PSC Genetic Stock Identification Workshop
May and September 2007

Logistics Workgroup

Final Report and Recommendations

Logistics Workgroup Members:

Bill Ardren	USFWS	Bill Johnson	ADFG
John Candy	DFO	Richard Kang	NMFS, Seattle
Brodie Cox	WDFW	George Nandor	PSMFC
Doug Dehart	USFWS	Eric Schindler	ODFW
Carlos Garza	NMFS, Santa Cruz	Lisa Seeb	ADFG
Allen Grover	CDFG	Eric Volk	ADFG
Denise Hawkins	WDFW	Bruce White	PSC

Logistics Workgroup Coordinator

Kenneth Johnson (Beaverton, OR)

October 22, 2007

TABLE OF CONTENTS

Preface

The Steering Committee for the PSC Genetic Stock Identification (GSI) Workshop (May and October, 2007) tasked the Logistics Workgroup to provide recommendations for several logistical aspects relating to the potential incorporation of genetic data into the management of ocean salmon fisheries. Standardization issues were foremost in terms of focus and importance. This included standardization of protocols used regionally for the collection and curation of genetic tissue samples and DNA extraction products. The second key area involved standardization issues related to establishing and maintaining a coastwide GSI database to support fishery management information needs.

This report outlines those recommendations. Denise Hawkins (geneticist – WDFW) served as Task Leader in specifically addressing differences between baseline and fishery samples, where standardized protocols are required, commodities to store, sample collection, preservation and storage methods, DNA extraction protocols, and storage duration. Likewise, John Candy (geneticist – DFO) served as Task Leader in developing recommendations for standardization of GSI data reporting and development of a coastwide GSI database.

A third independent report was prepared by George Nandor (PSMFC) in response to the Logistics Workgroup's recommendation that the Pacific States Marine Fisheries Commission be asked to host a new coastwide database for GSI data. His proposal discusses the long experience of PSMFC in managing a number of coastwide databases and its widely recognized data neutrality. Plans are outlined for the migration of the NMFS's GAPS baseline database to PSMFC and the expected costs of maintaining the new database for both baseline and fishery mixture samples.

Lastly, two other independent reports were prepared. Both address the basic disadvantage that GSI data presently can not provide age which is essential for stock-age-fishery specific estimates of exploitation rates. Eric Volk (ADFG) provided a review of the otolith thermal marking technology and explored its potential as a compliment to genetics in coastwide salmon management. In particular, hatchery of origin and brood year could be identified with a regional otolith thermal marking program of key indicator hatcheries.

Carlos Garza (NMFS), in turn, proposed a novel solution for the 'ageing problem' encountered with present GSI technology. His report discusses how 'Parental-based Tagging' would be able to provide the necessary age information as well as hatchery of origin. This is a bold new approach that will require further research and evaluation but could potentially serve as an alternative to CWTs for at some fishery management applications.

Findings of each report are presented in the Executive Summary, along with their key recommendations.

Executive Summary

I. GSI Sample Collection and Curation: Issues and Recommendations

Standardization of protocols can occur at many different steps and on many different levels. Sample collection protocols include the sampling design, the tissue collected, and the method of collection. There are also several methods of DNA extraction that can be used. Likewise, there are several options for what will actually be stored, the duration of storage, and the method of storage. Key points are summarized below, followed by recommendations.

In general, standardization is more important for baseline samples than for fishery mixture samples. The reason is that the two sample types are fundamentally different and thus have different requirements for standardization of collection, processing, and curation. Baseline samples are 'known origin' samples collected to genetically characterize populations that could contribute to a mixed stock fishery. These samples are required as references for assignment of stock of origin for mixed stock fishery samples, and have a high potential for reanalysis as new genetic markers are developed. In addition, there is often a wide variety of associated data collected such as location and date of catch, age, sex, length and other relevant biological data.

In contrast, fishery mixture samples are 'unknown stock-of-origin' samples collected as either landed or encountered fish from a group of mixed stocks during a fishery. These samples are compared to the baseline samples for determination of their most likely stock of origin for calculations of stock composition of the catch and estimations of fishery impacts on specific stocks. Fishery samples are rarely used for other purposes or reanalyzed.

Lastly, there are several ways to extract DNA. Chelex extractions and silica membrane purification columns are the most common methods in use. The chelex method is inexpensive and quick to carry out, but the extracted DNA is not purified and is not suitable for archiving. However, the chelex extractions perform well for most fishery sample applications. Silica membrane purification columns are substantially more expensive and time consuming, but the resulting DNA is purified and of high quality to facilitate archiving and sharing the DNA.

Summary Recommendations

The Logistics Workgroup recommends the following regarding collection and curation of samples for genetic analysis used in the management of ocean salmon fisheries:

1. Fishery mixture samples
    - representative of the fishery, and include at-sea sampling to provide catch location
    - bulk collection is fine, unless the genetic data will be used in conjunction with other information from individual fish (CWTs or scale ages, for instance)
    - chelex extraction method is fine
    - standardization of tissue collection and DNA extraction methods not necessary
    - store at a minimum until genotyping is completed and management applications are complete and agreed to by parties

2. Baseline samples
    - representative of the population
    - individual collection and storage to allow association with biological data
    - silica membrane purification column extraction to produce archival DNA
    - store indefinitely whenever possible

3. Tissue type
    - non-lethal sample to allow sampling of released fish
    - scales not recommended for genetic analysis, but good for ageing
    - fin or axillary process best tissue type (more important for baseline samples)

4. Storage and Curation
    - tissue and DNA both highly valuable and should be retained
    - no storage method recommended over another; alcohol dehydration most common
    - dry storage should be tested further for stability

II. Standardized GSI Reporting: Issues and Recommendations

The Logistics Workgroup was also tasked with recommendations on standardization issues relating to the development of a coastwide GSI database for use in ocean salmon management. This report addresses the three levels of GSI standardization necessary for significantly improving identification of salmon stocks taken in mixed stock fisheries for inter-agency use.

1. Standardization of inter-laboratory genotyping (allele calling)

At the lowest denominator, standardization issues for GSI data include inter-laboratory standardization of allele calling. Depending on their laboratory capabilities and preferences, agencies have developed different genetic markers types and sets of loci for GSI identification of salmon stocks. There are a variety of valid reasons for this. The complication, however, is that all laboratories must run the same suite of loci using standardized allele calling in order to reproduce exactly the same allelic "calls" for any given sample, independent of the instrument platform used. The data produced by a given laboratory must meet agreed standards for inter-laboratory concordance before being uploaded in a standardized format for the coastwide database (Moran et al 2005). High concordance has been obtained for both STR and SNP data.

The GAPS (Genetic Analysis of Pacific Salmon) database, housed at NMFS-Seattle, is the first coastwide baseline database with a standard set of microsatellite markers for Chinook salmon which can be run by all those who are part of the GAPS consortium (currently 12 laboratories). This database is now contains Chinook salmon genotypes from 13 loci, collected from 170 spawning ground locations (baseline) representing populations from south-eastern Alaska to California. Efforts are underway to expand the database to include the SNP baseline from the GAPS Chinook salmon collaboration and to include more baseline genotypes from other salmonid species for both STP and SNPs. At this point, there are no plans to include mixed-stock fishery data, but this would be possible with minimal changes to the current database schema.

2. Standardized coastwide GSI database

The concept for a coastwide GSI database requires four basic components:

1) The standardization of data formats which then creates the 'common language' necessary for sharing and understanding data collected by multiple entities on a regional scale.
2) A single entity selected to house and manage the database.
3) Strict data "neutrality" is required by the host site to ensure full confidence by all in data shared coastwide. This would require an agency that does not do either fishery management or data analysis to ensure data neutrality. This requirement would preclude long-term hosting of shared interagency data at NOAA due to perceived political bias that might affect access to the data or data integrity.
4) Additionally, the warehoused coastwide data must be readily available to all by Web-based data retrieval, including GIS supported applications to provide recovery location of baseline and fishery samples.

At the same time, it must be recognized that agencies have invested heavily in building GSI baseline and fishery mixture databases which might not become standardized:

1) Agencies may continue to maintain their respective data in internal formats that meet the respective need of their own programs and operations.
2) Agencies are required to develop the necessary software applications to convert their internal data sets into the coastwide formats approved for coastwide data exchange.
3) Agencies may need additional staff support to maintain the conversion software and provide their respective data sets to the coastwide data center on a timely basis.
4) In sum, standardization imposes a cost that must be recognized as an integral part of overall interagency co-management.

3. Standardization of GSI data protocols and coding with the coastwide CWT database

Both CWT and GSI data provide information on distribution and abundance of stock and stock complexes. As such, the Logistics Workgroup recognized that it may be advantageous *where possible,* to adopt CWT data standards established through PSC following the Pacific Salmon Treaty in 1985. These data standards and formats are well known and understood by researchers and fisheries managers alike. In addition, many CWT data fields are fully equivalent to GSI data fields and thus readily transferable. As such, the 'well seasoned' CWT data protocols can provide considerable assistance in establishing a comparable coastwide GSI database.

It must be noted that there are also differences between the CWT and GSI data sets where data fields are not similar or have different meaning. Estimates of fishery contribution, for example, are computed differently. Likewise, CWT production areas are largely based on historical management requirements while GSI regions are based on genetic similarity stock groupings.

The good news is that these differences are not a 'show stopper'. There are solutions available for likely most of the discrepancies that now exist. However, it will require the work of key personnel from both the fisheries and genetic communities to be tasked with designing the GSI database and integrate it to the degree possible with the existing CWT database formats.

Recommendations of the Logistics Workgroup:

1. Continued development of the standardized GAPS baseline to accept additional marker types and species.
   a) NMFS-Seattle laboratory is highly commended for outstanding development of the web-based GAPS database and data retrieval applications.
   b) Funding support for a coastwide GSI database should continue to be a high priority.
   c) Agency laboratories should accelerate efforts to standardize allele callings for a number of species with existing STR databases.

2. Further discussion recommended as to the future location of a coastwide GSI database.
   a) PSMFC is recommended as the preferred host site based on long experience in hosting regional databases and demonstrated data neutrality.
   b) This decision was not without some dissent and raised concerns that should be addressed in a larger forum. Issues raised are:
      - NMFS has the expertise and infrastructure in place to continue to serve as the host site and thus allow substantial efficiencies and cost savings.
      - Estimated NMFS costs for hosting the site are reportedly 2-3 fold lower than that of PSMFC. This raises the decision if data neutrality is worth PSMFC's extra cost.
   c) There is no support for retaining the GAPS baseline data at one site (i.e. NMFS) while developing a fishery mixed-stock database at a neutral site.

3. Further discussion recommended as to the configuration of a coastwide GSI database.
   a) The GAPS database should be modified to include fishery mixture sample data.
   b) There was no consensus on whether the new fishery mixture data would include both individual and mixture assignments.
   c) There is strong interest by some that individual assignments also be made available in the GSI database for fishery managers and other researchers. As such, this topic will need further exploration as to its practicality.

4. Further evaluation of CWT data standards and protocols for adoption for standardized GSI database.
   a) The challenges of establishing a coastwide GSI database have strong parallels to the development of the CWT database. As such, it is strongly recommended that the new GSI database incorporate all of those CWT data parameters that can be used.
   b) Key personnel from both the fisheries and genetics communities will be needed to merge the two databases to the degree possible. It will be a complicated process at some levels of detail.
   c) It is recommended that this effort be done under the PSC umbrella and by a Data Standards Workgroup dedicated to GSI data, similar to that done when standardizing the CWT data formats.
   d) Additionally, the warehoused coastwide data must be readily available to all by web-based data retrieval, including GIS supported applications to provide recovery location of baseline and fishery samples.

III. **Pacific States Marine Fisheries Commission's Possible Future Role in Hosting a GSI** Database for Fisheries Resource Management on the Pacific Coast

The Logistics Workgroup endorsed PSMFC (with some disagreement) as the organization to host and maintain a coastwide GSI database system for salmon stock identification in West Coast fisheries. PSMFC is willing to accept this task provided that the necessary funding can be found. The database would be maintained by the Regional Mark Processing Center (RMPC) which has administered the coastwide CWT database for over three decades.

1. Qualifications in Terms of Experience and Neutrality

The RMPC was designated in 1987 as the U.S. site to maintain the primary CWT database to expedite data exchange between the U.S. and Canada. Staff played a key role in helping standardize all CWT data formats through the PSC Data Standards Workgroup. The RMPC's role also includes regional coordination of CWT related marking programs.

PSMFC also maintains a number of other important regional data sharing projects and programs that are used to manage West Coast fisheries. Equally important, PSFMC is well known as a neutral site in terms of data management as it does not carry out data analyses nor manage any fisheries that are linked to the hosted databases.

2. Migration of NMFS's GAPS Database to PSMFC

The NMFS has done an outstanding job in developing the web-based GAPS database using Oracle. It is reportedly seamlessly portable to PSFMC, requiring only an Oracle license. As such, the port will be relatively easy as the RMPC has utilized Oracle for many years for managing the CWT database.

3. Design of the New GSI Database

RMPC staff would rely on the expertise of salmon geneticists and other fisheries scientists to help guide the formation of the database's details and the management and content of the genetic baseline and fishery mixture data sets. Standardization of the data fields in the database ideally would be done through the formation of a GIS Data Standards Workgroup, very similar to the current PSC Data Standards Workgroup tasked with standardizing CWT protocols and data exchange formats.

4. Cost Estimates

The estimated cost for 2008 would be $317,339. This includes start-up capital outlay costs for computer hardware and software, office supplies and furnishings, etc. Costs in 2009 and 2010 are estimated at $217,000 and $230,000, respectively.

IV. Use of Otolith Thermal Marking as a Compliment to Genetics in Coastwide Salmon Management: Issues and Recommendations

This report addresses the challenges of a coastwide otolith thermal marking program to compliment genetic stock identification methods across a spectrum of potential applications.

Otolith thermal marking is a widely used technique in the Pacific Rim nations of Canada, Japan, Russia and the United States for identifying hatchery-released salmonids by inducing structural patterns to their otoliths using short-term water temperature manipulations. The method is a practical means for 100% marking of hatchery salmon populations, offering distinct advantages over individual tagging of fish. Over a billion juvenile salmon (mainly pink and chum) are now marked annually in Alaska, Washington, and Canada.

1. Range of Otolith Thermal Marking Options

There is a wide range of options for using otolith thermal marks in conjunction with genetic analyses, ranging from the application of a single mark among all index hatcheries and brood years to unique marking of all hatcheries and brood years. However, there likely aren't sufficient patterns to uniquely mark all west coast hatcheries over several brood years.

At one end of the spectrum, a thermal marking program only offers information to confirm the fish's origins as a hatchery fish. At the other end, unique brood specific marks at each index hatchery would resolve specific sample origins and ages of hatchery origin fish.

2. Number of Otolith Marks Available

While at least 1000 otoliths patterns are possible for Chinook salmon, practical limitations associated with hatchery operations, fish development and visual recognition sharply reduce the actual number of available patterns. However, less than 200 marks are needed to provide four brood specific marks to all ~40 Chinook index hatcheries on the West Coast. This number is well within the practical limits of the otolith thermal marking technique.

3. Coordination

Given the broad use of otolith thermal marking programs today, any new multi-jurisdictional thermal marking plan linked with genetic stock identification will face important challenges in coordinating those efforts. This coordination will need to maintain control of assignment of all thermal mark codes as well as ensure that the marking program doesn't negatively impact other exiting programs on the West Coast as well as other international programs. It will be essential that a single coordinating body be established for both coordination and data management.

4. Costs

A significant cost issue for a broad scale otolith thermal marking program involving many hatcheries is that most will require some modification and upgrade to deliver specific water temperature events to large numbers of fish simultaneously. Small scale, portable operations

have been established for under $25,000 USD, but these systems usually require delivery of thermal events to several portions of the hatchery production in succession. This can, in effect, produce multiple marks at a given hatchery which increases issues associated with the number of marks and distinguishing between similar marks. As such, it is unwise to adopt thermal marking without adequate capital investment in the hatcheries for this purpose. A conservative estimate for capital upgrades to all Chinook index hatcheries would be $2-4 Million USD.

5. Recommendations

1. Establish objectives of the program in terms of how many distinct groups (hatcheries x brood years) will require unique marks for each species.

2. Identify funding for capital investment to hatcheries so that physical plant systems have the best chance of delivering a reliable and clear otolith thermal mark. This will reduce a variety of problems associated with detecting mark patterns and undoubtedly reduce errors in that process. The amount of that capital investment is tied to the overall objectives of the program, i.e. the number of unique marks required among all groups.

3. Utilize existing regional otolith laboratories for processing and recovery of otolith marks. Labs in Alaska, Canada and Washington State may be adapted to handle the increased sample load from this endeavor. These facilities are accustomed to rapid turn around times for samples, though additional funding would obviously have to accompany this increased load. Current prices for otolith sample processing and mark recovery range between $7 and $10 per sample.

4. Establish a coordinating body for assignment and quality control of otolith marking. An existing body within the NPAFC can serve as a model. In addition, close communication with this body should be maintained. An otolith mark database must be established as a stand alone or extension of existing mark or tag databases.

V. Parental-based Tagging as an Alternative to CWTs for Fishery Management

Carlos Garza (NMFS) presented Parental-based Tagging (PBT) as a 'bold' new approach that could potentially yield a considerable amount of information on not only stock origins but also age. This approach is based on genotyping parents, which can efficiently 'tag' a few thousand offspring with each parental genotyping. The numbers of loci needed increases linearly as the magnitude of the application increases. Thus this approach may be possible for some regional ocean fisheries. In this approach, increasing the number of loci is more important than maximizing the number of alleles to detect Mendelian incompatibilities between parents and offspring in the fishery. Large scale applications would require about 100 SNP loci, for example. With declining costs of SNP analysis, this approach may be feasible on a large scale in a few years, especially in areas where most fish are produced in hatcheries (e.g. California Chinook).

The method does not require information on individual matings, as a 'day bucket' of spawners yields required genotypes. A test of this method is underway at Feather River Hatchery on the Sacramento River in California. Even if not all individuals could be assigned to parents, the data can still be used to estimate stock proportions. If successful, one attractive feature of this method is that the ages of identified fish in the fishery are known.

This method, however, has several drawbacks. One is that since parents need to be genotyped, it would be applicable to only hatchery stocks. Another is that there is often a large error associated with collecting all parents used in hatchery spawning. Perhaps the largest problem would be the recovery of individuals to be genotyped in the ocean fisheries. Presently 20% of Chinook with clipped adipose fins are examined for CWTs. It would be prohibitive to screen 20% of marked fish for PBT. At present, releases for U.S. federal hatcheries and juveniles with CWTs are adipose clipped. One remedy might be to first estimate regional stocks with a reduced number of loci, then to fully genotype fish from hatcheries of interest. It was noted, however, that resorting samples is time consuming and can lead to error.

(Note: The above summary of Parental-based Tagging was taken from Stewart Grant's report on Carlos Garza's presentation to the Genetics Workgroup during the May, 2007 GSI Workshop in Portland, OR).

Carlos Garza's subsequent presentation during the second GSI Workshop in September 2007 (provided here-in) provides additional information on the PBT method and also addresses drawbacks.

I. GSI Sample Collection and Curation: Issues and Recommendations

Denise Hawkins (WDFW) – Task Leader
Logistics Workgroup

1.0 Introduction

The Logistics Workgroup of the Pacific Salmon Commission Genetic Stock Identification (GSI) Workshop was tasked to provide recommendations for several aspects of the collection and curation of samples and products integral to the incorporation of genetic data into the management of ocean salmon fisheries. This report outlines those recommendations, specifically addressing differences between baseline and fishery samples, where standardized protocols are required, commodities to store, sample collection, preservation and storage methods, DNA extraction protocols, and storage duration.

2.0 Differences between Baseline and Fishery Samples

2.1 Baseline Samples

Baseline samples are known origin samples collected to genetically characterize populations that could contribute to a mixed stock fishery. These samples are required as references for assignment of stock of origin for mixed stock fishery samples. Because these samples are of known origin, they are often useful for other analyses such as determination of population structure, monitoring of anthropogenic effects, and studies of population changes through time.

Baseline samples also have a high potential for reanalysis as new genetic markers are developed. It is cheaper and easier to genotype samples already collected than to resample populations, especially non-hatchery populations. Furthermore, as new markers are developed, there is a need to test for linkage with old markers (thus generating redundant information). This can only be done by genotyping all markers on the same individuals. In addition, there is often a wide variety of associated data collected with each individual baseline sample such as location and date of catch, age, sex, length and other relevant biological data.

2.2 Fishery Samples

Fishery samples are unknown stock-of-origin samples collected as either landed or encountered fish from a group of mixed stocks during a fishery. These samples are compared to the baseline samples for determination of their most likely stock of origin for calculations of stock composition of the catch and estimations of fishery impacts on specific stocks. The fishery samples are collected for the specific purpose of stock composition determinations and are rarely used for other purposes or reanalyzed after the stock composition estimates have been made. Depending on where and how the samples are collected, the exact location of catch for each sample is sometimes not known and individual biological data is sometimes not collected.

Due to the above outlined differences between baseline and fishery samples, the requirements for standardization of collection, processing and curation differ between the two types of samples.

3.0 Standardization

Standardization of protocols can occur at many different steps and on many different levels. Sample collection protocols include the sampling design, the tissue collected, and the method of collection. The specific sampling design should be determined by the questions being asked. Once the samples have been collected, there are several methods of DNA extraction that can be used. And finally, there are several options for what will actually be stored, the duration of storage, and the method of storage.

In general, standardization is more important for baseline samples than for fishery samples.

4.0 What to Collect and Store

The workgroup considered collection and storage requirements for: 1) the actual tissue sample, 2) the DNA extracted from that sample, and 3) the products generated through the process of the polymerase chain reaction (PCR) – the enzymatic process used to interrogate specific areas of the DNA. The PCR reactions for common genetic assays generate fluorescently labeled fragments that lose their intensity with long-term storage. They are useful only until the data are scored and verified, and we do not recommend a storage standard for PCR products. Individual laboratory-determined standard operating procedures regarding PCR products are sufficient.

In contrast, both the actual sample and the extracted DNA are highly valuable for long-term storage. Both can be easily shared among labs and can be used to generate data from multiple marker types.

5.0 Sample Type, Collection, and Preservation

Considerations concerning the sample include the quantity, tissue, collection method and storage method.

5.1 Quantity of Samples

The quantity of genetic samples required is determined by the number and type of analyses to be carried out. For baseline samples, it is important to collect enough material for reanalysis and sharing among labs.

5.2 Tissue Options

DNA can be extracted from many types of samples; Table 1 outlines the most common tissues collected and the advantages and disadvantages of each. Of the five tissues listed, collection of opercle punches, axillary processes, and fin clips provides the additional advantage of creating a short- to medium-term visual mark to identify fish that have been previously sampled. While scales can be very useful for determining age of fish, there are some issues with low yield of DNA and increased risk of cross-sample contamination that make them less valuable as samples for genetic analysis. Although DNA can be extracted from organ tissues, these samples

generally require lethal sampling methods and can't be used for analysis of fish that will be released. Therefore, we recommend collecting either fins or axillary processes for baseline samples due to their ease of collection and reliability for storage, whereas many tissue types are acceptable for fishery samples.

5.3 Tissue Collection Methods

There are also many methods of sample collection (Table 2). Which method is chosen is sometimes determined by how the samples will be used, with some of the methods outlined in Table 2 limited to specific types of samples. For baseline samples, we recommend the collection and storage of individual rather than pooled samples because of the ability to associate other biological data to individual samples, but bulk collections will work well for fishery samples if individual data are not associated with the samples. Many labs are beginning to test methods of collecting samples on paper to minimize the sampling supplies needed in the field and reduce handling effort in DNA extraction, however, these methods have not been in use long enough to evaluate the long-term stability of the stored samples.

5.4 Preservation and Storage Methods

Once the samples have been collected, options for preservation and storage exist with associated advantages and disadvantages (Table 3). Freezing samples is not suitable for long-term storage and repeated use because the freeze-thaw cycle accelerates the degradation of the DNA. However, freezing is the most common method of storing extracted DNA.

Dehydration of tissue samples with alcohol is the most common method of preservation and provides reliable long-term storage at room temperature. However, some labs are testing the reliability of dry storage for tissue samples (extracted DNA can also be stored dry) to decrease space and curation requirements.

6.0 DNA Extraction Methods

Depending on the type of tissue and collection method, there are several ways to extract DNA. Table 4 outlines five of the most common methods and compares their cost, handling and processing time, yield, and quality of DNA produced. Currently, chelex extractions and silica membrane purification columns are the most common methods in use. The two methods differ in cost, time and DNA quality. Chelex extractions are inexpensive and quick to carry out, but the DNA that is produced is not purified from other cellular components and is not suitable for archiving. Silica membrane purification columns are substantially more expensive and time consuming, but the resulting DNA is purified and of high quality. It is possible to automate some steps of the column extractions which can reduce the hands-on time required.

The Logistics Workgroup recommends silica membrane purification columns for baseline samples to facilitate archiving and sharing the DNA, but agrees that chelex extractions perform well for most fishery sample applications.

7.0 Storage Duration for Samples

Space requirements, personnel time for curation, and accessibility of samples all need to be considered when determining the length of time samples should be retained. Long-term storage requires increasingly larger amounts of space and personnel time as samples accumulate. It is not practical for some labs to retain samples for long periods of time. However, baseline samples and DNA have tremendous value and potential for use in other studies as outlined above and should be retained whenever possible.

While it might be desirable to reanalyze previous fishery samples when new markers are available that increase our resolution of specific stocks, these types of analyses are rarely carried out. Therefore, in cases where storage limitations exist, fishery samples need not be retained after the specific questions have been answered with the exception that fishery samples that are found to contain novel alleles should be retained for verification and standardization among labs.

8.0 Summary Recommendations

The Logistics Workgroup recommends the following regarding collection and curation of samples for genetic analysis used in the management of ocean salmon fisheries:

1. Fishery mixture samples
    - representative of the fishery, and include at-sea sampling to provide catch location
    - bulk collection is fine, unless the genetic data will be used in conjunction with other information from individual fish (CWTs or scale ages, for instance)
    - chelex extraction method is fine
    - standardization of tissue collection and DNA extraction methods not necessary
    - store at a minimum until genotyping is completed and management applications are complete and agreed to by parties

2. Baseline samples
    - representative of the population
    - individual collection and storage to allow association with biological data
    - silica membrane purification column extraction to produce archival DNA
    - store indefinitely whenever possible

3. Tissue type
    - non-lethal sample to allow sampling of released fish
    - scales not recommended for genetic analysis, but good for ageing
    - fin or axillary process best tissue type (more important for baseline samples)

4. Storage and Curation
    - tissue and DNA both highly valuable and should be retained
    - no storage method recommended over another; alcohol dehydration most common
    - dry storage should be tested further for stability

Table 1. Advantages and Limitations of Various Tissue Types for DNA Analyses.

Tissue	Advantages	Limitations
Opercle punch	Easy to collect; high DNA yield	Delamination can cause problems for bulk collections
Scale	Easy to collect	Cross contamination can cause problems; extraction consumes scale(s); low DNA yield; no indication of previous sampling
Axillary process (located above pelvic fin)	Easy to collect; maintains integrity in bulk collections; high DNA yield; easy to sample appropriate amount.	Amount of tissue limited by size of process (generally not a problem)
Fin clip	Easy to collect; high DNA yield; easiest on live fish	Delamination may cause problems with bulk collections
Heart, muscle, liver	Tissues may be available in allozyme archives; heart and liver have high DNA yield, muscle medium DNA yield; may be useful for multiple techniques	Generally lethal to collect

Table 2. Collection Methods

	Advantages	Limitations	Description
Vials, individually labeled (Alcohol, DMSO, Frozen)	Linkable to individual data (scales, ASL, etc.); most common technique	Vials must be individually labeled, filled with preservative or frozen, and capped. Each tissue must be cut by hand.	
Vials, bar-coded (Alcohol, DMSO, Frozen)	Linkable to individual data (scales, ASL, etc.); no individual labeling required; vials non-collection specific; used in Alaska and Canada	Must have field and lab barcode reader. Vials must be filled with preservative or frozen and capped. Each tissue must be cut by hand.	Factory-printed vials that do not repeat numbers.
Bulk (Alcohol, DMSO and Frozen)	Rapid and easy field collection, only single bottle must be labeled.	Not linkable to individual data (scales, ASL, etc.). Each tissue must be cut by hand.	Typically multiple individuals collected into a single 125 or 250 ml bottle of alcohol.
Blotter Paper (Dry)	Linkable to individual data (scales, ASL, etc.). Potentially easy to subsample by automation. Store at room temperature, no maintenance, low archive volume. Used in Pacific Northwest and California.	Must dry to preserve. Limited tissue types (fin).	Placed on paper and then in envelops or multiple individuals placed on single divided sheet
FTA paper (Dry) http://www.whatman.com/products/?pageID=7.31.31	Linkable to individual data (scales, ASL, etc.). Potentially easy to subsample by automation. Tissue collection and DNA extraction in single step.	Must dry to preserve. Tissue must be individually homogenized onto paper. Limited or untested for salmon.	Sample placed on FTA paper and dried. Paper contains extraction reagents, DNA sticks to the paper and can be released for analysis.

Table 3. Advantages and Limitations of Various Tissue Preservation Methods for DNA Analyses.

	Advantages	Limitations
Alcohol	Long term preservation well documented; rapid dehydration; easy/cheap to obtain; room temperature storage.	Requires hazmat shipping, approved storage and ventilation, alcohol levels must be monitored and maintained in vials
DMSO	Appropriate for shipping to "dry" communities.	Preservation properties less reliable than alcohol
Frozen	Long term preservation well documented; shipment on blue or wet ice does not require HAZMAT; useful for multiple techniques	Must remain frozen during storage; degradation over repeated freeze/thaw cycles; must guard against loss due to power or freezer failure
Dry	Dry, room temperature storage; easily cut for extraction; storage volume varies; may bypass extraction step	Must be dried after collection to avoid degradation; may be difficult to dry in wet environment; unknown length of preservation in different ambient conditions

Table 4. DNA Extraction Methods.

	Chelex	Phenol/chloroform	Silica purification columns (e.g. DNeasy, PureGene)	Magnetic Beads without robotics	Magnetic Beads with robotics
Chemical/disposable cost/sample	$0.05	$0.20	$2.00	$2.00	$2.00
Personnel time (192 fish)	2 hr	16 hr	5 hr	5 hr	2.5 hr
Total cost/sample	$0.36	$2.70	$2.78	$2.78	$2.39
Turnaround time (192 fish, 1 person)	3 hrs	16hr	6 hrs - overnight	5 hrs	5 hrs
Suitability for archive	No	Yes	Yes	Yes	Yes
DNA yield	High but dirty	High	Medium	Medium to high	Medium to high
Toxicity	Low	High	Medium	Low	Low
Automatable	No	No	Yes	No	Yes
Capital cost	$500	$3,000	$8,000	$3,000	$60,000

II. Standardized GSI Reporting: Issues and Recommendations

John Candy (DFO) – Task Leader
Logistics Workgroup

1.0 Objective

The Logistics Workgroup was also tasked with developing recommendations on standardization issues relating to the development of a coastwide GSI database for use in ocean salmon management. This report, under the leadership of John Candy (CDFO) with significant contributions from Paul Moran (NMFS-Seattle) on the inter-laboratory allelic standardization, addresses the three levels of GSI standardization necessary for significantly improving identification of salmon stocks taken in mixed stock fisheries.

1) Standardization of inter-laboratory genotyping (allele calling)
2) Standardized coastwide GSI database
3) Standardization of GSI data protocols and coding compatible with the coastwide CWT database.

2.0 Introduction

Depending on their laboratory capabilities and preferences, agencies have developed different genetic markers types and sets of loci for GSI. Unlike the earlier development of allozymes, where the number of loci was limited, there are many more DNA-based markers available to use. Microsatellites or Simple Tandem Repeats (STR) and Single Nucleotide Polymorphisms (SNPs) are two marker types which are currently in use and under continued development by fisheries laboratories along the West Coast of North America. Not all laboratories use the same assay techniques, and different marker types (and suites of loci) are most cost effective when multiplexed (multiple loci assayed simultaneously). As a result, each laboratory has found different ways to configure loci, which work best for them.

In order to standardize genetic data, all laboratories must run the same suite of loci using standardized allele calling (Moran et al. 2006). This allows each laboratory to reproduce exactly the same allelic "calls" for any given sample, independent of the instrument platform used. The data produced by a given laboratory must meet agreed standards for inter-laboratory concordance before being uploaded in a standardized format for the coastwide database (Moran et al 2005).

The GAPS (Genetic Analysis of Pacific Salmon) database is the first coastwide database with a standard set of STR markers which can be run by all those who are part of the GAPS consortium (currently 12 laboratories). This "state of the art" web-based database is currently housed at the NMFS laboratory in Seattle, and accessible to the GAPS collaborators in a beta version. It is capable of both data upload and data retrieval over the web. At present, version 2.1 of this database is populated with Chinook salmon genotypes from 13 STR loci, collected from 170 spawning ground locations (baseline) representing populations from south-eastern Alaska to California. Efforts are underway to expand the database to include the SNP baseline from the GAPS Chinook salmon collaboration and to include more baseline genotypes from other

salmonid species for both STRs and SNPs. There are no immediate plans to include mixed-stock fishery data in the GAPS database. However this would be possible with minimal modification of the current database schema.

3.0 Inter-laboratory Genotyping Standardization (Allele Calling)

Different instrument platforms and laboratory chemistries result in different allele size estimates and scoring for the same allele (Moran et al 2005). Genotyping standardization, or standardized allele calling, ensures that all laboratories are assigning the same "label" to alleles seen in their laboratory. There are now about 5 or 6 examples of successful allele calling standardizations between laboratories for a number of species. These are STR loci between:

WDFW and CDFO:	coho
WDFW and CDFO:	chum
NMFS Seattle and USFWS-Longview:	chum
CDFO and USFWS-Anchorage:	chum
CDFO and University of Alaska:	chum.

However, the most comprehensive and well documented allele calling standardization effort occurred for the GAPS STR markers (Seeb et al in press) and GAPS SNPs markers (see Narum et al. submitted) involving 9 and 7 laboratories respectively. This GAPS example will now be used to describe the standardization process. Similar processes were followed for these other standardization initiatives.

3.1 Criteria for Locus Selection

The first step in standardization requires that all laboratories decide on a common set of loci. Selection criteria for a STR locus consist of: 1) numbers of alleles, 2) allelic size range, 3) accuracy and "robustness" during laboratory processing, and 4) no obvious departures from Hardy-Weinberg (H-W) equilibrium or linkage disequilibrium. For the most part, the same criteria are used for SNPs except that the numbers of alleles or allelic size range are not a consideration. Instead, those SNPs loci that are diagnostic of a certain populations or stock groups are preferentially selected (intentional ascertainment bias for specific management applications, as opposed to neutral markers that reflect the underlying population genetics and historical demography).

For STRs, the information content is largely related to the number of alleles segregating at a given locus. More informative loci generally have large numbers of alleles (highly polymorphic). However, too many alleles can result in scoring errors and upper allele drop out, which complicates both standardization and interpretation. From 30-70 alleles per loci is generally considered the ideal number, though some researchers favor more loci with intermediate polymorphism, but faster and more robust automated genotyping. The allelic size range also determines which loci are non-overlapping and thus compatible for multiplexing with a single fluorchrome dye label. Most analytical techniques, including mixture models assume H-W equilibrium, and the alleles generated from different loci are independently derived. For the coastwide set of Chinook salmon standardized loci, initially 63 STR loci were "sponsored" for

inclusion in the standardized set from this initial pool, 25 loci were chosen as "candidates" and 13 were eventually chosen as "finalists" (Seeb et al. in press). To date, standardization has occurred for 41 SNPs (Narum et al. submitted).

3.2 Steps for Standardization

Once a standard set of loci have been selected, the next step in standardization is to exchange a "reference" set of tissue samples between laboratories, and assign each locus to a "curator laboratory" responsible for receiving, distributing and compiling allele information for the locus. The curator lab also maintains a "holotype" tissue for each allele (Seeb et al. in press). Next, each laboratory genotypes this standard set of samples with the common set of loci. Scores are compared and "labels" are assigned to each allele seen. The "label" is given the base-pair size most commonly seen and all laboratories adopt the same "label" for this allele. When a new allele is found, the curator laboratory sends out a "holotype" sample and a new "label" is identified.

3.3 Validation of Standardization

Validation of the standardization process ensures that all laboratories are identifying and correctly labeling the alleles. To test the allele calling error rate, a fresh tissue sample is collected and analyzed by all standardizing laboratories. This sample should represent a broad range of populations to capture a wide range of allele sizes (Seeb et al. in press). Since there is no independent measure of true allele size, the "modal score" or most common allele size found must be assumed to be "correct". For each locus concordance measured the percent departure from the modal score conditioned by the removal of missing data. For the "blind 2" sample involving 9 laboratories, the initial overall concordance was 95.5% and after some data handling errors were corrected this increased to 99.4% (Appendix table 1). A similar standardization occurred for 43 SNPs markers and 7 laboratories with an overall concordance of 98.8% (Appendix table 2). It was concluded that allele calling of STR markers could be reliably standardized by a number of geographically dispersed laboratories (Seeb et al in press). Clearly genotyping errors limit inter-laboratory standardization (as well as intra-laboratory standardization, for that matter).

There is broad consensus in the genetics community that per locus error rates are lower for SNPs than for STRs. It is also widely appreciated that, in general, more SNP loci are needed to provide comparable power. The net effect of this relationship is not yet clear. That is, four SNPs, on average, might be required to equal the average STR locus, but it is not yet clear if error rates will be four-fold higher for STRs than for SNPs. The bottom line is that inter-laboratory genetic standardization has reached a high level of sophistication among the West Coast salmon laboratories and intra-laboratory standardization for both marker classes is nearly as good as error rates within laboratories (see below).

3.4 Allelic Ladders

Typically, standardization has involved extensive debate over marker choice and the conditions for sharing of reference samples (holotypes) that need to be exchanged among laboratories.

Recent experience with allele ladders has obviated the need for sample distribution and has dramatically simplified inter-laboratory standardization (LaHood et al. 2002). Allele ladders are synthetic pools of amplified DNA samples that allow visualization of all or nearly all alleles in a single assay. With all the alleles arrayed according to size, it becomes trivial to simply match unknown alleles to rungs on the ladder. It is now possible to standardize new laboratories in a day or two by using the newly available GAPS allele ladders (LaHood 2007 PSC final report).

3.5 Ongoing Commitment to Standardization

SNP markers are defined by base changes at a specific nucleotide site. So long as SNP genotypes are contributed in a common format, such as that defined by the National Center for Biotechnology Information (NCBI; using the STRAND format defined in dbSNP; http://www.ncbi.nlm.nih.gov/projects/SNP/), SNP scores are standardized by definition across genotyping chemistries and platforms. Similar to STR databases, locus coordination and periodic blind testing will be required. However, SNPs will not require the ongoing agency commitment for allele curation and exchange of tissue samples that is required of STR databases.

4.0 Standardized Coastwide GSI Database

4.1 Basic Concept for a Coastwide GSI Database

There are a number of considerations for the development of a coastwide GSI database:

- Standardized formats create a 'common language' for sharing and understanding data collected on a coastwide scale by multiple agencies.
- Ideally a single entity will be selected to house and manage the coastwide database.
- Strict 'Data Neutrality' is required by the host site to ensure full confidence in shared data. This requirement would preclude long-term hosting of shared interagency data at NOAA due to perceived political bias that might affect access to the data or data integrity.
- Warehoused coastwide baseline and mixture data must be accessible to all via web-based data retrieval applications.

At the same time, it must be recognized that agencies have invested heavily in building GSI baseline and fishery mixture databases which might not become standardized:

- Agencies may continue to maintain their respective data in internal formats that meet the respective need of their own programs and operations.
- Agencies are required to develop the necessary software applications to convert their internal data sets into the coastwide formats approved for coastwide data exchange.
- Agencies may need additional staff support to maintain the conversion software and provide their respective data sets to the coastwide data center on a timely basis.
- In sum, standardization imposes a cost that must be recognized as a integral part of overall interagency co-management (Moran et al 2006).

4.2 Coastwide Database Development

4.2.1 Criteria for further Coastwide Database Development

In order to proceed with further coastwide database development, a number of criteria must be met. The Logistics Workgroup has accepted the following set of assumptions:

1) There is a desire to include additional markers and species as part of further coastwide GSI database development, as indicated by future expansion plans of the GAPS database to include Chinook salmon SNPs and other salmon STRs;
2) There is a desire to include marine fisheries data in a standardized coastwide database to provide assessment information to all agencies;
3) There are efficiencies to be gained by adopting a number of data standards used for the three decade old coastwide CWT database. The Expert Panel on CWTs recognized that GSI would be useful for complementing existing CWT programs, resulting in the possible combined CWT/GSI use in stock assessments in the fisheries (Hankin et al. 2005). An overview of the coastwide CWT system is provided by Johnson 1990.

Given this set of assumptions, the Logistics Workgroup strongly endorses further coastwide GSI database development. As such, further options must be considered to move forward.

4.2.2 Options for further Coastwide Database Development

The GAPS baseline is currently considered the "state of the art" coastwide database for Chinook salmon GSI data. As such, it should be strengthened and expanded, rather than developing something new. Following are options which impact the future of standardized database development:

- Option 1: Expand the GAPS database to handle additional species, markers and fisheries data at the NMFS laboratory in Seattle.

The NMFS site would be most familiar with the GSI database structure and this type of genetic data. By the same token, the NMFS site is less familiar with CWT data, or more specifically the Regional Mark Information System (RMIS), maintained by the Pacific States Marine Fisheries Commission (PSMFC) if this is a consideration for managing coastwide GSI database. The NMFS site is already in the process of extending the database to other species. One serious disadvantage, however, is that the NMFS site is generally not viewed as a "neutral" site. One advantage to hosting at NMFS is the existence of expertise and infrastructure that would allow substantial savings. Estimated costs are reportedly 2-3 fold higher for hosting at PSMFC, relative to previous and anticipated future costs for NMFS.

- Option 2: Move the GAPS database to a 'neutral party' (i.e. PSMFC) and expand the GAPS baseline to handle additional species, markers, and fisheries data.

The PSMFC site is considered a 'neutral site' and is the repository for all coastwide CWT data. They have developed the web-based RMIS (http://www.rmpc.org) for CWT data management

and user retrieval, but have no experience with genetic data. There also is a question of funding for expansion of GAPS at the PSMFC site. The GAPS database as currently implemented at the NMFS in Seattle, is seamlessly portable to PSMFC, requiring only an Oracle license.

- Option 3: Retain the GAPS baseline data at NMFS and build a fisheries mixed-stock database at a "neutral" site.

This is probably the least attractive option because it may duplicate effort at two sites and requires close integration between the two sites to ensure data compatibility. However, this leaves the baseline data with the developers and allows the mixed-stock data to be developed by a "neutral party". There is a question of further funding required to develop and maintain two databases, one located at NMFS and the other at PSMFC. Users need to weigh perceptions of "neutrality" and concerns for data integrity against relative costs.

4.3 Coordination with Other Entities

Coordination with genetic databases originating from the Yukon River and from other drainages outside the Pacific Salmon Commission also should be considered. Most funding for genetic data collection comes from non-GSI sources. It is important that future development of genetic methods for GSI and PSC harvest issues benefit from substantial existing data. Both U.S. and Canadian agencies support large databases that span the range of the species across the Pacific Rim. For example, the GAPs Chinook salmon datasets for STR markers now extends through Alaska, the Yukon, and into Russia. The GAPS database is already being used for studies of juvenile migration and habitat use, genetic effects of artificial propagation, marine mammal feeding ecology, and forensics (Moran et al 2006).

Many of these species-wide databases are being coordinated through the *ad hoc* Working Group on Stock Identification of the Scientific and Technical Committee of The North Pacific Anadromous Fish Commission (NPAFC; http://www.npafc.org//) whose goals are to develop, standardize, and disseminate genetic databases among the members (Canada, Korea, Russia, and United States). These databases are critical to research on salmon migration routes on the high-seas (e.g. BASIS; The Bering-Aleutian Salmon International Survey) as well as analyses such as the origin of salmon caught as bycatch in fisheries for walleye pollock in the Bering Sea.

5.0 Standardization between CWT and GSI data Formats

Both CWT and GSI data provide information on distribution and abundance of stock and stock complexes. As such, the Logistics Workgroup recognized that it may be advantageous *where possible,* to adopt RMIS data standards developed over the past 30+ years for CWT data (PSC TCDS (1989); Lapi et al. 1990). By using these standards, data formats are well known and understood by researchers and fisheries managers. As well, it provides a readily available standard set of data protocols to be followed by different agencies when uploading GSI data to a coastwide database.

Many CWT fields are common to GSI data fields. However, there are other fields which may not be similar, or have different meaning.

5.1 Comparison of CWT and GSI Database Structures

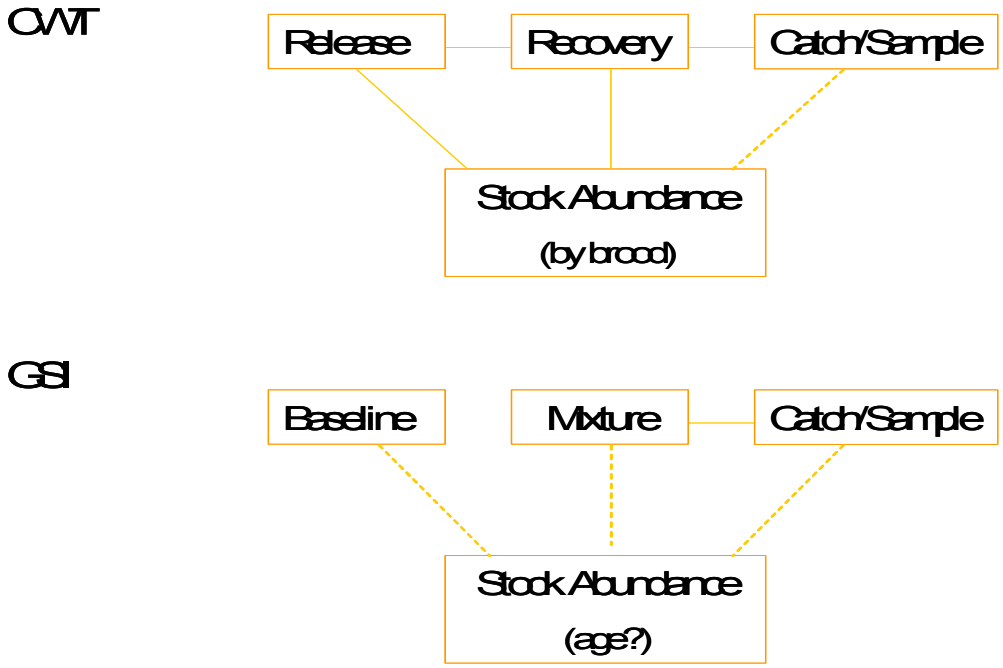

Figure 1 – Data Structures for CWT and GSI Databases.

In the most simplistic form, the CWT data consists of three main tables linked by a common tag code (Figure 1). These are the release data that is generated at the time of tagging juvenile fish during propagation. The recovery table fields consist of individual tag recovery information, while the catch/sample information pertains to sampling information where ideally 20% of the catch is scanned for tags. These three tables along with location codes allow estimates of abundance for a stock by brood year within a particular time/area/fishery stratum.

Alternatively, the GSI data uses baseline information of multilocus genotypes from known spawning ground samples and individual genotypes (mixtures) of unknown origin to estimate the stock composition of the sample using MLE or Bayesian techniques (indicated by dotted lines in Figure 1). The catch/sample information pertains to sampled proportion of catch and this is linked to the mixture table by a sample code. The stock composition estimate, combined with catch/sample information, provides estimate of abundance for a stock; however, no age or brood year information is available from this method. In addition, data field would have to capture pertinent information regarding the mixture analysis (e.g., mixture model type, and parameter settings).

5.2 CWT and GSI Expansion Factors

Both CWT and GSI use expansion factors to estimate the number of fish from a given stock found in a fishery. However, different methods are used to expand the CWT and GSI sampling data.

5.2.1 CWT Estimates of Stock Contribution

Estimates of CWT marked fish in a given time/area/fishery are based on a random sampling design, with sampling usually done at the landing site (i.e. docks or fish processing plants). The number of CWT recoveries in a given time/area sample is expanded based on the percent of the catch sampled. The 'expansion number' is the ratio of total catch divided by the sampled catch. Note: The following 'equations' are very basic. Minor adjustments to the expansion number are often made by correcting for unreadable tags or lost tags, etc)

$$\text{1) Estimated CWT Recoveries} = \frac{\text{Total Catch}}{\text{Sampled Catch}} \times \text{No. Observed Recoveries}$$

In turn, estimated total CWT contributions of tagged fish in a fishery for a given time/area stratum is calculated by adjusting for the number of total fish released divided by the number of tagged fish released from a hatchery:

$$\text{2) Total Contribution} = \frac{\text{No. Fish Released}}{\text{No. Fish Tagged}} \times \text{Estimated Recoveries}$$

5.2.2 GSI Estimates of Stock Contribution

Stock contribution is stratified by time/area/fishery. If reliable age data is available then the mixture data can be further stratified by age class for cohort analysis before generating stock composition estimates. Estimates of stock contribution for a given stratum requires only one expansion factor for GSI data since there is no release information with a "natural mark".

$$\text{Total Contribution} = \text{Total Catch} \times \text{Estimated Stock Proportion in Sample}$$

The catch is representatively sampled where the sample size depends on degree of stock composition complexity, amount of genetic differentiation between populations/regions, and level of precision and accuracy required. In an extreme case, the entire catch could be sampled for GSI, thereby eliminating sampling error completely. Sample sizes and accuracy has been discussed in detail by the Modeling and Sampling Workgroup

5.3 Standardized CWT and GSI Location Codes

Both CWT and GSI data capture location information in their respective data records. If a fishery is sampled for either genetic analysis or CWT head recovery, the standard set of CWT recovery

location codes such as time\gear\area would work equally well for both. Enhanced populations sampled for GSI could be standardized with existing CWT stock codes.

One significant difference is that CWT Production Areas are largely based on historical management requirements, while GSI regions are based on genetic similarity stock groupings. Consequently, CWT and GSI regional groupings do not always match one-to-one (Appendix Tables 4, 5, and 6).

5.3.1 CWT Location Codes

The CWT location coding scheme is based on a hierarchical seven level coding scheme that includes the following:

Level	Description
0	State or Province
1	Water type (marine or freshwater)
2	Sector
3	Region
4	Area
5	Location
6	Sub-location

In addition, a location code is further identified by five location types

Location Type	Description
1	Recovery site
2	Catch area
3	Release facility (hatchery, etc)
4	Release location
5	Stock

CWT location codes are standardized for the entire state or province and maintained by the lead state or province fishery agency (i.e. Alaska: ADFG; Canada: DFO; Washington WDFW; Oregon: ODFW; Idaho IDFG; California CDFG). Complete CWT data exchange specifications are provided on PSMFC's website: http://www.rmpc.org/files/PSC_V40_Specification.pdf .

5.3.2 GSI Location Codes

A GSI database may not benefit from existing CWT coding scheme where stock groupings and Production Area do not overlap. Genetic baseline samples are identified by latitude/longitude and typically organized and analyzed in a hierarchical hydrographic model. (basin, sub-basin, river, tributary, river reach). One possible reconciliation would be to create new GSI location codes using the existing CWT seven level coding scheme, maintaining the existing defined GSI regions. This would require that additional location code types would have to be added to the Location Type field (e.g. Type 6: GSI Recovery site; Type 7: GSI Catch area; etc.).

5.3.3 Draft Inventory of GSI Data Elements Essential for Coastwide Fisheries Management

The challenges of establishing a coastwide GSI database have strong similarities to the database standardization issues that faced the CWT program soon after the Pacific Salmon Treaty was implemented in 1985. Those issues were resolved with the first step being that each agency provided a written inventory of their respective CWT data elements. With that information, the PSC Data Standards Workgroup was tasked with developing a standardized CWT database that met the existing needs of both fishery management and researchers (Lapi et al. 1990).

A similar approach should be considered for establishing a coastwide GSI database that contains the essential data elements required for coastwide salmon fisheries management. As a first step, key personnel from fisheries agencies and genetics laboratories could be convened as a highly focused working group under the PSC umbrella. They would be charged with developing a written draft inventory of the data elements (i.e. field definitions, field sizes, required or optional reporting, etc) essential to a GSI database for coast wide fisheries management.

Individual data elements can be viewed as a shorthand name for identifying specific data that must be collected in order to produce useful management information. Each data field name should be further described with narrative explanations of its intent and purpose, scope, examples, and relationship to other data elements (i.e. metadata). A database would necessarily contain essential genetic stock id values as well as management parameters. Therefore, the information would need to come from experts in both those areas.

The nature of the data required cannot be fully ascertained until fishery managers and others have clearly stated the types of analysis and outputs they expect are required for fisheries management processes and other research purposes. For example, is individual stock attribution of recoveries essential? Is a stock mixture probability matrix essential? Therefore, some initial collaboration with the modeling and sampling experts will probably be required.

The final product should be a best effort draft document. The product could then be delivered to the PSC Workgroup on Data Standards for action. They are experts in formulating essential comprehensive and complex database definitions. They can also establish the reporting requirements and permanent processes needed to compile and update this database on an ongoing production basis. Once the Data Standards Workgroup is given a well designed draft GSI database inventory, they should then be able to create a useful and effective product, based on their past performance with standardizing the CWT database.

As a first effort for this report, three *very preliminary* Excel tables (Appendix Tables 4-6) were developed to show overlap between the CWT and current GSI data fields for recovery, catch/sample, and location data. As expected, many of the CWT and GSI fields do correspond 1:1. There are others which have similar elements but different meanings. This is particularly true for some of the location data fields. And again, there are many fields which are unique to GSI data as would be logically expected. Merging the two databases is clearly a worthwhile endeavor but it will require a focused effort of CWT and GSI personnel working together to sort through the data nuances to ‘get it right’.

6.0 Recommendations of the Logistics Workgroup:

1. Continued development of the standardized GAPS baseline to accept additional marker types and species.
   a) NMFS-Seattle laboratory is highly commended for outstanding development of the web-based GAPS database and data retrieval applications.
   b) Funding support for this effort should continue to be a high priority.
   c) Agency laboratories should accelerate efforts to standardize allele callings for a number of species with existing STR databases.

2. Further discussion recommended as to the future location of a coastwide GSI database.
   a) PSMFC is recommended as the preferred host site based on long experience in hosting regional databases and demonstrated data neutrality.
   b) This decision was not without some dissent and raised concerns that should be addressed in a larger forum. Issues raised are:
      - NMFS has the expertise and infrastructure in place to continue to serve as the host site and thus allow substantial efficiencies and cost savings.
      - Estimated NMFS costs for hosting the site are reportedly 2-3 fold lower than that of PSMFC. This raises the decision if data neutrality is worth PSMFC's extra cost.
   c) There is no support for retaining the GAPS baseline data at one site (i.e. NMFS) while developing a fishery mixed-stock database at a neutral site (i.e. PSMFC).

3. Further discussion recommended as to the configuration of a coastwide GSI database.
   a) The GAPS database should be modified to include fishery mixture sample data.
   b) There was no consensus on whether the new fishery mixture data would include both individual and mixture assignments.
   c) There is strong interest by some that individual assignments should be made available in the GSI database for fishery managers and other researchers. As such, this topic will need further exploration as to its practicality.

4. Further evaluation of CWT data standards and protocols for adoption for standardized GSI database.
   a) The challenges of establishing a coastwide GSI database have strong parallels to the development of the CWT database. As such, it is strongly recommended that the new GSI database incorporate all of those CWT data parameters that can be used.
   b) Key personnel from both the fisheries and genetics communities will be needed to merge the two databases to the degree possible. It will require significant effort at some levels of detail.
   c) It is recommended that this effort be done under the PSC umbrella by a Data Standards Workgroup dedicated to GSI data, similar to that done when standardizing the CWT data formats.
   d) Additionally, the warehoused coastwide data must be readily available to all by Web-based data retrieval, including GIS supported applications to provide recovery location of baseline and fishery samples.

7.0 References

Hankin, D.G., Clark, J.H., Deriso, R.B., Garza, J.C., Morishima, G.S., Riddel, B.E., Schwarz, C., and Scott, J.B. 2005. Report of the Expert Panel on the future of the coded wire tag recovery program for Pacific Salmon. Pacific Salmon Commission Tech. Rep. No. 18 230 pp. Available at: http://www.psc.org/pubs/CWT/EPfinalreport.pdf

Lapi, L., Hamer, M., and Johnson, B. 1990. Data organization and coding for a coastwide mark recovery data system. American Fisheries Society Symposium 7:720 724.

Johnson, J.K. 1990. Regional overview of coded wire tagging of anadromous salmon and steelhead in northwest America. American Fisheries Society Symposium 7:782-816.

LaHood, E.S., Moran, P., Olsen, J., Grant, W.S., and Park, L.K. 2002. Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. Molec. Ecol. Notes 2:187-190.

Moran, P. and 11 co-authors. 2005. Interlaboratory standardization of coast-wide Chinook salmon genetic data for international harvest management. Progress report from the Genetic Analysis of Pacific Salmonids (GAPS) consortium to the Chinook Technical Committee of the Pacific Salmon Commission, FY2004, FY2005, 44 p.

Moran, P., Teel, D.J., LaHood, E.S., Drake, J., and Kalinowski, S. 2006. Standardizing multi-laboratory microsatellite data in Pacific salmon: An historical view of the future. Ecology of Freshwater Fish 15:597-605.

Narum, S.R. and 16 authors (submitted) Differentiating populations at broad and fine geographic scales with microsatellites and SNPs. Molecular Ecology.

Seeb, L.W., Antonovich, A., Banks, M.A., Beacham, T.D., Bellinger, M.R., Campbell, M., Garza, J.C., Guthrie, C.M. III., Moran, P., Narum, S.R., Stephenson, J.J., Supernault, K.J., Teel, D.J., Templin, W.D., Wenburg, J.K., Young, S.F., and Smith, C.T. Submitted. Development of a Standardized DNA Database for Chinook Salmon. *Fisheries*

PSC TCDs (89) 1. Joint Technical Committee on Data Sharing Joint Working Group on Mark Recovery Databases. Information Content and Data Standards for a Coastwide Coded Wire Tag Database July 12, 1998.

Appendix Table 1. Concordant results from multi-lab analysis of the GAPS 13 microsatellite markers and nine laboratories (from Seeb et al in press).

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

Appendix Table 2. Concordant results from multi-lab analysis of 43 SNP loci (S. Narum, CRITFC)

Locus	Lab1	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7	Average
arf-188	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
ARNT	98.9%	82.2%	100.0%	80.5%	0.0%	98.9%	82.2%	95.0%
AsnRS-60	97.7%	95.3%	96.4%	97.8%	100.0%	95.6%	97.8%	97.4%
C3N3	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CYP17	100.0%	100.0%	97.6%	100.0%	0.0%	98.9%	100.0%	99.1%
E2-275	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
E9 BAC	100.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%
ETIF1A	96.6%	100.0%	100.0%	100.0%	97.6%	98.9%	100.0%	99.1%
FGF6A	98.9%	97.6%	100.0%	100.0%	100.0%	98.9%	100.0%	99.5%
GTH2B-550	100.0%	100.0%	100.0%	100.0%	97.7%	96.7%	100.0%	99.3%
GnRH-271	100.0%	100.0%	100.0%	100.0%	0.0%	100.0%	97.8%	99.6%
GPDH-338	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
HGFA-446	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
IGF-I.1-76	100.0%	100.0%	100.0%	100.0%	100.0%	98.9%	98.9%	99.5%
Ikaros-250	100.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%
il-1racp-166	100.0%	100.0%	100.0%	98.9%	65.9%	100.0%	100.0%	99.8%
LWSop-638	0.0%	0.0%	0.0%	100.0%	100.0%	0.0%	0.0%	100.0%
MetA	97.6%	92.5%	100.0%	100.0%	0.0%	93.4%	100.0%	98.2%
MHC1	98.9%	98.7%	98.9%	98.9%	0.0%	100.0%	100.0%	99.3%
MHC2	96.6%	100.0%	100.0%	100.0%	98.9%	92.2%	98.8%	97.5%
NOD1	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
P450	98.9%	100.0%	100.0%	100.0%	0.0%	97.7%	100.0%	99.3%
P53	98.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.8%
PGK-54	96.6%	97.7%	100.0%	100.0%	100.0%	96.6%	100.0%	98.3%
Prl2	98.8%	100.0%	100.0%	100.0%	91.0%	100.0%	100.0%	99.8%
RAG3	100.0%	100.0%	100.0%	100.0%	92.9%	100.0%	100.0%	100.0%
RFC2-558	95.5%	96.6%	100.0%	98.9%	100.0%	96.7%	97.8%	97.8%
SClkF2R2-135	97.8%	100.0%	100.0%	100.0%	100.0%	98.9%	100.0%	99.3%
S7-1	98.9%	100.0%	100.0%	100.0%	100.0%	98.8%	100.0%	99.5%
SL	97.8%	98.8%	100.0%	100.0%	83.8%	92.5%	100.0%	98.1%
SWS1op-182	100.0%	98.9%	100.0%	100.0%	90.9%	98.9%	98.9%	99.6%
TAPBP	97.6%	98.8%	100.0%	98.9%	0.0%	100.0%	98.9%	99.1%
TCL1	97.6%	97.6%	100.0%	100.0%	97.7%	100.0%	98.8%	99.1%
Tnsf	100.0%	100.0%	100.0%	100.0%	100.0%	92.1%	100.0%	98.4%
u202-161	96.6%	100.0%	100.0%	100.0%	100.0%	98.8%	100.0%	99.1%
u211-85	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
u212-158	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	98.9%	99.8%
u4-92	100.0%	100.0%	100.0%	100.0%	98.8%	98.9%	97.6%	99.7%
u6-75	98.9%	100.0%	100.0%	100.0%	100.0%	100.0%	98.9%	99.6%
unkn526	98.9%	100.0%	100.0%	100.0%	100.0%	98.9%	97.7%	98.9%
ZNF330-181	98.9%	97.8%	100.0%	98.9%	96.7%	96.6%	98.9%	98.7%
Average	98.92%	98.81%	99.82%	99.33%	97.25%	98.44%	99.03%	98.8%

Appendix Table 3. Comparison of DFO-MRP Production Areas and DFO-GSI regions.

Nation	Stock Group	DNA Region Name	MRP Production Area
CDN	Fraser	UPFR	UPFR
		MUFR	UPFR
		LWFR-F	LWFR
		LWFR-Sp	LWFR
		NOTH	TOMF
		SOTH	TOMF
		LWTH	TOMF
	Vancouver Island	ECVI	GSVI,GSMN,JNST
		WCVI	SWVI,NWVI
	Mainland Coast	SOMN	JNST,GSMN
		NOMN	RIVR,CCST,NCST
	Nass/Skeena	NASS	NASS
		Skeena	SKNA
	Queen Charlottes	QCI	QCI
	Transboundry	Taku	TRAN
		Stikine	SIAK,TRAN
		Alsek	TRAN
USA	Alaska	Alaska	ALASKA
	Washington	Puget Sound	GSMS,WA01,WA02,WA03,WA04,WA05
		Juan de Fuca	WA06
		Coastal Wash	UPWA.WAPA,GRAY
	Columbia	Low Col	LOCO
		Up Col-Sp	HEAD
		Up Col-Su/F	HEAD,DESC,BRGT
		Snake-Sp/Su	SNAK,BRGT
		Snake-F	SNAK
		Mid Col-Sp	HEAD
		Up Willamette	WILL
	Oregon	Oregon coastal	UPOR
		S.Oregon/Cal coast	LWOR,CALI
	California	Sacramento	SACR
		Up Klam/Trinity	CALI
		Cent Val-F	SACR
		Cent Val-Sp	SACR

Appendix Table 4. *Preliminary* comparison of required CWT and GSI data fields

COMPARISON OF REQUIRED CWT AND GSI DATA FIELDS

CWT RECOVERY FIELDS	CWT Required	GIS Required?
Record Code	Y	Y
Format Version	Y	Y
Submission Date	Y	Y
Reporting Agency	Y	Y
Sampling Agency	Y	Y
Recovery ID	Y	Y
Species	Y	Y
Run Year	Y	maybe
Recovery Date	Y	Y
Recovery Date Type	Y	maybe
(Sampling) Period Type	Y	maybe
(Sampling) Period	Y	maybe
Fishery	Y	Y
Gear	Y	Y
Ad-clip Selective Fishery	Y	maybe
Estimation Level	Y	N
Recovery Location Code	Y	Y
Sampling Site	Y	maybe
Recorded Mark	Y	Y
Sex	Y	Y
Weight (in Kg)	Y	Y
Weight Code	Y	Y
Weight Type	Y	Y
Length	Y	Y
Length Code	Y	Y
Length Type	Y	Y
Detection Method	Y	N
Tag Status	Y	maybe
Tag Code	Y	maybe
Tag Type	Y	maybe
Sequential Number	Y	N
Sequential Col. No.	Y	N
Sequential Row No.	Y	N
Catch Sample ID	Y	Y
Sample Type	Y	Y
Sampled Maturity	Y	Y
Sampled Run	Y	Y
Sampled Length Range	Y	N
Sampled Sex	Y	N
Sampled Mark	Y	N
Estimated Number	Y	Y
Unique GSI RECOVERY FIELDS		
Scale Sample (book #, orientation)		Y
GSI Assigned Stock (proposed new field)		Y
CWT Listed Stock (proposed new field)		Y
Brood Year (proposed new field)		Y
Baseline version/date		Y
Baseline size (# pops)		Y
Baseline range (# regions)		Y
Baseline Depth (# micros, # SNPS, # others)		Y
Mixture Model Type (MLE or bayes)		Y
Mixture Model Settings (# reps, genotypic tolerance, bootstraps, etc)		Y

Appendix Table 5. *Preliminary* comparison of CWT and GSI Catch/Sample Data Fields

COMPARISON OF REQUIRED CWT AND GSI DATA FIELDS

CWT Catch/Sample Data Fields	CWT Required	GIS Required?
Name	Y	Y
Record Code	Y	Y
Format Version	Y	Y
Submission Date	Y	Y
Reporting Agency	Y	Y
Sampling Agency	Y	Y
Catch Sample ID	Y	maybe
Species	Y	Y
Catch Year	Y	Y
(Sampling) Period Type	Y	maybe
Period	Y	maybe
First Period	Y	maybe
Last Period	Y	maybe
Fishery	Y	Y
Ad-clip Selective Fishery	Y	maybe
Estimation Level	Y	N
Catch Location Code	Y	Y
Detection Method	Y	maybe
Sample Type	Y	Y
Sampled Maturity	Y	N
Sampled Run	Y	maybe
Sampled Length Range	Y	N
Sampled Sex	Y	N
Sampled Mark	Y	N
Number Caught	Y	Y
Escapement Estim. Meth.	Y	N
Number Sampled	Y	Y
Number Estimated	Y	Y
Number Recovered Decoded	Y	N
Number Recovered No CWTs	Y	N
Number Recovered Lost CWTs	Y	N
Number Recovered Unreadable	Y	N
Number Recovered Unresolved	Y	N
Number Recovered Not Processed	Y	N
Number Recovered Pseudo Tags	Y	N
MR 1st Partition Size	Y	N
MR 1st Sample Size	Y	N
Mr 1st Sample Known Ad Size	Y	N
MR 1st Sample Obs Adclips	Y	N
MR 2nd Partition Size	Y	N
MR 2nd Sample Size	Y	N
MR 2nd Sample Known Ad Status	Y	N
MR 2nd Sample Obs Adclips	Y	N
Mark Rate	Y	N
Awareness Factor	Y	N
Sport Mark Incidence Sample Size	Y	N
Sport Mark Inc Sample Obs Adclips	Y	N
Unique GSI CATCH/SAMPLE FIELDS		
Tissue collection -tissue type		Y
Tissue collection -preservation method		Y

Note: While both CWT and GSI compute the 'Number Estimated', the values are not comparable because of different methods used to expand the sampled data.

Appendix Table 6. *Preliminary* comparison of CWT and GSI Location Data Fields

COMPARISON OF REQUIRED CWT AND GSI DATA FIELDS

CWT Location Data Fields	CWT Required	GIS Required?
Record Code	Y	Y
Format Version	Y	Y
Submission Date	Y	Y
Reporting Agency	Y	Y
Location Code	Y	Y
Location Type	Y	Y
Name (Short name - 20 chars)	Y	maybe
Latitude	Y	Y
Longitude	Y	Y
PSC Basin	Y	maybe
PSC Region	Y	maybe
EPA Reach	Y	maybe
Description (Long name- 100 char)	Y	Y
Fishery	Y	Y
Ad-clip Selective Fishery	Y	maybe
Estimation Level	Y	N
Catch Location Code	Y	Y

**New GSI location data fields probably aren't needed. The difficulty here is that the CWT regional groups aren't necessarily valid genetic groupings. There is a lot of overlap but not always. Hence adjustments will be needed to accommodate the genetic groupings.

III. **Pacific States Marine Fisheries Commission's Possible Future** Role in Hosting a GSI Database for Fisheries Resource Management on the Pacific Coast

George Nandor
Pacific States Marine Fisheries Commission
GSI Logistics Workgroup Member

1.0 Introduction

The Logistics Workgroup of the 2007 GSI Workshops held in Portland, Oregon and Vancouver, Canada recommends the Pacific States Marine Fisheries Commission (PSMFC) as the organization to host and maintain the proposed Genetic Stock Identification (GSI) Database system for salmon stock identification in West Coast fisheries. This would be done by the Regional Mark Processing Center program within PSMFC.

2.0 Overview of the Regional Mark Processing Center

The Regional Mark Processing Center (RMPC) is a unit within the Pacific States Marine Fisheries Commission that maintains the regional coded-wire tag (CWT) database. The RMPC is the central repository for all CWT marked and otherwise associated release, catch, sample, and recovery data regarding anadromous salmonids in the greater Pacific Coast Region of the Unites States of America. It also has a data sharing process with Canada under the auspices of the Pacific Salmon Treaty and the Pacific Salmon Commission.

2.1 Neutral Site for Data Sharing

The PSMFC is a neutral site for data sharing and also maintains numerous other regional data sharing projects and programs that are used to manage West Coast fisheries. These include the PIT Tag Information System (PTAGIS), Alaska Fisheries Information Network (AKFIN), California Cooperative Fish and Aquatic Habitat Data Program (CalFish), Pacific Fisheries Information Network (PacFIN), Recreational Fisheries Information Network (RecFIN) and others.

2.2 International Role for Sharing CWT Data through PSC Standardized Formats

The RMPC was originally formed in Oregon in 1970 through the Anadromous Fish Act. In 1977 the RMPC was transferred to the PSMFC and its role expanded from just data management to include regional coordination. In 1987 the Pacific Salmon Commission, formed by the Pacific Salmon Treaty, selected the RMPC as the site to maintain the primary coded wire tag database to expedite data exchange between the U.S. and Canada. This led to the standardization of all CWT data using the PSC format standards. Since then, there have been several data format upgrades (Version 4.0 is the current PSC standard). RMPC computers and data access methods have been upgraded over time to the current state of the art systems.

2.3 Coast-wide Role includes Regional Coordination with Data Management

The RMPC uniquely exists to provide essential services to international, state, federal, tribal and other fisheries organizations. These services include: regional coordination of salmonid tagging and fin marking programs; direction and management of region wide databases of information relating to the marking and coded-wire tagging of salmonids; development and maintenance of online computer applications for querying and reporting from the databases known collectively as the Regional Mark Information System (RMIS) and supporting and facilitating the ongoing needs of the member states of Pacific States Marine Fisheries Commission, the Regional Committee on Marking and Tagging (Mark Committee) and the Pacific Salmon Commission.

2.4 Overview of the CWT Program Supported by the RMPC

The CWT program is enormous in scale. Over 50 million juvenile fish are tagged annually and subsequently released to migrate to the Pacific Ocean. Large-scale ocean and freshwater sampling and recovery programs are conducted by the various states, federal and tribal entities. Approximately 150,000 - 310,000 tags are recovered yearly from commercial and sport fisheries plus escapement at hatcheries and spawning grounds. Over thirty years of this release and recovery data is recorded and stored at the RMPC. The RMPC also maintains a web site that provides tools to query the data and provides information about the program to both professional and casual users of the data.

This kind of long term experience with fisheries data housing and dissemination makes PSMFC very well qualified for hosting both the baseline and fishery (or recovery) GSI databases.

3.0 Proposal for Hosting the Future GSI Database

3.1 Migration of NMFS's 'GAPS' Database to PSMFC

The NMFS's Northwest Fisheries Science Center in Seattle has developed the first genetic database for salmon on the West Coast. The database, known as GAPS (Genetic Analysis of Pacific Salmon), currently houses baseline genotype data for Chinook salmon populations from south-east Alaska to central California. It is expected that baseline data for other salmon species will be added in the near future.

The GAPS database is web-based and 'state of the art' in terms of functionality and data retrieval tools. This includes the use of the widely used Oracle relational database software, thus making the baseline GAPS database easily transferable to other sites using Oracle. Work is also progressing on developing a geographic information system (GIS) with respect to recovery sites for genetic baseline samples.

The RMPC also utilizes Oracle's relational data management software to maintain its coast-wide CWT database. As such, RMPC staff will work closely with NMFS staff to port the existing GAPS baseline genetic database to the RMPC with the goal of maintaining as much functionality of the NMFS application as possible.

3.2 Design of the New GSI Database

RMPC staff would rely on the expertise of salmon geneticists and other fisheries scientists to help guide the formation of the database's details and the management and content of the genetic baseline and fishery mixture data sets. Standardization of the data fields in the database ideally would be done through the formation of a GIS Data Standards Workgroup, very similar to the current CWT Data Standards Workgroup. The group would be composed of representatives from the various interested regional fisheries agencies and may have many of the same representatives from the current workgroup. The group would define each required data field and the format of the data within each field, so that the data meets the needs of the fisheries managers when it is queried and retrieved.

3.3 Procedures for Including Genetic Samples from Salmon Fisheries

It is envisioned that sampling agencies and/or processing labs would forward all GSI fishery data directly to the RMPC. The data would then be loaded and validated against an extensive set of checks that would include whether data is in the proper format, that required data fields are present and completed, and that valid data is in each field. Once validated, the data would be moved into a relational database (Oracle) and made accessible through the newly proposed Regional GSI Data System (RGDS) via on-line query tools or by special request. The data would be publicly available and shared among all parties

.

4.0 Cost Estimates

A draft budget worksheet is provided in Section 6.0. If the program were to be funded in calendar year 2008, the estimated cost would be $317,339. This includes start-up capital outlay costs for computer hardware and software, office supplies and furnishings, etc. This very reasonable level of costs is possible because of the ability of the RMPC to take advantage of the economy of scale by piggy-backing and synchronizing staff efforts and resources with maintaining the existing coded wire tag program database.

5.0 Summary Statement

The Pacific States Marine Fisheries Commission is both capable and willing to host a new coast-wide GSI database for regional fishery management applications. The RMPC has over three decades of experience in managing CWT data for similar stock identification purposes, and thus understands the basic needs for both regional coordination and data management. In addition, PSMFC is well known for its neutrality with respect to data interpretation. In summary, the PSMFC has the ability and reputation to do this in a very professional and unbiased manner to fulfill the needs of many diverse parties in the United States and Canada.

It must be emphasized, however, that new funding will be necessary for PSMFC to accomplish this task.

6.0 Draft Budget for FY 2008

PSMFC Regional Mark Processing Center
(RMPC)
Draft GSI Budget
(January 1 - December 31, 2008)

Version: **10-Aug-07**

			1st Quarter	2nd Quarter	3rd Quarter	4th Quarter	Totals
Personnel							
61100	Salaries and Wages		1/1/08-3/31/08	4/1/08-6/30/08	7/1/08-9/30/08	10/1/08-12/31/08	
	RMPC Manager (13)		6,700	6,700	6,700	6,700	26,800
	Analyst/Programmer (11)	New GSI Position	15,000	15,000	15,000	15,000	60,000
	Administrative Assistant		2,000	2,000	4,000	4,000	12,000
	Total Salaries and Wages:		23,700	23,700	25,700	25,700	**$98,800**
62100	Employee Benefits: 28% + $713/month/employee						
	Rate = 28%		6,636	6,636	7,196	7,196	27,664
	plus (19 mos. @ $713/mo.)		3,209	3,209	3,565	3,565	13,547
	Total Benefits:		9,845	9,845	10,761	10,761	**$41,211**
	Total Salaries and Benefits:		**$ 33,545**	**$ 33,545**	**$ 36,461**	**$ 36,461**	**$140,011**
Services/Supplies							
63130	Communications Services		50	50	50	50	200
63160	Data Processing Fees/Computer Center Fees		1,800	1,800	1,800	1,800	7,200
63460	Photocopying/Printing		50	50	50	50	200
63520	Postage and Freight		30	30	30	30	120
63580	Rents/Leases		1,400	1,400	1,400	1,400	5,600
63610	Repairs and Maintenance		250	250	250	250	1,000
63640	Software and Support (Oracle, Windows Software, etc.)		5,500	1,000	1,000	500	8,000
63760	Supplies – Office		500	500	500	500	2,000
63850	Training		3,000				3,000
	Total Services and Supplies:		**$ 12,580**	**$ 5,080**	**$ 5,080**	**$ 4,580**	**$ 27,320**
Contractual							
64800	Contractual Services (300 hours/year @ $75/hr)		10,000	7,500	4,000	1,000	22,500
	Total Contractual Cost:		**$ 10,000**	**$ 7,500**	**$ 4,000**	**$ 1,000**	**$ 22,500**
Travel							
65100	Travel		1,000	1,000	1,000	1,000	4,000
	Total Travel Cost:		**$ 1,000**	**$ 1,000**	**$ 1,000**	**$ 1,000**	**$ 4,000**
Capital Outlay							
66100			87,000				
	Total Capital Outlay:		**$ 87,000.00**	**$ -**	**$ -**	**$ -**	**$87,000**
Total Direct Costs			**$ 144,125**	**$ 47,125**	**$ 46,541**	**$ 43,041**	**$ 280,831**
PSMFC Overhead		13.00%	**$ 18,736**	**$ 6,126**	**$ 6,050**	**$ 5,595**	**$ 36,508**
TOTAL BUDGET			**$ 162,861**	**$ 53,251**	**$ 52,591**	**$ 48,636**	**$ 317,339**

Capital Items:	Out Year Estimates:	
Database Server w/Storage - $50,000	2009 - $217,000	2012 - $257,000
Oracle Database - $30,000	2010 - $230,000	2013 - $275,000
Desktop Computer Workstation - $5,000	2011 - $240,000	2014 - $300,000
Office Furniture - $2,000		

Note:
This proposed level of funding for housing and maintaining the GSI database at PSMFC's RMPC is contingent upon having the existing RMPC Manager and 2 technical staff positions available to help support the GSI program, in parallel with the Coded Wire Tag database. If the Coded Wire Tag database program were to end, then an increased level of funding for the GSI database program at PSMFC would be required.

IV. Use of Otolith Thermal Marking as a Compliment to Genetics in Coastwide Salmon Management; Issues and Recommendations

Eric Volk
Alaska Department of Fish and Game

Objective

The purpose of this report is to address the most significant challenges of a coastwide otolith thermal marking program to compliment genetic stock identification methods across a spectrum of potential applications.

1.0 Introduction

Otolith thermal marking is a widely used technique for identifying hatchery-released salmonids by inducing structural patterns to their otoliths using short-term water temperature manipulations (Brothers, 1990; Volk et al., 1990, 1999). The method is a practical means for 100% marking of hatchery salmon populations, offering distinct advantages over individual tagging of fish.

1.1 Current Otolith Thermal Marking Programs

Large-scale thermal marking programs occur in Canada, Japan, Russia and the United States (Alaska, Canada and Washington State), with more than 1 billion juvenile salmon marked annually, 90% of which are pink and chum salmon. Hagen et al., (1995) demonstrated how otolith marking could be used as an in-season estimator of stock composition for fisheries managers in Alaska, with samples processed within 24 hours. Continued use of this technique to manage the Prince William Sound pink salmon fishery has demonstrated greater precision of hatchery contribution estimates with far smaller sample sizes and much faster results than traditional coded wire tagging programs could provide (Joyce and Evans, 2001). Similar success was demonstrated for in-season management of Canadian and U.S. sockeye (*O. nerka*) stocks in SE Alaska (Jensen and Milligan, 2001). The use of thermal marking for in-season management of Alaskan salmon fisheries continues.

1.2 Range of Otolith Thermal Marking Options

There is a wide range of options for using otolith thermal marks in conjunction with genetic analyses for coastwide salmon management, ranging from the application of a single mark among all index hatcheries and brood years to unique marking of all hatcheries and brood years. It is unlikely, however, that sufficient patterns exist to mark all west coast hatcheries over several brood years.

At one end of the spectrum, a thermal marking program only offers information to confirm the fish's origins as a hatchery fish. Where genetic analyses are unable to distinguish between closely associated hatchery and wild populations, this would solve a fundamental issue and probably enjoy some economies in sample processing because only those specimens in question from genetic analyses would need to be checked for thermal marks.

On the other hand, providing unique, brood specific marks at each index hatchery would resolve specific sample origins and ages of hatchery origin fish. The current CWT program accomplishes this. However, otolith thermal marking provides a practical means for marking the entire hatchery population instead of a much smaller fraction. Logistical challenges for implementation obviously increase across this spectrum.

Principle issues surrounding implementation of otolith thermal marking as a compliment to genetic analyses include: 1) The number of marks available for proposed objectives, 2) Coordination of mark application and recovery, and 3) Costs associated with executing this program at many hatcheries. These issues are, in some sense, inter-related and discussed below.

2.0 Number of Otolith Marks Available

At least 1000 patterns are possible for Chinook salmon otoliths incubated under ideal circumstances for thermal marking (Volk et al., 1994). The number was based upon marking in three otolith regions, beginning in the embryonic phase, and using a specific bar code rule developed for this purpose. However, practical limitations associated with hatchery operations, fish development and visual recognition may place important limits on the actual number of available patterns (Hagen, 1999). Nevertheless, less than 200 marks would be needed to provide four brood specific marks to all ~40 Chinook index hatcheries on the West Coast. This number is well within the practical limits of the otolith thermal marking technique.

A critical assumption in this is that a single, brood specific hatchery identifier can be effectively applied to the entire hatchery population in question. If physical plant limitations (i.e. delivering adequate amounts of chilled or heated water to the entire population simultaneously) effectively result in more than one mark for some hatcheries, problems of mark availability will be exacerbated.

The number of available marks must also be viewed in light of existing programs in the U.S. and Canada for the species in question. On the other hand, efficiencies may be gained where harvests of some hatchery populations do not overlap with others, allowing some marks to be repeated at widely spaced hatcheries. In the end, while sufficient marks probably exist for marking Chinook at index hatcheries, coordination with existing programs (see below) and adequate water supply ability will be important considerations.

3.0 Coordination

3.1 Need for Regional Coordination

Any multi-jurisdictional plan for broad application of otolith thermal marks faces important challenges in coordinating those efforts. Clearly, a single coordinating body is essential to maintain control of assignment of all thermal mark codes to all agencies involved in otolith thermal marking. These determinations ideally would be made by an inter-agency committee representing all thermal marking agencies. The complexity of this coordination would depend upon the scope of this effort (i.e., otolith thermal marks in support of genetic stock identification

work) in combination with existing otolith marking programs operated by agencies for their own purposes.

Otolith thermal marking is relatively inexpensive and conceptually simple on a small scale. As such, it enjoys wide application for research and local management issues. It is crucial that each of these efforts be considered in coordinating all applications on the West Coast. In an attempt to document and coordinate the widespread application of otolith thermal marks, a salmon marking working group was established under the Committee for Scientific Research and Statistics of the North Pacific Anadromous Fish Commission (Urawa et al, 2001).

This group can serve as a model for regional coordination and data management. In addition to maintaining a database of existing thermal marks, attempts are made to coordinate mark induction to avoid duplication. In principle, countries annually submit specific mark plans for induction to the upcoming brood year, so that obvious conflicts might be resolved prior to the commencement of marking. Following the marking season, summaries of actual marks induced are submitted and entered into a database. Users can query the database through the Internet, with a link from the NPAFC web site. Specific information on each nation's mark groups and induced patterns is summarized, including a digital image of most mark patterns. This provides a ready source of information for determining the origins of an unknown pattern.

A similar type of coordinating body on the West Coast will be vitally important to the success of an endeavor to thermally mark all index Chinook hatcheries to support genetic stock identification programs.

3.1 Mark Application

The basis for otolith thermal marking rests in the fundamental relationship between environmental temperature fluctuations and the appearance of regularly deposited otolith increments (Brothers, 1981; Campana and Neilson, 1982, 1985). The idea behind using short-term temperature manipulations to mark juvenile fish otoliths is to alter the appearance of one or more otolith increments to produce an obvious pattern of events.

There are a variety of methods for the orderly assignment of marks and all are variations on the theme of bars and spaces (Volk et al., 1994; Munk and Geiger, 1998). Patterns can be created such that alpha-numeric descriptors of patterns can be adapted to a database. Patterns should always be based upon relative spacing and number of bars rather than any absolute distances, as individual fish and otolith growth is variable. The coordinating body should determine the best approach for pattern assignment.

Quality control of actual mark application would most likely fall to the individual hatchery manager or a dedicated person hired in support of the manager to accomplish the marking. There are a host of problems that might interfere with faithful execution of the marking plan and it is unlikely that all can be eliminated. As a result, collection of "voucher" specimens from all appropriate groups should document the actual pattern induced, and any discrepancies of planned to actual events. Digital images of all voucher specimens should be entered to a common database established by the coordinating body.

3.2 Mark Recovery

The recovery of thermally induced otolith marks involves sectioning otoliths and interpreting the marks. Physically recovering otoliths can be done quickly. However, usual methods of removal invoke significant damage to the head of the fish which may pose secondary problems in some cases. Once otoliths are removed, their storage is simple, with tracking done using bar code labeled vials to associate with other samples.

Processing of otoliths should occur at the labs that are currently associated with major West Coast agencies' thermal marking efforts, including ADFG, CDFO and WDFW. A regional otolith lab should be considered. The principal activity of these labs is to recover the mark from the whole salmon otolith. Skilled technicians can section otoliths quickly. However, good quality otolith marks induced during hatchery rearing are the best way to reduce time consuming, high quality preparations needed to distinguish poorly induced patterns. Specific quality control measures should be adopted in the mark recovery process in recognition of documented error rates incurred during mark recovery (Volk et al., 1999; Joyce and Evans, 2001). In general, it appears that error rates have diminished as investigators have learned how to more effectively induce clear otolith marks.

4.0 Costs

A significant cost issue for a broad scale otolith thermal marking program involving many hatcheries is that most will require some modification and upgrade to deliver specific water temperature events to large numbers of fish simultaneously. Although some facilities may have separate water temperature regimes that could be effective for mark application, most will have to create separate water temperatures through heating or cooling. Obviously, the cost of this will be specific to individual hatcheries, but will likely be significant.

Small scale, portable operations have been established for under $25,000 USD, but these systems usually require delivery of thermal events to several portions of the hatchery production in succession. This can, in effect, produce multiple marks at a given hatchery which increases issues associated with the number of marks and distinguishing between similar marks.

Most thermal marking can be accomplished with an individual at each facility turning valves on and off at a specified time. However, where sufficient resources exist, computer controlled valves that operate automatically on a pre-determined schedule are preferred. These plumbing and electrical systems are available and feasible. However, one might expect that costs for marking would increase appreciably.

It is important to remember that a host of issues associated with pattern recognition, replicate marks and errors are directly related to the quality of the induced mark. This, in turn, is very often dependant upon the capacity and flexibility of the physical plant system that is the foundation for mark induction. Thus, it is unwise to adopt thermal marking without adequate capital investment in the hatcheries for this purpose. A conservative estimate for capital upgrades to all Chinook index hatcheries would be $2-4 Million USD. However, that figure is

totally dependant upon objectives of the program, mainly the number of unique marks required among the hatcheries.

5.0 Recommendations

5. Establish objectives of the program in terms of how many distinct groups (hatcheries x brood years) will require unique marks for each species.

6. Identify funding for capital investment to hatcheries so that physical plant systems have the best chance of delivering a reliable and clear otolith thermal mark. This will reduce a variety of problems associated with detecting mark patterns and undoubtedly reduce errors in that process. The amount of that capital investment is tied to the overall objectives of the program, i.e. the number of unique marks required among all groups.

7. Utilize existing regional otolith laboratories for processing and recovery of otolith marks. Labs in Alaska, Canada and Washington State may be adapted to handle the increased sample load from this endeavor. These facilities are accustomed to rapid turn around times for samples, though additional funding would obviously have to accompany this increased load. Current prices for otolith sample processing and mark recovery range between $7 and $10 per sample.

8. Establish a coordinating body for assignment and quality control of otolith marking. An existing body within the NPAFC can serve as a model and in any case, close communication with this body should be maintained. An otolith mark database must be established as a stand alone or extension of existing mark or tag databases.

6.0 Literature Cited

Brothers, E. B. 1981. What can otolith microstructure tell us about daily and sub-daily events in the early life history of fish? Rapports et Proces-Verbaux de Reunions, Counseil International pour l'Exploration de la Mer. 178: 393-394.

Brothers, E.B., 1990. Otolith Marking. Amer. Fish. Soc. Symp. 7: 183-202.

Calaprice, J.R., Lapi, L.A., and Carlson, L.J. 1975. Stock identification using X-Ray spectrometry and multivariate techniques. Int. North Pac. Fish. Comm. Bull. 32. 81-101.

Campana, S.E. and Neilson, J.D. 1982. Daily growth increments in otoliths of starry flounder (*Platichthys stellatus*) and the influence of some environmental variables in their production. Can. J. Fish. Aquat. Sci. 39: 937-942.

Campana, S.E. and Neilson, J.D. 1985. Microstructure of fish otoliths. Can. J. Fish. Aquat. Sci. 42: 1014-1032.

Hagen, P. 1999. A modeling approach to address the underlying structure and

constraints of thermal mark codes and code notations. NPAFC Doc. 395. 12p. Alaska Department of Fish and Game. Juneau, Alaska.

Hagen, P., Munk, K., VanAlen, B. and White, B. 1995. Thermal mark technology for in-season fisheries management: A case study. Alaska Fishery Research Bulletin 2: 143-155.

Jensen, K.A. and Milligan, P.A. 2001. Use of thermal mark technology for the in -season management of transboundary river sockeye fisheries. NPAFC Technical Report No. 3: 37-38.

Joyce, T.L. and Evans, D.G. 2001. Using thermally marked otoliths to aid the management of Prince William Sound pink salmon. NPAFC Technical Report No. 3: 35-36.

Munk, K.M. and Geiger, H.J. 1998. Thermal marking of otoliths: the RBr coding structure of thermal marks. NPAFC Doc. 367. 19p. Alaska Department of Fish and Game, Juneau, Alaska.

Urawa, S., Hagen, P.T., Meerburg, D., Rogatnykh, A., and Volk, E. 2001. NPAFC Technical Report No. 3. 13-15.

Volk, E.C., Schroder, S.L. and Fresh, K.L. 1990. Inducement of unique otolith banding patterns as a practical means to mass-mark juvenile pacific salmon. American Fisheries Society Symposium 7: 203-215.

Volk, E.C., Schroder, S.L., Grimm, J.J. and Ackley, H.S. 1994. Use of a bar code symbology to produce multiple thermally induced marks. Trans. Am. Fish. Soc. 123:811-816.

Volk, E.C., Schroder, S.L., and Grimm, J.J. 1999. Otolith thermal marking. Fisheries Research 43: 205-219.

V. Large Scale Parentage Inference as an Alternative to Coded-wire Tags for Salmon Fishery Management

John Carlos Garza and Eric Anderson
Southwest Fisheries Science Center, Santa Cruz, California

Salmon fishery management and stock assessment currently use cohort-based mortality models that rely on age-specific tag recovery data as the primary annual input. The current method of choice for large-scale tagging of Pacific salmonids is the coded wire tag (CWT), a small piece of mechanically-inserted metal (1.1 x 0.25 mm long) with a numerical code that is manually cut out of the fish's head post-mortem and read under a microscope. Approximately 1 billion of these tags and nearly 600 miles of wire have been implanted in Pacific salmonids in western North America over the last 30 years.

The CWT has been enormously useful in its 30+ years of use for understanding stock composition of fishery catch, ocean distribution of different salmon stocks and age-specific mortality of all causes. It has been a crucial component of the data used by the Pacific Salmon Commission (PSC) for estimation of fishery mortality of multiple individual salmon stocks in mixed fisheries to implement the Pacific Salmon Treaty allocation of catch between the US and Canada. The management/allocation models that are used by the PSC and other management bodies (e.g. the Pacific Fishery Management Council) are cohort-based and, thus, dependent upon the cohort of origin information garnered from CWT analysis.

However, coded wire tagging and the use of the CWT data in management of Pacific Salmon Treaty fisheries currently face significant challenges. Primary among them is that CWT programs requires enormous tagging effort for a small number of tag recoveries (<2%), and is generally only applied to hatchery stocks, due to logistical problems and potential mortality associated with tagging wild juveniles. In addition, the advent of mass marking and mark selective fisheries pose serious problems to the current management system. Historically, the removal of the adipose fin clip was sequestered as an external mark for the presence of a CWT. However, U.S. law now requires adipose fin clipping of all salmon from federally-supported hatcheries. This mass marking means that up to 80% of the adipose-clipped fish sampled in some salmon fisheries do not hold CWTs. In addition, mark-selective fisheries, which are the reason

for this mandate, cause a violation of one of the fundamental assumptions of the cohort analyses; that hatchery release groups experience the same mortality regimes as genetically-similar, naturally-spawning stocks. This assumption has not been adequately evaluated in Chinook salmon even in the absence of mark selective fisheries.

Genetic tagging methods have a long history in fishery genetics, having been applied to hatchery trout more than 20 years. These methods generally take one of two forms. The first is genetic stock identification (GSI) that uses genotype data and a baseline of allele frequencies to identify individuals to population/stock of origin or estimate stock proportions from a fishery mixture sample. The second is selective breeding of fish in a hatchery, such that all individuals include some unique allele or allelic combination.
However, such genetic methods have been restricted to providing population or hatchery level resolution and can not provide age of individual fish. In addition, genetic similarity between stocks/populations may limit the ability of GSI to separate stocks and selective hatchery breeding to produce genetic tags may run up against substantial operational constraints and other problems.

Several years ago, we proposed the idea of using large-scale parentage inference as an alternative to coded wire tags (Hankin et al. 2005). Originally termed *full parental genotyping*, and now referred to as *parentage-based tagging*, this genetic method does provide age of individual fish and provides exactly the same data as a CWT program, as well as significant additional information. Parentage-based tagging (PBT) is predicated upon the idea that sampling and genotyping the broodstock at a hatchery, or the spawning adults in a natural population, provides genetic tags that are recovered through parentage analysis, thereby providing a highly-efficient, transgenerational tag. For semelparous fish, the identification of parents also provides the age of that fish, not only to cohort or broodyear, but to exact date of fertilization. Since the "tagging" process requires genotyping the parents only and each female produces thousands of offspring, PBT is highly efficient, with one pair of genotypes providing thousands of tag releases. Juvenile fish are not handled at all for PBT.

The general operational routine for PBT is relatively straightforward, particularly in a hatchery setting:

1) Tissue sample broodstock adults at spawning,
2) Genotype the parental tissues with a standard set of molecular markers,
3) Create a reference "parent" database of all sampled adults,
4) Tissue sample catches from fisheries and adults in spawning escapement, and genotype these samples with same set of standard markers,
5) Query parent database to determine if parents were sampled,
6) Determine parental pair, if sampled, and thereby stock and age (cohort) of origin.

While it bears some similarity to standard GSI, PBT is fundamentally different in that it uses a type of matching algorithm to determine Mendelian compatibility of a sample with potential parental pairs present in the reference database. In contrast, GSI uses probabilistic evaluation of the alleles present in a sampled genotype to assess where its constituent alleles are present at highest frequency, as estimated from the data in the baseline database Statistical power for GSI is highly dependent upon the number of alleles at a locus, and less so on the number of loci. In contrast, the power for PBT is highly dependent upon the number of loci, since each locus provides an opportunity for a Mendelian incompatibility that excludes a fraction of the potential parental trios.

Such parentage analysis is a special case of the well-developed methods of pedigree reconstruction using genetic markers, which is the basis for most gene mapping and is used in legal situations to establish kin relationships. Traditionally, simple exclusion methods that rely on Mendelian incompatibilities were used in parentage analysis, but more recently maximum likelihood methods of analysis have become prevalent. However, the concept of performing parentage analysis on such a large scale, and in a mixed fishery context is novel. Implementation of this concept required the development of additional analytical methodology and further evaluation of the feasibility of such parentage analysis when there are such a large number of potential parent pairs. We undertook such development and evaluation for large-scale parentage inference in the last two years (Anderson and Garza 2006), and established the feasibility of performing PBT for salmon management.

In this work, we determined false positive rates (the probability that a trio identified as parents/offspring was done so incorrectly) for a wide variety of potential parentage inference situations. We determined the relationship between false positive rates and the amount of genetic data necessary, as well as evaluating the effects of genotyping error and the presence of close kin in the mixture samples. We also developed two new algorithms for more efficiently evaluating potential matches in the parent database.

In early stages of this simulation study, we quickly realized that the importance of having a low genotyping error rate and a high throughput, low cost genotyping system would mean that single nucleotide polymorphism (SNP) markers would need to be the basis for any large scale application of PBT. While SNP markers are currently not widely available in great numbers for all salmon species, they will be in the next several years. Microsatellite markers can certainly be used in parentage analysis, and are currently being so employed, but it is our strong contention that they are neither feasible nor optimal for use in any coastwide application of PBT. This is because of higher genotyping error rates, the lack of portability of data and the high cost, primarily due to staff time, necessary for genotype collection. Because of the sensitivity of parentage inference to genotyping errors, and the need for large genotypes collected at minimal cost, we have determined that large SNP genotypes are the molecular marker of choice for future applications of PBT. Because of this, all of our analytical and operational evaluation of PBT has centered on the use of large SNP genotypes for parentage identification.

The evaluation of genetic data and statistical power in the Anderson and Garza (2006) work found generally that approximately 100 SNP loci would result in a false positive rate of less than one per 10^{-13} parent/offspring trios examined. This is a rate that is essentially without errors from the genotype data, which would make it similar qualitatively to a CWT (although the error rate for CWTs due to problems with the tag coding is not well known). This analysis assumed a genotyping error rate that was similar to the highest one reported in the literature for SNP markers (1%) and for 90% power. Trying to assign the last 10% of offspring with high confidence raises the amount of data necessary by much more than 10%. It is also worth noting that these analyses assume a mean minor allele frequency of 20%, but it will be possible to high-

grade loci from among the large numbers that will be available in several years, such that the mean frequency is higher and the number of loci necessary lower.

One of the most important results of this work is the elucidation of a logarithmic relationship between the number of loci necessary for high accuracy parentage assessment and the number of potential parent/offspring trios that must be evaluated. This means that the number of SNP loci necessary for parentage analysis rises linearly as the number of parental trios possible rises exponentially. So the scope of parentage analyses necessary for coastwide implementation of a PBT program could never grow too large to be addressed with a relatively small and feasible number of genetic markers.

The presence of kin in the parent database does raise the probability of false positives for some kin relationships. What this means is that a family member may be mistaken for an actual parent in parentage inference when present. However, in general, only full siblings and double first cousins are problematic in the parentage analyses. In addition, from the point of view of cohort analysis, only false positives that incorrectly identify close kin as both parents will result in an error of importance (i.e. wrong age or hatchery). These are less likely errors than those that only misidentify one parent and therefore of less concern. Moreover, recording matings or the sorting of broodstock by date of spawning nearly eliminates the problem of false positives due to kinship.

The analyses in the Anderson and Garza (2006) study are actually quite conservative with respect to application in coastwide management of salmon fisheries. This is because those analyses were based upon the assumption that all of the fish that might need to be discriminated are part of a large undifferentiated (e.g. lacking population structure) population. When hatcheries or natural populations included in the parent database have non-zero values of F_{ST} or other genetic distance measures (i.e., there is structure present), it decreases the probability of false positive parental assignments to the individuals in the parent database that are from those differentiated populations. In general, the probability of a false positive parentage assignment for an individual fish decreases by an order of magnitude with an F_{ST} value of 0.05 between the population/hatchery broodstock of origin for the tagged fish and the population of origin of the

potential parent. Since there is substantial population structure in Chinook salmon and no two hatcheries have broodstock with non-significant F_{ST}, the probability that fish from different hatcheries might be identified as close kin is even lower than found in the simulation work.

One of the concerns that arises with a PBT program is the large number of samples that might need to be genotyped for such a program, particularly if it is necessary to achieve the number of tag recoveries for smaller stocks that are currently possible with an increased CWT insertion rate and a method for external identification of fish carrying CWTs. There are several ways that the amount of genotyping data that must be collected can be minimized and the cost of implementing PBT decreased to the point of feasibility. Reducing the amount of genetic data that needs to be collected is one way to achieve this. The amount of genetic data necessary to accurately infer parent/offspring trios is dependent upon the number of potential trios that must be evaluated in the parent database. In an ideal program, or a relatively small scale one, all matings could be recorded and associated with tissue samples and genotype data. This limits the number of trios that need be examined to those including actual mated pairs. However, accurate cataloguing of all mating information and its association with tissue samples is an enormous amount of effort by hatchery staff and would not be feasible for many hatchery programs.

However, there is a useful alternative that does not require the recording of all matings but still dramatically reduces the number of parent/offspring trios that must be examined, and therefore the amount of data for accurate inference. That method is to simply separate hatchery broodstock samples by day of spawning and preferably by sex as well. This simple step, which we refer to as day bins, will decrease the number of possible trios by at least an order of magnitude and therefore the amount of genetic data necessary. Another way to reduce the amount o data necessary is to use SNP panels that have only loci with high minor allele frequencies, since a marker with two alleles at equal frequency has the most power for pedigree reconstruction. In the Anderson and Garza (2006) study, all evaluation was with marker loci that had mean minor allele frequency above 0.2 (20%). Each increase in mean minor allele frequency of 0.1 for the marker panels decreases the false positive rate by an order of magnitude. In practice, however, it may be difficult to construct SNP marker panels with mean minor allele frequency greater than about 0.3.

Other ways to decrease costs are to decrease the tagging rate by sampling and genotyping less of the broodstock. However, since PBT generally requires both parents to be sampled to achieve identifications, the decrease must be done in such a way that all sampled broodstock are from matings in which both fish are sampled. Otherwise the decrease in sampled broodstock will have a disproportionate effect on the tagging rate, since fish with only one parent genotyped will not be tagged. Perhaps the most obvious way to decrease the genotyping burden is to simply incorporate more uncertainty into the management models, either through acceptance of a higher false positive rate (i.e. more identification errors) or through smaller sample sizes from mixed fisheries.

It is hard that this point to estimate the costs of a fully implemented PBT system relative to the current CWT system. Among the reasons for this are that the costs of the CWT program, both now and in 5 years (a realistic time frame for implementation of any alternative to CWT analysis), are very hard to determine, particularly with the advent of electronic detection and mark selective fisheries. In addition, whereas CWT analysis (and microsatellite-based genetic analysis) have relatively constant costs, the costs of high-throughput SNP genotyping are decreasing rapidly as new technologies are transferred from the field of human genetics to salmon genetics laboratories. However, preliminary analyses indicate that the cost of tagging with PBT is currently lower than the cost of tagging with CWTs and that the cost of tag recovery with PBT is higher than with CWTs.

A very attractive element of a PBT program is the abundant corollary data that results from such a tagging regime. The primary additional data that comes from such a program are the many large pedigrees for multiple salmon stocks. Such pedigrees will allow determination of near parametric values for variance in family size and marine survival, and the comparison of many parameters for hatchery and wild stocks. With the successful reconstruction of large pedigrees, this project will set the stage for future estimation of heritability of physical and life history traits in Chinook salmon, which in turn will allow the prediction of the consequences of different hatchery protocols and fishery regimes. This is also the first step in the mapping and identification of the genes responsible for characters such as fecundity, age at maturity, and run-timing, which will be of great interest to both geneticists and fishery managers.

Another attractive element of a PBT program is the prospect of integration with a traditional GSI program. Such an integrated PBT/GSI program would allow fishery sampling to proceed without respect to mark or tagging status, since all fish would provide some "tag" information. Fish from hatcheries where broodstock are sampled would be identified to hatchery and cohort of origin, and all other fish would be identified to stock of origin, or used in estimating mixture proportions. While there are some important logistical considerations that would need to be addressed for such an integrated program, the prospect of a genetic sampling program where every fish is "tagged" is sufficiently compelling that it may merit further evaluation.

Among the most important logistical challenges for implementation of a coastwide PBT program of any type is the need to find standardized SNP panels that have sufficient power for parentage analysis in all indicator hatcheries. The optimization of such panels will require a larger pool of markers available from which to choose and will require a broad multi-jurisdictional effort. However, preliminary analyses of data from the human genome project, where more than 3 million SNP markers have been described, indicates that it is certainly feasible to find such a set of markers.

From the point of view of the Logistics workgroup of the Pacific Salmon Commission GSI workshops, perhaps the most important step to be taken with respect to PBT is to ensure that the elements of the multi-jurisdictional database(s) proposed be able to accommodate the data and queries that would be necessary with a large-scale PBT tagging and sampling program. This includes the very large number of genotypes that would be collected in such a program, the ability to accept very large queries that include all potential parents for a given set of tagged fish, and the ability to integrate a PBT program with the developing GSI program.