

Final report: Developing DNA Markers for the Analysis of Chum Salmon Bycatch in Alaskan Trawl Fisheries

Submitted to: Pollock Conservation Cooperative Research Center
School of Fisheries and Ocean Sciences
245 O'Neill Building
University of Alaska Fairbanks
Fairbanks, AK 99775-7220

By: M.R. Garvin and A.J. Gharrett

Fisheries Division
School of Fisheries and Ocean Sciences
University of Alaska Fairbanks
Fairbanks, AK 99775-7220

January 2009

Abstract

Chum salmon (*Oncorhynchus keta*) populations have experienced declines in Western Alaska and in other areas of their geographic range. Possible reasons include regime shifts in climate, by-catch by other fisheries, and other environmental perturbations. Genetics offers powerful tools that can be used to estimate the origins of fish captured on the high seas in mixed stock fisheries. These tools can aid in reducing by-catch and may help to determine other possible causes for their declines. In this project, we developed and tested genetics methods that can be used to resolve stock mixtures of chum salmon. Although microsatellite variation is one approach to estimating compositions of stock mixtures, microsatellites require both coordination among labs that develop baseline to ensure compatibility and experience in conducting the analyses. The type of marker focused on in this study was single nucleotide polymorphism (SNP), which is becoming increasingly popular because, once the marker has been developed and baseline data acquired, SNP variation can be rapidly surveyed in many individuals. Also, the technology can be readily transported among labs and adapted to different types of instrumentation for analysis. In order to develop SNPs as a tool, however, tens or hundreds of informative SNP sites must be developed from the millions that exist in the chum salmon genome. Current discovery methods such as sequencing introduce ascertainment bias (a form of missampling), which can result in the development of uninformative SNPs or the failure to detect useful ones.

In this project, we developed or improved a number of methods for discovering and applying SNPs to make them more accessible to small laboratories and to reduce their cost. Specifically, we invented DEco-TILLING, a method for discovering useful SNPs rapidly and inexpensively; we improved a genotyping assay to screen SNPs in thousands of individuals for an order of magnitude less cost than the standard commercial assay; and we improved on a method to resolve the phase of SNPs linked close to each other in the same gene, which increases resolution and improves discovery efforts. We also demonstrated that the SNPs, which we discovered, are informative and increase precision of analyses of mixed stock fisheries of chum salmon over microsatellites alone. In a parallel study that was funded by the Bering Sea Fisherman's Association and was leveraged by this PCCRC project, we developed a baseline for microsatellite loci for the same populations that we developed SNP baseline. We now have funding from the Arctic-Yukon-Kuskokwim Sustainable Salmon Initiative to continue

developing both SNP and microsatellite baselines and to apply the baselines to analyze collections of immature chum salmon captured in the Bering Sea between 1998 and 2005.

Introduction

The problem

Salmon returns to western Alaskan systems declined in the last two decades, and although numbers increased recently in some areas, concern remains as to the cause(s) of the declines. Salmon, in particular chum salmon (*Oncorhynchus keta*), are critical to the livelihood and culture of rural Alaskans. In addition, chum salmon are the focus of other issues that include allocation among Alaskan users and between the U.S. and Canada. In order to understand the causes of these declines, it is necessary to learn about the marine distributions and timings of salmon stocks as well as the interannual variation in marine distributions.

Our approach

Substantial effort has been (and continues to be) devoted to genetic studies of North American chum salmon stocks, the objective of which is to resolve mixtures to their contributing components. An extensive allozyme baseline was developed to address those questions in the late 1900s. But sampling logistics, increasing costs of storing and processing samples, and the inadequacy of allozymes to provide fine-scale resolution in some applications have led most labs to abandon their allozyme operations. Allozyme variation is a consequence of changes in the DNA sequences that code for them. It is now possible and practical to examine DNA sequence variation directly. There are many types of DNA markers that can be applied to population genetics and stock identification questions. Two that have been applied to many species are mitochondrial DNA and microsatellite variation. More recently, single nucleotide polymorphism (SNP) analyses, which are nucleotide changes at single sites within the DNA sequence, are being developed for application to chum salmon. The near-term approach to stock identification will probably combine DNA marker types, but it is likely that SNPs or similar markers will eventually be the preferred marker because they reflect much of the variation in the genome and SNPs can easily be adapted to many different rapid laboratory analyses.

The molecular tools that are used to interpret a DNA marker (not just a type of marker, but every single marker) are developed individually for each marker (e.g., 10 SNPs would

require 10 development efforts). Subsequently, reference or baseline data must be acquired from the populations that may be contributors to a sample of fish intercepted in a catch. Once assembled, baseline data are scrutinized by simulation to determine the degree to which contributing populations or groups (often regional groups) of populations can be resolved from a mixture. These results are usually followed by development of additional markers and acquisition of additional baseline to increase resolution for particular applications. Many of the molecular methods for discovery and analysis of markers are still being developed; consequently, baseline information is unavailable for some markers and incomplete for other markers.

In this project, which was funded by the Pollock Conservation Cooperative Research Center (PCCRC) at the University of Alaska Fairbanks (UAF), we have been developing molecular markers for use in resolving natal origins of chum salmon. Our focus has been to determine the origins of chum salmon taken incidentally in the pollock fishery, but the molecular tools we have developed and the information we obtained are and will be applied to a variety of questions pertinent to chum salmon that are addressed by us and other agencies. Our original intentions were to develop and ground truth SNPs from: 1) a previous survey of mtDNA control region nucleotide sequences (Sato et al. 2004), which we did in collaboration with C. Smith (formerly with ADFG, Anchorage); 2) variation that we previously observed in a mtDNA restriction fragment analysis survey that was conducted over the entire chum salmon mtDNA genome (Churikov et al. 2001); 3) other available salmonid DNA sequence data available in the nucleotide sequence data base GenBank; and 4) sequences of the chum salmon major histocompatibility complex (MHC) genes, such as the MHC II B1 and B2 loci (Miller and Withler 1996). To accomplish the fourth approach, we planned to apply a relatively recently developed method called Eco-TILLING (Targeting Induced Local Lesions IN Genomes; Till et al. 2007). After some preliminary experiments, it became apparent that the MHC genes would not be as easy to develop into a tool as we had thought. But that effort lead us to a new approach, which we developed (DEco-TILLING; Garvin and Gharrett 2007), and which we used to discover informative SNPs efficiently.

In this paper, we detail the methods that we developed and improved during the course of this project. We also briefly describe the markers that we developed and show that the markers are informative for delineating chum salmon from different geographical regions. From our

complete suite of SNPs, we developed a limited chum salmon baseline for 72 populations, which we can use to make stock composition estimates. We have since added 3 additional loci to that suite and 16 populations to the baseline. Recently, we conducted simulations based on the SNP and microsatellite loci to evaluate their ability to discriminate chum salmon stocks. Finally, we estimated the composition of two mixed fishery samples, one collected near the Aleutian Islands and another collected west of St. Matthews Island.

Background

Salmon bycatch in the Bering Sea trawl fishery

Chum salmon returns to western Alaskan systems generally declined in the last two decades, and were particularly low from 1997 through 2002 (AYK Scientific Technical Committee 2005), although in recent years returns have rebounded in many areas. Because chum salmon are anadromous fish, the cause of these declines could be due to factors that effect either their freshwater or marine life history stages and may include regime shifts in climate or human impacts such as fishing. Chum salmon are economically and culturally important in western Alaska and rural Alaskans depend on them for subsistence. Reduced numbers contribute to a variety of sociological/allocation, conservation, and international treaty concerns. If the major causes for these declines can be identified, it may be possible to reverse reductions in numbers of returning salmon.

Salmon bycatch, especially of chum and Chinook (*O. tshawytscha*) salmon, is an important issue for Alaskan trawl fisheries. Chum salmon dominate the “non-Chinook” bycatch. National Marine Fisheries Service (NMFS) catch statistics indicate that chum salmon bycatch is prevalent in all sectors of the pollock (*Theragra chalcogramma*) fishery. Some geographical areas have concentrations of young chum salmon bycatch. A hotspot, which was identified by Geographical Information System (GIS) analysis (Ackley 1997), was incorporated into the fishery management plan to reduce chum salmon bycatch by defining a Chum Salmon Savings Area (CSSA), which was closed to trawling during the month of August and remained closed through October 14 if the annual chum trawl bycatch reached the limit set in the surrounding Catcher Vessel Operational Area (CVOA). The CSSA was closed in 2002, 2003, 2004, and 2005. In 2005, the Bering Sea-Aleutian Island (BSAI) chum salmon bycatch in the pelagic pollock trawl fisheries (705,963 fish) exceeded the previous high (242,191) in 1993. It declined

somewhat in 2006 (310,545) and substantially in 2007 (96,072) (Figure 1;

http://www.fakr.noaa.gov/sustainablefisheries/inseason/chum_salmon_mortality.pdf).

Industry and managers have realized for some time that salmon bycatch is a problem. Steps have been taken to reduce bycatch and industry took the lead in monitoring and policing itself through an “Intercooperative chum salmon bycatch management plan” in which they monitor bycatch and close statistical areas for some or all of the participants when the bycatch rate becomes too high. Although chum bycatch is closely monitored and GIS analyses of observer data and industry’s awareness provide invaluable spatio-temporal information, one of the underlying questions, the origins (and destination) of the fish, remains unaddressed. This knowledge is crucial in determining the ultimate impact of the bycatch.

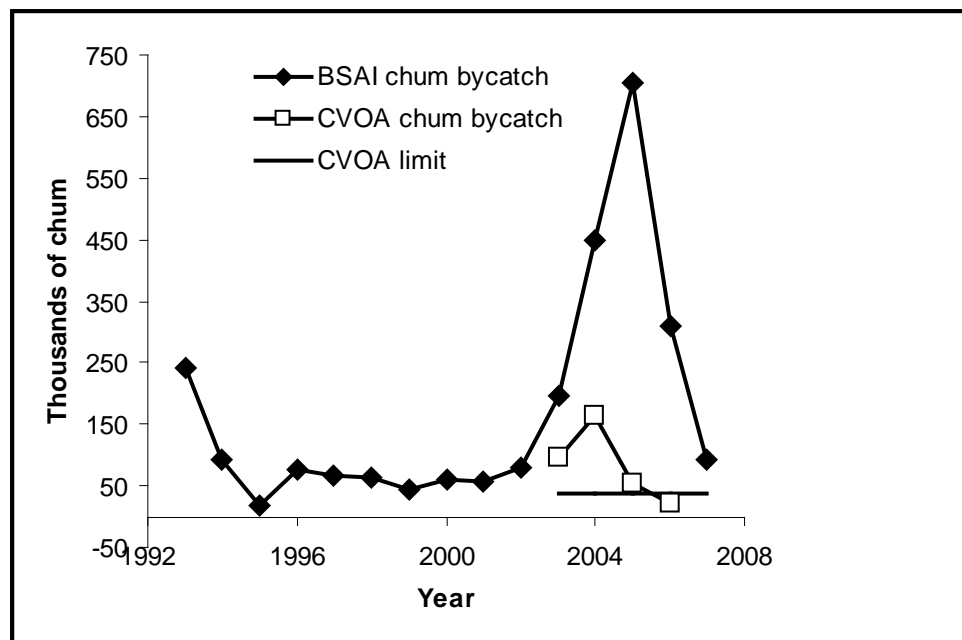


Figure 1. “Non-Chinook” salmon (predominantly chum salmon) bycatch by midwater pollock trawl gear caught in the Bering Sea/Aleutian Islands from 1993 to 2007.

A genetic approach to stock identification

Use of genetic differences, such as divergence in DNA sequences among populations, is the most promising method for stock identification of wild fish in the marine environment. The fish do not need to be captured, handled, and marked as they emigrate to the sea and the traits are heritable, so that they are passed from generation to generation and do not vary substantially from year to year as “environmental markers” like scale patterns may. Substantial effort has been

and continues to be devoted to genetic studies of North Pacific chum salmon stocks. The underlying idea is that reproductively isolated populations, such as salmon that home to natal streams to spawn, will diverge genetically over time as a consequence of random changes in their gene pools and differential local adaptive forces. One of the challenges is that in some areas such as lower-river and coastal western Alaska streams, post glacial colonization by chum salmon occurred relatively recently (on an evolutionary scale) and populations have not yet diverged substantially. Consequently, even genetically isolated populations may resemble each other (e.g., Flannery et al. 2007ab). The result is that it is sometimes difficult to identify fish from particular drainages; but in many instances, the geographic area from which fish originated can be deduced and it is possible to differentiate fish from different geographical regions. By increasing the numbers of informative genetic markers, resolution can be increased. Also, many of the markers that have been historically applied in stock identification are not strongly influenced by local adaptation, that is, they are relatively neutral selectively. Resolution among population or groups of populations can also be enhanced by focusing on markers that are involved in adaptation to local environments. Consequently, a reasonable approach is to identify and apply SNPs from the nuclear genome at loci or in regions of DNA that are likely to be involved in natural selection and would, therefore, show stronger divergence among populations than would neutral markers.

Stock identification is important in estimating origins in bycatch, but it is equally important for learning about the marine distributions of chum salmon stocks and the year-to-year variation in their distributions. From such information, we will be able to piece together the puzzle of the factors that influence marine survivals and learn how to predict the time and place of concentrations of young chum salmon.

Why SNPs?

Single Nucleotide Polymorphisms (SNPs) are gaining wide use as tools in many fields. Whereas SNPs suffer from the disadvantage of being only bi-allelic (two alleles per locus as compared to the dozens of alleles at many microsatellite loci), they have several major advantages over microsatellites. SNPs are the most abundant marker in genomes – approximately one SNP every 500 base pairs in chum salmon, which can be extrapolated to several million SNPs from which markers can be chosen. Most SNP discovery projects sequence individuals to identify markers. Given that there are millions of SNPs from which to choose and

that only a small percentage of these will be useful, the majority of SNPs that are discovered are uninformative. Methods that can streamline discovery of informative SNPs are essential for practical applications to multiple species, each of which requires independent discovery efforts.

One of the most common methods for detecting a particular SNP, after an assay has been designed specifically for it, is the TaqmanTM assay that involves recognition by a short DNA probe that is labeled with a fluorescent dye. The probe is a short stretch of DNA that is designed to complement the DNA sequence that includes the variable site. One probe is designed for each of the two SNP sequences; each probe carries a different dye. Probes are designed by the user and purchased, based on specifications that are provided from the sequences, from a commercial supplier. Both probes are added to each assay. Completely paired probes (including the variable site) are degraded by nuclease activity of the polymerase during the PCR reaction. The degraded probe releases a fluorescent dye that is detected by an instrument such as the ABI Prism[®] 7900. Incompletely paired probes are not cleaved and their dyes do not fluoresce. The homozygote (say **AA**) for one SNP allele will emit a single fluorescence; the homozygote for the other SNP allele (say **aa**) will emit a different color fluorescence; a heterozygote (**Aa**) will emit both colors.

A number of other instrumentation systems have also been developed to screen SNP variation. Some DNA “chips” can genotype enormous numbers of SNPs at a time. However, the chips are expensive, generally not reusable, and require that the suite of SNPs has been completely engineered before they can be produced. Most of the SNPs that will be used for fisheries applications have not yet been developed. Also, many of the methods designed for studies of humans are prohibitively expensive for the number of individuals that must be examined to address fisheries questions.

Another of the challenges in developing SNP markers is that multiple variable nucleotide sites are frequently discovered in the same short segment of DNA. Unfortunately, sites that are very close together are inherited as a unit, that is, they are statistically correlated. The complication is that eukaryotes have two copies of each chromosome (and the genes on those chromosomes). Ordinarily, the region examined for SNPs is only a small part of a single gene and the multiple SNPs identified within a gene are not inherited independently. For example, one SNP site may have two types (**A** and **a**) and the other has two types (**B** and **b**). An individual that has both kinds of SNPs at both sites (is a double heterozygote) might have one gene/chromosome carrying the composite **A-B** and the other carrying **a-b**. Alternatively, one

chromosome could carry **A-b** and the other **a-B**. It is not possible to distinguish between the two from the basic SNP assay, but correctly scoring each chromosome in a stock mixture is essential to stock identification applications. That kind of information is important for providing leverage in the algorithms that are used to resolve mixtures. To determine this linkage phase requires an additional analysis, which we developed. We also refined and modified other methods to improve efficiency of analysis and substantially reduce costs.

When this project began, the US Fish and Wildlife (USFWS) laboratory in Anchorage and Department of Fisheries and Oceans, Canada were developing a baseline for microsatellite loci and applying those data to analyses of Yukon River chum salmon stock mixtures. The Alaska Department of Fish and Game (ADFG) had begun to discover SNP markers and to develop a baseline for applications to Alaskan chum salmon management. Their focus was primarily, but not exclusively, in western Alaska.

Objective

The purpose of this study was to discover genetic markers that can be used to identify the population or geographic region of origin of chum salmon taken as by-catch by the pelagic trawl fishery and in other high seas samples. We focused on developing SNPs, which are single nucleotide variants in the DNA sequence. SNPs are distributed throughout the genome and the most abundant type of marker in the genome. However, their detection is based on sequence information, and very little sequence information was available for chum salmon. One of the problems with surveying sequence data for SNPs is that it is impractical to sequence a sufficient number of individuals that represent a sufficiently broad sample of a species range to avoid sampling errors (ascertainment bias) that result in devoting effort to uninformative SNPs or that fail to detect informative ones. Another problem was the high cost of SNP analysis, especially for small labs. A third challenge was how to use linked SNPs to the best advantage.

Consequently, our objectives evolved to:

1. To develop a method that can survey a large number of individuals for SNP variation rapidly and identify the most informative SNPs while reducing ascertainment bias;
2. To develop SNP analyses that can be applied economically;
3. To develop methods that can resolve linkage phase in closely linked SNPs;
4. To assemble and evaluate a limited SNP baseline;

and 5. To apply the baseline in an estimate of contributions to a mixed-stock sample.

Study area

We analyzed DNA of chum salmon, which were collected from populations that were distributed throughout most of the natural range of the species (Figure 2). The samples included populations from the Sea of Okhotsk, eastern and western Bering Sea, and the Pacific coast of North America. Many of the samples came from archives of allozyme samples.

Two mixed fishery samples of chum salmon, one sampled from the pelagic trawl fishery in August 2005 near the Aleutian Islands at 55° 20' N 167° 01' W and one sampled during BASIS surveys in August 2005 west of St. Matthews Island at about 59° 30' N 176° 00' W were analyzed.

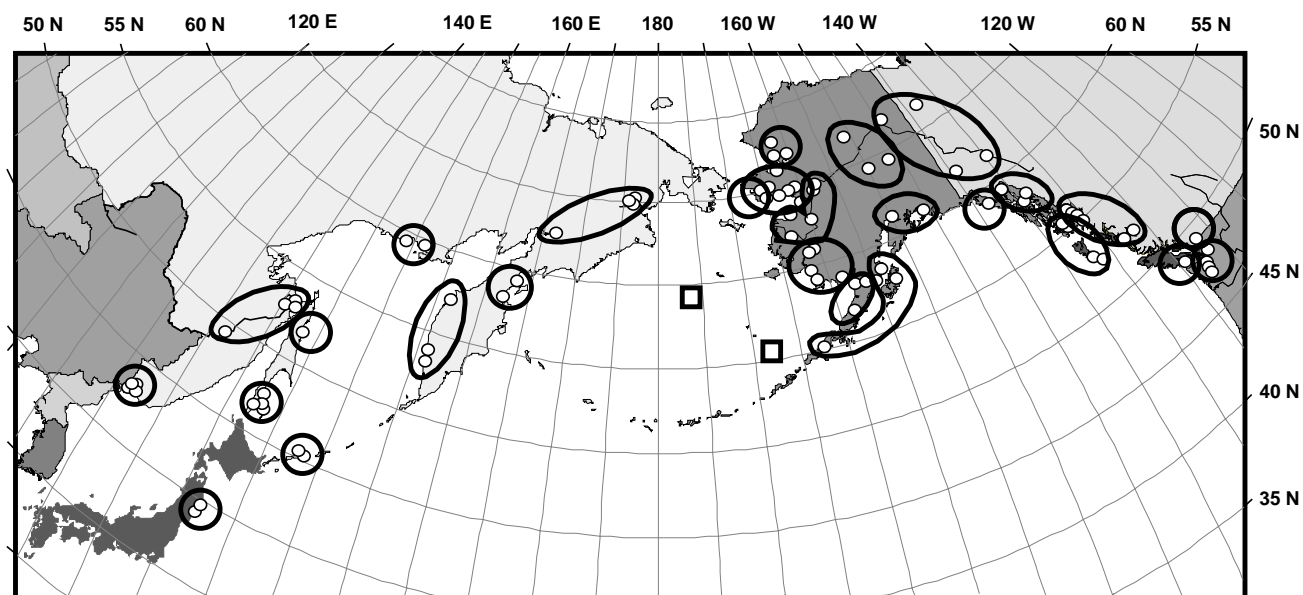


Figure 2. Populations and the groupings by geographical area used for analyses. The squares are locations of bycatch samples that we analyzed.

Experimental methods and results

The focus of this project was to develop and improve methods for discovering and detecting SNP variation followed by an evaluation of the SNPs that we developed. Consequently, our primary results are the methods. In this section, we briefly describe each of the methods that we developed and improved and then present results of the application of our efforts.

SNP discovery

DEco-TILLING – Most SNP assays are based on the detection of heteroduplexes. One of the properties of DNA is that the double stranded helix can be separated (denatured) and, under appropriate circumstances (buffer and temperature), complementary strands will reform the double helix. The restoration of double-stranded DNA is referred to as annealing. Any two complementary (carrying sequences of nucleotides that base-pair with each other) *strands* can anneal. The annealing process can tolerate a small amount of mispairing. If two sequences that differ at a single nucleotide site (a SNP) are denatured, the mostly complementary strands of the sequences can anneal, but the reformed double helix will have a mispaired “bubble” at the site where the two non-complementary *bases* occur. These are heteroduplexes and there are a variety of ways in which they can be detected physically or biochemically.

A technique called Eco-TILLING (Till et al. 2006) uses an extract of celery (*Apium graveolens*) (Yang et al. 2000), which has an enzyme called CEL1 that cleaves DNA only at single-stranded (e.g., mispaired) sites – our “bubbles”. Mispaired sites in heteroduplexes are partially digested by the celery extract and the resultant fragments are separated in a polyacrylamide gel and detected with a DNA sequencer. DEco-TILLING is a modification of Eco-TILLING. The important difference between the two methods is that Eco-TILLING depends on incomplete digests at random heteroduplex sites because the variety of fragments produced can be used both to identify and to survey mutations from the exact location of each mutation. Incomplete digests make it possible to estimate the location of the SNP in the sequence. DEco-TILLING, which we used for discovery, involves complete digests of both strands at all mispaired sites in a DNA heteroduplex. Estimates of location are less precise.

In DEco-TILLING, heteroduplexes are produced after PCR amplification by heating and cooling the amplified fragments. The annealed fragments are completely digested by the enzyme in the celery extract so that both strands of every PCR fragment are cleaved at the mispaired nucleotides. The fragments are then separated by size with electrophoresis on an agarose gel. Fragments smaller than the PCR target region indicate the presence of a SNP. If there is no mispairing (the PCR target region is homozygous), the fragment remains intact. The advantage of DEco-TILLING over Eco-TILLING for discovery is that neither a sequencer nor

polyacrylamide gels are required, the digest can be directly analyzed on a gel, and expensive fluorescent marker dyes are unnecessary.

Minimizing ascertainment bias – One of the problems with SNP discovery, which is often based on DNA sequences from a database or acquired *de novo*, is that it is impractical and expensive to sequence a large number of individuals. In addition, sequencing is generally limited to stretches of about 500 nucleotides. The restriction in numbers of individuals means that the geographic coverage of variation in the species is limited. Another consequence is that relatively infrequent variants may be discovered while abundant variants may remain undetected. These problems are two facets of ascertainment bias. By surveying pools of individuals from populations sampled throughout the species range, we were able to reduce ascertainment bias and choose newly detected SNPs that exhibited different levels of variation in different geographic regions.

Our goal is to identify SNPs that reflect geographic differences among populations, which means that we need both to discover the SNP variants and to evaluate their geographic variation. To discover informative SNPs that describe geographic variation, we conducted three rounds of DEco-TILLING. In the first round, we pooled DNA from five fish into each PCR reaction and tested eight pools from each population, which totaled 40 fish per population. We screened 12 populations (a total of 480 fish) for each locus that we evaluated. The populations represented the range of the species. Gels were scored as the number of the eight pools that were positive (included two different alleles) for each population (Figure 3). In order to interpret these results, we produced a family of curves that described the extent of allelic variation and differences among populations that could be demonstrated by eight pools of fish from a population (Figure 4). The curves are probability distributions that were determined by Monte-Carlo simulations, which simulated samples of an allele frequency pair (e.g., 0.95, 0.05) with replacement to obtain 10 alleles to represent a pool. Eight pools were generated to mimic the sampling design. Each iteration provided an estimate of the number of pools that had both types of alleles and produced heteroduplexes in the pool. By referencing these curves and evaluating 480 individuals, we substantially reduced ascertainment bias in identifying useful SNP sites.

For promising loci, we conducted the two more rounds of DEco-TILLING with DNA from single individuals. First, twenty individuals from pools that contained the variant were analyzed by a second round of DEco-TILLING. Homozygotes had no cleavage products. To

identify alternate homozygotes, we conducted a third round of DEco-TILLING in which DNA from one homozygous individual was added to PCR reactions of other homozygotes. Heteroduplexes observed in the third round of DEco-TILLING from mixtures of homozygotes identified alternate homozygotes. Finally, we sequenced two individuals – the common homozygote (**AA**) and the alternate homozygote (**aa**), which represented both alleles. This provided the information necessary for subsequent genotyping of large numbers of individuals. If an alternate homozygote could not be readily identified, we sequenced a heterozygote.

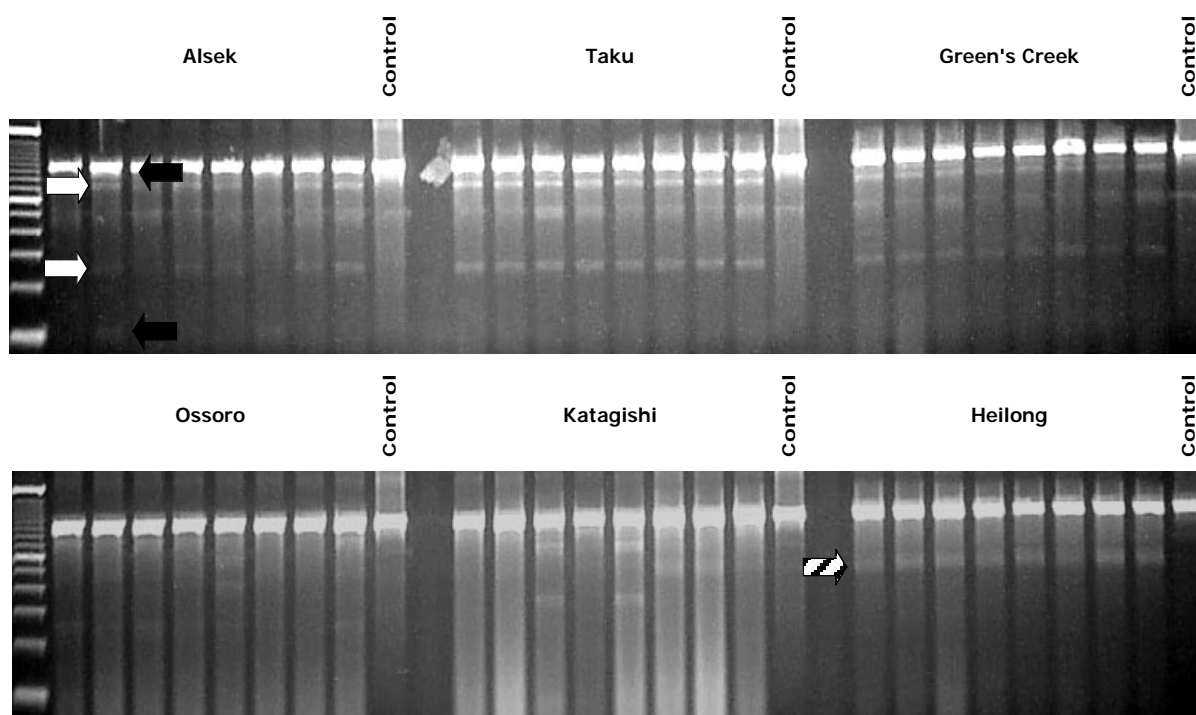


Figure 3. DEco-TILLING performed with DNA from six populations on a portion of the vitellogenin receptor locus. Each lane represents PCR products from five fish. There are eight lanes per population (40 fish). The final, ninth lane, is a control PCR product that was not subjected to DEco-TILLING. Three different SNP sites are represented by black arrows (site 1), a black and white hatched arrow (site 2), and white arrows (site 3). Note that the size of the two bands identified for each SNP, add to the size of the entire product. Site #2 is in the middle of the PCR product and, therefore, represents two bands of similar size.

Mitochondrial sites – We used previously published mitochondrial restriction site data (Churikov et al. 2001) that had been augmented by A.J. Gharrett and others (unpublished) in combination with the complete mitochondrial DNA sequence (K. Saitoh, Tohoku National Fisheries Research Institute, Shinjama, Shioyama 985-0001 Japan, personal communication) for chum salmon to identify the locations of key variable nucleotide sites in the mitochondrial

genome. The specific SNPs that that we developed were from five restriction sites that defined major branches of the mtDNA haplotype tree and that demonstrated divergence among populations from different geographic regions.

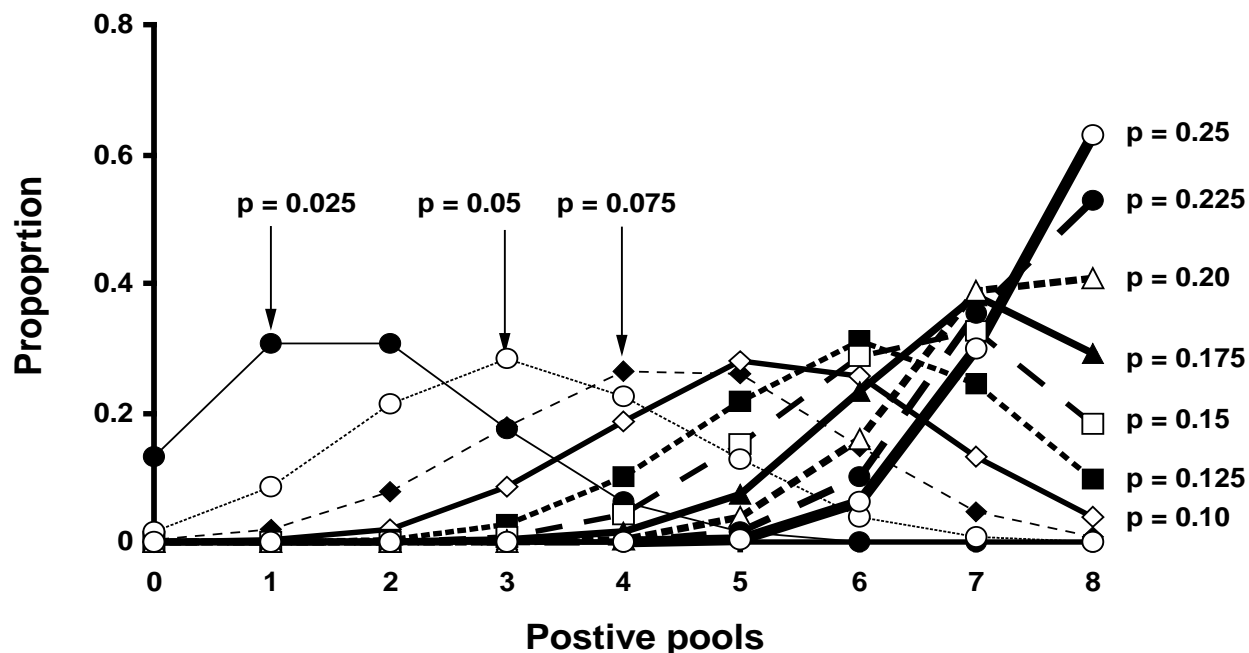


Figure 4. The distribution of the number of pools that have heteroduplexes for different underlying allele frequencies. Eight pools of five individuals (10 alleles) are tested. Heteroduplexes are observed in pools that have both allele types. Each curve describes the distribution for a different underlying less abundant allele frequency (say 0.025) in a population.

Nuclear sites – We focused our discovery efforts of nuclear loci in regions that were likely to be influenced by natural selection. Expression of nuclear genes is often modulated by sequences just upstream (5' untranslated regions) or just downstream (3' untranslated regions) of DNA sequence for the protein structure itself. In addition, genes in eukaryotes usually include sequences (introns) that are excised during processing of the RNA copy that carries the information to the cytoplasm where it is translated into proteins. The first intron is often involved in modulating gene expression. We focused our discovery efforts on the 3' and 5' untranslated regions, the first intron, and on genes that might show variation based on their function. Genes that are involved in regulating expression of other genes, such as an estrogen receptor protein, are also good candidates.

We surveyed 24 different PCR products that spanned a total of 28 kb (a 1 kb is 1000 nucleotides) and included populations that were sampled across the geographic range of chum salmon. We found 46 SNPs, five of which were discovered during sequencing and for which we had no information regarding their geographic distribution. Another 25 SNPs had little variation or no obvious differences in geographic distributions of variation. Sixteen of the SNPs appeared to be informative. Note that DEco-TILLING eliminated loci that had no, or low frequency, SNP sites at the first stage and eliminated the need to sequence numerous samples.

Eleven of the 16 informative SNP sites and a twelfth site that had been reported earlier by the Alaska Department of Fish and Game (ADFG)(Smith et al. 2005) were converted into genotyping assays. The remaining five sites are currently being developed for future surveys. These 12 nuclear SNPs represent 8 loci, which we labeled VT (Vasotocin I), IN (Insulin), RH (RH1OP), SP (Serpine), VR (Vitellogenin receptor), IS (Isotocin II), ER (Estrogen Receptor), and PL (Prolactin).

Choosing and improving the SNP assay

T_m shift—Following SNP discovery, the goal of analysis is to genotype each marker rapidly in as many individuals as possible. Because many individuals and SNPs are ordinarily analyzed in developing baselines and analyzing mixtures, the cost of analysis is critical. Current PCR-based methods such as the TaqmanTM are robust and easy to score, but are often not cost-effective. The T_m-shift assay developed by Roche Molecular Systems (Wang et al. 2005) is an inexpensive alternative. Instead of using probes that are engineered to recognize heteroduplexes of specific SNPs, the T_m-shift assay detects the difference in thermal stability between PCR products that differ for products from the different SNP alleles because the primers that recognize the different SNP sites are engineered to have different thermal stabilities. We developed the T_m-shift for analysis in our laboratory and have adapted Taqman assays to the T_m-shift assay for several SNPs.

LNA application – At times, the temperature stability differences that are detected by the T_m-shift assay can be subtle, and it can be challenging and time consuming to develop an assay. The conversion of sequence information to a functional assay was a serious impediment to our SNP development. The conformation of a specific category of nucleotide analog, called locked nucleic acids (LNAs), is rigidly constrained. The result of the constraint is that base pairing

between an LNA and its complementary nucleotide is very stable and substantially increases the temperature stability. By incorporating an LNA in the PCR primer at the SNP site, the assay improved markedly, and the time required to develop the assay was substantially reduced.

Determining linkage phase

Another challenge that is associated with discovery and analysis of SNP variation is linkage. If two sites are detected in a region scrutinized for SNPs, the sites are likely to be physically linked. Complex alleles (called haplotypes) that result from linked SNPs can provide more information than single, independently inherited SNPs (Davidson 2000, Drysdale et al. 2000, Rannala and Bertorelle 2001, Leblois and Slatkin 2007). For stock identification applications, the linkage phase must be determined empirically for each multiple heterozygote from the sample mixture. We developed a rapid method to resolve linkage phases in linked SNPs that improves on an older method (Eitan and Kashi 2002).

Our method is based on a modification of our T_m -shift SNP assay and a previously described method (Eitan and Kashi 2002). Primer pairs are designed so that the 3' end (the end to which nucleotides are added in the PCR reaction) of the PCR primers terminate at the first and second SNP sites in the haplotype, respectively (Figure 5). We refer to the two alleles at SNP site 1 as **A/a** and two alleles at SNP site 2 as **B/b**. A PCR product would only be generated if nucleotides at both SNP sites complemented the primers. In order to develop each assay, we designed four primers that could be run in four possible combinations (**A-B**, **a-B**, **A-b**, and **a-b**). Only one combination of primers can amplify a specific haplotype. In practice, one needs to run only two reactions on double heterozygotes, one that uses one of the primer pairs in either phase **A-b** or **a-B** and a second pair that uses primers in the phase **a-b** or **A-B**, in order to correctly identify the linkage phase of the double heterozygote (**AaBb**).

Assessment of markers

We used data from 20 SNP sites to conduct a preliminary evaluation of the markers with standard population genetics analyses. Those SNPs included eight mitochondrial markers and eight nuclear loci; the mitochondrion represents a single complex locus. Three of the nuclear loci included multiple SNPs that were scored as haplotypes. We surveyed variation at those SNP sites in 32 populations of chum salmon that represented 17 geographic regions (Figure 6) and

included about 2,200 fish. To ensure that we were dealing with “well behaved” data and that there were no anomalies in the data, we tested all populations and all composite genotype frequencies at nuclear loci for Hardy-Weinberg equilibrium and all nuclear locus pairs for linkage equilibrium. We observed no departures from Hardy-Weinberg equilibrium or linkage equilibrium.

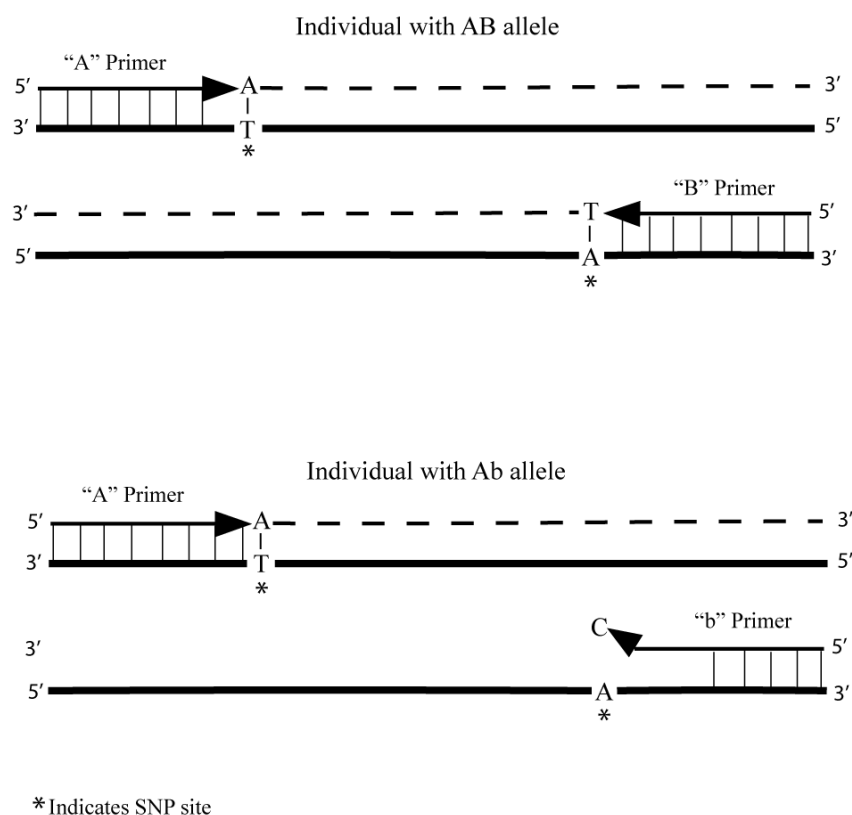


Figure 5. Representation of experimental determination of the linkage phase in loci with multiple SNP sites. Both primers terminate over a variable SNP site (shown with an asterisk). Only PCR reactions in which both primers match the underlying variant will result in exponential amplification. In the lower example, the “B” primer does not match the “b” allele and will therefore not allow a PCR product to be formed.

We partitioned the genetic variation obtained from the SNP analyses into components that quantify the relative proportion of the total variation that can be attributed to average departures from random mating within a population (F_{IS}), among populations within a geographic area (F_{SC}), and among geographic areas (F_{CT}) (Arlequin 3.0; Excoffier and Schneider 2005). The statistic F_{ST} is the total variation among populations.

The F_{SC} value is low, which indicates that there is relatively little variability among populations within geographic areas, although it is significant in most groups of populations. The F_{CT} (and F_{ST}) values show strong divergence and every locus is significant, which indicates that these loci describe divergence, which should be useful loci in discriminating among geographic regions. In addition, all of the loci that contain multiple (linked) SNPs have significant F_{SC} s, which supports the idea that complex SNP loci may be more informative than biallelic loci.

Table 1. Locus by locus F -statistics and estimates averaged over all loci with an AMOVA approach. Diploid and haploid data were combined from the variance values from the AMOVA tables as shown in Weir and Cockerham (1984).

Locus	F_{SC}		F_{CT}		F_{ST}		F_{IS}	
VT	0.006	*	0.131	***	0.136	***	0.027	
IN	0.006	*	0.080	***	0.086	***	0.039	
SP	0.008	*	0.082	***	0.089	***	-0.002	
RH	0.002		0.113	***	0.115	***	-0.052	
VR	0.040	***	0.145	***	0.179	***	0.019	
IS	0.038	***	0.045	***	0.081	***	0.013	
ER	0.007		0.444	***	0.449	***	0.003	
PL	0.015	**	0.087	***	0.101	***	0.053	*
MT	0.040	***	0.324	***	0.351	***	N/A	
AVG	0.023		0.171		0.190		0.01428	

* p<0.05 ** p<0.01 *** p<0.005

Simulations

Recently, we conducted simulations to determine the ability of the genetic data to resolve origins of mixtures. We based the simulations on genetic data from 88 populations for which we had both SNP (10 loci) and microsatellite data (109 loci – one is the haploid mtDNA). The populations spanned from the Puget Sound in the Pacific Northwest of North America to the Russian Far East and Honshu Island of Japan. The microsatellite data resulted primarily from our efforts, but data from a few populations were obtained by the Department of Fisheries and Oceans Canada (DFO). We assured the compatibility between DFO's and our data by stringent comparisons of several populations for which we both had data. We conducted simulations with three sets of data: (1) SNP data alone; (2) microsatellite data alone; and the combination of SNP and microsatellite data.

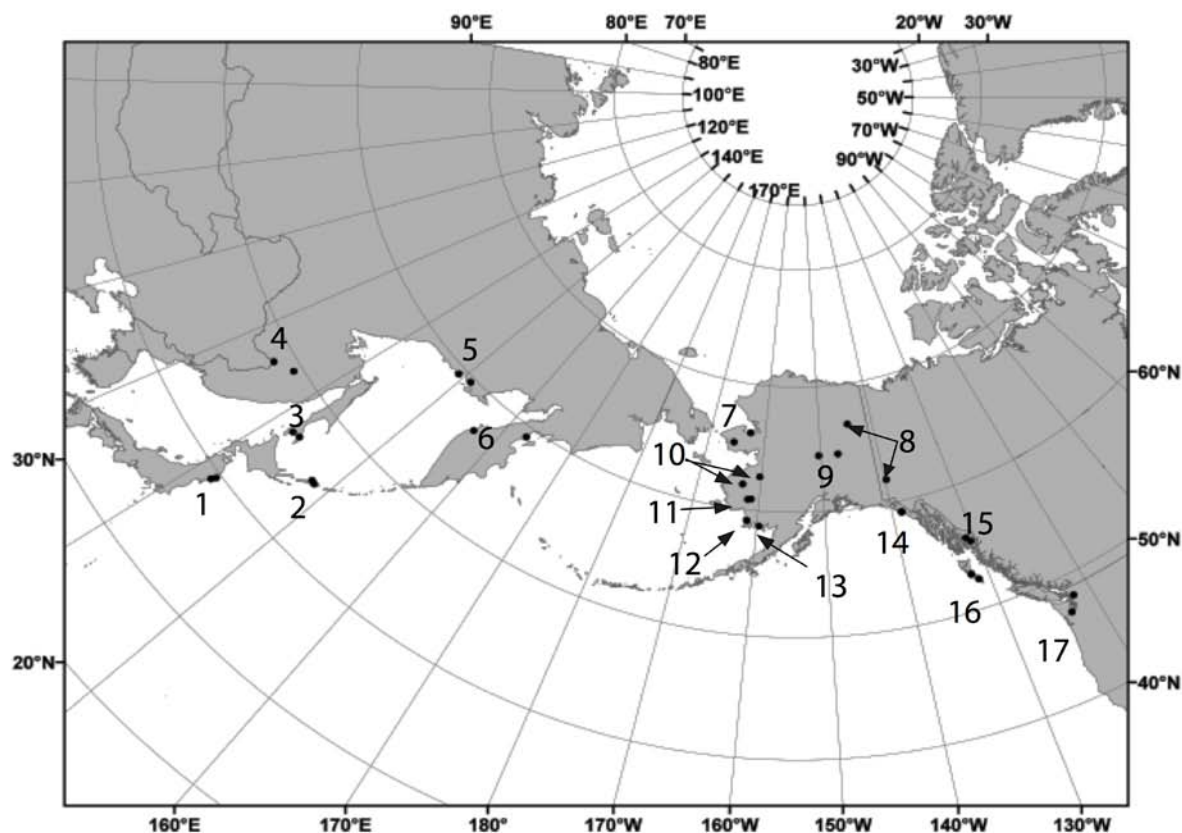


Figure 6. Locations of the 32 populations used for the statistical analyses in this study (black dots). Numbers correspond to the groupings we used for the hierarchical AMOVA analysis that is shown in Table 5.

How were the simulations conducted? The baseline was divided hierarchically into three regions, each of which had geographic subregions: Asia (two subregions), western Alaska north of the Alaska Peninsula (three subregions), and the populations that extended from the Alaskan Peninsula to the Pacific Northwest and referred to as Southeast (three subregions). Populations within each region were grouped into geographic areas, primarily by their geographic locations. The Asian region had ten areas, Western Alaska had eight areas, and the Southeast region had nine areas.

The simulations were conducted by *in silico* sampling a total of 200 individuals from populations within a geographic area (or subregion or region) and using the program SPAM (Debevec et al. 2000) to estimate the origins of that sample. Correct reassignment back to the regions, subregions, or areas (Table 2) from which they were drawn would indicate that the baseline carries sufficient information for reliable application to mixed-stock fisheries. Often the criterion of 90% (or even 95%) correct assignment is acceptable.

Table 2. Hierarchy of regions, subregions, and areas within subregions used in simulations to evaluate the ability of our genetic markers to discriminate among geographic units.

Region	Subregion	Geographical Area	Populations		
Asia	western Asia	Honshu	2		
		Primoriye	4		
		Amur River	4		
		southern Sakhalin Island	4		
		Tym River	1		
	eastern Asia	Kurile Islands	2		
		northern Okhotsk	2		
		western Kamchatka	3		
		eastern Kamchatka	3		
		Anadyr River	3		
Western Alaska	coastal	early Seward Peninsula	1		
		late Seward Peninsula	3		
		Norton Sound	8		
		lower Yukon	5		
		Kuskokwin River	3		
	middle/upper Yukon	middle Yukon	4		
		upper Yukon	5		
		Southeast	Bristol Bay	Bristol Bay	3
				southcentral Alaska	Alaska Peninsula
			Cook Inlet and Prince William Sound		2
Alsek River	2				
SE Alaska/N British Columbia	northern SE Alaska		3		
	mainland SE Alaska and British Columbia	7			
	island SE Alaska and British Columbia	4			
S British Columbia/Puger Sound	western Vancouver Island	1			
	Fraser River	1			
	Puget Sound	4			
			88		

SNP simulation results – SNPs performed very well in determining the regional origin of samples; all regional assignments exceeded 94%; and assignment to all subregions exceeded 85% for all subregions except Bristol Bay (83%). Assignment to areas exceeded 80% except for three geographic groups of areas: (1) west Kamchatka (78%) and east Kamchatka (75%); (2) Norton Sound (69%), lower Yukon (56%), middle Yukon (78%), and Kuskokwim (59%); and (3) northern Southeast Alaska (79%), and the mainland (72%) and island (76%) populations of Southeast Alaska and British Columbia. Most of the misassigned fish within each of these groups were attributed to other areas within the group or to adjacent areas. Notably, some Anadyr fish (4.4%) were misassigned to Seward Peninsula and Norton Sound populations. Also, it is apparent that the coastal western Alaska populations (Seward Peninsula, Norton Sound,

Lower Yukon, and Kuskokwim populations are quite similar and incompletely resolved with this suite of SNP loci.

Microsatellite simulation results – In general, the nine microsatellite loci alone performed better than the SNP loci alone. Regional assignments to Asian populations was less accurate (90%) than observed for SNPs alone (94%), but assignments to western Alaska and to southern populations exceeded 95%. All of the assignments to subregions exceeded 90% except for eastern Asian (85%) populations. Many of the assignments to areas within subregions exceeded 90%. The ones that did not exceed 85% were from the same groups that had poorer assignment levels for the SNP loci: (1) northern Okhostk (79%), western Kamchatka (83%), and eastern Kamchatka (77%); (2) Seward Peninsula (86%), Lower Yukon (66%), and Kuskokwim (71%); and (3) mainland Southeast Alaskan populations (80%). Misassignments were either to adjacent areas or for group (2) among the coastal western Alaskan populations.

Combined SNP and microsatellite simulation results – Analyses that used data from both sources of information were strongest. Addition of the SNP data to the analysis improved recognition of subregion assignments to geographic area (all exceeded 90%) an average of 2.3% over those observed for microsatellites alone, and the confidence intervals for the estimates also shrank. Assignments to area improved by an average of 2.5%. Assignments to most areas exceeded 85% and many exceeded 90% (Appendix I). The exceptions were the lower Yukon and Kuskokwim. The largest improvements were for the western Asian and Southeast Alaskan populations.

Clearly 10 SNP loci are insufficient to obtain the fine resolution that is desirable, but the results from those SNP loci alone and the improvement they contributed when added to the microsatellite data is encouraging.

Estimating mixture compositions

Compositions of two samples of immature chum salmon were estimated. Both samples were collected in August 2005, one from the pelagic trawl fishery at 55° 20' N 167° 01' W north of the Unalaska Island; and the other west and a bit south of St. Matthew Island at about 59° 30' N 175° 59' W (Figure 2). The program CBayes analysis uses Bayesian methods to estimate contributions, which iteratively takes into account the estimates it makes to adjust baseline probabilities. The SPAM program does not make such adjustments. The estimates from CBayes

are probably more accurate, but the results were similar, especially at the higher levels of hierarchy.

Bootstrap analyses estimated that the sample of the southern sample of immature chum salmon originated about equally from Asia, western Alaska and the Gulf of Alaska and Pacific Northwest populations (Appendix II). A surprisingly large proportion originated from the Pacific Northwest (~ 22%) and Southeast Alaska and northern British Columbia (~ 15%). Populations from both the western Asian range (Honshu, Primorye, and Sakhalin Island) and Kamchatka also were present. Bristol Bay chum salmon were the most abundant western Alaskan fish and very few middle and upper Yukon River fish were present.

The composition of the more northerly sample was strikingly different. Nearly 90% of the fish were of Asian origin, particularly Japan; western Asian populations accounted for more than 50 % and eastern populations more than 35%. The relatively few fish of North American origin were scattered among the geographic areas, and there estimates were generally not significantly different from zero.

How many SNP loci are needed?

Our results show that SNP data from nine nuclear loci and the mitochondrial genome only weakly resolve contributions to mixtures from several of the geographic areas that we examined. In addition, it would be convenient for analysis and automation if we did not have to use microsatellite loci. In order to examine the improvement that could be gained by increasing the number of nuclear SNP loci, we conducted simulations that used replicates of the nuclear loci. The idea is that increasing the number of loci by, say, two-, four-fold, or more will increase the possible resolution if the loci reflect divergence among populations or geographical areas. The nine loci that we evaluated in this report had some success in resolving differences. Moreover, they are examples of the divergence that exists when care is taken in selecting loci. By replication of the baseline information, we are simulating a situation in which our analysis has twice or four times as many *independent* loci that reflect the divergence.

The results of our analysis suggest that if we had an additional 27 nuclear SNP loci in addition to the mitochondrial data, nearly every area would approach 90% assignment or better. The misassignments that occur are generally to nearby areas (Appendix II).

Conclusions

The application of SNPs to specific questions requires efficient discovery of informative SNPs, rapid (and preferably) inexpensive genotyping, and often resolution of linked SNPs. Standard discovery methods do not address the problem of ascertainment bias and most current next-generation molecular technologies such as microarray- and pyro-sequencing (Marcel et al. 2007; Wikman et al. 2000) were developed for use in human diagnostics and drug discovery; and although these methods can exponentially increase the amount of genetic information available to a researcher, they are often impractical for population genetic studies, which need data derived from parallel rather than serial methods. In other words, these methods do not address the need for obtaining information from multiple loci in thousands of individuals, rather, they produce information from thousands of loci in a few (or one) individual, which again introduces ascertainment bias into the discovery process.

This study described a series of methods to apply SNPs in almost any laboratory. We began with discovery of informative SNPs. We applied our previously published method (DECO-TILLING) here to survey PCR amplicons in chum salmon that were produced with primers designed from chum salmon sequence and from sequence that was available from other species. We continued with a description of genotyping assays. As with SNP discovery, technological advances of methods to genotype SNPs in individuals have also increased the amount of available data; but, as with the SNP discovery methods, the costs to perform these assays can be quite high. We applied the T_m -shift method (Wang et al. 2005) to perform thousands of assays inexpensively. In addition, we described an improvement to the assay with the addition of a LNA to the 3' end of the SNP-specific primers. The result is an inexpensive assay that can be quickly adapted to genotyping SNPs.

We concluded our improvements of methodology with a method for resolving the linkage phase of linked SNPs, which can improve detection of divergence among populations and which are often discarded during SNP discovery. For example, the insulin, the isotocin II, and the vitellogenin receptor loci included two or more linked SNPs. The challenge generated by linked SNP sites is that haplotypes on single chromosomes cannot be scored in double (or multiple) heterozygotes with the primary SNP genotyping assay.

A genetic analysis of the loci that we developed in this study demonstrated that our new SNPs are informative. We analyzed samples from 17 geographic regions in this study, and

observed divergence among geographic regions that differ significantly from zero at all loci tested. In addition, the divergence among populations within geographic areas was significant ($P < 0.005$) only at loci that had multiple SNP sites.

In a process that begins with discovery, the application of a select set of SNPs may reduce the number of markers that is necessary to resolve population mixtures. Consequently, the costs of development and application of these methods can be reduced and analysis can be conducted in smaller labs, or possibly the field. Ultimately, improvements in this analysis would bring resource management agencies closer to their goal of managing populations in real time. These results indicate that our strategy identifies informative SNPs because we can demonstrate small, but significant divergence among populations as well as among groups.

Cost and time analysis

We conducted a cost and time analysis for three combinations of Eco-TILLING and DEco-TILLING for discovering and genotyping SNPs. The times and costs were based on three scenarios: (1) DEco-TILLING to discover SNPs and the T_m -shift primer assay to survey them, (2) Eco-TILLING to discover SNPs and the T_m -shift primer assay to survey them, and (3) Eco-TILLING both to discover SNPs and to survey them in order to make reasonable comparisons of the methods that might be used and because the published studies that use Eco-TILLING generally involve simultaneous screening and surveying of loci.

The calculations represent a small, but typical genetic study, which screens 480 individuals to identify SNPs as was described in this study, followed by a survey of 1,000 individuals at 6 different SNPs. The cost of the reagents for both methods is relatively small although it should be noted that the cost of Eco-TILLING is more than double that of DEco-TILLING when using the T_m -shift method for genotyping (Table 3). The major savings from the DEco-TILLING applications is time. It should also be noted that as many as 50 to 100 *unlinked* informative SNP sites may be necessary for accurate analysis, which suggests DEco-TILLING can provide a significant cost and time savings as compared to Eco-TILLING.

Table 3. Cost and efficiency analysis of DEco-TILLING and Eco-TILLING. Analysis is based on discovering SNPs using 480 samples as described in this study and subsequently screening 6 SNPs on 1,000 individuals. The cost of the sequence specific primers is based on the use of Universal dye-labeled primers (Till et al. 2006). The approaches are 1. DEco-TILLING for SNP discovery and Tm-shift for genotyping; 2. Eco-TILLING for SNP discovery and genotyping ; and 3. Eco-TILLING for SNP discovery and Tm-shift for genotyping .

Task	1	2	3
Reagent Costs			
Sequence specific labeled primers*	N/A	\$200.00	\$200.00
Sequence specific unlabeled primers	\$100.00	\$180.00	\$180.00
SNP site specific primers for genotyping	\$205.00	N/A	\$205.00
96 well Sephadex columns	N/A	\$180.00	\$180.00
Sequencing SNP sites	\$60.00	N/A	\$60.00
Total Cost	\$365.00	\$560.00	\$825.00
SNP Discovery Time			
Eco or DEco-TILLING pools for 6 DNA Sequences			
CEL1 Cleavage	6	1.5	1.5
Cleavage product purification	0	5	5
Volume reduction of purified product	0	6	6
Gel run time	9	27	27
Eco or DEco-TILLING individuals			
CEL1 Cleavage	6	1.5	1.5
Cleavage product purification	0	5	5
Volume reduction of purified product	0	6	6
Gel run time	9	27	27
Eco or DEco-TILLING individuals with homozygote			
CEL1 Cleavage	6	1.5	1.5
Cleavage product purification	0	5	5
Volume reduction of purified product	0	6	6
Gel run time	9	27	27
SNP Sequencing Time			
Sequencing forward and reverse strands	12	N/A	12
SNP Surveying Time			
Genotyping Assay For 1000 individuals	30	200	30
Total Hours	87	318.5	160.5

*Cost is based on using Universal primers for all 6 DNA sequences surveyed based on Till et al. 2006. If individual labeled primers are used, the cost increases by \$1,000.

Acknowledgements

Part of this work represents research conducted by M. Garvin for his M.S. thesis at the University of Alaska Fairbanks. The work was supported by the Pollock Conservation Cooperative Research Center (PCCRC) at the University of Alaska Fairbanks, the Bering Sea Fisherman's Association (BSFA), and the Experimental Program to Stimulate Competitive Research (EPSCoR) at the University of Alaska Fairbanks. Much of this work was done in cooperation with the NOAA Fisheries Auke Bay Laboratory. Samples were generously provided by the Russian Academy of Sciences Far East Branch and Institute of General Genetics, Kawerak, Inc., the Alaska Department of Fish and Game, the US Fish and Wildlife Service, and Department of Fisheries and Oceans Canada.

References

- Ackley, D. 1997. Bycatch patterns in the Bering Sea: Templates for area closures. Symposium on the Consequences and Management of Fisheries Bycatch. pp. 47-51.
- Churikov, D., M. Matsuoka, X. Luan, A.K. Gray, V.A. Brykov, and A.J. Gharrett. 2001. Assessment of concordance among genealogical reconstructions from various mtDNA segments in three species of Pacific salmon (genus *Oncorhynchus*). *Molecular Ecology* 10:2329-2339.
- Davidson, S. 2000. Research suggests importance of haplotypes over SNPs. *Nature Biotechnology* 18: 1134-1135.
- 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.
- Drysdale, C.M., D.W. McGraw, C.B. Stack, et al. 2000. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proceedings of the National Academy of Sciences of the USA* 97:10483-10488.
- Eitan, Y, and Y. Kashi. 2002. Direct micro-haplotyping by multiple double PCR amplifications of specific alleles (MD-PASA). *Nucleic Acids Research* 30:1-8.
- Excoffier, L.L., and S. Schneider. 2005 Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
- Flannery, B.G., J.K. Wenburg, and A.J. Gharrett. 2007a. Evolution of mitochondrial DNA variation within and among Yukon River chum salmon populations. *Transactions of the American Fisheries Society* 136:902-910.
- Flannery, B.G., J.K. Wenburg, and A.J. Gharrett. 2007b. Variation of amplified fragment length polymorphisms (AFLP) in Yukon River chum salmon, *Oncorhynchus keta*: Population structure and application to mixed-stock analysis. *Transactions of the American Fisheries Society* 136:911-925.

- Garvin, M., and A. Gharrett 2007. DEco-TILLING: An inexpensive method for SNP discovery that reduces ascertainment bias. *Molecular Ecology Notes* 7:735-746.
- Leblois, R., and M. Slatkin. 2007. Estimating the number of founder lineages from haplotypes of closely linked SNPs. *Molecular Ecology* 16:2237-2245.
- Miller, K.M., and R.E. Withler. 1996. Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics* 43:337-351.
- Rannala, B., and G. Bertorelle. 2001 Using linked markers to infer the age of a mutation. *Human Genetics* 18: 87-100
- Sato, S., H. Kojima, H. and 13 coauthors. 2004. Genetic population structure of chum salmon in the Pacific Rim inferred from mitochondrial DNA sequence variation. *Environmental Biology of Fishes* 69:37-50.
- Smith, C.T., J. Baker , L. Par, et al. 2005. Characterization of 13 single nucleotide polymorphism markers for chum salmon. *Molecular Ecology Notes* 5:259-262.
- Till, B., T. Zerr, E. Bowers, et al. 2006. High-throughput discovery of rare human nucleotide polymorphisms by Ecotilling. *Nucleic Acids Research* 34:1-12.
- Till, B.J., L. Comai, and S. Henikoff. 2007. Tilling and Ecotilling for Crop Improvement. In: *Genomics-Assisted Crop Improvement*, pp. 333-349. Springer Netherlands.
- Yang, B., X. Wen, N.S. Kosali, et al. 2000. Purification, cloning and characterization of the CEL1 nuclease. *Biochemistry* 39:3533-3541.

Products from the study

- Garvin, M.R., and A.J Gharrett. 2007. An inexpensive tool box for discovering, analyzing, and determining the linkage phase of SNPs: Application to variation likely to be subject to natural selection. **Poster** presented at annual meeting of the Society for Molecular Biology and Evolution 24-28 June, Halifax, Nova Scotia, Canada.
- Garvin, M.R., S. Fuller, R. Riley, S. Hall, R.W. Wilmot, and A.J. Gharrett. 2007. Stock identification is essential for determining the cause of salmon fluctuation. **Poster** presented at Symposium on the Sustainability of the Arctic-Yukon-Kuskokwim Salmon Fisheries, 6-9 February, Anchorage, Alaska
- Garvin, M.R. 2008. Informative SNP discovery and cost effective screening in any species using Chum salmon as an example. **Seminar** presented to Alaska Department of Fish and Game, 25 January 25th, Anchorage, Alaska.
- Garvin, M.R., and A.J Gharrett. 2007. DEco-TILLING: an inexpensive method for single nucleotide polymorphism discovery that reduces ascertainment bias. *Molecular Ecology Notes* 7: 735-746.
- Garvin, M.R., K. Saitoh, and A.J. Gharrett. Submitted. Application of SNP markers to population genetics studies: Discovery, genotyping, and linkage phase resolution. *Molecular Ecology Resources*.

Appendix I. Estimates of self assignment for geographic areas (Table 2) from the results of 100% simulations based on SNP and microsatellite data combined.

Area	Japn	Prim	Amur	SSak	Tym	Kurl	NOkh	WKam	EKam	Anad	Agia	SewP	Nort
Honshu	0.944	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
imoriye	0.014	0.948	0.002	0.002	0.002	0.007	0.002	0.002	0.001	0.001	0.000	0.000	0.000
Amur	0.001	0.002	0.911	0.002	0.003	0.001	0.005	0.002	0.003	0.002	0.000	0.000	0.000
S Sakhalin	0.002	0.002	0.003	0.917	0.003	0.006	0.003	0.003	0.003	0.001	0.000	0.000	0.000
Tym	0.005	0.004	0.011	0.006	0.929	0.006	0.010	0.016	0.014	0.003	0.000	0.000	0.000
Kutiles	0.003	0.003	0.000	0.002	0.001	0.939	0.000	0.001	0.000	0.000	0.000	0.000	0.000
N Okhotsk	0.001	0.002	0.004	0.003	0.003	0.001	0.850	0.006	0.006	0.005	0.000	0.000	0.000
W Kamchatka	0.004	0.005	0.013	0.010	0.019	0.007	0.041	0.885	0.089	0.011	0.001	0.001	0.001
E Kamchatka	0.002	0.003	0.006	0.006	0.007	0.003	0.017	0.021	0.823	0.004	0.001	0.001	0.001
Anadyr	0.001	0.002	0.002	0.003	0.004	0.002	0.010	0.008	0.007	0.903	0.002	0.002	0.002
E Seward P	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.001	0.907	0.006	0.003
L Seward P	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.001	0.002	0.012	0.874	0.007
Norton S	0.007	0.007	0.011	0.016	0.009	0.010	0.019	0.012	0.015	0.031	0.048	0.082	0.922
L Yukon	0.001	0.002	0.003	0.005	0.002	0.003	0.006	0.003	0.004	0.010	0.010	0.012	0.029
M Yukon	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.002	0.002	0.003	0.004
U Yukon	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002
Kuskokwim	0.002	0.002	0.003	0.004	0.003	0.002	0.008	0.004	0.006	0.013	0.013	0.012	0.025
Bristol Bay	0.001	0.001	0.002	0.004	0.003	0.001	0.004	0.003	0.004	0.004	0.002	0.002	0.002
Ak Pen	0.001	0.002	0.008	0.004	0.002	0.002	0.008	0.008	0.005	0.003	0.000	0.001	0.001
SC Alaska	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000
Alsek	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N SE ALaska	0.001	0.001	0.002	0.002	0.001	0.001	0.003	0.005	0.003	0.001	0.000	0.000	0.000
SE AK mainlan	0.001	0.002	0.006	0.004	0.003	0.002	0.007	0.009	0.007	0.002	0.000	0.001	0.000
SE Alaska isle:	0.000	0.001	0.001	0.001	0.001	0.000	0.001	0.002	0.002	0.000	0.000	0.000	0.000
S Cancouver I	0.000	0.000	0.001	0.001	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.000	0.000
Fraser R	0.001	0.003	0.004	0.003	0.002	0.001	0.002	0.004	0.004	0.001	0.000	0.000	0.000
Puget S	0.001	0.001	0.003	0.002	0.001	0.001	0.002	0.003	0.003	0.001	0.000	0.000	0.000

Appendix I (continued).

Area	LYuk	MYuk	UYuk	Kusk	BBay	AkPen	SCAk	Alsk	NSEA	Main	Isle	VCI	Fraser	PugS
Honshu	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
imoriye	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Amur	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
S Sakhalin	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.001	0.000	0.001	0.001	0.000	0.000	0.001
Tym	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001
Kutiles	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N Okhotsk	0.001	0.000	0.000	0.000	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.001
W Kamchatka	0.001	0.000	0.000	0.001	0.002	0.005	0.003	0.003	0.003	0.007	0.005	0.002	0.001	0.003
E Kamchatka	0.001	0.000	0.000	0.001	0.002	0.002	0.003	0.001	0.001	0.002	0.001	0.001	0.001	0.001
Anadyr	0.003	0.001	0.001	0.003	0.002	0.002	0.004	0.001	0.001	0.001	0.001	0.000	0.001	0.001
E Seward P	0.002	0.001	0.001	0.002	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
L Seward P	0.004	0.002	0.002	0.004	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Norton S	0.189	0.024	0.016	0.159	0.024	0.006	0.015	0.003	0.002	0.003	0.002	0.001	0.002	0.003
L Yukon	0.691	0.016	0.009	0.073	0.011	0.002	0.004	0.001	0.001	0.001	0.001	0.000	0.000	0.001
M Yukon	0.011	0.924	0.025	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U Yukon	0.003	0.019	0.939	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Kuskokwim	0.088	0.012	0.007	0.737	0.019	0.003	0.008	0.001	0.001	0.002	0.001	0.000	0.001	0.002
Bristol Bay	0.005	0.001	0.001	0.009	0.925	0.004	0.005	0.002	0.001	0.002	0.001	0.000	0.001	0.001
Ak Pen	0.001	0.000	0.000	0.001	0.003	0.933	0.010	0.010	0.010	0.015	0.007	0.002	0.003	0.005
SC Alaska	0.000	0.000	0.000	0.000	0.001	0.004	0.909	0.002	0.003	0.009	0.003	0.000	0.001	0.001
Alsek	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.932	0.000	0.001	0.000	0.000	0.000	0.000
N SE ALaska	0.000	0.000	0.000	0.000	0.001	0.006	0.005	0.007	0.917	0.032	0.016	0.002	0.003	0.005
SE AK mainlan	0.000	0.000	0.000	0.001	0.003	0.015	0.016	0.022	0.036	0.855	0.046	0.007	0.011	0.012
SE Alaska isle:	0.000	0.000	0.000	0.000	0.000	0.004	0.004	0.005	0.012	0.028	0.893	0.003	0.002	0.003
S Cancouver I	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002	0.005	0.004	0.967	0.005	0.006
Fraser R	0.000	0.000	0.000	0.000	0.001	0.005	0.004	0.004	0.005	0.024	0.010	0.006	0.943	0.038
Puget S	0.000	0.000	0.000	0.000	0.001	0.003	0.002	0.002	0.003	0.009	0.005	0.006	0.023	0.914

Appendix II. Estimates of contributions by region, subregion, and area (Table 2) for two bycatch samples of immature chum salmon. Analyses bootstrapped the baselines (SPAM; Debevec et al. 2000).

Region	North central Bering Sea			Aleutian Island		
	CBayes (no mtDNA)		SPAM	CBayes (no mtDNA)		SPAM
	Mean	Mean/SD	Point est.	Mean	Mean/SD	Point est.
Asia	89.00	31.90	84.46	32.49	8.07	31.29
W Alaska	5.66	2.62	9.03	30.27	7.99	33.48
Gulf of Alaska	5.34	2.88	6.51	37.24	9.52	35.23
Subregion						
W Asia	51.89	12.76	47.07	17.07	5.34	9.47
E Asia	37.11	8.65	37.39	15.42	4.49	21.82
W Alaskan coast	1.83	1.05	5.42	14.65	4.51	21.36
upper Yukon	3.70	2.33	3.32	3.04	1.86	2.52
Bristol Bay	0.13	0.29	0.29	12.58	4.39	9.60
SC Alaska	0.12	0.38	0.95	0.63	0.57	0.37
SE Alaska/N BC	2.80	1.80	2.57	14.61	4.36	17.08
S BC/Puget Sound	2.43	1.69	2.99	22.01	5.68	17.78
Area						
Honshu	40.38	8.38	17.59	5.20	2.76	2.73
Primorye	3.43	1.63	10.70	5.59	2.53	3.77
Amur	0.04	0.28	0.00	1.03	0.76	0.94
S Sakhalin	0.66	0.94	3.60	3.11	1.85	0.28
Tym	0.12	0.20	10.34	0.02	0.13	0.00
Kuriles	0.54	0.27	4.83	0.22	0.27	1.73
N Ohhotsk	11.08	2.34	4.03	1.93	1.35	1.09
W Kamchatka	14.73	2.96	18.27	9.16	3.29	14.34
E Kamchatka	13.37	2.99	10.58	0.30	0.34	2.60
Anadyr	4.59	1.74	4.51	5.91	2.57	3.80
Agiapuk	0.01	0.15	0.00	0.02	0.14	0.00
Seward Peninsula	0.05	0.24	0.00	0.49	0.35	0.32
Norton Sound	0.84	0.68	2.30	5.35	1.45	7.65
lower Yukon	0.41	0.43	1.36	5.42	1.27	8.39
middle Yukon	3.59	2.28	2.81	2.62	1.41	0.91
upper Yukon	0.11	0.32	0.51	0.42	0.47	1.61
Kuskokwim	0.55	0.46	1.77	3.37	1.12	5.00
Bristol Bay	0.13	0.29	0.29	12.58	4.39	9.60
Alaska Peninsula	0.09	0.31	0.95	0.59	0.54	0.37
SC Alaska	0.02	0.20	0.00	0.03	0.19	0.00
Alsek	0.02	0.21	0.00	0.02	0.20	0.00
N Southeast Alaska	1.76	1.34	1.89	3.68	1.71	4.92
Southeastern island	1.00	0.71	0.68	4.10	1.58	8.47
Southeastern mainland	0.05	0.27	0.00	6.82	2.47	3.69
Vancouver Island	0.01	0.15	0.00	2.15	1.16	3.94
Fraser River	1.07	0.79	2.15	3.69	1.15	6.59
Puget Sound	1.35	1.34	0.84	16.16	4.21	7.25