

Development of a Standardized DNA Database for Chinook Salmon

L. W. Seeb,
A. Antonovich,
M. A. Banks,
T. D. Beacham,
M. R. Bellinger,
S. M. Blankenship,
M. R. Campbell,
N. A. Decovich,
J. C. Garza,
C.M. Guthrie III,
T. A. Lundrigan,
P. Moran,
S. R. Narum,
J. J. Stephenson,
K. J. Supernault,
D. J. Teel,
W. D. Templin,
J. K. Wenburg,
S. F. Young, and
C. T. Smith



Seeb is research professor at the School of Aquatic and Fisheries Sciences, University of Washington, Seattle. She can be contacted at lseeb@u.washington.edu. Antonovich is a biometrician, Decovich is a fishery biologist, and Templin is a fisheries geneticist at the Alaska Department of Fish and Game, Division of Commercial Fisheries, Anchorage. Banks is marine fisheries geneticist and associate professor and Bellinger is a molecular geneticist at Oregon State University's Hatfield Marine Science Center, Newport. Beacham is a research scientist and Supernault is a research technician at the Department of Fisheries and Oceans Pacific Biological Station in Nanaimo, British Columbia. Blankenship was formerly with the National Marine Fisheries Service Southwest Fisheries Science Center in Santa Cruz, California, and is now a geneticist at the Washington Department of Fish and Wildlife, Olympia. Campbell is senior fisheries research biologist at the Idaho Department of Fish and Game, Eagle Fish Genetics Laboratory, Eagle. Garza is supervisory research biologist, National Marine Fisheries Service Southwest Fisheries Science Center, Santa Cruz, California. Guthrie is a fishery research biologist at the National Marine Fisheries Service Auke Bay Laboratories, Ted Stevens Marine Research Institute, Juneau, Alaska. Lundrigan was formerly with the University of Washington School of Aquatic and Fishery Sciences, Seattle, and is currently a fisheries biologist at the University of Victoria's Centre for Biomedical Research, Victoria, British Columbia. Moran is a research geneticist and Teel is supervisory fishery biologist at the National Marine Fisheries Service Northwest Fisheries Science Center, Seattle, Washington. Narum is lead geneticist and Stephenson is laboratory manager at the Columbia River Inter-Tribal Fish Commission, Hagerman Fish Culture Experiment Station, Hagerman, Idaho. Wenburg is director of genetics at the U.S. Fish and Wildlife Service in Anchorage, Alaska. Young is a fish and wildlife research scientist at the Washington Department of Fish and Wildlife, Olympia. Smith was formerly with the Alaska Department of Fish and Game Division of Commercial Fisheries and is now a fisheries geneticist at the U.S. Fish and Wildlife Service, Abernathy Fish Technology Center, Longview, Washington.

ABSTRACT: An international multi-laboratory project was conducted to develop a standardized DNA database for Chinook salmon (*Oncorhynchus tshawytscha*). This project was in response to the needs of the Chinook Technical Committee of the Pacific Salmon Commission to identify stock composition of Chinook salmon caught in fisheries during their oceanic migrations. Nine genetics laboratories identified 13 microsatellite loci that could be reproducibly assayed in each of the laboratories. To test that the loci were reproducible among laboratories, blind tests were conducted to verify scoring consistency for the nearly 500 total alleles. Once standardized, a dataset of over 16,000 Chinook salmon representing 110 putative populations was constructed ranging throughout the area of interest of the Pacific Salmon Commission from Southeast Alaska to the Sacramento River in California. The dataset differentiates the major known genetic lineages of Chinook salmon and provides a tool for genetic stock identification of samples collected from mixed fisheries. A diverse group of scientists representing the disciplines of fishery management, genetics, fishery administration, population dynamics, and sampling theory are now developing recommendations for the integration of these genetic data into ocean salmon management.

Desarrollo de una base de datos estandarizada de DNA para el salmón rey

RESUMEN: Se realizó un proyecto internacional con la participación de diversos laboratorios con la finalidad de desarrollar una base de datos estandarizada de DNA para el salmón rey (*Oncorhynchus tshawytscha*). Dicho proyecto surgió como respuesta a las necesidades del Comité Técnico Chinook de la Comisión del Salmón del Pacífico para identificar la composición poblacional del salmón rey que es capturado por la pesquería durante su migración. Un total de nueve laboratorios de análisis genéticos identificaron y reprodujeron cada uno 13 loci microsatélites. Con el objeto de probar que dichos loci fueran reproducibles entre laboratorios, se condujeron pruebas anónimas para verificar la consistencia de casi 500 alelos. Una vez estandarizada, se construyó una base de datos construida con información proveniente de más de 16,000 salmones que representan 110 poblaciones putativas distribuidas a lo largo del área de interés de la Comisión del Salmón del Pacífico, del sureste de Alaska hasta el Río Sacramento, California. La base de datos sirve tanto para identificar genéticamente los distintos stocks de salmón rey a partir de muestras combinadas provenientes de la pesquería como para diferenciar el linaje genético conocido más importante de esta especie. En la actualidad, un importante grupo de científicos especializados en disciplinas como el manejo y administración de pesquerías, genética, dinámica poblacional y teoría del muestreo están desarrollando recomendaciones para que esta base de datos genéticos se incorpore en el manejo del salmón.

INTRODUCTION

The Pacific Salmon Treaty was ratified between the United States and Canada in 1985, renegotiated in 1999, and extended to the Yukon River in 2002. Through the treaty, the two nations agreed to cooperate in the management, research, and enhancement of Pacific salmon. Pacific salmon migrate long distances during their marine period and are routinely intercepted in fisheries beyond the jurisdiction of the government in whose waters they spawn. The Pacific Salmon Treaty, through the Pacific Salmon Commission (PSC), serves as a means to coordinate management of the salmon resource and conduct conservation actions as required.

Chinook salmon (*Oncorhynchus tshawytscha*) are harvested throughout the year by commercial and sport fishers in the waters of southeast Alaska, British Columbia, and the Pacific Northwest. Fisheries typically harvest highly mixed stocks of Chinook salmon and are therefore under the jurisdiction of the Pacific Salmon Treaty. Quotas specified by the PSC are dependent on the abundance of Chinook salmon projected by its Chinook Technical Committee (CTC) using the Chinook salmon model (e.g., CTC 2005).

The model uses catch, escapement, coded-wire tag (CWT) recovery, and recruitment information to forecast relative abundance in treaty fisheries.

With the increased dependence on CWT recovery data, concern has been raised regarding the quality of the CWT data and inferences drawn from those data (Hankin et al. 2005). Historically only fish carrying a CWT had an adipose clip. However, with the advent of mass-marking of large numbers of hatchery fish using an adipose clip only (e.g., Mobrاند et al. 2005), recovery of CWT fish has been complicated. Recovery now requires both handling much larger numbers of individuals paired with electronic scanning of those adipose-clipped individuals to detect tags. As a further concern, abundance and harvest information is not available for most Chinook salmon stocks, so indicator stocks are used within the model to represent both larger groups of hatchery and wild stocks. The ability of the hatchery indicator stocks to accurately represent wild stocks has been generally supported in coho salmon (*O. kisutch*; e.g., Weitkamp and Neely 2002), but is still largely unknown for many stocks of Chinook salmon. The wild stocks may differ in ancestry, abundance, and timing

from the indicator stock assigned to them in the model.

In 2004, the Pacific Salmon Commission convened a panel of experts to examine limitations of the CWT program in the context of mass marking and mark-selective fisheries, as well as to evaluate the capacity of alternative technologies to improve assessment of Chinook salmon (Hankin et al. 2005). The expert panel concluded that alternative technologies could not by themselves replace CWTs, but that genetic stock identification (GSI) could indeed complement the existing CWT programs or be used in combination with other techniques such as otolith thermal marking to estimate the stock composition of a landed catch in a particular time/area fishery (Findings 11-13). However, they noted limitations in the GSI methods due to the lack of coastwide genetic baselines, which are databases of genotypes from breeding populations.

Finding 13. *Modern GSI methods can be used to estimate the stock composition of the landed catch in a particular time/area fishery. However, the accuracy and precision of data required to estimate stock-age-fishery specific exploitation rates using GSI*

methods is dependent upon a variety of factors. For example, microsatellite DNA-based GSI technology is not yet capable of generating consistent, replicable estimates due to the lack of a coastwide genetic baseline, the history of stock transfers within and among watersheds, and differences in methodologies and mixture separation algorithms. (Hankin et al. 2005)

Here we review a large multi-laboratory effort to develop and standardize a replicable DNA database for Chinook salmon.

History and need for coast-wide databases

Beginning in the 1980s, considerable effort and coordination was directed towards developing standardized allozyme (protein) baselines (e.g., Shaklee and Phelps 1990; White and Shaklee 1991; see Box 1). Standardization among laboratories involves identifying and adopting a common suite of loci (specific polymorphic DNA segments or their products) and adopting consistent names for the various alleles (see Box 1). Genetic analyses using allozyme databases were used extensively to estimate the stock contribution of Chinook salmon fisheries in the Columbia River, coastal Washington, and Strait of Juan de Fuca (e.g., Marshall et al. 1991; Shaklee et al. 1999). Collaborative work during the 1990s by multiple state, provincial, and federal agencies was directed towards enlargement of the database (Teel et al. 1999). The allozyme database grew to include comprehensive coverage of populations ranging from California through Alaska with representative populations from Russia (Teel et al. 1999). Although

this database was comprehensive and used by multiple laboratories, a number of limitations, including the requirements for lethal sampling, cryopreservation, lack of laboratory automation, and the finite number of loci, led researchers to the decision to replace the allozyme baseline with a DNA database.

Markers based on DNA, such as mitochondrial DNA (Cronin et al. 1993) and microsatellites (Banks et al. 2000; see Box 1), were also in use and shown to be valuable for resolving Chinook salmon population structure. However, unlike allozymes where a system was developed for sharing and transferring data among laboratories, microsatellite baselines evolved for use within single laboratories (Moran et al. 2006). The large number of available microsatellite loci resulted in little overlap among researchers' datasets. For example, at the beginning of 2003 over 60 loci were in use for Chinook salmon, but most (43) were used in only a single laboratory (Figure 1a). This was not limited to Chinook salmon, but was the norm for fisheries studies of many species. Furthermore, even when identical loci were being used, differences in chemistries and instrument platforms among laboratories produced variable size calls for the same allele from the same individual, which precluded easily merging datasets from multiple regions (Figure 1b). As a result, despite collection of a substantial amount of microsatellite data (e.g., Nelson et al. 2001; Beacham et al. 2003), progress in replacing the allozyme database was slow for multi-jurisdictional fisheries, and multiple databases proliferated.

Concerns regarding these multiple and independent databases were addressed at

various meetings of salmonid geneticists from 1999 to 2001, but progress was limited (LaHood et al. 2002). Finally, a workshop hosted by the CTC in 2003 resulted in funding for the database described herein. In the following year, a symposium on genetic databases for fishery management and conservation was held at the 2004 American Fisheries Society (AFS) Annual Meeting, and groups studying a number of divergent fishery resources began standardization efforts (e.g., Atlantic herring; Mariani et al. 2005). So although the advantages of DNA-based markers, and microsatellites in particular, were well documented beginning in the early 1990s (e.g., Wright and Bentzen 1994; Wirgin and Waldman 1994), coordination among laboratories lagged markedly.

The present database was constructed and evaluated through a two-year collaborative effort of an international multi-agency work group. Guiding principles required that the database would: (1) be subject to review by scientists from all interested agencies, (2) be freely available to all researchers managing or studying Chinook salmon, and (3) cover the range of Chinook salmon at a geographic scale appropriate to the management objectives of the PSC with the understanding that the database could be easily expanded to include the entire range of the species. It was anticipated that the database would allow for a wide variety of management and research applications on all life history stages throughout the species range in both freshwater and marine environments.

Box 1. Definition of terms commonly used in genetic stock identification of Pacific salmon.

Allele	An alternative form of a given gene or DNA sequence that differs from other alleles in DNA sequence or phenotype.
Allelic ladder	A pooled and diluted aliquot of multiple PCR products that originates from multiple individuals of known genotype. With the choice of appropriate individuals, a ladder with all the "rungs" necessary for successful interlaboratory allele indexing can be created.
Allozyme	Allelic form of a protein enzyme encoded at a given locus. Allozymes are usually distinguished by protein electrophoresis and histochemical staining techniques.
Locus (loci, plural)	The site that a particular gene or DNA sequence occupies on a chromosome.
Microsatellite	DNA sequences containing short (2–5 base pairs) repeats of nucleotides (e.g. GTGTGTGT).
Neighbor joining tree	A bottom-up clustering method used for the creation of phylogenetic trees that is based on the distance between each pair of populations or taxa.
PCR	The polymerase chain reaction or PCR amplifies a single or few copies of a piece of DNA across several orders of magnitude, generating millions of copies of the DNA.
SNP	Single nucleotide polymorphism; DNA sequence variation occurring when a single nucleotide (A, T, C, or G) differs between members of a species or within an individual between paired chromosomes.

METHODS AND MATERIALS

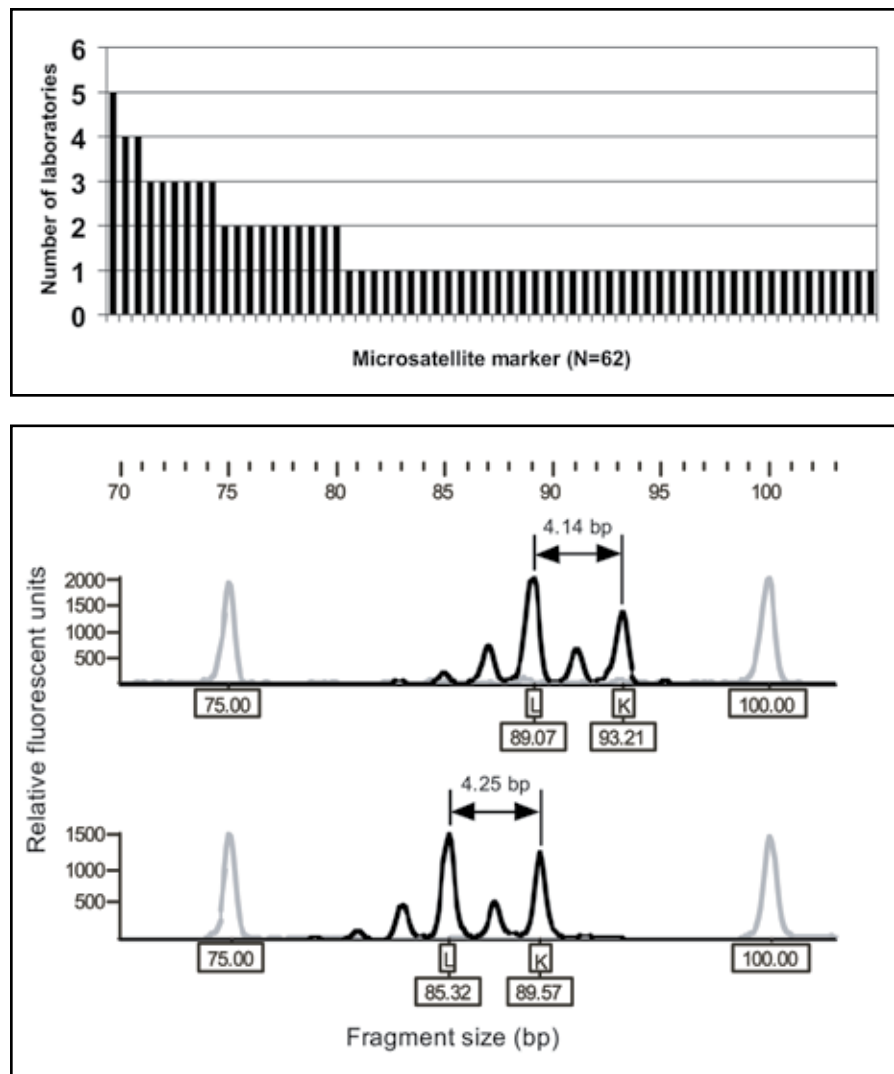
Identification of loci

During the first year of the study, each contributing laboratory analyzed an identical set of 500 individuals drawn from populations ranging from California, USA, to Kamchatka, Russia, for its own microsatellite locus panel. A total of 60 loci were evaluated, and “sponsorship” documents were created for each locus including description, source, analysis conditions, and locus variability. A subset of 25 loci drawn from the combined set was chosen based on qualitative selection criteria including the reliability of the PCR (polymerase chain reaction; see Box 1) amplification, and, to a lesser extent, allelic size range and number of alleles. Locus diversity across the range was generally not a consideration since too few individuals were assayed to reliably compute statistical measures. Each laboratory tested the 25 loci on a subset of 96 individuals drawn from the original 500 individuals. Scores for these 96 individuals were submitted to the coordinating laboratory (Northwest Fisheries Science Center, National Marine Fisheries Service). From this set, collaborators selected and tested the 15 most reliable loci for inclusion in the standardized baseline. A curator laboratory was identified for each locus with responsibility for receiving, compiling, and distributing allele information. A decision was also made to link alleles to tissues from a specific individual salmon that would reside in each laboratory. An additional set of locus documentation was prepared designating recognized alleles, assigning allelic nomenclature, and designating individuals for every recognized allele at each locus. These master allele lists were termed “curator documents.”

Genotyping evaluation

To assess the ability of each laboratory to accurately resolve all 15 loci, 2 sets of 96 “blind” samples were analyzed (first and second blind samples). These samples had never been previously genotyped in any of the laboratories. In both cases, test fish were drawn from marine fisheries and assumed to be highly mixed with broad representation of stocks. The samples for the first test were collected from British

Figure 1. a. Distribution of loci analyzed across laboratories at the beginning of the study. Most markers were assayed in only a single laboratory. Only three loci were used in more than half of the laboratories. b. Electropherograms from the same individual assayed on two different instruments. Identical size standards and sizing algorithm were used, yet differences in size are seen that reflect differences in amplification chemistries and electrophoresis hardware, polymer and running conditions employed in different laboratories. The figure illustrates a significant difference in the estimated size of the alleles, in addition to a different relative size of the repeat unit (i.e., 4.14 versus 4.25 bp; adopted from Moran et al. 2006).

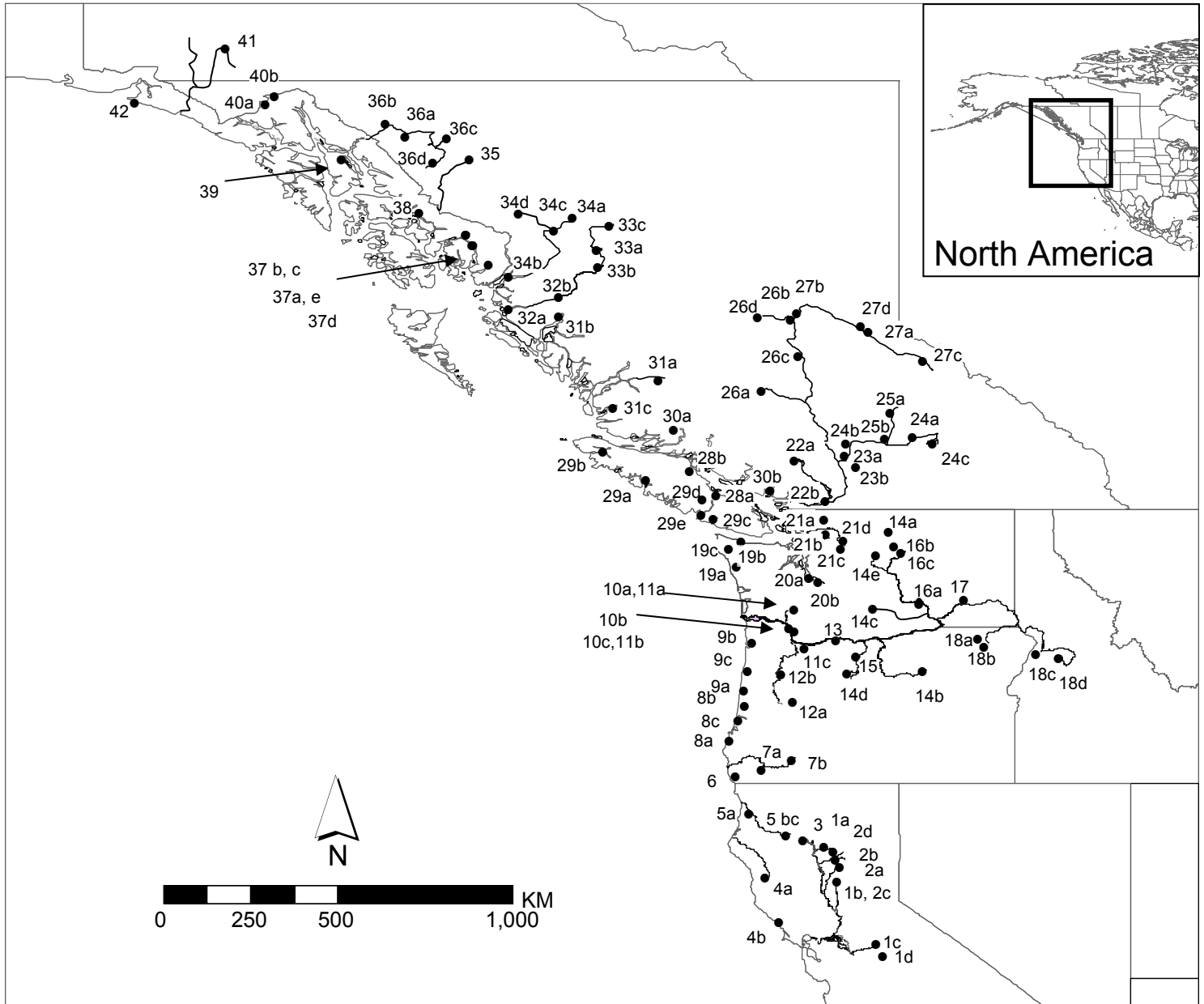


Columbia; the second set was collected from southeast Alaska. Allelic scores from each laboratory were adjusted to the recognized nomenclature, submitted to the coordinating laboratory, and compared across seven laboratories. Two additional laboratories that had not been involved in locus selection or the first blind test participated in the second blind test. The alleles were not sequenced, so the absolute length was unknown. Therefore, the modal score (the most common allele scored across the participating laboratories) was defined to be the “correct” score. Concordance of scores across laboratories or percent accuracy for each locus for each laboratory was based on this modal score.

Construction of the Database

Baseline populations were chosen that represent all previously identified genetic lineages of Chinook salmon from the southern end of the species range north to southeast Alaska with focus on major production areas and likely contributors to Pacific Salmon Treaty fisheries (Figure 2). Six laboratories (Alaska Department of Fish and Game [ADFG]; Department of Fisheries and Oceans Canada [DFO]; Columbia River Inter-Tribal Fish Commission [CRITFC]; Oregon State University [OSU]; Southwest Fisheries Science Center, National Marine Fisheries Service [SWFSC]; and Washington Department of Fish and Wildlife [WDFW])

Figure 2. Collection sites for Chinook salmon included in the DNA baseline. These sites represent populations potentially contributing to mixtures harvested in Pacific Salmon Treaty fisheries. Sites are designated by regional groups from Table 3.



contributed to the construction of the database.

Evaluation of the database

Definition of regional groups. The database was evaluated for its ability to correctly allocate to regional groups defined based on a combination of genetic similarity, geographic features, and management applications. Population estimates are combined into regional groups to provide the desired accuracy and precision of stock composition estimates. The initial list of regional groups for compositional analyses was enlarged from those used for the allo-

zyme baseline (Teel et al. 1999). In order to visualize how the present microsatellite data might be concordant with and/or improve upon this list, we calculated pairwise chord distances (Cavalli-Sforza and Edwards 1967) between all collections and then used PHYLIP (Felsenstein 2004) to create a neighbor joining tree (Saitou and Nei 1987; see Box 1). On such a tree, genetically similar populations will cluster together, facilitating the definition of regional groups.

Mixture simulation. Simulations were conducted to evaluate the application of these regional groups to genetic stock identification (compositional analyses) of

mixtures of Chinook salmon harvested in treaty fisheries. These simulations assessed whether the baseline of microsatellite allele frequencies provided sufficient information to identify regional groups in hypothetical mixtures.

Simulations were performed using the Statistical Package for Analyzing Mixtures (SPAM version 3.7; Debevec et al. 2000). Mixture genotypes were randomly generated from the baseline allele frequencies assuming Hardy-Weinberg equilibrium. The baseline allele frequencies were parametrically resampled to account for sampling variability. Each simulated mixture ($N = 400$) was entirely composed (100%)

of the regional group under study with equal contribution of all populations within the regional group. Then each simulated mixture was analyzed against the complete baseline. Bootstrap means for regional groups and 90% confidence intervals were derived from 1,000 simulations per group. Regional groups with mean correct estimates of at least 90% are considered highly identifiable in potential mixtures from treaty fisheries. Regional groups with mean correct estimates lower than 90% can still be considered identifiable in mixtures, but sources of misallocation should be considered when interpreting the results (e.g., Seeb et al. 2000).

RESULTS

Locus selection and evaluation of microsatellites

The first blind test revealed that data-handling errors, such as misalignment of individuals, were the largest source of disagreement among laboratories. In addition to data-handling errors, 2 of the 15 loci selected and tested stood out as being poorly standardized across 3 or more laboratories. These 2 loci were removed, and the final group of 13 loci was identified for inclusion in the baseline (Table 1). Beyond these sources of conflict, the results of the first blind comparison indicated that technical standardization had been accomplished for greater than 90% of the observed alleles at each locus. Furthermore, much of the remaining variability among laboratories was due to alleles being observed in the blind samples that had not been observed in the original 96 samples.

The results of the second blind test indicated that technical standardization had been accomplished for more than 95% of the observed alleles (Table 2a). However, a number of low-concordance values for specific loci indicated errors in the data from two of the laboratories. One laboratory obtained a concordance of only 44% for *Ogo4*, whereas other loci provided by the same laboratory showed 98% concordance or more. Another laboratory had 0% concordance at five loci and 82% at another locus. Data for the other seven loci from this second laboratory were close to the overall average. cursory checks in respective laboratories revealed data-handling errors that explained most of the discrepancies. After correction of these errors

the overall concordance for the laboratories at these loci was over 99% (Table 2b).

Baseline construction

Data provided by the collaborators were combined to create a large database suited to GSI compositional analyses of fisheries managed under the Pacific Salmon Treaty. A total of 220 sample collections and 16,394 individual fish representing 110 putative populations were included in Version 1.1 of the baseline (Table 3, Figure 2, Appendix 1). All individual fish were assayed across 13 loci; however, failures at particular loci sometimes resulted in incomplete multilocus genotypes. A target was set of 144 individuals per population with at least 120 genotypes per locus. In some populations, fewer than 120 individuals were available, so actual sample sizes were necessarily below the target.

The number of alleles varied markedly across loci, ranging from 9 (*Ots9*) to 74 (*Omm1080*; Table 1). Across all loci, 487 alleles were observed. A voucher set of tissues from a single individual for each of the alleles were identified and provided to all collaborators. For some alleles, insufficient tissue was available for sharing between all labs, so a second individual known to exhibit the same allele was substituted as the voucher. Additional alleles are expected as coverage of the baseline expands.

Simulations

Forty-two regional groups were defined based on genetic similarity, geographic features, and management applications. Simulations were performed to examine the accuracy and precision of this baseline simulating mixtures composed of individuals drawn entirely from populations of a single region. Almost all regions were defined as highly identifiable with bootstrap means of 1,000 simulated mixtures above 90% to the correct regional group (Table 3). Exceptions were the Deschutes Fall and Upper Stikine River regions which had 89.5% and 84.5% mean correct allocations, respectively. In addition, the lower bound of the 90% confidence interval for all regions was above a 90% threshold, with the exception of the Deschutes Fall, Upper Stikine River, Taku River, and Southeast Alaska Stikine River groups.

DISCUSSION

Database construction

Our primary goal of developing a replicable set of microsatellite loci for Chinook salmon was successfully completed. With an average above 90% correct allocation to regional groups, the DNA baseline provides an unprecedented resource for improved information and management of Chinook salmon during their ocean and freshwater life stages. Additional laboratories employing various chemistries and hardware will be able to use the database through the use of voucher specimens or by using the allelic ladders (see Box 1) that are currently under construction for all loci (e.g., LaHood et al. 2002). Allelic ladders consist of a collection of alleles covering a range of known sizes that can be used as an internal measurement to standardize the data. These ladders hold the promise of simplifying standardization as laboratories will not need to analyze voucher specimens for every allele.

All geographic regions and genetic lineages likely to contribute to fisheries of CTC interest are represented in the present baseline, and these data should be appropriate for complex fishery mixtures that include diverse populations from widespread locations. Although the current baseline is broad, it is not comprehensive. Efforts are currently underway to increase local coverage. Expanded baseline data will likely improve the accuracy of allocation to the regional groups, and, in at least some cases, are likely to provide a finer scale of estimation (e.g., sub-basins within major river systems). Fine-scale geographic allocation of mixtures and potential assignment of individual fish to population-of-origin may also provide important biological and life-history information such as migration timing and pathways and age-related changes in habitat use. Expansion of the baseline to include populations throughout Alaska, Canada, and Russia is underway. In addition to improved management, the baseline will be used to better understand the biology of the species and provide information about effective population size, evolutionary and demographic history, and population boundaries.

The results of the two blind tests were encouraging, but also signalled a need for continued vigilance in error checking and data manipulation. On the one hand,

Table 1. Microsatellite loci standardized for Chinook salmon. Reference, curator agency, and observed number of alleles are given for baseline Version 1.1.

Locus	Reference	Curator agency ¹	Observed number of alleles
<i>Ogo2</i>	Olsen et al. 1998	ADFG	26
<i>Ogo4</i>	Olsen et al. 1998	WDFW	20
<i>Oki100</i>	DFO unpublished ²	DFO	48
<i>Omm1080</i>	Rexroad et al. 2001	SWFSC	74
<i>Ots201b</i>	Grieg et al. 2003	ADFG	52
<i>Ots208b</i>	Grieg et al. 2003	CRITFC	54
<i>Ots211</i>	Grieg et al. 2003	ADFG	43
<i>Ots212</i>	Grieg et al. 2003	OSU	36
<i>Ots213</i>	Grieg et al. 2003	OSU	48
<i>Ots3M</i>	Grieg and Banks 1999	WDFW	19
<i>Ots9</i>	Banks et al. 1999	DFO	9
<i>OtsG474</i>	Williamson et al. 2002	CRITFC	19
<i>Ssa408</i>	Cairney et al. 2000	NWFSC	39

¹ Laboratory abbreviations: OSU, Oregon State University; SWFSC, Southwest Fisheries Science Center—National Marine Fisheries Service; DFO, Department of Fisheries and Oceans Canada; NWFSC, Northwest Fisheries Science Center—National Marine Fisheries Service; CRITFC, Columbia River Inter-Tribal Fish Commission; ADFG, Alaska Department of Fish and Game; WDFW, Washington Department of Fish and Wildlife.

² Personal communication, K. Miller, Department of Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada.

laboratories were able to achieve a very high degree of concordance through the standardization process and voucher sample comparisons. Furthermore, two new laboratories that had not been previously involved in locus selection, baseline construction, or the first blind test were able to achieve very high concordance without any significant errors. On the other hand, significant errors in data handling were still apparent in initial blind submissions from two of the original laboratories. These results, although not unique to microsatellite data (e.g., White and Shaklee 1991), suggest that ongoing quality control and error checking procedures both internally within each laboratory and among laboratories are warranted to ensure database integrity.

Table 2. Proportional genotyping accuracy by laboratory and locus for the second blind test of 13 microsatellite loci. Averages across locus and laboratory are given. A. Results as submitted. B. Results after correction for errors (see text).

A.

Locus	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	Lab 9	Average
<i>Ogo2</i>	0.987	1.000	1.000	1.000	0.993	0.000	1.000	0.993	1.000	0.886
<i>Ogo4</i>	0.439	1.000	1.000	0.995	1.000	0.000	0.995	0.994	0.990	0.824
<i>Oki100</i>	0.978	1.000	1.000	1.000	1.000	1.000	1.000	0.970	1.000	0.994
<i>Omm1080</i>	1.000	1.000	0.995	1.000	1.000	0.000	1.000	0.994	1.000	0.888
<i>Ots201b</i>	0.984	1.000	1.000	1.000	1.000	0.993	0.995	0.985	1.000	0.995
<i>Ots208b</i>	0.994	1.000	1.000	1.000	1.000	1.000	0.995	0.970	0.995	0.995
<i>Ots211</i>	1.000	1.000	0.994	1.000	0.993	0.955	0.994	0.985	0.994	0.991
<i>Ots212</i>	0.989	1.000	1.000	1.000	1.000	0.989	0.995	0.994	1.000	0.996
<i>Ots213</i>	0.987	1.000	0.982	1.000	0.985	0.000	1.000	1.000	1.000	0.884
<i>Ots3M</i>	1.000	1.000	0.988	0.994	1.000	0.000	1.000	1.000	0.995	0.886
<i>Ots9</i>	1.000	1.000	1.000	1.000	1.000	0.823	1.000	1.000	1.000	0.980
<i>OtsG474</i>	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
<i>Ssa408</i>	0.987	0.929	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991
Average	0.949	0.995	0.997	0.999	0.998	0.597	0.998	0.991	0.998	0.947

B.

Locus	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	Lab 9	Average
<i>Ogo2</i>	0.987	1.000	1.000	1.000	0.993	0.988	1.000	0.993	1.000	0.996
<i>Ogo4</i>	0.994	1.000	1.000	0.995	1.000	0.968	0.995	0.994	0.990	0.993
<i>Oki100</i>	0.978	1.000	1.000	1.000	1.000	1.000	1.000	0.970	1.000	0.994
<i>Omm1080</i>	1.000	1.000	0.995	1.000	1.000	0.938	1.000	0.994	1.000	0.992
<i>Ots201b</i>	0.984	1.000	1.000	1.000	1.000	0.993	0.995	0.985	1.000	0.995
<i>Ots208b</i>	0.994	1.000	1.000	1.000	1.000	1.000	0.995	0.970	0.995	0.995
<i>Ots211</i>	1.000	1.000	0.994	1.000	0.993	0.955	0.994	0.985	0.994	0.991
<i>Ots212</i>	0.989	1.000	1.000	1.000	1.000	0.989	0.995	0.994	1.000	0.996
<i>Ots213</i>	0.987	1.000	0.982	1.000	0.985	0.994	1.000	1.000	1.000	0.994
<i>Ots3M</i>	1.000	1.000	0.988	0.994	1.000	0.949	1.000	1.000	0.995	0.992
<i>Ots9</i>	1.000	1.000	1.000	1.000	1.000	0.979	1.000	1.000	1.000	0.998
<i>OtsG474</i>	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
<i>Ssa408</i>	0.987	0.929	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991
Average	0.992	0.995	0.997	0.999	0.998	0.981	0.998	0.991	0.998	0.994

Table 3. Mean correct allocations from 100% simulations for 42 regional groups of Chinook salmon populations. Bootstrap means of 1,000 simulated mixtures and upper and lower 90% bootstrap confidence intervals are given. Number of populations per region is also given.

Region number	Region name	Number of populations	Bootstrap mean	90% confidence interval	
				Lower	Upper
1	Central Valley fall	4	0.945	0.916	0.971
2	Central Valley spring	4	0.936	0.908	0.962
3	Central Valley winter	1	0.990	0.980	0.998
4	California Coast	2	0.985	0.975	0.995
5	Klamath River	3	0.984	0.970	0.995
6	North California/South Oregon Coast	1	0.972	0.956	0.987
7	Rogue River	2	0.941	0.912	0.968
8	Mid Oregon Coast	3	0.944	0.918	0.968
9	North Oregon Coast	3	0.957	0.933	0.979
10	Lower Columbia spring	3	0.971	0.952	0.987
11	Lower Columbia fall	3	0.973	0.956	0.987
12	Willamette River	2	0.983	0.968	0.994
13	Mid Columbia Tule fall	1	0.969	0.949	0.986
14	Mid and Upper Columbia spring	5	0.965	0.945	0.985
15	Deschutes fall	1	0.895	0.859	0.931
16	Upper Columbia summer/fall	3	0.963	0.937	0.985
17	Snake River fall	1	0.946	0.915	0.974
18	Snake River spring/summer	5	0.966	0.946	0.984
19	Washington Coast	3	0.951	0.928	0.971
20	South Puget Sound	2	0.988	0.977	0.997
21	North Puget Sound	4	0.971	0.955	0.985
22	Lower Fraser	2	0.984	0.972	0.994
23	Lower Thompson River	2	0.983	0.971	0.994
24	South Thompson River	3	0.976	0.962	0.990
25	North Thompson River	2	0.978	0.964	0.990
26	Mid Fraser River	4	0.980	0.966	0.991
27	Upper Fraser River	4	0.969	0.949	0.986
28	East Vancouver Island	2	0.979	0.966	0.991
29	West Vancouver Island	5	0.990	0.980	0.997
30	South BC Mainland	2	0.975	0.961	0.988
31	Central BC Coast	3	0.960	0.940	0.978
32	Lower Skeena River	2	0.950	0.927	0.971
33	Upper Skeena River	3	0.950	0.925	0.971
34	Nass River	4	0.939	0.913	0.963
35	Upper Stikine River	1	0.845	0.793	0.893
36	Taku River	4	0.920	0.884	0.954
37	Southern Southeast Alaska	5	0.969	0.950	0.986
38	Southeast Alaska, Stikine River	1	0.917	0.884	0.947
39	King Salmon River	1	0.987	0.977	0.995
40	Chilkat River	2	0.988	0.977	0.996
41	Alek River	1	0.980	0.967	0.992
42	Situk River	1	0.977	0.963	0.990

Power of the database

Our simulation results suggest that the present database provides greater resolution than was provided by the allozyme baseline. Only the Deschutes Fall and Upper Stikine regions fell below our target of 90% accuracy. Additional populations and tests are needed to evaluate whether these groups can be accurately identified or, alternatively, should be combined with adjacent regional groups. The similarity among the Taku and Stikine river trans-boundary populations was also observed with the allozyme dataset of Teel et al. (1999), and a single combined group for

these populations was used in the allozyme study of Guthrie and Wilmot (2004).

The neighbor-joining tree (Figure 3) suggested that some groups that had not been distinct from one another based on the allozyme baseline might be resolvable with the present microsatellite data. Overall, 37 regional groups were resolvable with the allozyme dataset (Teel et al. 1999) while 42 were recognized in this microsatellite dataset despite that fact that more populations were included in the allozyme baseline than the microsatellite baseline. Gains in resolution were most apparent in the California Central Valley where a

single allozyme group was separated into fall, spring, and winter groups.

Growth of and access to the baseline

More detailed analyses of the power of the database are ongoing. We expect future analyses to better define limitations of the present database and to identify where the addition of baseline collections and increased power via the addition of new markers are most needed. The addition of single nucleotide polymorphism (SNP) markers (see Box 1) which are easily standardized and may reflect selective

Figure 3. Neighbor-joining tree based on pairwise chord distances between collections of Chinook salmon (Appendix 1). Numbers and vertical lines on right indicate regional groups as designated in Table 3. Regional groups with only a single population are in bold.



variation (Smith et al. 2005; Smith et al. in press) is presently underway.

As part of the collaboration, a web-based application has been developed to allow the fisheries and research community to access the genotype data, curator documents, and supporting metadata. The Pacific Salmon Commission provided developmental funding to the Northwest Fisheries Science Center, National Marine Fisheries Service, Seattle, Washington (www.nwfsc.noaa.gov/research/divisions/cbd/standardization.cfm), and a test version of the web application is now being evaluated. The permanent location and support for the database are currently under discussion by the collaborating agencies and the Pacific Salmon Commission. Until the final location of the database is established, the data are available at the ADFG website (www.genetics.cf.adfg.state.ak.us/publish/data/PSCchinookver1_1.pdf).

CONCLUSIONS

Here we described a large collaborative study to develop a shared, web-accessible database of DNA markers to allow multiple agencies to conduct genetic stock identification studies of Chinook salmon. Replicate scoring of the microsatellite loci was successfully accomplished among nine geographically-dispersed laboratories. However, the study also revealed that ongoing coordination and error-checking will be required to ensure the integrity of the database as it expands. Growth of the database is presently underway via the addition of new collections and additional SNP markers. The database is now being used to estimate compositional analyses of PSC and other fisheries around the Pacific Ocean.

In May and September of 2007, the PSC convened a workshop focusing on the current and future capabilities, limitations, and use of genetic stock identification methods in ocean salmon management. A diverse group of scientists attended representing the disciplines of fishery management, genetics, administration, population dynamics, and sampling theory. Attendees were divided into four workgroups: genetics, management, logistics, and modeling/sampling. Reviews of the workshop are available at the PSC website (<http://psc.org/>). The final recommendations will be published when completed in 2008.

As genetic approaches to fisheries management become more commonplace, we anticipate the growth of this and similar types of databases that are publicly available through web applications. Ultimately, these databases will provide a wealth of information not only for fisheries applications, but also for researchers from a variety of disciplines investigating diverse aspects of salmonid life history, population genetics, and evolutionary theory.

ACKNOWLEDGEMENTS

This study was result of a collaboration among three National Marine Fisheries Service laboratories—Northwest Fisheries Science Center, Auke Bay Laboratory, and Southwest Fisheries Science Center; Washington Department of Fish and Wildlife; University of Idaho/Columbia River Inter-tribal Fisheries Commission/Idaho Department of Fish and Game; Oregon State University; Canadian Department of Fisheries and Oceans; and the Alaska Department of Fish and Game. The U.S. Fish and Wildlife Service, Anchorage, Alaska, and Idaho Department of Fish and Game, Eagle, Idaho, participated in the second blind sample. Primary funding for the study was provided by the U.S. Department of Commerce, NOAA, NMFS, funds appropriated to fund the technical work required for the U.S. section of the Pacific Salmon Treaty to implement the Chinook Salmon

Your Tags



Your Way

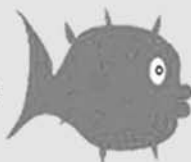
FLOY TAG

The World Leader & Innovator in Fish Tags - For Over 50 Years

- Shellfish, Lobster & Crustacean Tags
- T-Bar Anchor Tags, Spaghetti Tags, Dart Tags & More
- Net, Trap & Line Tags
- Laminated Disc and Oval Tags
- Dart, Fingerling, Streamer, Intramuscular Tags
- Guns and Tag Applicators, Extra Needles, etc.

...and almost any other kind of custom tagging solution you might need.

"Why Risk Your Research To The Copy-Cats



...When You Can Have The Original?"

**View our latest catalog at www.floytag.com, or email us at: sales@floytag.com
or call to discuss your custom tagging needs: **(800) 843-1172****