NPAFC
Doc. 1138
Rev. _____

PACSNP: Progress on the development and standardization of single nucleotide polymorphisms (SNPs) baseline for genetic stock identification of chum salmon.

By

J. E. Seeb[1], S. Abe[2], S. Sato[3] , W. D. Templin[4], S. Urawa[5],
K. I. Warheit[6], and L. W. Seeb[1]

[1]*University of Washington, School of Aquatic and Fishery Sciences, 1122 NE Boat Street, Box 355020, Seattle, WA USA 98195*
[2]*Laboratory of Breeding Science, Division of Marine Biosciences, Graduate School of Fisheries Sciences, Hokkaido University, 3-1-1 Minato, Hakodate 041-8611, Japan*
[3]*National Salmon Resources Center, Fisheries Research Agency, 2-2 Nakanoshima, Toyohira-ku, Sapporo 062-0922, Japan*
[4]*Gene Conservation Laboratory, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518*
[5]*North Pacific Anadromous Fish Commission, Suite 502, 889 West Pender Street, Vancouver, BC, V6C 3B2 Canada*
[6]*Washington Department of Fish and Wildlife, 600 Capitol Way N. Olympia Washington, 98501USA*

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

November 2008

This document may be cited as:
Seeb, J. E., S. Abe., S. Sato, W. D. Templin, S. Urawa, K. I. Warheit, and L. W. Seeb. 2008. PACSNP: Progress on the development and standardization of single nucleotide polymorphisms (SNPs) baseline for genetic stock identification of chum salmon. NPAFC Doc. 1138. 10 pp. (Available at http://www.npafc.org)

Abstract

Genetic markers have been extremely valuable in the study of the distribution and migration routes of salmon in the ocean. Of the various markers available**, s**ingle nucleotide polymorphisms (SNPs) assayed through high-throughput technologies are particularly appropriate for NPAFC applications because they are based on the actual DNA sequence and require little inter-laboratory standardization. Japanese and United States researchers met October 15-16, 2008, to discuss the development and standardization of the current SNP baseline for chum salmon. The collaboration (termed "PACSNP") focused on three major areas: 1) SNP discovery and genotyping, 2) joint projects, and 3) shared databases.

Introduction

Genetic studies have become an important part of North Pacific Anadromous Fish Commission (NPAFC) research activities. The NPAFC *ad hoc* Working Group on Stock Identification Studies has coordinated these studies with the goals to develop, standardize, and disseminate genetic databases among the parties. Genetic databases on chum salmon, in particular, have been coordinated through NPAFC. During the 1990's, large data sets of allele frequencies 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 and have been used extensively by NPAFC in BASIS and other research to determine distribution and migration routes of salmon in the ocean. Current work has focused on developing databases using DNA markers as a replacement for the allozyme markers.

United States and Japanese researchers met in Sapporo, Japan, at the National Salmon Resources Center for a two day meeting, October 15-16, 2008. The goal of the meeting was to discuss progress and development of a DNA database based on single nucleotide polymorphism (SNPs). SNPs are assayed through high-throughput technologies and are particularly appropriate for NPAFC applications. Unlike marker types based on fragment size, SNPs are based on the actual DNA sequence and require little inter-laboratory standardization. SNP data can be easily transferred between laboratories and instrument platforms. The participants informally adopted the name PACSNP for the initiative which may be expanded to other species of salmon in the future.

Summary of PACSNP Meeting

Discussions during the meeting focused on three major areas: 1) SNP discovery and genotyping, 2) Joint projects, and 3) Shared databases.

SNP discovery and genotyping. Current SNP discovery efforts in both United States and Japan are focused on nucleotide sequencing. Japanese researchers have developed DNA microarrays for mtDNA haplotype detection by hybridization. Researchers in the United States have begun using next-generation sequencing (454 technology, Barbazuk et al 2007). DNA sequence results from six runs of 454 sequencing of chum salmon cDNA are now available which

represent the bulk of the salmon transcriptome and should lead to the identification of tens of thousands of putative SNPs. An easy SNP detection kit and/or microarray for large numbers of SNPs would be beneficial for NPAFC studies.

Dr. Junko Stevens, from Applied Biosystems Incorporated, Foster City, California, USA, attended the meeting and presented an overview of TaqMan chemistry for SNP genotyping as well as SNP discovery through the process of high-resolution melt (HRM). She was able to give her presentations in both English and Japanese. The recently-developed medium density open array system by Fluidigm Corporation was also reviewed. This chip-based technology allows for the analysis of 96 samples for 96 SNPs in a medium-density array (96.96 chips) at a cost similar or lower to earlier single-plex or lower-density arrays.

It was agreed that additional loci are desirable to improve resolution to better address the level of resolution desired for NPAFC studies. United States researchers have found that SNPs identified for their ability to discriminate stocks in one region may have lower variability and limited resolution in other regions. This property, known as ascertainment bias, refers to the biased information content of marker frequencies based upon the geographic origin of the marker discovery (Smith et al. 2007). For example, heterozygosities for the 60 markers are highest in Alaskan populations and tend to be lower at the extremes of the range in Japan and the Pacific Northwest (see Figure 1,2). It was agreed that SNP discovery in the future for NPAFC applications would benefit from ascertainment panels including Japan and the Pacific Northwest of Washington.

Joint projects. The current joint Japanese and United States SNP baseline now includes 114 populations for 60 SNPs (Table 1). It was agreed that future work will build on this set of populations and set of 60 SNPs (Table 2, "PACSNP"panel). There was general agreement that to the extent possible, the SNP baseline should have a uniform distribution around the Pacific Rim and be as comprehensive as the allozyme baseline (Kondzela et al. 2002). The Washington Department of Fish and Wildwife (WDFW) is developing a baseline for its area, and the UW and WDFW are writing proposals to include British Columbia populations. All participants agreed that increase the representation of Russian populations was very important.

Shared databases. The participants agreed that a web-based database for the chum salmon SNP data would be very useful. Participants expressed an interest to have it open-accessed and linked to the NPAFC site similar to the current otolith mark website. Various possibilities will be explored in the future.

Acknowledgements

Funding for participant travel was provided by the University of Washington, School of Aquatic and Fishery Sciences, through a grant from the Gordon and Betty Moore Foundation to the Seeb Laboratory. The National Salmon Resources Center, Fisheries Agency of Japan, graciously hosted the meeting.

References

Barbazuk, W. B., S. J. Emrich, H. D. Chen, L. Li, and P. S. Schnable. 2007. SNP discovery via 454 transcriptome sequencing. Plant Journal 51(5):910-918.

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.

Elfstrom, C. M., C. T. Smith, and L. W. Seeb. 2007. Thirty-eight single nucleotide polymorphism markers for high-throughput genotyping of chum salmon. Molecular Ecology Notes 7(6):1211-1215.

Smith, C. T., and coauthors. 2005a. Characterization of 13 single nucleotide polymorphism markers for chum salmon. Molecular Ecology Notes 5(2):259-262.

Smith, C. T., and L. W. Seeb. 2008. Number of alleles as a predictor of the relative assignment accuracy of short tandem repeat (STR) and single-nucleotide-polymorphism (SNP) baselines for chum salmon. Transactions of the American Fisheries Society 137:751-762.

Smith, C. T., C. M. Elfstrom, L. W. Seeb, and J. E. Seeb. 2005b. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14(13):4193-4203.

Table 1. SNP loci adopted at PACSNP. Source of SNPs and comments are given.

	Locus[1]	Source[2]	Comments
1	*Oke_AhR1-278*	3	
2	*Oke_AhR1-78*	3	
3	*Oke_arf-319*	2	
4	*Oke_U401-143*[a]	3	Linked pair
5	*Oke_U401-220*[a]	3	Linked Pair
6	*Oke_CKS-389*	2	
7	*Oke_copa-211*	2	
8	*Oke_Cr30*[b]	1	mtDNA
9	*Oke_Cr386*[b]	1	mtDNA
10	*Oke_ND3-69*[b]	1	mtDNA
11	*Oke_ctgf-105*	3	
12	*Oke_DM20-548*	1	
13	*Oke_eif4ebp2-64*	2	
14	*Oke_FARSLA-242*	3	
15	*Oke_GHII-2943*[c]	3	Linked pair
16	*Oke_GHII-3129*[c]	3	Linked pair
17	*Oke_GnRH-373*	1	
18	*Oke_GnRH-527*	1	
19	*Oke_GPDH-191*	2	
20	*Oke_GPH-105*	3	
21	*Oke_GPH-78*	3	
22	*Oke_hnRNPL-239*	3	
23	*Oke_HP-182*	3	
24	*Oke_hsc71-199*	2	
25	*Oke_HSP90BA-299*	3	
26	*Oke_il-1racp-67*	2	
27	*Oke_IL8r-272*[d]	1	Linked pair
28	*Oke_IL8r-406*[d]	1	Linked pair
29	*Oke_KPNA2-87*	3	
30	*Oke_MAPK1-135*	3	
31	*Oke_MARCKS-362*	3	
32	*Oke_Moesin-160*	2	
33	*Oke_PP2A-635*	3	
34	*Oke_ras1-249*	3	
35	*Oke_RFC2-618*	2	
36	*Oke_RH1op-245*	2	

	Locus[1]	Source[2]	Comments
37	*Oke_serpin-140*	2	
38	*Oke_TCP1-78*	3	
39	*Oke_Tf-278*	3	
40	*Oke_Tsha1 196*	2	
41	*Oke_u1-519*	1	
42	*Oke_u202-131*	2	
43	*Oke_u212-87*	2	
44	*Oke_u216-222*	2	
45	*Oke_u217-172*	2	
46	*Oke_u200-385*	2	
47	*Oke_U302-195*	3	
48	*Oke_U502-241*	3	
49	*Oke_U503-272*	3	
50	*Oke_U503-302*	3	
51	*Oke_U504-228*	3	
52	*Oke_U505-112*	3	
53	*Oke_U506-110*	3	
54	*Oke_U507-286*[e]	3	Linked pair
55	*Oke_U507-87*[e]	3	Linked pair
56	*Oke_U305-130*	3	
57	*Oke_U305-307*	3	
58	*Oke_U509-219*	3	
59	*Oke_U510-204*	3	
60	*Oke_U511-271*	3	

[1] Linked SNP pairs are denoted with superscript letters.

[2]Sources are as follows: (1) (Smith et al. 2005a), (2) (Smith et al. 2005b), (3) (Elfstrom et al. 2007)

Table 2. Number of populations included in the 2008 PACSNP baseline.

Major Region	Region	Location	Number of Pops
Japan			
	Hokkaido	Japan Sea Coast of Hokkaido	2
		Nemuro Coast	2
		Okhotsk Sea Coast of Hokkaido	3
		Pacific Coast Hokkaido	5
	Honshu	Japan Sea Coast	3
		Pacific Coast Honshu	1
Korea			1
Russia			
		Amur River	2
		Anadyr River	1
		Eastern Kamchatka	3
		Western Kamchatka	4
Western Alaska			
	Kotzebue Sound		3
	Norton Sound		9
	Yukon River Summer	Lower	8
		Middle	3
	Yukon River Fall	Canada	4
		US Border	4
		US Tanana	5
	Kuskokwim Fall		2

Major Region	Region	Location	Number of Pops
	Kuskokwim Summer		10
	Bristol Bay	Togiak	1
		Nushagak	3
		Southwest Drainages	5
Alaska Peninsula/Kodiak			
	North Alaska Peninsula		7
	South AK Peninsula		6
	Kodiak Island		3
Gulf of Alaska			
	Susitna River		3
	Central Gulf	Prince William Sound	2
Southeast Alaska British Columbia			
	Northern Southeast		5
	Southeast Alaska Mainland		2
	Southern SE/N BC		3
Washington			
	Washington		2
	Fall		

Figure 1. Percentage of polymorphic loci in chum salmon populations in geographic order from west to east ranging from Japan (left most) across Alaska and the Pacific Rim to the Pacific Northwest (right most). SNPs were developed primarily for Alaska and show the highest level of polymorphism in that portion of the range.

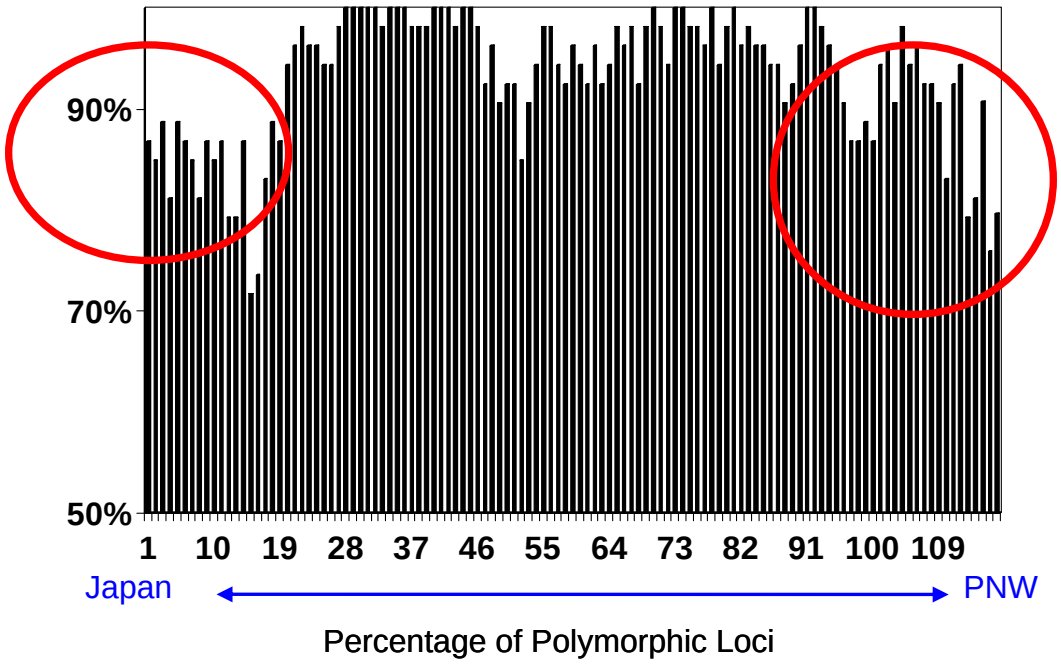
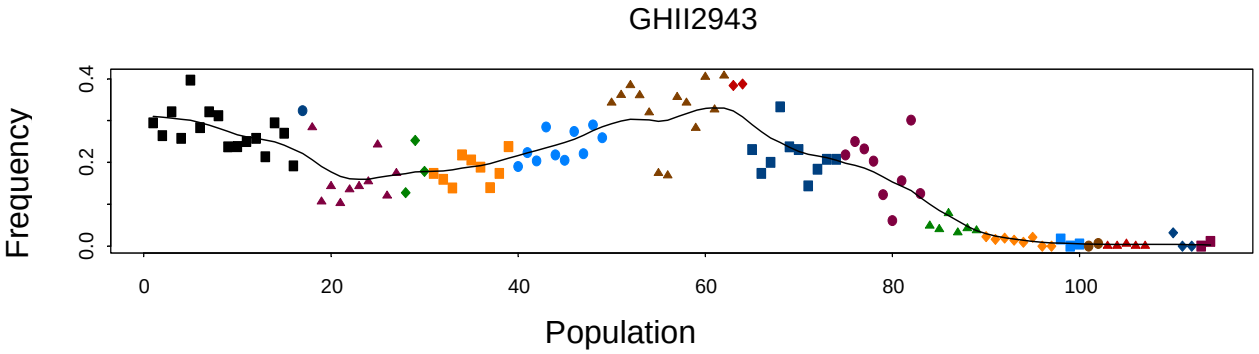
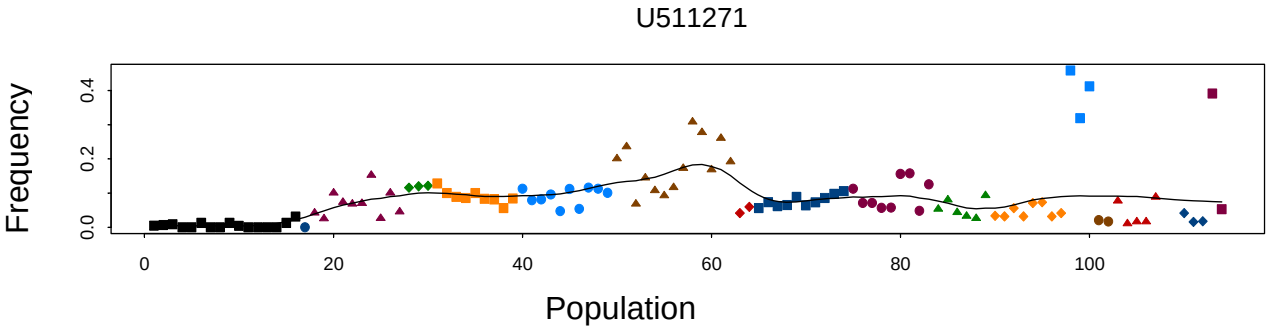


Figure 2. Two SNPs from the existing PACSNP panel of 60 SNPs. Frequency of common allele is on the y axis and the populations are ranked in geographic order across the Pacific Rim from west to east on the x axis.

A. SNP *GHI12943* provides an extreme example of ascertainment bias. This SNP was discovered in Alaska populations in the central region (roughly populations 25-80) and is highly variable there. This SNP shows limited resolving power in populations on the extremes of the range in British Columbia and Washington (population numbers > 100).



B. Ascertainment bias is not fatal to the value of a SNP. SNP *U511271* was also discovered in Alaskan populations and shows high variability in British Columbia and Washington (right most) and will be very useful in GSI studies in the southern portion of the range.



.