Title
CREATION AND UTILIZATION OF NOVEL GENETIC METHODS FOR STUDYING AND IMPROVING MANAGEMENT OF CHINOOK SALMON POPULATIONS

Author
Clemento, Anthony

Publication Date
2014-10-28

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

CREATION AND UTILIZATION OF NOVEL GENETIC METHODS FOR STUDYING AND IMPROVING MANAGEMENT OF CHINOOK SALMON POPULATIONS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

OCEAN SCIENCES

by

Anthony J. Clemento

December 2013

The Dissertation of Anthony J. Clemento
is approved:

Dr. John Carlos Garza, Chair

Dr. Jonathan Zehr

Dr. Grant Pogson

Dr. Eric Anderson

Dean Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Creation and utilization of novel genetic methods for studying and improving management of Chinook salmon populations

by

Anthony J. Clemento

As a major component of fisheries in the northern Pacific Ocean, Chinook salmon (*Oncorhynchus tshawytscha*) are of significant management concern. Their anadromous life history, in which adult fish migrate to their natal streams, leads to populations (stocks) that are genetically distinguishable and, ideally, would be managed independently. Many of these stocks, particularly at the southern end of the species' range, have experienced serious declines, which has motivated widespread hatchery production and supplementation. The physical coded-wire tagging (CWT) program currently used to track hatchery fish, and ultimately to supply information for cohort-based fishery harvest models, is increasingly ineffective and can no longer sustain the data demands of fishery managers and scientists. Also, current genetic tools utilizing microsatellite markers do not scale well to the enormous number of fish that need to be analysed, have error rates that are too high for individual- and pedigree-based methods, and genotype inconsistently across laboratories, creating an impediment to interjurisdictional collaboration. However, the next generation of genetic markers, single nucleotide polymorphisms (SNPs), do have low enough error rates and are amenable to the high-throughput genotyping required for ocean fishery stock identification and large-scale

tagging of hatchery fish via pedigree reconstruction. Here we describe the successful identification of 117 novel SNP loci using genomic data from a sister salmonid taxon and demonstrate their substantial power for discriminating five major stocks of salmon from the three largest basins on the Pacific coast of North America. We then assemble a panel of 96 SNP loci and genotype over 8000 individuals from 69 distinct populations for construction of a baseline for genetic stock identification (GSI) and show that it has, effectively, near-maximum power for discriminating most Chinook salmon stocks captured in mixed-stock fisheries off the coasts of California and Oregon. This baseline is used to confidently assign over 2000 ocean-caught Chinook to their source population and demonstrate over 99% concordance between the GSI assignments and identifications from CWTs recovered from these fish. The same panel of SNPs is also used to implement a large parentage-based tagging (PBT) experiment at one of the most productive hatcheries in the Central Valley of California. PBT involves genotyping reproducing adults and using their genotypes as intergenerational genetic tags that are recovered through parentage inference with their progeny. By genotyping over 12,000 individuals from six complete brood years, we show that the large number of resulting pedigrees effectively provide the same age and stock information as traditional CWTs, but also can be used to inform hatchery breeding practices, estimate the heritability of physical traits and eventually can serve as the basis for detailed linkage maps and associated mapping of quantitative trait loci. The genetic resources developed here are a substantial improvement over current methods and are fundamentally changing the way salmon populations are studied, monitored and managed.

To Danielle and Sylvia, my family and friends, without whom this would never have happened.

Acknowledgments

While many people deserve recognition for their contributions towards the successful completion of this dissertation, none more so than my advisor John Carlos Garza. When I was just another voice on the telephone asking for help, Carlos gave me the chance to come to Santa Cruz and work with the most amazing group of people and scientists that I have ever encountered. His unwavering confidence in my abilities, and unending tolerance of my moods, has made me a better scientist, collaborator and researcher. He has brought me into this new family of West Coast salmon geneticists and made me feel right at home.

I would also like to express my gratitude to my committee: Jon Zehr, Grant Pogson and Eric Anderson for their patience while I found my way to the end of this journey. I greatly appreciate the time they spent reading, editing and discussing my work and I give them full credit for being great role models of what a scientist should be. Special recognition to Eric, who has taught me to love the command line and has spent countless hours passing on his extreme knowledge of all things computer.

Many thanks to the entire Molecular Ecology team at the Santa Cruz Lab. I owe my lab skills to the patient tutelage of Libby Gilbert-Horvath and would not have enjoyed using them without the companionship of the original lab crew: Big (Aguilar) and Little (Martinez) Andy, Scott Blankenship and Cheryl Dean. I have also learned so much from the other graduate students and post-docs who have cruised through the lab, all of whom I now consider life-long friends: Hilary Starks, Vicky Pritchard, Eric Crandall,

and Dan Barshis. I need to send a special thanks to Devon Pearse for sharing an office those many years and sharing his insight and experience almost every day. On a level of her own, Alicia Abadía-Cardoso has been the best co-pilot ever as we have been on this graduate student journey together. I very much look forward to continued work with Martha, Diana, Cassie and Vanessa - you are all truly excellent.

I would be remiss if I did not mention my wife Danielle and my daughter Sylvia; their patience and tolerance during some of the more trying times was admirable and often it was only their love and caring that made everything alright. When there seemed to be no end in sight, they never lost faith in me and their confidence in my ability to git-er-done was unwavering. And finally, I thank my parents, who gave me every opportunity to succeed, and I can only hope that I have exceeded their expectations.

The text of this dissertation includes reprints of the following previously published and submitted material:

Chapter 1: Clemento, A.J., A. Abadía-Cardoso, H.A. Starks, and J.C. Garza. 2011. Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, *Oncorhynchus tshawytscha* Molecular Ecology Resources **11**(Suppl. 1):50-66.

I performed the majority of the DNA sequencing, data analysis and manuscript preparation for this project. H.A. Starks and A. Abadía-Cardoso assisted with some of the laboratory work and provided valuable suggestions for analyses and presentation of results. J.C. Garza (committee chair) directed and supervised the method development and provided assistance with the writing and editing of the final manuscript.

Chapter 2: Clemento, A.J., E.D. Crandall, J.C. Garza and E.C. Anderson. 2014. Evaluation of a SNP baseline for genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) in the California Current Large Marine Ecosystem Fishery Bulletin, *In Review*.

For this project, I performed all of the genotyping of baseline populations and carried out the majority of analyses and manuscript preparation. E.D. Crandall coordinated genotyping of the ocean fishery samples and compared results to coded-wire tag data. E.C. Anderson directed the power analyses for marker selection and evaluation of the final genotyping panel for population discrimination. Both E.C. Anderson (committee member) and J.C. Garza (committee chair) directed and supervised the method development and provided assistance with the writing and editing of the final manuscript.

This research was funded in part by a University of California Marine Council, Coastal Environmental Quality Initiative (CEQI) grant (2006-2007) and a California Sea Grant/ CALFED Science Fellows Program grant (R/SF-24, 2007-2011).

Introduction

The use of genetics to study and monitor populations is now ubiquitous as molecular markers have been discovered and made available for a wide range of species. While the concept of genetic research and monitoring can be interpreted broadly, most implementations utilize molecular data to discern population structure, study evolutionary and ecological processes, or measure population genetic parameters over time (Schwartz *et al.* 2006). Genetic markers can also be employed in lieu of physical tags for mark-recapture experiments (Palsbøl 1999) and as a tool for performing parentage assignment (Blouin 2003), which can yield insights into mating systems (Pearse 2001) and elucidate population dynamics (Hauser and Carvalho 2008). Pedigrees resulting from parentage assignment can also be used to estimate heritability and map genes involved in the inheritance of physical traits (e.g. size, growth, reproduction, migration) to their chromosomal locations (Fisher 1918, Lynch and Walsh 1998, Wu *et al.* 2007). Such data can provide a predictive framework for assessing the effects of environmental change on populations or the impacts of different management and conservation actions on captive or wild populations.

As a keystone species in the marine, terrestrial and freshwater ecosystems of

the West Coast of North America (Willson and Halupka 1995, Cederholm *et al.* 1999, Helfield and Naiman 2006), Chinook salmon (*Oncorhynchus tshawytscha*) provide an excellent system for ecological and population genetic investigations. Chinook salmon are anadromous, a life-history strategy characterized by adult migrations from the ocean to spawn in their natal stream, followed by subsequent juvenile migrations back out to the ocean. This spawning site fidelity creates genetically distinct populations and can provide an opportunity for local adaptation (Utter 1989; Taylor 1991). Chinook salmon exhibit broad variation in the duration of their juvenile freshwater residency, timing of adult spawning migrations and patterns of reproductive maturity (Taylor 1990, Groot and Margolis 1991). Two of the most common reproductive ecotypes described in the species, "spring-run" and "fall-run", often inhabit the same river systems. Spring-run fish are sexually immature as adult migrants, holding in deep river pools far up river drainages until they mature and spawn in fall and winter months; for fall-run fish sexual maturation is coincident with upstream migration in the Fall and consequent spawning in lower river reaches (Quinn 2005). Chinook populations are distributed around the Pacific Rim from the Central Coast of California in the east to Japan and coastal Russia in the west. They are the target of highly valuable commercial and recreational fisheries throughout the northern Pacific Ocean and continue to be a primary source of sustenance for Native American peoples.

Salmon, as with many other marine fish species (i.e. sardines and anchovies), naturally experience high variability in abundance. In the North Pacific, salmon abundance has been shown to be tied to naturally occurring climate oscillations on decadal

timescales (Mantua *et al.* 1997). The predominant oceanographic features of the northeastern Pacific are the eastern edges of the Alaskan subpolar gyre and the north Pacific subtropical gyre, which are alternately affected in the different phases of the Pacific Decadal Oscillation (PDO; Hare and Mantua 2000, Mantua and Hare 2002). Likely driven by changes in primary production associated with sea-surface temperature and upwelling (Cole 2000, Hinke *et al.* 2005), salmon fisheries in Alaska and along the Pacific Northwest coast have alternately experienced depressions corresponding with PDO cycles (Kruse 1998, Hare *et al.* 1999). While fishery scientists can utilize this type of oceanographic information for increased ecosystem-based management (Field and Francis 2006), salmon still face challenges not only in the marine environment but also in rivers and streams (Bisbal and McConnaha 1998).

Degradation of riverine spawning habitat, diversion of fresh water for human use, over-fishing, hatchery domestication selection, and highly variable ocean conditions have all been implicated in the recent declines of Chinook salmon populations in the southern portion of the species' range (Lindley *et al.* 2009). As a consequence, many Chinook salmon populations in the contiguous United States are now listed as threatened or endangered under the federal Endangered Species Act (Myers *et al.* 1998). In order to mitigate for the multiple impacts threatening Chinook salmon populations, state and federal agencies now produce millions of fish annually in hatcheries. These hatchery fish, intended to reduce variability in ocean abundance, provide fishing opportunities, and satisfy Native American treaty obligations, comingle with wild fish in the ocean and can compose the majority of the catch in certain times and places (Beamish

et al. 1997). However, the ecological consequences of releasing large numbers of hatchery fish are poorly understood (Levin *et al.* 2001), and may be severely compromising efforts to preserve wild populations (Hilborn 2011). Some natural populations may now be composed primarily of hatchery individuals or their offspring (Barnett-Johnson 2007).

The Central Valley of California was once the second largest source of Chinook salmon on the U.S. West Coast (after the Columbia River), despite being the southernmost drainage to support the species. Dominated by the Sacramento River to the north and the San Joaquin River to the south, the Central Valley historically maintained wild populations that numbered in the millions (Yoshiyama *et al.* 2001). As in most river systems that support the species, Chinook salmon from the Central Valley display a wide variety of life-history strategies, varying in the timing of migrations and sexual maturation (Yoshiyama *et al.* 1998). However, a majority of the historical spawning habitat for salmon has been eliminated as rivers are engineered for flood control and water is appropriated for agriculture, domestic water supplies, and hydroelectric production (Fisher 1994). Because spring-run Chinook salmon, which migrate upstream months before the fall-run form, generally penetrate further up into watersheds, the many large dams on Central Valley rivers disproportionately eliminated their primary spawning and holding habitats. Spring-run Chinook, which historically were more numerous than the fall-run in ocean fisheries (Yoshiyama *et al.* 2001), have experienced severe declines in California and are now listed as threatened under the California state and the federal endangered species acts. Six hatcheries now produce the majority of

Chinook (spring- and fall-run) that return to the Central Valley (Fisher 1994).

Ensuring sustainability and the persistence of salmon populations while providing fishing opportunities can be a complex task. Underestimation of the contribution from specific stocks can have serious conservation implications (e.g. overfishing and/or extinction of wild stocks), while overestimation can leave the resource underexploited, potentially costing the fishing industry and coastal communities millions of dollars (Michael 2010). Management of Pacific Ocean salmon fisheries off North America can be roughly divided into three regions: California and Oregon fisheries are managed by the Pacific Fishery Management Council (PFMC); fisheries in Washington, British Columbia, Canada and southeast Alaska are subject to the international Pacific Salmon Treaty, reported to and regulated by the Pacific Salmon Commission (PSC); and fisheries further north and west in Alaska are managed by the state, with salmon by-catch under the purview of the North Pacific Fishery Management Council. The primary method of assessing fishery impacts is through cohort analysis models. These models attempt to account for fishery mortalities on groups of fish (primarily of hatchery origin) born in the same year (and therefore the same age) through time. However, the uncertainty in the models coupled with the difficulty of estimating fishery impacts on highly age-structured populations has left scientists and managers in need of better data (Hankin 2005). Currently, the primary source of information for cohort analysis models comes from coded wire tags. Management of Chinook salmon fisheries in the eastern Pacific Ocean depends on an elaborate marking and coded-wire tagging program, implemented and monitored by State and Federal agencies. The primary focus

of this monitoring program is the millions of fish produced annually in hatcheries along the West Coast of North America. Data extrapolated from the program are used to parameterize stock-specific forecasting models and to estimate ocean abundance indices, which are then used to set fishing areas and seasons, determine quotas and legal gear and establish catch limits and size restrictions (Hyun, 2012). The accuracy of these models and the resulting abundance estimates are highly dependent on the quantity and quality of data input.

Genetic methods have long been used to study various aspects of salmon biology and ecology. The earliest genetic analyses of salmon utilized electrophoretically detectable protein polymorphisms known as allozymes, which were sufficient for discriminating populations at a relatively coarse geographic scale (Milner *et al.* 1985; Tessier *et al.* 1995; Allendorf and Seeb, 2000). With the introduction of polymerase chain reaction (PCR) and development of modern genetic techniques, a variety of new marker types became available to salmon scientists, including mitochondrial markers (mtDNA), amplified fragment length polymorphisms (AFLPs), minisatellites, and microsatellites (Beacham *et al.* 1996; Smith *et al.* 2001; Flannery *et al.* 2007; Clemento *et al.* 2009). For almost two decades, microsatellites have been employed in studies of population structure, behavioral ecology, and pedigree relationships, as well as for individual and genetic stock identification, because of their extensive polymorphism (Banks *et al.* 2000; Smith *et al.* 2005b/c; Seamons *et al.* 2004; Pearse *et al.* 2007; Smith *et al.* 2007). While their high variability provides sufficient statistical power for many population genetic applications, microsatellites can have high genotyping error and mutation rates. In ad-

dition, combining microsatellite data generated in different laboratories, or on different instrument platforms in the same laboratory, can require an onerous standardization procedure to account for subtle differences in electrophoretic conditions and resulting instrument output. This standardization process can add significant time and expense to multilateral database construction and collaborative research (Seeb *et al.* 2007).

As genomic resources for salmonids have expanded, single nucleotide polymorphism (SNP) markers have become an increasingly common choice for population genetic studies (Morin *et al.* 2004). A SNP is variation at a single DNA base at a known location in the genome. SNPs are abundant and can be found in both coding regions, where they may be targeted by selection, and in non-coding regions, where they are often assumed to not be the direct targets of natural selection (Vignal *et al.* 2002, Nosil *et al.* 2009). Since SNPs are generally bi-allelic, comparable power to highly polymorphic microsatellite markers is attained by using larger numbers of loci (Anderson and Garza 2006, Narum *et al.* 2008). At the same time, new technologies have yielded platforms (i.e. nanofluidics, microarrays) for efficient, high-throughput genotyping at large numbers of SNP markers. Data generation on these platforms requires significantly less time and money and resulting genotypes are subject to much lower error rates than microsatellites. While SNP development for Chinook salmon began around the middle of the last decade (Smith *et al.* 2005a, 2005b, 2006; Campbell and Narum 2008), only about 30 markers were available at the outset of the research described here. Discovery of larger numbers of SNPs (described in Chapter 1) is the first step towards implementing new SNP-based methods for genetic stock identification (GSI; Chapter 2) and

parentage-based tagging (PBT; Chapter 3) of Chinook salmon.

In Chapter 1, I further motivate the need for SNP development in Chinook salmon and describe the methodology we used to discover 117 novel SNP markers. Despite the broad importance of the species, very few genomic resources are available, so the utility of expressed sequence tag (EST) data from a more well-studied, somewhat closely-related, species was explored, steelhead trout (*Oncorhynchus mykiss*), for primer design. Although ESTs are by definition in coding regions, it can be expected that the resulting genomic DNA sequences will encompass introns, where SNPs may also be observed. I hypothesize that employing a balanced ascertainment panel for sequencing, with representatives drawn from a broad geographic range, will yield SNPs with increased power for population discrimination. Furthermore, implementing strict criteria about observed genotypes in the sequencing data should high-grade for SNPs with sufficient allele frequencies for GSI and PBT applications.

In Chapter 2, I describe the development of a coastwide genetic database for identifying Chinook salmon caught in fisheries in the California Current Large Marine Ecosystem. It is hypothesized that a single panel of 96 SNPs can be sufficient for providing the stock of origin for fish captured in large mixed-stock ocean fisheries. I explain the procedures used to select SNPs for inclusion in the panel and evaluate the power of the new baseline for genetic stock identification using valid statistical methods. The markers designed using the balanced ascertainment strategy described in Chapter 1 can be expected to be particularly effective for GSI, even if there are not large allele frequency differences between populations. I also demonstrate that inference

from the genetic data is comparable to that generated by physical tags for management applications.

In Chapter 3, I hypothesize that the same panel of SNP markers employed in Chapter 2 will be equally effective for intergenerational genetic tagging of a large hatchery population from the Feather River Hatchery, CA. The parentage-based tagging technique described here is fundamentally different from GSI; rather than assigning fish to their most likely management unit using allele frequencies, individuals are specifically identified by inferring parentage. The method is expected to be sufficiently powerful that the parent pairs assembled in pedigrees using the genetic data will match those recorded during spawning at the hatchery. Additionally, I demonstrate the utility of knowing large numbers of pedigrees for estimating the impacts of artificial propagation in the hatchery. Pedigrees can be used not only to assess important population genetic parameters (i.e. inbreeding) in unprecedented detail but also to estimate the heritability of observed life-history traits. It is likely that the tools developed and described here will substantially change the way Chinook salmon are managed, both in ocean fisheries and at hatcheries.

Chapter 1

Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, *Oncorhynchus tshawytscha*[1]

1.1 Abstract

Molecular population genetics of non-model organisms has been dominated by the use of microsatellite loci over the last two decades. The availability of extensive genomic resources for many species is contributing to a transition to the use of single nucleotide polymorphisms (SNPs) for the study of many natural populations. Here we describe the discovery of a large number of SNPs in Chinook salmon, one of the worlds most important fishery species, through large-scale Sanger sequencing of expressed se-

[1]published: Clemento, A.J., A. Abadía-Cardoso, H.A. Starks, and J.C. Garza. 2011. Molecular Ecology Resources **11**(Suppl. 1):50-66.

quence tag (EST) regions. More than 3MB of sequence was collected in a survey of variation in almost 132KB of unique genic regions, from 225 separate ESTs, in a diverse ascertainment panel of 24 salmon. This survey yielded 117 TaqMan (5' nuclease) assays, almost all from separate EST regions, which were validated in population samples from five major stocks of salmon from the three largest basins on the Pacific coast of the coterminous United States: the Sacramento, Klamath and Columbia Rivers. The proportion of these loci that was variable in each of these stocks ranged from 86.3 to 90.6% and the mean minor allele frequency ranged from 0.194 to 0.236. There was substantial differentiation between populations with these markers, with a mean FST estimate of 0.107, and values for individual loci ranging from 0 to 0.592. This substantial polymorphism and population-specific differentiation indicates that these markers will be broadly useful, including for both pedigree reconstruction and genetic stock identification applications.

1.2 Introduction

Chinook salmon (*Oncorhynchus tshawytscha*) is the largest species of Pacific salmonid and one of the worlds most commercially and recreationally valuable fishery species. Chinook salmon are anadromous, meaning that they hatch in rivers and streams, migrate to the ocean during either the first or second year of life, and then typically return to their natal stream to spawn. This homing creates geographic population structure and facilitates the potential local adaptation of populations and larger groups.

In the marine environment, stocks from different rivers, hatcheries and ecotypes, as well as fish of different ages, commingle, making it difficult to quantify catch composition or avoid stocks with depressed abundance in ocean fisheries. Degradation of riverine spawning habitat, diversion of fresh water for human use, over-fishing, hatchery domestication selection, and highly variable ocean conditions have all been implicated in the recent declines of populations in the southern portion of the species range (Lindley *et al.* 2009). As a consequence, many Chinook salmon populations in the contiguous United States are now listed as threatened or endangered under the federal Endangered Species Act (Myers *et al.* 1998). Populations in California have seen particularly severe reductions over the last decade, culminating with complete closures of the commercial fishery off California and Oregon in 2008 and 2009 (Lindley *et al.* 2009).

Population genetics has played a prominent role in salmon research and management over the last several decades. However, the predominant type of molecular genetic marker used has varied substantially over time. Prior to and immediately following the introduction of the polymerase chain reaction (PCR), allozymes were the primary type of genetic marker available for fish biologists (Myers *et al.* 1998; Waples *et al.* 2004). Following the introduction of PCR, came many other marker types including mitochondrial (mtDNA), amplified fragment length polymorphisms (AFLPs), minisatellites, and microsatellites (Beacham *et al.* 1996; Smith *et al.* 2001; Schlotterer 2004; Flannery *et al.* 2007; Clemento *et al.* 2009). Microsatellites, in particular, have been employed broadly in salmonids for studies of population structure, behavioral ecology, and pedigree relationships, as well as for individual and genetic stock identification, because

of their extensive polymorphism (Banks *et al.* 2000; Smith *et al.* 2005b/c; Seamons *et al.* 2004; Pearse *et al.* 2007; Smith *et al.* 2007). This variation provides substantial statistical power for many population genetic applications, but related to this is that microsatellites can have high genotyping error and mutation rates. In addition, combining microsatellite data generated in different laboratories, or on different instrument platforms in the same laboratory, may require a non-trivial standardization process to account for subtle but ubiquitous differences in electrophoretic conditions and resulting instrument output. This standardization process adds significant time and expense to multilateral database construction and collaborative research (Seeb *et al.* 2007).

More recently, single nucleotide polymorphism (SNP) markers have come to prominence (Morin *et al.* 2004). A SNP is a variation in the base present at a specific nucleotide site in the genome. SNPs are the most abundant polymorphism in vertebrate genomes, with a SNP present every 100-500bp on average (Vignal *et al.* 2002). They are common in both coding and non-coding regions of the genome and are typically biallelic, so analytic power similar to that provided by microsatellites is achieved by using larger numbers of loci (Anderson and Garza 2006; Narum *et al.* 2008; Glover *et al.* 2010). SNPs require substantially less laboratory staff time for allele calling and with the advent of new high-throughput genotyping technology, such as nanofluidics and spotted microarrays, data can be generated more quickly and at lower cost than for other marker types. Moreover, standardization requires only that laboratories agree to reporting standards and, ideally, that they use an identical, or overlapping, set of markers. While SNPs have seen extensive use in humans and model organisms, other

research communities have been slow to transition to SNP-based data collection, primarily because of the lack of genomic resources available for non-model species and the costs and effort involved in marker development.

In recent years, SNP development for Pacific salmonids has begun (Smith *et al.* 2005a, 2005b, 2006; Aguilar and Garza 2008; Campbell and Narum 2008; Campbell *et al.* 2009; Abadía-Cardoso 2011) and there are currently a handful of SNP assays available for all of the Pacific salmonid species. Nevertheless, many more are necessary for a number of the applications in which genetic markers are currently in use, including pedigree reconstruction, genetic stock identification, linkage map construction and QTL mapping. Moreover, since many of the existing assays were developed with specific applications in mind, they are frequently of limited utility in populations or phylogenetic lineages that were not part of the discovery process, due to ascertainment bias (Clark *et al.* 2005; Smith *et al.* 2007; Albrechtsen *et al.* 2010). The implementation of SNP-based methods, such as large-scale parentage inference (Garza and Anderson 2007), in California and other marginal parts of the species range requires many additional SNP assays.

SNP discovery typically involves examination of DNA sequence data from multiple individuals at the same locus, or identification of heterozygous nucleotide sites in a single individual. When only a small number of individuals from selected populations are used to discover or ascertain SNP variation, an ascertainment bias is introduced. This bias results in the allele frequency spectrum being shifted upward, with an underrepresentation of rare SNPs, which leads to overestimates of genome-wide heterozygosity

and population differentiation (Clark *et al.* 2005; Smith *et al.* 2007; Albrechtsen *et al.* 2010). This ascertainment sampling bias also results in SNPs that are not as polymorphic in other parts of the species range. Clark *et al.* (2005) recommend the use of standardized ascertainment criteria and a large ascertainment sample of known origin that includes individuals from outside of the primary focus range to reduce these biases and provide marker loci with broad utility.

Despite the species' importance, genetic resources for salmonids are still relatively scarce; there is not a complete genome sequence for any salmon species and there is not even a published linkage map for Chinook salmon. There is, however, a large library of expressed sequence tags (ESTs) from rainbow trout (*Oncorhynchus mykiss*), archived by the Gene Index Project (http://compbio.dfci.harvard.edu/tgi/). These ESTs include full or partial cDNA sequences that have been reverse transcribed from mRNA (Bouck and Vision 2007). Rainbow trout is in the same genus as Chinook salmon and previous work has shown that primers derived from *O. mykiss* can be successfully used to isolate DNA fragments and discover SNPs in other *Oncorhynchus species* (Smith *et al.* 2005a). Primers for conserved regions of known genes are also available in the literature (e.g. Moran 2002).

Here we describe the discovery, design of molecular assays, and evaluation of 117 new SNP assays for Chinook salmon, more than doubling the number of published SNP markers for use in the species. We sequenced genic regions in the Chinook genome from a geographically and phenotypically diverse ascertainment sample of 24 fish. We targeted 480 loci from ESTs of unknown function, as well as genes whose functions are

well described, and designed more than 150 5' exonuclease (TaqMan) assays. Assays were tested and validated by genotyping 337 individuals from five major lineages of the species, from the three largest rivers on the west coast of the coterminous United States (Sacramento, Klamath and Columbia Rivers) and the details of the resulting 117 validated assays are reported here.

1.3 Methods

1.3.1 Primer Design and PCR

Oligonucleotide primers were designed for 480 ESTs randomly selected from the *O. mykiss* Gene Index database of EST sequences. A secondary targeted gene approach was undertaken using primer information from 11 genes from published sources (e.g. Moran 2002) or in GenBank. Primers were designed using Primer3 (v.0.4.0; Rozen and Skaletsky 2000) and targeted EST segments 400-500 bp in length, so that genomic DNA fragments would generally be smaller than 1000bp, even if they contain introns. These primers (sequences are available from the authors upon request) were then used to amplify genomic DNA from a geographically and phylogenetically diverse ascertainment sample of 24 Chinook salmon, including fish from California populations with which we are actively working (Sacramento: Feather River-Spring and Fall, n=12; Sacramento: Butte Creek-Spring, n=2; Eel River, n=2; Klamath River, n=2), and also from elsewhere in the North American range of the species, including Washington (Columbia-Kalama River-Spring, n=2) and Canada (Thompson River-Spius Creek, n=2; Nanaimo Creek,

n=2). For the California samples, DNA was extracted from dried caudal fin clips using DNeasy 96 kits on a BioRobot 3000 (Qiagen, Inc.). For the other samples, previously extracted and frozen DNA was provided by collaborators.

Polymerase chain reaction (PCR) was carried out in 15μL single-locus reactions using Applied Biosystems (ABI) reagents as follows: 1.5μL of 10X buffer, 0.9μL of 1.5mM $MgCl_2$, 1μL of 2.5mM dNTPs, 1μL of 5mM primers (forward and reverse), 6.6μL of deionized water, 0.05μL of AmpliTaq 5U/μL DNA polymerase, and 4μL of genomic DNA. The thermal cycling routine employed was a modified step-down protocol with an initial denaturation of 95°C for 5 min, followed by 95°C for 3 min, 63°C for 2 min, 72°C for 1 min, repeated 13 times with a 1°C decrease in anneal temperature (63-50°C) each cycle, then 9 cycles of 95°C for 30 s, 51°C for 30 s, 72°C for 1 min, and 11 cycles of 95°C for 30 s, 51°C for 30 s, 72°C for 1 min (+10 s/cycle), with a final 5 min extension at 72°C. PCR products were visualized on 2% agarose gels by electrophoresis.

1.3.2 Sequencing and SNP Assay Development

If a locus showed a single band on the agarose gel for most individuals, PCR products from all 24 individuals were then sequenced, even for samples from which no band was visible, in an effort to preserve sample sizes. An EXO-SAP clean-up was performed prior to sequencing; 5μL of PCR product, 0.15μL of Exonuclease I (20U/μL), 1μL of shrimp alkaline phosphatase (1U/μL), 0.5μL of 10x buffer and 3.36L of deionized water were incubated at 37/μC for 60 min and then 80°C for 20 min with a final cool down to 4°C. Cycle sequencing reactions employed the BigDye Terminator sequencing

kit (v. 3.1; Applied Biosystems, Inc.) with standard conditions. Sequencing reaction products were then purified using 6% Sephadex columns and sequenced on an ABI 3730 DNA Analyzer using standard conditions. Sequences were assembled into contigs (24 individuals, forward and reverse sequences) and aligned with Sequencher 4.6 (Gene Codes Corporation) using the Dirty Data algorithm with a Minimum Match Percentage of 85% and a Minimum Overlap of 20 bp. Potential polymorphisms were visually verified on the chromatograms.

Only sites for which both homozygote genotypes and the heterozygote genotype were observed were chosen for assay development, so as to minimize identification of sequencing artifacts as polymorphisms and to ensure that the resulting SNP assays would have suitable minor allele frequencies for our intended applications. If all observed variable sites were heterozygous, we assumed that this was likely a duplicated gene and it was excluded from further analyses. In consensus sequences with multiple candidate SNPs, the site with the highest minor allele frequency (MAF) in the sequences from the Feather River populations was selected. The location of the SNP in either exonic or intronic sequences was evaluated (see Table 1.4) but was not used as a criteria for selecting the target variation. The contig sequence information was then sent to ABI for design of $5'$ exonuclease (TaqMan) assays. Taqman assays use two sequence-specific unlabeled primers and two allele-specific fluorescently-labeled probes to directly distinguish nucleotide variants (SNPs) in the target genomic DNA sample. These assays can be interrogated on a single locus, real time PCR instrument (e.g. ABI 7300 Real Time PCR System) or on a multiplex platform (e.g. Fluidigm BioMark/EP1

nanofluidic arrays).

Each Taqman assay was then evaluated on population samples of salmon from the three largest basins on the West Coast of the United States, in order to validate assay performance, refine allele frequency estimates and to evaluate the expected power of the markers for various applications. The five populations/stocks that were included are Feather River Spring-run, Butte Creek Spring-run, and Mokelumne River/Battle Creek Fall-run from the California Central Valley, Klamath and Trinity River Fall-run from northern California, and the Kalama and Cowlitz Rivers Spring-run stocks from the Columbia River basin in Washington. The 337 individuals from these five populations/stocks were genotyped with all designed assays on Fluidigm 96.96 Dynamic Arrays using the Fluidigm EP1 instrumentation and according to the manufacturers protocols. The Fluidigm system uses nano-fluidic circuitry to simultaneously genotype up to 96 samples with 96 loci in tiny reaction chambers embedded on the arrays (see Seeb *et al.* 2009 for a full description of the Fluidigm system methodology). Genotypes were called and the data compiled using the Fluidigm SNP Genotyping Analysis software. Each assay was assessed for plot quality and expected clustering patterns. The MAF and expected (H_E, unbiased) and observed (H_O) heterozygosity were calculated for each population. The software package GENETIX (Belkhir *et al.* 1996-2004) was used to estimate global F_{ST} (theta) with the estimator of Weir and Cockerham (1984), as an indicator of the power of the locus for genetic stock identification and related applications. Deviations from Hardy-Weinberg equilibrium proportions were evaluated with GENEPOP 4.0 (Rousset 2008).

1.4 Results

Of the 480 EST fragments targeted in the initial round of discovery, 244 yielded a single band when PCR products were electrophoresed in agarose and were further evaluated by sequencing; loci with multiple, weakly visible, or no PCR products were not considered further. Of the 244 loci that yielded PCR products, we successfully acquired sequence data for 225 EST fragments, with an average of 32 (of a maximum possible of 48) sequences per locus, when considering both forward and reverse strands (Table 1.1). The total length of the consensus sequences generated was 131.3kb, with a mean consensus length of 554 bp per gene. Eighty-seven loci (38.7%) yielded fragments substantially larger than the target fragment (for 12 of these loci forward and reverse sequences did not overlap), indicating the presence of one or more introns. Of the 225 EST loci for which sequence data were obtained, 177 contained some variation. In total, 661 variable sites were observed (including substitutions and insertions/deletions) and, of these, 611 were observed nucleotide substitutions that are potential SNPs. Only two nucleotides were observed at all but two of the 611 sites with substitutions present, with three bases observed at the other two sites. Fifty insertion/deletion polymorphisms were also identified, as were fifteen suspected microsatellites, but these were not considered targets for assay development at this time.

The mean density of observed mutations in the ∼131 kb of consensus EST sequence was 0.0046, or about one substitution every 215 bp. The mean length of fragments composing the consensus sequence was also weighted by the number of in-

Table 1.1: Summary of EST sequencing effort to identify genetic variation in populations of Chinook salmon (*O. tshawytscha*) from the west coast of North America. The weighted estimates account for unobserved variation in consensus sequence derived from less than 24 individuals.

	Total	**Mean per locus [range]**
EST loci successfully sequenced	225	
Base-pairs sequenced (all fragments)	3,024,916	12763.36 [382-45720]
Length of consensus sequence (bp)	131,287	553.95 [99-1566]
Weighted consensus (bp)	112,115	498.29 [72-1524]
Number of observed substitutions	611	2.72 [0 - 17]
Number of SNPs (all three genotypes observed)	228	1.01 [0 - 7]
Loci with no variable sites	48	
Insertions/deletions (indels)	50	
Transitions (A-G or C-T)	319	
Transversions (A-C or G-C or A-T or G-T)	290	
Sites with 3 nucleotides observed	2	
Possible duplicated genes	11	
Total number of substitutions + indels	661	
Density of substitutions in consensus sequence	0.0047	
Density of substitutions in weighted consensus sequence	0.0054	

dividuals for which each nucleotide was sequenced, so as to correct estimates of SNP density for undiscovered variation in the unsequenced individuals. This weighted consensus sequence length yielded a density estimate of 0.0054 or about one SNP every 183 bp. When only candidate SNPs (all three genotypes observed in sequences) are considered, density in the consensus sequence was 0.0017 (or about one SNP per 576 bp), whereas the weighted density was 0.0020 (or about one SNP per 492 bp).

Only nucleotide sites where all three genotypes were observed in the sequence data were considered as candidates for assay development. There were 228 of these putative SNPs present in the sequence data, with from one to seven present per gene, and sites in 112 genes also met the criteria for TaqMan assay design (SNP more than 40bp from either end of the sequence, with no additional variation or ambiguous sites within two bp of the target SNP). Fifteen of the original assays failed to produce reliable genotype data in the validation populations, which was defined as: no signal (all plots at the origin), a single cloud of plots with no distinct clusters, more than three clusters, or no heterozygote but both homozygote clusters within a population. For ten of the assays that failed, there were other variable sites in the genes in which they were located that met both the ascertainment and the assay design criteria. However, only five of these redesigned assays produced reliable genotypes. In addition, one of the assays that initially failed, produced reliable results with a manual assay redesign, for a grand total of 103 validated assays from the EST sequencing effort. In the small, secondary discovery effort, a total of 14 polymorphic sites, in 11 candidate genes, met the ascertainment and assay design criteria, and all yielded reliable genotype data. Multiple SNPs were

designed for the Aldolase and NAML genes. For the final 113 gene regions that contain the 117 validated assays, consensus sequences that indicate all of the observed nucleotide variation from the ascertainment sample were compiled and submitted to Genbank dbGSS (Accession Nos. HR308668-HR308783), while targeted SNP loci were uploaded to the NCBI dbSNP database (Accession Nos. 275518685-275518802; Table 1.2).

Table 1.2: Description of the 117 SNP assays developed in this project with the target polymorphism, primer and probe sequences, length of the consensus sequence in base pairs (bp), and GenBank (dbGSS) and NCBI (dbSNP) accession numbers indicated.

Assay name	Targets	Primers (5′-3′)	Probes (5′-3′)	bp	dbGSS	dbSNP
Ots_94857-232	T/C	F: GGCACTCTCCCTGGCTAGA R: CCCCATCACTTCTCTGGCTTTAAAT	VIC: CAGGATAATAACAAACAAG FAM: CAGGATAATAACGAACAAG	687	HR308668	275518685
Ots_94903-99	G/T	F: CCGTCTGAGTAGGAGGATCAATACA R: TTTGGATCCAGCTCTCCGTATAGA	VIC: CAAACCAGCAAACAT FAM: ACAAACCAGAAAACAT	314	HR308669	275518686
Ots_95442b-204	T/A	F: GTCTCTCTCTCTTTGCATCATTACACT R: GGACTCTTGAGCTGTCTGGCTATAT	VIC: TGGTTCCCCAAATTT FAM: TGATGGTTCCCCTAATTT	256	HR308670	275518687
Ots_96222-525	C/T	F: GCTCTTGCCCATCTGTAGGAT R: GGCGCAACATATGTATTAAGCAACT	VIC: TGTAGCTAATTTTAAGTTCTC FAM: AGCTAATTTTAAATTCTC	651	HR308671	275518688
Ots_96500-180	G/T	F: GATCATGTCAGATAGGATGCTGAAAGT R: CAGGTCTGGTCTACATCGAACAC	VIC: AAAACAAATCATTTTTCG FAM: AAAAACAAATAATTTTTCG	313	HR308672	275518689
Ots_96899-357	T/A	F: TCTCCTGAACTAATTTAGACCTCTGAATGT R: CCTCATATTGCTTTCATCTGAAGAGAGA	VIC: CTGAATGTTTTTTTTTAATCTTT FAM: CTGAATGTTTTTTTTTATCTTT	577	HR308673	275518690
Ots_97077-179	G/T	F: CCTGAACAAATACTTAACGCTCCAGTT R: GTAATAATACTTCACACCATTGCCACTTC	VIC: TCACAAATGTATCCTAAAGC FAM: CACAAATGTATACTAAAGC	288	HR308674	275518691
Ots_97660-56	T/A	F: TTCCCTAATCTGACGTACTACCAACT R: CGCCACTGACGTTCATTCCA	VIC: ACGAGACAGATATTC FAM: ACGAGACTGATATTC	455	HR308675	275518692
Ots_98409-850	C/T	F: CTGCGTTTCTGGAATGTTTTCAGT R: CAAACCTGTTACTGGCCAAATGAAA	VIC: TTGTTCACGAACCTTG FAM: TTGTTCACAAACCTTG	1072	HR308676	275518693
Ots_98683-796	T/A	F: GCAATGGCATGACAATGGAAGTC R: CACTGGCACTGGTGGAGATTA	VIC: CTCAGCCCCTATATTACAA FAM: CTCAGCCCCTATTTTACAA	895	HR308677	275518694
Ots_99550-204	C/T	F: TGACAGATTTCACCTTTAACTAGCTAAGC R: GCAACCTCTTTCACACTTCAGTAAC	VIC: AAGGCTTTGGTTGTTTG FAM: AAGGCTTTGATTGTTTG	356	HR308678	275518695
Ots_100884-287	T/C	F: CGGAAGACCAGATTCTCCAAGAGTA R: CGACCAAGTAGCGGCACTT	VIC: ATAGAACTACAATTCACATATAT FAM: AACTACAATTCGCATATAT	470	HR308679	275518696
Ots_101119-381	T/C	F: TTTTCTAGGACAGGTTGCTTGCA R: CCAGGTTTCTTTAGCCTACTTATTCTTTACA	VIC: TGCCACATGATAATTGA FAM: CCACATGGTAATTGA	1122	HR308680	275518697
Ots_101554-407	C/G	F: TGAAAGATATCAATTGTAGTAGTGGTGGTG R: ACACGCCAGTCCACAAGT	VIC: ATGGAGGATTGTGGTTGT FAM: ATGGAGGATTCTGGTTGT	417	HR308681	275518698
Ots_101704-143	T/G	F: ACTTCTTGAGCCAATCGGATGATG R: CCAGAGATAAACTAGTGGAGGAGATCA	VIC: CTTAGACGTCAGAGGTC FAM: CTTAGACGTCCGAGGTC	580	HR308682	275518699
Ots_101770-82	C/A	F: GCGACTTGACAACGAGGAGAA R: CCCTCTTCATAACGTTACCAAACAC	VIC: ACTTCCCGGAGCTGC FAM: ACTTCCCTGAGCTGC	783	HR308683	275518700
Ots_102195-157	T/C	F: TGGTCAGCGGTGTCTTTCAC R: CCCCGCATCTGTCAAATGGAT	VIC: TGCTGATTCAAGAAGAAGTA FAM: CTGATTCAAGGAGAAGTA	543	HR308684	275518701
Ots_102213-210	A/G	F: CATTCCATGACAATGATTGAAATCTAAAAACAC R: GAGTATCTCAATTGCAACACTATGGTATGT	VIC: CTGTATACAGTAAGAGTATTAAT FAM: ACAGTAAGAGCATTAAT	1074	HR308685	275518702
Ots_102414-395	A/G	F: GCCTACTGATAAATGTATGACAGTAATGGA R: CAATAACAAACAAGCTAGGAACAAAAGTGT	VIC: CACATAGTGTAGCTTTACTAC FAM: CACATAGTGTAGCTCTACTAC	1030	HR308686	275518703
Ots_102420-494	T/G	F: TGCCAACCTGGCCAGTTAC R: GCTTCCCTGCTTCCATGGT	VIC: CATGTGAACAACAAGCG FAM: CATGTGAACACCAAGCG	739	HR308687	275518704

Continued on next page

Table 1.2 – continued from previous page

Assay name	Targets	Primers (5′-3′)	Probes (5′-3′)	bp	dbGSS	dbSNP
Ots_102457-132	A/G	F: CCAGCAGAGACTGGGTTCAC R: TTCCCTACCGGCGAAACC	VIC: CAATTGTGCGTTGCCCCA FAM: ATTGTGCGTCGCCCCA	734	HR308688	275518705
Ots_102801-308	C/A	F: TGGGACAGAGGTGGGAATTGA R: CCCAAAGATGCTTAACTGAAGATGTG	VIC: AGGGACAGTTTCGCAGACG FAM: AAGGGACAGTTTCTCAGACG	670	HR308689	275518706
Ots_102867-609	A/G	F: CTCTGCCATTCATTTGGGCTTTG R: GTCTAAAGTGGTCCCCTTGGAT	VIC: ACAGAGAGAAGTCCCAGGTG FAM: AGAGAGAAGCCCCAGGTG	796	HR308690	275518707
Ots_103041-52	G/A	F: ACCACCCACCTCCTCAGA R: AGACAGAGAAAGTCGGGACACT	VIC: CATCCTGCTGGACCC FAM: CATCCTGTTGGACCC	691	HR308691	275518708
Ots_103122-180	T/C	F: CAAACGCGCACTCACACA R: TCACAATGGTACGATTTTACGACTCAA	VIC: CATCAACACAATCTGC FAM: CATCAACACGATCTGC	424	HR308692	275518709
Ots_104048-194	T/C	F: CAGCTGCTGCAGTCAATGAG R: GCTCCTTACCAGTGTTGTCAGT	VIC: CTGCCACCACCACCAC FAM: TGCCACCGCCACCAC	383	HR308693	275518710
Ots_104063-132	C/T	F: GCGTTACTGGTGTTATAAACGTTAGC R: GTTTATTTAATTATGAAGGACGATGTTGAAGTCA	VIC: CTTTCGTCCTTAGCACATAG FAM: CTTTCGTCCTTAACACATAG	874	HR308694	275518711
Ots_104216-70	G/T	F: AGTAGGAGTCGCAGCTATGGAA R: CCTGTGGTCGGAATGATGGT	VIC: TCTGCCCCGGCTCT FAM: TCTGCCACGGCTCT	546	HR308695	275518712
Ots_104415-88	C/T	F: CCTGAGCATCCCAGTTGAACT R: TGTTTTCAATACACTGCAATTTAGTTTTGGT	VIC: TCCTGAAAAACGACATCC FAM: CTGAAAAACAACATCC	434	HR308696	275518713
Ots_104569-86	T/G	F: CCTGCATGTTGTTCACGTTGTC R: CGGCCGGAGGGATCAC	VIC: TGGTCGCAGATGCC FAM: TGGTCGCCGATGCC	582	HR308697	275518714
Ots_105105-613	C/G	F: AGTACAAGTGCAGAGAATGACATCATG R: GGTGTTTTATTTTCCCATATATCTTTTAACTTTAAGCT	VIC: CCGAGCTTGAGTTAGGA FAM: CCGAGCTTGACTTAGGA	801	HR308698	275518715
Ots_105132-200	G/T	F: CGATGTACTGAGGGCAGTGT R: GAGTGGAGTTCCTTAATAATCATTGACCTT	VIC: CAAGAGTGGCATAAAA FAM: CAAGAGTGGAATAAAA	458	HR308699	275518716
Ots_105385-421	A/G	F: GACTGTCTTGGAACCGTTGCTA R: TCCCGGAACACACCAATGTC	VIC: CCTCCTGGGTATATCG FAM: CTCCTGGGCATATCG	676	HR308700	275518717
Ots_105401-325	T/G	F: GAACTGAGCGGCTGCTG R: CGCCTCCTGGTGTCTATCCT	VIC: CAAGATGAGACAGTTACAG FAM: CAAGATGAGACCGTTACAG	500	HR308701	275518718
Ots_105407-117	T/A	F: TGTGTACATCCGCGTAAATATTGAAGATAA R: CTGTGAGCTGCTGCAAACC	VIC: CAGGTTAGGAATGGTTG FAM: CAGGTTAGGATTGGTTG	476	HR308702	275518719
Ots_105897-124	T/C	F: CCTCAGTGTTATTTGTATATGATCATTTTGAAACATTT R: AGCCCAATGCATCTACTGAATTCAT	VIC: AACCAATAATATGAAACTGTG FAM: CCAATAATATGGAACTGTG	387	HR308703	275518720
Ots_106172-425	C/T	F: GCAGTCAGTGCGTTGATACG R: GGTGTAGACGTGAACAATGAGGATA	VIC: CTGATAACTACTTGGCGTGTGT FAM: TGATAACTACTTGGCATGTGT	466	HR308704	275518721
Ots_106313-729	G/A	F: TTGTTCAATGGGCATTAATGCATGTT R: TGCTTATGTGCAGATACTTGAGACAAA	VIC: AAGAGTCCAGCGTTACTT FAM: AAGAGTCCAGTGTTACTT	794	HR308705	275518722
Ots_106419b-618	T/G	F: CAAGGGCACATTGGCAGATTTT R: ACCGGACCAAAGCACACA	VIC: CAATGATTAATGATTAATCCTTC FAM: TGATTAATGATTCATCCTTC	806	HR308706	275518723
Ots_106499-70	C/G	F: ACTCTATCATCGGCAGGACCAT R: ACCGTAAGTGTGGTTGTGTTCATTA	VIC: CTCATTTTTCAGAATTGTATTC FAM: CTCATTTTTCAGAATTCTATTC	516	HR308707	275518724
Ots_106747-239	C/A	F: ATCGAGGATGCCTCAAAGACATC R: GTTAGACCCACCACCAGTCATC	VIC: CCCGCGGTGAGTAT FAM: CCCGCTGTGAGTAT	820	HR308708	275518725
Ots_107074-284	A/T	F: CCCACTTCCAGAGCCTGAA R: TTTTCCATGGCTGTGTGTACTGT	VIC: ACCGTAGCTGCACCTG FAM: CGTAGCAGCACCTG	399	HR308709	275518726

Continued on next page

Table 1.2 – continued from previous page

Assay name	Targets	Primers (5′-3′)	Probes (5′-3′)	bp	dbGSS	dbSNP
Ots_107220-70	C/G	F: GGGAACGCACAACATACTACAGTA R: AGAGATGAAACACTTCTATATCTATGGTGTGT	VIC: AGGTCATGAACTATAGATTCCATGTA FAM: TCATGAACTATAGATTCCATCTA	409	HR308710	275518727
Ots_107285-93	T/A	F: GCCCTTGTGACAATGCACTGTTATA R: AACATACACCAATACTTAGGTCTAGACAGT	VIC: AAGTAACGTATCAAATGGC FAM: AAAGTAACGTATCATATGGC	726	HR308711	275518728
Ots_107607-315	C/A	F: GTGATGAGAGGTTTCCGGAAAATCT R: GTGTTCTGGATTCCATTGTGCAAA	VIC: ATGGGAGACAGATAACT FAM: ATGGGAGACATATAACT	516	HR308712	275518729
Ots_107806-821	T/A	F: CTCCCTTGCTTTTGGTCATTGG R: TGCAGTGCTGAATTAGAGATTAATTTTTGTG	VIC: CAAAGAAAATCAAAATTT FAM: CAAAGAAAATCTAAATTT	997	HR308713	275518730
Ots_108007-208	A/T	F: CAGGCTTGTGTTAAGTAGGGAGAAA R: CATTGGACAAGACCGGGTAGTC	VIC: CAGTTTCACTTAATTTTAAAATG FAM: TTTCACTTAATTTAAAAATG	434	HR308714	275518731
Ots_108390-329	G/C	F: GAGGTTTGTTACTGTCACCCATAGA R: CCTGCTGTAGCAAACTGTCTCAAA	VIC: CTACTTATGTAGCATTTTAA FAM: CTACTTATGTAGGATTTTAA	1566	HR308715	275518732
Ots_108735-302	C/T	F: CCTTTTTCTTATTAGTTTTACTTCCCCAGAGA R: CAATTCCATTCTTGATTCTGTTTAACGGT	VIC: AAACAAACAACGCCTCATG FAM: AACAAACAACACCTCATG	324	HR308716	275518733
Ots_108820-336	G/A	F: TGAAATAAATTGTTCTGTTGATATGTGAATTTTGGA R: CAACGACACACCAACAACGT	VIC: ATTGCCCATCTCAGAATA FAM: AATTGCCCATCTTAGAATA	396	HR308717	275518734
Ots_109243-285	C/A	F: CGCTGCTACACTGTTTACTGTGATA R: GCACCTCTTAAATTGTAAGTAAAATGTTAACGA	VIC: TGGCAAGATGACCTTT FAM: AGTGGCAAGATTACCTTT	514	HR308718	275518735
Ots_109525-816	C/T	F: GCCAGATAGTAGCGTACATCATGAG R: CTCCCCATGTCCCTGAGTCT	VIC: CATGAGGCGTTCGGC FAM: ATGAGGCATTCGGC	1061	HR308719	275518736
Ots_109693-392	T/G	F: TCTCCCTCATTCCCATGTCATATCA R: GGGAACGTATCAGGTGAGTGT	VIC: TCCGTTAGTTCCTCCTGG FAM: TCCGTTAGTTCCTCCTGG	473	HR308720	275518737
Ots_110064-383	C/T	F: AACAAAGAATGTTAAACACCAAACAGGAA R: GTGCAAGGGACCTAGCTAATCC	VIC: CTACGTAATGAACGTTAGCT FAM: ACGTAATGAACATTAGCT	801	HR308721	275518738
Ots_110201-363	A/T	F: GTTTGGCTATTGAAATTATACATTAAAACATGTAGCT R: CCATGGCATCCTGTAAAGAACAACA	VIC: TGGATGCCAGTTTTAAAA FAM: TGGATGCCAGTTTAAAAA	558	HR308722	275518739
Ots_110381-164	A/G	F: CTCTTGTTTGCTATGGGAGATGTAGT R: CCGTATCCTAAACCCTTCACTGTT	VIC: ATTTGCGTCTTCTCCC FAM: TTGCGTCCTCTCCC	661	HR308723	275518740
Ots_110495-380	G/C	F: GCCTAGGTATGTACGAAACTTCACA R: AGGCTTTTTCAGATGGTCGTATGA	VIC: ATGGCCCCTGTCTATG FAM: ATGGCCCCTGTGTATG	825	HR308724	275518741
Ots_110551-64	C/A	F: GAGTGGTCAAGGTTTCAGTTTCTG R: GAAATGGACAGACACAAGGTCAAAC	VIC: ACGCTCGGAACATT FAM: ACGCTCTGAACATT	685	HR308725	275518742
Ots_110689-218	T/G	F: GTATAAACTAGAGTCCAGTGTTATGTTAATGTCTT R: CATGGCAGACAACAGTAGAGAATATGA	VIC: CACCAATCAATTAATTATT FAM: ACCAATCAATTCATTATT	397	HR308726	275518743
Ots_111084-96	A/G	F: AAAAGTTAATACTGGGTACAACCTCTGAAAA R: GGGACAGTAGTTGGGTCATTCAAAT	VIC: TTGAACCATTCTACTATTGGT FAM: AACCATTCTACCATTGGT	710	HR308727	275518744
Ots_111084b-619	C/A	F: TTGTGGAATTACACCTTCAGAGTTCAAT R: GCCTGTTTGGCTTTCTTAAACTGAT	VIC: CCATGGAAACGGACAAT FAM: TCCATGGAAACTGACAAT	710	HR308727	275518745
Ots_111312-435	C/T	F: CCATGCGCCTTTGAGGAAATTAA R: TTCATGGTCTTTATCCCCCCTACA	VIC: ACTCATACCTAGAGGTCACAT FAM: CTCATACCTAGAGATCACAT	475	HR308728	275518746
Ots_111666-408	C/T	F: GAGAATCTGGGATTGGTACATCCAT R: AAGCTCATGATACATGTATGAGTTATATTCTTCAAG	VIC: ATAGTATCACTAGTTAAAAAT FAM: ATAGTATCACTAATTAAAAAT	664	HR308729	275518747
Ots_111681-657	G/T	F: CTGAGCTTTTTCAACTTACTTGTTGGA R: GGCGCAGCAGCAACTG	VIC: TAGCGCAAACCCCGAACC FAM: CGCAAACACCGAACC	702	HR308730	275518748

Continued on next page

Table 1.2 – continued from previous page

Assay name	Targets	Primers (5′-3′)	Probes (5′-3′)	bp	dbGSS	dbSNP
Ots_112208-722	C/A	F: CTGCATGAACGTTAACTCAAATAAAAGGT R: AATGAGTTCTACTGACATTGTATACTAGAATAAGTATCA	VIC: TGTGAGGGCGGTCTT FAM: ATGTGAGGTCGGTCTT	944	HR308731	275518749
Ots_112301-43	T/C	F: GCATGGCTGCCCTAGAACA R: TCAGAACATTTCCTTCAGCTTCGT	VIC: CGTCGCATTCAGC FAM: CGTCGCGTTCAGC	397	HR308732	275518750
Ots_112419-131	A/T	F: GTGGGTAATCGATGCCAAAGAGAT R: TGGCAGTGTTTTCAACTAGCTTTG	VIC: AAGCGACTTGATTATC FAM: AGCGACATGATTATC	391	HR308733	275518751
Ots_112820-284	C/T	F: CATAGATGTTTATATGAAAAACCTCCCACTGT R: GCATCCAAAAAGACGTGTGTGTTT	VIC: ACTCACACTCGAGTGACT FAM: ACTCACACTCAAGTGACT	394	HR308734	275518752
Ots_112876-371	C/A	F: GCCTACAGCAAATTCAGCTACACAT R: TGGACCTTCAATCATCACAGCTT	VIC: CATCACAACGATGTGTG FAM: CACATCACAACTATGTGTG	1118	HR308735	275518753
Ots_113242-216	C/T	F: GAGGCCTAATGTCTCTTGTGACT R: GACATCTTCAACAAGTGTTCATTCACC	VIC: ATTACCAACGGAGAACC FAM: TTACCAACAGAGAACC	364	HR308736	275518754
Ots_113457-40	C/T	F: CCCAAGTGGTGAGTGTCAGT R: ACTACAACAGGTGTTGATAATAGAATCATTCTC	VIC: ATATGGATTGGAGAATAG FAM: CATATGGATTAGAGAATAG	555	HR308737	275518755
Ots_115987-325	T/G	F: GGAGGTGTAGTGAAATGGGAAGAT R: GCATTCAGTGAACCAGTAGTGCTAT	VIC: ATGCATAAAAGGTAATTGTG FAM: ATGCATAAAAGGTCATTGTG	631	HR308738	275518756
Ots_117043-255	C/A	F: TCTCAATCTTGACACAAACTGGCT R: TCGATCTGTTCTCGTGGTGTTTC	VIC: ACGTCAGAATGGATTCT FAM: AACGTCAGAATGTATTCT	628	HR308739	275518757
Ots_117138-545	T/G	F: GGTGGTGGCAGTATTTGTTATCATG R: GCAGTTACAGTCTGAGCTTGACAA	VIC: CAGTCAGACAGATACC FAM: CAGTCAGACCGATACC	661	HR308740	275518758
Ots_117242-136	A/G	F: GTGACAGGAGACAGAAAGAGACATT R: TGGTCCTCCCTGTCTCTATCTACTA	VIC: CAGCACATAACTTGACCTC FAM: AGCACATAACCTGACCTC	475	HR308741	275518759
Ots_117259-271	T/G	F: ACACCCACTTCAACCTCCATAAC R: GCCTCAGAGCTTAGCTTGGA	VIC: CTCTCCTGATCACTCTGT FAM: CTCTCCTGATCCCTCTGT	414	HR308742	275518760
Ots_117370-471	T/G	F: TGCAAACACAGAGGAAAGGGATTT R: GTTGGCTCCTTCAATTCAATTTGGA	VIC: ACGGAACAAATAAGACATTT FAM: CGGAACAAATAAGCCATTT	621	HR308743	275518761
Ots_117432-409	A/G	F: TCATCAAAACATGCCTCTTCTGTGT R: TGTTGAACCTGTCACTCTGTCTTC	VIC: TTTAGACTTTGCTCTATAACAG FAM: ACTTTGCTCCATAACAG	443	HR308744	275518762
Ots_118175-479	C/T	F: TGCGCGTCTCATTCAACCAT R: ACCTTACGTCCTAGGTAGGAAACA	VIC: AGAATGAAGTGAAAAGAA FAM: AGAATGAAGTAAAAAGAA	496	HR308745	275518763
Ots_118205-61	T/C	F: CCATACAGCCAGTCCAGGTG R: ACTGGACAGGGCTGGGT	VIC: TAGTAGCCCCTACACCTC FAM: TAGCCCCTGCACCTC	485	HR308746	275518764
Ots_118938-325	C/T	F: ATTTTCAAACAGGCATTTATCATTGGTGAA R: GGTCTGTCCCTCATTCTTTGCA	VIC: AGAGATGCAAAGTGGAGTT FAM: AGAGATGCAAAATGGAGTT	606	HR308747	275518765
Ots_120950-417	T/A	F: CAGACAGGTCACCATCACACT R: TGGTGAAGCTGTAGGAGAAGGA	VIC: CTGGACCAGAACTCTGA FAM: CTGGACCAGATCTCTGA	806	HR308748	275518766
Ots_122414-56	C/T	F: GCACCGTATCAACGAGCTCAT R: TGCATGGATTTCCTTTGTGTTGTTG	VIC: TGTATGACCTCTGACCTGT FAM: TGTATGACCTCTAACCTGT	423	HR308749	275518767
Ots_123048-521	A/C	F: CTCAACAGTGCACCTCCCTTAATT R: CCAAACACACCCTTCCATAATCTCT	VIC: TCACATCCAACTCAGTACT FAM: CATCCAACGCAGTACT	808	HR308750	275518768
Ots_123205-61	G/A	F: GCGCAAGAGGCGAAGATG R: GGTCACAGCCATGGTGTTG	VIC: CAAGTAGAATGCCTCCCCATA FAM: CAAGTAGAATGCCTTCCCATA	329	HR308751	275518769
Ots_123921-111	A/G	F: TCGCTAGGCAGAAATATAGGGTTCT R: GAGCATGGCGCTTGCA	VIC: TGCTAAATGGCATATATTAT FAM: CTAAATGGCACATATTAT	979	HR308752	275518770

Continued on next page

Table 1.2 – continued from previous page

Assay name	Targets	Primers (5′-3′)	Probes (5′-3′)	bp	dbGSS	dbSNP
Ots_124774-477	T/C	F: AGTTGTTCTTTTTATATTGTGTTTTTATTCCATTCCA R: GCCAAATAAAAACAAAGCATGAACACA	VIC: CCACCGCCATCTGATA FAM: CACCGCCGTCTGATA	710	HR308753	275518771
Ots_126619-400	T/C	F: GGATGGTTGTCATTTCTCTGCAAA R: CCGGGATACAATAATAATATTTGGTTAAGAGTTTTTT	VIC: AGAAAGTTCTAGAAATAATT FAM: AAAGTTCTAGGAATAATT	786	HR308754	275518772
Ots_127236-62	T/A	F: TGGAGAACTTGCACTGAATGTGAAA R: GCTGTTGGACCTTGACTTTAACAAATT	VIC: TCTCTTATCTGAGTTCTGC FAM: CTCTTATCTGTGTTCTGC	668	HR308755	275518773
Ots_127760-569	C/T	F: CTGCTGGCGCAGACATG R: CGTTATAGAGGATAGTTTGGAGGAAGGA	VIC: CCGGTTTACCGATTTG FAM: CGGTTTACCAATTTG	688	HR308756	275518774
Ots_128302-57	C/T	F: GGTTGCAGGGCAGAACTGT R: ACCCATCCAATAACCCATTTTCCTT	VIC: CCTGCAATACGACCAAC FAM: CTGCAATACAACCAAC	833	HR308757	275518775
Ots_128495b-45	A/T	F: GGACGCTGACCAGTACATAGG R: TTTCTCCCCAAAGTATTTAAGCCTACAC	VIC: AGTGTCAACCTTCTCTC FAM: TGTCAACCATCTCTC	153	HR308758	275518776
Ots_128693-461	C/T	F: TCAATGTTCATCAATGCACTTCCTGTA R: GCCTGCAGGAGAAGGTAGAGTTA	VIC: CACTCAGCTGGTACCCA FAM: ACTCAGCTGATACCCA	886	HR308759	275518777
Ots_128757-61	A/G	F: CGTGTCCGGCTTCTTTTATTTCATT R: GATGGGTATGTTAATCATATTACCAGCGTAA	VIC: TTGTGCATTTTCCCC FAM: TGTGCATTTCCCCC	377	HR308760	275518778
Ots_129144-472	C/A	F: CTGTTAGTGCAGAAGACGTAGCT R: GCAGAGCTATTGAGCCAAGTTACAA	VIC: TGGGTCTCGAGCCTGTA FAM: TGGGTCTCGATCCTGTA	635	HR308761	275518779
Ots_129170-683	C/A	F: AACCCTATGGGAACTCGTAGAACT R: GCTAGGAGTTCTCAAAAGGGTTCT	VIC: ATTAGAACTCGTAGAACTAT FAM: ATATTAGAACTCGTATAACTAT	795	HR308762	275518780
Ots_129303b-54	C/T	F: ACCTGGAGAGAAAGTTCAAGAGAGA R: GAGCTAGTAGAGGAAACAAAATAAACATTTCAAT	VIC: CCCTGGTGACCTCT FAM: CCCTGGTAACCTCT	711	HR308763	275518781
Ots_129458-451	T/C	F: TGGGACCCACATAAAGCAACTG R: GACATAAGACCCATTTAGCCCCTTTT	VIC: CATCTGGCAATGCCTT FAM: CATCTGGCAGTGCCTT	551	HR308764	275518782
Ots_129870-55	A/T	F: GCATGTAACACATTATTTGGCATATGTACT R: CAGTACACTGGAGATTTGCAATGTT	VIC: ATGCATTCACCTGTATTAT FAM: TGCATTCACCAGTATTAT	958	HR308765	275518783
Ots_130720-99	A/G	F: CGGTCATTGTAAATGTCAACGGTTT R: TGCTTGCATGTTCTTGGTGTAGTAA	VIC: CCTGTCTCATTCCC FAM: CTGTCCCATTCCC	542	HR308766	275518784
Ots_131460-584	T/C	F: CCTATTTTTGATAGGTCATAGTGAATGGGATAG R: CTGTACTCCTCCATTCCTTTTCACT	VIC: CTATCAAAGCAATACATTG FAM: CTATCAAAGCAGTACATTG	1283	HR308767	275518785
Ots_131802-393	T/C	F: TGATTGTCTCATGGCCAATTGTCA R: TGTAAATTCCACTTGGCAATCTTTGG	VIC: TGTTCGAGAATGAAGATGAGTAA FAM: TCGAGAATGAAGGTGAGTAA	489	HR308768	275518786
Ots_131906-141	A/T	F: GGCTCGAACCACCCAGTTTA R: TGCCCAACTGGTTTGCAATC	VIC: CACGGTTTACACTCCTATTA FAM: ACGGTTTACACTCCAATTA	408	HR308769	275518787
Ots_AldB1-122	C/T	F: GCCATGGAGGACTGGATGA R: GCCACCACTACTTGCTGAGAAAATA	VIC: ACCCACTTCGCCAACA FAM: ACCCACTTCACCAACA	469	HR308770	275518788
Ots_AldoB4-183	T/A	F: TTTGTGCGTAAAGTCAGGTAGTGT R: GTGCATGCCATGAGAACTTTGTTT	VIC: CTGTGTGTCTAAGACAAT FAM: CTGTGTGTCTATGACAAT	296	HR308771	275518789
Ots_CathD-141	T/C	F: CACTTGTTCTGCACACTACTTGTC R: CACACATGGATTTTGCCTGTCTAAA	VIC: TGGGAAGCAATCAA FAM: AATTGGGAAGCAGTCAA	484	HR308772	275518790
Ots_CRB-211	A/C	F: CAACGCGGGAATGGCTTTTAA R: GCCAGAGTCGCCAAAATAGTAGAAT	VIC: CTACCGTACTGAACTC FAM: CCGTACGGAACTC	1041	HR308774	275518792
Ots_EndoRB1-486	G/A	F: CCTTTGGGTCTGCTTGAGGTT R: GGAGCCAAATCCTAATGCTGAAGTA	VIC: TCCTTCTCACGCTTCT FAM: CTCCTTCTCATGCTTCT	1038	HR308775	275518793

Continued on next page

Table 1.2 – continued from previous page

Assay name	Targets	Primers (5′-3′)	Probes (5′-3′)	bp	dbGSS	dbSNP
Ots_Hsp90a	G/C	F: ACAGTATACCGGCTGCCTATTCATA R: GTCGTTTTTCATAGAAAATAGCTCACAGTT	VIC: ATTTGACTTGTCTTTTTG FAM: TTTGACTTGTGTTTTTG	373	HR308776	275518794
Ots_Myc-366	T/C	F: CCTTAGCTGCTCTTTGAAGTTGACT R: GGCTATAGAGTGTATTTACAGCATGCA	VIC: TCTCTGCTCATCTGTC FAM: CTCTGCTCGTCTGTC	409	HR308777	275518795
Ots_ALDBINT1-SNP1	T/C	F: CGCTGGGCATGGATGAGT R: GGCCAACACTGCTACTTCCT	VIC: CTACTGTTGTATTTTCTC FAM: CTGTTGTGTTTTCTC	474	HR308778	275518796
Ots_DESMIN19-SNP1	C/A	F: GGTCTGTCTGTCTGTCTATCTGTCA R: TGTGTGTCTTTGTTCATTCCTACCA	VIC: CCAGTCATGGGTCATT FAM: TCCAGTCATTGGTCATT	439	HR308779	275518797
Ots_NAML12-SNP1	A/G	F: TGCCACCTCAGTTTTAGTGTTATATCC R: AGCGCCAACCTGTCACT	VIC: AAACCATTTTCATTCTTTTG FAM: CCATTTTCACTCTTTTG	548	HR308780	275518798
Ots_NAML12-SNP2	C/A	F: GGCGGTTAGGTAGGATATGATTCC R: TCACGTAGCCTACCACAGATAAGT	VIC: TCCATAAGCGGGAAAA FAM: TTTCCATAAGCGTGAAAA	548	HR308780	275518799
Ots_BMP2-SNP1	C/T	F: ACTGCCACAGACACGAACTC R: GCCACTATCCACTCGTTCCA	VIC: CCCACTTCGCTGAAGT FAM: CCCACTTCACTGAAGT	638	HR308781	275518800
Ots_MTA-SNP1	C/T	F: GCCGAAAAATAAGCGATTAGTGATGA R: GCCCCATGGTAAACCTAATTAACCT	VIC: AATTGCCTCATTGGGTG FAM: AATTGCCTCATTAGGTG	220	HR308782	275518801
Ots_TF1-SNP1	G/T	F: CGGACAAAGAGCTACAGAAATGC R: CGTCCCTCTTCACGCATGA	VIC: CCGCCACCTTGGCT FAM: CGCCACATTGGCT	755	HR308783	275518802

All TaqMan assays were then used to genotype 337 fish from five major salmon stocks representing the Sacramento (Central Valley), Klamath and Columbia Rivers. A summary of the population genetic variability of the 117 validated assays can be found in Table 1.3. Mean MAF for all of the variable loci ranged from 0.194 in the Klamath/Trinity basin to 0.236 in the Feather River Spring-run stock. MAFs for individual loci in these five populations ranged from to 0.005 to 0.500. The proportion of polymorphic loci was nearly 90% in all populations and ranged from 90.6% in the lower Columbia stock to 86.3% in the Butte Creek Spring-run population. Thirty of the 117 loci were monomorphic in at least one population and four loci (Ots_102195-157, Ots_107220-70, Ots_117138-545, and Ots_123205-61) were not variable in any of the samples from the five populations/stocks. They are reported here, however, because either the ascertainment sequence data or additional genotype data (not shown) indicate these markers are variable in the species and may be useful in other parts of the range. Expected (unbiased) heterozygosity (H_E) for each variable locus ranged from 0.01 to 0.51 (mean = 0.33), while observed heterozygosity (H_O) ranged from 0.01 to 0.68 (mean = 0.33). Mean H_E was similar in all populations, ranging from 0.27 (Klamath) to 0.31 (Feather River and lower Columbia), and mean H_O followed the same pattern, ranging from 0.26 (Butte Creek) to 0.31 (Feather River). Overall F_{ST} for the individual loci and all five populations ranged from 0 to 0.592 and averaged 0.107 for all loci, indicating substantial differentiation in allele frequencies. Almost all loci were in Hardy Weinberg equilibrium in all populations; only two loci (Ots_127760-569 and Ots_109243-285) deviated from equilibrium in one and two populations, respectively (Table 1.3).

Table 1.3: Summary statistics for 117 SNP loci in five Chinook salmon populations. N is the number of individuals genotyped. H_E is expected (unbiased) heterozygosity and H_O is observed heterozygosity. F_{ST} is over all five populations. AF is the observed frequency of the minor allele from the Feather River stock in each population. Asterisks (*) indicate significant ($p<0.001$) deviations from Hardy-Weinberg equilibrium.

	Feather River (Spring) N=94			**Butte Creek (Spring) N=54**			**Central Valley (Fall) N=94**			**Klam/Trinity (Fall) N=48**			**L. Columbia (Spring) N=47**			
Assay	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	F_{ST}
Ots_94857-232	0.466	0.50	0.42	0.567	0.50	0.56	0.582	0.49	0.44	0.222	0.35	0.36	0.330	0.45	0.57	0.076
Ots_94903-99	0.158	0.27	0.25	0.078	0.15	0.12	0.137	0.24	0.25	0.010	0.02	0.02	0.255	0.38	0.47	0.048
Ots_95442b-204	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.030	0.06	0.07	0.150	0.26	0.21	0.124
Ots_96222-525	0.375	0.47	0.51	0.549	0.50	0.39	0.344	0.45	0.47	0.848	0.26	0.30	0.522	0.50	0.57	0.135
Ots_96500-180	0.194	0.32	0.32	0.149	0.26	0.21	0.096	0.17	0.15	0.622	0.48	0.32	0.315	0.44	0.33	0.183
Ots_96899-357	0.022	0.04	0.04	0.000	0.00	0.00	0.016	0.03	0.03	0.000	0.00	0.00	0.000	0.00	0.00	0.004
Ots_97077-179	0.289	0.41	0.44	0.104	0.19	0.08	0.190	0.31	0.34	0.239	0.37	0.35	0.178	0.30	0.22	0.021
Ots_97660-56	0.049	0.09	0.10	0.010	0.02	0.02	0.027	0.05	0.05	0.010	0.02	0.02	0.011	0.02	0.02	0.006
Ots_98409-850	0.054	0.10	0.11	0.031	0.06	0.06	0.075	0.14	0.13	0.000	0.00	0.00	0.000	0.00	0.00	0.021
Ots_98683-796	0.112	0.20	0.16	0.120	0.21	0.24	0.081	0.15	0.16	0.052	0.10	0.10	0.053	0.10	0.11	0.003
Ots_99550-204	0.350	0.46	0.48	0.202	0.33	0.37	0.287	0.41	0.38	0.245	0.37	0.40	0.163	0.28	0.28	0.020
Ots_100884-287	0.081	0.15	0.14	0.051	0.10	0.10	0.114	0.20	0.18	0.318	0.44	0.36	0.141	0.25	0.24	0.067
Ots_101119-381	0.429	0.49	0.53	0.573	0.49	0.40	0.544	0.50	0.44	0.032	0.06	0.06	0.064	0.12	0.13	0.232
Ots_101554-407	0.174	0.29	0.26	0.100	0.18	0.20	0.183	0.30	0.30	0.010	0.02	0.02	0.223	0.35	0.36	0.037
Ots_101704-143	0.367	0.47	0.46	0.696	0.43	0.39	0.456	0.50	0.43	0.659	0.45	0.41	0.426	0.49	0.43	0.069
Ots_101770-82	0.255	0.38	0.36	0.010	0.02	0.02	0.242	0.37	0.31	0.010	0.02	0.02	0.000	0.00	0.00	0.140
Ots_102195-157	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000
Ots_102213-210	0.228	0.35	0.39	0.138	0.24	0.28	0.196	0.32	0.37	0.240	0.37	0.40	0.021	0.04	0.04	0.038
Ots_102414-395	0.440	0.50	0.48	0.604	0.48	0.46	0.382	0.47	0.46	0.628	0.47	0.66	0.479	0.50	0.45	0.037
Ots_102420-494	0.333	0.45	0.51	0.220	0.35	0.36	0.277	0.40	0.45	0.010	0.02	0.02	0.422	0.49	0.49	0.088
Ots_102457-132	0.170	0.28	0.32	0.112	0.20	0.18	0.242	0.37	0.42	0.708	0.42	0.42	0.522	0.50	0.30	0.234
Ots_102801-308	0.261	0.39	0.41	0.344	0.46	0.44	0.258	0.39	0.34	0.309	0.43	0.57	0.163	0.28	0.28	0.010

Continued on next page

Table 1.3 – continued from previous page

Assay	Feather River (Spring) N=94			Butte Creek (Spring) N=54			Central Valley (Fall) N=94			Klam/Trinity (Fall) N=48			L. Columbia (Spring) N=47			
	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	F_{ST}
Ots_102867-609	0.462	0.50	0.48	0.373	0.47	0.51	0.376	0.47	0.49	0.948	0.10	0.10	0.678	0.44	0.47	0.193
Ots_103041-52	0.494	0.50	0.58	0.337	0.45	0.31	0.350	0.46	0.38	0.351	0.46	0.53	0.435	0.50	0.52	0.015
Ots_103122-180	0.005	0.01	0.01	0.020	0.04	0.04	0.000	0.00	0.00	0.174	0.29	0.30	0.522	0.50	0.42	0.404
Ots_104048-194	0.191	0.31	0.36	0.225	0.35	0.45	0.242	0.37	0.48	0.083	0.15	0.17	0.109	0.20	0.22	0.024
Ots_104063-132	0.054	0.10	0.11	0.179	0.30	0.26	0.038	0.07	0.05	0.436	0.50	0.57	0.128	0.23	0.21	0.188
Ots_104216-70	0.000	0.00	0.00	0.000	0.00	0.00	0.011	0.02	0.02	0.083	0.15	0.17	0.032	0.06	0.06	0.046
Ots_104415-88	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.010	0.02	0.02	0.096	0.18	0.19	0.086
Ots_104569-86	0.500	0.50	0.45	0.656	0.46	0.27	0.419	0.49	0.47	0.478	0.50	0.57	0.391	0.48	0.39	0.027
Ots_105105-613	0.495	0.50	0.53	0.522	0.50	0.48	0.690	0.43	0.39	0.128	0.23	0.26	0.478	0.50	0.51	0.145
Ots_105132-200	0.209	0.33	0.29	0.070	0.13	0.06	0.210	0.33	0.33	0.083	0.15	0.13	0.022	0.04	0.04	0.052
Ots_105385-421	0.054	0.10	0.11	0.060	0.11	0.12	0.038	0.07	0.08	0.000	0.00	0.00	0.245	0.37	0.45	0.099
Ots_105401-325	0.239	0.37	0.37	0.275	0.40	0.35	0.324	0.44	0.43	0.143	0.25	0.19	0.156	0.27	0.27	0.022
Ots_105407-117	0.348	0.46	0.41	0.510	0.50	0.58	0.277	0.40	0.47	0.073	0.14	0.15	0.160	0.27	0.28	0.102
Ots_105897-124	0.122	0.22	0.20	0.102	0.19	0.20	0.090	0.16	0.16	0.198	0.32	0.23	0.011	0.02	0.02	0.028
Ots_106172-425	0.000	0.00	0.00	0.083	0.15	0.17	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.084
Ots_106313-729	0.148	0.25	0.25	0.051	0.10	0.06	0.133	0.23	0.20	0.415	0.49	0.53	0.189	0.31	0.33	0.093
Ots_106419b-618	0.163	0.27	0.28	0.225	0.35	0.29	0.092	0.17	0.16	0.021	0.04	0.04	0.489	0.51	0.41	0.160
Ots_106499-70	0.304	0.43	0.37	0.163	0.28	0.24	0.237	0.36	0.34	0.587	0.49	0.52	0.457	0.50	0.49	0.098
Ots_106747-239	0.339	0.45	0.47	0.310	0.43	0.38	0.393	0.48	0.40	0.239	0.37	0.39	0.553	0.50	0.47	0.037
Ots_107074-284	0.429	0.49	0.49	0.333	0.45	0.38	0.378	0.47	0.44	0.677	0.44	0.52	0.544	0.50	0.35	0.056
Ots_107220-70	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000
Ots_107285-93	0.233	0.36	0.42	0.480	0.50	0.60	0.194	0.31	0.34	0.138	0.24	0.15	0.156	0.27	0.18	0.077
Ots_107607-315	0.255	0.38	0.40	0.177	0.29	0.24	0.236	0.36	0.43	0.117	0.21	0.23	0.261	0.39	0.30	0.011
Ots_107806-821	0.208	0.33	0.35	0.323	0.44	0.44	0.294	0.42	0.41	0.128	0.23	0.17	0.330	0.45	0.39	0.025
Ots_108007-208	0.317	0.44	0.43	0.202	0.33	0.37	0.275	0.40	0.46	0.652	0.46	0.52	0.422	0.49	0.31	0.098

Continued on next page

Table 1.3 – continued from previous page

Assay	**Feather River (Spring) N=94**			**Butte Creek (Spring) N=54**			**Central Valley (Fall) N=94**			**Klam/Trinity (Fall) N=48**			**L. Columbia (Spring) N=47**			
	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	F_{ST}
Ots_108390-329	0.444	0.50	0.48	0.448	0.50	0.44	0.353	0.46	0.47	0.630	0.47	0.57	0.234	0.36	0.38	0.060
Ots_108735-302	0.312	0.43	0.48	0.132	0.23	0.16	0.371	0.47	0.46	0.500	0.51	0.50	0.326	0.44	0.51	0.050
Ots_108820-336	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.362	0.47	0.51	0.362	0.47	0.43	0.337
Ots_109243-285	0.489	0.50*	0.35	0.511	0.51	0.43	0.550	0.50	0.37	0.630	0.47	0.35	0.500	0.51*	0.24	0.001
Ots_109525-816	0.000	0.00	0.00	0.000	0.00	0.00	0.005	0.01	0.01	0.031	0.06	0.06	0.032	0.06	0.06	0.015
Ots_109693-392	0.081	0.15	0.16	0.020	0.04	0.04	0.038	0.07	0.05	0.532	0.50	0.51	0.383	0.48	0.43	0.302
Ots_110064-383	0.500	0.50	0.48	0.548	0.50	0.37	0.429	0.49	0.47	0.609	0.48	0.43	0.554	0.50	0.54	0.011
Ots_110201-363	0.380	0.47	0.54	0.217	0.34	0.35	0.451	0.50	0.48	0.281	0.41	0.48	0.261	0.39	0.35	0.035
Ots_110381-164	0.194	0.32	0.34	0.194	0.32	0.35	0.278	0.40	0.38	0.348	0.46	0.57	0.294	0.42	0.24	0.014
Ots_110495-380	0.411	0.49	0.44	0.271	0.40	0.38	0.450	0.50	0.57	0.106	0.19	0.17	0.348	0.46	0.39	0.067
Ots_110551-64	0.244	0.37	0.36	0.200	0.32	0.32	0.178	0.29	0.31	0.000	0.00	0.00	0.202	0.33	0.32	0.045
Ots_110689-218	0.301	0.42	0.40	0.120	0.21	0.15	0.368	0.47	0.55	0.330	0.45	0.45	0.128	0.23	0.26	0.055
Ots_111084-96	0.484	0.50	0.44	0.568	0.50	0.36	0.467	0.50	0.47	0.844	0.27	0.22	0.815	0.30	0.33	0.118
Ots_111084b-619	0.000	0.00	0.00	0.019	0.04	0.04	0.000	0.00	0.00	0.067	0.13	0.13	0.261	0.39	0.34	0.196
Ots_111312-435	0.000	0.00	0.00	0.000	0.00	0.00	0.006	0.01	0.01	0.556	0.50	0.28	0.171	0.29	0.13	0.456
Ots_111666-408	0.378	0.47	0.52	0.455	0.50	0.50	0.357	0.46	0.43	0.958	0.08	0.08	0.144	0.25	0.24	0.267
Ots_111681-657	0.435	0.49	0.52	0.327	0.44	0.49	0.602	0.48	0.52	0.878	0.22	0.24	0.576	0.49	0.41	0.130
Ots_112208-722	0.494	0.50	0.41	0.413	0.49	0.48	0.432	0.49	0.48	0.915	0.16	0.17	0.389	0.48	0.42	0.140
Ots_112301-43	0.121	0.21	0.24	0.080	0.15	0.16	0.088	0.16	0.13	0.521	0.50	0.57	0.370	0.47	0.48	0.198
Ots_112419-131	0.344	0.45	0.42	0.673	0.44	0.35	0.452	0.50	0.39	0.532	0.50	0.55	0.394	0.48	0.45	0.054
Ots_112820-284	0.462	0.50	0.65	0.245	0.37	0.41	0.456	0.50	0.47	0.870	0.23	0.22	0.733	0.40	0.36	0.179
Ots_112876-371	0.478	0.50	0.52	0.310	0.43	0.42	0.360	0.46	0.42	0.222	0.35	0.31	0.457	0.50	0.39	0.034
Ots_113242-216	0.396	0.48	0.44	0.630	0.47	0.46	0.467	0.50	0.56	0.444	0.50	0.67	0.359	0.47	0.54	0.030
Ots_113457-40	0.233	0.36	0.38	0.226	0.35	0.22	0.299	0.42	0.32	0.032	0.06	0.06	0.217	0.34	0.35	0.043
Ots_115987-325	0.000	0.00	0.00	0.000	0.00	0.00	0.016	0.03	0.03	0.723	0.40	0.43	0.489	0.51	0.49	0.592

Continued on next page

Table 1.3 – continued from previous page

Assay	Feather River (Spring) N=94			Butte Creek (Spring) N=54			Central Valley (Fall) N=94			Klam/Trinity (Fall) N=48			L. Columbia (Spring) N=47			
	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	F_{ST}
Ots_117043-255	0.460	0.50	0.53	0.423	0.49	0.46	0.452	0.50	0.54	0.798	0.33	0.36	0.833	0.28	0.24	0.132
Ots_117138-545	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000
Ots_117242-136	0.447	0.50	0.45	0.143	0.25	0.20	0.456	0.50	0.54	0.229	0.36	0.46	0.207	0.33	0.28	0.089
Ots_117259-271	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.115	0.21	0.19	0.213	0.34	0.30	0.163
Ots_117370-471	0.169	0.28	0.31	0.040	0.08	0.08	0.275	0.40	0.33	0.128	0.23	0.26	0.085	0.16	0.13	0.057
Ots_117432-409	0.444	0.50	0.49	0.290	0.42	0.46	0.318	0.44	0.48	0.128	0.23	0.17	0.294	0.42	0.37	0.051
Ots_118175-479	0.170	0.28	0.32	0.245	0.37	0.29	0.270	0.40	0.36	0.766	0.36	0.43	0.239	0.37	0.43	0.211
Ots_118205-61	0.048	0.09	0.10	0.441	0.50	0.41	0.049	0.09	0.10	0.042	0.08	0.08	0.196	0.32	0.35	0.216
Ots_118938-325	0.183	0.30	0.21	0.163	0.28	0.33	0.196	0.32	0.35	0.063	0.12	0.13	0.652	0.46	0.57	0.206
Ots_120950-417	0.106	0.19	0.21	0.147	0.25	0.14	0.122	0.22	0.24	0.073	0.14	0.10	0.064	0.12	0.13	0.002
Ots_122414-56	0.425	0.49	0.48	0.387	0.48	0.51	0.226	0.35	0.34	0.611	0.48	0.60	0.630	0.47	0.61	0.107
Ots_123048-521	0.483	0.50	0.56	0.365	0.47	0.38	0.359	0.46	0.48	0.000	0.00	0.00	0.130	0.23	0.26	0.156
Ots_123205-61	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000	0.00	0.00	0.000
Ots_123921-111	0.272	0.40	0.41	0.106	0.19	0.21	0.222	0.35	0.31	0.000	0.00	0.00	0.064	0.12	0.13	0.086
Ots_124774-477	0.467	0.50	0.54	0.309	0.43	0.36	0.489	0.50	0.53	0.271	0.40	0.42	0.337	0.45	0.41	0.034
Ots_126619-400	0.406	0.49	0.46	0.276	0.40	0.39	0.353	0.46	0.38	0.359	0.47	0.37	0.319	0.44	0.38	0.001
Ots_127236-62	0.205	0.33	0.30	0.150	0.26	0.26	0.247	0.37	0.43	0.958	0.08	0.08	0.330	0.45	0.40	0.362
Ots_127760-569	0.023	0.04*	0.04	0.011	0.02	0.02	0.000	0.00	0.00	0.000	0.00	0.00	0.075	0.14	0.11	0.034
Ots_128302-57	0.286	0.41	0.40	0.150	0.26	0.26	0.165	0.28	0.24	0.447	0.50	0.47	0.138	0.24	0.15	0.070
Ots_128495b-45	0.352	0.46	0.46	0.457	0.50	0.45	0.397	0.48	0.43	0.261	0.39	0.34	0.467	0.50	0.40	0.014
Ots_128693-461	0.304	0.43	0.39	0.500	0.51	0.36	0.500	0.50	0.49	0.457	0.50	0.53	0.319	0.44	0.34	0.034
Ots_128757-61	0.473	0.50	0.62	0.451	0.50	0.51	0.450	0.50	0.50	0.115	0.21	0.23	0.394	0.48	0.53	0.070
Ots_129144-472	0.346	0.46	0.47	0.451	0.50	0.59	0.332	0.45	0.40	0.000	0.00	0.00	0.117	0.21	0.23	0.129
Ots_129170-683	0.247	0.37	0.34	0.118	0.21	0.24	0.137	0.24	0.23	0.458	0.50	0.42	0.170	0.29	0.21	0.084
Ots_129303b-54	0.407	0.45	0.43	0.471	0.27	0.24	0.362	0.41	0.33	0.083	0.51	0.51	0.245	0.44	0.49	0.043

Continued on next page

Table 1.3 – continued from previous page

	Feather River (Spring) N=94			**Butte Creek (Spring) N=54**			**Central Valley (Fall) N=94**			**Klam/Trinity (Fall) N=48**			**L. Columbia (Spring) N=47**			
Assay	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	AF	H_E	H_O	F_{ST}
Ots_129458-451	0.495	0.50	0.56	0.635	0.47	0.40	0.451	0.50	0.51	0.625	0.47	0.48	0.380	0.48	0.50	0.031
Ots_129870-55	0.299	0.42	0.36	0.337	0.45	0.47	0.301	0.42	0.41	0.021	0.04	0.04	0.160	0.27	0.28	0.070
Ots_130720-99	0.093	0.17	0.19	0.000	0.00	0.00	0.098	0.18	0.17	0.333	0.45	0.49	0.402	0.49	0.28	0.166
Ots_131460-584	0.170	0.28	0.27	0.120	0.21	0.24	0.231	0.36	0.37	0.322	0.44	0.56	0.457	0.50	0.40	0.071
Ots_131802-393	0.214	0.34	0.34	0.240	0.37	0.32	0.225	0.35	0.38	0.021	0.04	0.04	0.075	0.14	0.15	0.052
Ots_131906-141	0.100	0.18	0.18	0.220	0.35	0.32	0.112	0.20	0.22	0.692	0.43	0.53	0.163	0.28	0.24	0.278
Ots_AldB1-122	0.113	0.20	0.20	0.210	0.34	0.22	0.082	0.15	0.14	0.635	0.47	0.52	0.021	0.04	0.04	0.299
Ots_AldoB4-183	0.006	0.01	0.01	0.020	0.04	0.04	0.000	0.00	0.00	0.489	0.51	0.68	0.000	0.00	0.00	0.475
Ots_CathD-141	0.016	0.03	0.03	0.010	0.02	0.02	0.000	0.00	0.00	0.304	0.43	0.61	0.000	0.00	0.00	0.275
Ots_CRB-211	0.005	0.01	0.01	0.000	0.00	0.00	0.005	0.01	0.01	0.096	0.18	0.15	0.000	0.00	0.00	0.072
Ots_EndoRB1-486	0.437	0.50	0.67	0.659	0.45	0.41	0.278	0.40	0.36	0.010	0.02	0.02	0.267	0.40	0.44	0.194
Ots_Hsp90a	0.016	0.03	0.03	0.000	0.00	0.00	0.067	0.13	0.13	0.188	0.31	0.29	0.337	0.45	0.28	0.172
Ots_Myc-366	0.258	0.39	0.43	0.100	0.18	0.12	0.211	0.33	0.36	0.000	0.00	0.00	0.011	0.02	0.02	0.100
Ots_ALDBINT1-SNP1	0.133	0.23	0.24	0.147	0.25	0.22	0.192	0.31	0.32	0.734	0.39	0.32	0.750	0.38	0.41	0.369
Ots_DESMIN19-SNP1	0.218	0.34	0.37	0.170	0.29	0.17	0.207	0.33	0.28	0.174	0.29	0.30	0.149	0.26	0.21	-0.004
Ots_NAML12-SNP1	0.197	0.32	0.28	0.250	0.38	0.42	0.177	0.29	0.27	0.053	0.10	0.11	0.386	0.48	0.36	0.058
Ots_NAML12-SNP2	0.209	0.33	0.37	0.096	0.18	0.15	0.247	0.37	0.36	0.178	0.30	0.31	0.043	0.08	0.09	0.039
Ots_BMP2-SNP1	0.319	0.44	0.40	0.357	0.46	0.43	0.204	0.33	0.28	0.234	0.36	0.30	0.054	0.10	0.11	0.054
Ots_MTA-SNP1	0.239	0.37	0.35	0.271	0.40	0.38	0.390	0.48	0.49	0.042	0.08	0.08	0.266	0.39	0.40	0.071
Ots_TF1-SNP1	0.111	0.20	0.22	0.214	0.34	0.27	0.132	0.23	0.20	0.367	0.47	0.60	0.436	0.50	0.49	0.106
Mean	0.236	0.31	0.31	0.203	0.28	0.26	0.224	0.30	0.30	0.194	0.27	0.28	0.233	0.31	0.29	0.107
Polymorphic Loci (%)		88.0			86.3			88.0			88.0			90.6		

While full annotation of these gene fragments is beyond the scope of the present study, preliminary BLAST (Basic Local Alignment Search Tool, NCBI) results and annotation of the target SNP appear in Table 1.4. Note that we have included annotation not just for the loci described here (Reference 1), but also for an additional 24 loci (References 2, 3, 4 and unpublished) that are part of the final genotyping panel described in Chapter 2. To determine whether the target variation was in an intron or an exon, we aligned the genomic sequence from our Sanger sequencing effort with the EST sequence from which initial primers were designed. Over all 141 loci, 81 SNPs were found in exons while 48 SNPs were found in introns. For 12 loci, EST sequence was unavailable or gene annotation was insufficient to determine intron/exon boundaries. BLAST results revealed identity with a known gene or genomic fragment for 92 loci with E-values ranging from 1.2E-11 to zero (smaller numbers indicating higher similarity). The target SNP was found to be in the 5' or 3' untranslated region (UTR) of a gene for 25 loci, while an annotated translation (n.t.) was unavailable for 10 loci. Introns were found in a variety of locations with respect to the described gene or gene fragment; the introns for 24 loci were found within the coding sequence (CDS) of a gene while other were found up or downstream of the annotated region. Two of the described SNPs are within microsatellite (msat) repeats. Of the variation found in CDS exons, nine were synonymous substitutions while nine represented mutations that altered the resulting amino acid at that position (nonsynonymous); the status of the CDS SNP at Ots_AldoB4-183 could not be determined due to poor sequence homology.

Table 1.4: Preliminary BLAST results (BLAST hit and e-value) and annotation of the target SNP for the loci described here (Reference 1) and for an additional 24 loci (References 2, 3, 4 and unpublished) that are part of the final genotyping panel described in Chapter 2. Also included is whether the variation is present in an intron or exon and its location with respect to the described gene, either in coding sequence (CDS) or untranslated regions (UTR). No translation (n.t.) was available for 10 loci. For CDS exons, a single amino acid is indicated for synonymous substitutions, while both amino acids are included for non-synonymous substitutions. Reference codes are as follows: 1. Clemento *et al.* 2011; 2. Smith *et al.* 2005a; 3. Campbell and Narum 2008; 4. Smith *et al.* 2005b.

Assay name	**Ref**	**SNP**	**Genic location**	**BLAST # [E-value]**	**Description**
Ots_94857-232	1	intron	in MSAT repeat	AY543888 [3.8E-18]	*Salmo salar* clone Alu374 microsatellite sequence
Ots_94903-99	1	exon			
Ots_95442b-204	1	intron	CDS	NM_001123604 [0]	*Salmo salar* somatolactin (LOC100136491)
Ots_96222-525	1	intron	CDS	AB326306 [2.02E-45]	*Solea senegalensis* elongation factor 1 alpha isoform 42Sp50
Ots_96500-180	1	exon			
Ots_96899-357	1	intron	CDS	NM_001165332 [5.1E-56]	*Salmo salar* vaccinia related kinase 3 (vrk3)
Ots_97077-179	1	exon	3' UTR	BT045669 [3.04E-100]	*Salmo salar* BTG
Ots_97660-56	1	exon	5' of gene	EU025717 [7.07E-29]	*Salmo salar* single-strand selective monofunctional uracil
Ots_98409-850	1	intron	CDS	BT072224 [1.01E-45]	*Salmo salar* fizzy-related protein homolog
Ots_98683-796	1	intron	CDS	NM_001141554 [1.32E-83]	*Salmo salar* chymotrypsin-like (ctrl)
Ots_99550-204	1	exon	n.t. (pseudogene)	BT071884 [4.88E-69]	*Salmo salar* collagen alpha-2VI chain precursor
Ots_100884-287	1	exon	3'UTR	NM_001140998 [7.34E-173]	*Salmo salar* centrosomal protein 97 (cep97)
Ots_101119-381	1	intron			
Ots_101554-407	1	exon	3'UTR	BT058832 [8.48E-177]	*Salmo salar* NMDA receptor-regulated protein 1
Ots_101704-143	1	exon	n.t. (pseudogene)	BT071847 [3.89E-97]	*Salmo salar* histidyl-tRNA synthetase
Ots_101770-82	1	exon	CDS [Pro]	BT078766 [5.18E-77]	*Esox lucius* cartilage-associated protein precursor
Ots_102195-157	1	exon	CDS [Pro>Ser]	NM_001160495 [1.99E-119]	*O. mykiss* C type lectin receptor B
Ots_102213-210	1	intron	CDS	BT073541 [5.62E-76]	*O. mykiss* leukocyte cell-derived chemotaxin 2 precursor
Ots_102414-395	1	intron			
Ots_102420-494	1	exon	3'UTR	BT058671 [1.01E-123]	*Salmo salar* cathepsin K precursor
Ots_102457-132	1	exon	3'UTR (STS)	NM_001124228 [2.41E-64]	*O. mykiss* heat shock protein 70a (hsp70a)
Ots_102801-308	1	intron			
Ots_102867-609	1	exon			
Ots_103041-52	1	exon			
Ots_103122-180	1	exon			
Ots_104048-194	1	exon			
Ots_104063-132	1	intron	5' of gene	NM_001160489 [0]	*O. mykiss* mitochondrial complex I subunit NDUFB2
Ots_104216-70	1	exon			
Ots_104415-88	1	exon	3'UTR	NM_001140438 [1.54E-134]	*Salmo salar* asparagine-linked glycosylation 12 homolog

Continued on next page

Table 1.4 – continued from previous page

Assay name	Ref	SNP	Genic location	BLAST # [E-value]	Description
Ots_104569-86	1	exon			
Ots_105105-613	1	exon	CDS [Thr>Ser]	AF483538 [9.12E-90]	*O. mykiss* VHSV-induced protein mRNA
Ots_105132-200	1	exon	3' of gene	NM_001173959 [1.33E-139]	*Salmo salar* member of RAS oncogene family (rap2c)
Ots_105385-421	1	exon			
Ots_105401-325	1	exon	n.t.	ET306160 [9.74E-64]	*Salmo salar* genomic clone S0262B04
Ots_105407-117	1	exon	5'UTR	BT047755 [8.13E-133]	*Salmo salar* 60S ribosomal protein L36a
Ots_105897-124	1	exon	3'UTR	NM_001173890 [2.35E-127]	*Salmo salar* uridine 5-monophosphate synthase (pyr5)
Ots_106172-425	1	exon			
Ots_106313-729	1	intron	CDS	NM_001141707 [9.98E-129]	*Salmo salar* Wilms tumor 1 associated protein-like
Ots_106419b-618	1	intron	5' of gene	AF281332 [5.35E-77]	*O. mykiss* biotinidase fragment 1 mRNA
Ots_106499-70	1	intron	CDS	NM_001124329 [1.2E-11]	*O. mykiss* superoxide dismutase 1 (sod1)
Ots_106747-239	1	exon	CDS [Pro>Gln]	NM_001140141 [8.43E-95]	*Salmo salar* C10orf88 homolog
Ots_107074-284	1	exon			
Ots_107220-70	1	exon			
Ots_107285-93	1	intron	CDS	NM_001146578 [1.05E-73]	*Salmo salar* placental protein 25
Ots_107607-315	1	exon			
Ots_107806-821	1	intron	CDS	BT048833 [1.89E-82]	*Salmo salar* YIPF4
Ots_108007-208	1	exon			
Ots_108390-329	1	intron	3'UTR	BT045571 [3.92E-56]	*Salmo salar* ADP-ribosylation factor-like protein 9
Ots_108735-302	1	exon			
Ots_108820-336	1	exon			
Ots_109243-285	1	exon	3'UTR	NM_001140063 [2.15E-59]	*Salmo salar* Kunitz-type protease inhibitor 2 (spit2)
Ots_109525-816	1	exon	CDS [Ala]	NM_001124667 [1.46E-128]	*O. mykiss* prostaglandin synthase 2b (ptgs2b)
Ots_109693-392	1	exon	3'UTR	NM_001139789 [0]	*Salmo salar* nuclear transcription factor Y subunit gamma
Ots_110064-383	1	intron	CDS	NM_001165121 [4.29E-73]	*O. mykiss* lipopolysaccharide-induced tnf-alpha
Ots_110201-363	1	intron			
Ots_110381-164	1	exon			
Ots_110495-380	1	intron			
Ots_110551-64	1	exon	n.t.	AL954310 [8.23E-35]	*Danio rerio* DNA sequence in linkage group 8
Ots_110689-218	1	exon			
Ots_111084-96	1	exon			
Ots_111084b-619	1	intron			
Ots_111312-435	1	exon	n.t. (pseudogene)	BT072385 [0]	*Salmo salar* DNA topoisomerase 1
Ots_111666-408	1	intron	CDS	NM_001140911 [2.1E-50]	*Salmo salar* Tetraspanin-16 (tsn16)
Ots_111681-657	1	exon	CDS [Thr>Pro]	NM_001124224 [0]	*O. mykiss* NK2 homeobox 1b (nkx2.1b)
Ots_112208-722	1	intron	CDS	DQ784539 [0]	*O. mykiss* CW lactate dehydrogenase B gene
Ots_112301-43	1	intron			
Ots_112419-131	1	exon	3'UTR	NM_001139960 [2.38E-107]	*Salmo salar* junction plakoglobin (plak)

Continued on next page

Table 1.4 – continued from previous page

Assay name	Ref	SNP	Genic location	BLAST # [E-value]	Description
Ots_112820-284	1	exon			
Ots_112876-371	1	intron	CDS	HE608241 [0]	*O. mykiss* TPT1 gene for tumor protein
Ots_113242-216	1	exon	3'UTR	NM_001165318 [6.49E-153]	*Salmo salar* exostoses (multiple) 1c (ext1c)
Ots_113457-40	1	intron	n.t. (pseudogene)	BT071894 [3.63E-85]	*Salmo salar* C-ets-2
Ots_115987-325	1	intron			
Ots_117043-255	1	intron			
Ots_117138-545	1	exon			
Ots_117242-136	1	exon			
Ots_117259-271	1	exon	3' of gene	HM159472 [2.14E-41]	*Salmo salar* Foxl2-like protein (Foxl2) gene
Ots_117370-471	1	intron	CDS	AJ003200 [1.97E-55]	*Danio rerio* mRNA for ETS-domain transcription factor PEA3
Ots_117432-409	1	exon			
Ots_118175-479	1	exon			
Ots_118205-61	1	exon	CDS [Ala>Thr]	BT125411 [0]	*Salmo salar* Sjoegren syndrome autoantigen 1 homolog
Ots_118938-325	1	intron			
Ots_120950-417	1	exon			
Ots_122414-56	1	intron	n.t. (pseudogene)	BT072655 [2.71E-77]	*Salmo salar* histone deacetylase complex subunit SAP130
Ots_123048-521	1	intron	CDS	BT149991 [7.43E-56]	*Salmo salar* ribosomal protein S26 mRNA
Ots_123205-61	1	exon	3'UTR	NM_001139756 [6.84E-38]	*Salmo salar* transposase-like (LOC100194703)
Ots_123921-111	1	intron			
Ots_124774-477	1	exon	3'UTR	NM_001173779 [2.01E-165]	*Salmo salar* cAMP-responsive element-binding protein (cr3l2)
Ots_126619-400	1	intron	5' of gene	NM_001139980 [1.14E-68]	*Salmo salar* Acyl-CoA desaturase (acod)
Ots_127236-62	1	exon			
Ots_127760-569	1	exon	CDS [Ile]	NM_001140111 [6.59E-101]	*Salmo salar* Zinc finger protein 503 (zn503)
Ots_128302-57	1	exon	3'UTR	BT073572 [5.14E-138]	*O. mykiss* ribosomal protein L20
Ots_128495b-45	1	intron			
Ots_128693-461	1	intron	CDS	NM_001141075 [3.57E-94]	*Salmo salar* stathmin-like 4 (stmn4)
Ots_128757-61	1	exon	3'UTR	BT057575 [3.08E-176]	*Salmo salar* thymosin beta-11
Ots_129144-472	1	exon			
Ots_129170-683	1	intron			
Ots_129303b-54	1	intron			
Ots_129458-451	1	exon	3'UTR	NM_001139945 [7.48E-99]	*Salmo salar* FK506 binding protein 8 (fkbp8)
Ots_129870-55	1	exon			
Ots_130720-99	1	exon			
Ots_131460-584	1	intron	CDS	BT072067 [5.46E-74]	*Salmo salar* neural cell adhesion molecule L1-like precursor
Ots_131802-393	1	exon			
Ots_131906-141	1	intron	n.t.	AB258536 [1.62E-49]	*O. mykiss* Onmy-LDA gene for MHC class I antigen
Ots_AldB1-122	1	unk.	5' of gene	NM_001123627 [3.28E-17]	*Salmo salar* aldolase b and fructose-bisphosphate (aldob)
Ots_AldoB4-183	1	exon	CDS [poor match]	NM_001123627 [manual]	*Salmo salar* aldolase b (aldob)

Continued on next page

Table 1.4 – continued from previous page

Assay name	Ref	SNP	Genic location	BLAST # [E-value]	Description
Ots_CathD-141	1	exon	3'UTR (STS)	NM_001124711 [1.27E-170]	*O. mykiss* cathepsin D
Ots_CRB-211	1	intron	CDS	AF100933 [0]	*O. mykiss* carbonyl reductase dehydrogenase A
Ots_EndoRB1-486	1	unk.	n.t.	CU041379 [2.65E-36]	*Danio rerio* DNA sequence in linkage group 16
Ots_Hsp90a	1	unk.			Originally isolated from heat shock protein 90A
Ots_Myc-366	1	exon	CDS [Asp]	NM_001124699 [2.45E-122]	*O. mykiss* v-myc oncogene homolog
Ots_ALDBINT1-SNP1	1	exon	5'UTR	NM_001123627 [1.11E-66]	*Salmo salar* aldolase b (aldob)
Ots_DESMIN19-SNP1	1	unk.	n.t.	CR708193 [1.84E-14]	*Tetraodon nigroviridis* full-length cDNA
Ots_NAML12-SNP1	1	unk.			
Ots_NAML12-SNP2	1	unk.			
Ots_BMP2-SNP1	1	exon	CDS [Ser]	NM_001173834 [0]	*Salmo salar* bone morphogenetic protein 2 (bmp2)
Ots_MTA-SNP1	1	intron	CDS	DQ139342 [5.84E-101]	*O. tshawytscha* metallothionein A (metA)
Ots_TF1-SNP1	1	exon	CDS [Lys>Asn]	AF488833S1 [1.26E-147]	*Salmo salar* isolate Ssa-1 transferrin gene
Ots_ARNT-195	n/a	exon	3'UTR	NM_001124710 [0]	*O. mykiss* aryl hydrocarbon receptor translocator
Ots_RAG3	n/a	exon	3'UTR	OMU73750 [0]	*O. mykiss* RAG gene and 3'UTR
Ots_AsnRS-60	2	exon	CDS [Pro]	DQ025751 [3.71E-159]	*O. tshawytscha* Ots.AsnRS.82.35 genomic
Ots_aspat-196	3	intron	CDS	EF042601 [0]	*O. tshawytscha* aspartate aminotransferase gene
Ots_CD59-2	n/a	exon	5'UTR	NM_001124422 [1.38E-49]	*O. mykiss* CD59-like protein 2 (cd59-2)
Ots_CD63	n/a	exon	CDS [Gly]	NM_001124496 [1.78E-100]	*O. mykiss* Cd63 antigen (cd63)
Ots_EP-529	n/a	exon	CDS [Asp>Asn]	NM_001124693 [0]	*O. mykiss* ependymin (om-i)
Ots_GDH-81x	3	exon	3'UTR	EF042600 [0]	*O. tshawytscha* glutamate dehydrogenase gene
Ots_HSP90B-385	2	exon	CDS [Gly]	DQ908921 [0]	*O. tshawytscha* heat shock 90 kDa protein gene
Ots_MHC1	4	exon	CDS [Trp>Arg]	AF104585 [2.1E-110]	O. tshawytsha isolate Onts-B*3 MHC class I alpha 2
Ots_mybp-85	n/a	unk.			
Ots_myoD-364	3	exon	3'UTR	EF042596 [0]	*O. tshawytscha* myoD gene
Ots_Ots311-101x	3	intron	in MSAT repeat	AF393194 [7.7E-84]	*O. tshawytscha* msat OtsG311 sequence
Ots_PGK-54	n/a	exon	CDS [Lys>Met]	NM_001139794 [7.07E-71]	*Salmo salar* Phosphoglycerate kinase (pgk)
Ots_Prl2	4	intron	CDS	S66606 [0]	*O. tshawytscha* prolactin II
Ots_RFC2-558	2	unk.	n.t.	DQ025583 [6.37E-90]	*O. tshawytscha* clone Ots.RFC2.38.66
Ots_SClkF2R2-135	2	intron	CDS	DQ780892 [4.74E-112]	*O. tshawytscha* CLOCK1a (Clock1a)
Ots_SWS1op-182	2	intron	CDS	NM_001124321 [2.08E-98]	*O. mykiss* SWS1 opsin (LOC100135983)
Ots_TAPBP	n/a	exon	CDS [Ser]	NM_001124553 [6.37E-60]	*O. mykiss* tapasin long form
Ots_u07-07.161	n/a	unk.	n.t.	EU616651 [0]	*O. tshawytscha* isolate u07-07 SNP assay target
Ots_u07-49.290	n/a	unk.	n.t.	EU616659 [0]	*O. tshawytscha* isolate u07-49 SNP assay target
Ots_u4-92	2	unk.	n.t.	DQ025560 [8.66E-176]	*O. tshawytscha* clone Ots.u4.51.26
Ots_S71-336	n/a	unk.			
Ots_unk-526	n/a	intron	n.t.	GU817335 [1.35E-57]	*Salmo salar* clone BAC CHORI214-083H23

1.5 Discussion

We describe a large set of new genetic resources for Chinook salmon, one of the worlds most economically important fish species and a major component of north Pacific ocean fisheries. A large EST sequencing effort was undertaken that evaluated variation in 225 gene fragments, and in over 131 kb of genomic sequence, in an average of 16 individual salmon each. The resulting 117 SNP assays that were successfully validated in Chinook salmon were broadly polymorphic and had substantial power for biological inference. This effort more than doubles the number of published SNP assays available for this species. Applications for these markers include genetic stock identification (GSI) for mixed fishery and ecological applications, individual identification, linkage mapping and pedigree construction.

GSI is becoming a major component of salmon fishery management, with mixed ocean and inland fisheries being evaluated with molecular markers in all Pacific salmon species. GSI requires markers with frequency differences between populations in a reference baseline database (Seeb *et al.* 2007). The 117 SNPs described here have a mean estimated F_{ST} of 0.107 for the five Chinook salmon stocks evaluated here, which represent the three largest river systems in the coterminous United States: the Sacramento, Klamath and Columbia Rivers. Pairwise F_{ST} values for individual loci ranged from 0 to 0.592, indicating that some of these loci are subject to dramatically different evolutionary forces than others. Since a substantial fraction of these novel SNPs are located within coding regions of genes, it is possible that some of the polymorphisms

targeted by our assays are directly influenced by natural selection. More likely is that a number of these markers are located in genomic regions that have been affected by recent natural selection and the allele frequency differences are the result of hitchhiking effects (Barton 2000). Regardless, the substantial allele frequency differences between populations present at many of these SNP loci indicate that they will be very useful for stock discrimination, exceeding microsatellite loci in discriminatory power when weighted by the total number of alleles. Initial analyses (data not shown) indicate that by selecting the best set of 96 loci, which is convenient for our genotyping platform, from this and other published markers (Smith *et al.* 2005b; Campbell *et al.* 2008), assignment accuracy for GSI is as good or better with 96 SNPs than with the 13 standardized microsatellites currently in use (Seeb *et al.* 2007), for Chinook salmon ocean fishery mixtures south of the Columbia River basin.

Pedigree reconstruction, in the form of large-scale parentage inference, has been proposed as an alternative to physical tags for salmonids and other fishes (Hankin *et al.* 2005; Garza and Anderson 2007). This method, termed parentage-based tagging (PBT), involves genotyping reproducing individuals and using their genotypes as intergenerational genetic tags that are recovered through parentage inference with their progeny. PBT has some distinct advantages over traditional large-scale tagging programs such as coded-wire tags, including 1) individual-specific tag recoveries, 2) no tagging or handling of juvenile fish, with their associated very low recovery rates (<2 recoveries per 1000 tags in Chinook salmon; Hankin *et al.* 2005), 3) fish can be non-lethally sampled during seaward migration, in fisheries, and upon return to spawn, and 4) valuable corollary

data in the form of a large number of pedigrees (Garza and Anderson 2007). Over time, some of these pedigrees will become extensive and can serve as the basis for detailed linkage maps and associated mapping of quantitative trait loci (Boulding *et al.* 2008; Moen *et al.* 2008; Pemberton 2008). Goverment agencies have traditionally mitigated the terrestrial and aquatic ecosystem impacts responsible for salmonid population declines with production of fish in hatcheries and subsequent population supplementation. Millions of Chinook salmon originate in hatcheries each year and they are the majority of fish in some populations (Barnett-Johnson *et al.* 2007). Such genetic tagging, and the analysis of the associated pedigrees, could have considerable importance in understanding the effects of hatchery practices on life history parameters and fitness, since the entire production can be tracked by simply collecting genotypes from all broodstock at spawning.

PBT involves the identification of true parents from among very large sets of potential parents, which in turn requires the accurate evaluation of exceedingly small error rates to avoid biologically important rates of false positive pedigree reconstruction. Anderson and Garza (2006) describe novel importance sampling methods for estimating such probabilities and demonstrated that PBT can be used to accurately reconstruct parent/offspring trios in salmon using 80-100 SNP markers with a mean MAF of 0.20. The top 96 SNP loci described here have mean MAF of 0.22 (Klamath) to 0.28 (Feather) in the five focal populations, and the inclusion of other published SNP markers in an optimal set of 96 loci increases mean MAF even further for all populations (data not shown). With pedigree-based inference, minimizing genotyping errors is also critical,

since they can cause apparent Mendelian incompatibilities. The lower genotyping error and mutation rates of SNP markers combine to make them the preferred type of marker for the large scale data generation and parentage analyses necessary to implement PBT (Anderson and Garza 2006; Garza and Anderson 2007). Microsatellite markers are still useful for both GSI and small-scale pedigree reconstruction, particularly for inferring non-parent/offspring relationships, but the reduced staff time and other costs associated with SNP genotype generation, as well as the portability of SNP data, led Hankin *et al.* (2005) to recommend a transition to SNP markers for multilateral, collaborative research and management of Pacific salmon ocean fisheries.

The MAF requirements for SNP loci in GSI and PBT applications are different. GSI requires frequency differences between populations, which are maximized when loci are fixed for alternative alleles in different populations or lineages. In contrast, power for PBT is entirely dependent upon the mean MAF of the set of SNP loci employed, which is maximized when all loci have two alleles at equal frequency in the focal population. With our balanced ascertainment panel, we were able to discover a set of SNP loci with both MAFs in our focal California populations that exceed those necessary for PBT applications (Anderson and Garza 2006) and also with sufficient allele frequency differences to have high power with GSI. The eventual combination of GSI and PBT analyses on the same genotypic data in a single analytical framework will be a major advance in genetic tagging methodology. With such an integrated GSI/PBT system, all fish genotyped with the same set of markers will yield biological inference, either individual identification when parents are sampled (or a fish is recaptured), or

population assignment using a baseline reference database if they are not directly linked to other sampled individuals in a pedigree.

Our discovery effort employed a balanced ascertainment approach, which included an ascertainment sample with representatives of a number of Chinook salmon lineages, and a design criterion that targeted all loci with sufficient variation and generally did not discriminate on the basis of the population in which the variation was found. This strategy led to a set of loci that were similarly variable in all of the populations for which validation was pursued. The inclusion of fish in the ascertainment and validation samples that display variation in migration (yearling and sub-yearling outmigrants) and maturation strategies (fall-run and spring-run types), may also provide additional power for the discrimination of fish from stocks that are differentiated primarily due to these life history strategies (e.g. Central Valley-Fall vs. Central Valley-Spring). Nevertheless, these markers are an upwardly biased sample of the SNP MAF spectrum in the ascertainment populations, because of the three genotypes design criterion, and rare SNP alleles are underrepresented.

It is also important to note that our five validation populations are part of only one of the major lineages of Chinook salmon (Waples *et al.* 2004; Seeb *et al.* 2007), with the species extending across the North Pacific rim to Asia, and extensive differentiation throughout the range. However, the ascertainment sample also included representatives of populations from British Columbia lineages, so the SNP markers described here should be more broadly useful in the southeastern part of the species range. Still, these markers are expected to overestimate the mean MAF and proportion of polymorphic

loci in populations that are not part of the lineages in the ascertainment sample. Bias corrections can be used to better approximate marker polymorphism and differentiation for phylogenetically distance populations (Clark *et al.* 2005; Albrechtsen *et al.* 2010), but ultimately more SNP markers will need to be ascertained for applications in these Chinook salmon lineages.

Next-generation sequencing (NGS; pyrosequencing, single base extension) is a potentially powerful method for discovering SNPs and other genomic variation. Indeed, NGS is unparalleled for the identification of many potential candidate markers rapidly and at minimal expense. However, when genomic resources and sequence data for a target species exist, the most important components of SNP discovery become 1) validation of observed substitutions as true SNPs and not artifacts, 2) choosing an optimal set of polymorphisms for downstream applications using MAF and linkage criteria, and 3) avoiding ascertainment bias. In highly structured species, such as most salmonid fishes, it is critical to obtain sequence data from the same genomic regions in a diverse sample of individuals to minimize ascertainment bias. There is abundant existing genomic sequence in Oncorhynchus species for SNP discovery, we investigated only about 500 of the nearly 100K EST sequences in the *O. mykiss* Gene Index (and many of the *Salmo salar* Gene Index ESTs are likely informative as well), and about 80% of all the gene fragments we investigated had observed substitutions. Since it is easier to ensure identical genomic coverage across individuals with traditional Sanger sequencing than with NGS, and candidate SNPs discovered with NGS are typically resequenced in ascertainment panels prior to assay development anyway, it may be more economical

to employ traditional sequencing strategies to develop additional markers for population genetic applications in other parts of the Chinook salmon range. In contrast, the genomic needs for other applications, such as construction of microarray, linkage and physical maps, are likely to be best met with NGS strategies.

Chapter 2

Evaluation of a SNP baseline for genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) in the California Current Large Marine Ecosystem[1]

2.1 Abstract

Chinook salmon from the West Coast of North America are an economically and ecologically important species and a major component of North Pacific Ocean fisheries. Their anadromous life history strategy generates populations (or stocks) that are

[1]accepted with revisions, resubmitted, awaiting editorial approval: Clemento, A.J., E.D. Crandall, J.C. Garza and E.C. Anderson, Fishery Bulletin, 2014

frequently genetically differentiated from one another, although not visually discernable. In many cases, it is desirable to discern the stock of origin of an individual fish or the stock composition of a mixed sample to monitor stock-specific impacts and alter management accordingly. Genetic stock identification (GSI) provides such discrimination and we describe here a novel GSI baseline composed of genotypes from over 8,000 individual fish from 69 distinct populations at 96 single nucleotide polymorphism (SNP) loci. The populations included in the baseline represent the likely sources for over 99% of the fish encountered in ocean salmon fisheries off California and Oregon. This new genetic baseline permits GSI using rapid and cost effective SNP genotyping, and power analyses indicate that it has near maximum power for discriminating most Chinook salmon stocks to the level of resolution needed for fishery management by the Pacific Fishery Management Council. In an ocean fishery sample, GSI assignments of over 1000 fish, using our baseline, were highly concordant ($\sim$99%) at the reporting unit level to identifications from the physical coded wire tags recovered from the same fish. This SNP baseline represents an important advance in the technologies available for fishery management and ecological investigation of Chinook salmon at the southern end of their geographic range.

2.2 Acknowledgments

The authors would like to thank the entire Molecular Ecology and Genetic Analysis Team in the Fisheries Ecology Division of the SWFSC for their invaluable assistance with genotyping and analyses. Of critical importance to the successful completion of this project were the baseline samples provided to us by: California Department of Fish and Game (S. Harris), Hoopa Valley Tribal Fisheries Department (G. Kautsky), Oregon Department of Fish and Wildlife, Oregon State University Department of Fisheries and Wildlife (M. Banks), Idaho Department of Fish and Game (M. Campbell), Columbia River Inter-Tribal Fish Commission (S. Narum), NOAA Northwest Fisheries Science Center (P. Moran), U.S. Fish and Wildlife Service (M. Brown, D. Hawkins, and C. Smith), Washington Department of Fish and Wildlife (S. Blankenship and K. Warheit), University of Washington School of Aquatic and Fishery Sciences (L. Seeb), Department of Fisheries and Oceans, Canada (T. Beacham), and Alaska Department of Fish and Game (W. Templin). Fishery samples were collected by the California Department of Fish and Game, and provided to us by M. Heisdorf and M. Palmer-Zwahlen. We also thank T. Beacham and two anonymous referees for comments, which improved this manuscript. This project received funding from NOAAs Cooperative Fisheries Research Program and the Southwest Fisheries Science Center. A. Clemento also received support from a California Bay Delta Science Fellowship and the University of California Coastal Environmental Quality Initiative. Many of the baseline samples were collected and DNA extracted with funds from the Pacific Salmon Commission.

2.3 Introduction

Chinook salmon (*Oncorhynchus tshawytscha*) are found in rivers from central California around the North Pacific Rim to Russia (as well as those draining into the Bering Sea), and are the target of valuable commercial and recreational fisheries. A key component of the Chinook salmon life history is natal homing, whereby these anadromous fish typically return to spawn in the same river in which they were born. This homing generates populations (or stocks) that may be genetically differentiated from neighboring populations and can exhibit local adaption (Utter 1989, Taylor 1991). Recent population declines, particularly at the southern end of the species native range where many stocks are listed under the US Endangered Species Act (ESA; Federal Register 1990, 1999), have highlighted the need to refine the management and conservation of Chinook salmon. However, such refinements are challenging, because the migratory life history of salmon means that the many anthropogenic impacts occurring in rivers or in the ocean (e.g. fisheries, water diversion, or turbine entrainment) may affect multiple, intermingled stocks. In such cases, it may be necessary to discern the stock of origin of affected fish to monitor stock-specific impacts and design management strategies accordingly.

The use of pre-existing biological markers to distinguish salmon stocks has a long history. The traits used in these efforts have included morphometric and meristic characters (Fournier *et al.* 1984, Claytor and MacCrimmon 1987), scale patterns (Cook, 1982), parasite assemblages (Boyce 1985), and stable isotope ratios (Barnett-Johnson

et al. 2008). However, the most universally applicable methods have involved the use of genetic markers, since every fish has a unique genetic makeup. The first genetic markers widely used for identification in salmon were electrophoretically detectable protein polymorphisms known as allozymes (Milner *et al.* 1985, Shaklee and Phelps 1990, Tessier *et al.* 1995, Allendorf and Seeb 2000). With the advent of the polymerase chain reaction (PCR), many more types of genetic markers became available to discriminate salmon populations, including mitochondrial DNA polymorphisms (Cronin *et al.* 1993), minisatellites (Beacham *et al.* 1996, Miller *et al.* 1996), microsatellites (Seeb *et al.* 2007, Moran *et al.* 2013), amplified fragment length polymorphisms (Flannery *et al.* 2007) and, most recently, single nucleotide polymorphisms (SNPs; Smith *et al.* 2005a, Smith *et al.* 2005b, Aguilar and Garza 2008, Narum *et al.* 2008, Abadía-Cardoso *et al.* 2011, Clemento *et al.* 2011).

Genetic stock identification (GSI) typically proceeds in two steps. First, samples are collected from potential source populations and genotyped with a set of genetic markers in order to estimate population allele frequencies. These genotypes are called the baseline. Then, data from individuals sampled from a mixed-stock collection (a mixture) and genotyped with the same set of genetic markers are compared to the baseline to estimate the relative proportions of individuals from each of the represented source populations. Single individuals of unknown origin can also be assigned to specific populations. GSI inference is typically carried out using maximum likelihood or Bayesian methods (Smouse *et al.* 1990, Pella and Masuda 2000).

The first large-scale baseline for GSI of Chinook salmon utilized allozyme mark-

ers (Teel *et al.* 1999), but technical and logistical issues limited their future appeal. The allozyme database was supplanted in Canada by a microsatellite baseline developed by the Department of Fisheries and Oceans (Beacham *et al.* 2006), and more broadly by a microsatellite baseline database developed through a large, international collaboration (Seeb *et al.* 2007). This collaboration required enormous effort to standardize data across labs, as microsatellite allele names and sizes are not usually consistent between different labs and genotyping equipment. The Seeb *et al.* (2007) microsatellite baseline has been an effective tool for GSI but has a number of disadvantages: genotyping and scoring of microsatellites is labor-intensive; genotyping error rates can be relatively high, making the 13 microsatellites in that baseline inadequate for applications such as pedigree reconstruction (Anderson and Garza 2006, Garza and Anderson 2007, Abadía-Cardoso *et al.* 2013); missing data rates can also be quite high; and, finally, any new laboratory that wishes to use the baseline must undertake a costly standardization process. Additionally, it has now been demonstrated that SNPs, despite typically having only two alleles per-locus, do have sufficient power to be successfully employed in a GSI context with a modest number of genetic markers (Smith *et al.* 2007, Narum *et al.* 2008, Templin *et al.* 2011, Larson *et al.* 2013).

Early simulation studies suggested that the biallelic nature of SNPs would make them less useful than highly polymorphic microsatellites for population discrimination (Bernatchez and Duchesne 2000, Kalinowski 2004). However, SNPs are located throughout the genome and may be discovered in genetic regions with higher than average divergence (Nosil *et al.* 2009), increasing their utility for GSI. Moreover, SNPs do

not suffer from many of the disadvantages of microsatellites: SNP markers are amenable to the automated, high-throughput genotyping required for large projects; SNP genotyping error rates are very low, making them suitable for pedigree reconstruction; and, importantly, SNP assays do not typically require standardization between labs, so a SNP baseline is immediately useful to any group or agency that genotypes a mixture sample with the markers used in the baseline (Seeb *et al.* 2011).

Here, we describe the development and evaluation of a new baseline of SNP marker data for Chinook salmon in the southern part of their native range for use in ecological investigation in the California Current Large Marine Ecosystem (and its tributaries) and in fisheries managed by the Pacific Fishery Management Council (PFMC). We introduce a panel of 96 SNP markers and a baseline of nearly 8,000 salmon from 68 Chinook salmon populations ranging from California to Alaska. We describe the procedures used to select these SNP markers from amongst a larger number of candidates and document the resulting patterns of genetic differentiation between various populations. We evaluate the power of the new baseline for GSI by both self-assignment and simulated mixture analyses, focusing on stocks commonly encountered in PFMC fisheries. Finally, we analyze 2,090 fish sampled in 2010 from the sport and commercial fisheries off the coast of California and compare the results of these analyses to the coded wire tag (CWT) data from these fish to demonstrate the effectiveness of the baseline for classifying individuals to specific management units.

2.4 Methods

2.4.1 Baseline Populations

Populations were selected for inclusion in the baseline to provide broad geographic coverage across the range of Chinook salmon in the coterminous United States, from Washington to California, while also allowing for the identification of fish from elsewhere in the species' geographic range. Adult fish were sampled on spawning grounds, in terminal fisheries or at hatcheries over the last decade and were provided by numerous contributors (see Acknowledgments and Warheit *et al.* 2013). We included populations expected to be encountered in ocean fisheries off California and Oregon, as well as populations with special management status (e.g. ESA-listed). Accordingly, the major lineages of Chinook from California and Oregon are emphasized in the baseline, as were populations distinguished by life history strategy (spring-run, fall-run, winter-run, etc.), but representatives of the major lineages from further north were also included. DNA was extracted from samples for California populations using DNEasy Blood and Tissue kits on a BioRobot3000 (QIAGEN, Inc., Valencia, CA) according to the manufacturers protocols, while DNA from populations in Oregon, Washington, Canada and Alaska was extracted by the contributors (see Acknowledgments) using various methods. Sample sizes ranged from 44 to 1409 individuals per population and averaged 116. The 1409 fish from the Trinity River Hatchery were initially genotyped with our SNP panel for another purpose, but were included here in total to provide a comprehensive reference sample for identifying this important group. Excluding this disproportionately large

sample, the average number of individuals per population was 97. In total, the baseline includes 7,984 Chinook salmon from 68 distinct populations (Table 2.1).

Each population in the baseline belongs to a single reporting unit, a designation established in previous GSI work that reflects a combination of "genetic similarity, geographic features and management applications" (Seeb *et al.* 2007). Reporting units are generally composed of multiple populations that share genetic similarity or are subject to similar management regimes. The 68 Chinook salmon populations in our baseline fall into 38 distinct reporting units (Table 2.1) and some reporting units in Alaska and Canada are represented by only a single population.

Coho salmon (*Oncorhynchus kisutch*) are occasionally misidentified as Chinook salmon in ocean fisheries and in ecological sampling. We included a collection of 47 coho salmon from California as the 69th population in the baseline to assist in identifying coho salmon that have been incorrectly identified as Chinook salmon.

Table 2.1: Populations and reporting groups in the single-nucleotide polymorphism baseline for genetic stock identification of Chinook salmon from the West Coast of North America. Shown are the names used on the phylogeographic tree (Figure 2.1), the total number of individuals sampled (n), the number used in the training set (n_t), estimates of unbiased expected (Exp.) and observed (Obs.) heterozygosity (Hz), and the mean number of alleles (A); also shown are the proportion of individuals that self-assign (Assign.) to the population (pop.) from which they were sampled and the proportion that self-assign to the correct reporting (rep.) group, as well as the mean F_{ST} for each population within and between reporting groups. Note that mean summary values shown were calculated excluding the coho salmon sample.

Reporting Group	**Population**	**Tree Name**	**n**	**n_t**	**Exp. Hz**	**Obs. Hz**	**A**	**Assign. to pop.**	**Assign. to rep. group**	**Mean F_{ST} w/in**	**Mean F_{ST} btw.**
Central Valley spring	Butte Creek spring	CVsp_Butte	425	26	0.357	0.33	1.99	0.68	0.93	0.017	0.196
	Mill Creek spring	CVsp_Mill	145	23	0.371	0.377	1.99	0.48	0.8	0.012	0.173
	Deer Creek spring	CVsp_Deer	119	12	0.367	0.346	1.99	0.5	0.8	0.013	0.174
	Up. Sacramento R. sp.	CVsp_UpSac_late	372		0.368	0.355	1.99	0.26	0.78	0.008	0.175
Central Valley fall	Feather R. Hatchery sp.	CVfl_FeatherRHsp	470	47	0.373	0.374	1.99	0.44	0.87	0.009	0.179
	Feather R. Hatchery fall	CVfl_FeatherRHfl	146	23	0.37	0.371	1.98	0.18	0.85	0.004	0.19
	Butte Creek fall	CVfl_Butte	188		0.369	0.355	2	0.13	0.91	0.003	0.187
	Mill Creek fall	CVfl_Mill	97	12	0.366	0.358	1.98	0.14	0.95	0.004	0.2
	Deer Creek fall	CVfl_Deer	70		0.363	0.347	1.98	0.29	0.9	0.005	0.195
	Mokelumne River fall	CVfl_Mklmne	95	27	0.37	0.373	1.98	0.26	0.94	0.005	0.198
	Battle Creek fall	CVfl_Battle	141	23	0.369	0.351	1.99	0.29	0.89	0.005	0.188
	Up. Sac. R. late-fall	CVfl_UpSac	93	23	0.367	0.364	2	0.54	0.93	0.01	0.193
Central Valley winter	Sacramento R. winter	Sac_win	295	19	0.297	0.289	1.97	1	1	-	0.263
California Coast	Eel River	CACoast_Eel	95	12	0.327	0.321	2	0.89	0.96	0.029	0.203
	Russian River	CACoast_Russian	94		0.368	0.372	2	0.84	0.98	0.029	0.156
Klamath River	Iron Gate Hatchery	Klamath_IronGH	117	12	0.326	0.345	1.97	0.97	0.99	0.053	0.232
	Trinity River Hatchery	Klamath_TrinityHsp	1409	12	0.318	0.312	2	0.93	0.97	0.053	0.243
N. California/	Smith River	nCal_sOR_Smith	159		0.377	0.381	1.99	0.77	0.87	0.014	0.138
S. Oregon Coast	Chetco River	nCal_sOR_Chetco	94	11	0.372	0.367	1.99	0.73	0.86	0.014	0.137
Rogue River	Cole Rivers Hatchery	Rogue_ColeRHsp	141	11	0.367	0.362	2	0.62	0.86	0.006	0.155
	Applegate Creek	Rogue_Applgt	92		0.369	0.361	2	0.5	0.77	0.006	0.153
Mid Oregon Coast	Coquille River	mOR_Coquille	47		0.352	0.343	1.99	0.72	0.83	0.039	0.151
	Umpqua River spring	mOR_Umpqua	137	11	0.386	0.375	2	0.63	0.64	0.055	0.119
	Siuslaw River	mOR_Siuslaw	93		0.345	0.348	1.98	0.4	0.46	0.04	0.146

Continued on next page

Table 2.1 – continued from previous page

Reporting Group	Population	Tree Name	n	n_t	Exp. Hz	Obs. Hz	A	Assign. to pop.	Assign. to rep. group	Mean F_{ST} w/in	Mean F_{ST} btw.
North Oregon Coast	Nestucca Hatchery	nOR_NestuccaH	48		0.338	0.328	1.96	0.71	0.83	0.029	0.16
	Alsea River	nOR_Alsea	131		0.335	0.309	2	0.47	0.76	0.042	0.159
	Nehalem River	nOR_Nehalem	93		0.316	0.317	1.96	0.97	0.99	0.059	0.193
	Siletz River	nOR_Siletz	93		0.331	0.33	1.98	0.69	0.81	0.031	0.163
Willamette River	N. Santiam Hatchery	Will_NSantiamH	93		0.324	0.327	1.95	0.8	0.99	0.014	0.181
	McKenzie Hatchery	Will_McKenzHsp	48		0.334	0.376	1.94	0.69	0.96	0.014	0.175
Deschutes River fall	Lower Deschutes River	Deschutes_fl	94		0.366	0.357	2	0.56	0.56	-	0.145
Low. Columbia R. fall	Cowlitz Hatchery fall	COlow_CowHfl	141		0.365	0.374	1.99	0.79	0.79	-	0.142
Lower Columbia R. spring	Cowlitz Hatchery sp.	COlow_CowHsp	44	11	0.368	0.37	1.97	0.67	0.7	0.029	0.16
	Kalama Hatchery sp.	COlow_KalamaHsp	48	12	0.372	0.359	1.99	0.5	0.61	0.029	0.134
Mid Columbia R. tule	Spring Creek Hatchery	COmid_SpringCH	142		0.322	0.331	1.97	0.97	0.97	-	0.206
Upper Columbia R. summer/fall	Hanford Reach	COup_Hanford	92		0.355	0.353	1.99	0.36	0.76	0.002	0.166
	Priest Rapids Hatchery	COup_PriestHsumfl	48		0.361	0.359	1.99	0.25	0.83	0.002	0.164
	Wells Hatchery	COup_WellsHsumfl	48		0.355	0.369	1.99	0.46	0.92	0.004	0.175
Mid/Up. Columbia River spring	Wenatchee River	COmup_Wenatchee	48		0.209	0.202	1.88	0.85	0.85	0.048	0.26
	Cle Elum Hatchery	COmup_CleEHsp	48		0.262	0.255	1.95	0.94	0.96	0.048	0.219
Snake River fall	Lyons Ferry Hatchery	Snake_LyonsFHfl	119	12	0.359	0.36	2	0.45	0.45	-	0.158
Snake River spring/summer	Rapid River Hatchery	Snake_RapRHsumsp	48		0.191	0.194	1.84	0.85	0.94	0.034	0.272
	McCall Hatchery	Snake_MCallHsumsp	48		0.199	0.196	1.84	0.75	0.96	0.034	0.278
Washington Coast	Forks Creek Hatchery	WACoast_ForksCH	93		0.35	0.345	1.99	0.89	0.94	0.042	0.143
	Quinalt Lake fall	WACoast_Quinalt	48		0.348	0.341	1.98	0.9	0.96	0.042	0.152
South Puget Sound	Soos Creek Hatchery	sPuget_SoosCH	142		0.36	0.358	2	0.91	0.91	-	0.158
North Puget Sound	Kendall Hatchery sp.	nPuget_KendlHsp	48		0.326	0.336	1.95	0.92	0.96	0.042	0.17
	Marblemount H. sp.	nPuget_MrblHsp	48		0.343	0.337	1.99	0.92	0.94	0.042	0.156
Lower Fraser River	Harrison River	Fraser_Harris	48		0.329	0.326	1.98	0.96	0.96	0.152	0.169
	Birkenhead Hatchery	Fraser_BirkenH	91		0.259	0.255	1.84	1	1	0.152	0.231
Lower Thompson R.	Spius Creek Hatchery	Thompson_SpiusCH	46	11	0.271	0.275	1.89	1	1	-	0.201
Eastern Vancouver I.	Big Qualicum Hatchery	eVancI_BigQual	48		0.352	0.338	2	0.83	0.83	-	0.145
Western Vancouver I.	Robertson Hatchery	wVancI_RobHfl	48		0.341	0.364	1.98	0.96	0.96	-	0.152

Continued on next page

Table 2.1 – continued from previous page

Reporting Group	Population	Tree Name	n	n_t	Exp. Hz	Obs. Hz	A	Assign. to pop.	Assign. to rep. group	Mean F_{ST} w/in	Mean F_{ST} btw.
Lower Skeena River	Lower Kalum River	lSkeena_Kalum	48		0.303	0.303	1.96	0.77	0.77	-	0.156
Mid Skeena River	Morice River	mSkeena_Morice	47		0.279	0.276	1.89	0.81	0.91	0.013	0.173
	Kitwanga River	mSkeena_Kitwanga	48		0.291	0.29	1.94	0.54	0.75	0.013	0.175
S. Southeast AK	LPW - Unuk R. stock	sSEAK_Unuk	48		0.301	0.29	1.94	0.79	0.79	-	0.165
Gulf AK Alsek R.	Goat Creek	AlsekAK_Goat	48		0.243	0.245	1.69	0.96	0.96	-	0.248
Gulf AK Karluk R.	Karluk River	KarlukAK	47		0.23	0.22	1.73	1	1	-	0.237
Taku River	L. Tatsamenie Lake	Taku_LilTats	48		0.271	0.265	1.92	0.9	0.9	-	0.188
NSE AK Chilkat R.	Pullen Creek Hatchery	nSEAK_PullenCH	48		0.26	0.276	1.77	0.98	0.98	-	0.209
Gulf AK Situk R.	Situk River	SitukAK	48	12	0.244	0.248	1.77	0.94	0.94	-	0.21
Copper River	Sinona Creek	CopperAK_Sinona	47		0.229	0.226	1.63	0.98	0.98	-	0.244
Susitna River	Montana Creek	SusitnaAK_Montana	48		0.21	0.201	1.73	0.92	0.92	-	0.249
Lower Kuskokwim/ Western AK	George River	WestAK_George	47		0.234	0.229	1.78	0.43	0.98	0.004	0.239
	Kanektok River	WestAK_Kanektok	48		0.241	0.232	1.81	0.38	0.96	0.001	0.233
	Togiak River	WestAK_Togiak	48		0.241	0.229	1.79	0.4	0.94	0.005	0.233
Mid Yukon River	Kantishna River	Yukon_Kantishna	48		0.208	0.204	1.67	0.94	0.94	-	0.269
Coho salmon	California Coho	Coho	47		0.089	0.094	1.33	1	1	-	0.463
		total	8031	mean	0.32	0.317	1.93	0.69	0.88	0.028	0.188

2.4.2 Markers and Genotyping

We compiled 192 Taqman© (Life Technologies Corporation, Carlsbad, CA), or 5' nuclease, SNP genotyping assays from previously published discovery efforts (Smith *et al.* 2005a, 2005b; Campbell and Narum 2008; Narum *et al.* 2008; Clemento *et al.* 2011) to test their scorability and power for GSI. Taqman© technology combines standard PCR primers targeting the genomic region around a SNP with two different fluorescent probes that identify the two nucleotide bases present at the SNP. Per the manufacturers recommendation, a multiplex pre-amplification reaction was used to increase the copy number of targeted genomic regions. Multiplex PCR products were diluted with 15μL of 2mM Tris and frozen. Samples were then genotyped on 96.96 Dynamic Genotyping Arrays^{TM} using an EP-1 genotyping system (Fluidigm Corporation, South San Francisco, CA) according to the manufacturers protocols. Fluidigm Dynamic Arrays use integrated fluidic circuitry and PCR volumes of ~9nL to simultaneously determine the genotype at 96 SNP loci for 96 samples (two of which are no-DNA template controls). Genotypes were determined using the Fluidigm SNP Genotyping Analysis Software (version 2.1.1). Genotype determination using quantitative PCR methods involves discerning, on a two dimensional graph, clusters of the fluorescence intensity of the probes for the two alleles; the two homozygote clusters have fluorescence primarily from only one probe, while a heterozygote cluster has similar intensities from both.

2.4.3 Marker Selection

A panel of 95 SNP markers was selected from amongst the 192 candidates, reserving one marker for a species identification assay (see below). The risk of high grading bias (i.e. wrongly inflating the apparent resolving power of a group of loci for GSI) is particularly great when selecting a panel of markers to distinguish between populations that are closely related, as are many in our baseline. To avoid high grading bias, we employed the Training-Holdout-Leave-One-Out (THL) procedure of Anderson (2010), which requires that the data be split into training and holdout sets. Training-set genotypes are used to select the loci included in the baseline and can be included in the eventual baseline, but they are not used to evaluate its performance. Rather, performance of the baseline is determined with simulation and self-assignment using only the holdout set, which was not used in any way to select the baseline loci. We chose a training set of 372 individuals drawn from 22 populations (14 from California, three from Oregon, three from Washington, one from British Columbia and one from Alaska) for initial genotyping with all 192 loci.

For each locus, k, the observed relative frequencies, p_{ik} and q_{ik}, of the two SNP alleles were calculated for each population, i, in the training set. These values were then used to compute the expected probability of misassignment, $P(\mathrm{Mis}_{ijk})$, between every pair of populations i and j using only a single locus i:

$$P(\mathrm{Mis}_{ijk}) = 0.5\ [\ \delta(\mathrm{p}_{ik} \leq \mathrm{p}_{jk})\mathrm{p}_{ik}{}^2 + \delta(\mathrm{p}_{ik}\mathrm{q}_{ik} \leq \mathrm{p}_{jk}\mathrm{q}_{jk})2\mathrm{p}_{ik}\mathrm{q}_{ik} + \delta(\mathrm{q}_{ik} \leq \mathrm{q}_{jk})\mathrm{q}_{ik}{}^2 +$$

$$\delta(\mathrm{p}_{ik} \geq \mathrm{p}_{jk})\mathrm{p}_{jk}{}^2 + \delta(\mathrm{p}_{ik}\mathrm{q}_{ik} \geq \mathrm{p}_{jk}\mathrm{q}_{jk})2\mathrm{p}_{jk}\mathrm{q}_{jk} + \delta(\mathrm{q}_{ik} \geq \mathrm{q}_{jk})\mathrm{q}_{jk}{}^2]$$

for all k, where $\delta(\mathrm{x}) = 1$ if the condition x is true and 0 otherwise. The values of $P(\mathrm{Mis}_{ijk})$ were used to rank the loci for their suitability for resolving between populations i and j in GSI; lower $P(\mathrm{Mis}_{ijk})$ implies better resolving power.

The rankings obtained from $P(\mathrm{Mis}_{ijk})$ were combined with other criteria in a non-automated process to select the final panel of loci (Table 2.2). Each SNP assay was evaluated for scorability and evidence of Hardy-Weinberg (H-W) or linkage disequilibrium. Assays with overly dispersed clusters, more than three clusters, or inadequate spacing between clusters were excluded. Loci with significant deviations from equilibrium expectations were also removed. SNPs with large allele frequency differences between populations are particularly effective for GSI, while SNPs with high minor allele frequencies (MAFs) are most useful for parentage analysis (Anderson and Garza 2006). The remaining 168 loci were then ranked by their MAFs in hatchery populations to be included in pedigree reconstruction studies (see Discussion). Previous simulations indicated that about 100 loci with a MAF greater than 0.2 would be required to achieve the necessary statistical power to assign parentage with sufficiently low false-negative and false-positive rates (Anderson and Garza 2006). However, the observed MAFs for many loci were in fact greater than 0.2 (and as high as 0.5), meaning that the desired statistical power could be achieved with fewer loci. We therefore selected the 70 loci with the highest MAF in the Feather River population, the primary target for subsequent parentage investigations. The $P(\mathrm{Mis}_{ijk})$ rankings were then utilized to select 25 additional loci that were useful for distinguishing between difficult-to-resolve populations and reporting units. Finally, an assay to discriminate between Chinook and coho

salmon was included as the 96th assay for the 96.96 genotyping arrays.

Table 2.2: List of the 96 single nucleotide polymorphism loci used to construct the baseline for genetic stock identification of Chinook salmon from the West Coast of North America, including dbSNP accession numbers (at the NCBI on-line repository for short genetic variations) and source reference (SR) where available: 1. Clemento *et al.* 2011; 2. Smith *et al.* 2005a; 3. Campbell and Narum 2008; 4. Smith *et al.* 2005b.; 5. Narum *et al.* 2008

Locus	**dbSNP**	$\mathbf{S}_R$	**Locus**	**dbSNP**	$\mathbf{S}_R$	**Locus**	**dbSNP**	$\mathbf{S}_R$
Ots_94857-232	ss275518685	1	Ots_110495-380	ss275518741	1	Ots_131906-141	ss275518787	1
Ots_96222-525	ss275518688	1	Ots_110551-64	ss275518742	1	Ots_AldB1-122	ss275518788	1
Ots_96500-180	ss275518689	1	OkiOts_120255-113	unpubl.	-	Ots_AldoB4-183	ss275518789	1
Ots_97077-179	ss275518691	1	Ots_111312-435	ss275518746	1	Ots_Myc-366	ss275518795	1
Ots_99550-204	ss275518695	1	Ots_111666-408	ss275518747	1	Ots_ALDBINT1-SNP1	ss275518796	1
Ots_100884-287	ss275518696	1	Ots_111681-657	ss275518748	1	Ots_NAML12-SNP1	ss275518798	1
Ots_101119-381	ss275518697	1	Ots_112208-722	ss275518749	1	Ots_ARNT-195	unpubl.	-
Ots_101704-143	ss275518699	1	Ots_112301-43	ss275518750	1	Ots_RAG3	n/a	5
Ots_102213-210	ss275518702	1	Ots_112419-131	ss275518751	1	Ots_AsnRS-60	ss48398657	2
Ots_102414-395	ss275518703	1	Ots_112820-284	ss275518752	1	Ots_aspat-196	ss65917744	3
Ots_102420-494	ss275518704	1	Ots_112876-371	ss275518753	1	Ots_CD59-2	unpubl.	-
Ots_102457-132	ss275518705	1	Ots_113242-216	ss275518754	1	Ots_CD63	unpubl.	-
Ots_102801-308	ss275518706	1	Ots_113457-40	ss275518755	1	Ots_EP-529	unpubl.	-
Ots_102867-609	ss275518707	1	Ots_117043-255	ss275518757	1	Ots_GDH-81x	ss65917741	3
Ots_103041-52	ss275518708	1	Ots_117242-136	ss275518759	1	Ots_HSP90B-385	ss65713207	2
Ots_104063-132	ss275518711	1	Ots_117432-409	ss275518762	1	Ots_MHC1	ss49851328	4
Ots_104569-86	ss275518714	1	Ots_118175-479	ss275518763	1	Ots_mybp-85	unpubl.	-
Ots_105105-613	ss275518715	1	Ots_118205-61	ss275518764	1	Ots_myoD-364	ss65917726	3
Ots_105132-200	ss275518716	1	Ots_118938-325	ss275518765	1	Ots_Ots311-101x	ss65917748	3
Ots_105401-325	ss275518718	1	Ots_122414-56	ss275518767	1	Ots_PGK-54	n/a	5
Ots_105407-117	ss275518719	1	Ots_123048-521	ss275518768	1	Ots_Prl2	ss49851322	4
Ots_106499-70	ss275518724	1	Ots_123921-111	ss275518770	1	Ots_RFC2-558	ss48398670	2
Ots_106747-239	ss275518725	1	Ots_124774-477	ss275518771	1	Ots_SClkF2R2-135	ss48398694	2
Ots_107074-284	ss275518726	1	Ots_127236-62	ss275518773	1	Ots_SWS1op-182	ss48398635	2
Ots_107285-93	ss275518728	1	Ots_128302-57	ss275518775	1	Ots_TAPBP	n/a	5

Continued on next page

Table 2.2 – continued from previous page

Locus	**dbSNP**	$\mathbf{S}_R$	**Locus**	**dbSNP**	$\mathbf{S}_R$	**Locus**	**dbSNP**	$\mathbf{S}_R$
Ots_107806-821	ss275518730	1	Ots_128693-461	ss275518777	1	Ots_u07-07.161	unpubl.	-
Ots_108007-208	ss275518731	1	Ots_128757-61	ss275518778	1	Ots_u07-49.290	unpubl.	-
Ots_108390-329	ss275518732	1	Ots_129144-472	ss275518779	1	Ots_u4-92	ss48398636	2
Ots_108735-302	ss275518733	1	Ots_129170-683	ss275518780	1	Ots_BMP2-SNP1	ss275518800	1
Ots_109693-392	ss275518737	1	Ots_129458-451	ss275518782	1	Ots_TF1-SNP1	ss275518802	1
Ots_110064-383	ss275518738	1	Ots_130720-99	ss275518784	1	Ots_S71-336	n/a	5
Ots_110201-363	ss275518739	1	Ots_131460-584	ss275518785	1	Ots_unk_526	n/a	5

2.4.4 Population Genetics Analyses

The 7669 samples that were not in the training set for locus selection were genotyped with the final panel of 96 SNPs and used as the holdout set in subsequent power analyses (see next section). This holdout set was also used for standard population genetics analyses. We tested each locus-population pair for deviations from H-W equilibrium using the complete enumeration method (Louis and Dempster 1987) in GENEPOP vers. 4.0 (Rousset 2008). Similarly, in each population, all pairwise locus combinations were investigated for linkage disequilibrium (LD). Default Markov chain parameters were used, except for the number of batches which was increased to 500 to reduce the standard error to acceptable levels (< 0.02; Rousset 2008).

F_{ST} was estimated (with θ of Weir and Cockerham 1984) between all pairs of populations using the software package GENETIX vers. 4.05 (Belkhir 1996-2004). The dataset was permuted 1000 times to determine the significance of F_{ST} estimates. Phylogeographic trees were constructed with Cavalli-Sforza and Edwards' (1967) chord distance (DCE) and the neighbor-joining algorithm in PHYLIP vers. 3.69 (Felsenstein 2005) and were visualized with DENDROSCOPE (Huson *et al.* 2007). Majority-rule consensus values were calculated from 10,000 bootstrap samples of the data using the PHYLIP component CONSENSE. The F_{ST} values and genetic distances computed should provide an inflated estimate of the isolation between populations because the SNP loci used in the analysis are not a random sample from the genome, as some were chosen for their power in resolving population pairs in our baseline. Nonetheless,

these estimates are useful for assessing the relative genetic differentiation among these populations.

2.4.5 Power Analyses

Three different methods were used to assess the power of the SNP baseline for GSI. First, we performed a self-assignment analysis, and subsequently generated and analyzed simulated mixtures using two different procedures.

In self-assignment, allele frequencies for each potential source population are estimated from the samples. Then, for each individual, the probability of its genotype occurring in each population (assuming H-W and linkage equilibria) is calculated, and the individual is assigned to the population for which its genotype probability is highest. We used the likelihood method of Rannala and Mountain (1997), implemented in the software GSI_SIM (Anderson *et al.* 2008), to compute the genotype probabilities, employing a leave-one-out procedure that excludes the gene copies of the individual being assigned and recalculates population allele frequencies prior to assignment. Analogous to the THL procedure of Anderson (2010), both the training and holdout sets were included for estimating population allele frequencies. However, assignments of the training set individuals were excluded from the results to avoid any high-grading bias of assignment accuracy (Anderson 2010).

Analysis of simulated mixed fisheries is a common method for evaluating the resolving power of a baseline for stock identification (Fournier *et al.* 1984, Wood *et al.* 1987, Kalinowski 2004, Beacham *et al.* 2006). In many studies, samples from simulated

fisheries consisting entirely of fish from one population are analyzed; so called 100% simulations. However, such simulations do not typically assess how well the baseline will perform on samples from fisheries that exploit more than one stock. Therefore, we conducted simulations using 20 different mixing proportion vectors, the population composition of which was constructed by using the baseline to estimate mixing proportions from one of 20 different month-by-area strata from GSI data collected from commercial fisheries off the coast of California and Oregon in 2010 and 2011 (E. Crandall *et al.* unpubl. data). These vectors reflect mixing proportions we expect to encounter in PFMC fisheries. For a given value of the mixing proportion vector of all populations, a replicate simulation consisted of: 1) simulating the number of fish from each population in a sample size of 200 by drawing a multinomial random variable with cell probabilities equal to the mixing proportion vector; 2) simulating the genotypes of the individuals from each population in the mixture sample using two different techniques (cross-validation over gene copies [CV-GC] and K-fold cross-validation [K-Fold], see below); 3) calculating the maximum likelihood estimator (MLE) of the mixture proportions for all the populations from the simulated sample using the baseline, which contains all training and holdout individuals; and 4) estimating the mixing proportion of each reporting unit by summing the mixing proportion estimates of its constituent populations. For each of the 20 values of the mixing proportion vectors, 20,000 replicates were conducted using CV-GC, while 1,000 replicates were conducted using K-Fold. For both methods, the 5% and 95% quantiles of the distribution of the MLE of reporting unit proportions were calculated from the replicates for each mixing proportion vector.

Simulations were undertaken in two different ways. With CV-GC, genotypes were simulated by randomly sampling gene copies from the holdout set (to avoid high-grading bias) and those same gene copies were removed from the baseline when calculating the likelihood of population origin for the simulated individual (see Anderson *et al.* 2008). With K-fold, genotypes were simulated by drawing entire individuals without replacement (jackknife) from the holdout set to form the mixture sample. Those sampled individuals were not included in the baseline, but all unsampled individuals from the holdout set were included in the baseline for estimating the mixing proportions.

2.4.6 Mixed Fishery Samples

Samples from 2,090 salmon landed in fisheries in 2010 were collected by the California Department of Fish and Game (CDFG) at California ports. Just over half of these fish carried coded wire tags (CWTs) that identified their population of origin. All samples were genotyped with our panel of 96 loci. Individuals successfully genotyped at fewer than 60 loci were removed from further analysis. Failed genotypes were ones that either clustered with negative controls during scoring or fell outside of defined heterozygote and homozygote clusters, likely indicating sample contamination (Smith *et al.* 2011, Larson *et al.* 2013). We also used an individual heterozygosity (iHz; the proportion of heterozygous loci for each fish) criterion of iHz > 0.56 to identify and exclude samples potentially contaminated by DNA from other samples. Simulations of contaminated genotypes using observed allele frequencies indicated little overlap in the distribution of iHz for contaminated and uncontaminated samples (data not shown)

and that uncontaminated samples rarely had iHz > 0.56. We used the maximum likelihood framework in GSI_SIM to estimate the mixing proportion of different populations amongst the 2,090 fish, and then used that MLE as the prior for calculating the posterior probability of population of origin for each fish. Posterior probabilities of originating from different reporting units were obtained by summing the population-specific probabilities over all populations in a reporting unit. Individuals were then assigned to the reporting unit with the highest posterior probability.

Since all fish will be assigned to a maximum *a posteriori* (MAP) population regardless of true origin, we employed a simulation method similar to that in Cornuet *et al.* (1999), but modified to account for missing data, to detect fish that might originate from a population that is not in the baseline, or has an otherwise aberrant genotype. Briefly, for each fish from the fishery assigned to a population, the allele frequencies from the MAP population were used to simulate 10,000 genotypes with an identical pattern of missing data (if any) as the fish that was assigned. The log-probability of each simulated genotype was computed, given that it came from the population it was simulated from, and then the distribution of those values was compared to the log-probability, L_a, of the actual assigned fishs genotype, given the allele frequencies in the MAP population, on the basis of a z-score (L_a minus the mean of the simulated values, all divided by the standard deviation of the simulated values). The z-score calculation is done conditional on the exact pattern of missing data and is implemented in the C programming language as part of the GSI_SIM software. A low-confidence assignment was defined to be one that had a z-score $<$ -3.0 and either a reporting unit posterior

probability less than 0.9 or fewer than 90 loci successfully genotyped. Fish with low confidence assignments were left in an unassigned category.

2.5 Results

2.5.1 Genotyping and Basic Population Genetics

We successfully genotyped 8,031 samples from 69 populations for the baseline and submitted the data to the Dryad Digital Repository (http://www.datadryad.org). All individuals were retained in the baseline, regardless of missing data, as we desired a realistic representation of missing data patterns for subsequent power analyses. One locus failed to amplify entirely in the Copper River population, while three loci failed in the coho salmon sample. Unbiased estimates of heterozygosity (Nei 1978) ranged from 0.194 in the Snake River-Rapid River Hatchery stock to 0.381 in the Smith River population. The coho salmon in the baseline had very low heterozygosity (0.094). Observed heterozygosity and mean number of alleles were generally lower for populations from north of the Columbia River (Table 2.1), likely due to ascertainment bias. Significant deviations from HWE ($P < 0.0001$) were observed at various loci in 17 populations, but represented $< 0.3\%$ of all observations. Only the Butte Creek spring-run, Trinity River Hatchery spring-run and Smith River populations were not in HWE at more than two loci, with five, five and four significant tests, respectively. Similarly, only three loci deviated from HWE in more than two populations: Ots_u07_07.161 in three populations, Ots_111312-435 in six and Ots_111666-408 in four. Only one population (Trinity River

Hatchery spring-run) displayed significant LD ($P < 0.001$) at more than 1% of locus comparisons (1.14%) and, over all populations, the percentage of significant comparisons was 0.16%. Only two locus pairs were significant in more than five populations: Ots_AldB1-122 and Ots_AldoB4-183, known to be in same gene complex, were in LD in 42 populations, while Ots_Myc-366 and Ots_unk-526 displayed LD in eight populations.

A large range in the degree of differentiation between populations was observed (Table 2.1). Mean FST across all populations (excluding coho salmon) was 0.183, indicating that approximately 18% of genetic variation was partitioned between population samples. Within reporting units containing more than one population (n = 18), pairwise F_{ST} was between 0.000 and 0.152 with a mean value of 0.018. Ten pairwise comparisons, all within reporting units, were not significantly different from zero ($P < 0.01$). Between reporting units, F_{ST} values ranged from 0.005 to 0.411 with a mean value of 0.188. The least differentiated populations were the fall-run populations from California's Central Valley, as has been observed with other genetic datasets (Williamson and May 2005; Seeb *et al.* 2007).

Genetic structuring of the Chinook salmon populations in the baseline is displayed in an unrooted neighbor-joining dendrogram (Figure 2.1). Relationships are in strong agreement with expectations based on geography and previous studies (Waples *et al.* 2004, Beacham *et al.* 2006, Templin *et al.* 2011, Moran *et al.* 2013); populations are generally organized north to south along the main branch, with populations from within the same drainage usually clustering together. Populations from California's Central Valley are monophyletic relative to the remainder of the populations but are

characterized by short branch lengths, small distances between nodes and low bootstrap support. Central Valley spring-run and fall-run populations are also monophyletic, with the exception of the Feather River Hatchery spring-run, which is included in the fall-run reporting unit due to a history of substantial introgression between the runs and the consequent difficulty of genetically distinguishing them from fall-run fish (Garza *et al.* 2008). Sacramento River winter-run fish are quite distinct due to a well-documented recent bottleneck (Hedrick *et al.* 1995), and have one of the longest branches on the tree, with bootstrap support of 100%. Rivers from Northern California and coastal Oregon also form a monophyletic group. Columbia River populations are dispersed throughout the tree, although populations from the same reporting unit generally share a common branch, as do populations from Alaska.

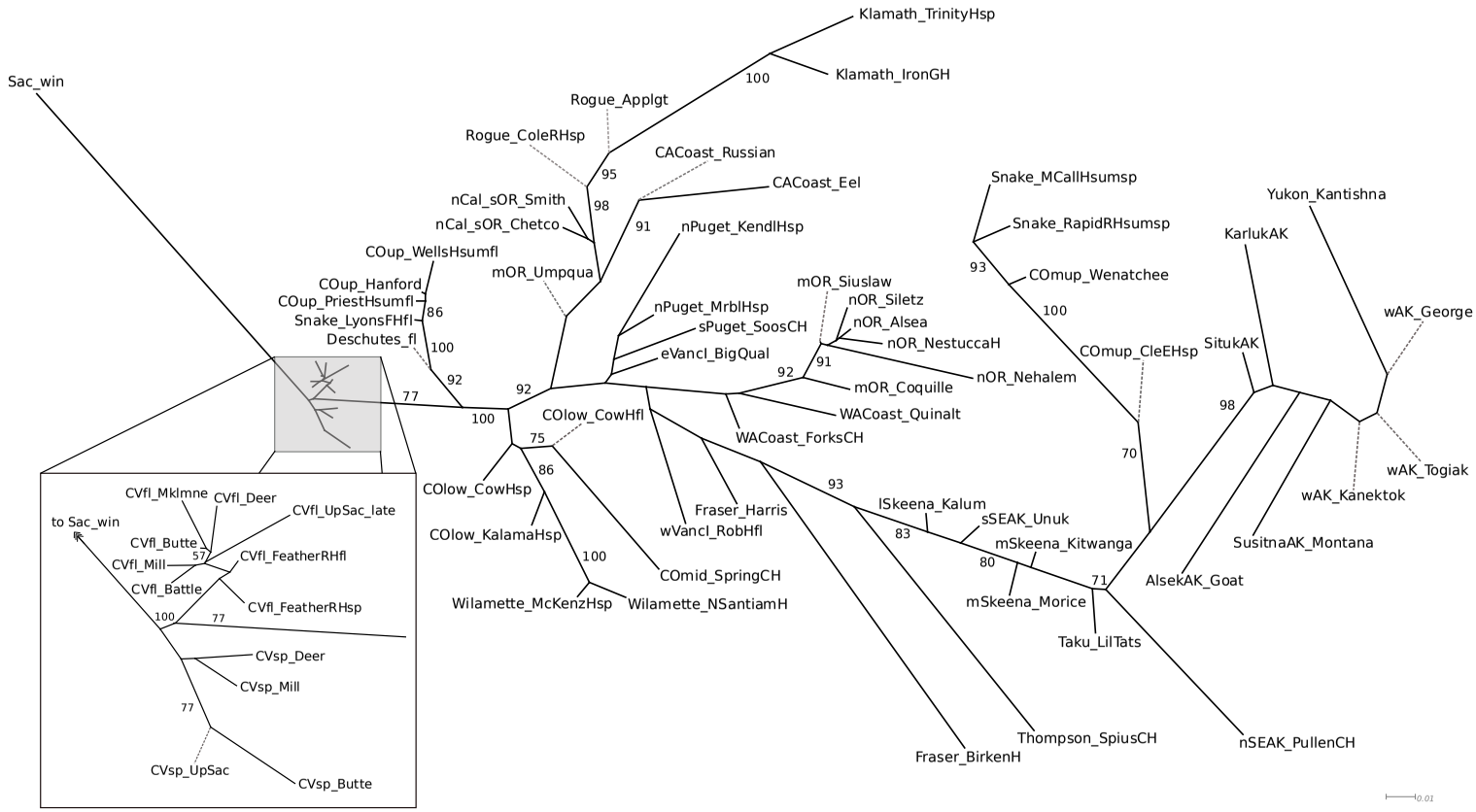

Figure 2.1: Unrooted neighbor-joining tree based on chord distances of 67 Chinook salmon populations from California to Alaska in the GSI baseline (see Table 2.1 for population details). Dashed lines indicate the position of populations which fall at tree junctions or have very short branch lengths. Sinona Creek and the coho salmon were omitted for missing data.

2.5.2 Assignment and Mixture Estimation Accuracy

The 7,669 individuals remaining after removal of training set fish were subjected to self-assignment using GSI_SIM (Table 2.1). Correct assignment to population ranged from 13% for Butte Creek fall-run to 100% for five different populations. The reporting units with the lowest correct assignment rates to population were the Central Valley fall-run, Upper Columbia River summer-/fall-run and Lower Kuskokwim/Western AK, averaging 28%, 36% and 40% respectively. The lowest rate of correct assignment to reporting unit was for the Siuslaw River population from the Mid Oregon Coast, with over half of the individuals assigning to populations in the North Oregon Coast reporting unit. The largest change in correct assignment percentage from population to reporting unit was for the Central Valley fall-run, which increased to 91%.

The results of the mixture simulations for the eight reporting units most frequently found in California and Oregon fisheries appear in Figure 2.2. Results for the remaining reporting units are not shown, as they are relatively uninformative, due to the rarity with which populations from north of the Columbia River are encountered at the southern end of the California Current marine ecosystem—an observation corroborated by historical CWT data: in the three decades since 1983, only 0.5% of all CWTs recovered from Chinook salmon in California ocean fisheries were from stocks outside of California or Oregon (data from www.rmpc.org). Accurate estimates of the mixing proportions were obtained for fishery samples simulated either by CV-GC or by K-Fold. The mean maximum likelihood estimate of the proportion of each reporting unit was

generally highly correlated with the true proportion, indicating that any bias was very small. For five reporting units (Central Valley fall-run, Sacramento River winter-run, Klamath River, California Coast, and Rogue River), the 5% and 95% quantiles for reporting-unit mixing proportions corresponded closely to the quantiles one would obtain with perfect identification of all fish (gray regions in Figure 2.2). The somewhat wider GSI quantile intervals observed for the Central Valley spring-run reporting unit were likely due to its similarity to the Central Valley fall-run reporting unit, combined with the fact that the spring-run is typically at much lower abundance than the fall-run. Likewise, the genetic similarity of Mid Oregon Coast and Northern California/Southern Oregon Coast fish made it difficult to accurately estimate mixing proportions for these reporting units; however, the estimates were still quite good and largely unbiased. Thus, despite the enlarged quantile intervals for Central Valley spring-run and the Mid Oregon versus Northern California reporting units, the results from both simulation methods indicated that the SNP baseline is capable of providing estimates of the true mixing proportions for most reporting units that are nearly as accurate as one would expect given perfect identification of each fish.

2.5.3 Fishery Sample

Of the 2,090 samples from California fisheries in 2010, 85 were excluded because they did not yield acceptable genotypes (< 60 successfully genotyped loci), and two samples were duplicates. Eight fish exceeded the iHz threshold of 0.56 and were removed due to potential contamination. Seven fish were identified as coho salmon

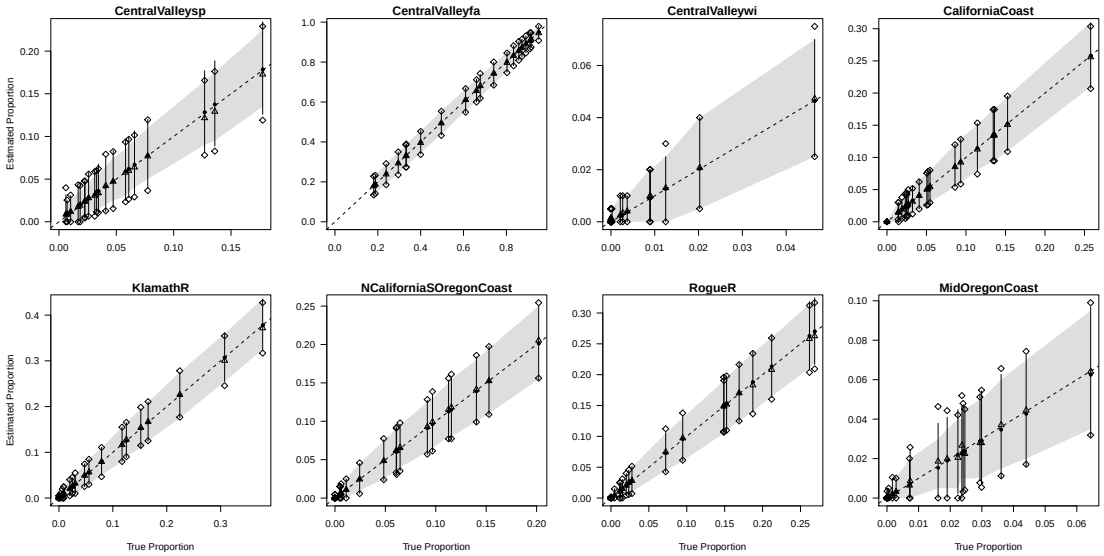

Figure 2.2: Estimates of mixing proportions from cross-validation over gene copies (CV-GC) and K-Fold simulations for the eight most abundant reporting units in California Chinook salmon fisheries. The x-axis gives the true proportion of fish from each reporting unit, and the y-axis gives the estimated proportion. The dashed line is the y=x line. Shaded regions give the range between the 5% and 95% quantiles of estimates that would be achieved with perfect assignment of fish to reporting unit; i.e., they represent the uncertainty due to the fact that fishery proportions are estimated with a finite sample (in our simulations, a sample of 200 fish). The 5% and 95% quantiles of the estimates using genetic data from the CV-GC and the K-Fold methods are shown with vertical line segments and open diamonds, respectively. The mean over 20,000 CV-GC simulation replicates and 1,000 K-Fold replicates are given by filled circles and open triangles, respectively. These points fall along the dotted line when the estimator is unbiased.

Table 2.3: Genetic stock identification (GSI) results of assigning 2010 California Chinook salmon fishery samples to their source populations using the single nucleotide polymorphism baseline, and concordance with coded-wire tag (CWT) recoveries.

Stock	**# from GSI**	**# with CWT**	**# GSI/CWT matches**	**% GSI/CWT agreement**
California Coast	30	1	0	0.00%
Central Valley fall	1581	958	957	99.90%
Central Valley spring	7	1	0	0.00%
Klamath River	108	50	49	98.00%
Lower Columbia River spring/fall	1	0	0	-
Mid Columbia River tule	7	2	2	100.00%
Mid Oregon Coast	14	1	0	0.00%
N. California S. Oregon Coast	58	25	25	100.00%
Rogue River	154	11	5	45.45%
Snake River fall	1	1	1	100.00%
Up. Columbia River summer/fall	8	2	2	100.00%
Total	1969	1052	1041	98.95%

through both GSI assignment and with the species-diagnostic assay. Another 18 samples did not meet assignment confidence criteria (mean z-score of -3.99 and a mean of 75 successfully genotyped loci) and were also excluded. For the remaining 1,969 fish, assignment probabilities to reporting unit ranged from 36.4% to 100% (mean 98.5%) and z-scores ranged from -4.12 to 2.68 (mean -0.04). Central Valley fall-run fish dominated the stock composition, accounting for over 80% of sampled fish, followed by the Rogue River (7.79%), the Klamath River (5.46%) and eight other stocks with less than 5% (Table 3). Of the assigned fish, 1,052 contained coded wire tags that were recovered. Genetic assignment to reporting unit disagreed with CWT origin for only 11 fish (1.05%) and, of these mismatches, six were fish with Klamath or Smith River tags that were assigned to the genetically similar Rogue River reporting unit.

2.6 Discussion

Here we describe one of the first large-scale SNP baselines for genetic stock identification of Chinook salmon and the first designed for use with fisheries in the California Current Large Marine Ecosystem off the West Coast of the coterminous United States. Chinook salmon are an economically and ecologically important species and are a major component of North Pacific Ocean fisheries. We genotyped over 8,000 individual fish from 69 distinct populations at 96 SNP loci to construct the baseline. The reporting units included in the baseline represent the likely sources for over 99% of the fish typically encountered in PFMC fisheries off California and Oregon. Furthermore, mixture analyses and self-assignment indicate that the baseline has near maximum possible power for discriminating Chinook salmon stocks at the reporting unit level. Mixture proportion estimates of Central Valley fall-run, Central Valley winter-run, California Coastal, Klamath River, and Rogue River reporting units (Figure 2.2) are no more variable than estimates that would be obtained if every fish carried an unambiguous reporting-unit tag. Mixing proportion estimates for Central Valley spring-run, Northern California/Southern Oregon, and Mid-Oregon Coast reporting units are somewhat more variable, but still appear to be nearly unbiased. In the ocean fishery sample, assignments of over 1,000 individuals to reporting unit, using our baseline, were highly concordant (98.95%) with the CWTs recovered from the same fish. This SNP baseline therefore represents an important addition to the technologies available to Chinook salmon managers and researchers.

2.6.1 Methodological Considerations

Management of Pacific Ocean salmon fisheries off North America can be roughly divided into three regions: California and Oregon fisheries are managed by the Pacific Fishery Management Council (PFMC); fisheries in Washington, British Columbia, Canada and southeast Alaska are subject to the international Pacific Salmon Treaty, reported to and regulated by the Pacific Salmon Commission (PSC); and fisheries further north and west in Alaska are managed by the state, with salmon by-catch under the purview of the North Pacific Fishery Management Council. The genetic baseline described here was designed primarily to identify fish caught in PFMC ocean fisheries and in ecological investigations in the southern portion of the California Current ecosystem and its associated tributary rivers and streams. We have shown that it performs well in this area but, due to an ascertainment strategy during SNP discovery that included individuals from the Columbia River and British Columbia (Clemento *et al.* 2011), the baseline also has sufficient statistical power to identify the source of some fish from elsewhere in the species' North American range. We observed high rates of self-assignment to reporting unit for all regions represented in the baseline, even though some reporting units are clearly composed of populations with minimal differentiation from each other. Moreover, the utility of our baseline could be effectively extended by simply genotyping the same panel of SNPs on additional populations in those regions, even though heterozygosity and mean number of alleles (Table 2.1), and presumably statistical power, in our baseline is reduced for populations from Canada and Alaska.

Other SNP baselines for Chinook salmon have also been described or are being constructed. Templin *et al.* (2011) describe a 45 SNP locus baseline of populations in the northern and western parts of the species range, designed primarily for GSI of populations from western and southcentral Alaska. This same baseline was also used to probe the seasonal distribution and migration pattern of Chinook salmon in the Bering Sea and North Pacific Ocean (Larson *et al.* 2012). Despite the presence of 14 populations from California, Oregon and Washington in that baseline, the authors appropriately emphasize that resolution of those southern populations is sufficient only for broad-scale assignments. Similarly, Warheit *et al.* (2013) describe the marker selection for eventual development of a SNP baseline for application to PSC fisheries. While the existence of multiple regional baselines is likely to expand, it will still benefit the entire community of fishery managers and scientists to carefully design marker panels with as much overlap as possible. It is conceivable that two or three panels of 96 SNPs could provide the level of resolution needed for identification throughout the species range. Alternatively, as next generation sequencing techniques mature, genotyping-by-sequencing (GBS) approaches might yield data for GSI at lower cost than current genotyping techniques. Such a GBS approach could be used to simultaneously genotype all of the SNPs in each of the regional baselines, allowing mixed-stock analysis throughout the species' range.

Inclusion of the species-diagnostic marker and coho salmon sample in the baseline provided insight into the prevalence of misidentification of coho salmon in ocean fisheries. In the 2010 fishery off California, seven fish sampled as Chinook salmon were found to be coho salmon. Without such methods to identify coho salmon, they are as-

signed by the baseline with erroneously high confidence to a northern, low-heterozygosity Chinook salmon population (data not shown). This problem is characteristic of most statistical methods for performing GSI: if an individuals true population of origin is not included in the baseline, then even if all the populations in the baseline are very poor candidates for the fishs origin, that fish might still be assigned with high posterior probability to one of the populations. This occurs when one population is much more likely than any of the other incorrect populations, even though it is not a likely origin for the individual on an absolute scale. We introduced a simulation-based z-score method, implemented in GSI_SIM, to identify fish that have likely not originated from populations in the baseline. An alternative, Bayesian nonparametric approach to dealing with fish from populations not in the baseline identifies those fish and estimates the allele frequencies in their (unrepresented) source population (Pella and Masuda 2000). That approach is particularly appropriate when large numbers of fish are sampled from each of the populations that are not included in the baseline and when the unrepresented populations are quite divergent from all those in the baseline. We chose the z-score approach over the Bayesian nonparametric approach for three main reasons: 1) it is computationally fast and simple, as there are no convergence problems that might be difficult to detect; 2) our baseline was sufficiently comprehensive for stocks contributing to PFMC fisheries that it is unlikely that large numbers of fish would originate from any single unrepresented population, let alone a highly divergent one; 3) our approach should be more appropriate for identifying fish whose genotypes are aberrant due to genotyping complications or sample contamination. Regardless of which method is used, all GSI

estimation should include some analysis to identify fish that either are from populations not included in the baseline or that have aberrant genotypes for another reason.

GSI is highly dependent on source populations being sufficiently genetically differentiated from one another for discrimination. In situations where hatchery broodstock transfers, supplementation, or other processes increase straying and gene flow between fish populations, differentiation decreases and it can become more difficult to use GSI. Such is the case in the Central Valley of California, where average FST between populations in the fall-run reporting unit is 0.006 and in the spring-run reporting unit is 0.013. In the dendrogram (Figure 2.1), this region is characterized by extremely short branch lengths, small inter-nodal differences and weak bootstrap support. Extensive straying of hatchery salmon due to off-site juvenile releases (California Hatchery Scientific Review Group 2012) and water operations (Fisher 1994) have eliminated historical differentiation between populations of fall-run Chinook salmon (Williamson and May 2005). Introgression between fall-run and spring-run fish at the Feather River Hatchery, and likely elsewhere within the basin, has reduced differentiation between these two phenotypes, with mean F_{ST} of 0.025 between fall-run and naturally spawning spring-run populations.

Sampling of different stocks for baseline construction in the presence of high stray rates is not entirely straightforward, particularly when populations are largely sympatric and not visually distinguishable. For example, there is clearly a single Central Valley fall-run fish sampled as winter-run in our baseline. These types of occurrences are almost inevitable given the high degree of disturbance and hatchery supplementation

over much of the species range. One approach is to move fish with discrepant genotypes from the baseline populations in which they were sampled to the ones to which they assign using GSI (e.g. Banks *et al.* 2000). However, such a procedure can introduce an upward bias in the predicted accuracy of the baseline, if, in fact, the removed fish actually do belong to the populations from which they were sampled, but simply have unlikely genotypes at the genetic markers used for baseline construction. We chose to be conservative in both 1) accepting a slightly lower rate of predicted resolution obtained by not removing mis-categorized fish, and 2) avoiding an upwardly biased predicted GSI accuracy if the fish removed are not mis-categorized.

2.6.2 Implications for Management

Accurately estimating the proportion of fish from different populations in mixed-stock ocean fisheries has important applications for harvest management and conservation. Stocks comingled in ocean fisheries can vary widely in productivity and abundance. Without precise information on their ocean distribution (as can be provided by GSI), managers have few options for protecting depressed or at-risk stocks from fishery impacts other than shutting down or curtailing fisheries over broad areas, as is currently done. For example, in 2008 and 2009, the largest closures on record of fisheries in California and Oregon were enacted to protect the severely reduced Central Valley fall-run stock (Lindley *et al.* 2009). The economic impact of fishery closures is substantial, resulting in millions of dollars of lost income for fishermen, coastal communities and retailers (Michael 2010).

Management of Chinook salmon in California, Oregon, and Washington, and in PSC-managed fisheries depends heavily on information generated by an elaborate CWT program (Hankin *et al.* 2005). Tiny wire tags are mechanically implanted into the heads of juvenile fish, with each tag bearing a code that identifies the release group and source hatchery (or stock) of the fish. Tagging of naturally spawned juvenile fish has generally proven unsuccessful (Beacham *et al.* 1996), so tagged hatchery stocks are used as proxies to estimate fishery impacts for groups of natural stocks. Aside from the largely unvalidated assumption that such proxies accurately reflect fishery impacts on the associated natural stocks (Hankin *et al.* 2005), the physical effects of tagging fish and removing their nerve-rich adipose fin (Buckland-Nicks *et al.* 2012) as an associated external mark can increase disease transmission (Elliott and Pascho 2001), interfere with homing (Morrison and Zajac 1987, Habicht *et al.* 1998) and swimming ability (Reimchen and Temple 2004) and may impact size-at-return for adult salmon (Vander Haegen *et al.* 2005). Moreover, extremely low recovery rates mean that CWT data are often quite limited and there is frequently great uncertainty associated with the resulting estimates derived from them (Hankin *et al.* 2005).

GSI has been advanced as an alternative to CWTs in fishery management for several decades. Our direct comparison of CWT and genetic assignments demonstrates that our baseline is capable of identifying fish to reporting unit with accuracy comparable to CWTs. Furthermore, using GSI, considerably more fish can be identified to reporting unit, including fish from natural stocks. Confident genetic assignments were obtained for ∼94% of fish from the 2010 fishery sample, whereas only 1,052 of those fish

carried coded wire tags and this number is inflated partially due to oversampling of fish believed to carry CWTs.

Fishery management decisions rely heavily upon cohort-based ocean harvest models (cf., O'Farrell *et al.* 2012), which require information on both stock of origin and age of fish impacted in fisheries. Since GSI does not provide fish age, it is not by itself an adequate alternative to CWTs. Nonetheless, new statistical methods capable of integrating GSI, length data, and scale- or otolith-based age data have recently been developed and shown to provide important inference in PFMC fisheries that are not available from CWTs alone (Satterthwaite *et al.* 2013). Moreover, pedigree-based genetic tagging does supply age for salmon (Anderson and Garza 2006, Garza and Anderson 2007). This method, termed parentage-based tagging (PBT), can identify the actual parents of a genotyped individual through parentage analysis if they have been genotyped with the same genetic markers. If the parents date of spawning is known, as it typically is in a hatchery, then the reconstructed pedigrees yield the offsprings precise age and any associated parental spawning information.

Importantly, both PBT and GSI can be undertaken with the same SNP genotypes, and the SNPs used in our GSI baseline are sufficiently powerful for PBT with salmon from California to Washington (Anderson 2012). This interoperability of genotype data enables an integrated program that uses both GSI and PBT simultaneously, providing identification for all fish in a fishery or ecological sample and yielding significantly greater inference than either method alone. For example, GSI cannot distinguish between spring-run and fall-run fish from the Feather River Hatchery in California, but

PBT discriminates them, almost without error, from any mixture. Likewise, though it is difficult to implement PBT in natural populations, the same SNP genotypes used in a PBT analysis permit accurate identification (via GSI) of fish from the naturally spawning, ESA-listed California Coastal Chinook Salmon ESU.

2.7 Conclusions

The advent of high-throughput SNP genotyping has already revolutionized human genetics (Jenkins and Gibson 2001), providing previously unattainable resolution (e.g, Novembre *et al.* 2008), and is poised to do the same for fisheries biology and management. Here, we use a careful and statistically valid power analysis of SNP genotypes from a large number of Chinook salmon populations concentrated at the southern end of their native range to show that SNPs can provide a powerful baseline for genetic stock identification (see also Larson *et al.* 2012) in fisheries and ecological investigation in the California Current and its tributaries in California and Oregon. We predict that these advances in genetic resources and methods will foster fundamental improvements in the way salmon populations are studied, monitored and managed.

Chapter 3

Large-scale genetic tagging experiment in a hatchery population of Chinook salmon (*Oncorhynchus tshawytscha*) allows for pedigree-based inference

3.1 Introduction

Studies on natural selection, behavioral ecology and the population biology of plants and animals often require tracking individuals, groups or populations over a period of time. This is generally achieved by marking or tagging individuals for subsequent recapture or detection. Physical tags have been used to elucidate the migration and dispersal patterns of birds over the last century (Baldwin 1921, Nickell 1968, Greenwood and Harvey 1982, Moore and Dolbeer 1989), however, mark-recapture experiments are

also common in studies of fish (Metcalfe and Arnold 1997, Jones *et al.* 1999), mammals (Hoskinson and Mech 1976, Ormiston 1985, Bethke *et al.* 1996) and even insects (Stern *et al.* 1965, Sumner *et al.* 2007). These tagging experiments have been used for a broad range of applications, which include: investigating behavioral responses to changing conditions, estimating the effects of natural selection and delineating the distribution of populations.

While physical tagging has a long history, the increasing availability of genomic resources has made genetic tagging methods (Palsbøll 1999) a viable alternative for a variety of species. In its simplest form, genetic tagging is analogous to physical tagging, with the 'mark' being the first time a genotype is encountered and a 'recapture' occurring by matching the original genotype to a subsequent sample. The use of genetic information in lieu of traditional tags to identify individuals has been demonstrated in taxa as diverse as whales (Palsbøll 1997), bears (Woods *et al.* 1999), and martens (Mowat and Paetkau, 2002). Genetic information can also be used to assign membership of individuals to their most likely population of origin. This method, termed genetic stock identification (GSI) in fisheries, and called population assignment in the field of molecular ecology, requires collecting baseline allele-frequency data from potential source populations and then uses maximum likelihood (Smouse *et al.* 1990) or Bayesian methods (Pella and Masuda 2000) to determine the probability that the sample originated from each population; sample and baseline genotypes are collected for the same set of genetic markers. GSI has been successfully applied in studies of highly-structured salmon populations for almost three decades, but is limited if groups are not sufficiently

differentiated (Beacham *et al.* 1985, Teel *et al.* 1999, Beacham *et al.* 2006, Seeb *et al.* 2011, Clemento *et al. in review*).

Yet another way that genetic data can be used as a tagging methodology is in the inference of relationships between individuals, primarily first order relatives such as parent-offspring or siblings. Parentage analysis has been used to address a diverse range of ecological questions, including dispersal, hybridization, fitness, relatedness and estimation of population size (DeWoody 2005). In an early genetic mark-recapture experiment in turtles, parentage was used to reconstruct and subsequently recapture a paternal genotype that was not directly observed (Pearse 2001). Many methods are also available to reconstruct sibships between individuals without parental information (e.g. Wang 2004), as well as to identify parents and offspring in the wild (Jones *et al.* 2009). Assignment of parentage with molecular markers generally utilizes Mendelian incompatibilities between offspring and putative parents to exclude unlikely trios, since a true offspring must carry one of the maternal and one of the paternal alleles. There is a variety of software available for actually using genetic data to infer parentage, however, many are limited in their computational capacity and ability to handle large and complex datasets in a reasonable amount of time (Jones *et al.* 2009). However, new algorithms have now been developed to perform truly large-scale parentage inference, allowing for practical extension of these genetic tagging methods to high fecundity organisms like salmon (Anderson and Garza, 2006, Anderson, 2012). Additionally, recent development of large numbers of single nucleotide polymorphism (SNP) markers (Clemento *et al.* 2011, Abadía-Cardoso *et al.* 2011), which are amenable to efficient high-throughput

genotyping, now allow for the practical analysis of the large number of individuals in salmonid populations (Abadía-Cardoso *et al.* 2013, Steele *et al.* 2013).

Chinook salmon (*Oncorhynchus tshawytscha*) are a highly valued species of Pacific salmonid and are the target of large commercial and recreational fisheries. Chinook salmon are anadromous, wherein adult fish migrate from the ocean to spawn in their natal river, and must therefore contend with impacts in both freshwater and marine environments. Over the last century, many Chinook salmon populations have been reduced or even extirpated by the construction of large dams, extensive water extraction for agriculture and human consumption, overfishing and variable ocean conditions (Myers *et al.* 1998). This has resulted in listings under the Endangered Species Act (ESA; FedReg 1990, 1999), particularly in the southern portion of the species range (e.g. California, Oregon and Washington). In order to mitigate for the multiple impacts threatening Chinook salmon populations, state and federal agencies now produce millions of fish annually in hatcheries. These hatchery fish – primarily intended to reduce variability in ocean abundance and provide fishing opportunities – comingle with wild fish in the ocean and can compose the majority of the catch in certain times and places.

Ensuring sustainability and the persistence of salmon populations while providing fishing opportunities can be a complex task. Overestimation of the contribution from specific stocks can have serious conservation implications, while underestimation can leave the resource underexploited, both potentially costing the fishing industry and coastal communities millions of dollars (Michael, 2010). Generally, management of Pa-

cific salmon ocean harvest in the coterminous United States falls under the purview of the Pacific Salmon Commission (PSC) and Pacific Fishery Management Council (PFMC), while NOAA Fisheries is responsible for controlling harvest of threatened and endangered population segments under the ESA. These entities employ a variety of methods to set fishing areas and seasons, determine quotas and legal gear and establish catch limits and size restrictions. Stock-specific forecasting models are used to estimate ocean abundance indices, which are then used to set harvest limits, first at the international level (PSC) then for local ocean fisheries and finally for terminal fisheries in rivers (Hyun, 2012). The accuracy of these cohort-based models and the resulting abundance forecasts are highly dependent on the quantity and quality of data; estimates of age and stock specific mortality rates and their distribution in the fishery catch, are critical inputs to these models. Currently, the primary source of information for fishery management comes from coded wire tagging of a limited number of hatchery stocks.

The need to identify stock-specific fishery impacts led, in the 1950s, to clipping of particular fins (adipose, anal, maxillary), in an attempt to identify production from different hatcheries or regions. By the 1970s, managing agencies began to use cohort information in fishery management models and turned to the use of coded wire tags (CWTs) in juvenile fish to indicate stock and cohort of origin (Jefferts *et al.* 1963). CWT data has been used to estimate "exploitation rates by age, maturation rates, adult equivalents, marine survival rates, total mortality" and even to infer exploitation patterns of untagged natural stocks (Morishima 2004). CWTs are small pieces of metal (0.5 - 1mm long) mechanically implanted into the heads of juvenile fish. Each tag bears a

group-specific code that identifies the release cohort and source hatchery (or stock). Tag recovery is accomplished through identification of fish carrying a tag (usually removal of the adipose fin), followed by removal of its head and shipment to a laboratory, where the tag is manually extracted and read under a microscope. "Harvest from a cohort is [then] estimated by expanding the number of CWTs recovered according to the fraction of the catch sampled, the fraction of the cohort carrying CWTs, the fraction of heads from recaptured fish that reach a laboratory, and the fraction of dissected heads from which a CWT is decoded (Bernard and Clark 1996)." However, due to limited tag recoveries (often less than 1%) and assumptions about the equivalence of tagged and untagged fish, there is frequently great uncertainty associated with the output of management models (Hankin *et al.* 2005).

Prior to 1996, only fish with CWTs were given adipose fin clips, but another major challenge to the continuing use of CWTs are recent state and federal regulations, which require adipose fin clips on a majority of hatchery production (Hankin *et al.* 2005). This will increasingly result in large numbers of adipose fin-clipped but untagged salmon and has already "decreased the effectiveness of the current program, added costs without gaining information, increased the numbers of fish that samplers handle and mutilate[,] and decreased the value [of] these fish to retailers (Alexandersdottir *et al.* 2004)." This problem has necessitated the use of secondary, electronic tag detection methods at considerable increased cost and effort to the entire program. CWTs are also subject to loss at uncertain rates, which effectively increases the number of clipped but untagged fish (Johnson 2004). Moreover, the tagging and marking process may

cause subtle injuries to juvenile fish that can affect performance and survival at later life stages (Morrison and Zajac 1987, Habicht *et al.* 1998, Reimchen and Temple 2004).

Given the declining effectiveness of the current CWT program, the PSC has recommended validation of alternative tagging strategies (Hankin *et al.* 2005). One of the most promising technologies, and the one evaluated here, is parentage-based tagging (PBT; Garza and Anderson 2007). Utilizing a novel statistical genetic framework for large-scale parentage analysis, genotypes collected from parental breeding generations in hatcheries are used to tag the offspring cohort. Subsequent non-lethal sampling of fish during their seaward migration, in fisheries, or upon return to spawn (either at hatcheries or instream) is followed by high-confidence parentage assignment (Anderson 2012), allowing accurate pedigree reconstruction, and identifying stock and cohort of origin in the process. Since a pair of Chinook salmon can produce thousands of offspring, the tagging of juveniles through genotyping of parents is highly efficient. This methodology generates the same information as the current coded wire tag (CWT) program, which currently provides the bulk of the cohort-specific fishery mortality data for salmon in the northeast Pacific. The ability to accurately identify offspring of spawning fish through parentage analyses means that a pair of parental genotypes translates into many genetic tags in the next generation and has broad potential application for population assessment of fish and other high fecundity species.

Described here is a large-scale, intergenerational genetic tagging experiment with a hatchery population of Chinook salmon from the Feather River, CA, USA. I first examine whether the same panel of SNP markers, successfully used to construct

a coastwide baseline for GSI (Chapter 2; Clemento *et al. in review*), can also be used to confidently reconstruct pedigrees of individuals that have undertaken an ocean migration. The accuracy of assignments is determined by comparing them with recorded cross information in order to evaluate whether genetic tagging data is comparable with that derived from the physical tags currently deployed in the system. Reconstructed parent-offspring trios are used to assess interannual variability in the age structure of offspring cohorts as well as the age structure and relative reproductive success (i.e variation in family size) of spawning broodstock. Data on the physical characteristics of parents and offspring allow for estimates of the heritability of length at maturity and correlations between female body size and the number of her offspring returning to spawn. Inbreeding and relatedness in spawning populations is assessed and the effects of parental relatedness on their reproductive success is evaluated. This research also provides the first evidence that PBT can identify parentage of offspring in large mixed-fishery samples. I demonstrate that parentage-based genetic tagging provides not only a powerful and efficient means of tagging large numbers of individuals, but also generates novel population information that can be used to inform hatchery and fishery management.

3.2 Methods

3.2.1 Study Site

The Feather River is one of the largest tributaries to the Sacramento River in the northern part of California's Central Valley. Historically, the Feather River supported runs of both the fall-run and spring-run ecotypes of Chinook salmon. The spring-run phenotype is characterized by adults that are sexually-immature when they migrate upstream during the Spring. These fish hold in deep pools throughout the summer and then mature and spawn and die throughout the fall and early winter months. In the fall-run phenotype, sexual maturation is coincident with upstream migration and spawning during the fall months. Prior to human modification of the watershed, spring-run fish spawned in the upstream reaches of the Feather River, spatially separated from the fall-run fish spawning further downstream (Department of Water Resources 2004). In 1968, however, construction of Oroville Dam, a principal feature of the California State Water Project, was completed on the mainstem Feather River. This dam blocks upstream passage of spring-run Chinook salmon (Fry and Petrovich 1970), confining them to spawn in the same downstream reaches where fall-run Chinook also spawn. As a consequence, introgression between the two types has been widespread on the currently available spawning grounds (Yoshiyama 1996, Williamson and May 2005).

Additionally, propagation of Chinook salmon at the Feather River hatchery has contributed to introgression between the two run types. For the first three decades of operation at the hatchery, little was done to distinguish or isolate the two run types:

mature fish were simply spawned as they arrived at the hatchery. Fish that entered the hatchery in September were considered to be "spring-run" and were spawned together, while those that entered in October were spawned together as "fall-run" (Department of Water Resources 2004). This practice did little to maintain the reproductive isolation of the two runs because both spring run and fall run fish typically mature between the months of October and December. Since 2003, the California Department of Fish and Wildlife (CDFW), who operates the Feather River Hatchery (FRH; Oroville, CA), has made a concerted effort to limit the amount of introgression between the two runs. Specifically, they devised a plan that is meant to exclude potentially fall-run fish from breeding with fish displaying the spring-run phenotype. During May and June at the hatchery, early-arriving, sexually immature fish are marked with an externally visible tag and released back into the river. Fish that arrive after July 1 are not admitted to the hatchery and so remain in the river untagged and are ultimately assumed to be fall-run. Fish are again allowed to swim up the ladder into the hatchery in late September/early October where they are sorted based on the presence or absence of the external tag which identifies individuals that expressed the spring-run phenotype. Tagged, early-arriving females are then mated one-to-one with spring-run males, the eggs incubated in daily lots, and the fish subsequently reared to the fry life stage and released in various locations in the drainage.

3.2.2 Hatchery Sampling

Caudal fin clips were collected from all returning fish (spawned and unspawned) by CDFW personnel at the FRH and dried on blotter paper. Comprehensive sampling and genotyping of the spring-run Chinook broodstock took place for the six years from 2006 to 2011 (Spring-run/Spring-origin in Table 3.1), while the fall-run broodstock (Fall-run/Non-spring-origin in Table 3.1) was also genotyped in 2008. Coded-wire tag data was used to retrieve samples from spring-run offspring that were collected as fall-run spawners in 2009, 2010 and 2011 (Fall-run/Spring-origin in Table 3.1), for assignment to spring-run parents. A small subset of 2012 spring-run fish, whose offspring were to be used to reintroduce Chinook salmon to the San Joaquin River, were also analyzed for parentage. Metadata, including gender, spawn date, fork length (mm) and spawning partner (spring-run only, 2006-2009) was recorded for each fish. In total, samples from 12,817 Feather River Hatchery Chinook salmon were collected and genotyped (Table 3.1).

Table 3.1: Summary of sampling and genotyping effort and success at the Feather River Hatchery, Oroville, CA. Included are the year of spawning, the hatchery designation of the run (Spring or Fall), the number of individuals genotyped (males and females), the number of individuals excluded for missing genotypes at more than 10 loci, and the number of individuals spawned, as reported by the hatchery. Population genetic parameters of unbiased heterozygosity (Hz), observed heterozygosity (Ho), the inbreeding coefficient (F_{IS}; values with an asterisk are significantly different from zero, $p<0.05$, 1000 permutations) and the mean individual relatedness coefficient (Rxy; see text) were calculated for each genotyped broodstock year and spawn run. Spring origin fish among Fall-run spawners were identified using coded-wire tag data and included as putative offspring during parentage analysis.

			Genotyped		**Excluded**		**Matings**					
Spawn Year	**Spawn Run**	**Run Origin**	**n** [♂]	**n** [♀]	**n** [♂]	**n** [♀]	**n** [♂]	**n** [♀]	**Hz**	**Ho**	**F_{IS}**	**Mean indiv. Rxy**
2006	Spring	n/a	593	553	144	47	590	590	0.375	0.377	-0.0057	-0.0066
	Fall	n/a	-	-	-	-	3390	3431	-	-	-	-
2007	Spring	n/a	731	692	82	86	701	701	0.374	0.373	0.0031	-0.0046
	Fall	n/a	-	-	-	-	1432	2233	-	-	-	-
2008	Spring	Spring	711	718	268	200	387	390	0.370	0.368	0.0044	0.0064
	Fall	non-Spring	1572	1716	297	154	1463	1680	0.370	0.367	0.0082*	0.0074
	Fall	Spring	-	-	-	-	-	-	-	-	-	-
2009	Spring	Spring	357	464	12	3	399	480	0.372	0.369	0.0077*	0.0005
	Fall	non-Spring	-	-	-	-	2838	2946	-	-	-	-
	Fall	Spring	268	359	31	17	-	-	-	-	-	-
2010	Spring	Spring	711	615	73	50	611	611	0.373	0.374	-0.0038	-0.0004
	Fall	non-Spring	-	-	-	-	7628	4961	-	-	-	-
	Fall	Spring	388	344	53	54	-	-	-	-	-	-
2011	Spring	Spring	551	577	9	18	529	529	0.371	0.375	-0.0116	0.0031
	Fall	non-Spring	-	-	-	-	6141	6887	-	-	-	-
	Fall	Spring	327	388	75	92	-	-	-	-	-	-
2012	Spring	Spring	90	92	0	1	90	90	-	-	-	-
		Total	12817		1766			**mean**	0.372	0.372	0.0003	0.0008

3.2.3 DNA Extraction and Genotyping

Collected tissue was sub-sampled and furnished to the SWFSC Santa Cruz Lab for analysis. DNA was extracted from dried tissue using Qiagen DNEasy 96 kits on a BioRobot 3000 (Qiagen, Inc.) according to the manufacturer's recommended protocols. All individuals were then genotyped at the 96 SNP loci described in Chapter 2 (Table 2.2. A multiplex pre-amplification reaction was used to increase copy number of targeted genomic regions. Unlabeled primers (no fluorescent probes) for the panel of 96 loci were combined and diluted to 50nM; the 5uL multiplex PCR contained 1.25uL of this pooled assay mix, 1.25uL of extracted DNA and 2.5uL of 2X Multiplex Master mix (Qiagen). The pre-amp thermal cycling routine consisted of 95°C for 15 min followed by fourteen cycles of 95°C for 15 seconds and 60°C for 4 minutes and a final hold at 10°C. Multiplex PCR product was diluted with 15uL of 2mM Tris and frozen. Samples were then genotyped on 96.96 Dynamic arrays (Fluidigm Corporation) using a Fluidigm EP1 according to manufacturer's protocols. Genotypes were called and the data collected using the Fluidigm SNP Genotyping Analysis software (vers. 2.1.1). Individuals with missing data at 10 or more loci were excluded from further analyses.

3.2.4 Population Genetic Analyses

Observed (Ho) and unbiased expected heterozygosity (Hz; Nei 1987) were calculated for each brood year using the Microsatellite Toolkit (Park 2001). The inbreeding coefficient (Fis), a measure of increased homozygosity due to inbreeding, was calculated for each brood year using the software package Genetix (Belkhir 2004) and significance

assessed with 1000 permutations of the dataset.

3.2.5 Pedigree Reconstruction

Upon release as juveniles (yearlings) back into the Feather River, Chinook salmon from the FRH migrate to the ocean and then return to spawn at age two, three and four. As such, the spring-run broodstock from 2006, 2007, 2008 (including fall-run), 2009 and 2010 was used as the potential parents of fish returning to spawn in 2008, 2009, 2010, 2011 and 2012. The software package SNPPIT (Anderson, 2012) was employed to perform parentage assignments. SNNPIT is a powerful and efficient tool for assigning parentage, which proceeds in two steps. First, the software assembles all possible pairs of parents and uses Mendelian exclusion to exclude pairs that cannot be the parents of the individual to be assigned. Each offspring is then assigned to the most likely parent pair from amongst those with few enough Mendelian incompatibilities. The software then employs Monte Carlo simulation with a novel importance sampling algorithm to calculate a p-value and associated false discovery rate (FDR) for each parentage assignment. Genotyping error rate was assumed to be 0.005 per gene-copy (1% per locus) for a majority of the loci used. Using observed Mendelian incompatibilities in reconstructed trios, however, genotyping error rates were estimated directly for four loci, adjusting the value for Ots_AldB1-122 to 0.0094, Ots_105401-325 to 0.0265, Ots_112208-722 to 0.027 and Ots_101704-143 to 0.011.

Each brood year from 2008-2012 was assigned parentage separately, however, all previous years were included as potential parental sources. This provided a test for

false-positive assignments, as the life-history of these animals (returning as two-year olds at the youngest) should preclude assignments to the year directly preceding the brood being assigned. No parentage assignments are expected to indicate that a returning spawner is only one year old.

Two independent SNPPIT runs were conducted for each group of offspring being assigned to parents. In the first run, metadata on spawn date and sex of the parents was ignored, such that all possible pairs of individuals within a year were considered possible parent-pairs, even if it was reported that they were spawned on different days or were the same sex. The second run limited possible parent pairs to only males and females spawned on the same day. Comparison of the two runs identified some minor metadata errors, and additionally verified correct assignments. Assignments for individuals with an FDR > 0.01 were conservatively excluded, meaning that no more than one assignment in a hundred is expected to be incorrect by chance alone. Parentage assignments were compared to recorded crosses for the spring-run years 2006-2009.

3.2.6 Age Structure, Reproductive Success and Length-at-spawning

The age of returning adults was determined for the 2008-2011 spring-run broodstock and the small sample from 2012 and ranged from two to four. Offspring from fish spawned in 2006 (hereafter, the 2006 cohort) could be identified when they returned at age two, three, four and five in 2008, 2009, 2010 and 2011, respectively; fish from the 2007 cohort could be identified returning at age two, three, four and five in 2009, 2010, 2011 and 2012, respectively; fish from the 2008 cohort could be identified

returning at age two, three and four in 2010, 2011 and 2012 respectively; and fish from the 2009 cohort could be identified at age two in 2011 and age three in 2012. The proportion of fish returning at age two, three and four years old from the 2006 and 2007 cohorts was compared using z-tests. Note that the 2012 sample is only a small subset of the total 2012 spring broodstock, and therefore the 2008 cohort was excluded from this analysis as one would expect four year old fish to be under-represented.

Parentage assignments were also used to examine the age structure of the spring-run spawners in 2010 and 2011. This first required removing the individuals that were included in these groups from the fall-run spawn groups as potential offspring of 2006, 2007, 2008 and 2009 parents (fall-run/spring-origin). For the remaining individuals (those actually spawned as spring-run in 2010 and 2011), the number of individuals that were age two, three and four were identified and z-tests used to compare the relative proportions of the three age classes between years.

The distribution of family size was examined using the inferred parent-offspring trios for fish spawning in 2006, 2007, 2008 and 2009. This analysis included only those parents with at least one offspring detected via pedigree reconstruction. Reproductive success was estimated by counting the number of offspring per parent that returned to the hatchery in any year. As the number of offspring per parent pair was not normally distributed, a Kruskal-Wallis test was used to detect differences between the 2006 and 2007 cohorts for which all age classes (two, three and four year olds) were likely observed. While the hatchery reports primarily one-to-one matings (only one male and one female per cross), the relative reproductive success of males versus females versus pairs was

examined across years, again including only those families with one or more offspring. Additionally, parentage reconstructions were used to discern the structure of sibships in the 2012 sample to be used as broodstock for reintroduction to the San Joaquin River.

Because the hatchery records the length of each fish, and parent-offspring relationships were identified by parentage analysis, the heritability of length-at-spawning was investigated for the dominant three-year old age class. The slope of the parent-offspring regression line was used to estimate heritability (h^2). The mean length of each parent pair was compared to all of their offspring and to male and female offspring separately. The relative contribution of fathers' and mothers' lengths to offspring length was also analyzed separately, specifically looking at the contribution of fathers to sons and mothers to daughters. Since larger females also generally produce a larger number of eggs, the length of each mother was compared to her reproductive success and the regression again fit with a linear model.

3.2.7 Relatedness

For each collection of spring-run spawners (2006-2011) and the single collection of fall-run spawners (2008), the relatedness coefficient (Rxy of Queller and Goodnight 1989) was calculated between all pairs of individuals in each collection using the software KINGROUP (Konovalov *et al.* 2004). Rxy provides a measure of the probability that the shared alleles between two individuals are identical by descent (IBD); higher values of Rxy suggest an increased degree of relatedness with a maximum value of 1 indicating identical genotypes. For each sample, a histogram of Rxy values was plotted, and the

mean, standard deviation and skew calculated. The distribution of Rxy values was also compared to a normal distribution with same mean and standard deviation as that observed in the sample. Since the hatchery kept records of the matings from 2006-2009, the distribution of Rxy values among pairs that achieved reproductive success (defined as those that had at least one offspring return to the hatchery in a subsequent year) could be compared to those that did not. Again, the mean, standard deviation and skew were calculated and the distribution of Rxy values for successful versus unsuccessful matings plotted. As relatedness data appeared to be normally distributed, a two-sided t-test was used to examine whether the mean of Rxy values were significantly different for successful versus unsuccessful parent pairs, for each year and over all years. Finally, the correlation between the size of each full-sib family and the degree of relatedness (value of Rxy) between the parents was investigated with a simple linear regression.

3.2.8 Fishery Samples

In 2010, CDFG collected samples at California ports from 2090 salmon landed in commercial and recreational fisheries. About half of these fish carried CWTs that identified their population of origin and age. Using SNPPIT, the FRH broodstock collections were searched for parents of these port-sampled fish and compared to the genetics-based recaptures to the CWT data. Again, assignments with FDR values $>$ 0.01 were excluded as low confidence. Samples were also collected from the commercial salmon fleet in 2010, 2011 and 2012, primarily for analysis with a genetic stock identification (GSI) baseline (Clemento *et al. in review*), however these collections also contain

offspring from the FRH hatchery. CWT data were unavailable for the GSI collections. In total, DNA was extracted from 24,242 fishery samples, genotyped with our panel of 96 SNP loci and examined for parentage among the sampled FRH brood stock samples. The age structure of FRH fish in the four ocean samples was determined using the parentage reconstructions.

3.3 Results

A total of 12,817 Chinook salmon collected at the Feather River Hatchery from 2006 to 2012 were genotyped with our panel of 96 SNP loci (Table 3.1). Genotypes from 1766 samples were excluded due to missing data (>10 missing loci), leaving 11,051 samples for further analysis. These analyzed samples fell into three categories: spring-run spawners (sample sizes ranged from 181 in the partial 2012 sample to 1255 in 2007 with a mean of 923); fall-run spawners (2008 only, with a sample size of 2837); and fall-run spawners of spring-run origin as determined by CWT data from 2009 to 2011 (with an average sample size of 584 per year). The last category of individuals (fall-run spawn/spring origin) were included as potential offspring, but were excluded from the parent broodstock sample for pedigree reconstruction and calculation of population genetic statistics.

3.3.1 Population Genetic Parameters

Estimates of unbiased heterozygosity ranged from 0.370 in the 2008 samples to 0.375 in the 2006 sample and averaged 0.372, while observed heterozygosity ranged

from 0.367 in the 2008 fall-run collection to 0.377 in the 2006 collection with a mean of 0.372 (Table 3.1). The inbreeding coefficient, Fis, ranged from -0.0116 in 2011 to 0.0082 in the 2008 fall-run sample; values for both the fall-run sample and the 2009 sample were significantly different from zero ($P < 0.05$). The overall degree of relatedness was estimated by first calculating the mean value of Rxy between each individual and all other individuals and then taking the mean of these individual values in each collection. This mean individual relatedness ranged from -0.0066 in 2006 to 0.0074 in the 2008 fall-run sample and averaged 0.0008 over all collections.

3.3.2 Hatchery Pedigree Reconstruction

Two independent pedigree reconstructions were performed: in the first, assembly of the possible parent pairs was not limited to individuals of the opposite sex or with the same reported spawn date, while in the second these factors were used to limit the space of possible parent pairs. In the analysis unconstrained by gender or spawn date, a total of 2791 parent-offspring trios were identified. Fifteen of these trios were not present in the pedigree reconstruction limited by gender and spawn date; three assignments were to parents with different spawn dates while the remaining assignments identified two parents of the same gender. These assignments were of high confidence, with low FDR scores (mean, 0.0028) and high maximum posterior probabilities (mean, 0.99), indicating that they are likely correct trios with errant metadata. Additionally, for three of the unique parent pairs in this group multiple offspring were assigned, further supporting the idea that they were true parental pairs, however, without additional

information to resolve the apparent discrepancies, these 15 assignments were excluded from further analyses.

The remaining 2776 parent-offspring pairs had FDR values ranging from 0 to 0.0098 (mean, 8.81×10^{-5}), with p-values ranging from 0 to 0.04 (mean, 5×10^{-4}) and posterior probabilities of the parent/offspring trio relationship ranging from 0.5418 to 0.9999 (mean, 0.9934). An FDR of 0.0098 can be interpreted as an expectation of 27 misassignments (0.98% of the 2776), although only 13 assignments exceeded an FDR of 0.002 (at which only 6 assignments are expected to be incorrect). In neither the constrained nor the unconstrained pedigree reconstruction were any offspring assigned to parents from the immediately preceding year, suggesting a low false positive rate. The assignment of offspring to the correct parent pairs was also confirmed by the hatchery recordings of the mated individuals. Parentage assignments recovered 1203 correct parental pairings from among the 1874 recorded at the hatchery (64.2%). Additionally, 354 of the recorded pairs not identified by parentage analysis were from 2009, for which there was only a limited sample of 2012 individuals that would comprise the dominant three-year old age class (Table 3.2).

Table 3.2: Summary of offspring (offs.) recoveries using PBT from four spawn years of Chinook salmon from the Feather River Hatchery, CA. Reported are the number (n) of males and females with recovered offspring, the proportion of parent pairs included in the parent database (see text), the number of mated pairs recorded at the hatchery and the number of those matings confirmed by parentage analysis. A scaled estimate of the number of expected recoveries had all parents been included in the database is also shown.

Spawn Year	**Males with Offs.**	**Females with Offs.**	**% Parent Pairs Included**	**Recorded Mates [hatchery]**	**Confirmed Mates [PBT]**	**Offs. recovered [# analyzed] in year**					**Total [8851]**
						2008 [3808]	**2009 [1385]**	**2010 [1828]**	**2011 [1649]**	**2012 [181]**	
2006	233	234	0.660	237	155	3	503	16	0	0	522
2007	459	459	0.783	608	520	0	47	1240	213	0	1500
2008	261	261	0.665	658	510	0	0	26	562	4	592
2009	126	127	0.964	372	18	0	0	0	34	128	162
sum	1079	1081	**sum**	1874	1203	3	550	1282	809	132	2776
					scaled	5	822	1647	1153	139	3766

The 2776 offspring assigned parentage accounted for 31.4% of the potential 8851 offspring sampled at the hatchery from 2008 to 2012. However, 3800 of the unassigned offspring were from the large, fall- and spring-run sample in 2008, for which only the small two-year old age class could be identified (as offspring of 2006 spawners). Excluding the 2008 offspring and evaluating only the years for which the parents of putative three-year olds were available (2009-2012), parentage was assigned for 55% of offspring. It must also be considered that a substantial number of parent pairs were not available for parentage assignment because they were excluded prior to analysis for excessive missing data (>10 loci; Table 3.1). For each day of spawning in each year, the number of genotyped parent pairs (all males x all females) was calculated and subtracted from the number of excluded parent pairs (excluded males x excluded females), weighted by the proportion of females spawned on that day. Summed over the spawn year, this provided an estimate of the percentage of parent pairs included in the parent database for analysis. As offspring are not assigned to single parents here, each parent pair excluded for missing data was a missed opportunity to assign parentage to an offspring. As the most likely source of missed assignments, the proportion of parent pairs in the database was used to scale observed offspring recoveries (Table 3.2). For example, had all parent pairs been included in the parent database for assignment of the 2009 offspring, one could expect to have recovered parentage assignments for 822 individuals. Using the scaled estimates of offspring recoveries from 2009 to 2012, the analysis is expected to have assigned parentage for an additional 985 fish, or a total of 3761 fish, which would be 74.58% of the 2009-2012 offspring available for recovery.

3.3.3 Age Structure

Using the reconsructed pedigrees, the age at which fish return to spawn was assessed for the 2006 and the 2007 cohorts. Of the 522 fish assigned to parents from 2006, three (0.57%) returned at age two (100% males), 503 (96.4%) at age three (36.8% males and 63.2% females), and 16 (3.07%) at age four (43.8% males and 56.2% females). Of the 1500 fish assigned to parents from 2007, 47 (3.13%) returned at age two (97.9% males and 2.1% females, 1240 (82.67%) at age three (53.7% males and 46.3% females), and 213 (14.2%) at age four (35.2% males and 64.8% females; Figure 3.1). While z-tests identified significant differences between the two cohorts for the proportion of two-year olds ($z = -3.24$, $P < 0.01$), three-year olds ($z = 7.81$, $P < 0.01$), and four-year olds ($z = -6.91$, $P < 0.01$), both mixtures were dominated by the three-year old age class. Two-year old females were uncommon in both cohorts.

Again utilizing the reconstructed pedigrees, the full age structure of the spring-run spawning broodstock was examined for the years 2010 and 2011. However, the tallies of parentage assignments for these two years in Table 3.2 contain individuals that were actually spawned with the fall-run (fall-run/spring-origin from Table 3.1). In order to get a true picture of the relative proportion of two-, three-, and four-year olds in the 2010 and 2011 spawning populations, parentage assignments of fall-run/spring-origin individuals were removed and the age structure evaluated anew. After exclusion of the fall-run spawners for 2010, 814 assignments remained, representing 61.5% of the spring-run spawners in that year. For 2011, 529 assignments were retained, representing 46.9%

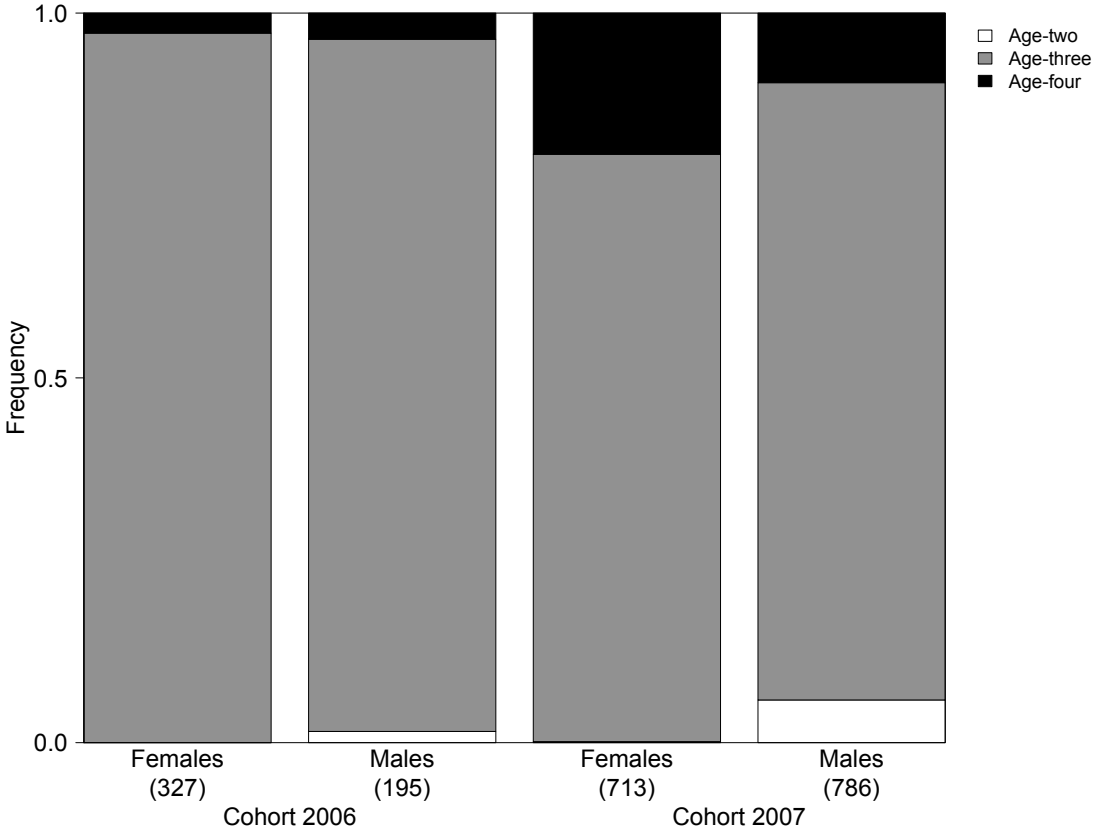

Figure 3.1: Age structure of returning adults (male and female) for two cohorts (2006 and 2007) from the Feather River Hatchery, CA. Numbers in parentheses indicate the total number of fish in each category, while white bars denote two-year olds, grey bars three-year olds and black bars four-year old fish.

of the spring-run spawners. Among the 2010 spring-run spawners, six (0.74%) were age two (100% males), 799 (98.16%) were age three (54.1% males and 45.9% females), and nine (1.10%) were age four (44.4% males and 55.6% females). Among the 2011 spring-run spawners, 16 (3.02%) were age two (100% males), 347 (65.6%) were age three (52.2% males and 47.8% females), and 166 (31.4%) were age four (33.1% males and 66.9% females; Figure 3.2). Z-tests also detected significant differences between the two spawn groups for the proportion of two-year olds ($z = -16.1129$, $P < 0.01$), three-year olds ($z = 16.493$, $P < 0.01$), and four-year olds ($z = -3.23$, $P < 0.01$).

3.3.4 Variance in Family Size and Reproductive Success

Parentage reconstruction yielded 2776 parent-offspring trios derived from 1083 unique parent pairs and distributed in 1081 pedigrees (only two males were found to have spawned with multiple females over the study period). A total of 1079 males and 1081 females successfully produced offspring that returned to the hatchery as adults. The mean number of offspring for successful parent pairs was 2.6 (range, 1-13; Figure 3.3). Among successful parent pairs, 37.9% had only a single offspring return and only one parent pair yielded thirteen offspring, the largest full-sibling family detected. For the 2006 cohort, 39.6% of the number of hatchery-reported matings (Table 3.1) yielded recovered offspring, while 65.5% of the reported number of 2007 spawners achieved reproductive success. In 2007, more parent pairs had two offspring return than one offspring, otherwise the distribution of family sizes across years was comparable (Figure 3.4). A significant difference in the pattern of reproductive success was found between the 2006

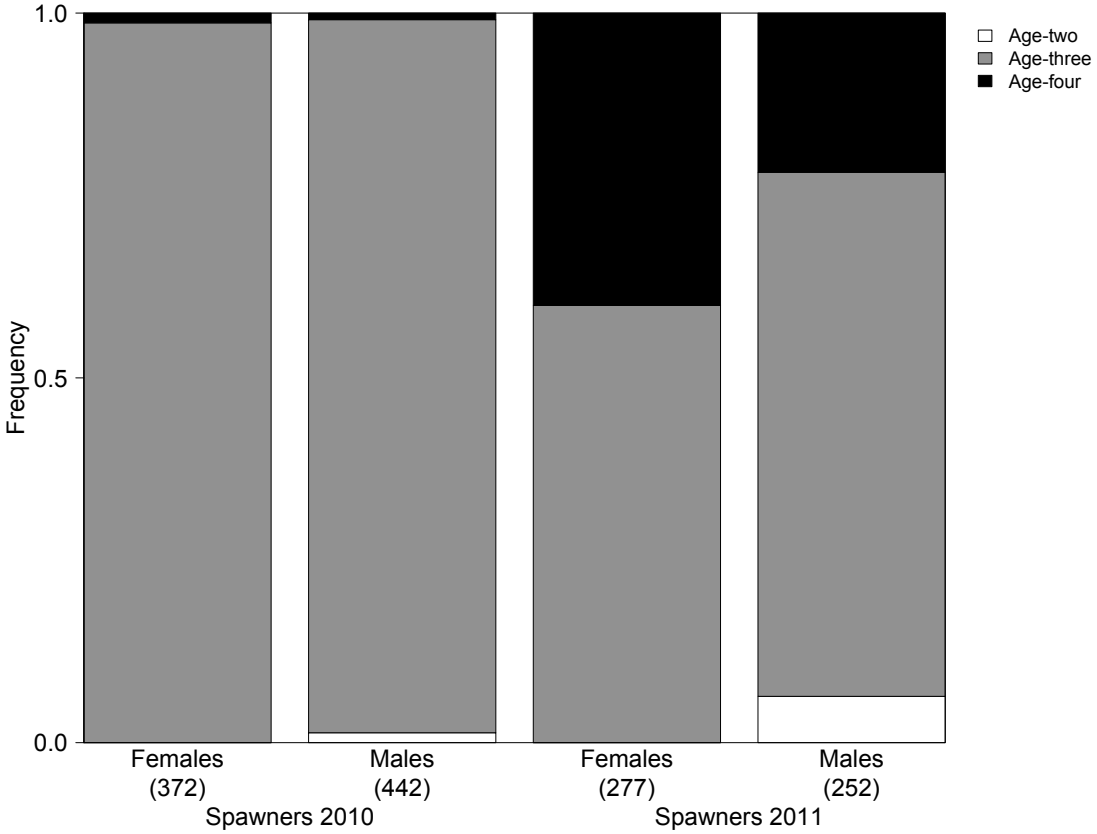

Figure 3.2: Age structure of spawning adults (male and female) for two years of spawner broodstock from the Feather River Hatchery, CA. Numbers in parentheses indicate the total number of fish in each category, while white bars denote two-year olds, grey bars three-year olds and black bars four-year old fish.

and 2007 cohorts (chi-squared = 44.67, $P < 0.001$). Among the over 2000 fish used as broodstock for a species reintroduction to the San Joaquin River, 102 pedigrees were assembled, containing a single family of four full-sibs, four families of three full-sibs, and 19 families of two full-sibs, with the remainder as singletons.

For salmon, as with most fishes, larger female body size generally allows for production of a larger number of eggs (Groot and Margolis 1991). If an increase in the number of eggs provides more opportunities to have offspring return, a correlation between female body size and reproductive success may be expected. Using the estimates of reproductive success from the reconstructed pedigrees and multiple regression, I found a highly significant correlation ($P < 0.001$) between reproductive success and mothers with lengths greater than 787mm (Figure 3.5).

3.3.5 Heritability of Length-at-spawning

Using the reconstructed families and the known lengths of sampled fish, the following regressions on length were examined for 3-year old offspring: parental mean-all offspring, parental mean-male offspring, parental mean-female offspring, father-son, mother-daughter (Figure 3.6). A positive, highly significant correlation was detected for all comparisons, however variability was also high. For all 3-year old offspring, the mean parental length explained approximately 3% of the observed variation ($F_{1,2302} = 68.59$, $R^2 = 0.029$, $P < 0.001$). The mean length of the parent pair explained more of the observed variation in the length of female offspring ($F_{1,1160} = 43.39$, $R^2 = 0.036$, $P < 0.001$) than of male offspring ($F_{1,1139} = 22.8$, $R^2 = 0.020$, $P < 0.001$). Among

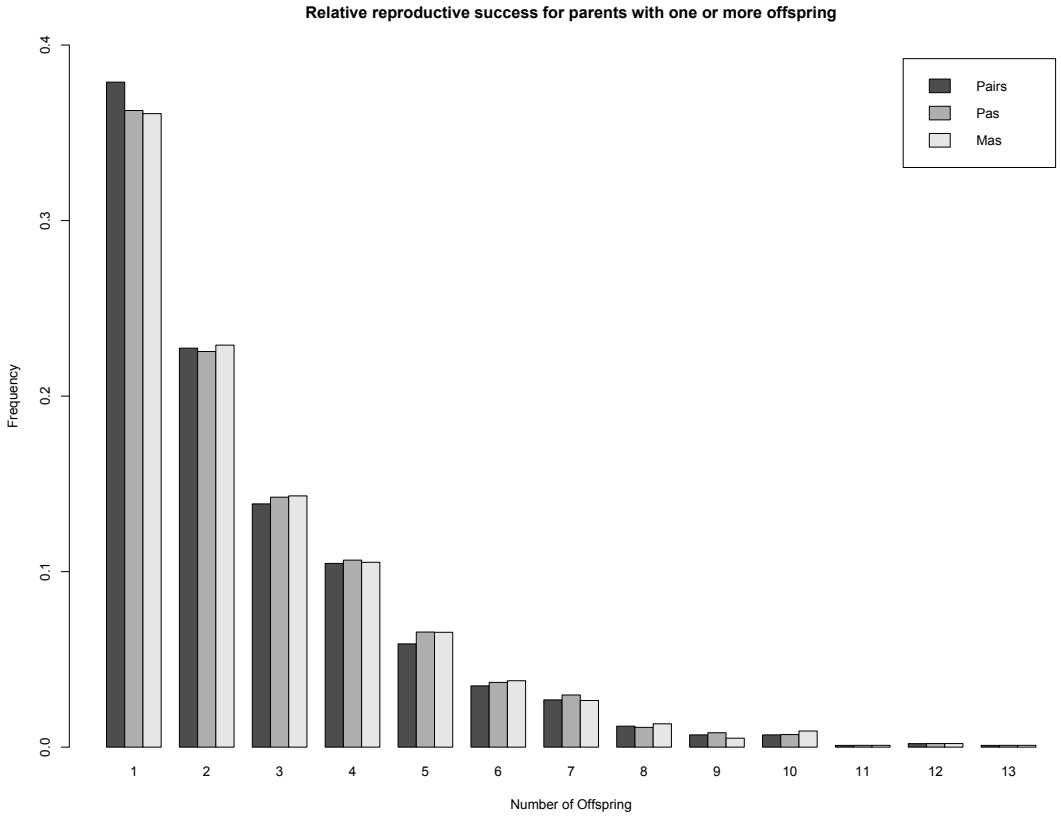

Figure 3.3: Number of offspring that returned to the hatchery for females (white bars), males (grey bars) and mated pairs (dark bars) over all study years. The similarity over comparisons is expected as generally one male is spawned with one female at the hatchery.

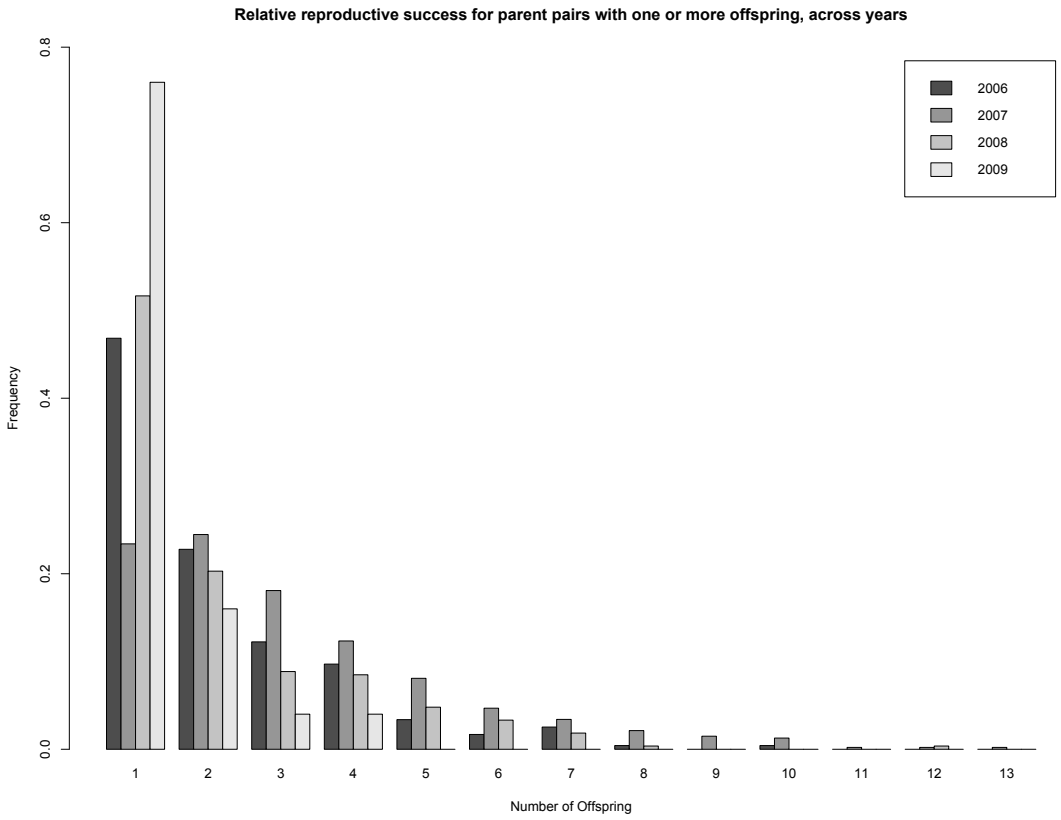

Figure 3.4: Number of offspring (full-siblings) that returned to the hatchery for parents spawned in each study year, 2006-2009. Note that offspring of 2009 spawners are under-represented as sampling permitted assignment of only two-year old fish.

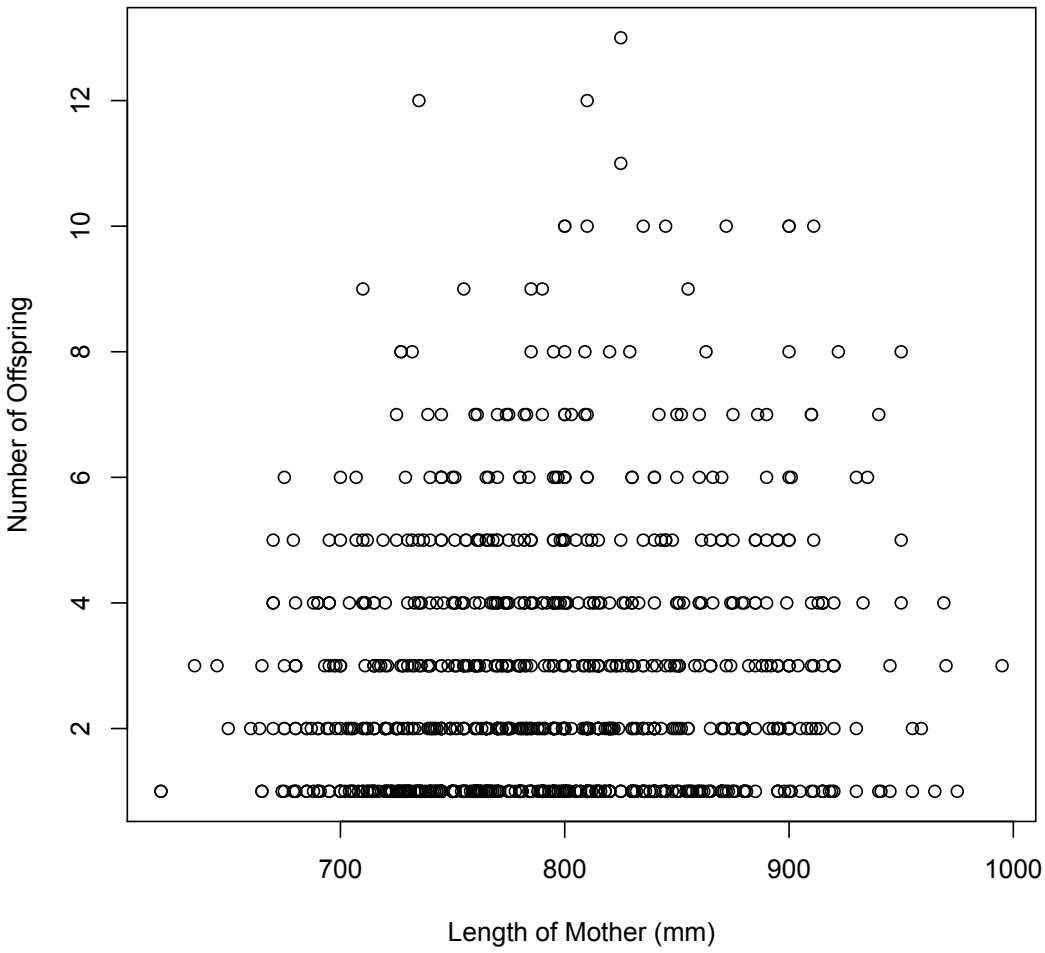

Figure 3.5: Relationship between the length of a mother and the number of her offspring that returned to the hatchery as adults at ages two, three or four. The size of full-sibling families here ranges from one to thirteen.

Table 3.3: Heritability (h^2) of length-at-maturity estimated as the slope of the length-length regression line between different comparisons of parents and offspring (Figure 3.6). Mean is the average length of the parents. The regression goodness of fit (R^2) and standard error (SE) are also reported.

Parent	**Mean**	**Mean**	**Mean**	**Male**	**Female**
Offspring	**All**	**Male**	**Female**	**Male**	**Female**
h^2	0.189	0.139	0.156	0.062	0.110
R^2	0.029	0.020	0.036	0.010	0.026
SE	18.93	24.27	19.47	15.98	15.60

offspring of the same gender as the parent, the mother's length explained more of the variation in the length of her female offspring ($F_{1,1160} = 31.11$, $R^2 = 0.026$, $P < 0.001$), than did the father's length of his male offspring ($F_{1,1139} = 11.2$, $R^2 = 0.010$, $P < 0.001$). Heritability (h^2) was calculated as the slope of the length-length regression line. Of the comparisons examined here, heritability was highest for the mean parent length as realized by all offspring ($h^2 = 0.189$), followed by the mean parent length and female offspring ($h^2 = 0.156$). The heritability of the mother's length by her female offspring ($h^2 = 0.110$) was higher than the heritability of the father's length by his male offspring ($h^2 = 0.062$; Table 3.3).

3.3.6 Relatedness

The Rxy estimator was used to calculate relatedness between all pairs of individuals within each of the sample collections. Over all samples, Rxy ranged from -0.62 to 0.83 (mean, 0.003), while the mean of all pairwise Rxy values within each broodstock collection (2006-2011) ranged from -0.0067 in the 2006 spring sample to 0.0074 in the 2008 fall sample. The estimator was normally distributed for all collections and overall,

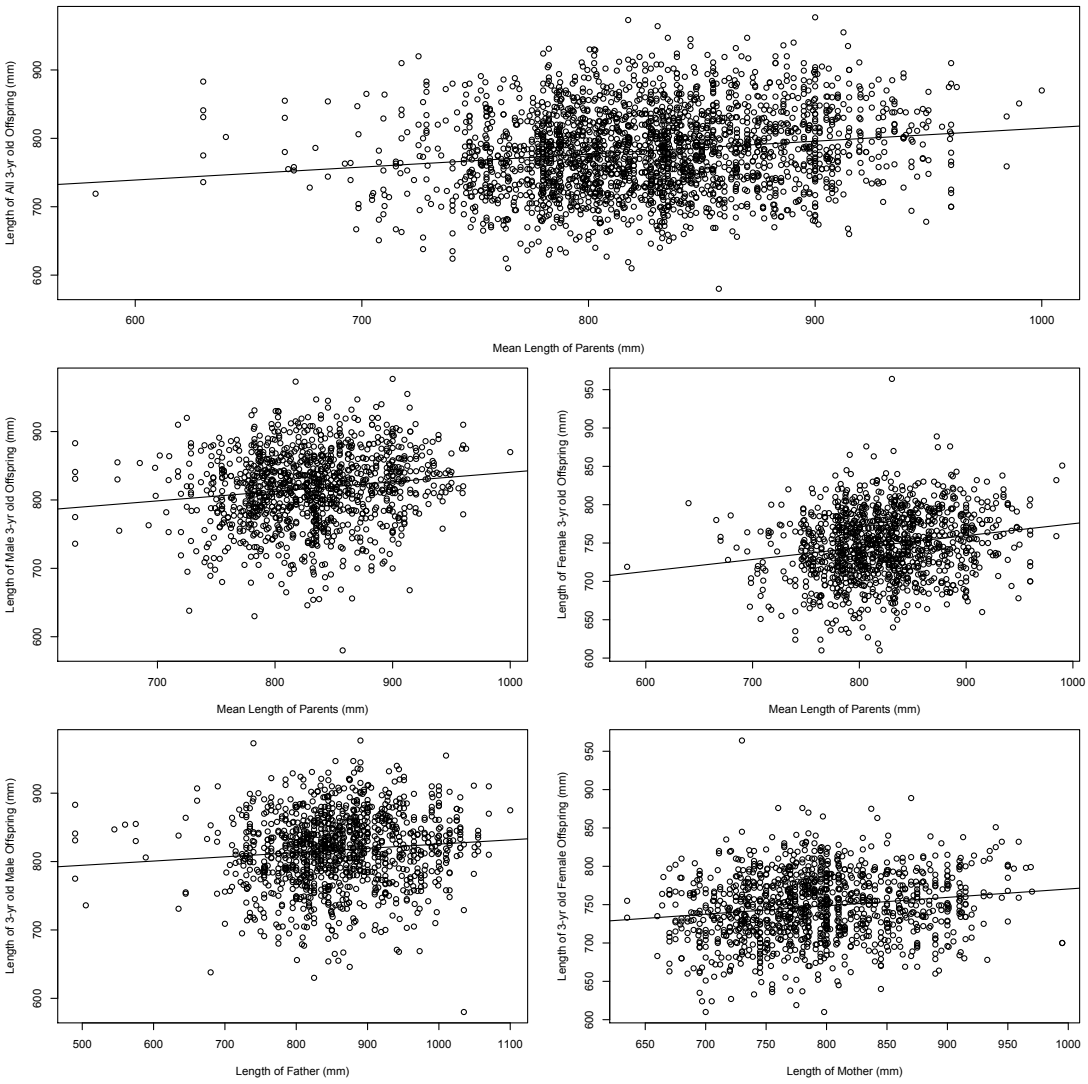

Figure 3.6: Linear regression of parental length on the length of their 3-year old adult offspring. Independent comparisons were made for: mean parent length and all offspring, male offspring, and female offspring, as well as, fathers and male offspring and mothers and female offspring.

while the skew was low but positive for all but the fall-run 2008 collection (range, -0.017 to 0.068; mean, 0.044; Figure 3.7). A positive skew suggests an asymmetry towards Rxy values greater than zero (i.e. a longer tail of higher relatedness estimates).

The distribution of relatedness between parents that successfully had offspring return to the hatchery as adults was compared to that of parent pairs with no reproductive success for the 2006, 2007, 2008 and 2009 spring-run brood years. These were the four samples for which the mated pair was recorded during spawning at the hatchery, however, only for the years 2006 and 2007 was the full age structure of the cohort (two-, three- and four-year olds) recovered through parentage assignments. It is likely that a small proportion of the parent pairs from 2008 deemed unsuccessful, may yet have four-year old offspring return in 2012, and likewise, many of the unsuccessful parent pairs from 2009 will have three-year old offspring return in 2012 and four-year olds in 2013. Data were again, approximately normally distributed. For all successful parent pairs (across years), Rxy ranged from -0.34 to 0.34 (mean, -0.0083) with a skew of 0.002; for unsuccessful pairs, Rxy ranged from -0.33 to 0.43 (mean, 8×10^{-4}) with a skew of 0.084 (Figure 3.8). For all years 2006-2009, mean relatedness was larger for unsuccessful parent pairs (range, -0.0296 - 0.0039) than for successful parent pairs (range, -0.0178 - -0.0035). T-tests detected no significant differences (mean p-value = 0.326) in the mean of Rxy values between successful and unsuccessful parent pairs in any year, or overall. With the exception of 2008, skew was also more positive in the sample of unsuccessful spawners (range, -0.047 - 0.211) than among successful spawners (range, -0.475-0.148). A weak negative correlation was detected between the relatedness of successful spawn-

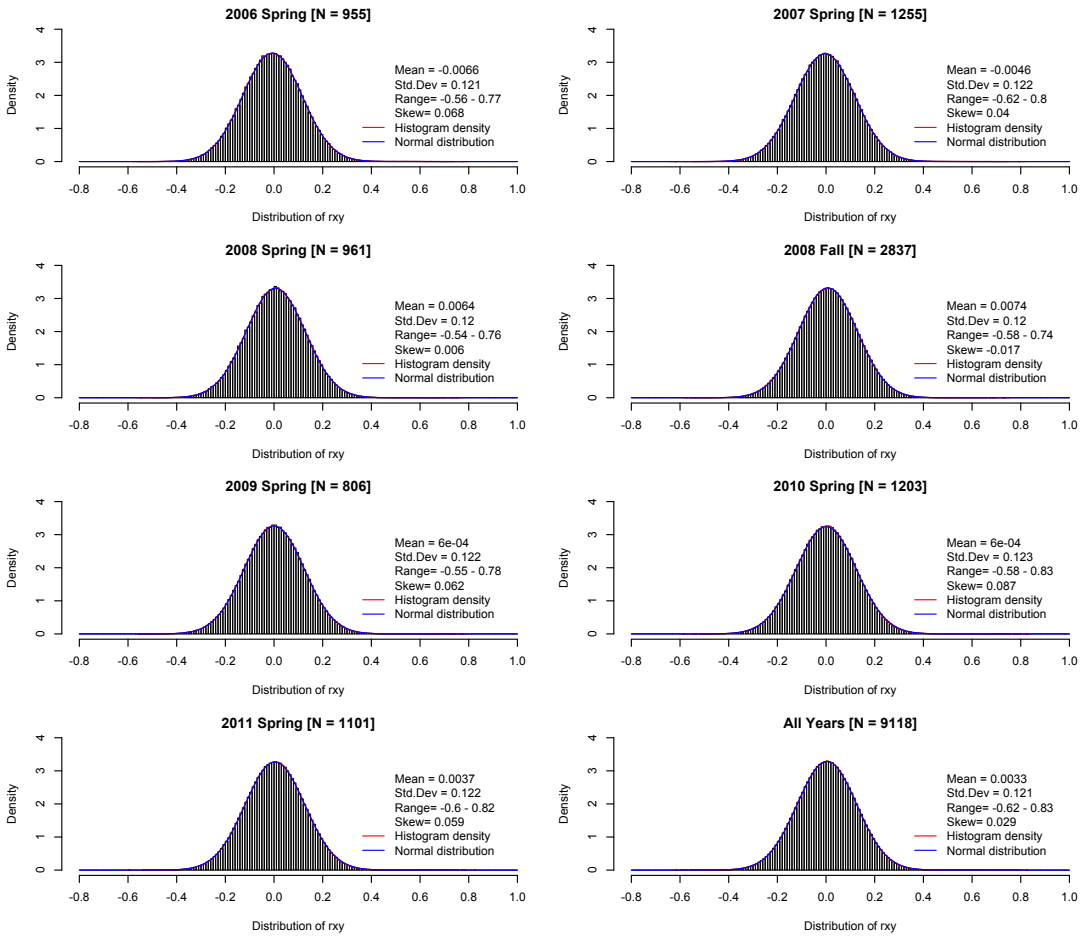

Figure 3.7: Distribution of the relatedness coeffcient (Rxy; Queller and Goodnight 1989) between all possible pairs of individuals in each collection of spawning broodstock and over all samples. Values are normally distributed, so the range, mean, standard deviation (Std. Dev.) and skew are reported.

ing pairs (all years) and the number of offspring they produced ($F_{1,1651} = 4.011$, $R^2 = 0.002$, $P < 0.05$), suggesting that less related parents may realize greater reproductive success (Figure 3.9).

3.3.7 Fishery Samples

A total of 24,242 Chinook salmon were sampled in four fishery collections from 2010 to 2012; of these, 874 (3.6%) were excluded for excessive missing data (>10 loci), leaving 23,368 samples for parentage analysis. A total of 771 fish sampled in ocean fisheries were assigned to FRH parents (Table 3.4). Over all assigned fishery samples, mean FDR was 0.001 and the posterior probability of the parent-offspring relationship ranged from 0.879 to 0.999 (mean, 0.996). Of the 2090 samples collected at California ports in 2010, 1855 were successfully genotyped, and CWTs were recovered for 1108 (515 from the Feather River). Recovered CWTs identified 61 individuals from the FRH spring-run, 40 of which were confirmed by parentage analysis (65.6%). Nine additional individuals that were assigned to spring-run parents presumably had lost or unreadable CWTs (the hatchery reports 100% tagging of spring-run offspring). One individual which assigned to spring-run parents with high confidence (FDR = 0) contained a CWT indicating a Coleman National Fish Hatchery (located on Battle Creek, a more northern tributary to the Sacramento River) origin, however, the cross between the genetically assigned parents had been recorded at the FRH in 2007, strongly suggesting a misread or errantly placed CWT. Importantly, only two-year old ocean-caught fish in 2010 were available for assignment to the fall-run broodstock, which was sampled only in 2008.

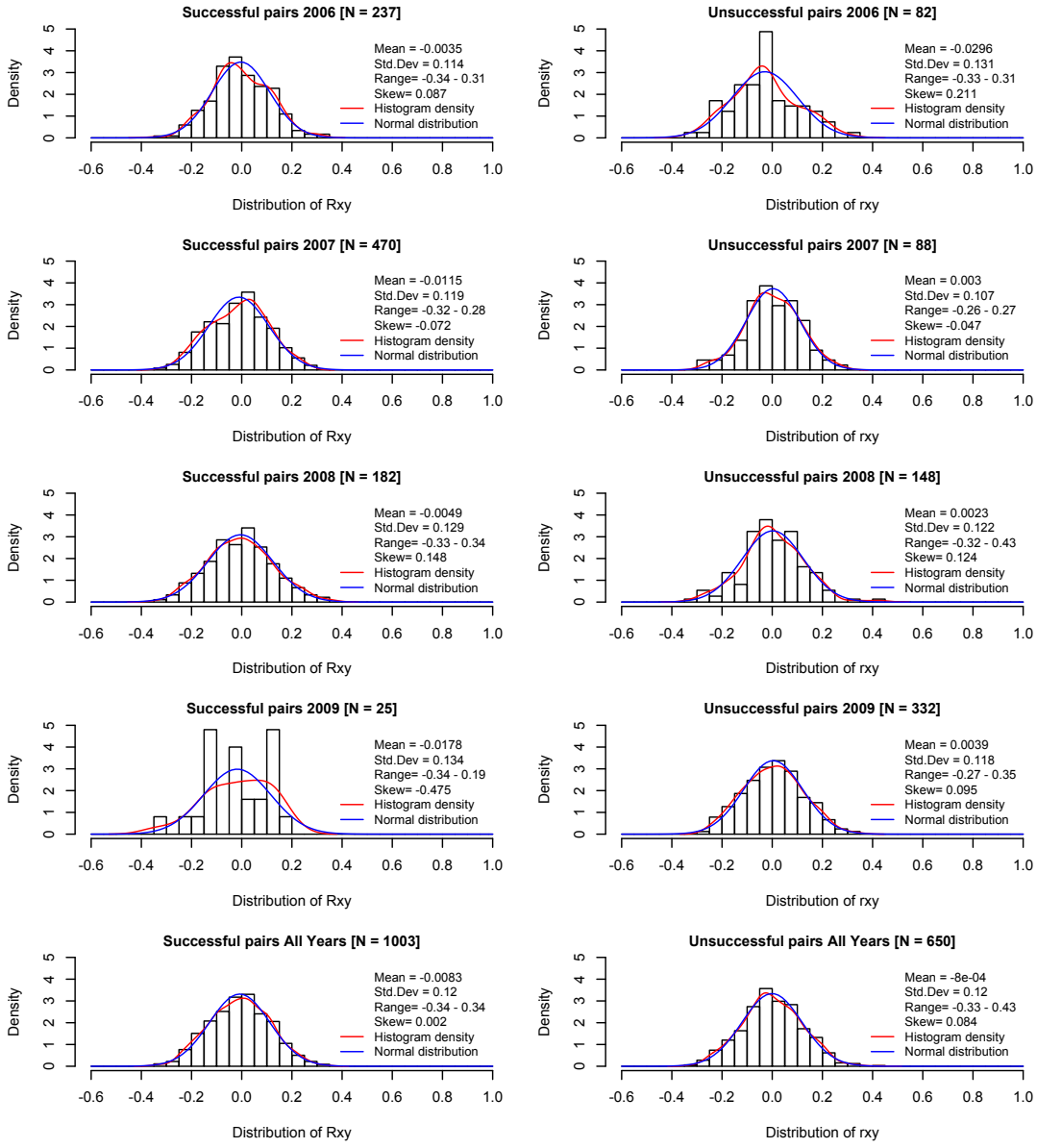

Figure 3.8: Mated pairs were recorded at the FRH for spring-run spawners from 2006-2009. Parentage assignment allowed for the comparison of the distribution of relatedness (Rxy) among pairs that successfully had offspring return to the hatchery as adults (left side) and those that did not (right side). Again, values were normally distributed, and the range, mean, standard deviation (Std. Dev.) and skew are reported.

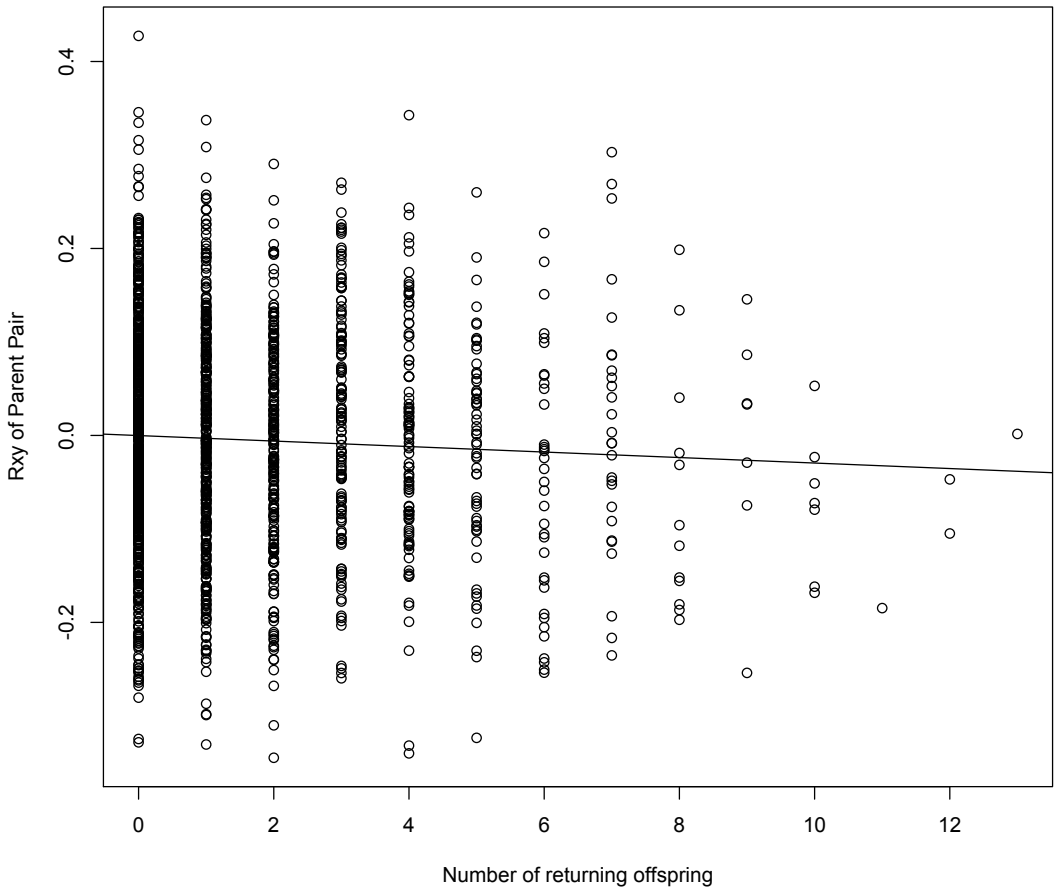

Figure 3.9: Linear regression of the degree of relatedness between a parent pair (as estimated by Rxy) and the number of offspring that returned to the hatchery in subsequent years. This includes Rxy values for parents that had no offspring return.

Parentage analysis identified 14 two-year old fish of the 454 individuals with fall-run CWTs and an additional 26 fall-run offspring with no tags (the hatchery tags only 25% of fall-run fish).

For the three samples collected by the commercial fleet in 2010, 2011 and 2012, 21,513 individuals were successfully genotyped. For the 2010 sample, parentage analysis identified 134 FRH offspring (88.8% two-years old and 11.2% three-years old), of which 85 were assigned to 2008 fall-run parents. Two-year old fish are not generally targeted by the commercial fleet, however, this sample was collected specifically for the analysis described here and employed a catch-and-release strategy. In the 2011 sample, all but two (two-year olds) of the 449 assigned individuals were found to be three-years old and 84.4% of assignments were to 2008 fall-run parents. Finally, in the 2012 sample, the 99 parentage assignments indicated 87.9% three-year olds and 12.1% four-year olds, with 83.3% of four-year old fish coming from 2008 fall-run parents (Table 3.4).

Table 3.4: Summary of sampling effort and genotyping success in samples collected from mixed-stock ocean fisheries. Individuals with missing data at more than 10 loci were excluded. The age composition (comp.) of individuals in each sample was determined by assigning them to parents from the Feather River Hatchery (FRH). The number of the reported assignments (Assmts.) that were to fall-run parents sampled only in 2008 is also shown, as well as the age-class represented by those offspring (offs.) recoveries. For the 2010 samples collected at California ports, coded-wire tag information was available for comparison to genetic recaptures.

Sample Collection	**Sampled [n]**	**Excluded [n]**	**Total CWTs**	**FRH CWTs [fall]**	**FRH CWTs [spring]**	**Age comp. (all)**			**Assmts. to 2008 fall run**	**Age-class of fall offs.**	**Assmts. without CWTs**
						2-yr old	**3-yr old**	**4-yr old**			
CA Ports 2010	2091	236	1108	454	61	43	46	0	40	2-yr old	35
Fishery 2010	5042	425	-	-	-	119	15	0	85	2-yr old	-
Fishery 2011	7924	88	-	-	-	2	447	0	379	3-yr old	-
Fishery 2012	9185	125	-	-	-	0	87	12	10	4-yr old	-
totals	24242	874				164	595	12	514		

3.4 Discussion

The current study describes the first implementation and verification of a large-scale genetic tagging and pedigree reconstruction experiment in Chinook salmon, a keystone species in the marine, terrestrial and freshwater ecosystems of the West Coast of North America (Willson and Halupka 1995; Cederholm *et al.* 1999; Helfield and Naiman 2006). Chinook are also the target of highly valuable commercial and recreational fisheries throughout the northeastern Pacific Ocean and receive substantial management attention. Our experimental design involved sampling entire parental generations from one of the largest hatchery programs in California, USA, and subsequently recovering their offspring as they returned to the hatchery two, three, and four years later. Using a panel of 96 SNP markers and new, highly-efficient algorithms for parentage reconstruction, offspring were assigned to their parents with high accuracy, as confirmed by recorded mate pairs and physical tags. Pedigrees were then used to calculate informative population genetic parameters and investigate the potential for heritability of important life-history traits. Hatchery offspring carrying genetic tags were also recovered from large mixed-stock fishery samples, demonstrating the effectiveness of the methodology for providing the necessary stock- and cohort-specific information to current ocean harvest models.

3.4.1 Technical Issues

Sample quality was unexpectedly poor in some years and some collections; extracted DNA was quite degraded and did not yield acceptable genotypes for 13.8% of FRH broodstock samples. While DNA degradation has been documented in carcass recoveries of naturally spawning salmon (Baumsteiger, 2009), our lab generally observed adequate DNA quality when sampling live fish. In the years other than 2008 (and to a lesser extent in 2006), the proportion of successfully genotyped fish was closer to expectations. As described, samples were collected from all fish encountered at the hatchery, including some that may have died in the holding pens while awaiting spawning. It may be that many of the individuals that failed genotyping in 2008 had been dead for some time, at which point natural processes of decay begin to degrade DNA. To that point, disproportionately more fish were genotyped than reported spawned in 2008 as compared to other years (Table 3.1) and many of the individuals that failed genotyping did not appear in mate pair records. Alternatively, spring-run fish may have encountered warmer temperatures while holding in the river in 2008, which can decrease available dissolved oxygen and encourage fungal infections to the detriment of exposed fish (Pauley 1967, Allen *et al.* 1968). Anecdotally, samplers noted fungal and algal growth on the caudal fins from which samples were collected, for some of these individuals. Subsequent sampling efforts should target intact tissue as much as possible and avoid areas potentially contaminated by fungus or algal growth.

3.4.2 Parentage Assignments

Pedigrees were reconstructed with high confidence, as indicated by high maximum posterior probabilities and low FDR scores, and their accuracy was confirmed by the records of mated pairs at the hatchery. For the years in which sampling allowed for the recovery of the dominant three-year old age-class (2009-2012), the proportion of offspring assigned to parents was almost 55%. However, the proportion of parent pairs included in the parent database is critical to understanding the relative success of assigning parentage. For example, an offspring would not be assigned parentage if one or both parents were either not sampled or had been excluded for excessive missing data. If it is assumed that adult sampling at the hatchery was comprehensive, then the primary source for unassigned parentage is likely the exclusion of parents with low-quality genotypes. Using the excluded proportion of possible parent pairs to correct observed assignments, it is expected that $\sim$75% of offspring would have been assigned parentage had their parents been retained in the parent database.

There are a number of possible explanations for the remaining proportion of offspring that were not assigned to parents. Some parents may have spawned outside of the study period (e.g. the parents of 4 year-old fish collected in 2009 would have spawned in 2005), while others may simply have not been sampled, as in 2006 and 2009, where the number of individuals genotyped was less than the number reported spawned at the hatchery. Unassigned individuals could themselves be strays from elsewhere in the Central Valley, however over the study period, only three individuals contained CWTs

that indicated a source other than the FRH and therefore do not account for a large number of missing assignments. The missing parents may also have spawned naturally in the river outside of the hatchery, so they would not have been sampled and could not be assigned offspring. Finally, the missing parents of some spring-run fish could have spawned as part of the fall-run hatchery broodstock. The rate at which fall-run spawners contribute subsequent spring-run spawners can be estimated by examining the pattern of assignments to 2008 parents, when both the spring-run and fall-run broodstock were sampled. Of the 592 individuals assigned to parents from 2008 (both runs), 353 were subsequently spawned as spring-run in their collection year (as opposed to being fall-spawn/spring-origin). And of these 2010, 2011 and 2012 spring-run spawners, almost 32% derived from parents that were spawned as part of the fall-run. This is likely an upper-bound for the expected proportion of unassigned spring-run spawners from fall-run parents, as four year-olds are underrepresented in the partial 2012 sample. Regardless, this suggests that fall-run parents, not wild spawning fish in the river, are the most likely source for spring-run fish that were not assigned parentage.

3.4.3 Heritability of Length-at-maturity

Body size is an important morphological trait for salmon at various stages in their complex life history. At juvenile life stages, larger coho salmon (*O. kisutch*) have been shown to experience higher over-winter survival (Quinn and Peterson, 1996), higher rates of return at maturity (Bilton *et al.* 1982), and increased marine survival (Holtby *et al.* 1990). In adults, larger females have higher fecundity while for males,

larger body size has been correlated with increased social status and greater access to spawning opportunities for males in multiple salmon species (Keenleyside and Dupuis 1988, Fleming and Gross 1994; Quinn and Foote 1994). Given these potentially strong selective pressures on length, it is not surprising then that length-at-maturity should have a heritable component (Ricker 1972). The estimates here of heritability of length-at-maturity are within the range reported over multiple studies of salmonids (Carlson and Seamons 2008) and directly comparable to those reported for captive stocks of Atlantic salmon (*Salmo salar*; Refstie and Steine 1978) and Chinook salmon (Winkelman and Peterson 1994). While these estimates of heritability of length-at-maturity are not as high as those reported for age of maturity in Chinook salmon (Hankin *et al.* 1993) or more recently, spawn timing in steelhead trout (*O. mykiss*; Abadía-Cardoso 2013), they do still provide evidence that a genetic component is available to be be acted upon by selection.

The relationship between parent and offspring length was highly variable. This may be attributed primarily to the fact that length, as determined by growth, must have a significant environmental component. While larger size can confer a competitive advantage, growth is ultimately dependent on habitat and resource (food) availability, and offspring encounter a different regime than their parents. It is possible that the inherited component may provide the potential to reach some maximum length (similar to the ultimate length described by Bertalanffy, 1938), however resources must be available to reach this potential. The relationship between parent and offspring length-at-maturity may also be confounded in this study by the age of maturity, which likely has a strong

heritable component in Chinook salmon (Withler *et al.* 1987; Hankin *et al.* 1993). The observed positive correlation could arise if, for example, four year-old parents were more likely to give rise to offspring that returned at four years of age, or two year-old males were more likely to have two year-old returning offspring. Without knowing the age of both parents and offspring, I was not able to examine this effect directly, however, it is clear from Figure 3.6 that the small (500-600mm), predominately two-year old male offspring are descended from parents representing a diverse range of lengths and likely ages. Furthermore, analysis indicated that three-year old fish predominate in both cohorts and spawn years, suggesting that the majority of the length comparisons here are between three-year old parents and three-year old offspring.

3.4.4 Age Structure of Returning Adults and Spawning Broodstock

Reconstructed pedigrees allowed for examination of the age structure in two cohorts (following a group of offspring born in the same year through time) and two full spring-run spawn groups. While age-structure of the spawning population is undoubtedly a product of interannual cohort strength, it is still an important parameter that will benefit from baseline data collection. Though only a small sample of years, evidence indicates high interannual variability in the relative distribution of two- three- and four-year old fish in both cohorts and spawners. This variability is highlighted by the large difference in the proportion of four-year old spawners beween 2010 and 2011. Without baseline data on age structure at the FRH (which this study is providing for future years there), this may represent a large pulse of primarily female four year-old

spawners in 2011 or uncharacteristically few in 2010. In either case, if age-at-return does indeed have a strong heritable component in Chinook (Withler *et al.* 1987; Hankin *et al.* 1993), these differences could lead to significant changes in the age structure of not just the 2010 and 2011 cohorts, but also the age structure of spawn groups in subsequent years. No five year-old fish were recovered with PBT, although only one opportunity (in 2011) was available for detecting this age-class. While CWT data did indicate that some five-year old individuals were present in the dataset (data not shown), these fish were collected in years for which their parents would have been spawned outside of the study period.

Two-year old males, also called jacks, were detected in all cohorts and spawn years. This alternative life history strategy is well described in the species (Myers *et al.* 1988), however it may be increasing in frequency due to fishing pressure (Ricker 1981, Hard *et al.* 2008) and release of hatchery-reared fish (Unwin and Glova 1997). It has also been suggested that the random mating practices at hatcheries are imposing a powerful selective force towards younger age-at-return, by including the spawning of jacks, which in the wild, experience reduced opportunities for spawning and low reproductive success. A mating regime that more closely resembles the natural spawning hierarchy favoring large males is recommended (Hankin *et al.* 2009). At the FRH there is not an explicit policy against spawning jacks, however, in practice they are discarded at a higher rate than large fish and so receive fewer opportunities for reproductive success (personal observation). A single female returning at two-years of age (termed a jill) was also detected. While the presence of jills is uncommon, if returning two-

year old females increase in frequency in the future, their presence could exacerbate the shift towards early age-at-maturity. PBT offers a powerful tool for monitoring age structure in hatcheries and, in the future, will allow for quantitative genetic study of the inheritance of the trait.

3.4.5 Inbreeding and Reproductive Success

Inbreeding is a potentially serious negative consequence of artificial propagation of salmonids in hatcheries (Wang *et al.* 2001; Waples 1991). At the level of individuals, inbreeding results from matings between family members (i.e. siblings, cousins, aunts/uncles, etc.). In the wild, salmon use their ability to identify kin (Quinn 1985; Olsen 1998) to avoid matings with close relatives (Landry *et al.* 2001; Rajakaruna *et al.* 2006). Using parentage analysis to identify two-generation pedigrees, I assessed the precise relationship of mated individuals for the 2012 collection of FRH broodstock to be used for reintroduction in the San Joaquin River, CA. Matings between siblings has been shown to have serious consequences on fitness (Kincaid 1983, Wang *et al.* 2001) and marine survival (Thrower and Hard 2009), and so was a primary concern of project managers and scientists (Broodstock selection document; available at: http://restoresjr.net/program_library/02-Program_Docs/StockSelectionStrategy2010Nov.pdf). The analysis found no matings between full-siblings, however, almost 20% of individuals spawned had a full-sibling in the broodstock. This was much higher than anticipated and has motivated additional safeguards to evaluate and correct for related individuals in future reintroduction efforts. It is important to note, that this type of individual-based

analysis in large populations would be infeasible with any other tagging technology.

In future years of this project pedigree-based mate evaluation will be the norm, however in the meantime, genetic estimates of relatedness can be used to evaluate inbreeding at a population level. In this sense, inbreeding results from matings between individuals that are more related than average, as opposed to having a specific known relationship (Queller and Goodnight 1987). Little evidence was found for high levels of inbreeding in any of the broodstock samples analyzed. Estimates were highest for the large fall-run collection, but the overall distribution of relatedness conformed to expectations. The finding that reproductive success is correlated with lower levels of parental relatedness is novel in Chinook salmon, but has been shown to be a major determinant of survival for small captive stocks of coho salmon (*O. kisutch*; Conrad *et al.* 2013). However, the high variability in this relationship suggests that breeding practices in large hatchery programs intended to limit close kin matings may not impact reproductive success as much as stochastic environmental effects. Management guidelines at the FRH call for one-to-one matings between males and females, however the expected number of offspring from inbred matings may be unchanged. For the FRH, the almost identical patterns of reproductive success for males and females and the identification of only two half-sibling relationships confirms that the desired mating scheme is being implemented in practice at the hatchery. A similar analysis of breeding practices in a California steelhead program revealed that hatchery procedures concerning re-use of males and spawning of two-year olds were vastly different in practice than as specified in management goals (Abadía-Cardoso *et al.* 2013).

Using the results of the reproductive success analysis, a positive correlation between female body size and the number of her offspring that return in subsequent years was detected. For females, larger body size allows for the production of more eggs and increased chance of reproductive success. It is somewhat surprising that this effect is detected with adult offspring, after the many high-mortality stages (emergence, outmigration, ocean entry, etc) encountered during their life history. However, increased survival of offspring from large females may be mediated by the heritable component of size. Offspring size at early life stages has been shown to be largely determined by maternal size in Chinook (Heath *et al.* 1999), and here I show that correlations with parental size persist into adulthood. This indicates that the size advantage conferred upon offspring by their parents may have important implications for future survival and potential reproductive success.

3.4.6 Fishery Assignments

Parentage-based tagging has been proposed as an alternative to coded-wire tags for management of Pacific salmonids (Hankin *et al.* 2005, Garza and Anderson 2007). While CWTs are used in components of hatchery management, their primary purpose is to identify the stock and age of individuals captured in mixed-stock ocean fisheries, for input into the cohort-based mortality models used by management agencies (i.e. PSC, PFMC). Here I was able to perform a direct comparison between the genetic and traditional tagging methods, as 100% of the FRH spring-run receives CWTs. Parentage-based analysis identified the majority of individuals containing coded-wire

tags, despite a large number of excluded parent-pairs for two-year old fish. PBT also identified nine additional fish that should have had an FRH spring-run tag and one individual with a CWT reported from the wrong hatchery. This suggests that CWT error/loss rates may be as high as 14%, which is much higher than what is generally reported and expected (Johnson 2004). If indicative of the CWT program in general, error rates of this magnitude would undoubtedly influence the output of fishery harvest models.

Analysis of the ocean fishery samples further demonstrates the ability of PBT to provide stock-specific age distribution for fish encountered by the commercial fleet – the exact data needed for current management models. Furthermore, the high confidence of assignments shows that the statistical tools for assigning parentage, as well as the statistical power of the SNP panel, scale to the magnitude of the problem. In a high fecundity species like salmon, pedigree reconstruction can be extremely challenging because of the sheer number of possible parent-offspring trios that must be evaluated. For example to assign parentage for the 2012 ocean fishery sample, 7×10^9 possible trios were examined; in the analysis unconstrained by spawn date or sex, this number was 6.7×10^{10}. This is the largest parentage analysis reported for a salmonid species using SNPs (Abadía-Cardoso *et al.* 2013; Steele *et al.* 2013), and would not have been computationally possible with the previous generation of tools for assigning parentage (Anderson and Garza 2006, Jones *et al.* 2009, Hauser *et al.* 2011, Anderson 2012).

3.5 Conclusions

This study describes the large-scale genetic tagging of a hatchery Chinook salmon population by pedigree reconstruction. I demonstrate the power of SNP markers for accurate parentage assignment in this high fecundity species and show that genetic tags are capable of providing data comparable to current physical tags for fishery management. This tagging methodology also provides multigenerational pedigrees, which can be used to investigate population features, and how they change over time or in response to management actions. As illustrated here, pedigrees can be used to measure heritability of phenotypic traits, variance in reproductive success, and age structure in a population. This work also establishes a baseline for a variety of population genetic parameters, to which future generations can be compared. In subsequent years, as two- and three-generation pedigrees accumulate, I will investigate in even greater detail the quantitative genetic component of heritable life-history traits. This information will be used to formulate future management strategies and direct scientific investigations. The experiment described here should provide ample evidence that adoption of parentage-based tagging at hatcheries is not only technically feasible, but can provide important inference to guide genetic management of populations.

Conclusions and Future Directions

Chinook salmon is the largest species of Pacific salmonid and is the focus of highly valuable fisheries throughout the northern Pacific Ocean. Their complex life history exposes them to impacts in both the freshwater and marine environments and has led, in some cases, to severe population declines. Government agencies have traditionally mitigated the terrestrial and aquatic ecosystem impacts responsible for salmonid population declines with production of fish in hatcheries and subsequent population supplementation; millions of Chinook salmon originate in hatcheries each year and can be the majority of fish in some populations. Wild and hatchery stocks comingled in ocean fisheries can vary widely in productivity and abundance and the proportion of fish from different populations in mixed-stock ocean fisheries has important implications for harvest management and conservation. Without precise information on their ocean distribution, managers have few options for protecting depressed or at-risk stocks from fishery impacts other than shutting down or curtailing fisheries over broad areas. Hatchery fish are currently accounted for in ocean fisheries through the use of coded-wire tags. These tags provide the age and source stock of fish, which is then used in cohort-based models to inform fishery management decisions. However, the coded-wire tagging pro-

gram is aging and inefficient, suffering from extremely low tag recovery rates, and it has been recommended that alternative methods be explored. The work described here provides a powerful alternative to the coded-wire tagging program, capable of increasing both the quantity and quality of data used to manage this important resource on the West Coast of North America. Furthermore, the genetic methods employed here provide a broad range of corollary benefits, primarily in the form of large numbers of multi-generational pedigrees, which can be used not only to better monitor and manage hatchery supplementation programs, but also to understand the heritable basis of a wide range of important physical traits in the species.

The current generation of genetic tools for studying Pacific salmonids depends primarily on microsatellite markers. We have detailed here the numerous shortcomings of microsatellites for our desired applications and have shown that a transition to SNP markers will provide the high-throughput capacity, low-error rates and simple data portability necessary for the next generation of management methodologies. Despite limited genomic resources for Chinook salmon, our sequencing effort using ESTs from steelhead trout was very successful, yielding 117 new SNP assays and more than doubling the number of SNP markers described for the species. Furthermore, our balanced ascertainment and sequencing strategy generated SNPs with both high minor allele frequencies in our focal populations and sufficient power for discriminating populations on a coastwide scale. Many of these markers are already in broad use for genetic investigations throughout the species' North American range.

We then assembled a panel of 96 SNPs and provided a comprehensive power

analysis demonstrating its ability to identify Chinook salmon caught in the California Current Large Marine Ecosystem to their management unit or population of origin. Again, because of the balanced ascertainment strategy employed during SNP development, the baseline is also useful in fisheries north of the Columbia River. In a direct comparison with data from coded wire tags, we show that the GSI baseline provides results that are 99% concordant with the physical tags. Furthermore, using GSI, considerably more fish can be identified to reporting unit, including fish from natural stocks. In the future, this baseline can be easily extended, simply by genotyping new populations with the same set of SNP markers. Work is already underway to use GSI assignments together with GPS locations of sampling to correlate specific stocks with oceanographic conditions and underwater features. Efforts are also being made to incorporate GSI information into current ocean harvest models.

Finally, our research demonstrates that the same panel of SNP markers, which effectively provides coastwide (California, Oregon and Washington) resolution for GSI, also retains abundant power for large-scale parentage analysis. Since PBT does provide age and stock information and the entirety of hatchery production can be tracked by simply collecting genotypes from broodstock at spawning, cohort-based ocean harvest models stand to benefit tremendously from increased tagging and recovery rates. Such genetic tagging, and the analysis of the associated pedigrees, will also have considerable importance in understanding the effects of hatchery practices on life history parameters and fitness. As pedigrees become extensive we will be able to estimate the heritability of important traits in even greater resolution and they will serve as the basis for detailed

linkage maps and associated mapping of quantitative trait loci. The ultimate goal is an integrated GSI/PBT program, where all fish genotyped with the same set of markers can yield biological inference, either individual identification when parents are sampled (or a fish is recaptured), or population assignment using a baseline reference database if they are not directly linked to other sampled individuals in a pedigree. If implementation of PBT expands to all hatcheries, as is currently happening, we can expect that the advances in genetic resources and methods described here will foster fundamental improvements in the way salmon populations are studied, monitored and managed.

References

Abadía-Cardoso, A., A.J. Clemento, and J.C. Garza. 2011. Discovery and characterization of single nucleotide polymorphisms in steelhead/rainbow trout, *Oncorhynchus mykiss*. Molecular Ecology Resources **11**(Suppl. 1):31-49.

Abadía-Cardoso, A., E.C. Anderson, D.E. Pearse, and J.C. Garza. 2013. Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (*Oncorhynchus mykiss*). Molecular Ecology **22**:4733-4746.

Aguilar, A., and J.C. Garza. 2008. Isolation of 15 single nucleotide polymorphisms from coastal steelhead, *Oncorhynchus mykiss* (Salmonidae). Molecular Ecology Resources **8**:659-662.

Albrechtsen, A., F.C. Nielsen, and R. Nielsen. 2010. Ascertainment biases in SNP chips affect measures of population divergence. Molecular Biology and Evolution **24**:1-20.

Alexandersdottir, M.A., G. Hoffmann, G. Brown, and P. Goodman. 2004. Technical review of the CWT system and its use for Chinook and coho salmon management. Available from: www.psc.org/info_codedwiretagreview.htm

Allen, R.L., T.K. Meekin, G.B. Pauley, and M.P. Fujihara. 1968. Mortality among Chinook salmon associated with the fungus *Dermocystidium*. Journal of the Fisheries Research Board of Canada **25**:2467-2475.

Allendorf, F., and L.W. Seeb. 2000. Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. Evolution **54**:640-51.

Anderson, E.C., and J.C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics **172**:2567-2582.

Anderson, E.C., R.S. Waples, and S.T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences **65**:1475-1486.

Anderson, E.C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Molecular Ecology Resources **10**:701-710.

Anderson, E.C. 2012. Large-scale parentage inference with SNPs: an efficient algorithm for statistical confidence of parent pair allocations. Statistical Applications in Genetics and Molecular Biology **11**:12p.

Baldwin, S.P. 1921. Recent returns from trapping and banding birds. The Auk **38**:228-237.

Banks, M.A., V.K. Rashbrook, M.J. Calavetta, C.A. Dean, and D. Hedgecock. 2000. Analysis of microsatellite DNA resolves genetic structure and diversity of Chinook salmon (*Oncorhynchus tshawytscha*) in Californias Central Valley. Canadian Journal of Fisheries and Aquatic Sciences **57**:915-927.

Barnett-Johnson, R., C.B. Grimes, C.F. Royer, and C.J. Donohoe. 2007. Identifying the contribution of wild and hatchery Chinook salmon (*Oncorhynchus tshawytscha*) to the ocean fishery using otolith microstructure as natural tags. Canadian Journal of Fisheries and Aquatic Sciences **64**:1683-1692.

Barnett-Johnson, R., T.E. Pearson, F.C. Ramos, C. Grimes, and R.B. MacFarlane. 2008. Tracking natal origins of salmon using isotopes, otoliths, and landscape geology. Limnology and Oceanography. **53**:1633-1642.

Barton, N.H. 2000. "Genetic hitchhiking". Philosophical Transactions of the Royal Society of London, Biological Sciences **355**:1553-1562.

Baumsteiger J., and J.L. Kerby. 2009. Effectiveness of salmon carcass tissue for use in DNA extraction and amplification in conservation genetic studies. North American Journal of Fisheries Management **29**:40-49.

Beacham, T.D., R.E. Withler, and A.P. Gould. 1985. Biochemical genetic stock identification of pink salmon (*Oncorhynchus gorbuscha*) in southern British Columbia and Puget Sound. Canadian Journal of Fisheries and Aquatic Sciences **42**:1474-1483.

Beacham, T., R. Withler, and T. Stevens. 1996. Stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) using minisatellite DNA variation. Canadian Journal of Fisheries and Aquatic Sciences **53**:380-394.

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

Beamish R.J., C. Mahnken, and C.M. Neville. 1997. Hatchery and wild production of Pacific salmon in relation to large-scale, natural shifts in the productivity of the marine environment. ICES Journal of Marine Sciences **54**:1200-1215.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 1996-2004. GENETIX 4.05, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Bernard, D.R., and J.E. Clark. 1996. Estimating salmon harvest with coded-wire tags. Canadian Journal of Fisheries and Aquatic Sciences **53**:2323-2332.

Bernatchez, L., and P. Duchesne. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Canadian Journal of Fisheries and Aquatic Sciences **57**:1-12.

Bertalanffy, L.V. 1938. A quantitative theory of organic growth (inquiries on growth laws). Human Biology **10**:181-213.

Bethke, R., M. Taylor, S. Amstrup, and F. Messier. 1996. Population delineation of polar bears using satellite collar data. Ecological Applications **6**:311-317.

Bilton, H.T., D.F. Alderdice, and J.T. Schnute. 1982. Influence of time and size at release of juvenile coho salmon (*Oncorhynchus kisutch*) on returns at maturity. Canadian Journal of Fisheries and Aquatic Sciences **39**:426-447.

Bisbal, G.A. and W.E. McConnaha. 1998. Consideration of ocean conditions in the management of salmon. Canadian Journal of Fisheries and Aquatic Sciences **55**:2178-2186.

Blouin, M.S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. TRENDS in Ecology and Evolution **18**:503-511.

Bouck, A., and T. Vision. 2007. The molecular ecologist's guide to expressed sequence tags. Molecular Ecology **16**:907-924.

Boulding, E.G., M. Culling, B. Glebe, P.R. Berg, S. Lien, and T. Moen. 2008. Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. Heredity **101**:381-391.

Boyce, N.P., Z. Kabata, and L. Margolis. 1985. Investigations of the distribution, detection, and biology of *Henneguya salminicola* (Protozoa, Myxozoa), a parasite of the flesh of Pacific salmon. Canadian Technical Report of Fisheries and Aquatic Sciences **1405**, 55 p.

Buckland-Nicks, J.A., M. Gillis, and T.E. Reimchen. 2012. Neural network detected in a presumed vestigial trait: ultrastructure of the salmonid adipose fin. Proceedings of the Royal Society B: Biological Sciences **279**:553-563

California Hatchery Scientific Review Group. 2012. California Hatchery Review Report. Prepared for the U.S. Fish and Wildlife Service and Pacific States Marine Fisheries Commission. 102 p. Available from: http://swfsc.noaa.gov/publications/FED/01067.pdf; appendices at: http://cahatcheryreview.com/reports/

Campbell, N.R., and S.R. Narum. 2008. Identification of novel single-nucleotide polymorphisms in Chinook salmon and variation among life history types. Transactions of the American Fisheries Society **137**:96-106.

Campbell, N.R., K. Overturf, and S.R. Narum. 2009. Characterization of 22 novel single nucleotide polymorphism markers in steelhead and rainbow trout. Molecular Ecology Resources **9**:318-322.

Carlson, S.M., and T.R. Seamons. 2008. A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. Evolutionary Applications **1**:222-238.

Cavalli-Sforza, L.L., and A.W. Edwards. 1967. Phylogenetic analysis. Models and estimation procedures. American Journal of Human Genetics **19**:233-57.

Cederholm, C.J., M.D. Kunze, T. Murota, and A. Sibatani. 1999. Pacific salmon carcasses: essential contributions of nutrients and energy for aquatic and terrestrial ecosystems. Fisheries **24**:6-15.

Clark, A.G., M.J. Hubisz, C.D. Bustamante, S.H. Williamson, and R. Nielsen. 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Research **15**:1496-1502.

Claytor, R., and H. MacCrimmon. 1988. Morphometric and meristic variability among North American Atlantic salmon (*Salmo salar*). Canadian Journal of Zoology **66**:310-317.

Clemento, A.J., E.C. Anderson, D. Boughton, D. Girman, and J.C. Garza. 2009. Population genetic structure and ancestry of *Oncorhynchus mykiss* populations above and below dams in south-central California. Conservation Genetics **10**:1321-1336.

Clemento, A.J., A. Abadía-Cardoso, H.A. Starks, and J.C. Garza. 2011. Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, *Oncorhynchus tshawytscha*. Molecular Ecology Resources **11**(Suppl. 1):50-66.

Clemento, A.J., E.D. Crandall, J.C. Garza, and E.C. Anderson. 2013. Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) in the California Current Large Marine Ecosystem. Fishery Bulletin *In review*.

Cole, J. 2000. Coastal sea surface temperature and coho salmon production off the northwest United States. Fisheries Oceanography **9**:1-16.

Conrad, J.L., E.A. Gilbert-Horvath, and J.C. Garza. 2013. Genetic and phenotypic effects on reproductive outcomes for captively-reared coho salmon, *Oncorhynchus kisutch*. Aquaculture **404-405**:95-104.

Cook, R. C. 1982. Stock identification of sockeye salmon (*Oncorhynchus nerka*) with scale pattern recognition. Canadian Journal of Fisheries and Aquatic Sciences **39**:611-617.

Cornuet, J., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics **153**:1989-2000.

Cronin, M. A., W. J. Spearman, and R. L. Wilmot. 1993. Mitochondrial DNA variation in Chinook (Oncorhynchus tshawytscha) and chum salmon (O. keta) detected by restriction enzyme analysis of polymerase chain reaction (PCR) products. Canadian Journal of Fisheries and Aquatic Sciences **50**:708-715.

Department of Water Resources. 2004. The effects of the Feather River Hatchery on naturally spawning salmonids. SP-F9 Final Report. Available from: http://orovillerelicensing.water.ca.gov/wg-reports_envir.html

DeWoody, J.A. 2005. Molecular approaches to the study of parentage, relatedness and fitness: practical applications for wild animals. Journal of Wildlife Management **69**:1400-1418.

Elliott, D.G., and R.J. Pascho. 2001. Evidence that coded-wire-tagging procedures can enhance transmission of *Renibacterium salmoninarum* in Chinook salmon. Journal of Aquatic Animal Health **13**:181-193.

Federal Register. 1990. Endangered and Threatened Wildlife and Plants; Listing of the Sacramento River Winter-run Chinook Salmon as Threatened. Federal Register **55**: 4962. Office of the Federal Register, National Archives and Records Administration (NARA), College Park, MD.

Federal Register. 1999. Endangered and Threatened Species; Threatened Status for Three Chinook Salmon Evolutionarily Significant Units (ESUs) in Washington and Oregon, and Endangered Status for One Chinook Salmon ESU in Washington. Federal Register **64**:14308-14328. Office of the Federal Register, NARA, College Park, MD.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author at: http://evolution.genetics.washington.edu/phylip.html. Department of Genome Sciences, University of Washington, Seattle.

Field, J.C. and R.C. Francis. 2006. Considering ecosystem-based fisheries management in the California Current. Marine Policy **30**:552-569.

Fisher, F.W. 1994. Past and present status of Central Valley Chinook salmon. Conservation Biology **8**:870-873.

Flannery, B.G., J.K. Wenburg, and A.J. Gharrett. 2007. Variation of amplified fragment length polymorphisms in Yukon River chum salmon: population structure and application to mixed-stock analysis. Transactions of the American Fisheries Society **136**:911-925.

Fleming, I.A. and M.R. Gross. 1994. Breeding competition in a Pacific Salmon (coho: *Oncorhynchus kisutch*): measures of natural and sexual selection. Evolution **48**:637-657.

Fournier, D.A., T.D. Beacham, B.E. Riddell, and C.A. Busack. 1984. Estimating stock composition in mixed stock fisheries using morphometric, meristic, and electrophoretic characteristics. Canadian Journal of Fisheries and Aquatic Sciences **41**:400-408.

Fry, D.H., and A. Petrovich. 1970. King salmon spawning stocks of the California Central Valley, 1953-1969. California Department of Fish and Game, Anadromous Fish Administrative Report **21**.

Garza, J.C., and E.C. Anderson. 2007. Large scale parentage inference as an alternative to coded-wire tags for salmon fishery management. In: PSC Genetic Stock Identification Workshop (May and September 2007): Logistics Workgroup final report and recommendations, p. 48-55. Pacific Salmon Commission, Vancouver, British Columbia, Canada. Available from: http://www.psc.org/info_genetic_stock_id.htm

Garza, J.C., S.M. Blankenship, C. Lemaire, and G. Charrier. 2008. Genetic population structure of Chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. 82p. Available from: http://www.yubaaccordrmt.com/Studies%20%20Reports/CVChinDraftFinalReport-Garza.pdf

Glover, K.A., M.M. Hansen, S. Lien, T.D. Als, B. Høyheim, and Ø. Skaala. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genomics **11**:2-12.

Greenwood, P.J., and P.H. Harvey. 1982. The natal and breeding dispersal of birds. Annual Review of Ecology and Systematics **13**:1-21.

Groot, C., and L. Margolis. 1991. *Pacific Salmon Life Histories*. Vancouver: UBC Press.

Habicht, C., S. Sharr, D. Evans, and J.E. Seeb. 1998. Coded wire tag placement affects homing ability of pink salmon. Transactions of the American Fisheries Society **127**:652-657.

Hankin, D.G., J.H. Clark, R.B. Deriso, J.C. Garza, G.S. Morishima, B.E. Riddell, C. Schwarz, and J.B. Scott. 2005. Report of the expert panel on the future of the coded wire tag recovery program for Pacific salmon. Pacific Salmon Commission Technical Report **18**: 230p. Available from: http://www.psc.org/pubs/psctr18.pdf

Hankin, D.G., J. Fitzgibbons, and Y. Chen. 2009. Unnatural random mating policies select for younger age at maturity in hatchery Chinook salmon (*Oncorhynchus tshawytscha*) populations. Canadian Journal of Fisheries and Aquatic Sciences **66**:1505-1521.

Hard, J.J., M.R. Gross, M. Heino, R. Hilborn, R.G. Kope, R. Law, and J.D. Reynolds. 2008. Evolutionary consequences of fishing and their implications for salmon. Evolutionary Applications. Special Issue: Evolutionary perspectives on salmonid conservation and management **1**:388-408.

Hare, S.R., N.J. Mantua, and R.C. Francis. 1999. Inverse production regimes: Alaska and West Coast Pacific Salmon. Fisheries **24**:6-14.

Hare, S.R., and Mantua, N.J. 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. Progress in Oceanography **47**:103-145.

Hauser, L., and G.R. Carvalho. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries **9**:333-362.

Hauser, L., M. Baird, R. Hilborn, L.W. Seeb, and J.E. Seeb. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Molecular Ecology Resources **11**(Suppl. 1):150-161.

Heath, D.D., C.W. Fox, and J.W. Heath. 1999. Maternal effects on offspring size: variation through early development of Chinook salmon. Evolution **53**:1605-1611.

Hedrick, P.W., D. Hedgecock, and S. Hamelberg. 1995. Effective population size in winter-run Chinook salmon. Conservation Biology **9**:615-624.

Helfield, J.M., and R.J. Naiman. 2006. Keystone interactions: salmon and bear in riparian forests of Alaska. Ecosystems **9**:167-180.

Hilborn R. 1992. Hatcheries and the future of salmon in the Northwest. Fisheries **17**: 5-8.

Hinke, J.T., G.M. Watters, G.W. Boehlert, and P. Zedonis. 2005. Ocean habitat use by Chinook salmon in coastal waters of Oregon and California. Marine Ecology Progress Series **285**:181-192.

Holtby, L.B., B.C. Andersen, and R.K. Kadowaki. 1990. Importance of smolt size and early ocean growth to interannual variability in marine survival of coho salmon (*Oncorhynchus kisutch*). Canadian Journal of Fisheries and Aquatic Sciences **47**:2181-2194.

Hoskinson, R.L., and L.D. Mech. 1976. White-tailed deer migration and its role in wolf predation. The Journal of Wildlife Management **40**:429-441.

Huson, D.H., D.C. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics **8**:460.

Hyun S., R. Sharma, J.K. Carlile, J.G. Norris, G. Brown, R.J. Briscoe, and D. Dobson. 2012. Integrated forecasts of fall Chinook salmon returns to the Pacific northwest. Fisheries Research **125-126**:306-317.

Jefferts, K.B., P. Bergman, and H. Fiscus. 1963. A coded wire identification system for macro-organisms. Nature **198**:460-462.

Jenkins, S., and N. Gibson. 2001. High-throughput SNP genotyping. Comparative and Functional Genomics **3**:57-66.

Johnson, K.J. 2004. Regional overview of coded wire tagging of anadromous salmon and steelhead in northwest America. Regional Mark Processing Center, Pacific States Marine Fisheries Commission, 205 SE Spokane Street, Suite 100, Portland, Oregon 97202-6413, USA. Available from: http://www.rmpc.org/files/RegionalOverviewProfPaper-30May04.pdf

Jones, G.P., M.J. Milicich, M.J. Emslie, and C. Lunow. 1999. Self-recruitment in a coral reef fish population. Nature **402**:802-804.

Jones, A.G., C.M. Small, K.A. Paczolt, and N.L. Ratterman. 2010. A practical guide to methods of parentage analysis. Molecular Ecology Resources **10**:6-30.

Kalinowski, S.T. 2004. Genetic polymorphism and mixed-stock fisheries analysis. Canadian Journal of Fisheries and Aquatic Sciences **61**:1075-1082.

Keenleyside, M.H.A., and H.M.C. Dupuis. 1988. Courtship and spawning competition in pink salmon (*Oncorhynchus gorbuscha*). Canadian Journal of Zoology **66**:262-265.

Kincaid, H.L. 1983. Inbreeding in fish populations used for aquaculture. Aquaculture **33**:215-227.

Konovalov, D.A., C. Manning, and M.T. Henshaw. 2004. KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. Molecular Ecology Notes **4**:779-782.

Kruse, G.H. 1998. Salmon run failures in 1997-1998: A link to anomalous ocean conditions? Alaska Fishery Research Bulletin **5**:55-63.

Landry, C., D. Garant, P. Duchesne, and L. Bernatchez. 2001. Good genes as heterozygosity: the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). Proceedings of the Royal Society Biological Sciences **268**:1279-1285.

Larson, W.A., F.M. Utter, K.W. Myers, W.D. Templin, J.E. Seeb, C.M. Guthrie III, A.V. Bugaev, and L.W. Seeb. 2012. Single-nucleotide polymorphisms reveal distribution and migration of Chinook salmon (*Oncorhynchus tshawytscha*) in the Bering Sea and North Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences **70**:128-141.

Levin P.S., R.W. Zabel, and J.G. Williams. 2001. The road to extinction is paved with good intentions: negative association of fish hatcheries with threatened salmon. Proceedings of the Royal Society Biological Sciences **268**:1153-1158.

Lindley, S.T., C.B. Grimes, M.S. Mohr, W. Peterson, J. Stein, J.T. Anderson, L.W. Botsford, D.L. Bottom, C.A. Busack, T.K. Collier, J. Ferguson, J.C. Garza, A.M. Grover, D.G. Hankin, R.G. Kope, P.W. Lawson, A. Low, R.B. MacFarlane, K. Moore, M. Palmer-Zwahlen, F.B. Schwing, J. Smith, C. Tracy, R. Webb, B.K. Wells, and T.H. Williams. 2009. What caused the Sacramento River fall Chinook stock collapse? NOAA Technical Memorandum NMFS-SWFSC-447, 121p.

Louis, E.J., and E.R. Dempster. 1987. An exact test for Hardy-Weinberg and multiple alleles. Biometrics **43**:805-811.

Lukacs, P.M., and K.P. Burnham. 2005. Review of capture-recapture methods applicable to noninvasive genetic sampling. Molecular Ecology **14**:3909-3919.

Lynch and Walsh 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland MA: Sinauer Associates, Inc.

Mantua, N.J., and S.R. Hare. 2002. The Pacific decadal oscillation. Journal of Oceanography **58**:35-44.

Mantua, N.J., S.R. Hare, Y. Zhang, J.M. Wallace, and R.C. Francis. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society **78**:1069-1079.

Metcalfe, J.D., and G.P. Arnold. 1997. Tracking fish with electronic tags. Nature **387**:665-666.

Michael, J. 2010. Employment impacts of California salmon fishery closures in 2008 and 2009. Business Forecasting Center, University of the Pacific, 3601 Pacific Avenue, Stockton, CA 95211. Available from: http://forecast.pacific.edu/BFC%20salmon%20jobs.pdf

Miller, K.M., R.E. Withler, and T.D. Beacham. 1996. Stock identification of coho salmon (*Oncorhynchus kisutch*) using minisatellite DNA variation. Canadian Journal of Fisheries and Aquatic Sciences **53**:181-195.

Milner, G.B., D.J. Teel, F.M. Utter, and G.A. Winans. 1985. A genetic method of stock identification in mixed populations of Pacific salmon, *Oncorhynchus spp.* Marine Fisheries Review **47**:1-8.

Moen, T., B. Hayes, M. Baranski, *et al.* 2008. A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. BMC Genomics **9**:223.

Moore, W.S., and R.A. Dolbeer. 1989. The use of banding recovery data to estimate dispersal rates and gene flow in avian species: case studies in the red-winged blackbird and common grackle. The Condor **91**:242-253.

Moran, P. 2002. Current conservation genetics: building an ecological approach to the synthesis of molecular and quantitative genetic methods. Ecology of Freshwater Fishes **11**:30-55.

Moran, P., D.J. Teel, M.A. Banks, T.D. Beacham, M.R. Bellinger, S.M. Blankenship, J.R. Candy, J.C. Garza, J.E. Hess, S.R. Narum, L.W. Seeb, W.D. Templin, C.G. Wallace, and C.T. Smith. 2013. Divergent life-history races do not represent Chinook salmon coast-wide: the importance of scale in Quaternary biogeography. Canadian Journal of Fisheries and Aquatic Sciences **70**:415-435.

Morin, P., G. Luikart, and R. Wayne. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution **19**:208-216.

Morishima, G.S. 2004. In a nutshell: coded wire tags and the Pacific Salmon Commissions fishery regimes for Chinook and southern coho salmon. Report to the Pacific Salmon Commission. Available from: http://www.psc.org/pubs/CWT/
CWTWebPapers/SpecificForWorkshop/morishima2004.pdf

Morrison, J., and D. Zajac. 1987. Histologic effect of coded wire tagging in chum salmon. North American Journal of Fisheries Management **7**:439-441.

Mowat, G., and D. Paetkau. 2002. Estimating marten *Martes americana* population size using hair capture and genetic tagging. Wildlife Biology **8**:201-209.

Myers, J.M., R.G. Kope, G.J. Bryant, D. Teel, L.J. Lierheimer, T.C. Wainwright, W.S. Grant, F.W. Waknitz, K. Neely, S.T. Lindley, and R.S. Waples. 1998. Status review of chinook salmon from Washington, Idaho, Oregon, and California. NOAA Technical Memorandum NMFS, NMFS-NWFSC-35, 7p.

Narum, S.R., M.A. Banks, T.D. Beacham, M.R. Bellinger, M.R. Campbell, J. Dekoning, A. Elz, C.M. Guthrie, C. Kozfkay, K.M. Miller, P. Moran, R. Phillips, L.W. Seeb, C.T. Smith, K. Warheit, S.F. Young, and J.C. Garza. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Molecular Ecology **17**:3464-3477.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89**:583-590.

Nickell, W.P. 1968. Return of northern migrants to tropical winter quarters and banded birds recovered in the United States. Bird-Banding **39**:107-116.

Nosil, P., D. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Molecular Ecology **18**:375-402.

Novembre, J., T. Johnson, K. Bryc, Z. Kutalik, A. Boyko, A. Auton, A. Indap, K. King, S. Bergmann, M. Nelson, M. Stephens, and C.D. Bustamante. 2008. Genes mirror geography within Europe. Nature **456**:98-101.

O'Farrell, M.R., M.S. Mohr, A.M. Grover, and W.H. Satterthwaite. 2012. Sacramento River winter Chinook cohort reconstruction: analysis of ocean fishery impacts. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-491, 74p.

Olsén H. 1998. Present knowledge of kin discrimination in salmonids. Genetica **104**:295-299.

Ormiston, B.G. 1985. Effects of a subminiature radio collar on activity of free-living white-footed mice. Canadian Journal of Zoology **63**:733-735.

Palsbøll, P.J., J. Allen, M. Bérubé, P.J. Clapham, T.P. Feddersen, P.S. Hammond, R.R. Hudson, H. Jørgensen, S. Katona, A.H. Larsen, F. Larsen, J. Lien, D.K. Mattila, J. Sigurjónsson, R. Sears, T. Smith, R. Sponer, P. Stevick, and N. Oien. 1997. Genetic tagging of humpback whales. Nature **388**:767-769.

Palsbøll, P.J. 1999. Genetic tagging: contemporary molecular ecology. Biological Journal of the Linnean Society **68**:3-22.

Park, S.D.E. 2001. Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. thesis: University of Dublin

Pauley, G.B., 1967. Prespawning adult salmon mortality associated with a fungus of the genus *Dermocystidium.* Journal of the Fisheries Research Board of Canada **24**:843-848.

Pearse, D.E., C.M. Eckerman, F.J. Janzen, and J.C. Avise. 2001. A genetic analogue of mark-recapture methods for estimating population size: an approach based on molecular parentage assessments. Molecular Ecology **10**:2711-2718.

Pearse, D.E., C.J. Donohoe, and J.C. Garza. 2007. Population genetics of steelhead (*Oncorhynchus mykiss*) in the Klamath River. Environmental Biology of Fishes **80**:377-387.

Pella, J., and M. Masuda. 2000. Bayesian methods for analysis of stock mixtures from genetic characters. Fishery Bulletin **99**:151-167.

Pemberton, J.M. 2008. Wild pedigrees: the way forward. Proceedings of the Royal Society Biological Sciences **275**:613-621.

Queller, D.C., and K.F. Goodnight. 1989. Estimating relatedness using genetic markers. Evolution **43**:258-275.

Quinn, T.P., and C.A. Busack. 1985. Chemosensory recognition of siblings in juvenile coho salmon (*Oncorhynchus kisutch*). Animal Behaviour **33**:51-56.

Quinn, T.P., and C.J. Foote. 1994. The effects of body size and sexual dimorphism on the reproductive behaviour of sockeye salmon, *Oncorhynchus nerka*. Animal Behaviour **48**:751-761.

Quinn, T.P., and N.P. Peterson. 1996. The influence of habitat complexity and fish size on over-winter survival and growth of individually marked juvenile coho salmon (*Oncorhynchus kisutch*) in Big Beef Creek, Washington. Canadian Journal of Fisheries and Aquatic Sciences **53**:1555-1564.

Quinn, T.P. 2005. *The Behavior and Ecology of Pacific Salmon and Trout*. Vancouver: UBC Press.

Rajakaruna, R.S., J.A. Brown, K.H. Kaukinen, and K.M. Miller. 2006. Major histocompatibility complex and kin discrimination in Atlantic salmon and brook trout. Molecular Ecology **15**:4569-4575.

Rannala, B., and J.L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences USA **94**:9197-9201.

Refstie, T., and T.A. Steine. 1978. Selection experiments with salmon: genetic and environmental sources of variation in length and weight of Atlantic salmon in the freshwater phase. Aquaculture **14**:221-234.

Reimchen, T.E., and N.F. Temple. 2004. Hydrodynamic and phylogenetic aspects of the adipose fin in fishes. Canadian Journal of Zoology **82**:910-916.

Ricker, W.E. 1981. Changes in the average size and average age of Pacific Salmon. Canadian Journal of Fisheries and Aquatic Sciences **38**:1636-1656.

Rousset, F. 2008. GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources **8**:103-106.

Rozen, S., and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Bioinformatics Methods and Protocols: Methods in Molecular Biology (eds. Krawetz S, Misener S), pp. 365-386. Humana Press, Totowa, NJ. Software available from: http://fokker.wi.mit.edu/primer3/input.htm

Satterthwaite, W., M. S. Mohr, M. R. OFarrell, E. C. Anderson, M. A. Banks, S. J. Bates, M. R. Bellinger, L. A. Borgerson, E. D. Crandall, J. C. Garza, B. J. Kormos, P. W. Lawson, and M. L. Palmer-Zwahlen. 2013. Use of genetic stock identification data for comparison of the ocean spatial distribution, size-at-age, and fishery exposure of an untagged stock and its indicator: California Coastal versus Klamath River Chinook. Transactions of the American Fisheries Society *In press*.

Seamons, T.R., P. Bentzen, and T.P. Quinn. 2004. The mating system of steelhead, *Oncorhynchus mykiss*, inferred by molecular analysis of parents and progeny. Environmental Biology of Fishes **69**:333-344.

Schlotterer, C. 2004. The evolution of molecular markers - Just a matter of fashion? Nature Reviews Genetics **5**:63-69.

Seeb, L.W., A. Antonovich, M.A. Banks, T.D. Beacham, M.R. Bellinger, S.M. Blankenship, M.R. Campbell, N.A. Decovich, J.C. Garza, C.M. Guthrie III, T.A. Lundrigan, P. Moran, S.R. Narum, J.J. Stephenson, K.J. Supernault, D.J. Teel, W.D. Templin, J.K. Wenburg, S.F. Young, and C.T. Smith. 2007. Development of a standardized DNA database for Chinook salmon. Fisheries **32**:540-552.

Seeb, J.E., C.E. Pascal, R. Ramakrishnan, and L.W. Seeb. 2009. SNP genotyping by the 5'-nuclease reaction: advances in high throughput genotyping with non-model organisms. Pages 277-292 in A. Komar, editor. Methods in Molecular Biology, Single Nucleotide Polymorphisms, 2d Edition. Humana Press.

Seeb, J.E., G. Carvalho, L. Hauser, K. Naish, S. Roberts, and L.W. Seeb. 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in non-model organisms. Molecular Ecology Resources **11**(Suppl. 1):1-8.

Shaklee, J.B., and S.R. Phelps. 1990. Operation of a large-scale, multi-agency program for genetic stock identification. In Fish-marking techniques (N. C. Parker, A. E. Giorgi, R. C. Heidinger, D. B. Jeter, Jr., E. D. Prince, and G. A. Winans, eds.), p. 817-830. Proceedings of the American Fisheries Society Symposium **7**, Bethesda, MD.

Smith, C.T., R.J. Nelson, C.C. Wood, and B.F. Koop. 2001. Glacial biogeography of North American coho salmon (*Oncorhynchus kisutch*). Molecular Ecology **10**:2775-2785

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

Smith, C.T., J.E. Seeb, P. Schwenke, and L.W. Seeb. 2005b. Use of the 5'-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. Transactions of the American Fisheries Society **134**:207-217.

Smith, C.T., W.D. Templin, J.E. Seeb, and L.W. Seeb. 2005c. Single Nucleotide Polymorphisms (SNPs) provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook salmon caught in Yukon River fisheries. North American Journal of Fisheries Management **25**:944-953.

Smith, C.T., L. Park, D. VanDoornik, L.W. Seeb, and J.E. Seeb. 2006. Characterization of 19 single nucleotide polymorphism markers for coho salmon. Molecular Ecology Notes **6**:715-720.

Smith, C.T., A. Antonovich, W.D. Templin, C.M. Elfstrom, S.R. Narum, and L.W. Seeb. 2007. Impacts of marker class bias relative to locus-specific variability on population inferences in Chinook salmon: a comparison of single-nucleotide polymorphisms with short tandem repeats and allozymes. Transactions of the American Fisheries Society **136**:1674-1687.

Smith, M.J., C.E. Pascal, Z. Grauvogel, C. Habicht, J.E. Seeb, and L.W. Seeb. 2011. Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. Molecular Ecology Resources **11**(Suppl. 1):268-277.

Smouse, P.E., R.S. Waples, and J.A. Tworek. 1990. A genetic mixture analysis for use with incomplete source population data. Canadian Journal of Fisheries and Aquatic Sciences **47**:620-634.

Steele, C.A., E.C. Anderson, M.W. Ackerman, M.A. Hess, N.R. Campbell, S.R. Narum, and M.R. Campbell. 2013. A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. Canadian Journal of Fisheries and Aquatic Sciences **70**:1046-1054.

Stern, V.M., E.J. Dietrick, and A. Mueller. 1965. Improvements on self-propelled equipment for collecting, separating, and tagging mass numbers of insects in the field. Journal of Economic Entomology **58**:949-953.

Sumner, S., E. Lucas, J. Barker, and N. Isaac. 2007. Radio-tagging technology reveals extreme nest-drifting behavior in a eusocial insect. Current Biology **17**:140-145.

Taylor EB. 1990. Phenotypic correlates of life-history variation in juvenile Chinook salmon, *Oncorhynchus tshawytscha*. Journal of Animal Ecology **59**:455-468.

Taylor, E.B. 1991. A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. Aquaculture **98**:185-207.

Teel, D.J., P.A. Crane, C.M. Guthrie III, A.R. Marshall, D.M. Van Doornik, W. Templin, N.V. Varnavskaya, and L.W. Seeb. 1999. Comprehensive allozyme database discriminates Chinook salmon around the Pacific Rim. (NPAFC document 440) 25 p. Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, Alaska USA 99518.
Available from: http://www.npafc.org/new/publications/Documents/PDF%201999/440(USA).pdf

Templin, W.D., J.E. Seeb, J.R. Jasper, A.W. Barclay, and L.W. Seeb. 2011. Genetic differentiation of Alaska Chinook salmon: the missing link for migratory studies. Molecular Ecology Resources **11**(Suppl. 1):226-246.

Tessier, N., L. Bernatchez, P. Presa, and B. Angers. 1995. Gene diversity analysis of mitochondrial DNA, microsatellites and allozymes in landlocked Atlantic salmon. Journal of Fisheries Biology **47**:156-163.

Thrower, F.P., and J.J. Hard. 2009. Effects of a single event of close inbreeding on growth and survival in steelhead. Conservation Genetics **10**:1299-1307.

Unwin, M.J., and G.J. Glova. 1997. Changes in life history parameters in a naturally spawning population of chinook salmon (*Oncorhynchus tshawytscha*) associated with releases of hatchery-reared fish. Canadian Journal of Fisheries and Aquatic Sciences **54**:1235-1245.

Utter, F.M., G.B. Milner, G.Stahl, and D.J. Teel. 1989. Genetic population structure of Chinook salmon, *Oncorhynchus tshawytscha*, in the Pacific Northwest. Fishery Bulletin **87**:239-264.

Vander Haegen, G.E., H.L. Blankenship, A. Hoffmann, and D.A. Thompson. 2005. The effects of adipose fin clipping and coded wire tagging on the survival and growth of spring Chinook salmon. North American Journal of Fisheries Management **25**:1161-1170.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics, Selection, Evolution **34**:275-305.

Wang, S., J.J. Hard, and F. Utter. 2001. Salmonid inbreeding: a review. Reviews in Fish Biology and Fisheries **11**:301-319.

Wang, J.L. 2004. Sibship reconstruction from genetic data with typing errors. Genetics **166**:1963-1979.

Waples, R.S. 1991. Genetic interactions between hatchery and wild salmonids: lessons from the Pacific Northwest. Canadian Journal of Fisheries and Aquatic Sciences **48**(Suppl. 1):124-133.

Waples, R.S., D.J. Teel, J.M. Myers, and A.R. Marshall. 2004. Life-history divergence in Chinook salmon: Historic contingency and parallel evolution. Evolution **58**:386-403.

Warheit, K.I., L.W. Seeb, W.D. Templin, and J.E. Seeb. 2013. Moving GSI into the Next Decade: SNP Coordination for Pacific Salmon Treaty Fisheries. FPT 13-09. Washington Department of Fish and Wildlife. 47 + iii pp.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:358-1370.

Williamson, K.S., and B. May. 2005. Homogenization of Fall-Run Chinook salmon gene pools in the Central Valley of California, USA. North American Journal of Fisheries Management **25**:993-1009.

Willson, M.F., and K.C. Halupka. 1995. Anadromous fish as keystone species in vertebrate communities. Conservation Biology **9**:489-497.

Winkelman, A.M., and R.G. Peterson. 1994. Genetic parameters (heritabilities, dominance ratios and genetic correlations) for body weight and length of Chinook salmon after 9 and 22 months of saltwater rearing. Aquaculture **125**:31-36.

Withler, R.E., W.C. Clarke, B.E. Riddell, and H. Kreiberg. 1987. Genetic variation in freshwater survival and growth of chinook salmon (*Oncorhynchus tshawytscha*). Aquaculture **64**:85-96.

Wood, C.C., S. McKinnell, T. Mulligan, and D. Fournier. 1987. Stock Identification with the maximum-likelihood mixture model: sensitivity analysis and application to complex problems. Canadian Journal of Fisheries and Aquatic Sciences **44**:866-881.

Woods, J.G., D. Paetkau, D. Lewis, B.N. McLellan, M. Proctor, and C. Strobeck. 1999. Genetic tagging of free-ranging black and brown bears. Wildlife Society Bulletin **27**:616-627.

Wu, R., C.X. Ma, and G. Casella. 2007. *Statistical Genetics of Quantitative Traits: Linkage, Maps and QTL*. New York, NY: Springer.

Yoshiyama, R.M., E.R. Gerstung, F.W. Fisher, and P.B. Moyle. 1996. Historical and present distribution of Chinook salmon in the Central Valley drainage of California. In: Sierra Nevada Ecosystem Project: Final report to Congress, vol. III. Centers for Water and Wildland Resources, University of California, Davis, CA, p.309-361.

Yoshiyama, R.M., F.W. Fisher, and P.B. Moyle. 1998. Historical abundance and decline of Chinook salmon in the Central Valley region of California. North American Journal of Fisheries Management **18**:487-521.

Yoshiyama, R.M., P.B. Moyle, E.R. Gerstung, and F.W. Fisher. 2000. Chinook salmon in the California Central Valley: an assessment. Fisheries **25**:6-20.