One_MARCKS-241		
F: CCTATCACAGCTTGGTTGAGTTCAA		
R: TCCACCCGCTCATTTTTGTAAGAT		
VIC-TTGCTTAAAAGGTCTTCC		
FAM-TTGCTTAAAAGGTCATCC		
One_MHC2_190 [3]	A	29
One_MHC2_251 [3]	A	30
One_Ots213-181	A	
One_p53-534	A	
One_ins-107	B	23
One_Prl2	A	

Marker	Reference[1]	H-W
One_RAG1-103	A	
One_RAG3-93	A	
One_RFC2-102	B	
One_RFC2-285	B	
One_RH2op-395	A	
One_serpin-75	B	
One_STC-410	A	22
One_STR07	A	
One_Tf_ex11-750	A	
One_Tf_in3-182	A	
One_U301_92	A	
One_U401-224		20
F: GGGTGGAGACGAACGGATTC		
R: GTACGATTTTTTTGTAGCCCCAAGT		
VIC-CACCTGGAAAGGACTGA		
FAM-ACACCTGGAAATGACTGA		
One_U404-229		
F: GTTTGTGTGTTGGTGTTTGTCCTT		
R: CATTTATCTTGGTGGACGTGTGAGT		
VIC-CATGTTCTTCAGTGAACC		
FAM-ATGTTCTTCAATGAACC		
One_U502-167		
F: GCTTTTGTGCAATAGCTATGTTGCT		
R: GCAAAGGTAGGCAGCAGATTG		
VIC-CTTCTTGATCAATAACG		
FAM-CTTCTTGATCGATAACG		
One_U503-170		20
F: GATTCAGAATTGCCACGACAAAGAA		
R: GTGATTGGTACATGTCTGTCGAGTT		
VIC-AAGTACTAAAATCAGTTTTACATTG		
FAM-TACTAAAATCAGTTGTACATTG		
One_U504-141		
F: GCTATAGCTCACAGAGGATCCCA		
R: TATTGGCGGGTGAGGGATG		
VIC-TCAAGGACACAAACAA		
FAM-TCAAGGACAAAAACAA		
One_U508-533		
F: AGGCACAACCTCACATTTGGAA		

Marker	Reference[1]	H-W
R: CTCAAAGGGTCTGAATACTTATGTAAATAAGGT		
VIC-ACACTACAGCCTTATTC		
FAM-ACACTACAGCTTTATTC		
One_VIM-569	A	
One_ZNF-61		
F: CCATTCATGTTCTATTCAGATATATTTTGTGCA		
R: CCTAGCTAGAGCTCAACAATATGCA		
VIC-CTATGGACATGATCTTT		
FAM-TTCTATGGACATTATCTTT		
One_Zp3b-49	B	

[1] A) Elfstrom et al. (2006); B) Smith et al. (2005).
[2] mtDNA markers; composite haplotype loci were assembled for MSA analyses.
[3] MHC markers were significantly linked in more than 50% of collections. Composite phenotypes were assembled for MSA analyses.

Table 3. Descriptive statistics for SNPs used in the current ADF&G sockeye salmon baseline, including expected (H_e) and observed heterozygosity (H_o) for nuclear loci**,** and F_{ST} for all nuclear and mitochondrial markers and for the combined nuclear marker. Minimum and maximum values and overall F_{ST} are shown, while average heterozygosities include only nuclear markers. Superscripts indicate sets of markers which were pooled into a single locus.

SNP	H_e	H_o	F_{ST}
One_ACBP-79	0.472	0.406	0.121
One_ALDOB-135	0.286	0.252	0.116
One_ctgf-301	0.045	0.042	0.048
One_E2-65	0.338	0.302	0.110
One_GHII-2165	0.307	0.220	0.275
One_GPDH-201	0.492	0.447	0.083
One_GPDH2-187	0.210	0.172	0.168
One_GPH-414	0.447	0.383	0.138
One_hcs71-220	0.333	0.298	0.108
One_HGFA-49	0.307	0.277	0.088
One_HpaI-71	0.465	0.400	0.133
One_HpaI-99	0.204	0.157	0.218
One_IL8r-362	0.123	0.114	0.092
One_KPNA-422	0.378	0.339	0.098
One_LEI-87	0.478	0.420	0.114
One_MARCKS-241	0.032	0.029	0.073
One_MHC2_190[a]	0.491	0.305	0.356
One_MHC2_251[a]	0.491	0.334	0.303
One_Ots213-181	0.277	0.241	0.125
One_p53-534	0.071	0.061	0.125
One_ins-107	0.496	0.434	0.114
One_Prl2	0.500	0.447	0.096
One_RAG1-103	0.055	0.050	0.102
One_RAG3-93	0.160	0.143	0.104
One_RFC2-102	0.348	0.307	0.112
One_RFC2-285	0.099	0.088	0.100
One_RH2op-395	0.018	0.017	0.042
One_serpin-75	0.072	0.066	0.064
One_STC-410	0.456	0.353	0.220
One_STR07	0.460	0.393	0.145
One_Tf_ex11-750	0.488	0.387	0.206
One_Tf_in3-182	0.154	0.112	0.268
One_U301-92	0.277	0.252	0.089
One_U401-224	0.488	0.439	0.107
One_U404-229	0.123	0.103	0.162

SNP	H_e	H_o	F_{ST}
One_U502-167	0.046	0.044	0.049
One_U503-170	0.254	0.224	0.115
One_U504-141	0.389	0.351	0.089
One_U508-533	0.092	0.079	0.125
One_VIM-569	0.219	0.197	0.094
One_ZNF-61	0.415	0.352	0.152
One_zP3b-49	0.235	0.174	0.266
One_CO1[b]	N/A	N/A	0.254
One_Cytb_17[b]	N/A	N/A	0.498
One_Cytb_26[b]	N/A	N/A	0.255
One_CO1_Cytb17_26	N/A	N/A	0.295
One_MHC2_190_251	N/A	N/A	0.259
Minimum	0.018	0.017	0.042
Maximum	0.500	0.447	0.295
Average/Overall	0.288	0.243	0.149

[a] These SNP genotypes were combined into a single locus, *One_MHC2_190_251*, and treated as haploid data.
[b] These SNPs were combined into haplotypes and treated together as an mtDNA locus, *One_CO1_Cytb17_26*.

Table 4. Percent of total collections exhibiting significant linkage disequilibrium for the pairs of loci for which disequilibrium was most commonly observed.

			Significant linkage disequilibrium	
Criteria	Marker pair		Number of collections	Percentage of total
P < 0.01	*One_MHC2_190*	*One_MHC2_251*	320	55%
	One_GPDH	*One_GPDH2*	197	34%
	One_Tf_ex10-750	*One_Tf_ex3-182*	108	19%
	One_RF-112	*One_RF-295*	43	7%

Table 5. Log-likelihood *G* and associated test statistics for the homogeneity of allele frequency log-likelihood ratio tests over all loci across populations within regions and broad regional groupings. Because the number of populations is heterogeneous across regions, we also tabulate *G* divided by degrees of freedom (df) for each regional level.

Broad Regions	Regions	*G*	df	*P*	# of pops	*G* / df
Western Kamchatka	Western Kamchatka	2,927	328	0.00	9	8.92
Eastern Kamchatka	Eastern Kamchatka	6,376	533	0.00	14	11.96
Norton Sound	Norton Sound	1,417	82	0.00	3	17.27
Western Bristol Bay	Yukon Kuskokwim	7,685	410	0.00	11	18.74
	Togiak	1,436	164	0.00	5	8.75
	Igushik	271	123	0.00	4	2.21
	Wood	3,207	738	0.00	19	4.35
	Nushagak	3,566	328	0.00	9	10.87
	Western Bristol Bay Total	16,165	1,763	0.00	48	9.17
Eastern Bristol Bay	Kvichak	15,155	697	0.00	18	21.74
	Alagnak	1,730	123	0.00	4	14.07
	Naknek	2,954	492	0.00	13	6.00
	Egegik	1,093	123	0.00	4	8.89
	Ugashik	608	164	0.00	5	3.71
	Eastern Bristol Bay Total	21,540	1,599	0.00	44	13.47
Alaska Peninsula	North Peninsula	11,994	861	0.00	22	13.93
	South Peninsula	11,105	779	0.00	20	14.25
	Alaska Peninsula Total	23,098	1,640	0.00	42	14.08
Western Gulf	Western Gulf	177,933	4,715	0.00	116	37.74
Eastern Gulf	Eastern Gulf	105,112	4,018	0.00	99	26.16
WAAP		62,220	5,084	0.00	137	12.24
Coastwide Total		354,568	14,678	0.00	375	24.16

Table 6. Proportion of estimates correctly allocated back to reporting group of origin and 90% confidence intervals for mixtures of 400 fish simulated from baseline populations that contribute to each reporting region (100% simulations) using the program SPAM.

		90% Confidence Interval	
Region	Estimate	Lower	Upper
Western Kamchatka	0.969	0.949	0.986
Eastern Kamchatka	0.956	0.933	0.978
Norton Sound	0.946	0.913	0.973
Yukon Kuskokwim	0.908	0.862	0.949
Togiak	0.946	0.898	0.980
Igushik	0.860	0.779	0.929
Wood	0.938	0.881	0.981
Nushagak	0.912	0.862	0.954
Kvichak	0.950	0.924	0.973
Alagnak	0.977	0.961	0.990
Naknek	0.947	0.916	0.974
Egegik	0.913	0.864	0.954
Ugashik	0.855	0.784	0.914
North Peninsula	0.893	0.851	0.932
South Peninsula	0.917	0.882	0.948
Western Gulf of Alaska	0.927	0.896	0.955
Eastern Gulf of Alaska	0.967	0.946	0.985

Table 7. Proportion of estimates correctly allocated back to reporting group of origin and 90% credibility intervals for mixtures of 200 known fish that were removed from the baseline populations that contribute to each reporting region (100% proof tests) using the program BAYES with a flat prior.

		90% Confidence Interval	
Region	Estimate	Lower	Upper
Western Kamchatka	0.990	0.972	1.000
Eastern Kamchatka	0.974	0.934	0.996
Norton Sound	0.985	0.961	0.999
Yukon Kuskokwim	0.978	0.926	0.999
Togiak	0.987	0.960	1.000
Igushik	0.974	0.899	0.999
Wood	0.957	0.823	0.999
Nushagak	0.956	0.866	0.998
Kvichak	0.959	0.901	0.998
Alagnak	0.992	0.973	1.000
Naknek	0.972	0.933	0.997
Egegik	0.947	0.868	0.995
Ugashik	0.959	0.898	0.996
North Peninsula	0.980	0.935	0.999
South Peninsula	0.958	0.914	0.991
Western Gulf of Alaska	0.894	0.827	0.948
Eastern Gulf of Alaska	0.983	0.950	0.999

Figures

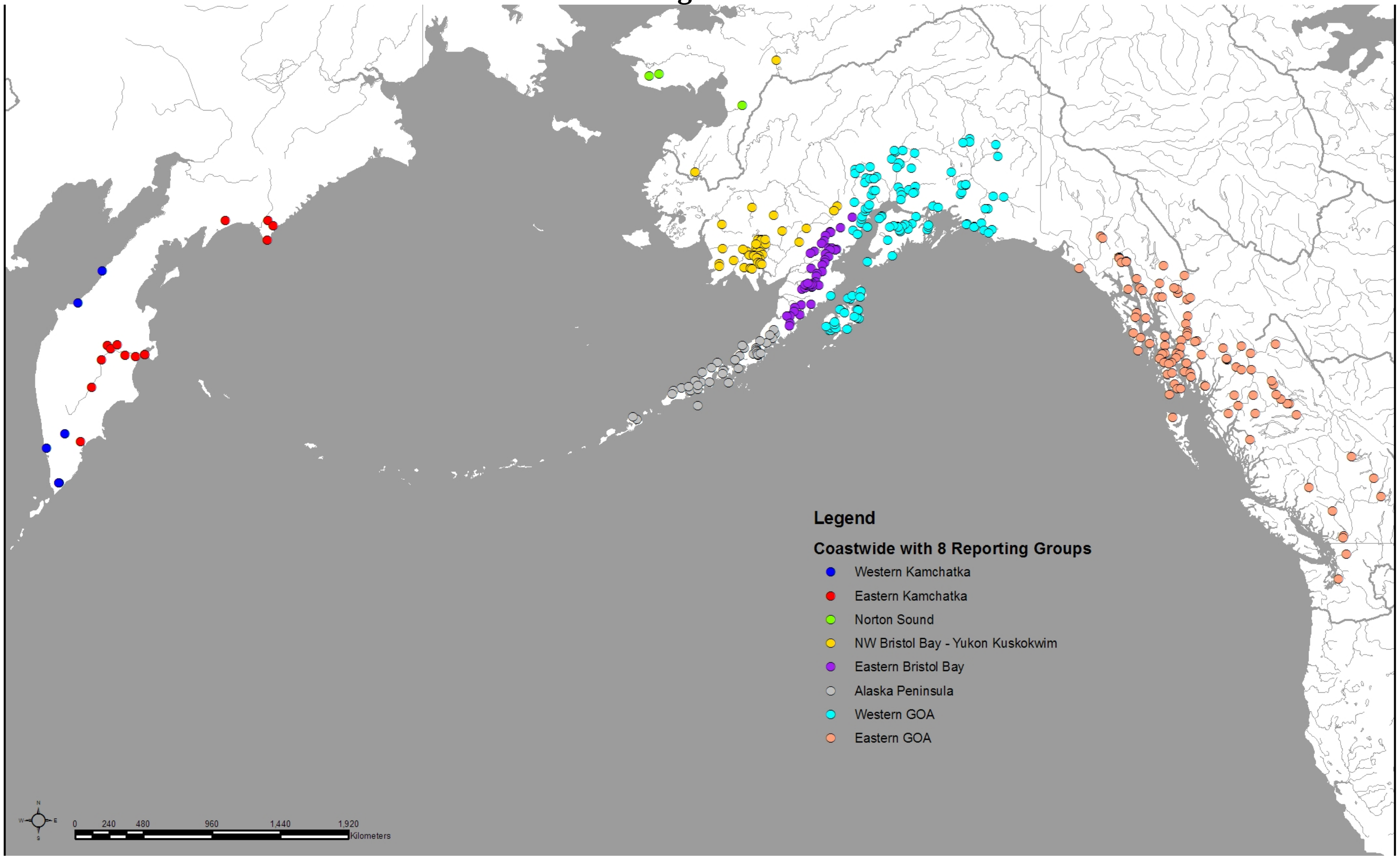

Figure 1. Locations where sockeye salmon were sampled for tissues suitable for genetic analysis from throughout the Pacific Rim. These tissues were screened for 42 nuclear and 3 mitochondrial single nucleotide polymorphism markers. This baseline, augmented with additional markers, will serve as a baseline to examine the potential power and precision of stock composition estimates from fishery samples taken under the Western Alaska Salmon Identification Program. Colors denote eight geographic regions that match the colors and regions in Figure 6. Western and Eastern Kamchatka, Norton Sound, and Eastern and Western Gulf of Alaska represent five of the proposed reporting groups. The remaining regions (Western Bristol Bay YK, Eastern Bristol Bay, and the Alaska Peninsula) are further subdivided into a total of 12 reporting groups as shown in Figures 2 and 7.

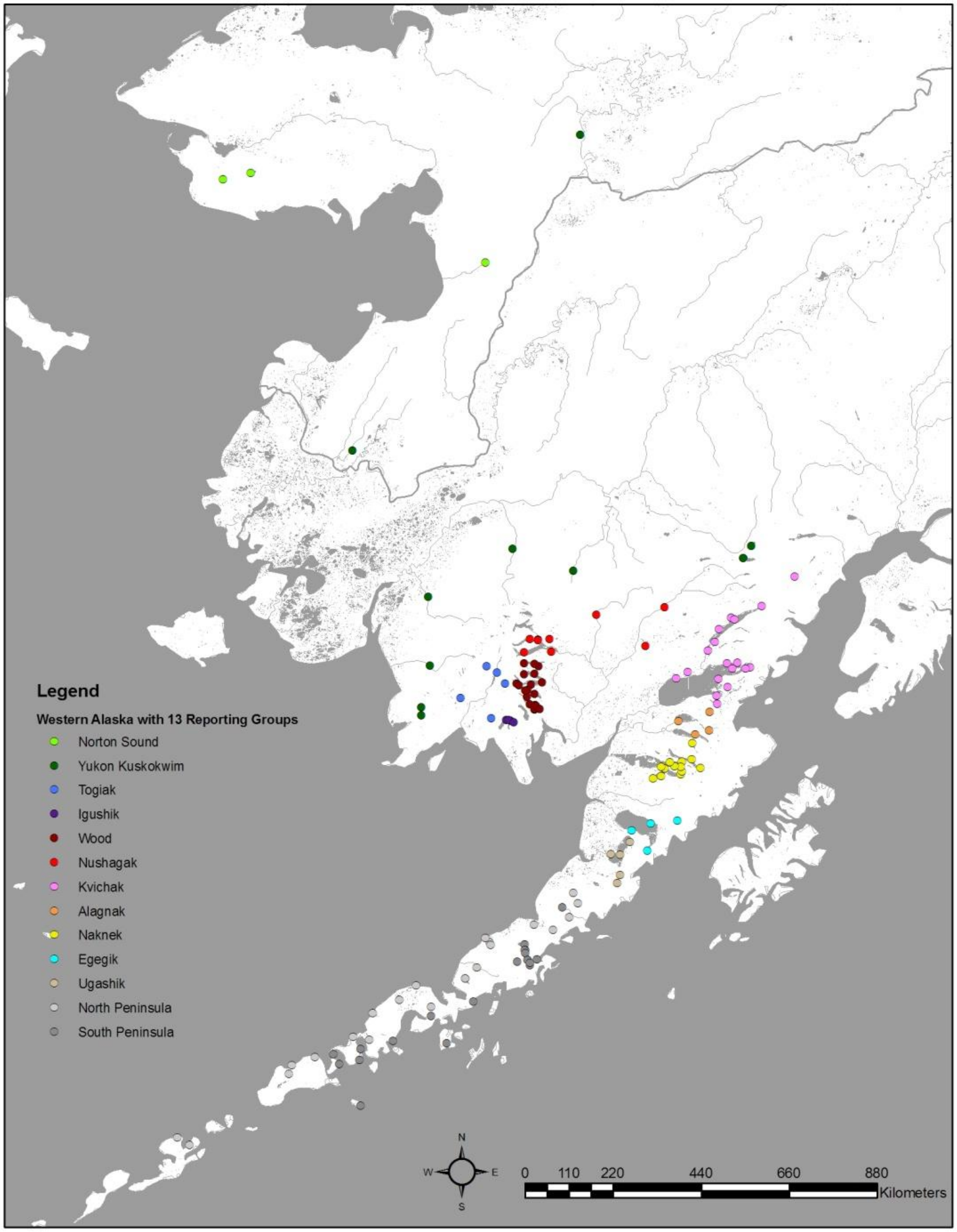

Figure 2. Sockeye salmon sample locations from Western Alaska and the Alaska Peninsula (WAAP) included in the SNP baseline. Colors denote the 13 WAAP reporting regions.

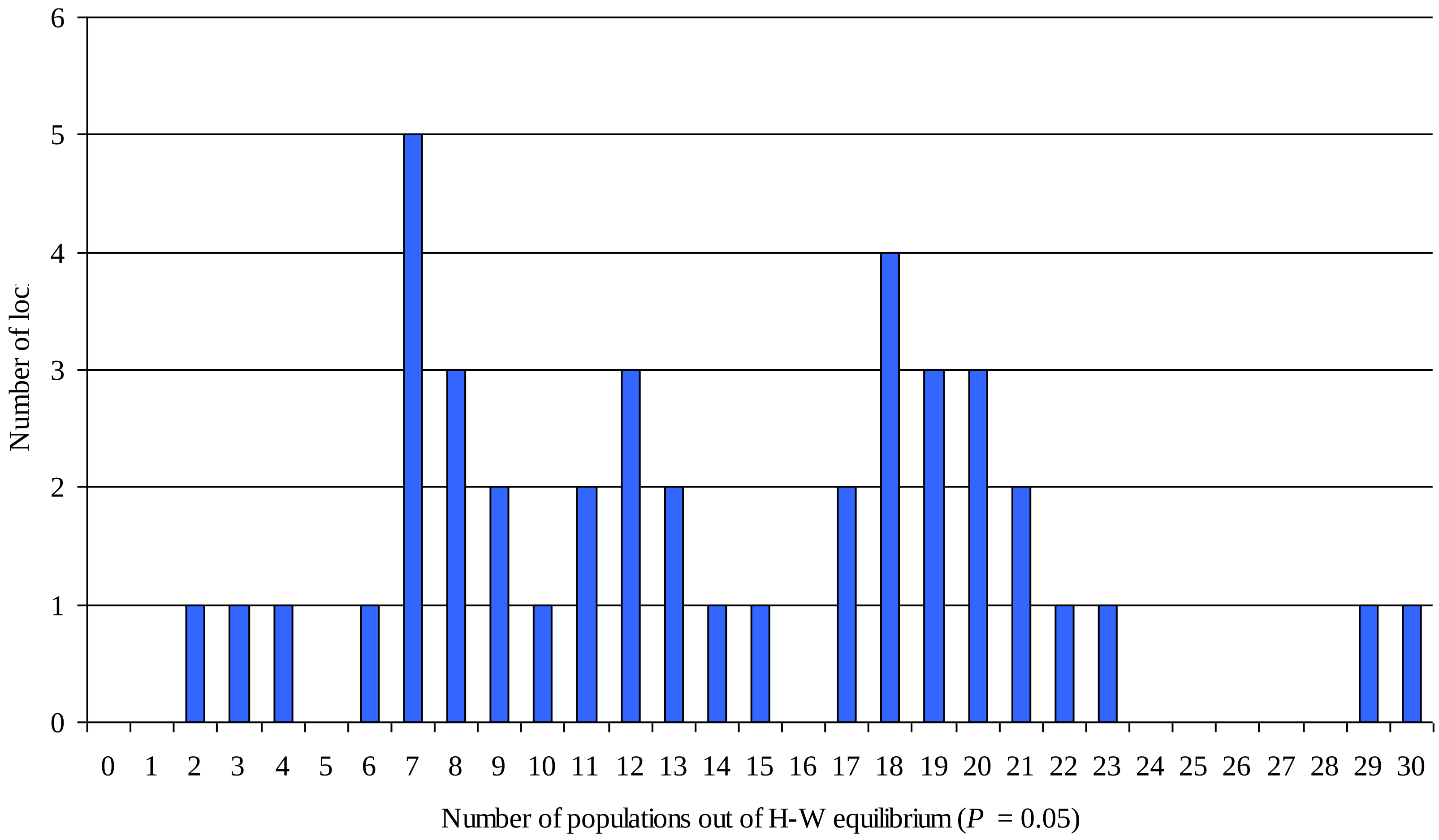

Figure 3. Number of loci that were out of H-W equilibrium (*P* = 0.05) for 0 to 30 populations. By chance, the one would expect 18.75 populations to be out of H-W expectation at this criterion (375 populations * 0.05). We review the loci that were out of H-W equilibrium at more that 23 populations in the text.

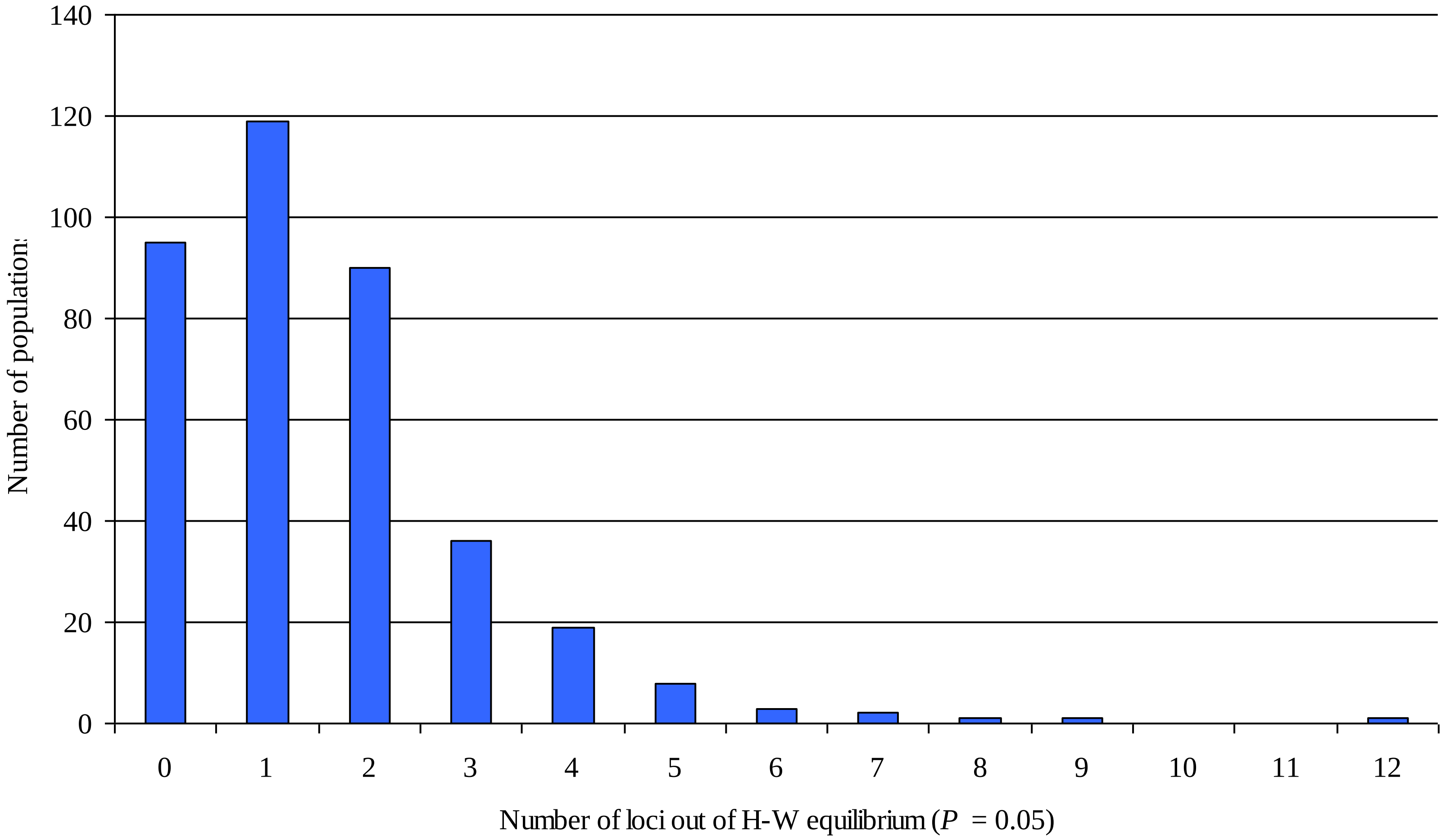

Figure 4. Number of baseline populations that were out of H-W equilibrium (*P* = 0.05) for 0 to 12 loci. By chance, the one would expect 2.1 loci to be out of H-W expectation at this criterion (i.e., 42 loci * 0.05). We review the populations that were out of H-W equilibrium at more that 5 loci in the text.

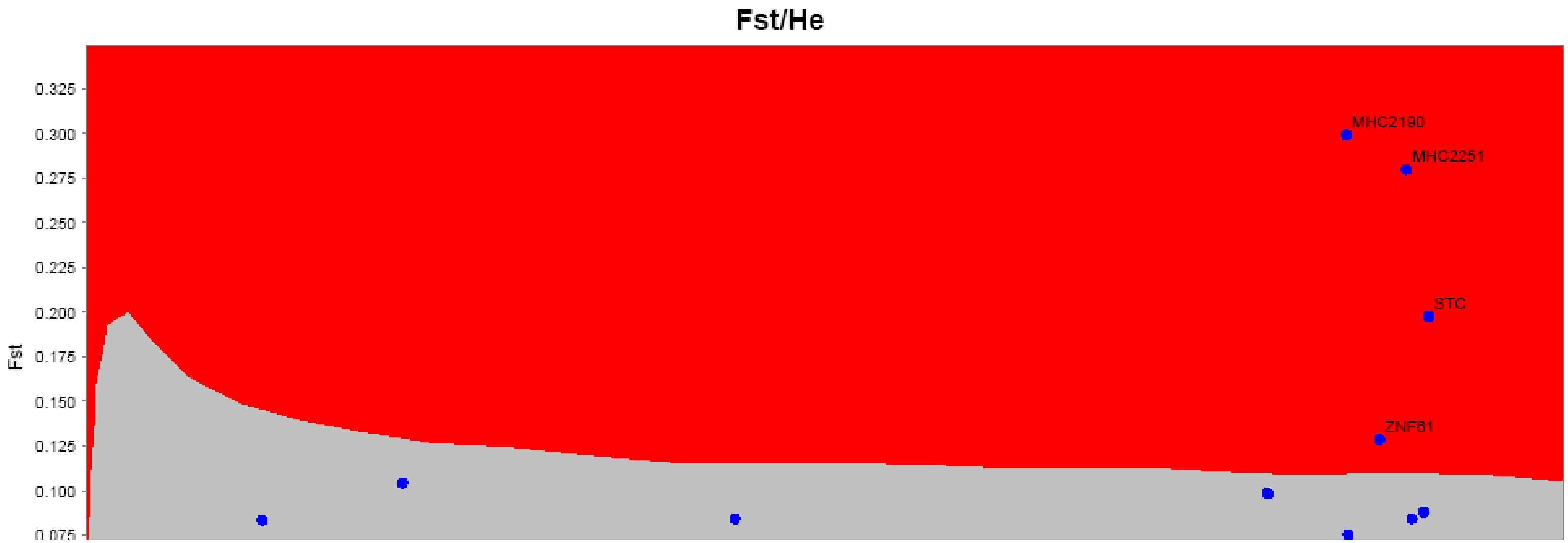

Figure 5. LOSITAN (Antao et al. 2008) graphical output showing the relationship between F_{ST} and H_e for SNP markers analyzed in select populations from western Alaska and the north Alaska Peninsula (method details in text). The expected distribution of F_{ST} and H_e under an island model of migration with neutral markers is shown in gray. Loci in the red area are candidates for positive selection and loci in the yellow area are candidates for balancing selection. Outlier loci are tagged with labels.

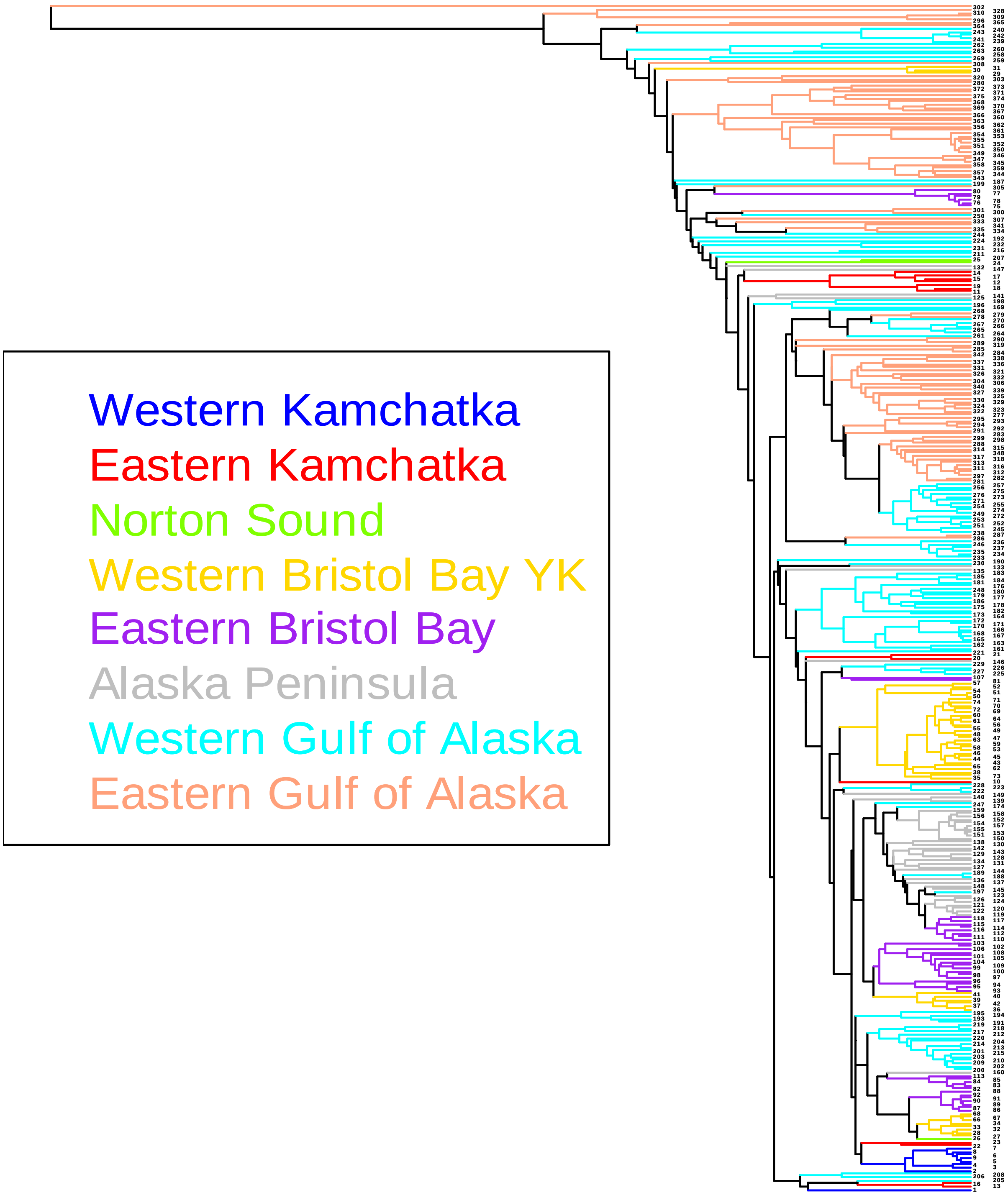

Figure 6. Unweighted pair-group method (UPGMA) tree of Cavalli-Sforza and Edwards chord distances among the 375 populations included in the coastwide 42 SNP baseline. Population numbers correspond to those in Table 1. Note the high variation within the Gulf of Alaska relative to the WAAP.

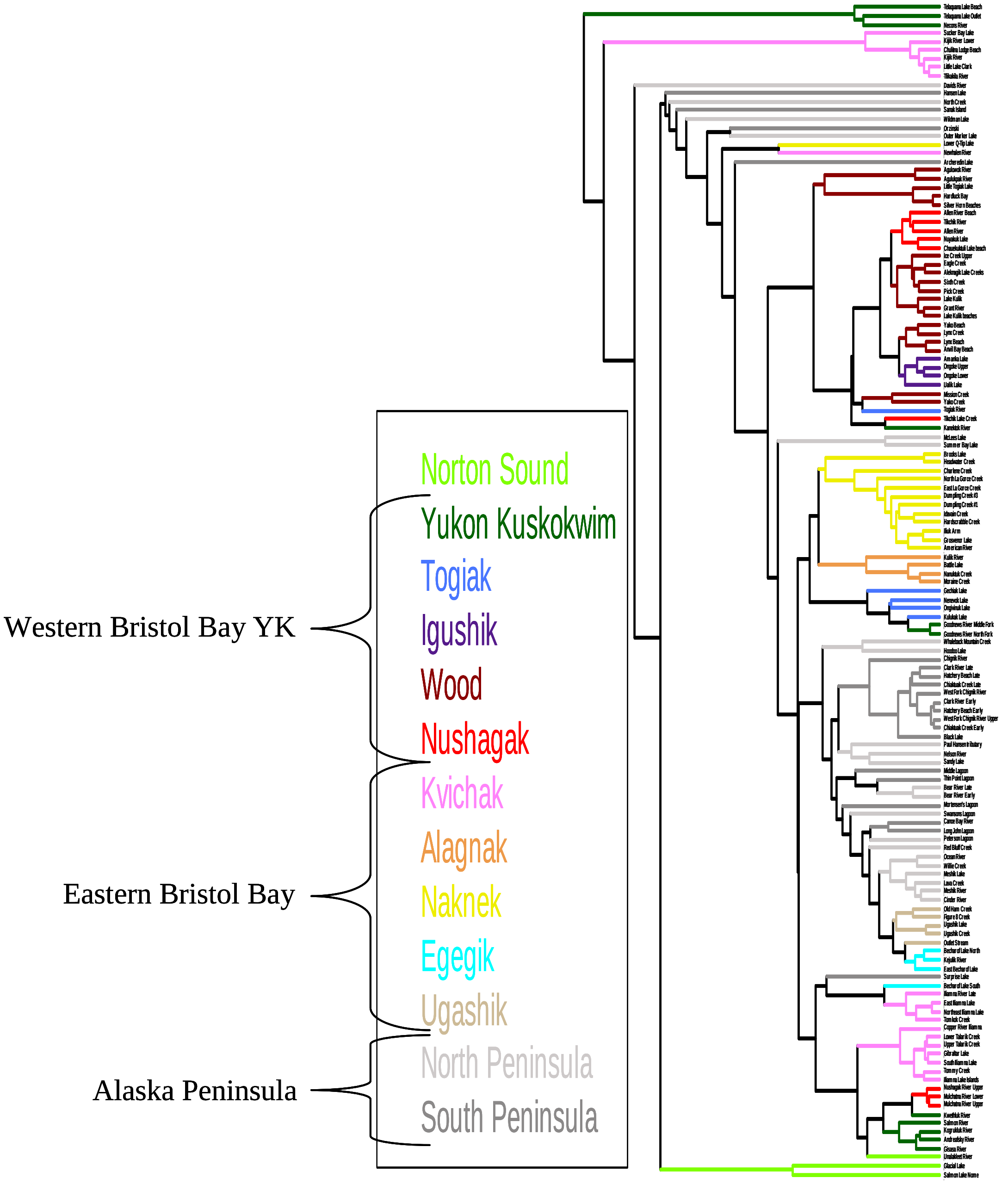

Figure 7. Unweighted pair-group method (UPGMA) tree of Cavalli-Sforza and Edwards chord distances among the 137 populations included in the WAAP portion of the coastwide 42 SNP baseline.

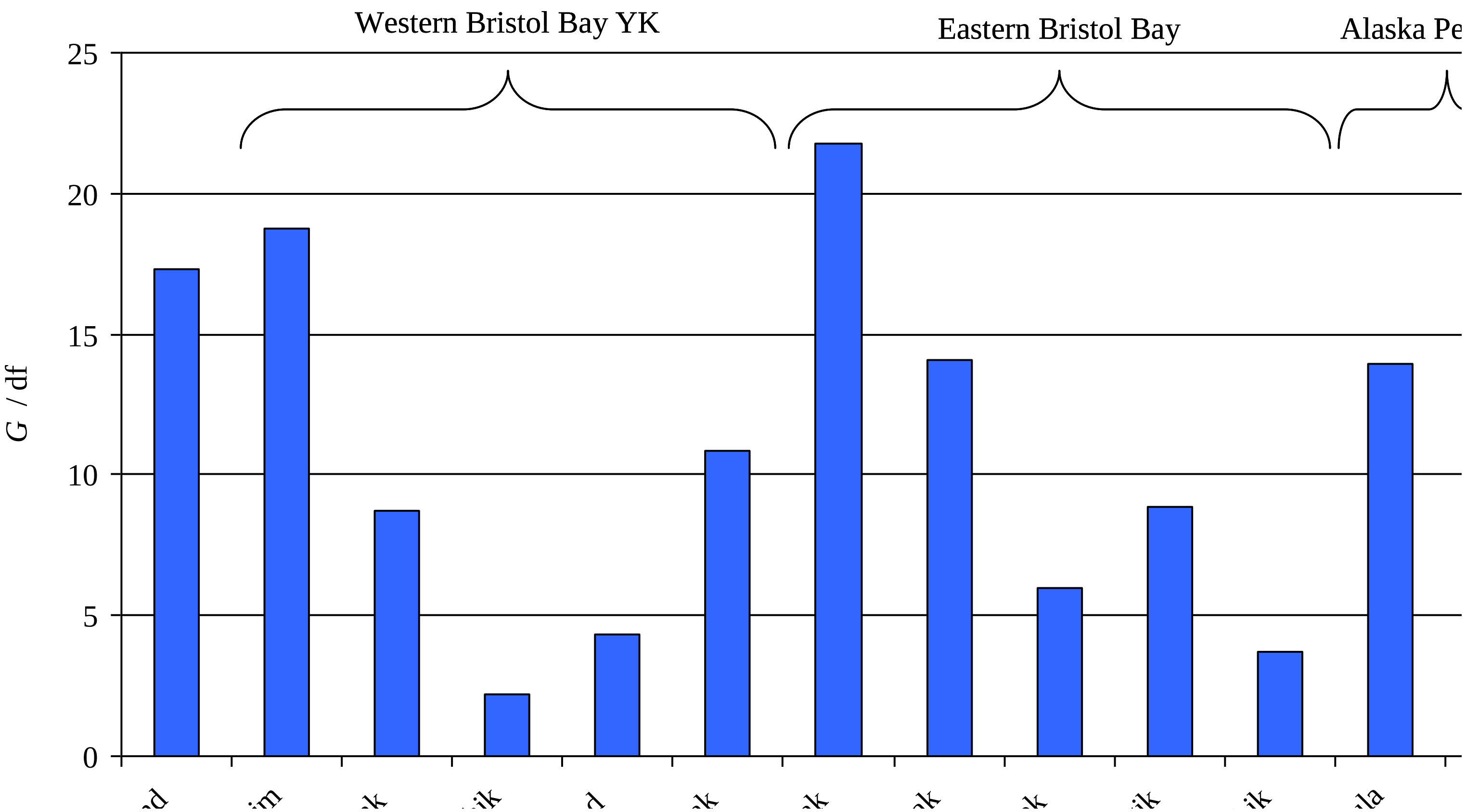

Figure 8. Log-likelihood ratio test statistics (*G*) divided by degrees of freedom (df) over all loci by reporting group within the WAAP area.

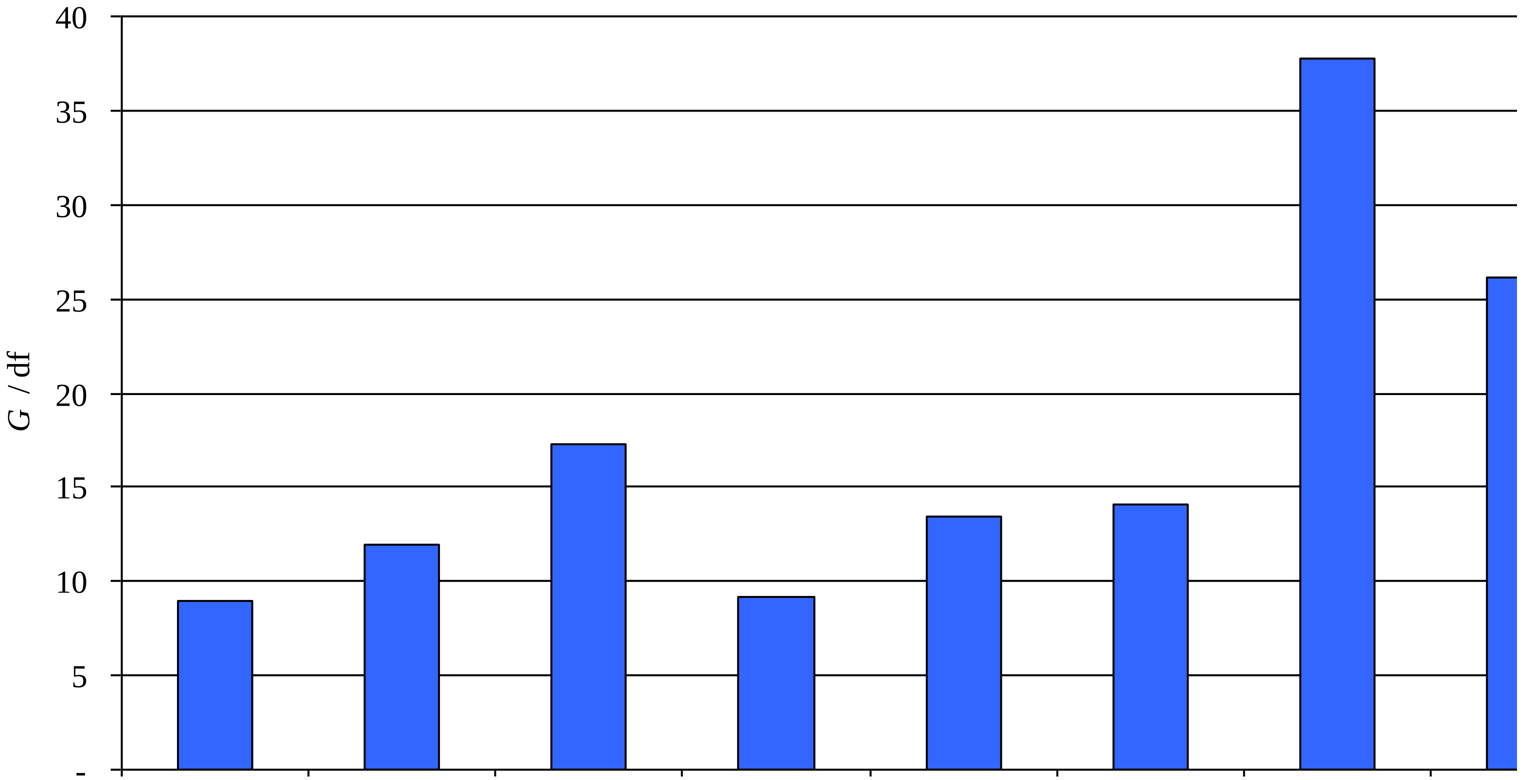

Figure 9. Log-likelihood ratio test statistics (*G*) divided by degrees of freedom (df) over all loci by region within the full baseline.

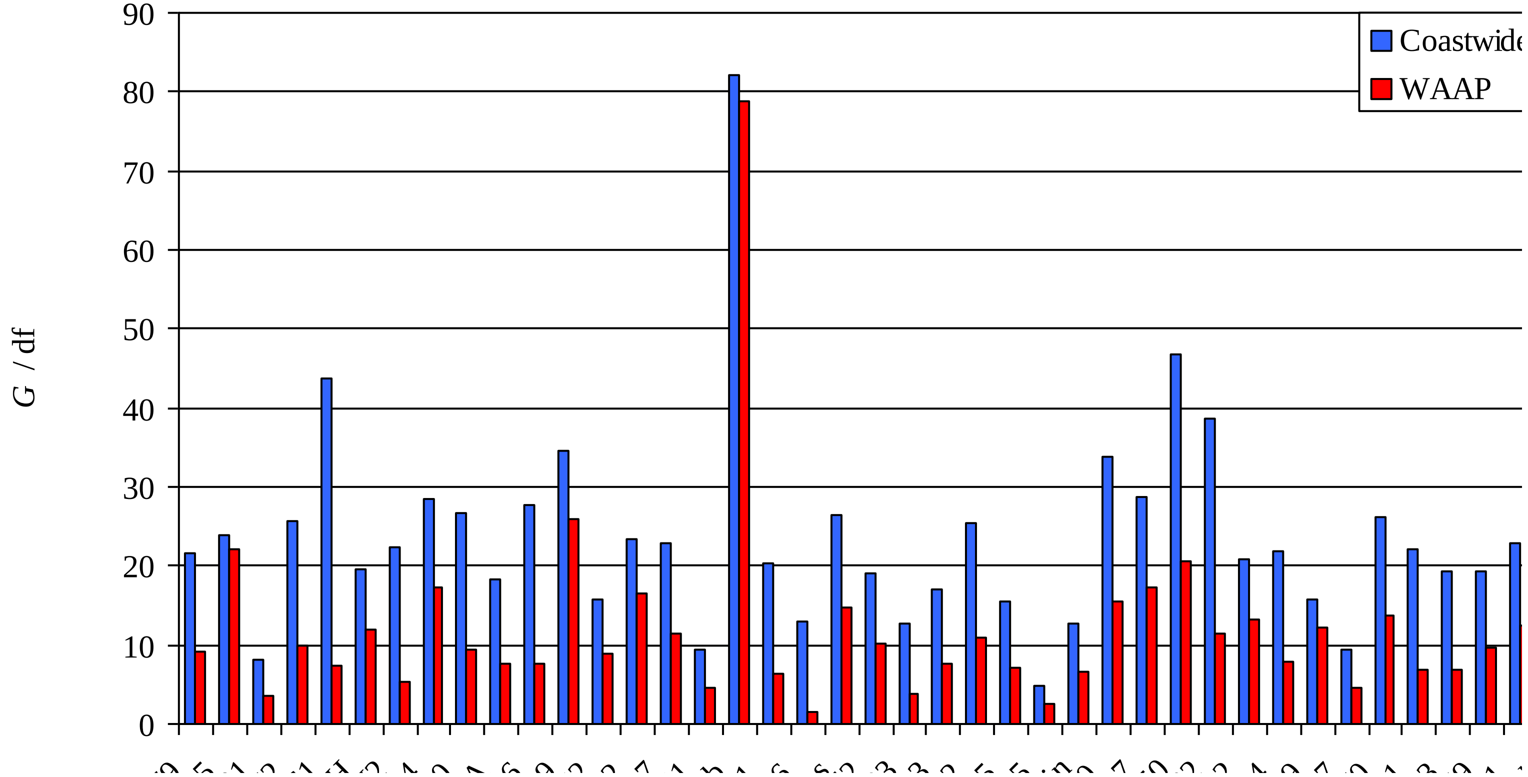

Figure 10. Log-likelihood ratio test (*G*) statistics divided by degrees of freedom (df) for each SNP marker for the populations within the full coastwide baseline and the more restricted WAAP baseline. Note the similar and high values for the *G* statistics for both geographic regions at the one MHC marker included in this analysis and the generally lower values for the *G* statistics in the WAAP area for the remaining markers.

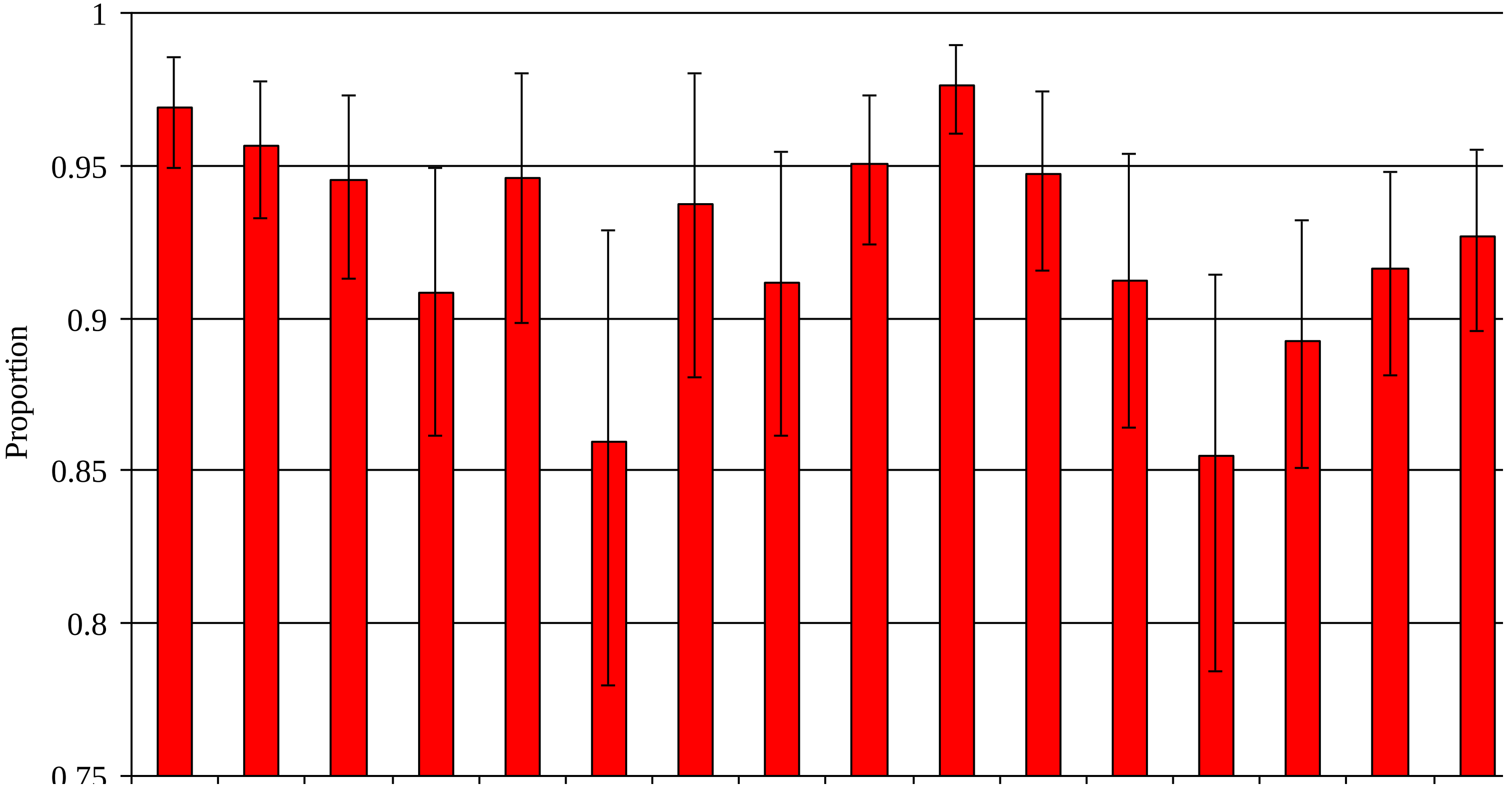

Figure 11. Proportion of estimates correctly allocated back to reporting group of origin and 90% confidence intervals for mixtures of 400 fish simulated from baseline populations that contribute to each reporting region (100% simulations) using the program SPAM.

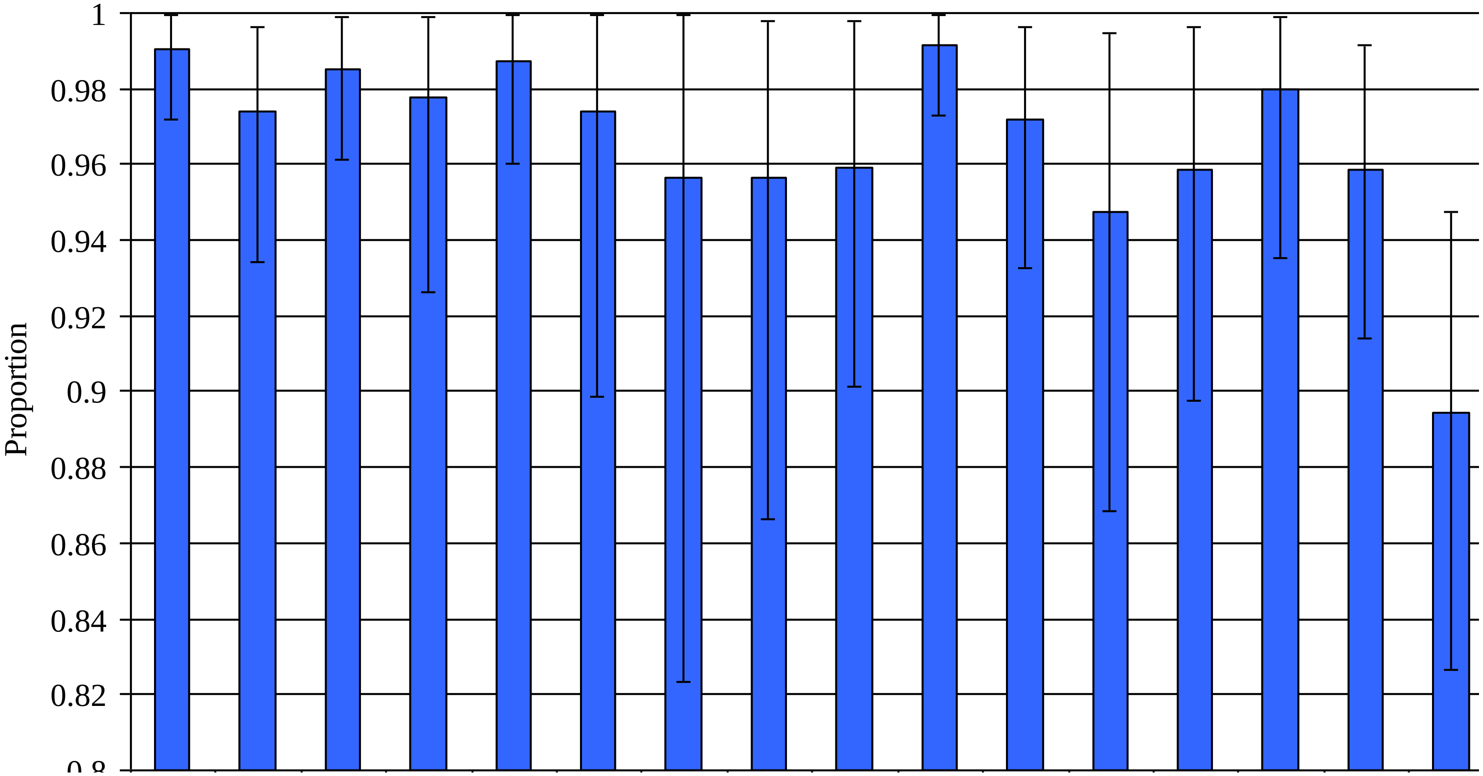

Figure 12. Proportion of estimates correctly allocated back to reporting group of origin and 90% credibility intervals for mixtures of 200 known fish that were removed from the baseline populations that contribute to each reporting region (100% proof tests) using the program BAYES with a flat prior.