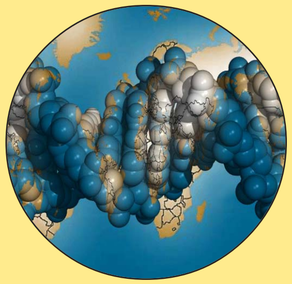


Single Nucleotide Polymorphisms (SNPs) provide high throughput and high resolution DNA data for sockeye salmon

J. Seeb, A. Antonovich, W. Templin, L. Seeb, C. Habicht, G. Johnston, K. Pronzati , E. Lardizabal, C. Elfstrom and C. Smith

Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, Alaska 99518, USA

Abstract

Studies of the ecology, migration, life history, or harvest of sockeye salmon often require markers for which a large number of individuals can be processed in a relatively short time. The advent of automated genotyping technologies makes some gene markers ideal for these studies. We describe the development and application of 26 single nucleotide polymorphism (SNP) genotyping assays that provide both high throughput and high resolution.

Introduction

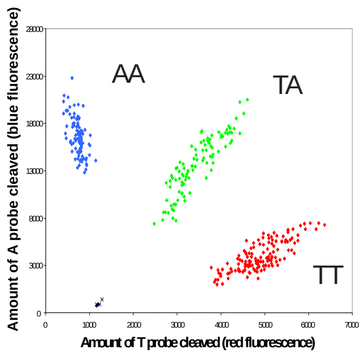
SNPs are biallelic markers that were first resolved in salmon in both nuclear and mitochondrial DNA using approaches such as restriction fragment length polymorphism assays or DNA sequencing. Although some high-resolution SNPs were identified, these approaches were generally not useful for large-scale studies of salmon populations because of time-consuming laboratory protocols.

Recent technical advances that produced substantial improvement in the rate and cost of SNP detection along with standard scores (A, C, T or G) lead many to predict that SNPs will be the marker of choice for studies of resource management. Advantages for SNPs include transportability of data from laboratory to laboratory, comparatively low rate of PCR error and genotyping error compared to other marker types such as microsatellites, relative ease of scoring, and high density of polymorphic markers throughout the genome. Additionally, SNP assays interrogate variation from both nDNA and the more rapidly evolving mtDNA. High-Fst outlier SNPs (e.g. SNPs under directional selection) may provide powerful signals that resolve population structure not apparent from the analysis of neutral markers.

A primary hurdle to SNP implementation is the development of high-throughput assays that resolve populations. Additionally, it is not clear how many biallelic SNPs will be required to equal the resolution offered by hypervariable microsatellites. We designed this study to develop and test high-throughput SNP markers for the identification of populations of sockeye salmon in high-seas and near-shore mixtures. We examined 26 nDNA and mtDNA SNPs in 12 populations distributed throughout the species range and compared the results to those obtained for 13 microsatellites.

Genotyping Without Gels

SNP genotyping involves simply amplifying target DNA in the presence of allele-specific dyes. The genotype of each fish is determined by the color of the resulting reaction. The color of each well in a 384 well plate is read by a scanner, and the results are displayed as a scatter plot. The simplicity of interpreting such a scatter plot allows 384 genotypes to be scored in under 5 minutes.



Methods

SNPs were developed using a targeted gene approach. Loci encoding functional genes were chosen based upon the expectation that historical mutations might occasionally produce high-Fst (outlier) loci that would provide especially powerful resolution like that observed with *major histocompatibility complex* (*MHCII*) SNPs previously reported. In this study, *transferrin* (iron binding serum protein) and *zona pellucida* (gamete recognition) SNPs provided resolution equal to or greater than that of *MHCII*. Three mtDNA SNPs were also included.

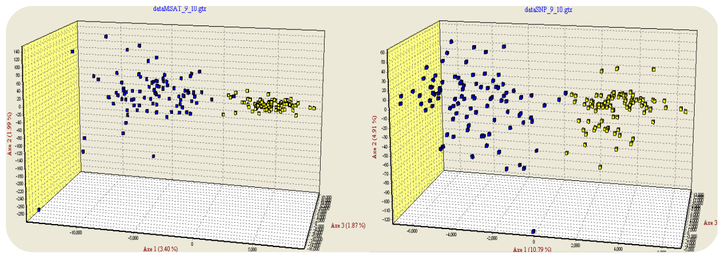
Data was collected for 26 SNP loci and 13 microsatellite loci from 12 test populations from throughout the species' range. Populations were organized in six proximal pairs. Resolving power of the two data sets was done by doing individual assignment tests on the proximal pairs and by using population assignment simulations on the entire 12 populations.

Individual Assignment

Individual assignment tests done using factorial correspondence analysis (Belkhir et al. 2004; www.univ-montp2.fr/~genetix/genetix/genetix.htm) were generally concordant between data types. Microsatellites performed slightly better in Washington and Norton Sound; SNPs were vastly superior in Bristol Bay.

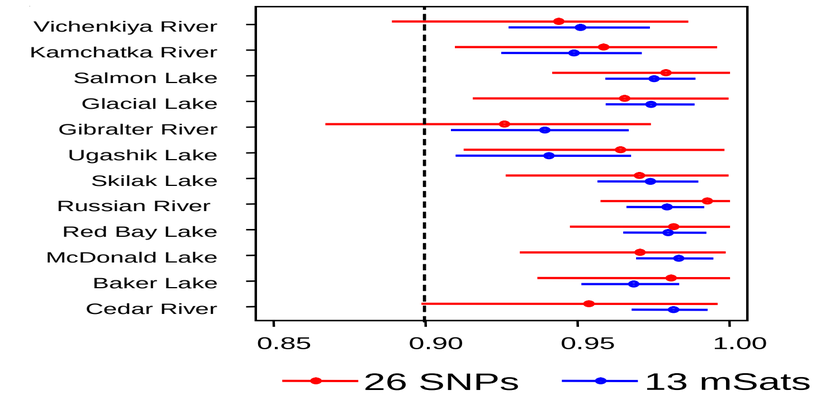
In this example in Southeast Alaska, McDonald Lake individuals are represented by blue boxes and Red Bay Lake individuals are yellow. Assignments are equally discrete for the two locus types; SNP assignments show more dispersion.

Microsatellite **SNP**

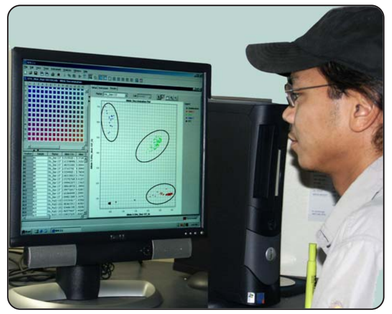


Population Assignment

The population assignment tests (Debevec et al. 2000; http://www.cf.adfg.state.ak.us/geninfo/research/genetics/genetics.php) were done by simulating a series of 12 mixtures, each composed of 100% of one of the component populations. Scores > 0.90 are commonly accepted values to validate a baseline for genetic stock identification purposes. Scores were generally ≥ 0.95; confidence limits were slightly larger around SNP estimates.



Conclusion



We find SNP markers to be a rapid and reliable method of stock identification. The expectation in the literature is that each individual microsatellite, with its many alleles, will contain five to eight times as much information as does any random neutral SNP. At the same time, outlier SNPs such as *MHCII* and others provide vastly more information than do random microsatellites. In this example, 26 SNPs provide population and individual assignment similar to 13 microsatellites. Economies of chemistry and throughput permit the analysis of about 75 SNPs per fish for the same cost of analyzing 13 microsatellites per fish. Agency and university consortia are now ramping up studies of SNP development that will soon provide enough additional markers to facilitate an array of high-throughput, relatively inexpensive studies at any scale. The markers discovered in this study have been exported to laboratories at NOAA Fisheries and the Russian Federal Research Institute of Fisheries for our combined effort to establish a common DNA database for sockeye salmon inhabiting the Pacific Rim.

Acknowledgments

This research was funded by North Pacific Research Board grants R0205 and R0303, a grant from the Southeast Alaska Sustainable Salmon Fund, and a grant from the Northern Fund of the Pacific Salmon Commission. A combined grant from ten Bristol Bay salmon processors enabled the first use of SNP data to provide in-season updates on the stock composition of sockeye salmon passing through test fisheries in the summer of 2005.