SNP Development and Lab Infrastructure Support for Genetic Stock ID

Final Report to the Pacific Salmon Commission
May 29, 2009

by

Shawn Narum & Nate Campbell

Columbia River Inter-Tribal Fish Commission
Hagerman Fish Culture Experiment Station
3059-F National Fish Hatchery Road
Hagerman ID 83332

Summary:

The objectives of this project included two key items that will improve GSI capabilities regionally and coastwide. First, additional single nucleotide polymorphism (SNP) markers were developed for use in the genetics community to differentiate important fall Chinook indicator stocks. A diverse sample of 32 individual Chinook salmon samples were sequenced in ten gene fragments, and a total of 64 variable sites were observed. One or two SNPs per gene were submitted for further development into Taqman genotyping assays, and 9 of 11 were successful. These assays have been provided to the genetics community for broader genotyping studies. Second, laboratory equipment/database were acquired to increase genotyping efficiency and throughput that will allow GSI estimates from fisheries to be produced in a timely manner. Specific items included an automated pipetting robot (Tecan EVO200) and a genotyping database (PROGENY).

Introduction

Genetic Stock Identification (GSI) with either microsatellite or single nucleotide polymorphism (SNP) markers have proven to be effective in determining the proportion of stock origin in mixed stock fisheries (Beacham et al. 2006; Smith and Seeb 2008). SNP discovery and assay development will provide additional genetic markers to increase power for Genetic Stock Identification (GSI) of mixed stock fisheries. This study was designed to develop additional SNPs quickly and efficiently, and follows from Recommendation 12 in the PSC Expert Panel Report (Hankin et al. 2005), that there is a need for support of an "immediate evaluation of a coordinated transition for all salmon species from GSI based on the use of microsatellite markers to GSI based on SNP markers."

Current Project Objectives

The specific objectives of this project were:

1) Develop additional SNP markers to differentiate summer/fall run Chinook salmon
2) Improve infrastructure for GSI capability with laboratory equipment and data processing/storage.

The first objective of this proposal was to develop SNP markers that specifically differentiate summer/fall stocks to improve stock resolution for GSI applications. Current microsatellite markers do not adequately distinguish populations of summer/fall run Chinook salmon in the Columbia River. Inclusion of SNP markers used in tandem with

existing microsatellite data will not only improve the resolution of stocks for GSI, but will also improve accuracy of stock assignments.

The second objective of this project was to improve genotyping throughput and data handling in the laboratory since thousands of samples are processed in the CRITFC genetics laboratory (Hagerman Fish Culture Experiment Station). Broad implementation of GSI applications will require genetics laboratories to efficiently genotype large numbers of samples and accurately track and store the data. The automated pipetting robot and genotype database that were acquired will greatly improve the capabilities of the CRITFC genetics laboratory for GSI applications.

SNP Discovery Methods

Expressed sequence tags (ESTs) and fully annotated genes were selected from GenBank and TIGR databases for primer design. Primer pairs were designed using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to generate a product between 300 and 600bp. Each set of primers was then tested by PCR on extracted genomic DNA from *O. tshawytscha* under the following conditions: 1x Thermophilic DNA polymerase buffer (Promega), 2mM $MgCl_2$, 0.25mM dNTPs, 0.1 mg/ml BSA, and 1U/rxn ampliTaq polymerase (Applied Biosystems; ABI) with 2uL genomic DNA (extracted using Qiagen DNeasy 96 kits ®) in 12uL total volume. Thermal cycler conditions for PCR were typically 94°C for 1 min, 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min/kb, with steps 2-4 repeated for 35 cycles followed by a cool down to 4°C (MJ Research and ABI thermal cyclers). Amplified fragments were

evaluated by agarose gel electrophoresis for clean amplification of a single product. Primer pairs that produced clean fragments were selected for amplification and sequencing of an ascertainment panel of 32 individuals per species with diverse genetic backgrounds (Table 1). Amplified fragments were successfully amplified and sequenced using Big Dye v1.1 chemistry and a 3730 DNA sequencer (ABI). Following electrophoresis, chromatogram data was first analyzed with the Sequencing Analysis program (ABI) to assign base calls and then Sequencher v.4.7 (Gene Codes) was used to align and edit the data. Sequencing analysis revealed some fragments that appear to be in duplicated genes and these fragments were excluded from further development.

Since nearly all gene fragments contained multiple SNPs, polymorphic sites were prioritized for further development by minor allele frequency, observation of all three possible genotypes, and adequate sequence data in flanking regions. One or two SNPs per gene were identified for further development using these criteria, and pertinent sequences were submitted to ABI for Custom Taqman™ allelic discrimination assays (5' exonuclease assays). Primer and probe mix for each submitted SNP was designed and produced by ABI. Genotyping was performed in 384 well plates using 1uL of dried *O. mykiss* genomic DNA and 5uL 1X TaqMan® Universal PCR Master Mix (No AmpErase® UNG, ABI) with 1X primer-probe mixture. Amplification was performed using MJ Research and ABI thermal cyclers with 384-well blocks using standard 2-step cycling for 50 cycles. End point reads of each reaction were collected using an ABI 7900HT instrument and data was analyzed using the allelic discrimination function of ABI SDS v.2.1 software.

Table 1. SNPs identified in 10 genes from 32 samples of Chinook salmon.

Sample	Genetic Lineage	Location	MapK prom: A151T	MapK 3' A309C	hsp27b1a: C150T	hsp47 3': A339G	hsp90BA 3': C252T	hsc71-5': G453A	hsc71 3': C488T	arp: T51A	arp: TT436AA	* hsf1b 3': 156(T)	* hsp70a 3': C319A
1	Col. R. stream-type	Carson Cr.	A/T	A/C	C	A	C/T	G	T	T/A	TT	9	C
2	Col. R. stream-type	Cowlitz Hat. sp	A		C	A	C/T	A	C	A	TT	9/10	C/A
3	Col. R. stream-type	Cowlitz Hat. sp	A/T	A/C	C	A	C	A	C	T	TT	9/10	A
4	Col. R. stream-type	Rapid River	T	A	C	A	T	G	C	T	TT	9	C
5	Northern range	King Salmon River	A/T	A	C	A	T	G	C/T	T	TT	9	C
6	Northern range	Bistraya River	T	A	C	A	T	G	T	T	TT	9	C
7	Northern range	Tahini River	T	A	T	A	C/T	G	C	T	TT	9/10	C
8	Northern range	Togiak River			C	A/G	T	G	T	T	TT		
9	Tule	Spring Cr. H			C/T	A	C/T		C	A	TT	9/10	C/A
10	Tule	Spring Cr. H	A		C	A	C/T		C	T	TT/AA		A
11	Col. R. ocean-type	Hanford Reach			C/T	A/G		G/A	C				
12	Col. R. ocean-type	Hanford Reach	A	C	C	A	C	G/A	C	T	AA	9/10	A
13	Col. R. ocean-type	Lyons Ferry Hat	A	A/C	T	A	C	G	C/T			9/10	A
14	Col. R. ocean-type	Lyons Ferry Hat		A/C	C/T	A	C	G	T	T	TT	9/10	C/A
15	Col. R. ocean-type	Methow R. summer	A	A	C		T	G	T	T/A	TT	9	C
16	Col. R. ocean-type	Methow R. summer	A	C	C/T	A	C/T	G	C/T	T	TT/AA	9/10	C/A
17	Col. R. ocean-type	Methow R. summer	A	C	C	A	T	G	T	T	TT	9	C
18	Col. R. ocean-type	Methow R. summer	A	A	C	A	T	G	T	T	TT	9	C
19	PNW coastal	North Fork Hat.	A	C	C	A	C	G	T	T	TT	9/10	C
20	PNW coastal	South Fork Hoh R.	A	A/C	T	A	C/T	G	C/T	T	TT	9/10	
21	PNW coastal	McQueston River			C	G	C	G	T				
22	PNW coastal	Salmon River			C	G		G	C/T				
23	PNW coastal	SF Umpqua	A	A	C	A	C	G	C	T	TT		C/A
24	PNW coastal	Nanaimo	A/T	A	C/T	A	C/T	G/A	C/T			9/10	A
25	PNW coastal	Nanaimo	A	A	C/T	A	C	G/A	C/T			9/10	C
26	PNW coastal	Nestucca	A	C	C	A	C	G	C	T	TT/AA	9	
27	PNW coastal	Soos	A	C	T	A	C/T	G/A	C	T	TT/AA	9/10	C
28	PNW coastal	Soos	A	A/C	C/T	A	C/T	G/A	C/T			9/10	C/A
29	Southern range	Sacramento R. (fa)	A/T	A/C	T	A	C/T	A	C	T/A	TT/AA		C/A
30	Southern range	Sacramento R. (fa)	A/T	C	T	A	C/T	G/A	C	T	TT/AA	9/10	A
31	Southern range	Eel	A		C	A	C/T		C	T	TT/AA	9	C
32	Southern range	Eel	T	A	C/T	A	C/T	G/A	C				C

* These SNPs failed assay design

Results & Discussion

A total of 64 variable sites were observed from sequencing 32 individuals in 10 gene fragments. One or two SNPs per gene were submitted for further development into Taqman genotyping assays, and 9 of 11 were successful. Assays were used to genotype samples from three populations (n = 32 per population) including East Fork Salmon River (stream-type adults), Secesh River (stream-type adults), and Hanford Reach (ocean-type adults). Markers were tested for linkage disequilibrium and deviation from Hardy-Weinberg in each population with GENEPOP (Raymond and Rousset 1995). There were no significant deviations from Hardy-Weinberg expectations for any locus or population tested. Significant linkage (corrected $p \leq 0.0104$; Narum 2006) was observed only between assays developed from the same gene. Genotypes from linked loci may be phased to improve statistical power for applications like assignment tests (Smith and Seeb 2008). This project provides additional SNPs that appear to be useful for GSI of Chinook salmon.

Quality Control

Genetic data was tested under standard quality control procedures in CRITFC's genetic laboratory. This includes confirmation of assay genotypes through concordance tests with sequence data, repetitive genotyping, positive and negative controls, and automated allele conversion.

Project Benefits / Monitoring and Evaluation

Both objectives of this study were intended to improve both regional and coastwide GSI applications. The additional SNPs that were discovered will benefit all agencies as these markers are available for the entire genetics community. Lab infrastructure support will bring the CRITFC laboratory closer to providing regional GSI estimates on a time frame and scale that will match other regional labs. The new equipment and database have greatly improved the infrastructure of the Hagerman laboratory.

References

Beacham, T. D., J. R. Candy, K. L. Jonsen, J. Supernault, M. Wetklo, L. Deng, K. M. Miller, and R. E. Withler. 2006b. Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim using microsatellite variation. *Transactions of the American Fisheries Society* 135:861-888.

Narum, SR (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics*, **7**, 783-787.

Raymond, M, and Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Smith CT, Seeb LW (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.