Which Genetic Markers and GSI Methods are More Appropriate for Defining Marine Distribution and Migration of Salmon?

Michael A. Banks and **David P. Jacobson**
Coastal Oregon Marine Experiment Station, Hatfield Marine Science Center,
Department of Fisheries and Wildlife, Oregon State University
2030 SE Marine Science Drive, Newport, Oregon 97365, USA

Keywords: Genetic marker choice, GSI, statistical power

Potential for increased statistical power in stock discrimination gained through employing more polymorphic marker types such as microsatellites is attractive for applications in determining the distribution and migration of salmon in the high seas. Increased sampling requirements owing to the many character types (alleles) that typify microsatellites, however, raise question about accuracy. While increasing the number of individuals characterized for baseline data addresses this issue, it is seldom feasible to characterize all likely contributing stocks. Furthermore, increasing population sample size (minimum n) requirements per stock inevitably will result in a reduction in the total number of potential contributing stocks that can be characterized in baselines, as funding for research is always finite. This approach thus results in diminishing returns because it increases the potential for more individuals in any study that may have originated from stocks other than those characterized in baselines. Methods invoking Baye's rule and a number of new approaches have shown promise in overcoming this enigma. This study reviews empirical findings from working with some of these new approaches (SPAM, Debevec et al. 2000; WHICHLOCI, Banks et al. 2003; and GMA, Kalinowski 2003). Primary goals were to determine how alternate genetic markers, sample sizes and population compositions effect precision and accuracy of results for known mixtures. Data includes microsatellites, major histocompatibility complex (MHC), and D-Loop mtDNA sequence variance (Table 1) from five alternate Chinook life history types of the Sacramento and San Joaquin rivers of central California, USA.

Our primary finding is that the statistical power of genetic stock identification is critically dependent on the information content of genetic markers (loci) employed. Power is maximized through ranking loci according to their discriminatory ability among populations under consideration and using a resource such as WHICHLOCI (Banks et al. 2003) to evaluate what minimum numbers of high ranking loci are necessary in order to achieve desired assignment accuracy and precision criteria (Fig. 1). Earlier studies have suggested that a modest number of loci, each with a modest number of alleles provides better assignment accuracy (Smouse and Chevillon 1998; Bernatchez and Duchesne 2000) and Cornuet et al. (1999) noted a converse relationship between number of loci (8–30) and number of individuals required (30–12) in order to achieve 100% assignment. Our studies, however, reveal that no simple predictor such as number of alleles or heterozygositiy of loci provides consistent prediction of individual based assignment performance. While more polymorphic and heterozygous loci generally rank higher, on occasion specific loci with as few as four or five alleles may rank high because of unique frequency and distribution of genotypes among populations under consideration. It is thus gainful to apply methods to assess information content of alternate loci and evaluate what minimum number of combined loci is necessary in order to achieve sufficient stock assignment power. For example, iteration over 10 data sets with a simulated sample size of 10,000 using the 16 best ranking loci among those employed in this study achieved 99.91% accuracy with a variance of 0.023, yet 99.18 accuracy is attained with just nine of these highest ranking loci with minimal variance increase (to 0.61).

Interestingly, when comparing the performance criteria of alternate loci choices for individual based population assignment (discussed above) with those attained for population based genetic stock identification methods such as GMA (Kalinowski 2003) or SPAM (Debevec 2000) both number of alleles and heterozygosity were better predictors of top performing loci than WHICHLOCI (Fig. 2). Note that sample size for data used in this study was only 19 for each life history type; a value well below the number (150–200) assessed as necessary to ensure reliable microsatellite allele frequencies (Banks et al. 2000). Ongoing research will test if these findings hold true for population based genetic stock identification methods using greater sample sizes per life history type.

Current advances in genotyping technology are facilitating the application of single nucleotide polymorphisms (SNPs) in genetic stock identification (GSI). At first glance, SNPs do not appear suited to GSI because most SNPs have only two alleles and can never have more than 4 alleles (because there are only four different bases in DNA), however, technology advances may allow effective characterization of several orders of magnitude more SNPs than the number of microsatellites that can be characterized for the same cost. These methods could thus harness similar

All correspondence should be addressed to M. Banks.
e-mail: Michael.Banks@oregonstate.edu

numbers of characters but from a broader component of the genome, with potential to provide even better stock discrimination performance, although linkage rears as a likely confounding factor. MHC is the only 'SNP' locus characterized among loci considered in this study, however, Rosenberg et al. (2003) provides insightful findings from their study of informativeness of alternate genetic markers for inferring ancestry among human populations. This study compares findings from 377 microsatellites and 8,714 SNPs and finds that random dinucelotide microsatellites are from five to eight times more informative than random SNPs, but also that 2–12% of SNPs have higher information content than the median for dinucleotides. We are only just beginning to accumulate sufficient data to allow similar comparison between microsatellites and SNPs for fisheries contexts, but the field is advancing fast, presenting exciting potential for high resolution stock identification that should provide enlightening perspective for marine distribution and migration of stocks in the near future.

Table 1. Genetic markers evaluated for information context in the context of genetic stock identification

Locus name	Number of alleles	Heterozygosity	WHICHLOCI rank	Reference
OtsG311	31	0.72026	1	Williamson et al. (2002)
Ots-107	25	0.76586	2	Nelsen & Beacham (1999)
OtsG409	41	0.87952	3	Williamson et al. (2002)
OtsG422	41	0.83782	4	Williamson et al. (2002)
Ots-209	31	0.8762	5	Greig et al. (2003)
OtsG253	25	0.83838	6	Williamson et al. (2002)
Ots-204	34	0.67852	7	Greig et al. (2003)
Ots-104	23	0.86344	8	Nelsen & Beacham (1999)
OtsG249	41	0.87342	9	Williamson et al. (2002)
Ots-211	33	0.7877	10	Greig et al. (2003)
OtsG83b	32	0.79526	11	Williamson et al. (2002)
Ots-213	27	0.70234	12	Greig et al. (2003)
Ots-201	20	0.8614	13	Banks et al. (pers. commun.)
Ots-215	20	0.78456	14	Banks et al. (pers. commun.)
OtsG78b	36	0.79196	15	Williamson et al. (2002)
Ots-208	27	0.83854	16	Greig et al. (2003)
Ots-2	13	0.63838	17	Banks et al. (1999)
One-13	10	0.72378	18	Scribner et al. (1996)
Ots-212	25	0.92508	19	Greig et al. (2003)
Ots-515	17	0.78348	20	Naish & Park (2002)
Ots-503	16	0.8062	21	Naish & Park (2002)
OtsG474	10	0.72574	22	Williamson et al. (2002)
OtsG432	24	0.8721	23	Williamson et al. (2002)
D-Loop	5	-	24	Nielsen et al. (1994)
Ots-3	10	0.76222	25	Banks et al. (1999)
Ots-501	11	0.68698	26	Naish & Park (2002)
Ots-528	9	0.64108	27	Naish & Park (2002)
MHC	4	0.59988	28	Kim et al (1999)
Ots-519	8	0.56574	29	Naish & Park (2002)
Ots-504	5	0.5448	30	Naish & Park (2002)
Ots-10	7	0.52104	31	Banks et al. (1999)
Ots-514	10	0.64692	32	Naish & Park (2002)

Fig. 1. Factorial correspondence analysis using data of 500 samples per Chinook life history type, simulated using WHICHLOCI (Banks et al. 2003) and analyzed using GENETIX (Belkhir et al. 2002). Square plots represent individual salmon samples colored in accord with their life-history type and positioned on the graph in relation to their genetic relationship. This is approximated from either: **A**. nine loci from the bottom of the rank depicted in Table 1 or **B.** nine loci from the top of this same rank**.** Tighter clustering of life-history types and zero overlap in B demonstrates the power gained through using loci with greater information content.

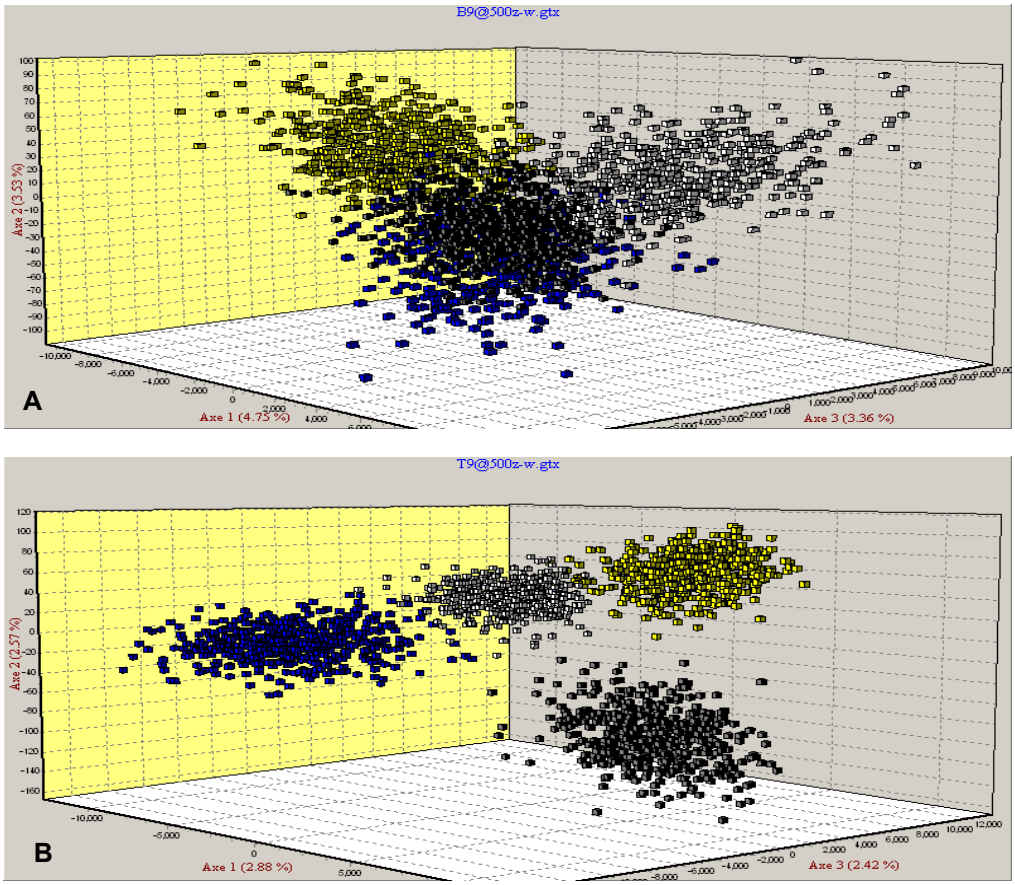
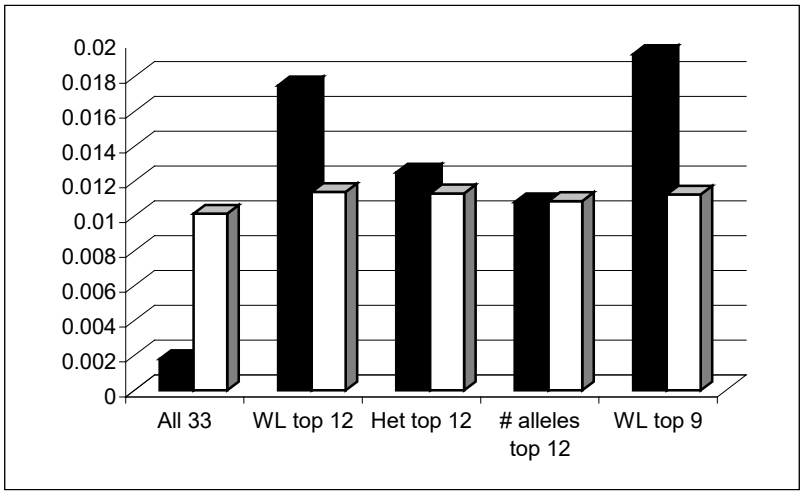


Fig. 2. Genetic stock Identification results using the population based GMA method of Kalinowski (2003). One thousand simulated fishery samples had N = 500. For fair comparison with individual based methods, relative component estimates were set with the two spring life history types and late fall each at a frequency equivalent to just one fish in the 500 fishery sample (0.002), setting the remainder for the components made up almost equal proportions of fall (0.442) and winter (0.522). Bars shown in black depict the absolute sum of the differences between observed and expected frequencies across all life history types and bars shown in white depict average standard deviation. Paired columns are displayed according to alternate loci combinations used: From left to right, all 33 loci together, 12 loci that ranked highest using WHICHLOCI, 12 loci with highest heterozygosities, 12 loci with the greatest number of alleles and lastly 9 loci that ranked highest using WHICHLOCI.



REFERENCES

Banks, M.A., Blouin, M.S., Baldwin, B.A., Rashbrook, V.K., Fitzgerald, H.A., Blankenship, S.M., and D. Hedgecock. 1999. Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tshawytscha*). J. Hered. 90: 281–288.

Banks, M.A., V.K. Rashbrook, M.J. Calavetta, C.A. Dean, and D. Hedgecock. 2000. Analysis of microsatellite DNA resolves genetic structure and diversity of chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. Can. J. Fish. Aquat. Sci. 57: 915–927.

Banks, M.A., W. Eichert, and J.B. Olsen. 2003. Which genetic loci have greater population assignment power? Bioinformatics 19(11): 1436–1438.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2002. *GENETIX 4.04, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions: CNRS UMR 5000, Université de Montpellier II, Montpellier, France. (http://www.univ-montp2.fr/%7Egenetix/genetix/genetix.htm)

Bernatchez, L., and P. Duchesne. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Can. J. Fish. Aquat. Sci. 57: 1–12.

Cornuet, J.-M., S. Piry., G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genet. 153: 1989–2000.

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. J. Hered. 91(6): 509–510.

Greig, C., D.P. Jacobson, and M.A. Banks. 2003. New tetranucleotide microsatellites for fine-scale discrimination among endangered chinook salmon (*Oncorhynchus tshawytscha*). Molecular Ecology Notes 3: 376–379.

Kalinowski, S.T. 2003. Genetic Mixture Analysis 1.0. Department of Ecology, Montana State University, Bozeman MT 59717. Available for download from http://www.montana.edu/kalinowski.htm

Kim, J.T., K.M. Parker, and P.W. Hedrick. 1999. Major Histocompatibility Complex differentiation in Sacramento River chinook salmon. Genetics, 151: 1115–1122.

Naish, K.A., and L.K. Park. 2002. Linkage relationships for 35 new microsatellite loci in chinook salmon *Oncorhynchus tshawytscha*. 33: 316–318.

Nelson, R.J., and T.D. Beacham. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. Animal Genetics 30: 228–229.

Nielsen, J.L., D. Tupper, and W.K. Thomas. 1994. Mitochondrial DNA polymorphism in unique runs of chinook salmon (*Oncorhynchus tshawytscha*) from the Sacramento-San Joaquin river basin. Cons. Biol. 8: 882–884.

Rosenberg, N.H., L.M. Li, R. Ward, and J.K. Pritchard. 2003. Informativeness of genetic markers for inference of ancestry. Am. J. Hum Genet. 73:1402–1422.

Scribner, K.T., J.R. Gust, and R.L. Fields. 1996. Isolation and characterization of novel salmonid microsatellite loci: cross-species amplification and population genetic applications. Can J. Fish. Aquat. Sci. 53: 833–841.

Smouse, P.E., and C. Chevillon. 1998. Analytical aspects of population-specific DNA fingerprinting for individuals. J. Hered. 89: 143–150.

Williamson, K.S., J.F. Cordes, and B. May. 2002. Characterization of microsatellite loci in Chinook salmon (*Oncorhynchus tshawytscha*) and cross-species amplification in other salmonids. Mol. Ecol. Notes 2: 17–19.