NPAFC
Doc. 824
Rev. _____

Development of a Pacific Rim baseline for chum salmon based on single nucleotide polymorphism markers (SNPs)

by

L. W. Seeb[1], C. T. Smith[1], W. D. Templin[1], R. L. Wilmot[2], and, J. E. Seeb[1]

[1]Gene Conservation Laboratory, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518
[2]National Marine Fisheries Service, Auke Bay Laboratory, 11305 Glacier Highway, Juneau, AK, USA 99801

Submitted to the
NORTH PACIFIC ANADROMOUS FISH COMMISSION
BY THE
UNITED STATES OF AMERICA

October, 2004

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:
Seeb, L. W., C. T. Smith, W. D. Templin, R. L. Wilmot, and, J. E. Seeb. 2004. Development of a Pacific Rim baseline for chum salmon based on single nucleotide polymorphism markers (SNPs). (NPAFC Doc. 824) 10 pages. Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518.

Abstract

Genetic stock identification using a comprehensive allozyme baseline has been extremely valuable for high-seas and coastal migration studies from throughout the Pacific Rim. However, newer DNA techniques provided significant advantages over allozymes in sampling, sample handling, and the potential for improved resolution. Of the various DNA markers, single nucleotide polymorphisms (SNPs) assayed through high-throughput technologies are particularly appropriate for NPAFC applications. Unlike microsatellite markers, they require no inter-laboratory standardization; SNPs are easily transferred between laboratories and instrument platforms. Here we evaluate the ability of six SNP markers to differentiate among eight populations representative of major chum salmon lineages from throughout the range. The results demonstrate that even these relatively few SNPs can accurately discriminate among many of the regional lineages of chum salmon. SNPs offer the promise of rapidly replacing the chum salmon allozyme baseline for BASIS and similar ~~highseas~~high-seas applications.

Background and Introduction

Genetic stock identification studies for chum salmon from throughout the Pacific Rim have been conducted for over two decades. During the 1990's, large data sets of allele frequency information from nuclear allozyme loci originating from Washington, British Columbia, Alaska, Canada, Russia, and Japan were completed. These data were collected by numerous agencies from around the Pacific Rim. The database was coordinated and disseminated through the North Pacific Anadromous Fish Commission (NPAFC) *ad hoc* Working Group on Stock Identification and reviewed in recent NPAFC documents (Crane et al. 2001; Kondzela et al. 2002).

The allozyme database has been widely applied by NPAFC nations and continues to be used to track the migration of juvenile chum salmon in the Gulf of Alaska and Bering Sea (e.g. Kondzela and Wilmot 2002) and in the Bering/Aleutian Salmon International Survey (BASIS) (e.g. Urawa et al. 2004). Other applications include estimates of the origin of chum salmon sampled from: 1) the Bering Sea trawl fishery for walleye pollock (*Theragra chalcogramma*) (Wilmot et al. 1998), 2) ~~highseas~~high-seas test fisheries (Winans et al. 1998; Urawa et al. 2000), 3) confiscated fishery samples from illegal fishing (Wilmot et al. 2000); and 4) commercial and test fishery catches from the south Alaska Peninsula and Kamchatka Peninsula (Seeb and Crane 1999a; Seeb et al. 2004).

Development of DNA Markers

A number of limitations are associated with the use of allozyme markers including lethal sampling, cryopreservation, and limited number of loci. Many classes of DNA markers do not share these limitations and provide additional advantages over allozyme markers. Concurrently with the application of the allozyme database, genetic markers based on DNA variation including mitochondrial DNA (mtDNA) PCR-RFLP, mtDNA sequence analyses, and microsatellite

markers were developed (e.g. Park et al. 1993; Scribner et al. 1998; Seeb and Crane 1999a; Abe et al. 2002; Beacham et al. 2004). However, developing high-throughput techniques has limited the applicability of some DNA marker types (e.g. mtDNA PCR-RFLP and mtDNA sequence analyses) while difficulties in transferability and standardization across laboratories and instrument platforms have inhibited others (e.g. microsatellite analyses). Standardization difficulties are particularly well illustrated by microsatellite markers.

Microsatellites are a class of marker for which the character states or alleles consist of different sized fragments. Microsatellites typically express a large number of alleles per locus (>10, ranging upwards of 60) which may provide better per locus resolution on average than markers with fewer numbers of alleles. Another attractive property of microsatellite markers is the potential for relatively high throughput. However, an important limitation of microsatellites is inter-laboratory standardization and transferability. Differences in physical and chemical conditions as well as instrument platforms among laboratories require considerable time and effort to standardize any microsatellite locus. This is particularly important in the context of multi-jurisdictional fisheries management and the mandate of the NPAFC *ad hoc* Working Committee on Stock Identification which is to develop and disseminate databases among the NPAFC nations. Despite over a decade of use, no inter-laboratory standardized microsatellite database exists for any species of Pacific salmonid.

Single nucleotide polymorphisms (SNPs) have been genotyped in salmon using a broad range of techniques (e.g. Park et al. 1993; Cronin et al. 1993; Gharrett et al. 2001). Although many of these SNPs were identified over a decade ago, the chemistries available for applying these markers were relatively slow and cumbersome. Recent developments have produced rapid SNP genotyping technologies which are now faster than those for any other marker class (Ranade et al. 2001; Melton 2003; Moriya et al. 2004; Smith et al. 2004). SNP data can be standardized to external DNA sequences and are thus automatically standardized across chemistries, instrument platforms, and laboratories.

SNP Baseline Development

We have initiated a significant effort to development SNP loci for chum salmon. To date we have identified 164 SNPs through DNA sequencing. Of these, 25 have been tested, and 13 have been configured for high-throughput analysis. Smith et al. (in press) report these 13 SNPs originating from five nuclear and two mitochondrial DNA (mtDNA) sequences (Table 1) including those polymorphisms previously applied by Sato et al. (2004). Smith et al. (in press) observed significant linkage disequilibrium for SNPs within loci (i.e. within the mtDNA, within the interleukin 8 receptor and within gonadotropin releasing hormone), but linkage disequilibrium was not detected among SNPs from different sequences.

Here we begin to evaluate the results of Smith et al. (in press) for their applicability to genetic stock identification studies on a Pacific Rim basis. Our analyses were based only on those loci without significant linkage disequilibrium including five nuclear and one mtDNA SNP: *Oke_CKS_2-389, Oke_GnRH_3-375*, *Oke_IgM_3-519, Oke_IL8r_1-46, Oke_DM20-548*, and *Oke_Cr30.* Eight populations from throughout the Pacific Rim ranging from Japan to Washington State comprised the analysis. These populations represented many of the known

lineages in chum salmon (Seeb et al. 2004). Populations included: Sasauchi River (Honshu Island, Japan), Tokachi River (Hokkaido Island, Japan), Bolshaya River (Kamchatka Peninsula, Russia), Sheenjek River (Upper Yukon drainage, United States), Anvik River (Lower Yukon drainage, United States), Volcano River (South Alaska Peninsula, United States), Wally H. Noerenberg Hatchery (WHN, northern Gulf of Alaska, United States), and Puget Sound (Washington State, United States).

A multi-dimensional scaling (MDS) plot based on chord genetic distances (Cavalli-Sforza and Edwards 1967) was constructed (Figure 1). Relationships depicted in the MDS closely follow those previously supported by allozyme data (Seeb and Crane 1999b; Seeb et al. 2004). We also conducted a simulation analysis using SPAM version 3.7 (Debevec et al. 2000) to test the power of the SNP markers to discriminate among the representative populations (Table 2). Conditional maximum likelihood estimates of population contributions were made for 1000 artificial mixtures with N=400 individuals. The expected mean contribution was equal to 100%. Mean contribution estimates were greater than 90% for four of the eight populations and above 90% for continent-of-origin for six of the eight populations. The simulations for two populations, Bolshaya and Volcano rivers, representing the Kamchatka and Alaska Peninsulas, misallocated between each other. These two groups also show close genetic affinities using allozymes (Seeb et al. 2004) and will require additional SNPs to successfully discriminate.

Conclusions

The results, not unexpectedly, presented here demonstrate that SNPs can accurately discriminate among many of the regional lineages of chum salmon. Simulation results and discriminating ability will improve as additional SNPs and populations are incorporated into the baseline. Based on studies in other organisms, we anticipate that 20-30 SNPs will be sufficient to accurately and precisely discriminate chum salmon at a scale useful for Pacific Rim analyses (Werner et al. 2004).

Data collected using these assays are particularly well suited to large-scale inter-jurisdictional studies such as have been conducted for chum salmon using the allozyme database. Since the markers directly reflect the underlying DNA sequences, they are readily combined with data collected across hardware, chemistry platforms, and laboratories. The repeatability of data as well as the rapid rate at which SNP data are generated suggests that SNP markers are well suited to mixture and migratory studies of chum salmon for a variety of Pacific Rim and NPAFC applications.

Literature Cited

Abe, S., Sato, S., Kojima, H., Ando, J., Anod, H., Wilmot, R.L., Seeb, L.W., Efremov, V., Leclair, L., Buchholtz, W., Jin, D-H., Urawa, S., Kaeriyama, M., Urano, A. 2002. Genetic differentiation among Pacific Rim populations of chum salmon inferred from mitochondrial DNA sequence variation. (NPAFC Doc 616) 27 pages. Laboratory of Animal Cytogentics, Center for Advanced Science and Technology, Hokkaido University, North 10, West 8, Kita-ku, Saporo 060-0810, Japan.

Beacham, T. D., K. D. Le, and J. R. Candy. 2004. Population structure and stock identification of chum salmon (*Oncorhynchus keta*) based upon microsatellite analysis. N. Pac. Anadr. Fish Comm. Tech. Rep 5: 31-33.

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

Crane, P. A., C. M. Kondzela, and L. W. Seeb. 2001. History and use of allozyme baselines for Pacific Rim chum salmon (NPAFC Doc. 561) 17 pages. Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK USA 99518.

Cronin, M. A., W. J. Spearman, R. L. Wilmot, J. C. Patton and J. W. Bickham. 1993. Mitochondrial DNA variation in chinook (*Oncorhynchus tshawytscha*) and chum salmon *(O. keta*) detected by restriction enzyme analysis of polymerase chain reaction (PCR) products. Canadian Journal of Fisheries and Aquatic Sciences 50: 708-715.

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: 509-511.

Gharrett, A.J., A.K. Gray, and V. Brykov. 2001. Phylogeographic analysis of mitochondrial DNA variation in Alaskan coho salmon, *Oncorhynchus kisutch*. Fishery Bulletin 99: 528-544.

Kondzela, C.M., P. A. Crane, S. Urawa, N. V. Varnavskaya, V. Efremov, X. Luan, W. D. Templin, K. Hayashizaki, R. L. Wilmot, and L. W. Seeb. 2002. Development of a comprehensive allozyme baseline for Pacific Rim chum salmon. (NPAFC Doc. 629) 23 pages. Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518.

Kondzela, C. and R. Wilmot. 2002. Origin of juvenile chum salmon from Gulf of Alaska coastal waters, 2001. Alaska Fishery Science Center Quarterly Report. July-August-September, 9 pp.

Melton, L. 2003. On the trail of SNPs. Nature 422: 917-923.

Moriya, S., A. Urano, S. Urawa, O. Suzuki, and S. Abe. 2004. Development of DNA microarray for rapid detection of mitochondrial DNA haplotypes of chum salmon. N. Pac. Anadr. Fish Comm. Tech. Rep. 5 28-30.

Park, L. K., M. A. Brainard, D. A. Dightman, G. A. Winans. 1993. Low levels of intraspecific variation in the mitochondrial DNA of chum salmon (*Oncorhynchus keta*). Molecular Marine Biology and Biotechnology 2: 362-370.

Ranade, K., M.-S. Chang, C.-T. Ting, D. Pei, C.-F. Hsiao, M. Olivier, R. Pesich, J. Hebert, Y.-D.I. Chen, V. Dzau, D. Curb, R. Olshen, N. Risch, D.R. Cox, and D. Botstein. 2001. High-throughput genotyping with single nucleotide polymorphisms. Genome Research 11: 1262-1268.

Sato, S., H. Kojima, J. Ando, H. Ando, R. L.Wilmot, L. W. Seeb, V. Efremov, L. Leclair, W. Buchholz, D. H. Jin, S. Urawa, M. Kaeriyama, A. Urano, and S. Abe. 2004. Genetic population structure of chum salmon in the Pacific Rim inferred Ffrom mitochondrial DNA sequence variation. Environmental Biology of Fishes 69: 37-50.

Scribner, K. T., P. A. Crane, W. J. Spearman, and L. W. Seeb. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: Analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 55: 1748-1758

Seeb, L. W. and P. A. Crane. 1999a. Allozymes and mitochondrial DNA discriminate Asian and North American populations of chum salmon in mixed-stock fisheries along the south coast of the Alaska Peninsula. Transactions American Fisheries Society 128: 88-103.

Seeb, L. W. and P. A. Crane. 1999b. High genetic heterogeneity in chum salmon in Western Alaska, the contact zone between northern and southern lineages. Transactions American Fisheries Society 128: 58-87.

Seeb, L.W., P. A. Crane, C. M. Kondzela, R. L. Wilmot, S. Urawa, N. V. Varnavskaya, and J. E. Seeb. 2004. Migration of Pacific Rim chum salmon on the high seas: insights from genetic data. Environmental Biology of Fishes 69: 21-36.

Smith, C. T., J. Baker, L. Park, L. W. Seeb, C. Elfstrom, S. Abe, and J. E. Seeb. In press. Characterization of 13 single nucleotide polymorphism markers for chum salmon. Molecular Ecology Notes.

Smith, C. T., J. E. Seeb, and L. W. Seeb. 2004. Development of 5'-nuclease reations for high-throughput SNP genotyping in salmon. N. Pac. Anadr. Fish Comm. Tech. Rep. 5: 45-46.

Urawa, S., T. Azumaya, P. A. Crane, and L. W. Seeb. 2004. Origin and distribution of chum salmon in the Bering Sea during the early fall of 2002: estimates by allozyme analysis. (NPAFC Doc. 794) 11 p. National Salmon Resources Center, Toyohira-ku, Sapporo 062-0922, Japan.

Urawa, S., M. Kawana, G. Anma, Y. Kamei., T. Shoji, M. Fukuwaka, K. M. Munk, K. W. Myers, and E. V. Farley, Jr. 2000. Geographic origin of high-seas chum salmon determined by genetic and thermal otolith markers. North Pacific Anadromous Fish Commission Bulletin 2: 283-290.

Werner, F. A. O., G. Durstewitz, F. A. Habermann, G. Thaller, W. Kramer, S, Kollers, J. Buitkamp, M. Georges, G. Brem, J. Mosner, and R. Fries. 2004. Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. Animal Genetics 35: 44-49.

Wilmot, R. L., C. M. Kondzela, C. M. Guthrie, and M. S. Masuda. 1998. Genetic stock identification of chum salmon harvested incidentally in the 1994 and 1995 Bering Sea trawl fishery. North Pacific Anadromous Fish Commission Bulletin 1: 285-299.

Wilmot, R. L., C. M. Kondzela, C. M. Guthrie III, A. Moles, J. J. Pella, and M. Masuda. 2000. Origins of salmon seized from the F/V Arctic Wind. (NPAFC Doc. 471) 18 pp. Auke Bay Fisheries Laboratory, Alaska Fisheries Science Center, NMFS, NOAA, 11305 Glacier Highway, Juneau, AK 99801-8626.

Winans, G. A., P. B. Aebersold, Y. Ishida, and S. Urawa. 1998. Genetic stock identification of chum salmon in ~~highseas~~high-seas test fisheries in the western North Pacific Ocean and Bering Sea. North Pacific Anadromous Fish Commission Bulletin 1: 220-226.

Table 1. Observed allele frequencies for the rarer allele at 13 SNP loci. Collections were taken from Sasauchi River (Honshu Island, Japan), Tokachi River (Hokkaido Island, Japan), Bolshaya River (Kamchatka Peninsula, Russia), Sheenjek River (Upper Yukon drainage, United States), Anvik River (Lower Yukon drainage, United States), Volcano River (South Alaska Peninsula, United States), Wally H. Noerenberg Hatchery (WHN, northern Gulf of Alaska, United States), and Puget Sound (Washington State, United States). From Smith et al. (in press).

	Population							
Locus:	Sasauchi	Tokachi	Bolshaya	Sheenjek	Anvik	Volcano	WHN	Puget Sound
~~n~~N	78	80	95	95	95	64	95	91
Oke_Cr30	0.347	0.182	0.032	0.000	0.011	0.079	0.000	0.000
Oke_Cr231	0.026	0.025	0.032	0.000	0.000	0.016	0.000	0.011
Oke_Cr386	0.883	0.722	0.063	0.000	0.011	0.078	0.000	0.000
Oke_ND3_2-287	0.885	0.734	0.032	0.000	0.011	0.078	0.000	0.000
Oke_CKS_2-389	0.135	0.222	0.382	0.274	0.274	0.375	0.483	0.099
Oke_GnRH_3-375	0.867	0.763	0.500	0.145	0.315	0.422	0.389	0.216
Oke_GnRH_4-529	0.000	0.000	0.043	0.126	0.168	0.023	0.006	0.016
Oke_IgM_3-519	0.026	0.081	0.274	0.069	0.150	0.234	0.391	0.450
Oke_IL8r_1-46	0.571	0.559	0.321	0.263	0.213	0.359	0.183	0.148
Oke_IL8r_2-136	0.563	0.526	0.289	0.263	0.216	0.359	0.174	0.144
Oke_IL8r_3-283	0.387	0.365	0.205	0.116	0.074	0.141	0.067	0.071
Oke_IL8r_4-417	0.545	0.519	0.299	0.263	0.216	0.359	0.176	0.159
Oke_DM20-548	0.149	0.237	0.479	0.606	0.426	0.391	0.472	0.330

Table 2. a. Mean contribution estimates for 1000 bootstrap resamplings of individual populations based upon data from six independent SNP loci (*Oke_CKS_2-389, Oke_GnRH_3-375, Oke_IgM_3-519, Oke_IL8r_1-46, Oke_DM20-548*, and *Oke_Cr30*). ~~a. Mean contribution estimates for 1000 bootstrap resamplings of individual populations.~~ Each simulated mixture (N=400) was composed of 100% of a single population. Bold cells indicate correct population allocation. b. Mean contribution estimates for continent-of-origin of the above simulations.

a.

	Population	Sasauchi	Tokachi	Bolshaya	Sheenjek	Anvik	Volcano	WHN	Puget Sound
1	Sasauchi R.	**0.910**	0.104	0.001	0.000	0.001	0.002	0.000	0.000
2	Tokachi R.	0.089	**0.868**	0.011	0.001	0.004	0.027	0.000	0.001
3	Bolshaya R.	0.000	0.008	**0.736**	0.003	0.033	0.196	0.037	0.002
4	Sheenjek R.	0.000	0.001	0.011	**0.950**	0.083	0.015	0.011	0.008
5	Anvik R.	0.000	0.003	0.025	0.037	**0.791**	0.056	0.015	0.015
6	Volcano R.	0.000	0.015	0.135	0.002	0.046	**0.666**	0.005	0.001
7	WHN	0.000	0.000	0.074	0.004	0.019	0.027	**0.912**	0.010
8	Puget Sound	0.000	0.001	0.007	0.004	0.025	0.011	0.020	**0.963**

b.

		Asia	North America
Asia			
1	Sasauchi R.	0.999	0.001
2	Tokachi R.	0.979	0.021
3	Bolshaya R.	0.748	0.252
North America			
4	Sheenjek R.	0.003	0.997
5	Anvik R.	0.038	0.963
6	Volcano R.	0.225	0.775
7	WHN	0.038	0.962
8	Puget Sound	0.003	0.997

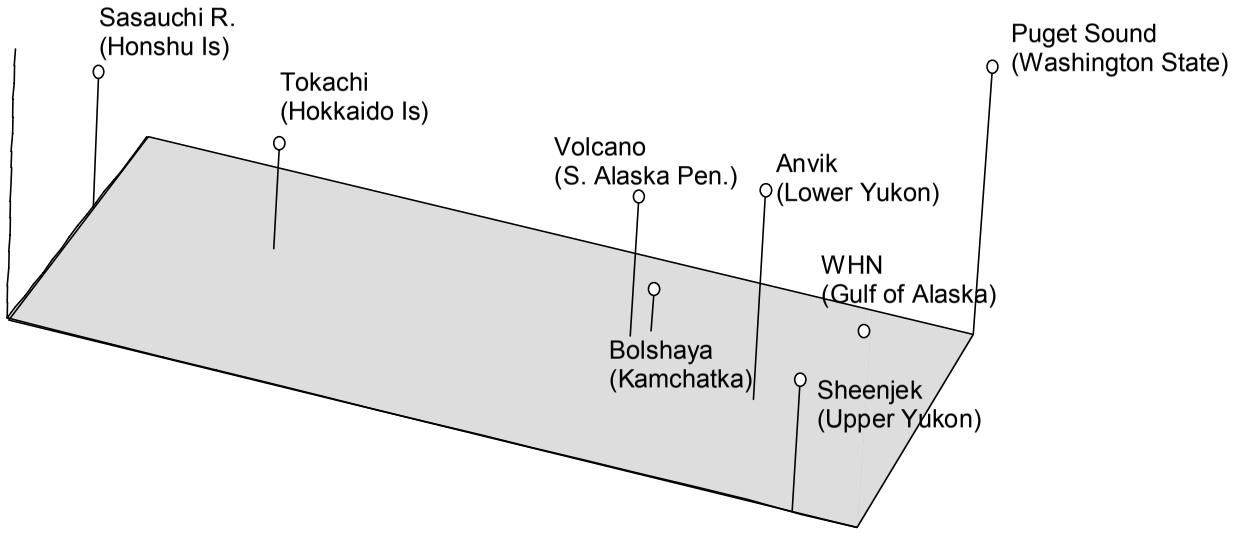

Figure 1. Multi-dimensional scaling based on Cavalli-Sforza and Edwards (1967) chord distances for eight representative populations of chum salmon.