

Comparisons of Maximum Likelihood and Bayesian Methods for Mixed-Stock Analysis

Blair G. Flannery, Penny A. Crane, & John K. Wenburg
 Conservation Genetics Laboratory
 U.S. Fish and Wildlife Service
 1011 E. Tudor Rd. Anchorage, AK 99503



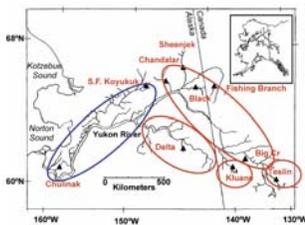
1. Introduction and Objective

Mixed-stock analysis (MSA) has long been used to apportion harvests. Initially, MSA relied solely on phenotypic characteristics, behavioral patterns, and physical tags, but, with the advent of allozyme analysis, has shifted to genetic data. Currently, a wide range of molecular tools are available for use in MSA studies, some of which access highly polymorphic loci (e.g. microsatellites). The use of highly variable loci for MSA in turn prompted the development of new statistical methods to deal with the increased variation.

In this study, we use allozyme, microsatellite, and amplified fragment length polymorphism (AFLP) data, collected from Yukon River chum salmon populations, to compare the performance of the statistical methods.

2. Methods

Populations and identified structure



Sample sizes for the above populations range between 75 and 96. Data from 17 allozyme loci, 11 microsatellite loci, and 28 AFLP loci have been collected for these populations. The population structure depicted above is supported by all marker classes.

Create mixtures

First, we simulated mixtures for analysis. Mixtures were simulated by randomly drawing from the multinomial distribution with N equal to the mixture sample size and the probabilities equal to the specified mixture composition; individual multilocus genotypes were simulated by sampling with replacement pairs of alleles at each locus using the observed baseline allele/phenotype frequencies, and assumptions of Hardy-Weinberg and gametic phase equilibrium.

Mixtures with 0, 20, 40, 60, 80, and 100% contribution of Canadian border chum salmon (Fishing Branch and Big Creek) were created for each marker class. This is called an accuracy graph and allows for determining the performance of apportionment over a range of mixture scenarios. These populations were selected because the ability to identify Canadian origin fish is important for meeting Pacific Salmon Treaty mandates. Twenty five mixtures were simulated for each increment for a total of 150 mixtures per marker class. The computational time associated with Bayesian mixture modeling prevented analysis of > 25 mixtures per increment.

Analyze mixtures

Mixtures were then analyzed using the maximum likelihood (ML) and Bayesian mixture modeling methods, as implemented in the computer programs SPAM 3.7 and Bayes, respectively.

For the ML approach, AFLP and allozymes were analyzed based on their observed allele/phenotype frequencies. To control the problem of sampling zeros, the microsatellites were analyzed after their allele frequencies were adjusted with the Rannala-Mountain Bayesian estimator.

Statistically test mixture estimates

The mixture estimates were then statistically tested (sign test) to determine which apportionment method produced estimates that were closer to the expected. Each mixture was analyzed by both the ML and Bayesian method to ensure that differences between estimates were the result of the statistical method.

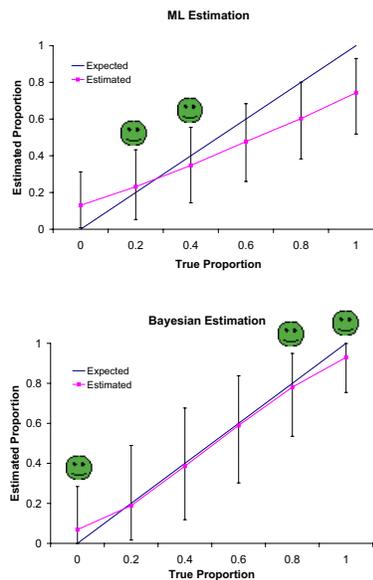


3. Results

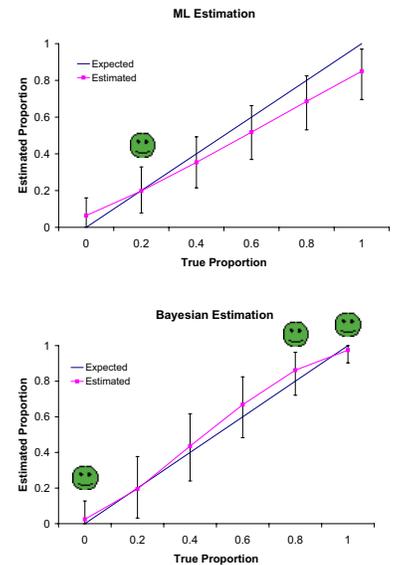
😊 = estimate significantly closer to the expected ($p < 0.05$)

Maximum Likelihood and Bayesian Mixture Modeling

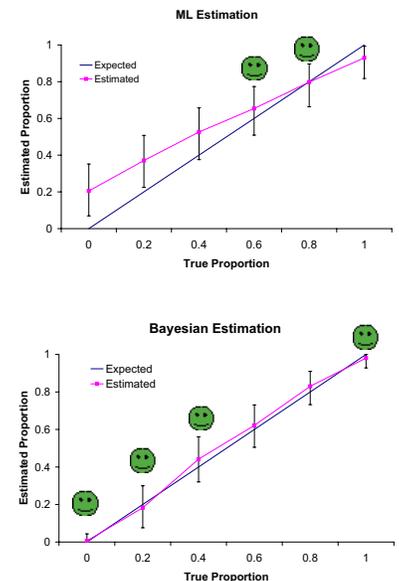
Allozymes



AFLP



Microsatellites



4. Conclusions

For microsatellites, Bayesian mixture modeling gave the best MSA results. The large bias for microsatellites using the ML method and discrepancies between mixture estimation and mixture simulation results (simulation results not shown) emphasizes the benefits and appropriateness of the Bayesian method for highly variable markers.

For moderately variable markers, there was not a dramatic improvement of the Bayesian method over ML. Bayesian does better at the extreme ranges of mixture contributions, but not necessarily in the middle, which suggests that the Bayesian method is not adversely affected by estimates being constrained between 0 and 1 as is the ML method.