

Single nucleotide polymorphism (SNP) genotyping using the 5'-nuclease reaction provides high-throughput/high-resolution stock identification of Chinook, sockeye, and chum salmon

by

Christian T. Smith, Lisa W. Seeb and James E. Seeb

Alaska Department of Fish and Game
Division of Commercial Fisheries
333 Raspberry Road
Anchorage, AK
USA 99518

Submitted to the

North Pacific Anadromous Fish Commission

by

The United States of America

October 2003

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:

Smith, C. T., L.W. Seeb, and J.E. Seeb. 2003. Single Nucleotide Polymorphism (SNP) genotyping using the 5'-nuclease reaction provides high-throughput/high-resolution stock identification of Chinook, sockeye, and chum salmon. (NPAFC Doc. 733). 11p. Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, AK 99518.

Abstract

Information on the oceanic migration patterns and relative marine survival of individual stocks is critical to our understanding of fluctuations of salmonid populations under changing climatic conditions. Migrations following stock-specific corridors may lead to differing marine survival and varying rates of return among stocks during periods of changing marine conditions. We are developing markers based on the 5'-nuclease reaction to genotype single nucleotide polymorphisms (SNPs) in large number of individuals from high seas collections of Chinook, sockeye, and chum salmon. These markers are being developed in order to utilize the wealth of previously described polymorphisms that have not been applied on large scales due to throughput constraints of older methodologies. Results to date clearly show that SNP genotyping is a rapid, cost effective, and high-resolution approach to baseline development and to the analysis of large numbers of samples from complex mixtures. Time and monetary requirements for running SNP genotyping assays are low relative to those for running other classes of genetic markers, and SNP data are standardized across laboratories and platforms much more readily than are those for other genetic marker classes. We anticipate that SNPs will become an increasingly important tool for stock identification studies of salmonids on the high seas.

Introduction

Pacific salmon of the genus *Oncorhynchus* spawn in lakes and rivers around the North Pacific Basin extending southward to approximately 40°N. Several anadromous species within this genus, including Chinook (*O. tshawytscha*), chum (*O. keta*) and sockeye salmon (*O. nerka*) originating from thousands of drainages throughout their respective ranges, may mix together in large aggregates on the high seas and in near-shore areas. Information on the oceanic migration patterns and relative marine survival of individual stocks is critical to our understanding of fluctuations of salmonid populations under changing climatic conditions. Migrations following stock-specific corridors may lead to differing marine survival and varying rates of return among stocks during periods of changing marine conditions. By comparing genetic attributes of collections of fish taken in high-seas and near-shore areas with those characteristic of potentially contributing stocks one can infer the origin of the collection and thus a point on the migratory route of that stock.

Analyses of genetic variation in aggregate collections of salmon and other fishes have driven the development of a collection of analytical procedures known as genetic stock identification (GSI). Since a collection of salmon taken in the North Pacific Ocean or Bering Sea may contain individuals from rivers throughout the species' range, the allele frequency baselines used to perform GSI must be composed of samples from throughout as much of a species' range as possible. A pre-requisite of genetic markers used for GSI is thus the availability of relatively high-throughput corresponding genotyping technologies. Allozymes and microsatellite DNA are two markers that have met the throughput requirements for various GSI projects (Seeb et al. 1998; Seeb et al. 1999; Seeb et al. 2000; Nelson et al. 2001; Beacham et al. 2003; Nelson et al. 2003) and are thus the two marker types for which the most data are presently available. One difficulty

that has arisen in using genetic markers for which relative electrophoretic mobilities are converted into genotypes (including both allozymes and microsatellites) is standardization of alleles of some loci among laboratories. The geographic ranges of Chinook, sockeye and chum salmon encompasses several political and management jurisdictions making collaboration among the respective management agencies crucial to the effective protection of these species. Interlaboratory standardization of microsatellites alleles is proving time consuming and expensive.

Single nucleotide polymorphisms (SNPs) are a class of genetic marker that have a long history of use in a broad range of taxa, including salmon (e.g. Bickham et al. 1995; Park et al. 1995, 1996; Gharrett et al. 2001). Theoretical considerations that make SNPs appealing as tools for conservation management (reviewed in Moran 2002; Brumfield et al. 2003) include the massive number of potential markers and the distribution these markers throughout genomic regions characterized by a broad spectrum of mutation and selective pressures. This versatility means that SNP data can be used to address a much broader range of questions than can data from more restricted marker classes.

Here we describe our ongoing effort to develop high-throughput SNP genotyping assays in Chinook, chum and sockeye salmon using the 5'-nuclease reaction. We investigate the utility of these assays as management tools by assessing whether or not SNPs previously genotyped using techniques such as PCR-restriction fragment length polymorphism (PCR-RFLP), density gradient gel electrophoresis (DGGE), and DNA sequencing are readily accessible via the 5'-nuclease reaction. We also assess the relative throughput potential for SNP genotyping using the 5'-nuclease reaction.

Methods

Our primary approach to SNP identification has involved sequencing a given locus in several individuals from around the North Pacific Ocean and Bering Sea in hopes of finding a polymorphic nucleotide position. The choice of loci to be examined for SNPs was based on previously reported polymorphisms at each locus in one or more *Oncorhynchus* species (Table 1). Each locus was sequenced in between 40 and 50 individuals taken from the collections shown in Fig. 1. Sequences generated in our laboratory and elsewhere were aligned and examined for SNPs using SeqMan (DNASTAR) and the SNP Pipeline (Buetow et al. 1999).

Once each SNP was identified, oligonucleotide probes and PCR primers for use in 5'-nuclease reactions were designed using either Primer3 (Rozen & Skaletsky), Primer Express (ABI), or Assays-by-Design (ABI). Genomic DNA was extracted from between 48 and 96 individuals from each of the localities shown in Fig. 1 using the Chelex method described by Small et al. (1998). Genotyping assays for each SNP were run on an ABI7900HT real-time sequence detection system in 384-well reaction plates with two wells in each plate as negative controls (no-template) and two wells as positive controls (one for each allele). Pipetting into 384-well plates was done using a BioRobot RapidPlate (Zymark). Thermal cycling was performed on either the ABI7900HT or on a DNA Engine Tetrad (MJ Research). Allelic discrimination analyses were performed using Sequence Detection Software 2.0 (ABI).

Results

We have successfully developed seven SNP assays for sockeye, ten for Chinook, and three for chum salmon (Table 1). All genotypes determined using the SNP assays were identical to those determined based on DNA sequencing. Geographic partitioning of variation at both mitochondrial (Fig. 1) and nuclear loci followed the patterns revealed by surveying these same SNP loci using PCR-RFLP, DGGE or DNA sequencing (citations in Table 1), which are generally similar to those observed for microsatellite and allozyme markers.

The thermal cycling times for these assays were under two hours. Following amplification, it took approximately five minutes to read and analyze a 384-well plate. Using the ABI7900HT and one additional 384-well thermal cycler block three times per day, a single technician was able to genotype 2280 individuals. Since thermal cycling was the limiting factor, the use of the 7900 plus a dedicated DNA Engine Tetrad would increase this number to 5700 genotypes per day.

The relative simplicity of the raw data analysis (Fig. 2) and the lack of an electrophoresis component rendered genotyping based on the present SNP assays faster and thus less expensive than analysis of microsatellites and allozymes (Fig. 3).

Discussion

Conversion of genetic markers from PCR-RFLP, DNA sequencing, and DGGE to high-throughput SNP genotyping assays was successful. These genotyping assays were rapid and inexpensive compared to the original analysis methods used to genotype these loci and therefore greatly increase the potential utility of each for use in fishery management.

We were able to run the present SNP assays in less time (and thus at a lower cost) than the microsatellite or allozyme assays largely because the present SNP genotyping assays require no electrophoresis step and because analysis of raw data generated by the present assays is relatively simple. The number of SNP loci required to perform GSI will depend on the question being addressed, but will likely be larger than the number of markers with greater allele numbers (such as AFLPs or microsatellites). This will impact the relative economy applying SNPs. In the absence of compelling evidence for any simple relationship between number of alleles and utility for GSI work, we agree with Banks et al. (2003) that an empirical approach is an appropriate way to find the optimal combination of markers to use for each program.

A limitation of using the 5'-nuclease assay to genotype SNPs is that each locus is analyzed separately. As the number of SNPs available in salmon species increases, finding a multiplexing solution will become increasingly important. Simultaneous analysis of 20 SNP loci was achieved using DNA microarrays in chum salmon (Moriya et al. 2002), however that protocol required exactly 20 times the PCR volume required for the present assays and would produce data an order of magnitude or more slowly than the present assays. Other SNP multiplexing assays that we would like to investigate in the

future include the oligonucleotide ligation assay (Jarvius et al. 2003) and the use of sequence-tagged molecular inversion probes (Hardenbol et al. 2003).

One final desirable attribute of the present genotyping assays is that in facilitating the detection of only one amplification product per reaction and by eliminating the multitude of potential inconsistencies associated with electrophoresis, SNP genotyping assays should be much easier to transfer among laboratories than are other markers. Furthermore, SNP data collected using the present technology are readily combined with SNP data collected using any other method. This portability, in combination with the relatively low time and monetary requirements for running SNP genotyping assays and the wealth of previously described polymorphisms that may be accessed using these new technologies, suggest that SNPs will become an increasingly important tool for stock identification studies of salmonids on the high seas, where large sample sizes and exhaustive baselines are the norm.

Acknowledgments

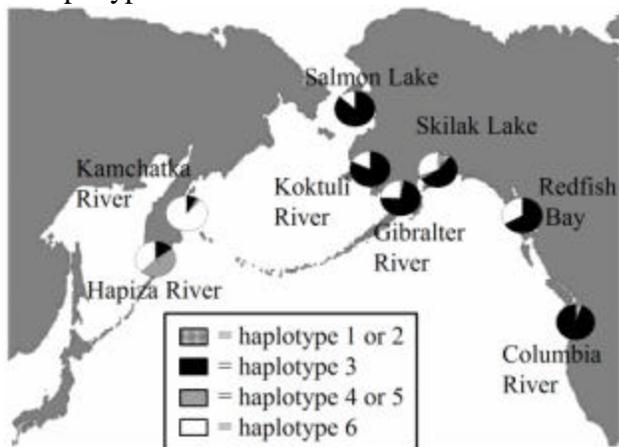
The authors are grateful for technical assistance provided by N. DeCovich, Z. Grauvogel, S. Wilson, G. Johnston, K. Pronzati, T. Dodge, and P. Grass. We would also like to thank M. Powell, S. Narum (Columbia River Inter-Tribal Fisheries Commission), T. Lundrigan and E. Iwamoto (North West Fisheries Science Center) for providing tissue samples. Unpublished DNA sequence data that helped us design some of the Chinook SNP assays were generously provided by Piper Schwenke (NWFSC). This work was funded by US National Marine Fisheries Service Grant NA96FW0196, North Pacific Research Board Project R0205, the Southeast Alaska Sustainable Fisheries Fund, and the Chinook Technical Committee of the Pacific Salmon Commission.

Table 1. Number of SNP genotyping assays developed for sockeye, Chinook, and chum salmon.

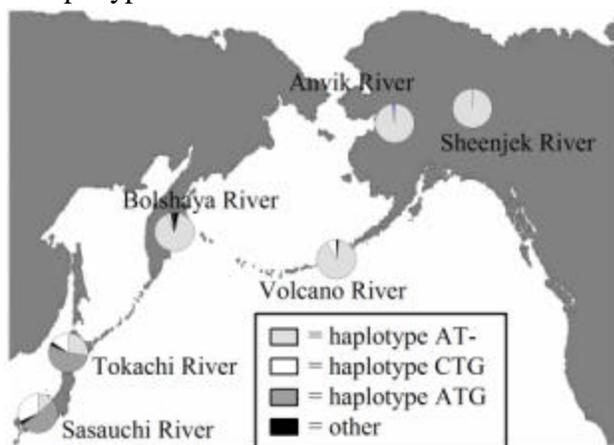
Locus	Sockeye	Chinook	Chum	Source
Mitochondrial DNA loci				
cytochrome c oxidase III		1		Piper Schwenke, NWFSC, unpublished data
cytochrome oxidase I	1			Gharrett et al. 2001
control region	1		3	Shedlock et al. 1992; Abe et al. 2002
cytochrome <i>b</i>	2			Bickham et al. 1995
Nuclear DNA loci				
prolactin II	1	1		Xiong et al. 1992
microsatellite Ots2		1		Blankenship et al. 2002
P53		1		Park et al. 1996
growth hormone II		1		Park et al. 1995
MHC class II	1	1		Grimholt et al. 1994; Miller et al. 2001; Kim et al. 1999
transferrin		1		Ford et al. 1999
somatolactin		1		Ford 1998; Piper Schwenke, NOAA, personal communication
MHC class I		1		GenBank accession #s: AF104583 - AF104588
unknown transcript	1			GenBank accession #s: CD510575, CD510588, CD510544
cytochrome p450A		1		Piper Schwenke, NWFSC, unpublished data
Total	7	10	3	

Fig. 1. Example of haplotype frequencies for mitochondrial DNA SNPs (from Table 1) observed in three *Oncorhynchus* species.

a) sockeye salmon compound cytochrome oxidase I, control region and cytochrome *b* haplotypes



b) chum salmon compound control region position 30, 231 and 386 (Abe et al. 2002) haplotypes



c) Chinook salmon cytochrome oxidase III haplotypes

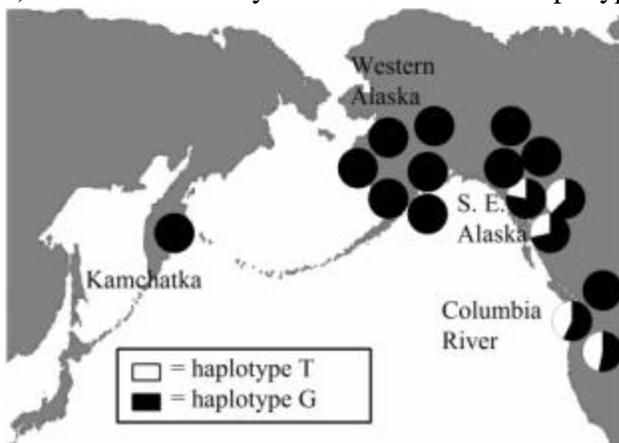


Fig. 2. Example of the data produced by a nuclear SNP genotyping assay. Each dot on the chart below represents an individual fish, whose genotype is determined by where its dot falls on the axis (which indicate the amount of each allele-specific probe cleaved during the course of the assay). The circles indicate clusters of a single genotype each. Mitochondrial assays produce two clusters while nuclear DNA assays produce three clusters. The “x”s represent unreadable samples (air-bubbles, failed PCRs etc.), and the dots in the bottom left corner are the negative controls (no template DNA added).

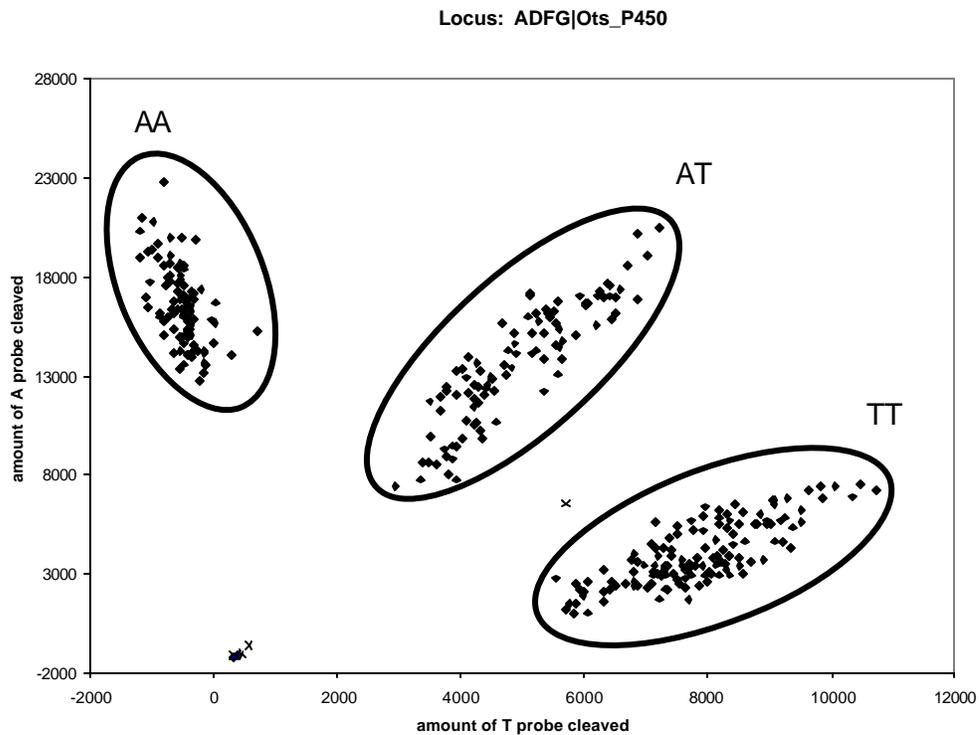
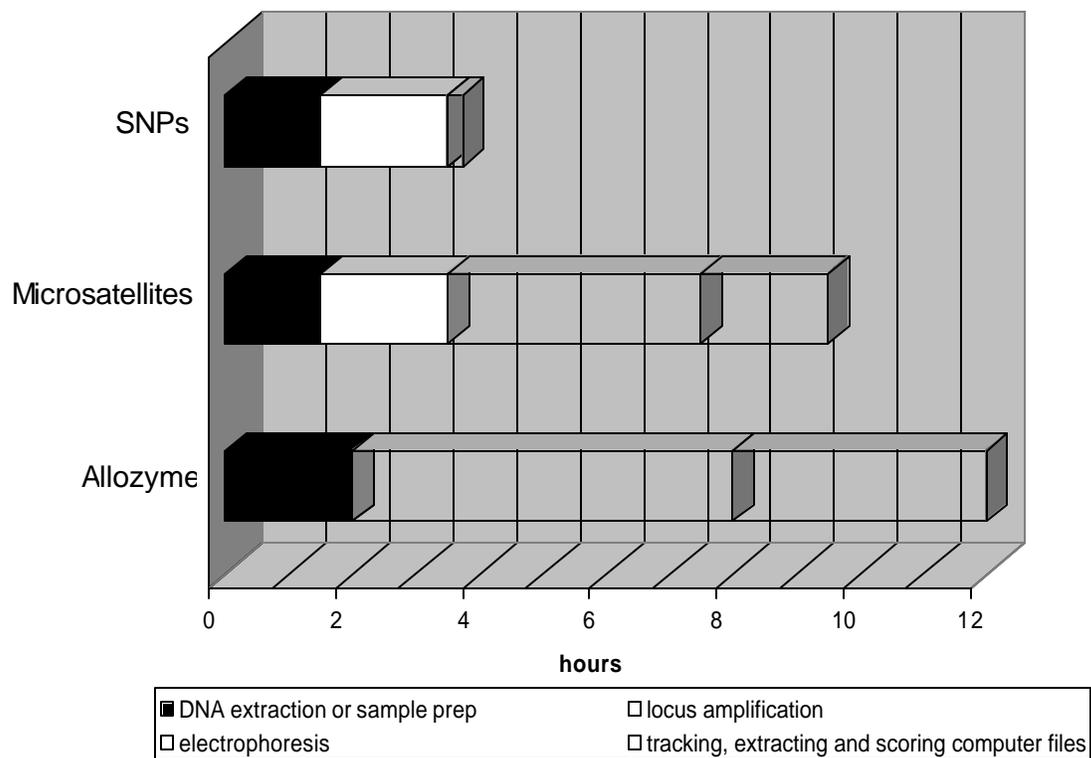


Fig. 3. Time required to analyze a single microtitre plate (380 individuals plus controls) for three different marker types.



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 Cytogenetics, Center for Advanced Science and Technology, Hokkaido University, North 10, West 8, Kita-ku, Sapporo 060-0810, Japan.
- Banks, M.A., Eichert, W. and Olsen, J.B. (2003) Which genetic loci have greater population assignment power? *Bioinformatics* **19**, 1436-1438.
- Beacham, T.D., Candy, J.R., Supernault, K.J., Wetklo, M., Deagle, B., Labaree, K., Irvine, J.R., Miller, K.M., Nelson, R.J. and Withler, R.E. (2003) Evaluation and application of microsatellites for population identification of Fraser River Chinook salmon (*Oncorhynchus tshawytscha*). *Fishery Bulletin* **101**, 243-259.
- Bickham, J.W., Wood, C.C. and Patton, J.C. (1995) Biogeographic implications of cytochrome *b* sequences and allozymes in sockeye (*Oncorhynchus nerka*). *Journal of Heredity* **86**, 140-144.
- Blankenship, S.M., May, B. and Hedgecock, D. (2002) Evolution of a perfect simple sequence repeat locus in the context of its flanking sequence. *Molecular Biology and Evolution* **19**, 1943-1951.
- Brumfield, R.T., Beerli, P., Nickerson, D.A. and Edwards, S.V. (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* **18**, 249-256.
- Buetow, K.H., Edmonson, M.N. and Cassidy, A.B. (1999) Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genetics* **21**, 323-325.
- Ford, M.J. (1998) Testing models of migration and isolation among populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolution* **52**, 539-557.
- Ford, M.J., Thornton, P.J. and Park, L.K. (1999) Natural selection promotes divergence of transferrin among salmonid species. *Molecular Ecology* **8**, 1055-61.
- Gharrett, A.J., Gray, A.K. and Brykov, V. (2001) Phylogeographic analysis of mitochondrial DNA variation in Alaskan coho salmon, *Oncorhynchus kisutch*. *Fishery Bulletin* **99**, 528-544.
- Grimholt, U., Olsaker, I., de Vries Lindstrom, C. and Lie, O. (1994) A study of variability in the MHC class II beta 1 and class I alpha 2 domain exons of Atlantic salmon, *Salmo salar* L. *Animal Genetics* **25**, 147-53.
- Hardenbol, P., Baner, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U. and Davis, R.W. (2003) Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnology* **21**, 673-678.
- Jarvis, J., Nilsson, M. and Landegren, U. (2003) Oligonucleotide ligation assay. In: Kwok, P.-Y., (Ed.) *Single Nucleotide Polymorphisms Methods and Protocols*. Totowa, New Jersey: Humana Press Inc. 215-228.
- Kim, T.J., Parker, K.M. and Hedrick, P.W. 1999. Major histocompatibility complex differentiation in Sacramento River Chinook salmon. *Genetics* **151**: 1115-22.
- Miller, K.M., Kaukinen, K.H., Beacham, T.D. and Withler, R.E. (2001) Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* **111**, 237-57.

- Moran, P. 2002. Current conservation genetics: building an ecological approach to the synthesis of molecular and quantitative genetic methods. *Ecology of Freshwater Fish* **11**: 30-55.
- Moriya, S., Abe, S., Urawa, S., Suzuki, O. and Urano, A. (2002) DNA microarray for rapid detection of mitochondrial haplotypes of chum salmon. (NPAFC Doc. 617) 8 pages. Research and Development Center, Nisshinbo Industries, Inc. 1-2-3, Onodai, Midori-ku, Chiba, 267-0056, Japan
- Nelson, R.J., Small, M.P., Beacham, T.D. and Supernault, K.J. (2001) Population structure of Fraser River Chinook salmon (*Oncorhynchus tshawytscha*): an analysis using microsatellite DNA markers. *Fishery Bulletin* **99**, 94-107.
- Nelson, R.J. Wood, C.C., Cooper, G., Smith, C.T., and Koop, B. (2003) Population structure of sockeye salmon of the central coast of British Columbia: implications for recovery planning. *North American Journal of Fisheries Management* **23**:703-720.
- Park, L.K., Moran, P. and Dightman, D.A. (1995) A polymorphism in intron D of the Chinook salmon growth hormone 2 gene. *Animal Genetics* **26**, 285
- Park, L.K., Moran, P. and Dightman, D.A. (1996) A Chinook salmon PCR-RFLP marker in the p53 locus. *Animal Genetics* **27**, 121-128.
- Rozen, S. and Skaletsky, H.J. (1997) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.
- Seeb, J.E., Habicht, C., Olsen, J.B., Bentzen, P., Shaklee, J.B. and Seeb, L.W. 1998. Allozyme, mtDNA, and microsatellite variants describe structure of populations of pink and sockeye salmon in Alaska. *North Pacific Anadromous Fisheries Commission Bulletin No. 1*, 300-318.
- Seeb L.W. and Crane PA (1999) 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., Habicht C., Templin W.D., Tarbox K.E., Davis R.Z., Brannian L.K. and Seeb J.E. (2000) Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the Exxon Valdez oil spill. *Transactions of the American Fisheries Society* **129**, 1223-1249.
- Shedlock, A.M., Parker, J.D., Crispin, D.A., Pietsch, T.W. and Burmer, G.C. (1992) Evolution of the salmonid mitochondrial control region. *Molecular Phylogenetics and Evolution* **1**, 179-192.
- Small, M.P., Beacham, T.D., Withler, R.E. and Nelson, R.J. (1998) Discriminating coho salmon (*Oncorhynchus kisutch*) populations within the Fraser River, British Columbia, using microsatellite DNA markers. *Molecular Ecology* **7**, 141-155.
- Xiong, F., Chin, R.A. and Hew, C.L. (1992) A gene encoding Chinook salmon (*Oncorhynchus tshawytscha*) prolactin: gene structure and potential cis-acting regulatory elements. *Molecular Marine Biology and Biotechnology* **1**, 155-64.