Genetic and phenotypic diversity in sockeye salmon, *Oncorhynchus nerka*

Caroline Storer

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Mater of Science

University of Washington

2012

Committee:

Thomas Quinn

Steven Roberts

James Seeb

William Templin

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The research described here is the collective efforts of work in two labs, the Seeb Lab and the Roberts Lab, on two different projects. The first chapter was motivated by a need for improved genetic tools for the management of sockeye salmon specifically those populations inhabiting Bristol Bay. It was initially designed to be a large-scale high-throughput SNP discovery project describing the use of new sequencing technologies for the quick development of hundreds of new molecular markers. However, due to limitations in the sequence data SNP discovery was less successful than anticipated. Therefore, the project was adapted to include the comparison and development of methods for evaluating SNPs for creating panels of high-throughput assays that are customized for performance, research questions, and resources. With the addition of this research, this study will help to advance the field of population genetics/genomics and will have significant implications for the way we select molecular markers for ecological, evolutionary, and applied research in non-model species.

The second chapter was motivated by a desire to better understand the process of senescence in Pacific Salmon. Pilot studies at the same study site suggested a molecular signal of senescence associated with stress. This pilot work inspired the research described here providing a deeper examination of different pathways involved in senescence using molecular tools.

ACKNOWLEDGEMENTS

The author wishes to sincerely thank the School of Aquatic and Fishery Sciences for their support and input specifically co-advisors Steven Roberts and Jim Seeb and additional committee members Tom Quinn and Bill Templin. This thesis would never have been completed without the encouragement, patience, and frivolity of my family and friends. This research was made possible with funding from the Alaska Sustainable Salmon Fund under Study 45908 from the National Oceanic and Atmospheric Administration, US Department of Commerce, administered by the Alaska Department of Fish and Game. Additional funding was provided by the Bristol Bay Regional Seafood Development Association. This project was made possible with support from the Alaska Department of Fish and Game and infrastructure support from the Gordon and Betty Moore Foundation. All samples were acquired through the generosity of Washington Department of Fish and Wildlife, Alaska Department of Fish and Game, and other contributors to the International Program for Salmon Ecological Genetics.

Chapter I: Rank and order: evaluating the performance of SNPs for individual assignment in a non-model organism

Abstract

Single nucleotide polymorphisms are valuable tools for ecological and evolutionary studies. In non-model species, the use of SNPs has been limited by the number of markers available. However, new technologies and decreasing technology costs have facilitated the discovery of a constantly increasing number of SNPs. With hundreds or thousands of SNPs potentially available, there is interest in comparing and developing methods for evaluating SNPs to create panels of high-throughput assays that are customized for performance, research questions, and resources. Here we use five different methods to rank 43 new SNPs and 71 previously published loci for sockeye salmon: F_{ST}, informativeness (I_n), average contribution to principal components (LC), and the locus-ranking programs BELS and WHICHLOCI. We then tested the performance of these different ranking methods by creating 48- and 96-SNP panels of the top-ranked loci for each method and used empirical and simulated data to obtain the probability of assigning individuals to the correct population using each panel. All 96-SNP panels performed similarly and better than the 48-SNP panels except for the 96-SNP BELS panel. Among the 48-SNP panels, panels created from F_{ST}, I_n, and LC ranks performed better than panels formed using the top-ranked loci from the programs BELS and WHICHLOCI. The application of ranking methods to optimize panel performance will become more important as more high-throughput assays become available.

Introduction

Molecular markers are widely used in the fields of ecology, evolution, and resource management (Sagarin et al., 2009; Schlotterer, 2004). Among the many types of markers, single nucleotide polymorphisms (SNPs) have received increased attention due to their potential value for the study of non-model organisms (Garvin et al., 2010; Seeb et al., 2011a). Their use in ecology and conservation has been demonstrated for several species including mammals, birds, fish, and insects (for example Emerson et al., 2010; Olsen et al., 2011; Quintela et al., 2010). Additionally, SNPs are abundant throughout the genome; some SNP technologies are robust and automated, enabling accurate and high-throughput genotyping of thousands of individuals (Morin et al., 2004; Seeb et al., 2011a).

The use of high-throughput SNP panels for the study of non-model organisms has primarily been limited by the cost and difficulties of discovering new SNPs, and consequently, the number of available assays has been low or nonexistent for many species. However, technological advances and innovative methodologies are enabling rapid SNP discovery (Geraldes et al., 2011; Hohenlohe et al., 2011). With decreasing technology costs (Martinez & Nelson, 2010), SNP discovery projects are becoming more common, and the number novel SNPs potentially available for conversion to high-throughput assays is rapidly growing (e.g., Hohenlohe et al., 2011; Karlsson et al., 2011).

Population studies in non-model organisms that used high-throughput assays for SNPs typically went through an initial discovery phase where every new assay was precious and every available marker was used (e.g., Morin et al., 2010; Campbell & Narum, 2011). Increasingly, many researchers are interested in developing SNP panels of 48 or more that are tailored to their specific research question (Dai et al., 2008) and study system (Morin et al., 2010). Panels of SNPs can be developed and optimized for laboratory performance (i.e., genotypes are easily distinguishable and reproducible), for genotyping platform, and for power to resolve population structure (Helyar et al., 2011). One approach for identifying loci with high information content for a panel has been to evaluate their ability to elucidate population structure (Winans et al., 2004). Additionally, locus selection programs such as WHICHLOCI (Banks et al., 2003) and BELS (Bromaghin, 2008) are used to rank and evaluate loci based on their performance for individual assignment and in some cases mixed stock analysis (e.g., BELS; Glover et al., 2010). However, there is some concern that upward bias in a SNP's rank can be introduced when using these programs with high-resolution loci (Anderson, 2010). Upward bias is potentially introduced because high-resolution loci are often both discovered and evaluated using the same data set. Although there is currently no consensus on how to rank molecular markers, especially SNPs, ranking and evaluating a SNP's value for a panel will be of increasing importance in ecology and resource management as the number of high-throughput assays continues to grow.

In sockeye salmon (*Oncorhynchus nerka*) this is already the case. At present, a limited set of 45 SNPs provides insight into life history, migration, and harvest (Gomez-Uchida et al., 2011; McGlauflin et al., 2011; Smith et al., 2011). However, the cultural and economic importance of this species across the Pacific Rim has increased demand for resolving power and created a need for more SNPs and higher resolution SNPs to increase population resolution. In

the Pacific Northwest of the U.S., where some stocks are currently listed for protection under the Endangered Species Act, more SNPs are needed to improve resolution of stock structure and provide new options for conservation and management (Campbell & Narum, 2011). In Bristol Bay, Alaska, the location of the world's largest fisheries for sockeye salmon, stakeholders seek improved SNP panels to better differentiate among stocks (c.f. Smith et al., 2011; Baker et al., 2009; Seeb et al., 2000). SNPs are also increasingly used for unraveling the complexity of distribution and migration patterns on the high seas (Seeb et al., 2011b; Habicht et al., 2010). New SNPs and ranking methods will be important for answering these questions and for the management of this valuable resource.

Our objective was to develop SNP panels that could provide improved resolution of sockeye salmon populations inhabiting Bristol Bay as well as provide additional information for studies of migration of mixed populations (Habicht et al., 2010). Here we both develop new high-throughput SNP assays for sockeye salmon and explore different ranking methods for these and all other SNP assays commonly in use. We successfully developed 5'-nuclease assays for 43 new SNP loci using next generation sequence (NGS) data and high resolution melt analysis (HRMA; Wu et al., 2008; Seeb et al., 2009). These new assays increase the number of published markers for sockeye salmon to well over 100. Additionally, we explore five different ranking methods to evaluate all of these loci: locus-specific values of F_{ST} (Weir & Cockerham, 1984), informativeness (I_n; Rosenberg et al., 2003), average contribution of a locus to principal components (LC), and locus-ranks from the programs BELS (Bromaghin, 2008) and WHICHLOCI (Banks et al., 2003). The ranks from each method were used to create 48- and 96-SNP panels to take advantage of base 48 array or multiplex platforms commonly in use (e.g., Seeb et al., 2009; Perkel, 2008). Panels were then tested for performance using empirical and simulated datasets (Rosenberg, 2005). All 96-SNP panels performed similarly better than the 48-SNP panels except for the BELS panel. Among the 48-SNP panels, panels created from F_{ST}, I_n, and LC ranks performed better than panels using the top-ranked loci from the programs BELS and WHICHLOCI. As more SNPs become available, the differences between methods may increase and have a greater impact on panel performance, warranting careful exploration of locus ranking and evaluation.

Materials and Methods

SNP discovery

Discovery methods were iterative and adapted for different transcriptome datasets as they emerged from our laboratory. First, primers were selected directly from chum salmon (*O. keta*) 454 assemblies (Seeb et al., 2011c). Additional SNP primers were selected from SOLiD sequence assemblies from sockeye salmon (Everett et al., 2011). These latter sequences originate from 10 fish from five locations (Figure 1.1 red circles; Table 1.1).

Primers were designed and tested for PCR amplification of a single product on a single pooled sample of DNA. Successful primers were then used to screen individuals for SNPs using HRMA as in McGlauflin et al. (2010). HRMA was performed following the manufacturer's instructions on Lightcycler 480 (Roche Diagnostics) platform using eight test fish from each of 24 locations (192 fish total; Figure 1.1 blue circles; Table 1.1). These locations were chosen to focus upon Bristol Bay populations and also include a few representatives from the eastern and western Pacific Ocean.

Candidate SNPs that were successfully detected using HRMA were selected for Sanger sequencing. Sequences where the identity of the SNP was confirmed by the presence of at least two genotypes were used for designing primers and probes for the 5'-nuclease assays. As a final validation step, each assay was then tested by genotyping the same panel of 192 fish used for HRMA. Assays that did not perform well or where the SNP deviated from Hardy-Weinberg expectations (HWE) were discarded (HWE was tested on a subset of populations for which we possessed additional samples of (N = 61-95)). The Sanger sequences used for 5'-nuclease assay design were used to annotate validated markers using the NCBI sequence database and Blastx. Only assays where the most similar sequence hit had an e-value less than 10^{-10} were annotated.

SNP assessment

Six pairs of population samples (hereafter referred to as assessment populations) were chosen from throughout the species' range to assess within and among region variability (Figure 1.1 green diamonds; Table 1.1). All fish from the 12 assessment populations were genotyped at 114 nuclear loci (Appendix A) using 5'-nuclease assays (Seeb et al., 2009). These SNPs included the 43 new SNPs described in this paper, 68 previously published SNPs for sockeye salmon (Campbell & Narum, 2011; Habicht et al., 2010; Elfstrom et al., 2006; Smith et al., 2005)

and three unpublished markers from the Department of Fisheries and Oceans Canada (Molecular Genetics Laboratory, Pacific Biological Station, Department of Fisheries and Oceans Canada).

Tissues (heart, liver, fin, or axillary process) or genomic DNA were obtained from archived samples at the University of Washington (UW), the Alaska Department of Fish and Game (ADF&G), and the Washington Department of Fish and Wildlife (WDFW). Genomic DNA was extracted as necessary using the DNeasy96 Blood and Tissue Kit (QIAGEN, USA).

The markers were first assessed using standard population genetic indices using the 12 assessment populations (Table 1.1) as follows. Populations were tested for deviations from HWE at each locus using chi-square tests as implemented in GenAlEx 6.2 (Peakall & Smouse, 2006). All critical values were corrected for multiple comparisons using a sequential Bonferroni correction (Rice, 1989). Allelic richness was calculated for each locus in each population using FSTAT v.2.9.3.2 (Goudet, 2005) to look for effects of ascertainment bias. Differences in average allelic richness among locations were tested for significance with an ANOVA. Linkage disequilibrium was tested in each collection for each pair of SNPs using Genepop 4 (Rousset, 2008). To check for genotyping error, 8% of each collection was genotyped again.

Population differentiation was measured as F_{ST} (Weir & Cockerham, 1984) at each locus using Genepop 4 and between population pairs across all loci using Arlequin 3.5 (Excoffier & Lischer, 2010). A principal coordinate analysis with six coordinates was performed in GenAlEx to visualize the genetic relationship among populations. Arlequin was also used to detect loci under directional selection (Excoffier et al., 2009) across the entire range using the hierarchical island model with six regions (Table 1.1), 20,000 simulations, 100 demes, and 50 groups. Detection of loci under directional selection is based on Beaumont and Nichols original work using heterozygosity and high differentiation to identify outlier loci as candidates for directional selection (Beaumont & Nichols, 1996). The value of loci detected to be under directional selection (outlier loci) was investigated by removing these loci from the data set and then re-measuring genetic differentiation between populations. Significance of differences in genetic differentiation measured with outlier loci and without outlier loci was tested using a Mantel test.

SNP ranking

Each locus was ranked according to five measures: F_{ST} (Weir & Cockerham, 1984), informativeness as calculated by Rosenberg (I_n; Rosenberg et al., 2003), average contribution of a locus to principal component analysis (LC), BELS ranking (Bromaghin, 2008), and WHICHLOCI (Banks et al., 2003). We additionally considered the ranking approach GAFS of Topchy et al. (2004); GAFS was not implemented because of its similarity to BELS and computational costs (Helyar et al., 2011). Each method used is summarized in Table 1.2. F_{ST}, LC, and I_n are all measures of genetic diversity based on differences in allele frequencies observed at a locus, while BELS and WHICHLOCI are scores based on maximizing the likelihood of assigning a genotype to the correct population. F_{ST} and LC are essentially measures of genetic diversity. Informativeness (I_n) has been shown to be correlated with F_{ST} by Rosenberg et al. (2003). Informativness's relationship to LC was determined using a Spearman's rank correlation. The LC was determined using a multivariate locus comparison method developed by Moazami-Goudarzi and Laloë (2002) and implemented in S-Plus (MathSoft, Inc, 2000). Here, locus contribution was determined for the first five principal components.

BELS and WHICHLOCI provide each locus a rank based on the accuracy of individual assignment for that locus and the value lost when the locus is removed from the panel in a jackknife fashion. Loci that result in the greatest loss in individual assignment performance when removed receive the highest score. Both of these locus-ranking programs were run with resampling for a simulated population size of 200 individuals and with 250 iterations. No critical population was defined. In WHICHLOCI minimum correct assignment was set at 95.0%. In BELS, the performance measure was designated to maximize mean individual assignment accuracy for 100% correct assignment. For BELS, the role of locus input order was explored by running the analyses with four different locus orders: alphabetical, reverse alphabetical, and two randomly generated locus orders. Differences in locus ranks for each input order were tested in a pairwise fashion using the Wilcoxon Signed Rank test.

Initially, each locus was ranked using all individuals available (full set) for the twelve SNP assessment populations (Table 1.1). However, to reduce the potential for upward bias introduced when loci are ranked and assessed using the same individuals, Anderson's Simple Training and Holdout method (Anderson, 2010) was implemented. Half of each assessment population was randomly selected for locus ranking (training set). For odd numbered population

sizes the extra individual was assigned to the training set. The remaining individuals (holdout set) were reserved for panel testing. Significance of differences in locus ranks using the full population set and the training population set were tested using the Wilcoxon signed rank test.

Panel testing

SNP panels were designed to assess the value of increasing the number of markers included in a panel and to evaluate the different measures for ranking SNPs using the 12 assessment populations. Two panel sizes were selected, 48 and 96 SNPs, to test for differences in resolving power when the number of markers was increased. These panel sizes represent the capabilities of high-throughput genotyping platforms commonly in use (e.g., Perkel, 2008).

Using the training set, five 48-SNP and five 96-SNP panels were created from the top ranked loci for each locus measure. Additionally, a sixth panel of top ranked loci based on their average rank and a seventh panel of randomly selected loci were created for both the 48- and 96-SNP panels. Each panel was tested for performance with two different methods. Using the program ONCOR (Kalinowski et al., 2008), assignment tests were performed assigning holdout set individuals from each assessment population (Table 1.1) to a baseline of the training set individuals that had been used for SNP ranking. Since the origin of assigned individuals was known, the probability of assignment to the population of origin was reported for assignment accuracy.

The second method used to assess panel performance was a simulation of individual assignment described by Rosenberg (2005) as implemented by Ackerman et al. (2011). These simulations use the allele frequencies for user-described populations to assign a simulated individual back to the correct population and report the probability that this assignment is correct. Here individuals were simulated using allele frequencies from holdout set individuals for each population. For each panel, individual assignment was simulated 500 times with 1000 individuals in R, and the frequency of correct assignment (f_{ORCA} ; Rosenberg, 2005) was reported. Differences in panel performance for both assessment methods were tested for using an ANOVA and the post hoc Tukey's Honestly Significant Difference test ($\alpha = 0.01$).

In addition to panel testing, we examined the value of using the full set of loci and the change in assignment accuracy with decreasing panel size after the subsequent removal of loci. Beginning with the full set of 110 polymorphic loci, ONCOR was used to determine probability

of correct assignment similarly to 96- and 48- SNP panel assessment. Loci were then excluded five at a time by lowest average rank (Appendix A) until only the five top ranked loci remained for individual assignment.

Results

SNP Discovery

We developed 5'-nuclease assays for SNP genotyping of sockeye salmon from both ascertainment sources: chum salmon contigs originating from 454 assemblies (Seeb et al., 2011c) and sockeye salmon contigs originating from SOLiD assemblies (Everett et al., 2011). Over 1800 potential primers were initially tested; of these, only 515 passed the initial PCR test (Table 1.3). This test ensured that PCR amplification would occur and that only one product would amplify. Unsuccessful amplification could be a result of the primer annealing on or near intron-exon boundaries. Using HRMA, templates that did not contain a putative SNP or that contained multiple polymorphisms were eliminated. Multiple polymorphisms in the same template were attributed to paralogous sequence variation, known to be problematic in tetraploid-origin salmonids (Seeb et al., 2011c). Putative SNPs derived from SOLiD sequence had a failure rate three-fold of that observed in putative SNPs derived from 454 sequence (Table 1.3). Many of the remaining 148 putative SNPs were polymorphic in the majority of the test populations, and resequencing confirmed the identity of at least two genotypes in 93 of these. We attempted to design 5'-nuclease assays for all 93; only 43 had differentiable genotypes that met HWE (Table 1.3).

Where possible, validated SNPs were annotated to an NCBI sequence with greatest similarity over 10^{-10} (Appendix B). However, several validated SNPs could not be annotated due to low sequence similarity. These sequences likely represent novel genes with no annotated reference in public databases.

SNP assessment

Each fish from the 12 populations used for SNP assessment was genotyped at all 114 nuclear loci (Appendix A). These were all of the 5'-nuclease assays for sockeye salmon, with reliable laboratory performance, that were available at the time.

Re-genotyping discrepancies were less than 1% in all populations. Individuals missing genotypes at more than 10% of the loci were excluded from analyses. Sample sizes reported in Table 1.1 are all individuals included in post-genotyping analyses after removal of individuals with missing information. Poor tissue quality, which hampers genotyping ability due to the degradation of DNA, was the most likely cause for the low genotyping success in some fish. Four loci were monomorphic in all 12 populations and were removed from subsequent analyses (Appendix A). All remaining 110 loci were retained in the data set.

In four collections, there were deviations from HWE at a single locus after correction for multiple tests: Ozernaya River at *One_zP3b*, Damdochax Creek at *One_U1202-105*, Pick Creek at *One_Tf_ex3-182*, and Baker Lake at *One_ U1102-220*. In each of these collections there was a rare homozygote genotype at each of these loci and average minor allele frequencies less than 0.03 with one exception. There was an excess of heterozygotes in the Baker Lake collection for *One_U1102-220*. Mean allelic richness varied across locations and ranged from 1.8 in Lake Kulik to 1.95 in Main Bay ($F = 2.74$, $P = 0.002$). Significant deviation from linkage equilibrium was observed in over half of the collections for only three pairs of loci: *One_aldB-152* & *One_ALDOB-135*, *One_GPDH-201* & *One_GPDH2-187*, and *One_MHC2_190* & *One_MHC_251*. These loci were treated as independent for the remaining analyses because we wanted ranking and panel testing to include all available loci for the species. Retaining these loci in the data set could lead to upward bias in assignment success and redundant information. However, there are only three pairs of loci which were not linked in all populations warranting their retention in downstream analyses.

Across all collections, the average F_{ST} was 0.114 for all 110 polymorphic SNPs (Appendix A). There was significant genetic differentiation between all population pairs ($P < 0.001$) except for the pair from Southcentral Alaska (Table 1.1; populations 36 & 37); however, the level of differentiation between and among regions is variable as indicated by the heat map (Figure 1.2). The genetic relationships among populations can be seen in the PCA (Figure 1.3) where the population pairs generally cluster and are separated clinally from east to west on principal coordinate 1 (44.5% of the variation observed among the collections). Population differentiation across the species range may be driven by the five loci detected to be under diversifying selection (Figure 1.4). Two of these were new loci described in this paper. When loci under directional selection were removed from the data set the same pattern of genetic

differentiation was observed with only the pair of Southcentral populations remaining indistinguishable. However, all F_{ST} values were lower without the loci under directional selection and there was a significant difference in genetic differentiation measured when these outlier loci were removed ($Z = 0.94$, $P < 0.01$).

SNP ranking

Informativeness values (I_n) were highly correlated with the locus contribution (LC) (r_s=0.93, $P < 0.001$; Figure 1.5) using a Spearman rank correlation. I_n was also highly correlated (r_s=0.99) to F_{ST} as shown by Rosenberg et al. (2003). Most loci were ranked differently using each method for both the full population set and the training set (Figure 1.6). The greatest differences in rank were observed for loci with small heterozygosities (e.g., *One_gadd45-269* and *One_parp3-170*). Often these loci received a high rank (low number) from BELS and a lower rank (high number) from F_{ST}, I_n, and LC measures (e.g., *One_redd1-414* and *One_serpin-75*). BELS rank did vary with input order (Figure 1.7), but all of the top-ranked loci remained top-ranked loci and the variation in locus rank was not significantly different between ranks from the different input orders ($P = 0.59 – 0.97$).

Although there were differences in locus ranks, the 96-SNP panels contained many of the same loci. When using the full population set to determine locus rank only 3-7 loci differed between the five 96-SNP panels created using the different ranking methods (Table 1.2). Up to 13 loci differed between the five panels when ranks were determined using only the training set, which contained half as many individuals. However, F_{ST} and I_n panels shared all but one locus. There were fewer loci shared between the 48 SNP panels. The F_{ST}, I_n, and LC 48-SNP panels from the full population set had up to 11 different loci, and each of these shared only a little over half of their loci with the BELS and WHICHLOCI 48-SNP panels. The F_{ST}, I_n, and LC 48-SNP panels from the training set were more similar with only 3 - 7 different loci. Only 16 loci differed between these panels and the WHICHLOCI panel, while the BELS panel had the most unique loci, sharing as few as 12-20 loci with another panel. Despite differences in panel composition with two different population sizes (full set vs. training set) there was no significant difference in average locus rank ($P = 0.96$). Since the purpose of splitting the SNP assessment populations into a training set for SNP ranking and a holdout set for assessing SNP performance was to reduce upward bias only training set ranks were used for panel testing.

Panel testing

Using empirical data from the holdout set of individuals when assigned to a baseline of training-set individuals in ONCOR the average probability of correct assignment was 0.83 for the 96-SNP panels and 0.70 for the 48-SNP panels. The average probability of correct assignment was higher using simulated individuals (*f*ORCA) for both the 96-SNP panel, 0.96, and the 48-SNP panel, 0.85. There was a significant difference in mean assignment scores using empirical ($F = 48$, $P < 0.0001$) and simulated ($F = 27409$, $P < 0.001$) data (Figure 1.8). In the empirical data there was greater variation in probability of correct assignment and fewer significant differences between panel performances (Table 1.4). Most of the 96-SNP panels performed similarly and significantly better than the 48-SNP panels ($P < 0.001$) except for the BELS and randomly generated 96-SNP panels which performed similarly to the 48-SNP F_{ST} and I_n panels only when using the empirical data (Table 1.4). The 48-SNP panels tested empirically performed similarly to at least one other 48-SNP panel except for the BELS 48-SNP panel which had the lowest average probability of correct assignment (0.49). Of the 48-SNP panels all the panels performed significantly differently ($P < 0.001$) using the simulated data. The 48-SNP F_{ST}, I_n, and LC panels had the highest average probability of correct assignment (0.87 – 0.88) and the panels with the lowest average were the randomly generated panel (0.84) and the BELS panel (0.72).

When all 110 polymorphic loci were used for individual assignment, the average probability of correct assignment was 0.85. The average probability of correct assignment decreased as loci were removed but remained above 0.7 until only 40 loci remained (Figure 1.9). The range of probabilities for correct assignment also increased as loci were removed from the data set. The 25^{th} quantile nearly flanks the average probability of correct assignment until only 75 loci remain. The probability of correct assignment at the 75^{th} quantile remained nearly as high as 1.0 for some individuals until only 30 loci remain (Figure1. 9). The greatest changes in assignment accuracy began when dropping from 20 loci (0.85) to 15 loci (0.64).

Discussion

SNP discovery

Our goal was to expand the battery of 45 commonly used SNPs into sets of 48 or 96 to better utilize the medium density arrays commonly in use. We successfully developed and

validated 43 new SNP assays. Using HRMA, we were able to quickly and affordably evaluate putative SNPs. SNP candidates were eliminated if HRMA revealed that there was no SNP present in the amplified region or if there were multiple variants (suggesting paralogous variation). Unsuccessfully amplified loci may have been adjacent to intron-exon boundaries resulting in PCR failure. This source of candidate SNP drop-out is difficult to avoid when using transcriptome without a reference genome to identify intron-exon boundaries, and the lack of a reference genome may continue to present challenges for future candidate SNP discovery. However, despite these challenges, improved NGS technologies and improved bioinformatics will continue to accelerate SNP discovery in non-model organisms (Hohenlohe et al., 2011). One drawback of our approach was that sequence assembly using short reads and transcriptome sequences, especially without a reference genome, was difficult and computationally exhaustive. Some false positives, especially in the SOLiD-derived transcriptome, were probably dependent on the method of assembling the short reads (Everett et al., 2011). Based on the difficulty of SNP discovery here and the availability of new technologies and techniques we no longer use SOLiD sequencing for SNP discovery.

One facet of SNP discovery that warrants attention is ascertainment bias, which is introduced during the SNP discovery process because the variation being captured is usually only representative of a small number of individuals (Ackerman et al., 2011; Brumfield et al. 2003; Luikart et al., 2003; Morin et al., 2004). Concerns about ascertainment bias have been previously addressed (e.g., Rosenblum & Novembre, 2007), and there appears to be a growing consensus that the effects of ascertainment bias are nearly negligible when parsing out relationships between populations when more SNPs are used (Smith et al., 2007). In this study, ascertainment bias for some SNPs would have been introduced during the initial SNP detection step where sequences from only a few individuals in Bristol Bay Alaska were used. However, using populations across the species' range for SNP validation was meant to ensure the capture of SNPs to resolve Bristol Bay populations while also providing geographically broad resolution. Despite a limited number of ascertainment fish, there does not appear to be a strong signal for ascertainment bias in this study. Allelic richness, which can be a signal of ascertainment bias, does not vary much across the range of populations surveyed. The significant variation among regions may reflect underlying differences in genetic diversity between populations as there is no clear geographic trend in mean allelic richness.

SNP assessment

Most populations were easily differentiated except for the Main Bay Hatchery-Coghill Lake pair in Southcentral Alaska. The exception may be attributed to the fact that fertilized eggs from Coghill Lake fish were introduced into the Main Bay Hatchery population during the last three decades (PWSAC Hatcheries, www.pwsac.com/mbh.htm). The high F_{ST} values observed between all other populations and regions in this study reflect the large geographic range surveyed in addition to the extreme philopatry of the species which results in strong genetic differences across even small geographic scales (Quinn, 2005). Over 40% of observed genetic variation is accounted for in the first principal component of genetic distance, which parses apart Washington, British Columbia, Southcentral Alaska and the more western collections. The second principal component primarily differentiates among the western collections: Russia, Bristol Bay and the Alaska Peninsula. This suggests that there are different suites of SNPs that are better for resolving population structure across different geographical scales (Helyar et al., 2011). One approach to identifying an additional suite of SNPs would be to rank loci by their contribution only to a specific principal component that is pulling apart populations of interest.

Linkage disequilibrium (LD) was observed in some loci and in some locations, but only between loci where linkage relationships or LD were noted in other studies (e.g., the MHC SNPs in McGlauflin et al. (2011)). The treatment of linked loci is often dependent on the application and decided by the primary investigator. Often combining linked loci can provide increased resolution (Creelman et al., 2011); however some programs used for genetic analyses such as population assignment cannot use this phased data which consists of multi-allelic haplotypes. In some cases, linked loci appear to provide similar information, measuring the same allele frequencies across populations. Although these loci may have similar resolving power, they may only provide redundant information (i.e., providing the power to differentiate between the same populations) in which case one locus might be dropped from the loci set without losing resolution. Developing more standardized methods for parsing the difference between the value of a locus for its resolving power and its value due to uniqueness of information will become important for creating highly optimized SNP panels.

In previous studies, using a subset of these SNPs, the MHC loci have often been identified as candidates for natural selection (Gomez-Uchida et al., 2011; McGlauflin et al., 2011); however, this was not the case here. Those studies surveyed populations across a much

smaller geographic range (Creelman et al., 2011) and for different life history types (McGlauflin et al., 2011) suggesting that the MHC loci might be displaying a signature of local adaptation. In this study, strong genetic differentiation across a large geographic range may dwarf a signal of selection at the MHC loci that may occur at smaller geographic scales.

Studies have shown that SNPs indicated to be under directional selection (Beaumont & Nichols, 2006; Excoffier et al., 2009) can greatly improve the resolution of population structure and the accuracy of individual assignment (e.g., Ackerman et al., 2011; Freamo et al., 2011). This warrants the exploratory use of these methods for locus assessment. Many of the loci indicated to be under selection were also some of the most informative; the added value of including these loci was demonstrated most recently by Ackerman et al. (2011) where the inclusion of these non-neutral markers significantly improved individual assignment. Here we found that the removal of outlier loci did significantly decrease F_{ST} values, but the relationships between populations remained the same. In studies where there is less natural variation between populations the value of including outlier loci in individual assignment would most likely be higher. Despite concerns regarding the influence of these markers in population genetic studies, it is evident that non-neutral markers are valuable for population identification.

SNP ranking

SNPs can be ranked in a variety of ways. Computer programs such as WHICHLOCI generate optimized SNP panels using genetic data, rigorous statistical algorithms and general objective functions. Alternatively, ranking procedures developed for specific applications might consider everything from laboratory performance to accuracy of individual assignment. Unsurprisingly, sample size does impact ranking as we observed greater variation in locus ranks using the training set (Figure 1.6), which had half as many individuals as the full assessment populations. However, many of the highest ranking loci remained highest ranking loci (e.g., *One_apoe-83)*. Interestingly, differences in locus ranks based on diversity indices (F_{ST}, I_n, and LC) versus the likelihood-based ranking programs BELS and WHICHLOCI were greater using the training set. Potentially, sample size may have a greater impact on ranking when using these programs. Although the ranking strategies used here are not novel, we believe that showing a comparison of ranking approaches for the same data is informative and will be of value once researchers have access to hundreds of SNPs.

Panel testing

As expected, the addition of more loci significantly improved individual assignment in both panel testing (Figure 1.8) and when loci were dropped sequentially from individual assignment by rank (Figure 1.9). More interestingly, it appears that as the number of available loci increases, the ranking approach will become more important as evidenced by differential performance of the 48-SNP panels using both empirical and simulated data (Table 1.4). One would expect that, if we were creating a 96-SNP panel from over 200 markers we would see similar differences in 96-panel performance to those that we observed for 48-panel performance in this study.

We used two approaches for evaluating panel performance. The greater variation in probability of correct assignment observed using empirical data can be attributed to differences in individual quality. In the empirical data some of the samples may be of poor tissue quality, be missing genotypes, or might not have diagnostic genotypes, making them difficult to assign back to populations of origin. With simulated individuals there is none of this natural variation in individual quality explaining the low variance and higher average probability of correct assignment. Because sample quality is not a source of variation in the simulated data it provides a better idea of which panel performs best based solely on SNP composition while using empirical data provides a better idea of actual panel performance in a study. There would be less variation in the probability of correct assignment with larger sample sizes for SNP ranking (training set) and evaluation (holdout set).

For both panel-testing approaches and for both panel sizes there was a pattern in panel performance. F_{ST}, I_n, and LC panels were often the most similar and had the highest average probability of correct assignment. The similarity between these panels is expected since these ranking methods are all highly correlated (e.g., Figure 1.5). These three panels were also similar to the WHICHLOCI panel and the panel based of average loci rank across all five ranking methods (Table 1.4). The BELS panels had the lowest average probability of correct assignment, which was even lower than the panel of randomly selected loci. BELS has difficulty ranking loci when assignment accuracy is set to be 100% (Helyar et al., 2011), possibly accounting for the panel's poor performance. Despite the poor performance of the BELS panels, there is continuity in how BELS ranked loci; the highest ranked loci remained the highest ranked over multiple runs (Figure 1.7). Some of the highest ranked loci were also highly ranked for F_{ST}, I_n, and LC (e.g.,

One_apoe-83). The stability of highest locus ranks and variability of mid- and low-performing locus ranks might be an artifact of the program's intent to determine a minimum set of loci that maximizes performance. Once the best performing loci, such as the top 40, have been identified, the addition or removal of the remaining loci results in minimal changes in performance resulting in arbitrary ranks.

Conclusions

The popularity of a given type of molecular marker has changed repeatedly over recent history. Regardless of the marker type or discovery method, there is continued interest in developing methods for ranking and evaluating markers, hence the design of locus selection programs such as BELS and WHICHLOCI. SNPs have recently become a marker of choice for several non-model species, and there is growing interest in methods to evaluate the ever-increasing number of SNPs. Here we not only describe an effective method for SNP discovery in the culturally and commercially important non-model sockeye salmon, but we also demonstrate how common locus-ranking methods perform differently when developing a SNP panel. Although our investigations explore the role of loci for use across a large geographic scale with high overall differentiation, the same methods can be applied and optimized for finer geographic resolution. The steps outlined here provide a starting place for developing a minimum panel for maximum assignment for any specific system or question. With more SNPs to choose from, investigators will need to take criteria for ranking and optimizing SNP panels into careful consideration. Here we recommend panels of 48 or 96 SNPs that will expand the options for improved management and conservation of the iconic sockeye salmon.

Table 1.1 All collection location and sample sizes sorted by application.

Application	Region	Map #	Location	n
SNP discovery ascertainment	Bristol Bay, Alaska	1	Yako Creek	2
		2	Yako Beach	2
		3	Silverhorn Bay Beach	2
		4	Lake Kulik	2
	Southcentral Alaska	5	Mendeltna Creek	2
SNP discovery validation	West Kamchatka, Russia	6	Hapiza River	8
	Bristol Bay, Alaska	7	Deer Creek	8
		8	Tikchik River	8
		9	Upper Nushagak-Klutapuk Creek	8
		10	Pick Creek	8
		11	Upper Talarik Creek	8
		12	Ualik Lake tributary	8
		13	Becharof Creek	8
		14	Margot Creek	8
	Alaska Peninsula	15	Hatchery Beach, Chignik	8
		16	Broad Creek	8
		17	Cinder River	8
		18	Bear Lake	8
		19	Meshik River	8
	Southcentral Alaska	20	Yentna River slough	8
		21	Susitna River slough	8
		22	Coghill Lake	8
	Southeast Alaska	23	Klukshu River, Alsek	8
		24	Hugh Smith Lake	8
		25	McDonald Lake	8
	British Columbia, Canada	26	Scud River	8
		27	Taku River mainstem	8
		28	Slamgeesh Lake	8
		29	Meziadin Lake Beach	8
SNP assessment	West Kamchatka, Russia	30	Bolshaya River	90
		31	Ozernaya River	93
	Bristol Bay, Alaska	32	Lake Kulik	68
		33	Pick Creek	84
	Alaska Peninsula	34	Bear Lake	93
		35	Cinder River	89
	Southcentral Alaska	36	Coghill Lake	89
		37	Main Bay	61
	British Columbia, Canada	38	Upper Tatshenshini River	88
		39	Damdochax Creek	85
	Washington	40	Issaquah Creek	87
		41	Baker Lake	93

Table 1.2. Descriptions of the different approaches used for ranking SNP loci.

Ranking approach	Description	Reference
F_{ST}	Scaled among-population variance in allele frequency	Weir & Cockerham 1984
Locus contribution (LC)	Average contribution of each locus to principal components	Moazami-Goudarzi & Laloe 2002
Informativeness for assignment (I_n)	Estimates potential for an allele to be assigned to one population in comparison to an average population	Rosenberg et al. 2003
BELS	Ranks a locus' performance for maximizing mixture estimation accuracy during individual assignment	Bromaghin 2008
WHICHLOCI	Determines locus efficiency for correct population assignment and propensity to cause false assignment	Banks et al. 2003

Table 1.3. Summary of SNP discovery and validation. The number of primers that amplified a single product are shown for the first validation procedure: PCR test. The number of primer pairs that had melt curves with putative SNPs are shown for HRMA validation. The Sanger sequencing validation procedure shows the number of sequenced HRMA products that confirmed the SNPs identity for a primer pair. SNPs that were successfully genotyped from these sequences in a 5'-nuclease genotype are shown for the final validation procedure.

		Validation procedure			
Sequence Source	Primer Pairs	PCR test	HRMA	Sanger sequence	5'-nuclease genotype
Chum 454[1]	308	108	71	47	19
Sockeye SOLiD[2]	1536	407	77	46	24
Total	1844	515	148	93	43

[1] Seeb et al. 2011
[2] Everett et al. 2011

Table 1.4. P-values from post hoc Tukey's Honestly Significant Difference test for comparisons of performance of 96- and 48- SNP panel using empirical data (A) and simulated data (B). 96- and 48- SNP panels were generated using the following measures: genetic differentiation (F_{ST}), Rosenberg's informativness (I_n), average contribution of locus to principal components (LC), ranks from the locus selection programs BELS and WHICHLOCI (WL), average rank based on the five preceding measures, and randomly generated ranks. Non-significant p-values are indicated in bold.

A.

	96 F_{ST}	96 I_n	96 LC	96 BELS	96 WL	96 Random	96 AVG	48 F_{ST}	48 I_n	48 LC	48 BELS	48 WL	48 Random
96 I_n	**1.00**												
96 LC	**1.00**	**1.00**											
96 BELS	0.01	0.01	0.01										
96 WL	**1.00**	**1.00**	**1.00**	**0.05**									
96 Random	**0.93**	**0.92**	**0.93**	**0.71**	**0.99**								
96 AVG	**1.00**	**1.00**	**1.00**	0.02	**1.00**	**0.96**							
48 F_{ST}	0.00	0.00	0.00	**1.00**	0.00	**0.21**	0.00						
48 I_n	0.01	0.01	0.01	**1.00**	0.05	**0.68**	0.02	**1.00**					
48 LC	0.00	0.00	0.00	**0.99**	0.00	0.05	0.00	**1.00**	**0.99**				
48 BELS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
48 WL	0.00	0.00	0.00	0.03	0.00	0.00	0.00	**0.26**	0.04	**0.63**	0.00		
48 Random	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.53**	
48 AVG	0.00	0.00	0.00	**0.90**	0.00	0.01	0.00	**1.00**	**0.91**	**1.00**	0.00	**0.90**	0.00

Table 1.4. Continued.
B.

	96 F_{ST}	96 I_n	96 LC	96 BELS	96 WL	96 Random	96 AVG	48 F_{ST}	48 I_n	48 LC	48 BELS	48 WL	48 Random
96 I_n	**1.00**												
96 LC	**0.77**	**1.00**											
96 BELS	0.00	0.00	0.00										
96 WL	**0.28**	**0.83**	**1.00**	0.00									
96 Random	0.00	0.00	0.00	0.00	0.00								
96 AVG	0.02	**0.24**	**0.95**	0.00	**1.00**	0.00							
48 F_{ST}	0.00	0.00	0.00	0.00	0.00	0.00	0.00						
48 I_n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
48 LC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
48 BELS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
48 WL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
48 Random	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
48 AVG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

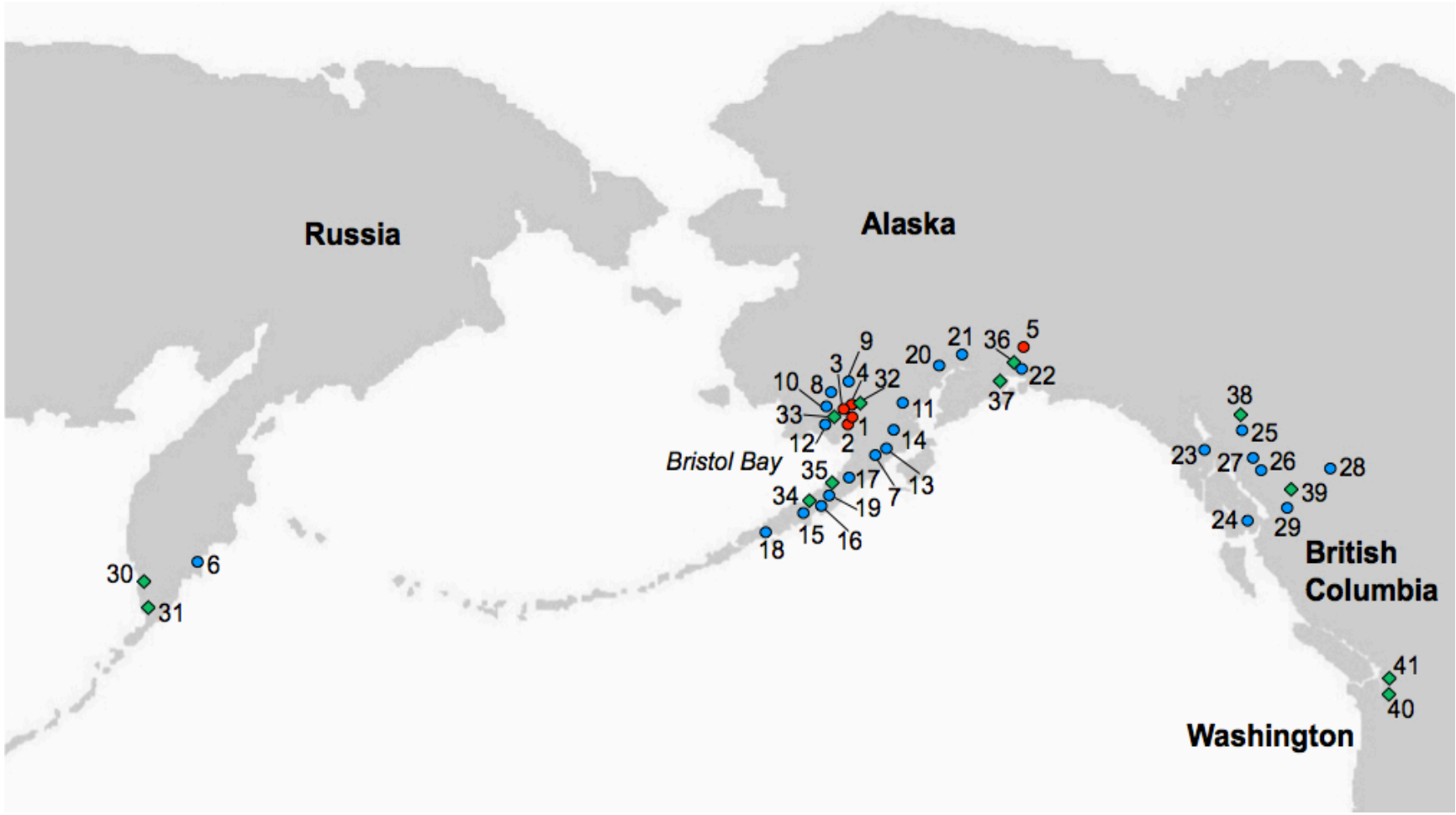

Figure 1.1. Locations of samples collected for SNP discovery. See Table 1.1 for location names corresponding to numbers. Sockeye salmon collected for SOLiD sequencing and initial SNP ascertainment indicated by red circles. Samples collected for SNP validation indicated by blue circles. Collections used for SNP assessment and ranking at all 114 SNP loci marked with green diamonds.

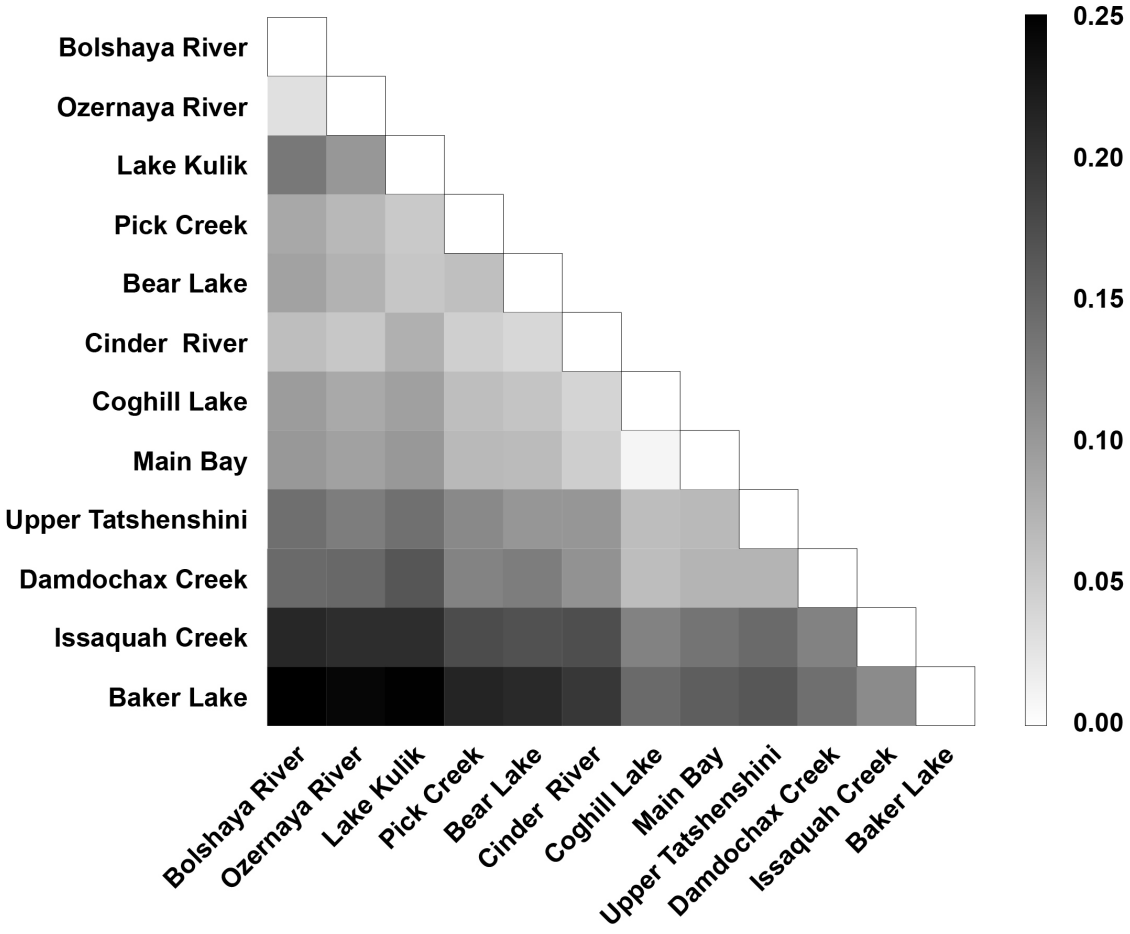

Figure 1.2. Matrix of pairwise F_{ST} values for all population comparisons**.** Values calculated for all 110 polymorphic SNPs. Shading reflects degree of divergence and corresponds to F_{ST} values indicated in the legend (right). Collections are in geographic order from Russia to Washington.

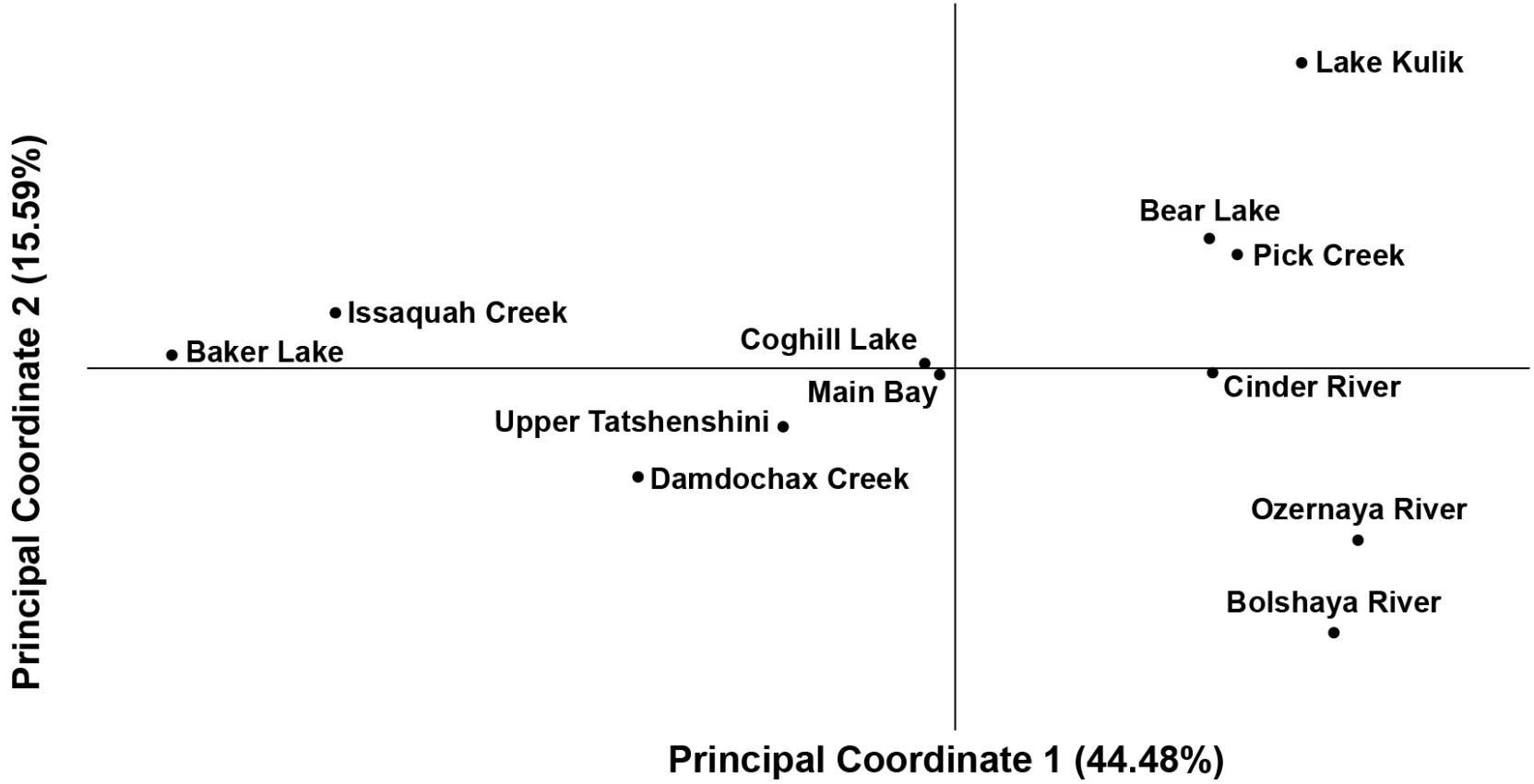

Figure 1.3. Principal coordinate analysis of SNP assessment populations. The first and second principal coordinates shown based on population allele frequencies for 110 SNPs. The percentage of variance accounted for by each coordinate is given in parentheses.

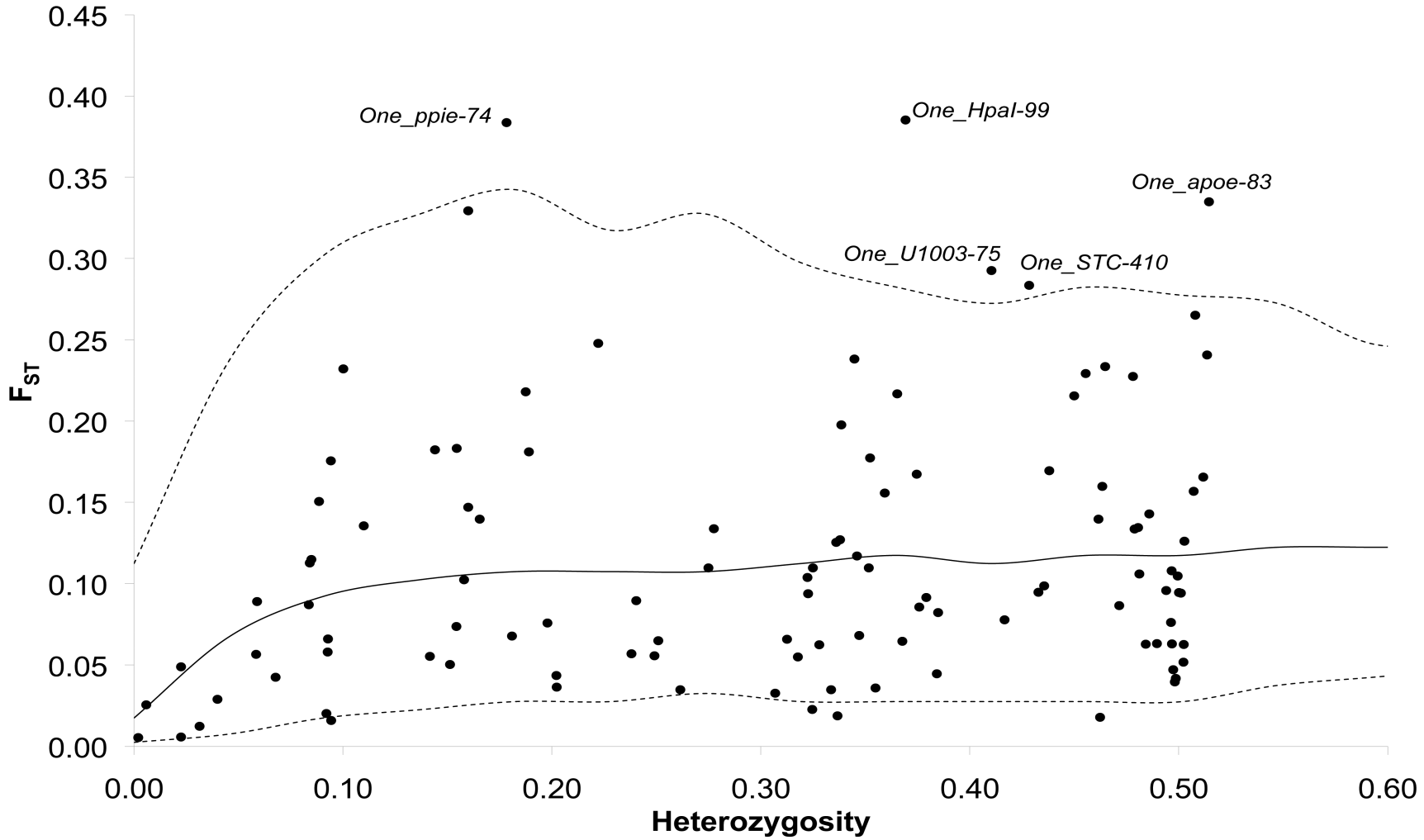

Figure 1.4. Heterozygosity and F_{ST} for assessment populations. Values were calculated for all 110 sockeye salmon SNPs and the upper and lower 99th quantiles are denoted with dashed lines. The 50th quantile is denoted with a solid line. Loci lying outside of the upper 99th quantile are labeled and considered to be candidates for directional selection.

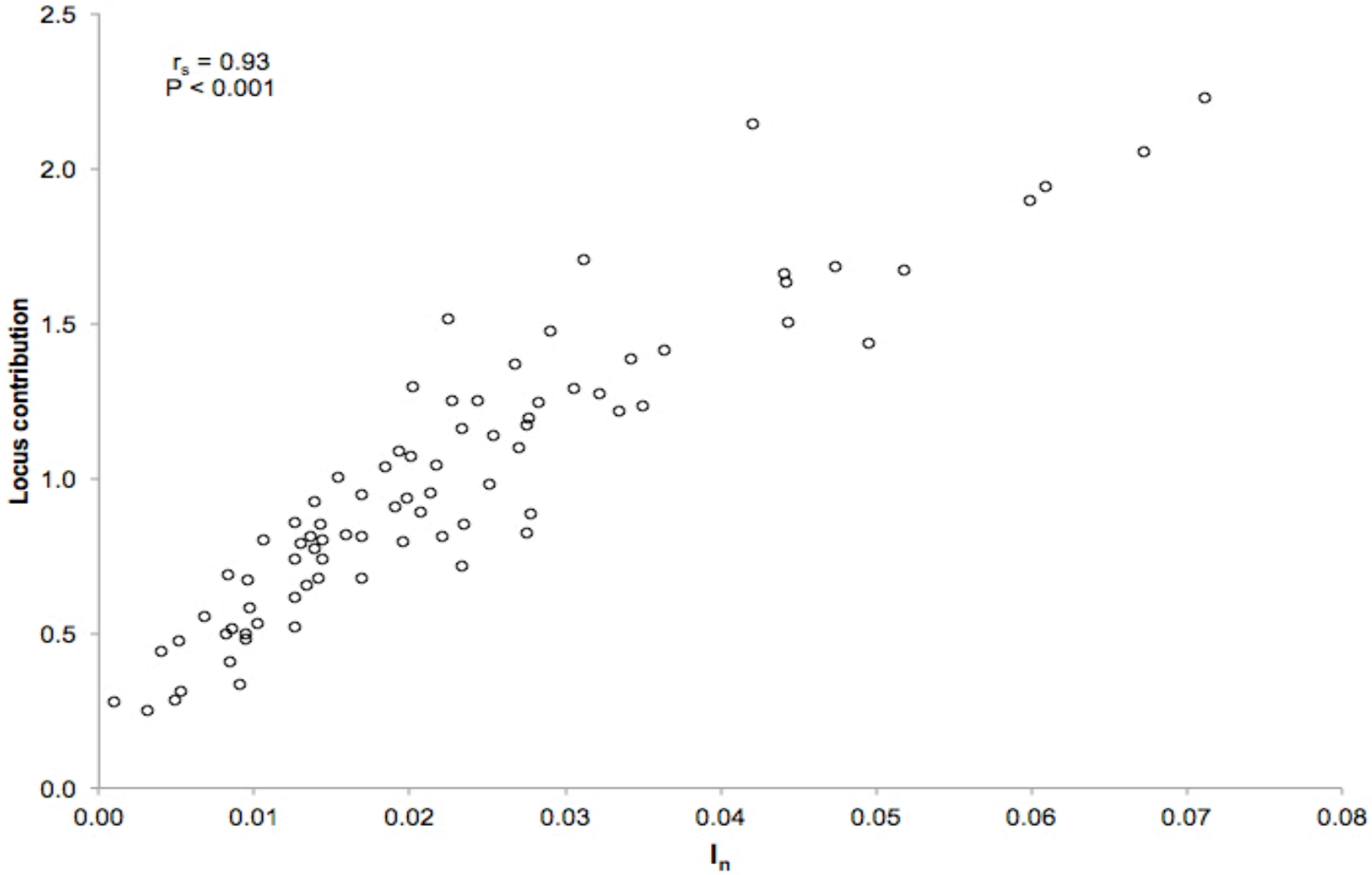

Figure 1.5. Spearman rank correlation between the average contribution of a locus to principal component analysis (LC) and the locus informativeness (In) calculated for the 12 assessment populations and 110 SNPs.

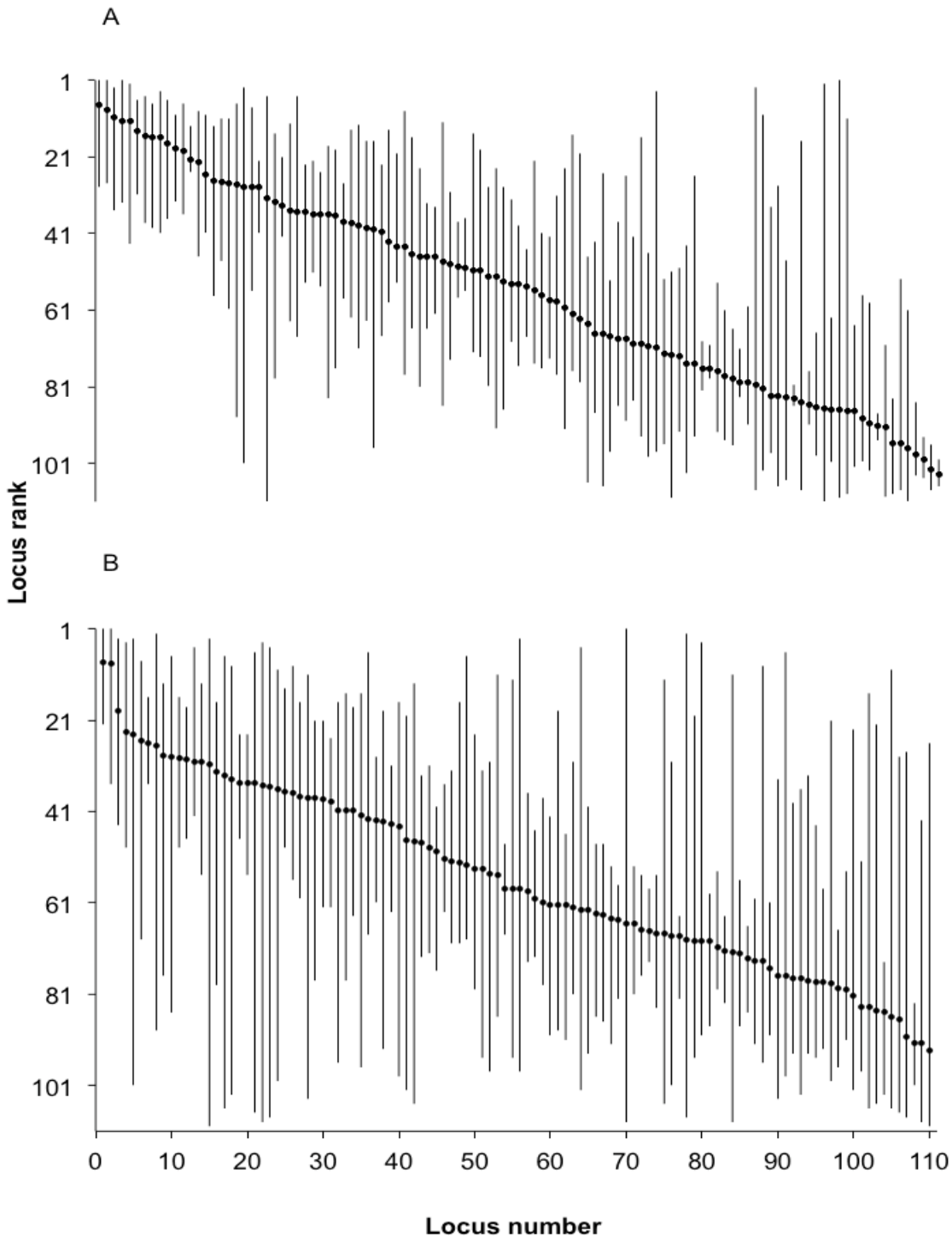

Figure 1.6. Loci ordered from left to right by highest average locus rank (locus number) as in Appendix A for the full 12 assessment populations (A) and for training-set individuals only (B). Average locus rank indicated by closed circles with bars extending from the highest and lowest rank for that locus from the different ranking procedures.

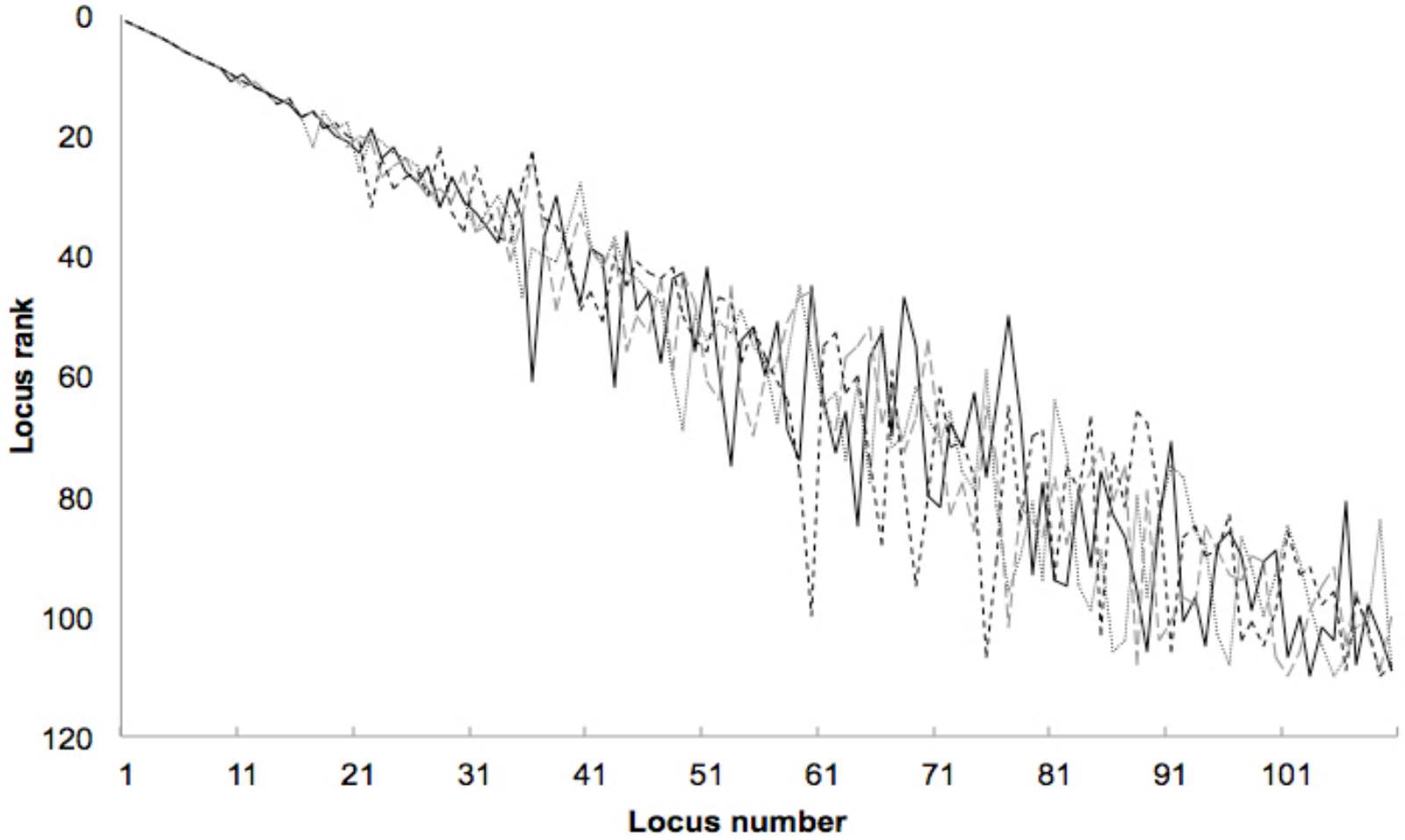

Figure 1.7. Difference in BELS locus ranks with input order. Input orders: alphabetical (dotted line), reverse alphabetical (solid line), and two randomly generated locus orders (black dashes and grey dashes). Locus number corresponds to average locus rank (Appendix A).

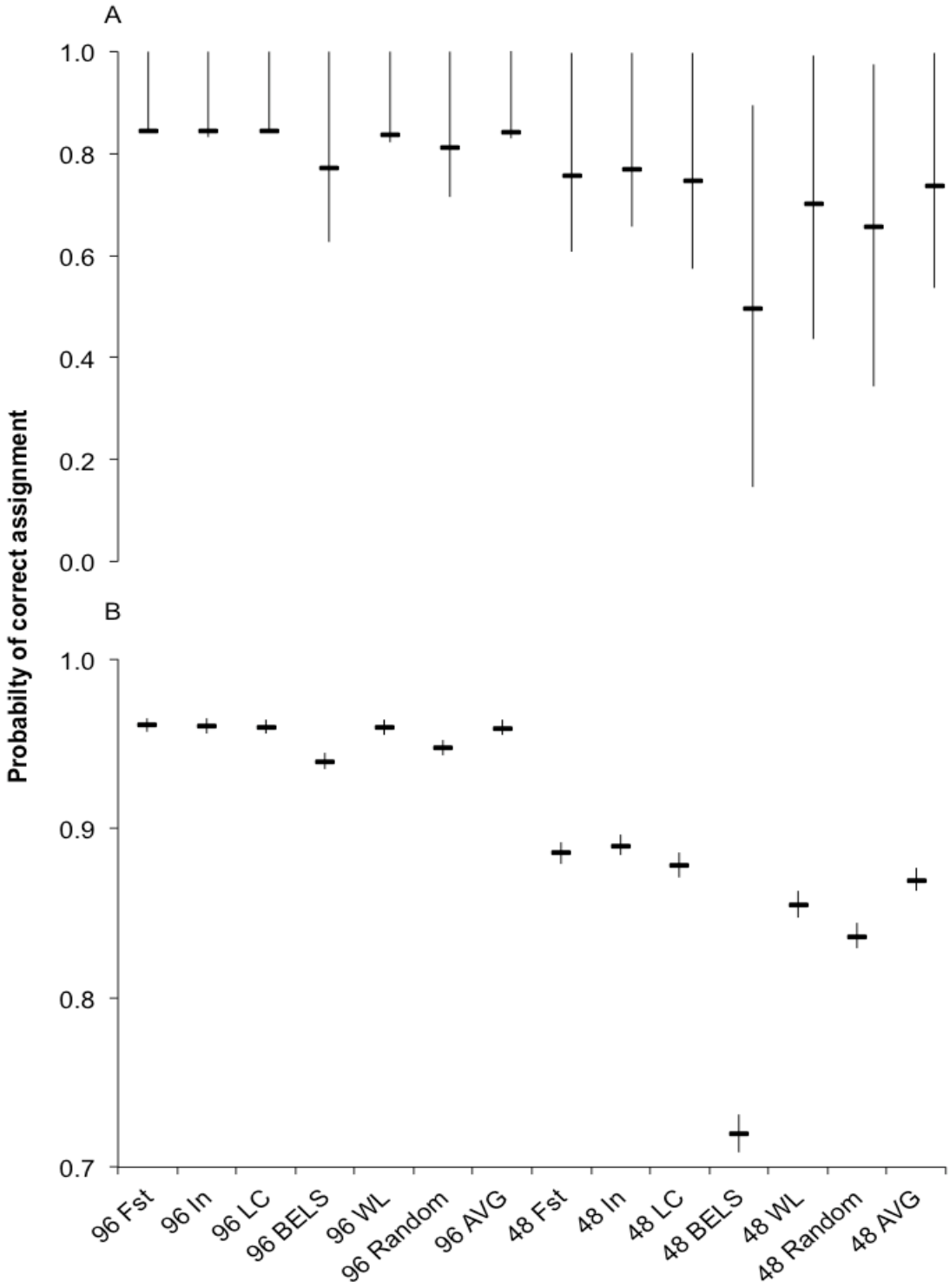

Figure 1.8. Probability of correct assignment for 48- and 96-SNP panels using empirical data (A) and simulated data (B). Each panel contains the highest ranked loci for each ranking approach: F_{ST}, informativeness (I_n), average contribution of a locus to principal components (LC), the locus-selection program BELS, and the locus-selection program WHICHLOCI (See Table 1.2). The random panel contains loci chosen at random. Whiskers extend to the 1st and 3rd quartile around the mean.

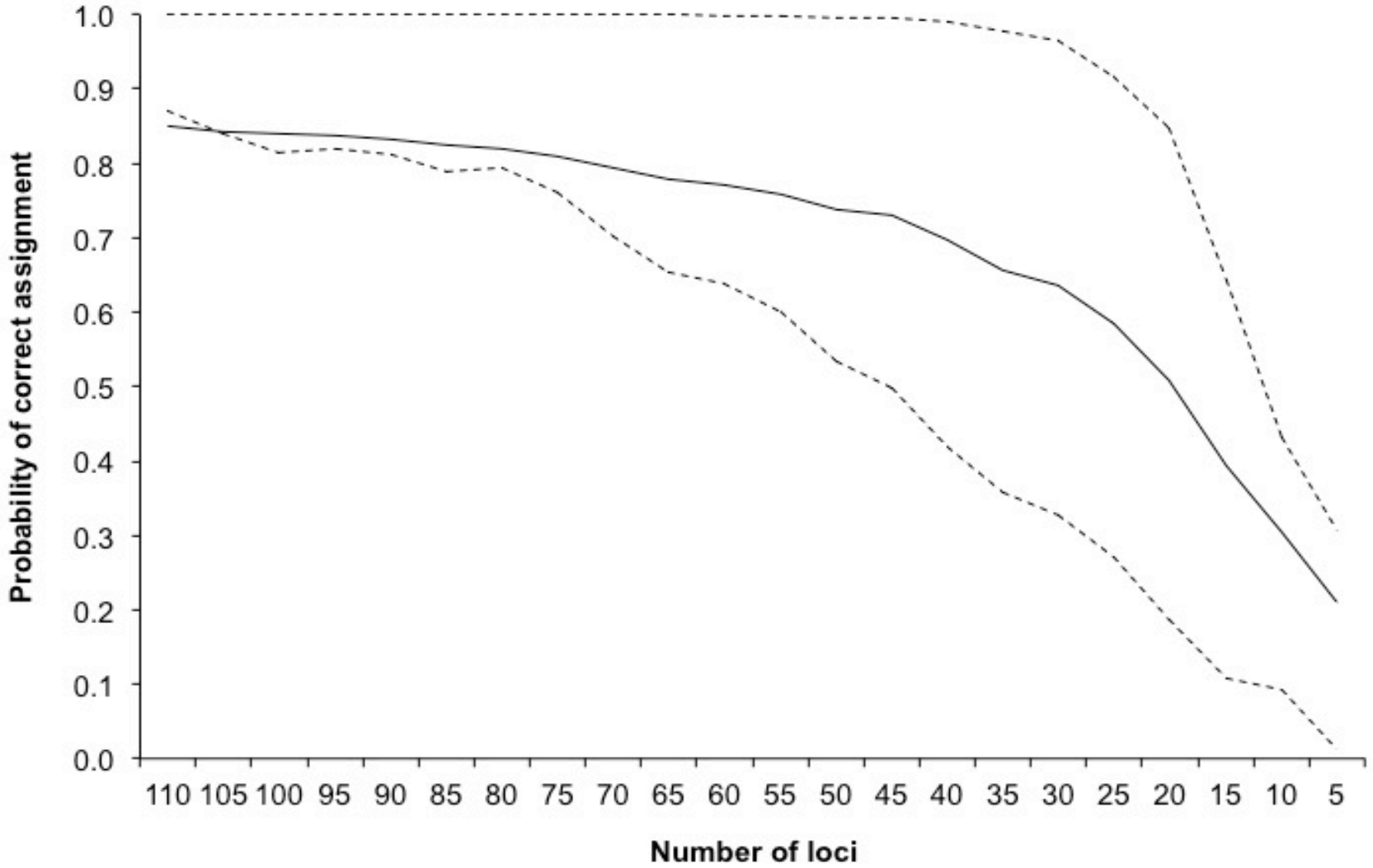

Figure 1.9. Probability of correct assignment with decreasing number of loci. Loci were removed five at a time by lowest rank.

Chapter II: Characterizing differences in gene expression patterns associated with variability in senescence

Introduction

Aging and Senescence

Aging is a complex process, regulated by multiple cellular and molecular mechanisms. Generally, aging is characterized by cellular damage exceeding the work capacity of repair mechanisms (Kim, 2007) and the associated decline in cell replication (Hodes, 1999). The accumulation of these age-related changes, ultimately leading to death, is referred to as senescence (Finch, 1994). During senescence an organism is more susceptible to disease and has diminished capacity to carry out fundamental physiological processes (Chen, 2007). Two primary processes that have been identified as regulating senescence are telomere dynamics and oxidative stress.

Telomeres are a series of tandem nucleotide repeats that cap the ends of chromosomes, protecting them from degradation and the loss of genetic information that directs cellular processes. Telomeres shorten with each cell division, though they can be repaired by the enzyme telomerase. As cells age average telomere length decreases and eventually reaches a critical length, inhibiting further cellular replication and triggering senescence (Aubert & Lansdrop, 2008; Armanios et al., 2009). While telomere dynamics are clearly linked to life expectancy (see Monaghan 2010 for discussion), it is unclear whether aging causes telomere shortening or whether telomere shortening causes aging (Hornsby, 2006). Despite this uncertainty, the study of telomere dynamics has provided new insights into understanding life span and survival (e.g., Salomons et al., 2009) and has prompted research of new anti-aging medical treatments (Jaskelioff et al., 2010).

Oxidative stress refers to cellular damage incurred by reactive oxygen species such as hydrogen peroxide and superoxide. This includes DNA (and telomere) damage, the detrimental oxidation of fatty acids and amino acids, and even the inactivation of enzymes through the oxidation of co-factors (Shigenaga et al., 1994). Reactive oxygen species are natural byproducts of cellular metabolism, but they can also be generated in response to different types of environmental stress. In most cases antioxidants are able to ameliorate the negative impacts of reactive oxygen species by converting them into less detrimental molecules (Bayir, 2005). However, when reactive oxygen species activity

supersedes antioxidant activity the resulting accumulation of cell damage causes senescence (Sohal et al., 2002). It is clear that reactive oxygen species can be detrimental to a wide range of cellular functions from transcription to cell signaling and that oxidative stress increases with age (Kregel & Zhang, 2007). Understanding how the accumulation of these changes mechanistically impacts organismal lifespan remains to be shown, beyond correlations between oxidative stress and senescence.

Research regarding the mechanisms of senescence is ongoing and continues to uncover new intricacies to the aging process. For example, protein aggregation, which is the accumulation of misfolded proteins, has received growing attention for its role in cellular senescence. While protein aggregation can be facilitated by reactive oxygen species, new research indicates that there is a genetic basis for the likelihood of a protein to aggregate in the amino acid sequence (Villar-Pique, 2011). As technologies improve for studying these different mechanisms a better understanding of biological senescence is being gained on both genetic and cellular scales.

Pacific Salmon as a Model for Senescence

Semelparous Pacific salmon (*Oncorhynchus* spp.) are unusual in that they undergo rapid senescence, declining from a visually healthy individual entering a stream to death after spawning in a matter of days or a few weeks. In salmon, as well as other rapidly senescing species such as octopi, lampreys, and many insects, this process is associated with an energetic investment in reproduction in conjunction with starvation and stress (Finch, 1994). The physiological trade-offs resulting from energetic investment in reproduction instead of foraging include decreased immune function (Finch, 1994), increased oxidative stress (Gotz et al., 2005), and up to 80% loss in stored energy (Gende et al., 2004a; Morbey et al., 2005). In addition to the depletion of stored fat, senescing salmon also lose a significant portion of protein stores (Brett, 1995). At the same time, cellular damage occurs in all major organ systems including the central nervous system where neurons and neurites begin to breakdown (Maldonado et al., 2002).

The combined effects of these physiological changes and the influence of environmental variables contribute to the rate at which senescence occurs but there are genetic controls over senescence rates as well. For example, rates of senescence can vary

significantly among discrete breeding populations (Carlson et al., 2007). In some cases, variation in in-stream longevity among populations has been attributed in part to bear predation, which can account for up to 80% of mature salmon mortality in some small streams. In shallow streams, bears kill a higher proportion of salmon (Quinn et al., 2001) and tend to feed on newly arrived fish (Gende et al., 2004b), and in these streams the salmon senesce more rapidly and die sooner than in streams where predation risk is lower (Carlson et al., 2007). In addition, salmon arriving at a given streams tend to live longer (i.e., senesce slower) than those arriving later in the season (Perrin and Irvine, 1990; Hendry et al., 1999; Doctor and Quinn, 2009).

To investigate the mechanisms driving senescence in Pacific salmon, we assessed five different physiological processes that might be involved in senescence (olfaction, immune response, reproduction, memory, and aging). We tested the null hypothesis, that gene expression levels would not differ between fish in a more advanced state of senescence compared to less senescent fish, against two alternatives. Under the first scenario, gene expression might be lower in fish as they approach death, as the processes supporting migration (olfaction and memory), reproduction, and survival are no longer needed. This was contrasted against the alternative hypothesis, that levels of gene expression would be elevated, as would be consistent with the fish trying to optimize their performance even as death approached. Our second objective was to develop quantitative expression assays to measure differences in the number of gene transcripts (i.e., expression) for genes associated with aging and senescence in sockeye salmon.

Methods

Sampling

All sampling took place at Hansen Creek, a tributary of Lake Aleknagik in southwestern Alaska, where previous studies on sockeye salmon senescence have taken place (Carlson et al., 2004, 2007; Doctor & Quinn, 2009). On August 1, 2009, 14 male sockeye salmon were collected with seine nets as they schooled at the mouth of the stream in the lake prior to entering Hansen Creek and were termed “pre-senescent”. Based on other tagging studies (e.g., Doctor & Quinn, 2009) we would expect such individuals to enter the stream within a few days to a week, and die of senescence in

about 10 d after entering the stream unless they died from predation or other premature cause. With hand nets, 14 other live but "senescent" salmon were sampled in-stream on August 10 and 12. These fish showed visible signs of approaching death (frayed fins, emaciated bodies, unhealed wounds, and the rough skin commonly seen in senescent salmon). All sockeye salmon sampled for gene expression were euthanized using IACUC compliant protocols. After euthanasia brains were removed and stored in RNAlater (QIAGEN) for subsequent processing.

When possible, photos of sampled senescent salmon were taken and ranked based on extent of senescence as determined by degree of pigment loss and extent of tissue degradation (condition). Fish with severely damaged dorsal fins and pigment loss throughout the caudal fin extending to the anal fin were classified as most senescent and assigned the rank of 3 (e.g., 44; Figure 2.1). Fish with a moderately damaged dorsal fins and pigment loss isolated to the caudal fin were assigned the rank of 2. Fish with minimal fin damage and pigment loss were classified as least visibly senescent and assigned the rank of 1. Senescent fish condition was then used to correlate with gene expression values.

Gene expression sample preparation

RNA was isolated from brain tissue (~50 mg) from the 28 salmon (14 pre-senescent and 14 senescent fish) using Tri-Reagent (Molecular Research Center) per the manufacturer's protocol. To eliminate possible DNA carryover, total RNA was DNase-treated using the Turbo DNA-free Kit (Ambion) according to the manufacturer's "rigorous" protocol. Removal of genomic DNA was confirmed via real-time PCR on DNAse-treated total RNA (data not shown). Total RNA was stored at -80C. RNA was then reverse transcribed using M-MLV Reverse Transcriptase (Promega) and oligo dT primers (Promega) according to the manufacturer's protocol. cDNA was diluted 2-fold with molecular biology grade water and stored at -20C.

Gene selection and primer design

Genes were selected based on physiological responses of interest during the process of senescence and the availability of sequence data. No gene expression assays

existed for genes of interest so new assays were designed using public sequence archives from Genebank. Nucleotide cDNA sequences or expressed sequence tags from sockeye salmon, or the closely related species, *O. mykiss* for five genes (Viperin, NMDA, OMP1, TERT, and GnRHp) were selected for primer design (Table 1). There were no available nucleotide sequences for telomerase reverse transcriptase (TERT) so a nucleotide sequence for TERT in a closely related species (*O. mykiss*) was identified using tblastn (NBCI) and the amino acid query sequence for TERT (ADM83418). The resulting top blast hit (e-value: 1E-39) corresponded to an EST in the same species (CX246542) and was used for primer design. Primer3 (Untergasser et al., 2007) was used to design primers that amplify a 150-250 bp region of genes of these genes. The ribosomal RNA gene 18S gene was chosen as a normalizing gene for expression analysis because this gene has been shown to be constituently expressed in a number of plant (Nicot et al. 2005) and animal (Bas et al. 2004) species including zebrafish (McCurley & Callard 2008) under most conditions.

Real time quantitative PCR

Quantitative, real-time PCR was carried out on the BioRad CFX96 thermocycler. All samples were run in duplicate for each gene in 20 ul reactions containing 0.5uL cDNA, 0.2 uM forward/reverse primers, 1X Sso Fast EvaGreen Master Mix (BioRad), and nuclease free water. Cycling parameters were as follows: reaction activation at 98C for 2 minutes and 40 cycles of denaturation at 98C for 2 seconds and annealing/extension, followed by final sample melt from 65C to 95C at 0.2C increments. Average cycle threshold (C_t) for each sample and gene amplification efficiency were determined using Real-Time PCR Miner Software (Zhao & Fernald 2005). Relative mRNA concentration (R_0) for each gene for each individual was calculated using the following equation: $R_0 = 1/(1+E)^{C_t}$, where E is the average gene efficiency and C_t is the cycle number at threshold. The R_0 for each gene was normalized to 18S R_0 (a constituently expressed gene) from each individual. Using the normalized R_0, fold increase over the minimum R_0 value for each gene was calculated for all individuals (fold over minimum).

Gene expression analyses

Differences in fold over minimum (F/M) values between pre-senescent and senescent sockeye salmon for each gene were tested for using a student's t-test $\alpha = 0.050$ (Microsoft Excel for Mac 2011). Variance in expression between pre-senescent and senescent fish was compared using a F-test. Additionally, correlations between expression for each gene were tested to determine if individual fish displayed expression patterns across genes (i.e., did fish with high expression for one gene also have high expression for other genes) using Pearson's correlation in SPSS 17.0 (SPSS Inc.). Principal component analyses (PCA) were implemented for further comparisons between pre-senescent and senescent fish for all five genes using the package FactoMineR (Husson et al. 2008) in R 2.9.2 (R Development Core Team 2009). FactoMineR replaces missing values with the mean of each variable. The ranks of senescent fish condition as determined by photographs were also correlated with gene expression using the PCA function in FactoMineR. Significance of each principal component in explaining individual variance was computed using a Monte Carlo permutation test in R using the function "ordi.monte" from the collection of R functions BIOSTATS (McGarigal 2009). Missing values were imputed in order to run significance tests.

Results

Senescent salmon condition

Senescent salmon displayed varying degrees of physical senescence at the time of sampling (Figure 2.1). Although not all senescent salmon were photographed, five fish had considerable pigment loss and fin damage and were therefore assigned the rank of 3. Four fish had minimal pigment loss and pigment damage and were assigned the rank of 1. Only two fish received the rank of 2, falling between minimally senescent and severely senescent.

Gene expression analyses

Limited RNA and cDNA recovery reduced the sample of pre-senescent salmon to 11 for comparison with the 14 senescent salmon. Expression was more variable in senescent fish than pre-senescent fish for all five genes ($F = 5081$ $P << 0.01$; Table 2.2).

All genes examined were also expressed at higher levels in the brain tissue of senescent fish compared to pre-senescent samples but only in TERT ($P = 0.035$) and Viperin ($P = 0.016$) were the differences significant (Figure 2.2). Expression was positively correlated in all pairwise comparisons between genes and in 5 of 10 comparisons the correlations were significant (Table 2.3).

Based on PCA, senescent fish expression was more variable and could not be significantly differentiated from pre-senescent fish (Figure 4). Only principal component 1, accounting for 61.06 % of individual variance, was significant ($P << 0.01$). All five genes were significantly and positively correlated with principal component 1 (Table 2.4, Figure 2.3). The condition of senescent fish (Figure 2.1) was not significantly correlated with expression of any gene (i.e., more visibly senescent fish did not have higher expression values) but the condition vector was most closely associated with GnRHp, the gene associated with reproduction.

Discussion

Variation and senescence

In this study physiological processes associated with reproduction and senescence (olfaction, immune function, reproduction, memory, and aging) were assessed by targeted gene expression analysis. When all genes were analyzed collectively in a principal component analysis pre-senescent and senescent fish could not be significantly differentiated despite some clustering (Figure 2.4). This result is not surprising as significantly greater expression in senescent salmon was only detected for two genes: TERT (aging) and Viperin (immune function; Figure 2.3). Only 36.15% of the variation explained by the first principal component (Table 2.4) was attributed to TERT and Viperin (collectively) possibly explaining the dampened signal of senescence from these two genes when analyzed collectively. However, all five genes showed higher average gene expression and greater variance ($P << 0.01$; Figure 2.4) in expression among senescent salmon collected at their spawning grounds than in pre-senescent salmon collected as they were entering Hansen Creek. Additionally, all genes were significantly and positively correlated with the first principal component, the only significant principal component, which mainly pulls out senescent fish from the primary cluster of individuals.

This suggests that all genes are displaying a signal of senescence albeit some more strongly than others.

There was a wide range of physical condition of senescent fish (Figure 2.1) with some appearing only minimally senescent and others with severe pigment loss and tissue degradation. Greater variation in expression in senescent fish may reflect these differences in health at the time of sampling (Figure 2.1), as we had no way to determine precisely how soon they would have died. Fish with whiter, bleached skin tended to have higher levels of expression for many genes (e.g., 44, 62, and 63) such as NMDA, GnRHp, and Viperin, however there was no correlation between condition and gene expression. Both gene expression and physical differences in sampled fish highlight the diversity of aging and possibly reflect minor sampling differences between fish such as day sampled. Some fish did appear to have higher overall expression (e.g., 64; Table 2.2) and there was a positive correlation in individual expression across all gene comparisons (Table 2.3), however the correlation was only significant for five of the ten comparisons.

Variation in both gene expression and physical health in salmon at the spawning site may also reflect the natural variation in rates of senescence within a population (Carlson et al., 2004; Doctor & Quinn, 2009). Fish may be spawning at the same time, but they could have been in-stream for different numbers of days; some senescing over a period of days and others spending several weeks alive. Although normalizing gene expression values to a constitutively expressed gene (18S) should reduce the signal of individual variation, it is possible that during rapid senescence in sockeye even 18S may not be a suitable normalizer. In one instance 18S gene expression was less stably expressed than other housekeeping genes after chemical exposure of zebrafish (McCurley & Callard, 2008). Here we observed an inter-assay coefficient of variation of 16.1 % for 18S C_t values after fish 9, 11, and 12 were excluded from analyses due to low RNA and cDNA recovery. This CV is within in the generally accepted range of 15 - 30% for gene expression assays from NOAA and USGS, suggesting that the use of 18S as a normalizing gene is acceptable.

Telomerase reverse transcriptase

Telomeres shorten as organisms age due to cell replication resulting in DNA damage and loss associated with cellular aging. Telomerase reverse transcriptase (TERT), also referred to as telomerase, is the catalytic subunit of the enzyme telomerase which is responsible for telomere repair and extension. TERT is currently of interest because increased telomerase activity has been linked to longevity in several species such as lobster (Klapper et al., 1998) and mice (Banks et al., 2010). Additionally, it is thought that the reactivation of telomerase can subdue or deter aging as a result of telomere loss (Bernardes de Jesus & Blasco, 2011). However, unregulated increased telomerase activity has been linked to cell proliferation in cancer (see Greider, 1998). Our understanding of the role of telomerase in longevity and senescence is limited to a few model species such as mice and humans and by the difficulties of studying aging. Since this is the first time TERT gene expression has been observed in an adult salmonid and a rapidly senescing animal it is difficult to interpret the significance of higher telomerase expression.

One possible explanation for high telomerase gene expression is that the gene is no longer being actively translated into the protein telomerase and therefore, telomeres are no longer being actively repaired, contributing to cellular senescence. There is potential for future research examining telomerase and telomere dynamics in salmon as they are rapidly senescing and have population-specific aging rates (Carlson et al., 2007). Moreover, salmon that arrive on the spawning grounds early in the season have slower senescence rates than those arriving later (Perrin & Irvine, 1990; Carlson et al., 2007; Doctor et al. 2009), and there is a high heritability for breeding date (e.g., Quinn et al., 2000; 2011) so there may be adaptive variation in senescence within as well as among populations. One first step might be to examine telomere length inheritance (similarly to Olsson et al., 2011) and determine if different telomere lengths are inherited by salmon that senesce at different rates.

NMDA-type glutamate receptor

The NMDA (N-methyl-D-aspartate) glutamate receptor is one of three primary groups of glutamate receptors responsible for excitatory neurotransmission (Nakanish, 1992) and is involved in synaptic plasticity and memory (Li & Tsien, 2009). Although

the difference in expression of NMDA between pre-senescent and senescent salmon observed here was not significant statistically, the gene may be of interest because homing requires olfactory memory of natal streams (Dittman et al., 1996; Dittman et al., 1997). Additionally, the NMDA glutamate receptors have been linked to neurodegenerative disorders in humans (Li & Tsien, 2009). In sockeye salmon, neuronal degeneration has been noted during up-stream migration (i.e., senescence) and at the spawning grounds (Carruth et al., 2002). Determining the mechanisms of neuronal degeneration at a gene level is difficult in salmon without complete and annotated genomic resources. However, it could be investigated using immunohistochemical techniques such as in-situ hybridization to "observe" where or if genes are expressed during migration and rapid senescence.

OMP1

Salmon use olfaction to return to their natal streams during their spawning migration using olfactory receptor neurons (ORN) to transmit signals of specific odors of amino acids and bile salts to their brain (Doving et al., 1985; Carruth et al., 2002). ORN function depends on the presence of olfactory marker proteins (OMP) which, when removed from mice, dampened and impaired odor detection (Ivic et al., 2000; Youngentob et al., 2001). Although the role of OMP in salmon homing has not been elucidated, they are present in sockeye salmon olfactory epithelium tissue (Kudo et al., 2009). The salmon sampled for this study were nearing the end of their spawning migration but olfaction was still very active, as evidenced by the fact that sockeye salmon displaced from their spawning grounds within Hansen Creek returned to their breeding site (Stewart et al., 2004). In addition, male salmon use olfaction to distinguish pre- from post-ovulatory females (Honda 1982) and male salmon seek reproductive opportunities throughout their lives on the spawning grounds. Olfaction probably plays an important role in both pre-senescent and senescent fish, consistent with the lack of significant differences in expression of OMP1 between pre-senescent and senescent salmon.

GnRHp

Part of the complexity of senescence in salmon is that while they are rapidly deteriorating in physical health they are also preparing to reproduce. The gonadotropin-releasing hormone precursor (GnRHp) is part of the GnRH axis which plays a critical role in reproduction. In mammals, there is often a pulse in the GnRH axis resulting in a downstream increase in GnRH expression associated with reproduction. This axis has been the subject of study in the development in fish (Palevitch et al., 2007) and in the migration of Pacific salmon (Parhar et al., 1994). For example, in chum (Onuma et al., 2010) and sockeye salmon (Kitani et al., 2003), GnRH was upregulated in adults during their return migration in the ocean. Onuma et al. (2010) hypothesized that the strong signal of GnRH expression in these oceanic salmon was linked to the initiation of migration. Unlike these studies, which observed differences in GnRH after several weeks, we sought to detect changes in expression over a much shorter time scale (days). This may explain why significant differences in GnRHp expression were not detected here. Also, Onuma et al. (2010) sampled salmon at sea, not in freshwater during the final weeks or days of their lives as we did. Because the GnRH pathway plays an important role in reproduction, there may be some regulation of this pathway even in senescent fish.

Viperin

Viperin (virus inhibitory protein, endoplasmic reticulum-associated, interferon-inducible) is an anti-viral protein involved in the innate immune response. Gene homologs exist for several species of vertebrates and invertebrates and it was first characterized in teleosts after exposure of rainbow trout to fish rhabdovirus (Boudinot et al., 1999). More recent studies have shown that viperin expression varies in response to different viruses and even bacteria in humans (Dang et al., 2010). We found that Viperin expression was higher in senescent salmon. It is possible that expression is lower in pre-senescent salmon because mRNA is being actively translated into proteins to ward off viral infections until successful reproduction can be insured. Viperin expression also appears to be higher in fish that are more visibly senescent (e.g. 44, 62, and 63; Figure 2.1) suggesting that body bleaching may be a physical indicator of a heightened immune response. It would be interesting to couple Viperin gene expression data with

morophometric/colormetric data and physiological samples such as viral load in the future.

Conclusions

Understanding the molecular mechanisms and controls of senescence remains the goal of aging research. Pacific salmon provide a unique opportunity to study this process as they undergo rapid senescence, declining from a visually healthy individual entering a stream to death in a matter of days or a few weeks. Understanding Pacific salmon senescence not only provides insight into the aging process, but is also important for understanding the survival of this commercially and culturally important species. To investigate the process of senescence in sockeye salmon we used publically available genomic resources to develop five new gene expression assays for different physiological pathways. These new assays can be used for investigating these processes in sockeye salmon at any point in their life cycle. Here we successfully used them to investigate changes in pathways related to reproduction, homing, immune response, aging, and memory. Higher expression of all five genes in these pathways in senescent salmon suggests that there are many processes contributing to senescence and further highlights the complexity of this process. The significantly higher expression of immune and aging genes warrants further investigation of these pathways and their role in aging.

Table 2.1. Primer data for the normalizing gene 18S and the five genes of interest in the process of senescence with their associated physiological response.

Gene	Accession #	Response	Forward Primer	Reverse Primer	Amplicon size
18S ribosomal RNA	AF308735	-	AAACGGCTACCACATCCAAG	TCCCGAGATCCAACTACGAG	247
Viperin	NM_001124253	immune	ACTCAACCCTGTACGCTGGA	TGGATCTTTCCTCCCCTCTC	244
NMDA-type glutamate receptor 1 subunit (NMDA)	AB292234.1	memory	GAGAGCTGTTTTTCCGTTCG	GAAGATCCCTGCTGCTATGC	239
olfactory marker protein 1 (OMP1)	AB490250.1	olfactory	GACCCCTGACCTCACACACT	GTACATGACCTTGCGGACCT	169
telomerase reverse transcriptase (TERT)	CX246542	aging	AAAGGCTTTGTGATGGCACT	CGTTTTCTTCCCAAACTGGA	151
GnRH Precursor (GnRHp)	D31868	reproduction	ATTGGTCGTATGGGTGGCTA	TCTTGAATGCTCCATCATCG	226

Table 2.2. Fold over minimum values for each fish and each gene. Greater shading within a cell reflects a larger F/M value and greater expression of that gene in that fish. The condition of each fish ranked by extent of senescence (Figure 2), with 3 being the most visibly senescent and 1 being the least visibly senescent, for all fish with pictures available. Sample 9, 11, and 12 were excluded from analyses due to low RNA and cDNA recovery.

ID	Status	Condition	TERT	NMDA	OMP1	GnRHp	Viperin
On09_01	Pre-senescent	n/a	1.87	1.30	5.66	1.97	1.00
On09_02	Pre-senescent	n/a	9.23	7.17	28.87	13.60	17.56
On09_03	Pre-senescent	n/a	1.00	1.82	10.73	1.00	39.34
On09_04	Pre-senescent	n/a	8.70	6.77	32.11	18.90	66.74
On09_05	Pre-senescent	n/a	7.48	6.21	31.16	6.41	50.51
On09_06	Pre-senescent	n/a	6.63	3.76	11.80	18.52	20.04
On09_07	Pre-senescent	n/a	2.44	1.00	19.63	10.36	2.42
On09_08	Pre-senescent	n/a	11.92	4.91	20.97	55.23	128.84
On09_09	Pre-senescent	n/a				51.04	
On09_10	Pre-senescent	n/a	22.05	12.85	40.26	156.07	141.55
On09_11	Pre-senescent	n/a	227.63	60.13		834.93	2912.57
On09_12	Pre-senescent	n/a	1051.26	92.85	2551.34	7660.39	3129.04
On09_13	Pre-senescent	n/a	8.26	9.32	18.94	56.62	51.35
On09_14	Pre-senescent	n/a		8.85		447.95	203.26
On09_44	Senescent	3	9.77	18.23	106.98	139.64	31337.70
On09_46	Senescent	1	4.80	14.38	39.47	84.20	10216.74
On09_47	Senescent	2	10.91	2.84	10.90	27.99	1625.29
On09_52	Senescent	3	8.82	5.86	43.50	33.32	4439.32
On09_53	Senescent	1	10.32	5.38	40.96	5.95	373.25
On09_54	Senescent	n/a	15.41	5.66	48.82	26.35	91.70
On09_55	Senescent	1	44.41	6.93		582.22	221.66
On09_57	Senescent	1	8.54	1.92	19.85	1.76	13.69
On09_59	Senescent	3	19.43	8.27	75.03	196.27	808.18
On09_60	Senescent	2	14.76	4.41	22.47	13.57	336.91
On09_61	Senescent	n/a	20.02	8.07	52.67	69.13	7297.69
On09_62	Senescent	3	30.63	34.37	178.40	1014.51	4759.96
On09_63	Senescent	3	16.05	4.84	1.00	541.83	18223.95
On09_64	Senescent	n/a	80.75	56.04	9014.49	149.66	18099.33

Table 2.3. Correlation between expression values for each gene. Significant correlation values are indicated with an asterisk.

	TERT	NMDA	OMP1	GnRHp	Viperin
TERT	-	0.815*	0.835*	0.399	0.319
NMDA		-	0.825*	0.401*	0.514*
OMP1			-	0.25	0.382
GnRHp				-	0.202
Viperin					-

Table 2.4. Each of the five genes used for this study ordered by greatest contribution to principal component 1. Correlations and measures of significance (p – value) are provided.

Gene	Contribution	Correlation	P - value
NMDA	27.61	0.950	3.82 E-13
TERT	29.58	0.918	1.02 E -10
OMP1	25.10	0.875	1.01 E -08
Viperin	6.57	0.583	2.33 E-03
GnRHp	11.13	0.448	2.47 E-02

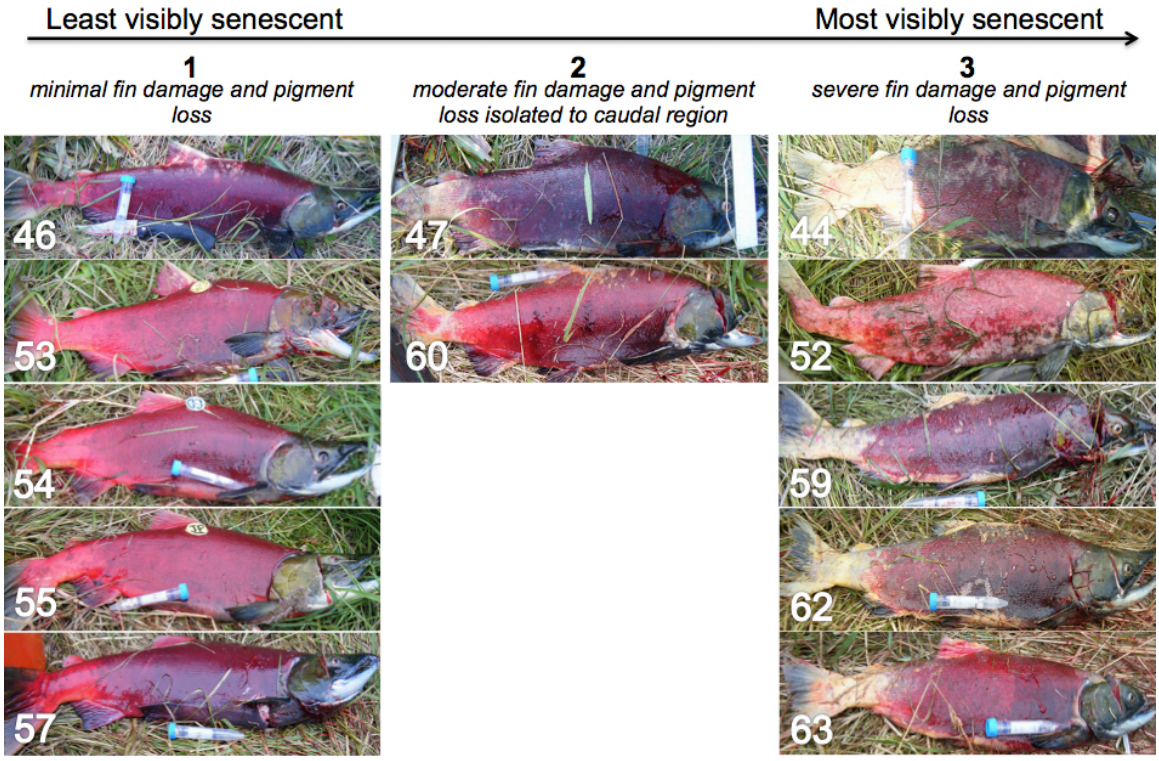

Figure 2.1. Images of senescent salmon sampled at their spawning grounds ranked by condition. Each image is labeled with its corresponding ID from Table 2.

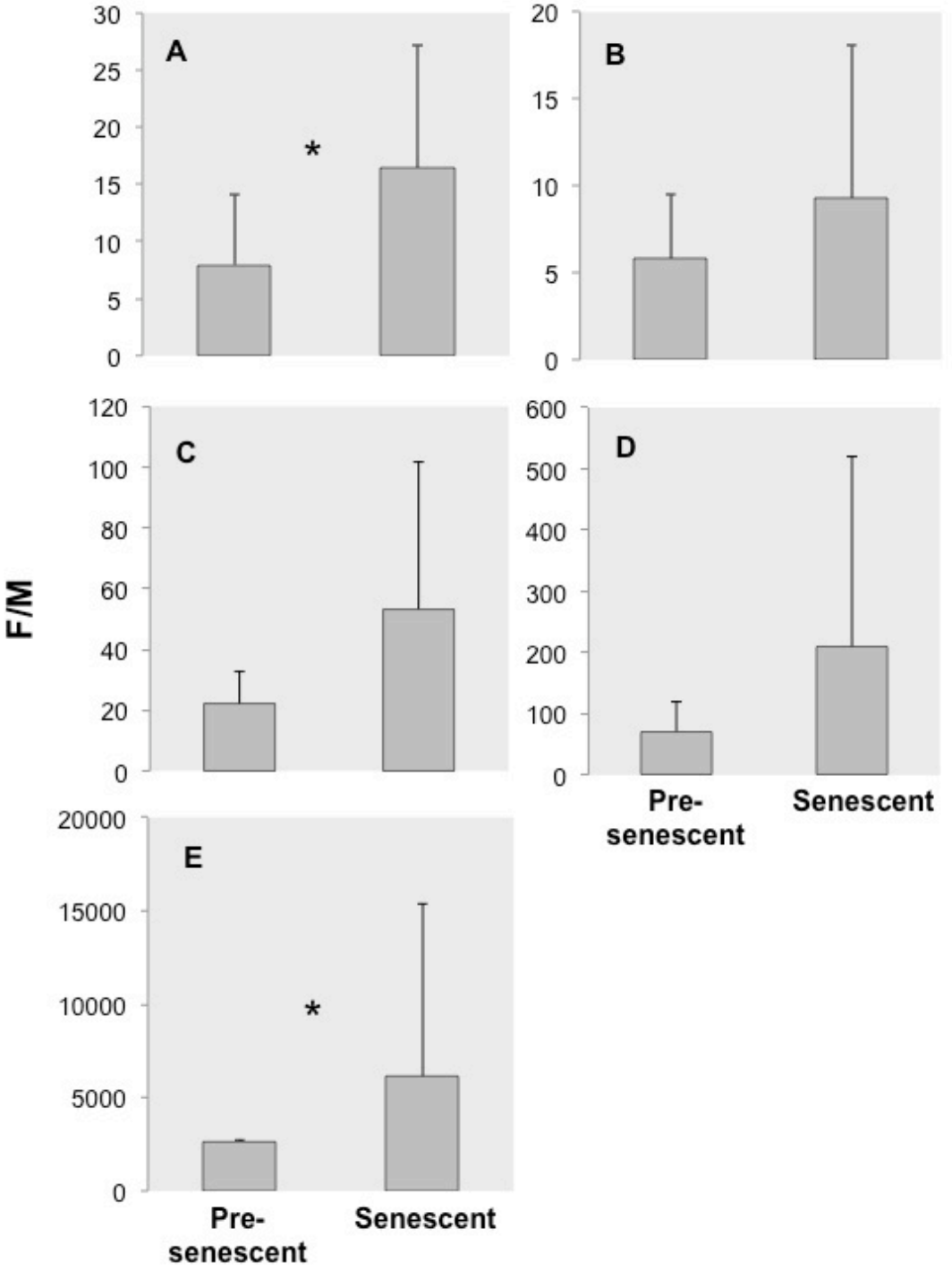

Figure 2.2. Average fold increase over the minimum R_0 value (F/M) for TERT (A), NMDA (B), OMP1(C), GnRHp (D), and Viperin (E) in pre-senescent and senescent salmon. Significance ($P \leq 0.05$) is denoted with a star.

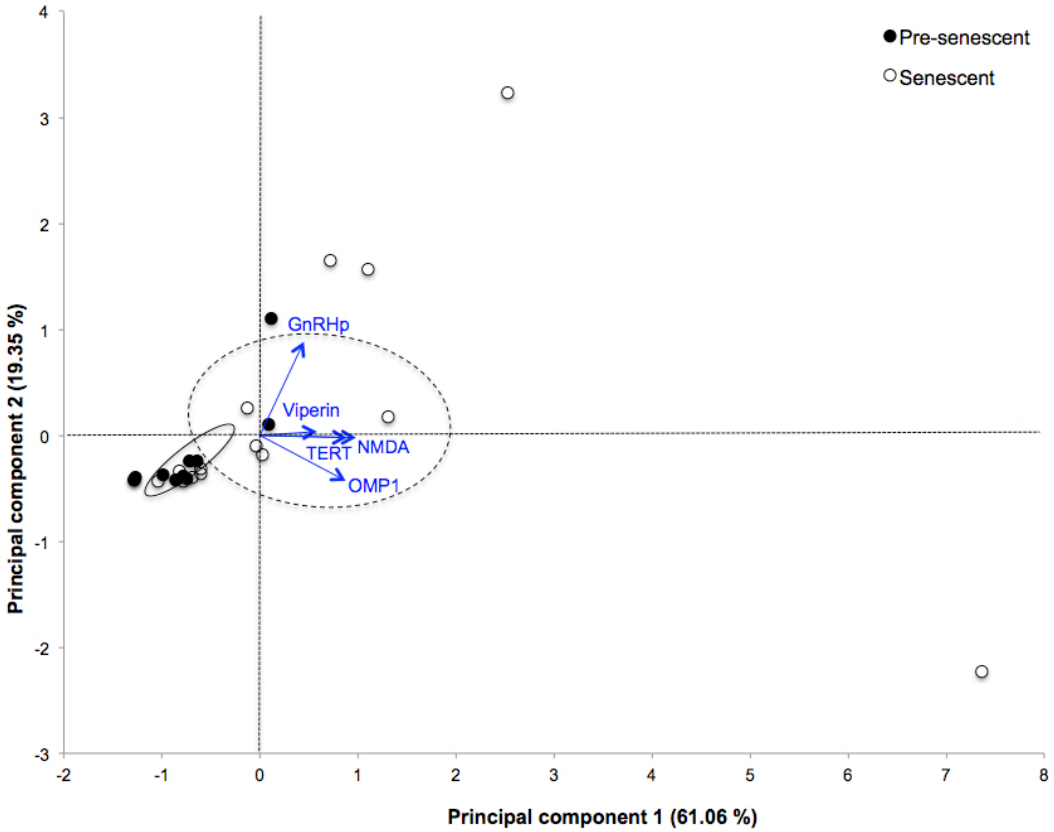

Figure 2.3. Principal component analysis ordination of pre-senescent (closed circles) and senescent (open circles) individuals for all five genes encircled by 95% confidence ellipses. Vectors of gene expression shown in blue. Only principal component 1 is significant ($P < 0.01$).

REFERENCES

Ackerman MW, Habicht C, Seeb LW (2011) Single-Nucleotide Polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. Transactions of the American Fisheries Society 140: 865-881.

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Molecular Ecology Resources 10: 701-710.Aubert G, PM Lansdrop (2008) Telomeres and Aging. Physiological Reviews 88, 557-579.

Armanios M, JK Alder, EM Parry, B Karim, MA Strong, & CW Grider (2009) Short telomeres are sufficient to cause the degenerative defects associated with aging. The American Society of Human Genetics 85, 823-832.

Baker T, Sands T, West F, Westing C, Brazil C (2009) Management of the Nushagak district sockeye salmon fishery: how 50 years of data helps. In: Krueger C, Zimmerman C, editors. Pacific salmon: Ecology and Management of Western Alaska Populations. Bethesda, Maryland: American Fisheries Society. pp. 433-461.

Banks MA, Eichert W, Olsen JB (2003) Which genetic loci have greater population assignment power? Bioinformatics 19: 1436-1438.

Banks WA, JE Morley, SA Farr, et al. (2010) Effects of a growth hormone-releasing hormone antagonist on telomerase activity, oxidative stress, longevity, and aging in mice. Proceedings of the National Academy of Sciences 107: 22272-22277.

Bas A, G Forsberg, S Hammarstrom, & ML Hammarstrom (2004) Utility of the Housekeeping Genes 18S rRNA, b-Actin and Glyceraldehyde-3-Phosphate-Dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. Scandinavian Journal of Immunology 59: 566–573.

Bayir H (2005) Reactive oxygen species. Critical Care Medicine 33, S498-S501.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London Series B-Biological Sciences 263: 1619-1626.

Bernardes de Jesus, B & MA Blasco (2011) Aging by telomere loss can be reversed. Cell Stem Cell doi: 10.1016/j.stem.2010.12.013.

Boudinot P, P Massin, M Blanco, S Riffault, A Benmansour (1999) Vig-1, a new fish gene induced by the rhabdovirus glycoprotein, has a virus-induced homologue in humans and shares conserved motifs with the MoaA family. J Virol 73: 1846e52.

Brett, J.R. 1995. Energetics. *In* Physiological Ecology of Pacific Salmon. *Edited by* C Groot, L Margolis, and W C Clarke. University of British Columbia Press, Vancouver. pp. 1-68.

Bromaghin JF (2008) BELS: backward elimination locus selection for studies of mixture composition or individual assignment. Molecular Ecology Resources 8: 568-571.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution 18: 249-256.

Campbell NR, Narum SR (2011) Development of 54 novel single-nucleotide polymorphism (SNP) assays for sockeye and coho salmon and assessment of available SNPs to differentiate stocks within the Columbia River. Molecular Ecology Resources 11: 20-30.

Carlson, S.C., Rich, H.B., Jr, and Quinn, T.P. 2004. Reproductive lifespan and sources of mortality for alternative male life history strategies in sockeye salmon, *Oncorhynchus nerka*. Canadian Journal of Zoology **82**: 1878-1885.

Carlson SM, R Hilborn, AP Hendry, & TP Quinn (2007) Predation by bears drives senescence in natural populations of salmon. PLos One 2, e1286.

Carruth LL, RA Dores, TA Maldonado, DO Norris, T Ruth, & RE Jones (2000) Elevation of plasma cortisol during the spawning migration of landlocked kokanee salmon (*Oncorhynchus nerka kennerlyi*). Comp Biochem Physiol C127: 123–131

Carruth LL, RE Jones, & DO Norris (2002) Cortisol and pacific salmon: a new look at the role of stress hormones in olfaction and home-stream migration. Integrative and Comparative Biology 42:574–581.

Chen J, CN Hales, & SE Ozanne (2007) DNA damage, cellular senescence and organismal ageing: causal or correlative? Nucleic Acids Research 35: 7417-7428.

Creelman EK, Hauser L, Simmons RK, Templin WD, Seeb LW (2011) Temporal and geographic genetic divergence: characterizing sockeye salmon populations in the Chignik watershed, Alaska, using single-nucleotide polymorphisms. Transactions of the American Fisheries Society 140: 749-762.

Dai Z, Papp AC, Wang D, Hampel H, Sadee W (2008) Genotyping panel for assessing response to cancer chemotherapy. Bmc Medical Genomics 1.

Dang W, M Zhang, Y Hu, & L Sun (2010) Differential regulation of *Sciaenops ocellatus* viperin expression by intracellular and extracellular bacterial pathogens. Fish & Shellfish Immunology 29:264-270.

Dittman AH & TP Quinn (1996) Homing in Pacific salmon: mechanisms and ecological basis. J. Exp. Biol. 199:83–91.

Dittman, A.H., Quinn, T.P., Nevitt, G.A., Hacker, B., and Storm, D.R. 1997. Sensitization of olfactory guanylyl cyclase to a specific imprinted odorant in coho salmon. Neuron **19**: 381-389.

Doctor, K.K., and Quinn, T.P. 2009. The potential for adaptation-by-time in sockeye salmon (*Oncorhynchus nerka)*: The interactions of body size and in-stream reproductive lifespan with date of arrival and breeding location. Canadian Journal of Zoology **87**: 708-717.

Elfstrom CM, Smith CT, Seeb JE (2006) Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon. Molecular Ecology Notes 6: 1255-1259.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, et al. (2010) Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences of the United States of America 107: 16196-16200.

Evans TG, E Hammill, K Kaukinen, et al. (2011) Transcriptomics of environmental acclimatization and survival in wild adult Pacific sockeye salmon (*Oncorhynchus nerka*) during spawning migration. Molecular Ecology 20: 4472-4489.

Everett MV, Grau ED, Seeb JE (2011) Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. Molecular Ecology Resources 11: 93-108.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity 103: 285-298.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567.

Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Molecular Ecology Resources 11: 254-267.

Finch CE (1994) Longevity, senescence, and the genome. University of Chicago Press, Chicago, USA.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. Molecular Ecology Resources 10: 915-934.

Gende, S.M., Quinn, T.P., Willson, M.F., Heintz, R., and Scott, T.M. 2004a. Magnitude and fate of salmon-derived nutrients and energy in a coastal stream ecosystem. Journal of Freshwater Ecology **19**: 149-160.

Gende SM, TP Quinn, R Hilborn, AP Hendry, & B Dickerson (2004b) Brown bears selectively kill salmon with higher energy content but only in habitats that facilitate choice. Oikos 104: 518-528.

Geraldes A, Pang J, Thiessen N, Cezard T, Moore R, et al. (2011) SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. Molecular Ecology Resources 11: 81-92.

Glover KA, Hansen MM, Lien S, Als TD, Hoyheim B, et al. (2010) A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. Bmc Genetics 11.

Gomez-Uchida D, Seeb JE, Smith MJ, Habicht C, Quinn TP, et al. (2011) Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. Bmc Evolutionary Biology 11.

Gotz ME., Malz CR., Dirr A., et al. Brain aging phenomena in migrating sockeye salmon *Oncorhynchus nerka*. J Neural Transm. 2005;112:1177–1199.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86: 485-486.

Greider CW (1998) Telomerase activity, cell proliferation, and cancer. Proceedings of the National Academy of Sciences 95:90-92.

Habicht C, Seeb LW, Myers KW, Farley EV, Seeb JE (2010) Summer-Fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. Transactions of the American Fisheries Society 139: 1171-1191.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, et al. (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular Ecology Resources 11: 123-136.

Hendry AP, OK Berg, & TP Quinn (1999) Condition dependence and adaptation by- time: breeding date, life history, and energy allocation within a population of salmon. Oikos 85: 499-514.

Hodes RJ (1999) Telomere length, aging, and somatic cell turnover. The Journal of Experimental Medicine 190: 153-156.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular Ecology Resources 11: 117-122.

Honda H (1982) On the female pheromones and courtship behavior in the salmonids, *Oncorhynchus masou* and *O. rhodurus*. Bulletin of the Japanese Society of Scientific Fisheries 48: 47-49.

Hornsby PJ (2006) Short telomeres: cause or consequence of aging? Aging Cell 5: 577–578.

Francois Husson, Julie Josse, Sebastien Le and Jeremy Mazet (2009) FactoMineR: Factor Analysis and Data Mining with R. R package version 1.12. http://CRAN.R-project.org/package=FactoMineR

Ivi, L, MM Pyrski, JW Margolis, LJ Richards, S Firestein, & FL Margolis (2000) Adenoviral vector-mediated rescue of the OMP-null phenotype in vivo. Nat. Neurosci. 3:1113–1120.

Jaskelioff M, FL Muller, JH Paik, et al. (2010) Telomerase reactivation reverses tissue degenerationin aged telomerase-deficient mice. Nature 10.1038/nature09603.

Kalinowski ST, Manlove KR, Taper ML (2008) ONCOR: A computer program for genetic stock identification. http://www.montana.edu/kalinowski/Software/ONCOR.htm

Karlsson S, Moen T, Lien S, Glover KA, Hindar K (2011) Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Molecular Ecology Resources 11: 247-253.

Kim SK (2007) Common aging pathways in worms, flies, mice, and humans. The Journal of Experimental Biology 210:1607-1612.

Kitani T, S Matsumoto, H Yamada, M Amano, M Iwata, & H Ueda (2003) Changes in GnRH levels in the brain and pituitary gland during migrations of sockeye salmon and chum salmon. Fish Physiology and Biochemistry 28:269-270.

Klapper W, K Kuëhne, KK Singh, K Heidorn, R Parwaresch, & G Krupp (1998) Longevity of lobsters is linked to ubiquitous telomerase expression. Federation of European Biochemical Societies 439: 143-146.

Kregel KC & HJ Zhang (2007) An integrated view of oxidative stress in aging: basic mechanisms, functional effects, and pathological considerations. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology 292: R18-R36.

Kudo H, Y Doi, H Ueda, & M Kaeriyama (2009) Molecular characterization and histochemical demonstration of salmon olfactory marker protein in the olfactory epithelium of

lacustrine sockeye salmon (*Oncorhynchus nerka*). Comparative Biochemistry and Physiology 154:142-150.

Li F & JZ Tsien (2009) Clinical Implications of Basic Research: Memory and the NMDA receptors. N Engl J Med, 361:302.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics 4: 981-994.

Maldonado TA, RE Jones & DO Norris (2002) Timing of neurodegeneration and beta-amyloid (Ab) peptide deposition in the brain of aging kokanee salmon. Journal of Neurobiolgy 53: 21–35.

Martinez DA, Nelson MA (2010) The Next Generation Becomes the Now Generation. Plos Genetics 6.

McCurley AT & GV Callard (2008) Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. BMC Molecular Biology 9: doi:10.1186/1471-2199-9-102.

McGlauflin MT, Smith MJ, Wang JT, Young SF, Chen N, et al. (2010) High-resolution melting analysis for the discovery of novel single-nucleotide polymorphisms in rainbow and cutthroat trout for species identification. Transactions of the American Fisheries Society 139: 676-684.

McGlauflin MT, Schindler DE, Seeb LW, Smith CT, Habicht C, et al. (2011) Spawning habitat and geography influence population structure and juvenile migration timing of sockeye salmon in the Wood River Lakes, Alaska. Transactions of the American Fisheries Society 140: 763-782.

Moazami-Goudarzi K, Laloe D (2002) Is a multivariate consensus representation of genetic relationships among populations always meaningful? Genetics 162: 473-484.

Morin PA, Luikart G, Wayne RK, Grp SNPW (2004) SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution 19: 208-216.

Morin PA, Pease VL, Hancock BL, Robertson KM, Antolik CW, et al. (2010) Characterization of 42 single nucleotide polymorphism (SNP) markers for the bowhead whale (*Balaena mysticetus*) for use in discriminating populations. Marine Mammal Science 26: 716-732.

Miller KA, S Li, K Kaukinen, et al. (2011) Genomic signatures predict migration and spawning failure in wild Canadian salmon. Science 331: 214–217.

Morbey YE, CE Brassil, & AP Hendry (2005) Rapid senescence in Pacific salmon. American Naturalist 166: 556–568.

Nakanishi S (1992) Molecular diversity of glutamate receptors and implications for brain function. Science 258: 597-603.

Nicot N, J Hausman, L Hoffmann & D Evers (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of Experimental Botany 56:2907-2914.

Olsen MT, Volny VH, Berube M, Dietz R, Lydersen C, et al. (2011) A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Arctic ringed seal (*Pusa hispida hispida*). Molecular Ecology Resources 11: 9-19.

Olsson M, A Pauliny, E Wapstra, T Uller, T Schwartz, & D Blomqvist (2011) Sex differences in sand lizard telomere inheritance: paternal epigenetic effects increases telomere heritability and offspring survival. PLos One 6, e17473.

Onuma TA, K Makino, H Ando, M Ban, M Fukuwaka, T Azumaya, & A Urano (2010) Expression of GnRH genes is elevated in discrete brain loci of chum salmon before initiation of homing behavior and during spawning migration. General and Comparative Endocrinology 168:356-368.

Palevitch O, K Kight, E Abraham, S Wray, Y Zohar, & Y Gothilf (2007) Ontogeny of the GnRH systems in zebrafish brain: in situ hybridization and promoter-reporter expression analyses in intact animals. Cell Tissue Research 327:313-322.

Parhar IS, N Koibuchi, M Sakai, M Iwata, S Yamaoka (1994) Gonadotropin-releasing hormone (GnRH): expression during salmon migration. Neuroscience Letters 172:15-18.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.

Perkel J (2008) SNP genotyping: six technologies that keyed a revolution. Nature Methods 5: 447-453.

Perrin CJ & JR Irvine (1990) A review of survey life estimates as they apply to the area-under-the-curve method for estimating the spawning escapement of Pacific salmon. Canadian Technical Report of Fisheries and Aquatic Sciences 1733: 1-49.

Quinn TP, L Wetzel, S Bishop, K Overberg, & DE Rogers (2001) Influence of breeding habitat on bear predation, and age at maturity and sexual dimorphism of sockeye salmon populations. Canadian Journal of Zoology 79: 1782-1793.

Quinn TP (2005) The behavior and ecology of Pacific salmon and trout. University of Washington Press, Seattle.

Quintela M, Berlin S, Wang B, Hoglund J (2010) Genetic diversity and differentiation among *Lagopus lagopus* populations in Scandinavia and Scotland: evolutionary significant units confirmed by SNP markers. Molecular Ecology 19: 2380-2393.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43: 223-225.

Rosenberg NA (2005) Algorithms for selecting informative marker panels for population assignment. Journal of Computational Biology 12: 1183-1201.

Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured populations: A case study in the eastern fence lizard. Journal of Heredity 98: 331-336.

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

Sagarin R, Carlsson J, Duval M, Freshwater W, Godfrey MH, et al. (2009) Bringing molecular tools into environmental resource management: untangling the molecules to policy pathway. Plos Biology 7: 426-430.

Salomons HM, Mulder GA, van de Zande L, Haussmann MF, Linkskens MHK, & Verhulst S (2009) Telomere shortening and survival in free-living corvids. Proceedings of the Royal Society of London. Series B, Biological Sciences 276, 3157–3165.

Schlotterer C (2004) The evolution of molecular markers - just a matter of fashion? Nature Reviews Genetics 5: 63-69.

Seeb LW, Habicht C, Templin WD, Tarbox KE, Davis RZ, et al. (2000) Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the Exxon Valdez oil spill. Transactions of the American Fisheries Society 129: 1223-1249.

Seeb JE, Pascal CE, Ramakrishnan R, Seeb LW (2009) SNP genotyping by the 5 '-nuclease reaction: advances in high-throughput genotyping with nonmodel organisms. Single Nucleotide Polymorphisms: Methods and Protocols, Second Edition 578: 277-292.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, et al. (2011a) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Molecular Ecology Resources 11: 1-8.

Seeb LW, Seeb JE, Habicht C, Farley EV, Jr., Utter FM (2011b) Single-nucleotide polymorphic genotypes reveal patterns of early juvenile migration of sockeye salmon in the eastern Bering Sea. Transactions of the American Fisheries Society 140: 734-748.

Seeb JE, Pascal CE, Grau ED, Seeb LW, Templin WD, et al. (2011c) Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. Molecular Ecology Resources 11: 335-348.

Shigenaga MK, TM Hagen, BN Ames (1994) Oxidative damage and mitochondrial decay in aging. Proceedings of the National Academy of Sciences 91: 10771–10778.

Smith CT, Elfstrom CM, Seeb LW, Seeb JE (2005) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14: 4193-4203.

Smith CT, Antonovich A, Templin WD, Elfstrom CM, Narum SR, et al. (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 MJ, Pascal CE, Grauvogel Z, Habicht C, Seeb JE, et al. (2011) Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. Molecular Ecology Resources 11: 268-277.

Sohal RS, RJ Mockett, & WC Orr (2002) Mechanisms of aging: an appraisal of the oxidative stress hypothesis. Free Radical Biology and Medicine 33: 575–586.

SPSS for Windows, Rel. 11.0.1. 2001. Chicago: SPSS Inc.

Stewart IJ, Carlson SM, Boatright CP, Buck GB, Quinn TP (2004) Site fidelity of spawning sockeye salmon (*Oncorhynchus nerka* W.) in the presence and absence of olfactory cues. Ecology of Freshwater Fish, 13, 104–111.

Topchy A, Scribner K, Punch W (2004) Accuracy-driven loci selection and assignment of individuals. Molecular Ecology Notes 4: 798-800.

Untergasser A, H Nijveen, X Rao, T Bisseling, R Geurts, & JAM Leunissen (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research 35:W71-W74.

Villar-Pique A, NS de Groot, R Sabate, et al. (2011) The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity. Journal of Molecular Biology doi:10.1016/j.jmb.2011.12.014

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. Evolution 38: 1358-1370.

Winans GA, Paquin MM, Van Doornik DM, Baker BM, Thornton P, et al. (2004) Genetic stock identification of steelhead in the Columbia River Basin: an evaluation of different molecular markers. North American Journal of Fisheries Management 24: 672-685.

Wu S-B, Wirthensohn MG, Hunt P, Gibson JP, Sedgley M (2008) High resolution melting analysis of almond SNPs derived from ESTs. Theoretical and Applied Genetics 118: 1-44.

Youngentob SL, FL Margolis, & LM Youngentob (2001) OMP gene deletion results in an alteration in odorant quality perception. Behav. Neurosci. 115: 626–631.

Zhao S, & Fernald RD (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. Journal of Computational Biology 12: 1047-1064.

Appendix A. All 114 loci in alphabetic order with descriptive statistics (Ho, He, & FST) for full 12 SNP-assessment populations. Average rank based on training set individuals and five ranking approaches: genetic differentiation (F_{ST}), Rosenberg's informativness (I_n), average contribution of locus to principal components (LC), and ranks from the locus selection programs BELS and WHICHLOCI. The numeral 1 indicates the highest rank. Locus ranks for each approach are based on a training set of the 12 SNP-assessment populations.

Avg. Rank	Locus name	Reference	H_o	H_e	F_{ST}	F_{ST} Rank	I_n Rank	LC Rank	BELS Rank	WHICHLOCI Rank
43	*One_ACBP-79*	Elfstrom et al. 2006	0.41	0.41	0.126	40	33	47	73	46
18	*One_agt-132*	this study	0.4	0.39	0.231	10	9	9	103	38
106	*One_aldB-152*	Campbell & Narum 2011	0.31	0.32	0.021	102	101	93	29	107
99	*One_ALDOB-135*	Elfstrom et al. 2006	0.25	0.25	0.032	87	80	88	54	91
2	*One_apoe-83*	this study	0.34	0.34	0.316	4	1	2	1	35
	One_bckB-137	Campbell & Narum 2011				monomorphic				
90	*One_c3-98*	this study	0.1	0.09	0.015	104	103	98	46	34
	One_ccd16-131	this study				monomorphic				
47	*One_CD9-269*	this study	0.32	0.34	0.086	54	63	41	32	70
74	*One_cetn1-167*	this study	0.43	0.43	0.1	55	55	61	84	84
12	*One_CFP1*	DFO	0.28	0.28	0.204	20	18	18	47	45
104	*One_cin-177*	Campbell & Narum 2011	0.45	0.47	0.051	75	74	103	87	85
55	*One_ctgf-301*	Elfstrom et al. 2006	0.06	0.06	0.057	88	95	82	13	12
63	*One_dds-529*	Campbell & Narum 2011	0.45	0.43	0.084	74	76	49	30	81
107	*One_DDX5-86*	this study	0.43	0.45	0.019	103	104	108	28	108
95	*One_E2-65*	Smith et al. 2005	0.25	0.24	0.053	91	86	76	44	95
70	*One_gadd45-269*	Campbell & Narum 2011	0	0	0	109	109	106	2	1
86	*One_gdh-212*	Campbell & Narum 2011	0.48	0.46	0.06	71	69	85	66	75
24	*One_GHII-2165*	Elfstrom et al. 2006	0.26	0.26	0.226	12	14	10	100	44
82	*One_ghsR-66*	Campbell & Narum 2011	0.37	0.38	0.082	66	54	77	80	76
101	*One_GPDH-201*	Smith et al. 2005	0.46	0.47	0.043	85	87	97	52	98
66	*One_GPDH2-187*	Smith et al. 2005	0.23	0.22	0.084	53	48	58	86	71
10	*One_GPH-414*	Elfstrom et al. 2006	0.36	0.36	0.218	9	7	8	85	36

Appendix A. Continued.

57	*One_HGFA-49*	Smith et al. 2005	0.32	0.34	0.065	65	73	43	37	74
73	*One_HpaI-71*	Elfstrom et al. 2006	0.43	0.44	0.103	58	58	74	74	72
8	*One_HpaI-99*	Elfstrom et al. 2006	0.24	0.23	0.366	2	2	3	89	37
108	*One_hsc71-220*	Elfstrom et al. 2006	0.35	0.34	0.037	92	88	94	83	101
92	*One_Hsp47*	DFO	0.28	0.29	0.062	94	94	72	39	89
48	*One_Ig-90*	this study	0.1	0.09	0.062	67	70	62	45	17
93	*One_IL8r-362*	Habicht et al. 2010	0.15	0.14	0.049	83	79	87	36	103
71	*One_ins-107*	Smith et al. 2005	0.46	0.45	0.073	81	75	59	53	59
37	*One_KCT1-453*	this study	0.3	0.29	0.12	37	53	29	35	61
105	*One_KPNA-422*	Elfstrom et al. 2006	0.33	0.33	0.018	106	106	104	10	104
	One_LEI-87	Elfstrom et al. 2006				monomorphic				
44	*One_leptin-92*	Campbell & Narum 2011	0.45	0.45	0.088	50	45	46	72	31
89	*One_lpp1-44*	this study	0.32	0.31	0.059	79	77	70	61	90
64	*One_MARCKS-241*	Habicht et al. 2010	0.04	0.04	0.029	100	102	95	5	10
6	*One_metA-253*	Campbell & Narum 2011	0.13	0.14	0.204	16	16	19	69	8
60	*One_MHC2_190*	Elfstrom et al. 2006	0.4	0.41	0.131	45	42	79	90	51
28	*One_MHC2_251*	Elfstrom et al. 2006	0.35	0.35	0.204	13	11	22	104	40
59	*One_Mkpro-129*	Campbell & Narum 2011	0.47	0.46	0.059	70	62	56	38	79
72	*One_ODC1-196*	this study	0.44	0.45	0.058	73	72	55	57	77
20	*One_Ots213-181*	Elfstrom et al. 2006	0.31	0.31	0.158	24	25	28	42	55
26	*One_p53-534*	Elfstrom et al. 2006	0.09	0.08	0.111	35	47	38	56	9
78	*One_parp3-170*	Campbell & Narum 2011	0.01	0.01	0.026	108	108	107	20	2
45	*One_PIP*	DFO	0.45	0.45	0.092	43	41	40	76	48
5	*One_ppie-74*	Campbell & Narum 2011	0.1	0.11	0.364	3	5	4	101	7
38	*One_Prl2*	Elfstrom et al. 2006	0.41	0.43	0.158	27	19	34	93	43
46	*One_psme2-354*	this study	0.3	0.31	0.102	48	35	52	59	63
102	*One_rab1a-76*	this study	0.2	0.2	0.044	99	99	100	15	106

Appendix A. Continued.

3	*One_RAG1-103*	Elfstrom et al. 2006	0.08	0.07	0.166	21	44	12	3	14
67	*One_RAG3-93*	Elfstrom et al. 2006	0.17	0.17	0.064	52	50	48	81	87
49	*One_redd1-414*	Campbell & Narum 2011	0.44	0.45	0.089	69	68	51	7	68
52	*One_RFC2-102*	Smith et al. 2005	0.3	0.3	0.113	39	30	37	98	69
79	*One_RFC2-285*	Smith et al. 2005	0.07	0.06	0.043	95	91	91	50	20
84	*One_RH2op-395*	Elfstrom et al. 2006	0.02	0.02	0.006	107	107	109	25	11
103	*One_rpo2j-261*	Campbell & Narum 2011	0.28	0.29	0.03	98	96	102	22	105
76	*One_sast-211*	Campbell & Narum 2011	0.08	0.09	0.058	76	83	101	51	30
75	*One_serpin-75*	Smith et al. 2005	0.03	0.03	0.013	105	105	96	12	21
80	*One_spf30-207*	Campbell & Narum 2011	0.3	0.3	0.056	82	85	90	4	86
25	*One_srp09-127*	Campbell & Narum 2011	0.13	0.14	0.1	47	49	44	14	29
88	*One_ssrd-135*	Campbell & Narum 2011	0.48	0.48	0.038	90	93	81	9	96
15	*One_STC-410*	Elfstrom et al. 2006	0.29	0.31	0.268	6	3	6	110	28
21	*One_STR07*	Elfstrom et al. 2006	0.37	0.37	0.22	7	6	15	107	39
30	*One_SUMO1-6*	Campbell & Narum 2011	0.28	0.3	0.144	28	21	27	62	54
17	*One_sys1-230*	Campbell & Narum 2011	0.4	0.38	0.258	11	8	7	106	33
36	*One_taf12-248*	Campbell & Narum 2011	0.07	0.07	0.107	49	51	39	68	6
19	*One_Tf_ex11-750*	Elfstrom et al. 2006	0.36	0.39	0.148	29	24	32	41	47
42	*One_Tf_in3-182*	Elfstrom et al. 2006	0.14	0.14	0.132	30	26	63	105	13
109	*One_tshB-92*	Campbell & Narum 2011	0.09	0.09	0.022	101	100	105	43	109
13	*One_txnip-401*	Campbell & Narum 2011	0.07	0.07	0.148	32	37	42	34	5
110	*One_U1002-101*	this study	0	0	0.006	110	110	110	26	110
22	*One_U1003-75*	this study	0.28	0.29	0.275	8	4	5	109	50
65	*One_U1004-183*	this study	0.38	0.39	0.104	51	40	69	94	58
77	*One_U1009-91*	this study	0.31	0.32	0.068	64	65	65	65	82
1	*One_U1010-81*	this study	0.11	0.1	0.319	1	10	1	8	22
85	*One_U1012-68*	this study	0.29	0.29	0.099	63	56	73	88	80
83	*One_U1013-108*	this study	0.23	0.23	0.053	72	71	67	64	83
100	*One_U1014-74*	this study	0.2	0.19	0.035	93	89	99	23	102

Appendix A. Continued.

39	*One_U1016-115*	this study	0.4	0.41	0.134	38	34	31	63	53
4	*One_U1017-62*	this study	0.07	0.07	0.219	18	31	16	49	4
27	*One_U1024-197*	this study	0.24	0.24	0.126	36	39	36	17	60
54	*One_U1101*	this study	0.35	0.35	0.079	68	66	50	48	57
16	*One_U1102-220*	this study	0.2	0.15	0.178	17	22	20	79	23
34	*One_U1103*	this study	0.05	0.05	0.083	44	64	54	24	18
56	*One_U1104-138*	this study	0.02	0.02	0.049	84	98	86	19	3
35	*One_U1105*	this study	0.27	0.29	0.171	15	15	26	97	56
97	*One_U1201-492*	this study	0.49	0.48	0.044	97	97	78	21	100
58	*One_U1202-1052*	this study	0.35	0.34	0.086	60	52	45	70	73
68	*One_U1203-175*	this study	0.39	0.39	0.094	56	57	53	92	64
41	*One_U1204-53*	this study	0.29	0.3	0.125	25	20	24	102	65
61	*One_U1205-57*	this study	0.08	0.08	0.083	61	78	89	60	19
91	*One_U1206-108*	this study	0.31	0.32	0.034	96	92	92	6	99
31	*One_U1208-67*	this study	0.37	0.4	0.129	42	38	25	27	62
23	*One_U1209-111*	this study	0.19	0.17	0.25	5	12	11	108	42
53	*One_U1210-173*	this study	0.14	0.14	0.052	86	82	71	11	25
50	*One_U1211-97*	this study	0.14	0.14	0.068	80	67	66	31	24
51	*One_U1212-106*	this study	0.42	0.44	0.103	41	32	68	95	32
87	*One_U1214-107*	this study	0.17	0.18	0.074	62	60	83	71	92
32	*One_U1215-82*	this study	0.42	0.43	0.154	26	23	17	96	41
98	*One_U1216-230*	this study	0.45	0.47	0.058	78	81	75	67	97
96	*One_U301-92*	Elfstrom et al. 2006	0.24	0.24	0.066	77	84	80	58	93
11	*One_U401-224*	Habicht et al. 2010	0.45	0.44	0.117	31	27	23	16	49
7	*One_U404-229*	Habicht et al. 2010	0.14	0.13	0.136	33	28	35	18	16
33	*One_U502-167*	Habicht et al. 2010	0.09	0.1	0.129	34	43	33	78	15
62	*One_U503-170*	Habicht et al. 2010	0.28	0.29	0.089	46	46	57	91	67
94	*One_U504-141*	Habicht et al. 2010	0.35	0.37	0.041	89	90	84	33	94

Appendix A. Continued.

29	*One_U508-533*	Habicht et al. 2010	0.1	0.12	0.183	22	29	21	40	78
81	*One_vamp5-255*	Campbell & Narum 2011	0.28	0.29	0.108	59	61	64	75	88
14	*One_vatf-214*	Campbell & Narum 2011	0.11	0.11	0.178	19	36	13	55	27
69	*One_VIM-569*	Elfstrom et al. 2006	0.25	0.24	0.105	57	59	60	82	66
	One_zn706-68	Campbell & Narum 2011				monomorphic				
40	*One_ZNF-61*	Habicht et al. 2010	0.35	0.36	0.16	23	17	30	99	52
9	*One_zP3b-49*	Smith et al. 2005	0.27	0.27	0.183	14	13	14	77	26

Appendix B. Forward and reverse primer and probe sequences for all newly developed SNP markers featured in this paper with SNP characterization and gene annotation when possible.

Locus name	NCBI accession	Gene	E-value	VIC/ FAM	Primer sequences	Probe sequences
One_agt-132	445261120	Angiotensinogen (angt)	5.00E-70	A/C	F:GACCCAGATCAACAACTTCATCCA R:TGGTTGAGCTAAGGTCCTTGAAC	VIC: ACAGGAAAATCACGAGCCT FAM:CAGGAAAATCCCGAGCCT
One_apoe-83	445261132	apolipoprotein E (apoe)	1.00E-46	C/T	F:CGCCATGGACAAGGTCAAG R:GGCACAGTGCTTCCAAACC	VIC:TTTAGACGGCGGTCTC FAM:ATTTAGACAGCGGTCTC
One_c3-98	445261138	complement component C3	1.00E-75	C/T	F:GAGTGTGGAACTGGTTCTTGTTG R:GCCGGCAGGGCATCA	VIC:GTTGATGGACCACCTGGT FAM:TTGATGGACCACTTGGT
One_ccd16-131	445261145	Coiled-coil domain-containing protein	7.00E-23	C/T	F:CCGTGACCTGTTGAACTTTGTTTAG R:TCACGTTCTTGGAAAACAGC	VIC:AAGGAGAAAGTTGCCGAGCT FAM:ATAAGGAGAAAGTTACCGAGCT
One_CD9-269	445261152	similar to CD9 antigen	1.00E-41	C/T	F:ACGCTCTGAGGTGATATGAAACAC R:CATCCGACGTCAACATCCAAAC	VIC:TGGAATGGAGAAATC FAM:ATGGAATGAAGAAATC
One_cetn1-167	445261159	Centrin-1 (cetn1)	5.00E-56	A/C	F:CAGAAATCCTGACTGTTAAAACAATG CA R:CTGCTCGTTGATCTCTCCATCTC	VIC:TTGACGAAGCAGACCGA FAM:TTGACGAAGCCGACCGA
One_DDX5-86	445261166	Probable ATP-dependent RNA helicase	2.00E-72	C/T	F:CTCCCACATTGATCTGGACGTA R:TGCCACTTGGCCCAAAGAG	VIC:AGGACTTCCTGAAGGAC FAM:AGGACTTCCTAAAGGAC
One_Ig-90	445261172	Ig kappa chain V region K29-213	3.00E-11	C/G	F:GGATTGTGGTAACTCTGACAGTAGT R:CATCTAAATTCAGTGGCAGTGGGTTA	VIC:CTCCTGCATCTTCAGCC FAM:CCTGCATGTTCAGCC
One_KCT1-453	445261179	3-ketoacyl-CoA thiolase, mitochondrial	2.00E-58	G/T	F:GGGAAAGTATGCTGTGGGATCAG R:GGTTCCTCAGTGAGTGTTCTCTATG	VIC:TGGTCAGGGTATCGCCATA FAM:TGGTCAGGGTATCTCCATA
One_lpp1-44	445261186	Lipid phosphate phosphohydrolase 1	2.00E-38	C/T	F:GGTCCAATAGGGAGCTCAGACA R:GGGAATGAACCAGACATGTGAATG	VIC:TTGTGCTTTCCTGACCTAT FAM:TTGTGCTTTCCTAACCTAT
One_ODC1-196	445261193	ODC1 ornithine decarboxylase gene	1.00E-121	C/T	F:AACTCTGCGTCTGTCTGCTT R:TCAGATGGTTCATTATGACAGCAACA	VIC:CGAACAGGGCTGGATG FAM:CGAACAGGACTGGATG
One_psme2-354	445261200	Proteasome activator complex subunit 2	5.00E-14	A/G	F:TGGTCCTTCAGGTACTTTTCAGAGA R:CAAATGCCAATTCTCACCACATGA	VIC:TGATGCAGTAGCTAAAG FAM:ATGCAGTGGCTAAAG

Appendix B. Continued

One_rab1a-76	445261207	no hits	G/T	F:TCGCCATATTCTCTCTCCCTATCC R:ATCCACTCAGACCCATATCTACCAA	VIC:TGTGGAGCAAGGTAACT FAM:TGTGGAGCAATGTAACT
One_U1002-101	445261214	no hits	G/T	F:GCCAACCCTATACTGTACGGATTTTT R:TCCGTTGCATTGTCCATCCA	VIC:TCGTTCCAAAGAATGTTGTG FAM:CGTTCCAAAGAATTTTGTG
One_U1003-75	445261221	no hits	C/T	F:TCACGAGCCCCAGTCAGA R:CGGGTTTCGGTGGTTTAGTATTCTA	VIC:AGAGACTACTTCCTTTTTG FAM:AGAGACTACTTCTTTTTTG
One_U1004-183	445261228	no hits	A/G	F:GGTGTGACTGCTGTGTTTAATTGC R:ACCATCATTACACAGCAATTCTGAGT	VIC:AAGTTCCCTGTATTTCTT FAM:TCCCTGCATTTCTT
One_U1009-91	445261235	no hits	A/G	F:CTCTGTCCTTGAACTGTTGTCTGTT R:GCCGCTGCTACTCTTCCT	VIC:CATGTTCTGTATGGACCC FAM:TGTTCTGTGTGGACCC
One_U1010-81	445261242	no hits	A/G	F:CAGCCCCTCGAGGTAACTG R:GTTGAGACAACAAAACGTCTACTGT	VIC:CACACCAACGTTATGTAGAG FAM:CACCAACGTTGTGTAGAG
One_U1012-68	445261250	no hits	C/T	F:TCTATTACCATACAGGCCCAGTACA R:CCTTTTGTGTCTTCCAGTCATGTGA	VIC:TGACGGGTGTTCCTGATAA FAM:TGACGGGTGTTCTTGATAA
One_U1013-108	445261258	no hits	G/T	F:TCTGTGCTCTCCTCCAGGAT R:CGAAACTGAGGAGTGCTCTGA	VIC:ACGGAATTCCTGTTGCCCT FAM:ACGGAATTCCTTTTGCCCT
One_U1014-74	445261265	no hits	C/T	F:TCCCCTGCAGCAACTGTTTT R:GGCAGAGACGGCATCCT	VIC:TTGACCTGCGCCAGTAT FAM:TTTTGACCTGCACCAGTAT
One_U1016-115	445261272	no hits	-/T	F:GGATTTTTGACTTGACCGTTTTGTGT R:ATTAACATGTGCAAAGGGAGAATGC	VIC:AATGGCAGTTTTTTATTTGA FAM:ATGGCAGTTTTTTTATTTGA
One_U1017-62	445261279	no hits	A/T	F:CAGAGAAGGACGTACCATTGATACAT R:CCGGTAGATTGGCGTTGCT	VIC:CAGAAAAACTGGTACTTGTT FAM:CAGAAAAACTGGTTCTTGTT
One_U1024-197	445261286	no hits	G/T	F:CTGAACTGATCTACCGCTCTGT R:GGAACAGATACTCCAGGAGAGATGA	VIC:ACCTGACCCAACAAA FAM:ACCTGACACAACAAA
One_U1101	445261293	no hits	C/A	F:CTATGACATGTTTATTTTAATTAGCCACCAACT R:AGTATAGCTAGGGAACCTTTCGATCTT	VIC:TGGACGTATGTCATATTT FAM:TGGACGTATGTAATATTT
One_U1102-220	445261300	no hits	C/T	F:TCCCTCTGCTGGAGAACTACAG R:GGAACAGCAGTCCTGAGTACAG	VIC:CCAGTAGTGTTTTCTG FAM:CAGTAGTGCTTTCTG
One_U1103	445261307	no hits	G/A	F:CCCAGCCGCCATGTGTA R:TGTAGTTCAGCCACCATCTTTGG	VIC:TCGGCGAAAACT FAM:TCGGCAAAAACT
One_U1104-138	445261314	no hits	G/T	F:GGAACAGAACACTGAGAATGAATGC R:GGGAATATGTCGACTGCTCACT	VIC:CCTTCTCAGAGGGTAGAGA FAM:CCTTCTCAGAGGTTAGAGA

Appendix B. Continued.

One_U1105	445261321	no hits	T/A	F:GCCTTAATAGTGTCTTCTGATCCCTTT R:CCCTCTGTTGTCCAGACTCTTAG	VIC:CCTGTTTTTTTTAAAAGAC FAM:TCCTGTTTTTTTTTAAAGAC
One_U1201-492	445261328	no hits	A/G	F:GCTTATGACGGAGAAGAGATGCA R:AGGATACTGAAGCCCAGAGACA	VIC:AAGACTTCCTCCAGGCTC FAM:ACTTCCCCCAGGCTC
One_U1202-1052	445261341	no hits	T/C	F:CGATTTGAGTCTCCAATGGTCTCT R:ATTCCTATGGTTAACATCAATTCTATAAAGTCAT	VIC:CAAACTTTTTCATCTACATTTA FAM:ACTTTTTCATCCACATTTA
One_U1203-175	445261348	no hits	G/A	F:CCCGGAGACATACTTGATGCA R:GGAGGACCTGCAGGATCAC	VIC:CCATAGTTGCTGGGCTT FAM:CTCCATAGTTACTGGGCTT
One_U1204-53	445261354	no hits	C/T	F:GTAAAACCCTTCATGTTGGCCATT R:CTCCATGTCTGAATGTCCCATCA	VIC:ATGCATACACGCTGATGC FAM:ATGCATACACACTGATGC
One_U1205-57	445261361	no hits	A/G	F:AGTAAATGGTTATTCACGTAACGGATAAG R:CAGGACAGTTCCACATTCTAACAGA	VIC:AGTTATCATGGTCATCTCT FAM:AGTTATCATGGTCGTCTCT
One_U1206-108	445261368	no hits	G/T	F:CTGAGATGGTGCTTTCTGAGGATA R:TGGATGAAAGGGAAATTCTGTCAACA	VIC:AACATTGAGCTTCCC FAM:ATAACATTGATCTTCCC
One_U1208-67	445261376	no hits	A/C	F:ACTTGAATGTCTGTTTCGTAGGTGAT R:ACACAGTTGACAGTGGAGCAA	VIC:CCCAATGTGATTGTCAC FAM:CCAATGTGCTTGTCAC
One_U1209-111	445261383	no hits	C/T	F:GTCACGTAATCACGAGAAAGATACTAAATGT R:TCTGCGTCTCCAGAGAGGTT	VIC:CTCACATCGAGATGATC FAM:TCACATCGAAATGATC
One_U1210-173	445261388	no hits	A/G	F:ACAAAGTCTCTCTCTGAGTAGGAGTAC R:CAAAGTATCTCAGAGTGCTGATCTAGGA	VIC:CCCTCCTATTCATTATGATTGT FAM:CCTCCTATTCATTACGATTGT
One_U1211-97	445261395	no hits	C/T	F:GCGTGTCCTCCCATTAGAAGA R:CTGCAGAAGTACAGCATCTATCTGA	VIC:CTGTTTCAGTGTGCTTG FAM:CTGTTTCAGTATGCTTG
One_U1212-106	445261402	no hits	A/G	F:CGTAATGACCTACCACCATATCAGT R:TGGCATGACTTTAACAATTCCCAAAAAA	VIC:TTTTGACATACAAAAAATA FAM:TTTGACATACAGAAAATA
One_U1214-107	445261409	no hits	A/C	F:CCAAATGTACTCCATGTTGGTTAGC R:TGCCTGAGTATTAAGCTATATCATTGAAGTTTT	VIC:TAGTGACCTATTAAATTGC FAM:TGACCTATTCAATTGC
One_U1215-82	445261416	no hits	A/C	F:GTTGCTTGGTTTCGTTTGGAGTAG R:CTCCAGAAGAGGAATACCACAGTTC	VIC:AATGAGACAAAGTATTTGGT FAM:AATGAGACAAAGTCTTTGGT
One_U1216-230	445261423	no hits	A/T	F:TGGGATCGGACGTCAATAGATTTC R:GTAATACAGAGTGAGCGTGATACATTGT	VIC:CCTGGCTACTAAGTAAC FAM:CTGGCTACAAAGTAAC