Selection and Adaptation of Fitness-related traits in Coho Salmon

Miyako Kodama

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Kerry Ann Naish, Chair

Jeffrey Hard

David Beauchamp

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

University of Washington

University of Washington

Abstract

Selection and Adaptation of Fitness-related Traits in Coho Salmon

Miyako Kodama

Chair of the Supervisory Committee:
Associate Professor Kerry Ann Naish
School of Aquatic and Fishery Sciences

Selection is the primary force driving phenotypic differentiation and adaptive evolution, and understanding how and to what extent selection drives adaptive evolution remains a key challenge in evolutionary biology. A number of commercially important species have been long exposed to deliberate, continuous selection focusing on the development or enhancement of desired traits, and the phenotypic characteristics of these species have been widely altered. Domesticated populations are excellent models to study the mechanism of adaptive evolution. In particular, such populations could be utilized to study the direct and indirect results of concerted selection on specific phenotypes, and to demonstrate how adaptive phenotypic differentiation following certain selective pressures are determined at specific sites in the genome. Such knowledge is useful as it provides insights into whether and how populations will evolve in response to natural and human-induced selection and how certain actions might mitigate or facilitate such changes.

Here, I examined how selection acts on wild and domesticated coho salmon (*Oncorhynchus kisutch*) and investigated the genetic basis of adaptive evolution following domestication selection. This dissertation is two-fold and consists of four chapters. The aim of the first part of this dissertation (Chapter 1) was to investigate the temporal variation in selection by comparing the mode, direction and strength of selection on fitness related traits between two cohorts of wild coho salmon; specifically, I examined selection on date of return and body length using a constructed pedigree of the wild population at Big Beef Creek in Washington State, in order to illustrate how selection operates in wild and drives adaptive evolution over time. The aim of the second part of the dissertation (Chapter 2, 3, 4) was to construct a dense linkage map for this species, and to identify genomic regions that have responded to domestication selection in selectively bred salmon using this map. In Chapter 2, I investigated the relationship between chromosome arrangements and the retention of recently diverged or undifferentiated duplicated regions by deriving a linkage map for coho salmon and comparing this map with those of Chinook salmon, rainbow trout, and Atlantic salmon. Using this linkage map, I investigated the genetic architecture of growth-related traits, and examined the genetic basis for a tradeoff between age at sexual maturity and growth during crucial decision periods in coho salmon (Chapter 3). In Chapter 4, I determined whether there was evidence for selection at genomic regions linked with selectively bred traits in a domesticated population, and identified loci that played a role in driving adaptive phenotypic evolution in response to selective breeding. The work depicted herein demonstrated how phenotypic variation for important fitness-related traits might be maintained in wild, and provided insights into how phenotypic differentiation resulting from certain selective pressures was determined at specific sites in the genome.

Table of Contents

List of Figures

List of Tables

Acknowledgements

There are many people who have helped me complete this dissertation. First of all, I would like to express my deepest gratitude to my advisor, Kerry Naish. Her excellent mentorship and patience were invaluable for my professional growth as a scientist. She has been a role model as a successful female researcher, an educator and also a person, and she has been inspirational and motivational throughout my graduate career. I would also like to thank Jeff Hard, Dave Beauchamp and Katie Peichel for serving as my committee members; they have been always encouraging, supportive and generous with their time and knowledge.

I would also like to express appreciation to my colleagues at School of Aquatic and Fisheries Sciences at the University of Washington. Special thanks go to my colleagues from Merlab for their good advice, collaboration and wonderful friendships: namely, Carl Ostberg, Charlie Waters, Daniel Drinan, Daniel Peterson, Eleni Petrou, Isadora Jimenez-Hidalgo, Jocelyn Lin, Kristen Gruenthal, Maureen Hess, Marine Brieuc, Melissa Baird, Shannon O'Brien, Teresa Sjostrom and Todd Seamons. I would also like to thank Daisuke Sasatani, Eric Ward, Hiroki Oura, Joe Anderson and Kotaro Ono for their invaluable comments and suggestions while completing this dissertation.

I am grateful for NOAA Fisheries, School of Aquatic and Fishery Sciences and Graduate Opportunities and Minority Achievement Program (GO-MAP, University of Washington) for their financial support. I am also appreciative of the staff members at the School of Aquatic and Fisheries Sciences at the University of Washington; I would particularly like to thank Scott Schafer, Robin Weigel, Addi Daisley, Samantha Scherer and Amy Fox for the help they provided.

Finally, I would like to express my appreciation to my friends and family for their emotional support and encouragement. Special thanks go to my good friends in Seattle for making my life here enjoyable and memorable with numerous dinners and get-togethers. My parents, Kazuyoshi and Kayoko Kodama, as well as my sister, Erika Nakashima have always been there for me when I needed their support, and I am beyond thankful for their unconditional love and encouragement. I would also like to thank my parents-in-law, Karen Christensen and Claus Henriksen to travel back and forth from Denmark, helping me dedicate more time to completing this dissertation. Lastly, I would like to express my appreciation to my beloved husband, Simon Sigurd Henriksen and our beautiful son, Noah Henriksen for their love and support throughout my academic endeavors, especially at the final stage of completing my Ph.D. Thank you all!

General Introduction

Selection is the primary force driving phenotypic differentiation and adaptive evolution (Simpson 1953; Kingsolver *et al.* 2001). If selection consistently favors certain traits, such traits may become widespread within a population, resulting in adaptive phenotypic differentiation. At the molecular level, selection occurs when a certain allele become widespread because of its effect on the individuals. Understanding how and to what extent selection drives adaptive evolution remains a key challenge in evolutionary biology. Specifically, understanding how selection operates on phenotypic traits in wild and identifying the genomic regions directly targeted by selection are imperative to fully understand the genetic basis of adaptive evolution (Fisher 1930; Lande and Arnold 1983; Kimura 1984; Orr 1998). Such knowledge is particularly of concern in conservation and management, as well as in commercial breeding (Crandall and Bininda-Emonds 2000; Moritz 2002; Goddard and Hayes 2009; Allendorf *et al.* 2010). To enable more effective conservation and management of wild populations, such knowledge would be useful because it provides insights into whether and how populations will evolve in response to natural and human-induced changes (Crandall and Bininda-Emonds 2000; Moritz 2002; Allendorf *et al.* 2010) and how conservation actions might mitigate this change. For animal and plant breeders, it is of commercial interest to learn how artificial selection promotes phenotypic differentiation because such knowledge helps achieve enhancement of desired traits in domesticated populations (Goddard and Hayes 2009).

The study of adaptive evolution in salmonid species is of considerable interest in conservation and management due to their high commercial and cultural values. Many populations of these species have been either extirpated or declined due to anthropogenic influences (National Research Council 1995; Gustafson *et al.* 2007). Salmonid fishes display a substantial degree of variation in life history traits, and are adapted to a wide diversity of habitats (Taylor 1991; Waples 1994; Fraser *et al.* 2011). Many of these traits (e.g. growth rate, age at sexual maturity, return timing to the natal stream, resulting size at spawning) have been shown to be under selection (van den Berghe and Gross 1989; Seamons *et al.* 2004, 2007; Serbezov *et al.* 2010; Anderson *et al.* 2010). Such traits tend to evolve quickly under consistent directional selection (Crozier *et al.* 2008), thus making salmonids an ideal study subject to investigate how selection promotes adaptive evolution.

Within the salmonids, coho salmon (*Oncorhynchus kisutch*) provides a simple study subject to study selection and adaptive evolution. This species is known to have the capacity to evolve quickly (Crozier *et al.* 2008), and a highly domesticated population selectively bred for large size and early maturation is available for this study. Access to such resources provides an exciting opportunity to learn how deliberate selection promotes adaptive evolution. Specifically, the domesticated population could be compared to its wild counterpart, in order to study the mechanism and the extent to which selection causes divergence between domesticated and source populations. Such examinations may facilitate discovery of the loci that are involved in driving adaptive evolution, and provide insights into the rate and nature of adaptive response to deliberate, continuous selection targeting desired traits.

The overall aims of this study were to examine how selection acts on wild and domesticated coho salmon and to investigate the genetic basis of adaptive evolution. First, the temporal variation in selection was investigated by comparing the mode, direction and strength

of selection on fitness related traits between two cohorts of wild coho salmon (Chapter 1). Such comparisons provided insights into how selection operates in wild and may drive adaptive evolution over time. Second, the molecular basis of various growth-related traits was determined by constructing dense, sex-specific linkage maps (Chapter 2) and identifying the genomic regions that were linked with those traits in coho salmon (Chapter 3). Third, combined with information from Chapter 2 and 3, loci that were under the direct influence of selection during domestication were identified through genome scans using the domesticated, hatchery and wild source populations of this species (Chapter 4). Together, the work depicted herein provided insights into how phenotypic variation for important fitness-related traits is maintained in wild, revealed the direct and indirect results of concerted selection on specific phenotypes, and demonstrated how adaptive phenotypic responses are determined at specific sites in the genome.

References

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart, 2010 Genomics and the future of conservation genetics. Nat. Rev. Genet. 11: 697-709.

Anderson, J. H., P. L. Faulds, W. I. Atlas, G. R. Pess, and T. P. Quinn, 2010 Selection on breeding date and body size in colonizing coho salmon, *Oncorhynchus kisutch*. Mol. Ecol. 19: 2562-2573.

Van den Berghe, E. P., and M. R. Gross, 1989 Natural selection resulting from female breeding competition in a Pacific salmon (coho: *Oncorhynchus kisutch*). Evolution. 43: 125-140.

Crandall, K. A., and O. R. P. Bininda-Emonds, 2000 Considering evolutionary processes in conservation biology. Trends Ecol. Evol. 15: 290-295.

Crozier, L. G., A. P. Hendry, P. W. Lawson, T. P. Quinn, N. J. Mantua et al., 2008 Potential responses to climate change in organisms with complex life histories: evolution and plasticity in Pacific salmon. Evol. Appl. 1: 252-270.

Fisher, R. A., 1930 *The genetical theory of natural selection*. Clarendon Press, Oxford.

Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen, and E. B. Taylor, 2011 Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity. 106: 404-420.

Goddard, M. E., and B. J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381-391.

Gustafson, R. G., R. S. Waples, J. M. Myers, L. A. Weitkamp, G. J. Bryant et al., 2007 Pacific salmon extinctions: quantifying lost and remaining diversity. Conserv. Biol. 21: 1009-1020.

Kimura, M., 1984 *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri et al., 2001 The strength of phenotypic selection in natural populations. Am. Nat. 157: 245-261.

Lande, R., and S. J. Arnold, 1983 The measurement of selection on correlated characters. Evolution. 37: 1210-1226.

Moritz, C., 2002 Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst. Biol. 51: 238-254.

National Research Council, 1995 Science and the endangered species act. National Academies Press.

Orr, H. A., 1998 Using quantitative trait locus data. Genetics. 149: 2099-2104.

Seamons, T. R., P. Bentzen, and T. P. Quinn, 2007 DNA parentage analysis reveals inter-annual variation in selection: results from 19 consecutive brood years in steelhead trout. Evol. Ecol. Res. 9: 409-431.

Seamons, T. R., P. Bentzen, and T. P. Quinn, 2004 The effects of adult length and arrival date on individual reproductive success in wild steelhead trout (*Oncorhynchus mykiss*). Can. J. Fish. Aquat. Sci. 61: 193-204.

Serbezov, D., L. Bernatchez, E. M. Olsen, and L. A. Vøllestad, 2010 Mating patterns and determinants of individual reproductive success in brown trout (*Salmo trutta*) revealed by parentage analysis of an entire stream living population. Mol. Ecol. 19: 3193-3205.

Simpson, G. G., 1953 The Baldwin effect. Evolution. 7: 110-117.

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

Waples, R. S., 1994 Genetic considerations in recovery efforts for Pacific salmon. Conserv. Biol. 8: 884-886.

Chapter 1: Temporal variation in selection on body length and date of return in a wild population of coho salmon, *Oncorhynchus kisutch**

Miyako Kodama[1], Jeffrey J Hard[2] and Kerry A Naish[1]

[1]School of Aquatic and Fishery Sciences, University of Washington

[2]National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, Washington

*This chapter has been published as Temporal variation in selection on body length and date of return in a wild population of coho salmon, *Oncorhynchus kisutch*, Kodama M, Hard JJ and Naish KA, BMC Evolutionary Biology 12 (1), 116,

Abstract

A number of studies have measured selection in nature to understand how populations adapt to their environment; however, the temporal dynamics of selection is rarely investigated. The aim of this study was to assess the temporal variation in selection by comparing the mode, direction and strength of selection on fitness related traits between two cohorts of coho salmon (*Oncorhynchus kisutch*). Specifically, we quantified individual reproductive success and examined selection on date of return and body length in a wild population at Big Beef Creek, Washington (USA).

Reproductive success and the mode, direction and strength of selection on date of return and body length differed between two cohorts sampled in 2006 and 2007. Adults of the first brood year had greater success over those of the second. In 2006, disruptive selection favored early and late returning individuals in 2-year-old males, and earlier returning 3-year-old males had higher fitness. No evidence of selection on date of return was detected in females. In 2007, selection on date of return was not observed in males of either age class, but stabilizing selection on date of return was observed in females. No selection on body length was detected in males of both age classes in 2006, and large size was associated with higher fitness in females. In 2007, selection favored larger size in 3-year-old males and intermediate size in females. Correlational selection between date of return and body length was observed only in 2-year-old males in 2006.

We found evidence of selection on body length and date of return to the spawning ground, both of which are important fitness-related traits in salmonid species, but this selection varied over time. Fluctuation in the mode, direction and strength of selection between two cohorts are likely to be due to factors such as changes in precipitation, occurrence of catastrophic events (flooding), the proportion of younger- versus older-maturing males, and sex ratio and densities of spawners.

Introduction

A number of studies have measured selection in nature in an effort to understand how populations adapt to their environment over time (Siepielski *et al.* 2009; Kingsolver and Diamond 2011). Reviews collating the estimates of selection in natural populations have debated their magnitude, mode and temporal stability (Conner 2001; Hoekstra *et al.* 2001; Kingsolver *et*

al. 2001; Kingsolver and Pfennig 2007; Siepielski *et al.* 2009; Kingsolver and Diamond 2011). However, generalizations have been limited because many of the studies lacked temporal replication, comprised small sample sizes that reduced their statistical power to detect selection, or measured fitness components instead of total fitness (Kingsolver *et al.* 2001; Kingsolver and Pfennig 2007). Nonetheless, two patterns that do emerge are that phenotypic selection is often strong enough to cause evolutionary changes in relatively few generations, and that directional selection frequently prevails over stabilizing or disruptive selection (Hoekstra *et al.* 2001; Kingsolver and Pfennig 2007). Several factors may interact to reduce response to directional selection (Kingsolver and Diamond 2011), including the fact that selection can vary in strength, direction and mode over time (Siepielski *et al.* 2009; Bell 2010).

Studies indicate that temporal changes in selection are attributable to variation in environmental and ecological factors, such as changes in climate, sex ratio or density (Siepielski *et al.* 2009, 2010; Bell 2010). Long-term investigation into patterns of selection provides insight into how phenotypic variation in fitness-related traits is maintained, and how populations adapt to variable environmental conditions (Siepielski *et al.* 2009). Such information also allows predictions on whether and how populations evolve in response to human perturbations, thus facilitating effective conservation and management of exploited species (Hendry *et al.* 2003; Stockwell *et al.* 2003; Carlson and Seamons 2008; Waples and Hendry 2008).

Salmonid fishes provide an ideal study system to study temporal changes in selection. These species are philopatric; in principle, it is possible to sample an entire population at maturity, and the assessment of lifetime reproductive success can be used to accurately measure selection (Sandercock 1991; Quinn 2005). Also, several life history traits have been shown to be subject to selection in these species (van den Berghe and Gross 1989; Seamons *et al.* 2004, 2007; Ford *et al.* 2008; Serbezov *et al.* 2010; Anderson *et al.* 2010). For example, several studies have demonstrated that return timing in salmon is often linked to fitness and responds to selection (Seamons *et al.* 2004, 2007; Ford *et al.* 2008; Anderson *et al.* 2010). Variability in this trait is typically affected by environmental conditions such as rainfall, temperature or flow regime of their natal river, all of which tend to fluctuate annually (Quinn 2005). Despite such fluctuations, however, the earlier return of males to the spawning grounds ("protandry") is commonly reported in salmonids (Morbey 2000). This behavior may maximize mating opportunities in males, as females may not be reproductively active later in the season (Quinn 2005). However, protandrous arrival may not always be beneficial, as its fitness advantages depend on factors such as female availability, the number of competitors present or post-arrival mortality (Morbey and Abrams 2004).

Body size in salmonids is also an example of a trait that is acted upon by selection and might be linked to fitness (Fleming 1998; Garant *et al.* 2001; Seamons *et al.* 2007; Theriault *et al.* 2007; Ford *et al.* 2008; Anderson *et al.* 2010). Male salmon exhibit high variability in size at maturity, and such variation results in several alternative tactics to achieve fertilization (Gross 1985, 1991; Holtby and Healey 1990; Foote *et al.* 1997; Fleming 1998; Quinn 2005; Koseki and Fleming 2007). Age at maturity is also linked to size variation; smaller, younger-maturing males may adopt sneaking to gain access to spawning females (Gross 1985; Holtby and Healey 1990;

Fleming 1998). On the other hand, large, late-maturating males may engage in fighting and outcompete smaller males to gain access to spawning females (Gross 1985; Fleming 1998). In female salmon, variability in size and age tends to be smaller than in males, and studies suggest that large body size may increase their reproductive success (Fleming 1998; Beacham and Murray 1993; Fleming 1996; Andersson 1994; Steen and Quinn 1999; Dickerson, Quinn, and Willson 2002; van den Berghe and Gross 1984; Olsen and Vollestad 2003; van den Berghe and Gross 1986; Morita and Fukuwaka 2006; Foote 1990; Kitano 1996).

Within the salmonids, coho salmon (*Oncorhynchus kisutch*) provides a simple study subject for studies on the effect of selection because it is strictly anadromous, and has shorter generation time and simpler age structure compared to most other salmonids; this species typically returns from the ocean at three years of age, while some younger-maturing males, "jacks," return at two years (Quinn 2005). Like most *Oncorhynchus*, this species is semelparous (Quinn 2005); thus, there are no repeat spawners that may hinder precise separation of parental and offspring generations, which could further complicate the assessment of the magnitude and mode of selection based on estimates of individual reproductive success.

The aim of this study was to investigate the temporal variation in selection by comparing the mode, direction and strength of selection on fitness related traits between two cohorts of wild coho salmon. Here, we followed the convention of Siepielski *et al.* (2009) and defined the dynamics of temporal variation in selection as "the interannual differences in selection on a given trait within a population." Specifically, we examined selection on date of return and body length in the wild population at Big Beef Creek, Washington. Information obtained was used to illustrate how selection operates in nature and provide insights into the temporal dynamics of selection. Pedigree reconstruction based on 11 highly polymorphic DNA microsatellite loci was conducted using more than 3000 individuals, and lifetime reproductive success of two parental brood years was quantified. Younger- and older-maturing males of this species exhibit alternative tactics to maximize reproductive success (sneaking versus fighting). Therefore, the mode, direction and strength of selection acting on males of two age classes and females of a single age class were estimated separately. Estimated selection in this study are the results of both natural and sexual selection, as selection estimates were obtained from regression analyses with lifetime reproductive success as a fitness measure.

Materials and Methods

Study Area and sampling

This study was conducted at Big Beef Creek (47°39'N, 122°46'W) in Washington State, USA, situated on Hood Canal in Puget Sound. Big Beef Creek is routinely monitored as an indicator stream for long-term ecological studies in the region, and the amount of water discharge (cubic feet per second) from the creek is measured on a daily basis (USGS).

The creek supports a healthy native run of anadromous coho salmon, which is indigenous to the system and has been monitored by the Washington Department of Fish and Wildlife (WDFW) over the past 30 years. As hatchery practices have never been conducted at this creek, we expect no or little effects on the current genetic structure of the population. A weir is placed

at the mouth of the creek, and only naturally spawned fish (all hatchery fish are externally marked) are passed over the weir for subsequent spawning. Therefore, intrusion of hatchery fish into the wild population at this creek has been prevented as far as possible.

Returning adults of coho salmon were sampled every year from 2006 to 2010. Records indicate that peak river entry of coho salmon to this creek occurs from late-October to mid-November (Weitkamp *et al.* 1995). The brood class of this species returns on average at 3 years, but early maturing males return at 2 years; therefore, sampling provided 2 full cohorts, for parental brood years 2006 and 2007 and their adult offspring returning in 2008 to 2010.

As each individual passed the weir, date of return, body length (length from the tip of snout to fork of tail) and sex were recorded, and a fin clip was obtained for DNA analysis. In coho salmon, returning males smaller than 35cm in body length are typically 2 years old, and returning males larger than 45cm are 3 years old (Kinsel *et al.* 2009). Scales from all individuals in the range 35cm and 45cm were collected and read by WDFW to verify exact age. Once measurements were taken, individuals were allowed to swim upstream for spawning. Data from Big Beef Creek coho suggest that retuning coho spawn for up to three weeks after river entry (Kinsel *et al.* 2009).

This study was reviewed and approved by the University of Washington's Institutional Animal Care and Use Committee (IACUC).

Microsatellite analysis and parentage

Genomic DNA from sampled individuals was extracted using the DNeasy extraction kit (QIAGEN, Valencia, CA, USA) following the manufacturer's procedures. Extracted DNA was used to amplify 11 microsatellite loci (Supplementary Material S1.2) via multiplex Polymerase Chain Reaction (PCR) using a QIAGEN Multiplex PCR kit. Specifically, reaction mixtures consisted of 10-200ng genomic DNA, 1x QIAGEN Multiplex PCR Master Mix, 0.03 µM - 0.4µM of each primer, making up a total volume of 10µl (Supplementary Material S1.2). Cycling conditions consisted of a 15-min, initial activation step at 95ºC, 30 cycles of 30-s denaturing step at 94ºC, 90-s annealing step at 57 - 60 ºC and 90-s extension step at 72 ºC, and a 30-min, final extension step at 60ºC (Supplementary Material S1.2). The forward primer of every locus was labeled with fluorescent dye. Individuals were genotyped using a 96-capillary system Molecular Dynamics MegaBACE 1000 automatic genotyper (GE Healthcare, Piscataway, NJ, USA), and Genetic Profiler version 2.2 was used to determine fragment sizes of all loci. To calculate our genotyping error rate, 96 individuals from samples obtained in 2006 were randomly chosen, re-extracted and genotyped. The genotyping error rate was estimated by calculating the percentage of allele calls that were different between two analyses.

MICROCHECKER v. 2.2.3 was used to estimate the frequency of null alleles, as well as to screen for large allele dropout and accidental scoring of stuttering (Van Oosterhout *et al.* 2004). GenePop v. 4.0.10 was used to perform exact tests for deviations from Hardy-Weinberg equilibrium (Raymond and Rousset 1995). GenAlEx v.6.41 was used to calculate observed and expected heterozygosities, as well as Weir and Cockerham's F_{IS} values (Peakall and Smouse 2006). FRANz v. 1.9.999 was used to calculate exclusion probabilities for two-parent and single-

parent assignments based on the 11 loci used for the analyses, as well as to perform pedigree reconstruction (Riester *et al.* 2009). This program uses a log-ratio, the parent-pair log-odds ratio (LOD) score to assign parentage, and it estimates statistical confidence for each assignment using Markov Chain Monte Carlo (MCMC) sampling. All adults returned in 2006 and 2007 were considered as candidate parents for individuals returned in 2008 to 2010. Only individuals that were genotyped at more than 6 loci were included in the analysis. Parentage assignments that had a posterior probability higher than 0.99 were used for further analysis.

We performed two tests to assess our assignment error rate. First, FRANz v. 1.9.999 was used to perform pedigree reconstruction with unlikely parent-offspring pairings: 1) individuals sampled in 2006 as candidate parents and individuals sampled in 2007 as candidate offspring, and 2) individuals sampled in 2007 as candidate parents and individuals sampled in 2006 as candidate offspring. Paternity and maternity assignments that had more than 99% posterior probability from these trials were used to calculate our parentage error rate. Second, COLONY v. 2.0 was used to perform pedigree reconstruction for the first parental brood year (2006) and their candidate offspring as a comparison (Wang 2004; Jones and Wang 2009; Wang and Santure 2009). This program implements a maximum likelihood method to assign parentage among individuals using multilocus genotype data, and provides the approximated probability for each assignment. Both sexes were allowed to be polygamous. Only parentage assignments that had probability of more than 0.99 were employed and compared with assignments obtained from FRANz v. 1.9.999 that had probability of more than 0.99. It was not possible to perform a similar comparison on the whole dataset, because COLONY was computationally intensive.

Selection analysis

Reproductive success (RS), defined as the number of returning adult offspring produced by each parent, was calculated using the estimated parentage assignments, and was used as a measure of fitness for selection analysis. The Mann-Whitney U-test was used (Zar 1999) to test differences in reproductive success of 2-year-old males, 3-year-old males and females between the two parental brood years. Body length and date of return were standardized within each sex and cohort to a mean of zero and to a standard deviation of one, denoted as zl (body length) and zd (date of return). Relative fitness, w, was calculated for each individual by dividing its reproductive success by the within-sex, within-age (in males) and within-brood year mean (Lande and Arnold 1983; Brodie *et al.* 1995). Selection analyses were performed using the zero-inflated regression model implemented in the pscl package in R (Zeileis *et al.* 2008), with negative binomial error distribution and the canonical log link. Zero-inflated models describe the data better than the simple generalized linear model, as relative fitness exhibited overdispersion and a large number of true zero values (parents with no returning offspring). In order to impose relative fitness as a response variable, all zero-inflated models in our analysis included an offset term of the logarithm within-sex, within-age and within-brood year mean reproductive success. Regression was performed separately on 2-year-old males, 3-year-old males and females in each brood year. The initial model for all groups included five coefficients ($z_l, z_d, z_l^2, z_d^2, z_l z_d$) for each of the explanatory variables:

$$w = \alpha + z_l + z_d + z_l^2 + z_d^2 + z_l z_d$$

w is relative fitness, α is the y-intercept of the fitness function, and zl and zd are standardized body length and date of return for each individual, respectively. Originally, stream discharge on each individual's return date was also included in the initial model, however it was removed because discharge itself did not explain variation in relative fitness in all groups of both cohorts. Final model selection using the Akaike information criterion (AIC) was based on a stepwise method recommended by Zuur *et al.* (2009). Final models were validated by fitting the residuals against each explanatory variable.

Univariate cubic splines were calculated following Schluter (1988), in order to visualize the relationship between reproductive success and measured traits in both males and females. Specifically, these analyses were performed using the generalized additive model, GAM function, implemented in R with negative binomial error distribution and an additional overdispersion parameter, θ (Hilborn and Mangel 1997). θ was calculated for 2-year-old males, 3-year-old males and females within each parental brood year according to the equation:

$$\theta = \frac{\text{mean (RS)}^2}{\text{var (RS)} - \text{mean (RS)}}$$

The smoothing parameter, λ, was estimated for each curve using the generalized cross-validation (GCV) criterion implemented in the GAM function, available from the mgcv library in R (Wood 2001).

Environmental influence on return date

It is known that upstream migration is influenced by precipitation in coho salmon (Quinn 2005). Spearman's rank correlation test was performed to investigate the relationship between the number of arrived spawners counted on a daily basis and the amount of daily water discharge in each parental brood year, as neither variable was normally distributed (Zar 1999).

Results

Characteristics of sampled fish

A total of 3512 returning adults were sampled from 2006 to 2010. On December 3rd in 2007, an atypically large flood breached the weir, possibly allowing some late returning adults to enter the stream unsampled. A total of 1678 individuals were sampled in 2006 and 2007; these were considered candidate parents. In 2008 to 2010, 1834 individuals were sampled; 1240 individuals were likely to be the offspring of the individuals collected in 2006 and 2007, as 3-year-old individuals sampled in 2008 (n = 423) and 2-year-old individuals sampled in 2010 (n = 171) were candidate offspring of individuals returning in 2005 and 2008, respectively. However, all individuals collected were genotyped and included in the analyses.

The number of fish returning in 2007 (n = 1177) was greater than that in 2006 (n = 501; Table 1.1). The sex ratio (male: female) was greater in 2006 (1.2) than in 2007 (0.8; Table 1.1). There were consistently more males present in 2006, whereas the male to female sex ratio remained low for the majority of the season in 2007 (Supplementary Material S1.1). In 2006, the proportion of 2- to 3-year-old males was 0.6:1, but this ratio decreased to 0.06:1 in 2007 (Table 1.1). The date of return differed between the two years; individuals arrived from November 2nd

to December 15th in 2006, but arrived earlier from October 1st to November 20th in 2007. The mean date of return by sex ranged from November 6th (males) to 9th (females) in 2006, and from October 23rd (males) to 26th (females) in 2007 (Table 1.1). Individuals tended to be smaller in 2007 (Table 1.1).

Population genetic statistics

Genotyping error rate was small, with the error rate per locus ranging from 0 to 1.0% (Supplementary Material S1.2). Across years, no consistent presence of null alleles was detected in all loci, and no large allele dropout or accidental scoring of stuttering were detected. 99.1% of the collected samples (3481 individuals) were successfully genotyped at more than 10 loci, and 99.8% of the collected samples (3507 individuals) were successfully genotyped at more than 6 loci. All loci were moderately to highly polymorphic, with the number of alleles ranging from 9 to 45 and observed heterozygosity ranging from 0.73 to 0.96 (Supplementary Material S1.3). Among 55 tests (11 loci in each year from 2006 - 2010), significant deviation from Hardy-Weinberg equilibrium was observed in 25 tests (Supplementary Material S1.3). Such results may be due to the large sample sizes in 2007 and 2009; however, observed and expected heterozygosity was similar across all loci in all years (Supplementary Material S1.3). F_{IS} values tended to be small, ranging from -0.03 to 0.07 across all loci in all years (Supplementary Material S1.3).

Parentage analysis

The exclusion probabilities for two-parent and single-parent assignments were > 0.99999 according to FRANz, indicating that the microsatellite dataset provided sufficient power to perform parentage analyses. Results from the tests on the error in our parentage assignment suggest that our error in assigning parents was between 1.7% and 3.6%. Specifically, when pedigree reconstruction was performed with individuals sampled in 2006 as candidate parents and individuals sampled in 2007 as candidate offspring, 3.6% of the assignments (84 out of 2354 assignments) calculated by FRANz had greater than 99% posterior probability. When pedigree reconstruction was performed with individuals sampled in 2007 as candidate parents and individuals sampled in 2006 as candidate offspring, 3.3% of the assignments (33 out of 1002 assignments) calculated by FRANz had more than 99% posterior probability. Assignments with more than 99% probability obtained by FRANz were compared with assignments obtained by COLONY for the 2006 brood year, and 1.7% of mismatches in assignments were observed (25 out of 1490 assignments).

Among 980 returning individuals that could be the offspring of adults returning in 2006, 470 individuals (48.0%) were assigned to both parents, 47 individuals were assigned to a father only (4.8%), 33 individuals were assigned to a mother only (3.4%), and 430 individuals (43.9%) were not assigned to any parents (Table 1.2). Among 251 returning individuals that could be the offspring of adults returning in 2007, 46 individuals (18.3%) were assigned to both parents, 14 individuals were assigned to a father only (5.6%), 5 individuals were assigned to a mother only (2.0%), and 186 individuals (74.1%) were not assigned to any parents (Table 1.2). Among a total of 516 individuals that had both parents assigned, 7 individuals were assigned to parents that

were sampled across different years (in 2006 or 2007). Given that such matings were impossible, further analyses were performed with and without these assignments. Because the results and their significance did not differ, these assignments were excluded from further analyses.

The majority of the population produced no or few offspring in both brood years, while some individuals produced a large number of offspring (Figure 1.1). In 2006, 71.2% of 2-year-old males, 42.8% of 3-year-old males and 47.6% of females produced no returning adult offspring. In 2007, 86.7% of 2-year-old males, 92.3% of 3-year-old males and 92.2% of females produced no returning adult offspring. Average reproductive success in all three groups was higher for the 2006 brood year than the 2007 brood year (Table 1.1). However, a significant difference in reproductive success between 2006 and 2007 was only detected in 3-year-old males and females (3-year-old males; Mann-Whitney *U*-test, $P < 0.001$; females; Mann-Whitney *U*-test, $P < 0.001$), and no such difference was detected in 2-year-old males (Mann-Whitney *U*-test, $P = 0.052$).

Selection analysis (Date of Return)

In the 2006 brood year, a positive quadratic selection gradient ($P < 0.001$) on return timing was observed in 2-year-old males, indicating that disruptive selection favored both early and late returning individuals (Table 1.3); the univariate cubic spline also supported this finding (Figure 1.2a). A linear gradient in 3-year-old males was negative ($P < 0.01$), indicating that earlier returning individuals had greater reproductive success (Table 1.3). This trend is also illustrated by the univariate cubic splines (Figure 1.2a). Neither the linear nor quadratic gradients were significant in females (Table 1.3), although the cubic spline showed that early returning individuals tended to produce more offspring (Figure 1.2a).

In the 2007 brood year, there was less evidence of selection on return timing. No statistical support for selection on date of return was observed in males of either age class (Table 1.3). No strong relationship was detected via univariate cubic splines in 3-year-old males. However, the cubic spline implied that both early and late returning individuals tended to produce more offspring for 2-year-old males (Figure 1.2b). The pattern for 2-year-old males may reflect the lack of power to detect selection in this dataset, as 2-year-old males were rare in 2007 (104 individuals in 2006, 30 individuals in 2007). In females, a significant negative quadratic gradient ($P < 0.05$) was observed, indicating that females returning in the middle of the season had higher fitness than early or late returning females (Table 1.3). However, no strong relationship was detected via univariate cubic splines in females (Figure 1.2b).

Selection analysis (Body Length)

In the 2006 brood year, no significant support for selection was detected for body length in males of either age class (Table 1.3). However, the univariate cubic spline showed that large individuals tended to produce more offspring in both groups (Figure 1.2c). A significant negative bivariate selection gradient on date of return and length ($P < 0.001$) for 2-year-old males (Table 1.3) indicated that higher reproductive success of younger- or late-returning males depended on their size. No significant gradient was detected for older males or for females. For females, the linear gradient on length was positive ($P < 0.05$; Table 1.3), indicating that directional selection

favored larger individuals. Graphical representation of the relationship between body length and reproductive success also supported the trend that larger females tended to produce more offspring (Figure 1.2c).

In the 2007 brood year, no significant selection gradient on length was detected in 2-year-old males (Table 1.3), and no trend was observed via univariate cubic splines (Figure 1.2d). In 3-year-old males, selection favored large size, as the linear gradient was positive ($P < 0.001$; Table 1.3); the cubic splines also showed that larger individuals tended to produce more offspring in this group (Figure 1.2d). The quadratic gradient in females was negative ($P < 0.01$), indicating that intermediate-sized females had higher fitness than bigger-sized or smaller-sized females (Table 1.3). The univariate cubic spline also implied that intermediate-sized individuals tended to produce more offspring (Figure 1.2d).

Environmental influence on return date

In 2006, the majority of the population returned at the beginning of the run (Figure 1.3c). In contrast, fish in 2007 returned episodically over a series of peaks throughout the run (Figure 1.3d). A significant correlation was found between count of daily arrivals and the amount of daily water discharge for both males and females in 2007 (males, $\rho = 0.31$, $P < 0.05$; females, $\rho = 0.30$, $P < 0.05$). No such correlation was detected in in 2006.

Discussion

The aim of our study was to determine the temporal variation in the mode, magnitude and direction of selection in a natural coho salmon population using measures of individual fitness. Our results have shown that reproductive success, as well as selection on body size and date of return to the spawning ground differed markedly between two parental cohorts. Adults in the 2006 brood year had higher reproductive success than those in 2007, and direction and strength of selection differed for both traits between two cohorts. In the first parental cohort, there was significant selection on date of return for 2- and 3-year-old males (disruptive selection on return timing and directional selection favoring earlier return, respectively), and on body length for females (directional selection favoring larger size). In the second parental cohort, there was significant directional selection favoring larger size in 3-year-old males and stabilizing selection on both date of return and size in females. The only evidence for bivariate selection on timing and size was in 2-year-old males in 2006. Observed fluctuations in selection may be due to factors such as changes in precipitation, proportion of 2-year-old versus 3-year-old males, sex ratio and spawner density, as well as the occurrence of catastrophic events (flooding). Although stream discharge itself did not explain variation in relative fitness in all groups, results suggest that precipitation affects return timing of Big Beef Creek coho in some years, potentially explaining the variation in the intensity and the direction of selection on this trait.

Pedigree quality and incorrect assignments can affect the estimated reproductive success and selection gradients (Araki and Blouin 2005). Calculated exclusion probabilities showed that our microsatellite dataset was sufficient for identifying parents, and the addition of more markers is less likely to improve our estimates. Low error rates in our parentage assignment affirmed that microsatellites used in our analyses provided enough statistical power to exclude non-candidate

parents. Nevertheless, our overall assignment rate of offspring to one parent or both parents was approximately 50% in the first cohort and 25% in the second cohort. Such results may have occurred due to genotyping errors, to unsampled parents in the candidate parent generations, or to immigration from a non-natal stream during the candidate offspring generations. It is possible that a large number of parents were not sampled at the weir, especially in 2007 when there was a late flood, and this could have led to reduction in assignment success. However, traps were maintained well past the return season and flooding occurred after this period. Additionally, if the number of unsampled parents returned randomly during the season, we would expect to see a larger number of assignments to single parents than to both parents, especially given the number of individuals with unassigned parents. This outcome was not the case. Assignment to single parents varied between two to six percent, and these figures are comparable between both cohorts. While we cannot rule out inefficiency in trapping, especially in 2007, the most plausible explanation for unassigned fish in the offspring generations is that they were immigrants. Parentage estimates in similar systems (Seamons *et al.* 2004, 2007; Theriault *et al.* 2007) acknowledged that not all parents were sampled. In these studies, there was a greater proportion of single parent assignments (32 to 42%) than we observed in this study. All information combined indicates that our sampling was not biased, and the results obtained here were sufficient to gain insights into the factors associated with reproductive success and the process of natural selection on adult size and return timing. Because the estimates of selection were calculated on the parental generation using individual reproductive success, the rate of immigrants in the offspring generation has no bearing on selection estimates. However, we do acknowledge that offspring of the adults that might have emigrated to neighboring streams were not measured in our study, and we might have underestimated individual reproductive success. We assumed that offspring straying was not affected by parental phenotype.

In both 2006 and 2007, the majority of the fish did not produce any returning adult offspring. However, overall reproductive success of the 2006 parental cohort was higher than the 2007 parental cohort across both sexes and age groups. Such a drastic decrease in reproductive success may be explained by the substantial autumn flood that occurred in late 2007. In this year, all individuals arrived by November 20th, therefore nests of all females were susceptible to disturbance from the flood that occurred on December 3rd. This large flood likely scoured egg pockets of the nests, destroying fish eggs during incubation (Lapointe *et al.* 2000). A decrease in reproductive success may also be explained by density dependent effects; there was a greater number of spawners present in 2007. Long-term data from Big Beef Creek coho salmon has shown that there is an inverse relationship between the number of female spawners and the number of coho smolts produced by each female (Kinsel *et al.* 2009). This relationship is likely due to overspawning mortality because greater density of female spawners results in nest destruction and use of less suitable habitat (Fukushima *et al.* 1998; Essington *et al.* 2000; Blanchfield and Ridgway 2005).

The direction and strength of selection differed for both traits across all groups between the two parental brood years, and these results might be attributable to variation in return numbers, sex ratios and differences in environment. In 2006, precipitation did not occur until the beginning of November, delaying the return timing of the population (Figure 1.3c). Over 80% of

the population returned in the first seven days of the return season, and the male-female sex ratio was particularly high at the beginning of the season (Supplementary Material S1.1). Selection favoring early return in 3-year-old males suggests that earlier returning individuals may have had increased opportunities for mating under these circumstances (Quinn 2005; Morbey 2000). Changes in male to female sex ratio during the season revealed that there were consistently more males present; an absence of selection in females thus implies that they were able to secure high-quality sites for nests and obtain mates throughout the spawning season. Interestingly, males of both age classes returned from November 2nd to 22nd and from December 12th to 14th. These were the periods of high density in the stream, possibly explaining significant disruptive selection on date of return in 2-year-old males. As the ability of older, larger males to monopolize access to females tends to decrease at higher densities of spawners (Fleming and Gross 1994), "sneaking tactics" employed by 2-year-old males were likely more effective during these periods.

In 2007, the return period was protracted, and individuals returned earlier over a series of peaks from October 1st to November 20th (Figure 1.3d). The male to female sex ratio remained low for the majority of the season (Supplementary Material S1.1), therefore earlier return might not have been necessary for males to maximize mating opportunities. In females, however, the low male to female sex ratio and the high number of returning females suggest that the opportunity for intrasexual competition among females would have been high. Therefore, return in mid-season may have been important to both secure high-quality sites for nests and gain mates.

Mating success of male salmon is often determined by female choice and intrasexual competition, and studies suggest that there is a positive relationship between male social status and body length (Quinn and Foote 1994; Quinn *et al.* 2001; Quinn 2005). Unlike in 2006, males were more abundant in 2007; therefore, intrasexual competition between males in 2007 might have resulted in directional selection, favoring larger size in 3-year-old males of this cohort. Although no selection on size was detected in 2-year-old males in 2007, this result may be due to a lack of power because these males were rarer in 2007. Mating success of small, younger-maturing males ("jacks") is frequency-dependent relative to large, older-maturing males (Gross 1985; Quinn 2005), and jacks enjoy reproductive advantages when their form is rarer (Gross 1985, 1991; Foote *et al.* 1997; Fleming 1998; Quinn 2005; Koseki and Fleming 2007). Therefore, it is possible that 2-year-old males had reproductive advantages in 2007, as they were uncommon (39% of all males in 2006; 6% of all males in 2007). In females, selection favored large size in 2006 but intermediate size in 2007. These results were unexpected, as numerous reproductive advantages are known to accompany greater size (van den Berghe and Gross 1989; Fleming 1998; Beacham and Murray 1993; Fleming 1996; Andersson 1994; Steen and Quinn 1999; Dickerson, Quinn, and Willson 2002; van den Berghe and Gross 1984; Olsen and Vøllestad 2003; van den Berghe and Gross 1986; Morita and Fukuwaka 2006; Foote 1990; Kitano 1996). However, several studies have suggested that selection may not always favor large size, because efforts of attaining large size also increase the cost associated with growth rate (Healey 1986; Holtby and Healey 1986; Seamons *et al.* 2007).

Observed patterns of selection can be compared to other studies on natural populations of steelhead trout (19 cohorts; Seamons, Bentzen, and Quinn 2007), coho salmon (two cohorts; Ford *et al.* 2008), and a colonizing population of coho salmon (three cohorts; Anderson *et al.* 2010). All studies estimated selection on body length and return date in males and females. Consistent directional selection toward later return date was observed by Ford *et al.* (2008). In contrast, our findings support the results of Seamons *et al.* (2007) and Anderson *et al.* (2010), who observed fluctuation in the mode, direction and strength of selection on return date in both sexes. Variation in selection might have been observed because of the large number of replicates in Seamons *et al.* (2007) and Anderson *et al.* (2010). As we observed, environmental conditions (e.g., precipitation) likely affect return date in salmon; thus, accurate patterns of selection on this trait may only be obtained using sufficient replicates. Anderson *et al.* (2010) observed consistent directional selection on body length, favoring large size in both sexes. In contrast, and similar to our study, Seamons *et al.* (2007) and Ford *et al.* (2008) observed fluctuations in the mode and direction of selection on this trait in both sexes. Such results may reflect temporal variation in intraspecific competition. In particular, Ford *et al.* (2008) observed changing proportions of younger- and older-maturing males between two cohorts; different frequencies of these males potentially resulted in variation in intensity of intraspecific competition, leading to fluctuating selection. Additionally, Seamons *et al.* (2008) attributed changes in sex ratio and breeding density in both sexes as a potential cause of fluctuating selection. Our findings support both studies because changes in the proportion of younger- and older-maturing males, sex ratio and breeding density likely caused fluctuations in selection.

Because coho salmon at Big Beef Creek is a long established population, the distribution of studied traits was presumed to be at a stable optimum, and weak selection was expected. However, strong selection was observed in some years, and the direction and strength of selection were not consistent between two cohorts in this study. The extensive review on selection in natural populations found that the direction and strength of selection vary substantially and that quadratic selection is typically quite weak in wild (Kingsolver *et al.* 2001; Kingsolver and Diamond 2011). Established wild populations may be "chasing" fluctuating optima, and that selection can be quite strong in some years, and not so in others, and that the direction of selection varies over time (Siepielski *et al.* 2009). Such fluctuations in selection may act to maintain phenotypic variation in the traits in question, and we may have observed this process in our study.

We found evidence of selection on body length and date of return to the spawning ground, both of which are important fitness-related traits in salmonids. Reproductive success and the mode, direction and strength of selection widely varied between two parental brood years; differences may be due to factors such as annual changes in precipitation, occurrence of catastrophic events (flooding), the proportion of younger- versus older-maturing males, sex ratio and densities of spawners. As these factors are seldom consistent each year, long-term studies may be important to gain insights into anticipated evolutionary change.

Acknowledgements

We thank Matthew Gillum and Eric Kummerow for field sampling and collection of phenotypic data, and Mara Zimmerman and Clayton Kinsel for help assessing WDFW unpublished data. We also thank Isadora Jimenez-Hidalgo for assistance with molecular analyses. Dave Beauchamp, Todd Seamons and Joseph Anderson provided thoughtful discussions and helpful suggestions. Kotaro Ono and Eric Ward provided invaluable comments on statistical methods. Funding for this study was provided by NOAA Fisheries/Federal Columbia River Power System (FCRPS) Biological Opinion Remand Funds, School of Aquatic and Fishery Sciences and Graduate Opportunities and Minority Achievement Program (GO-MAP), University of Washington.

Figures

Figure 1.1 - Reproductive success (RS) of (a) 2006 and (b) 2007 parental cohorts of coho salmon. The number of adult offspring produced by 3-year-old males (black), 2-year old males (gray) and females (white). Results are displayed as the proportion of parents producing a given number of adult offspring.

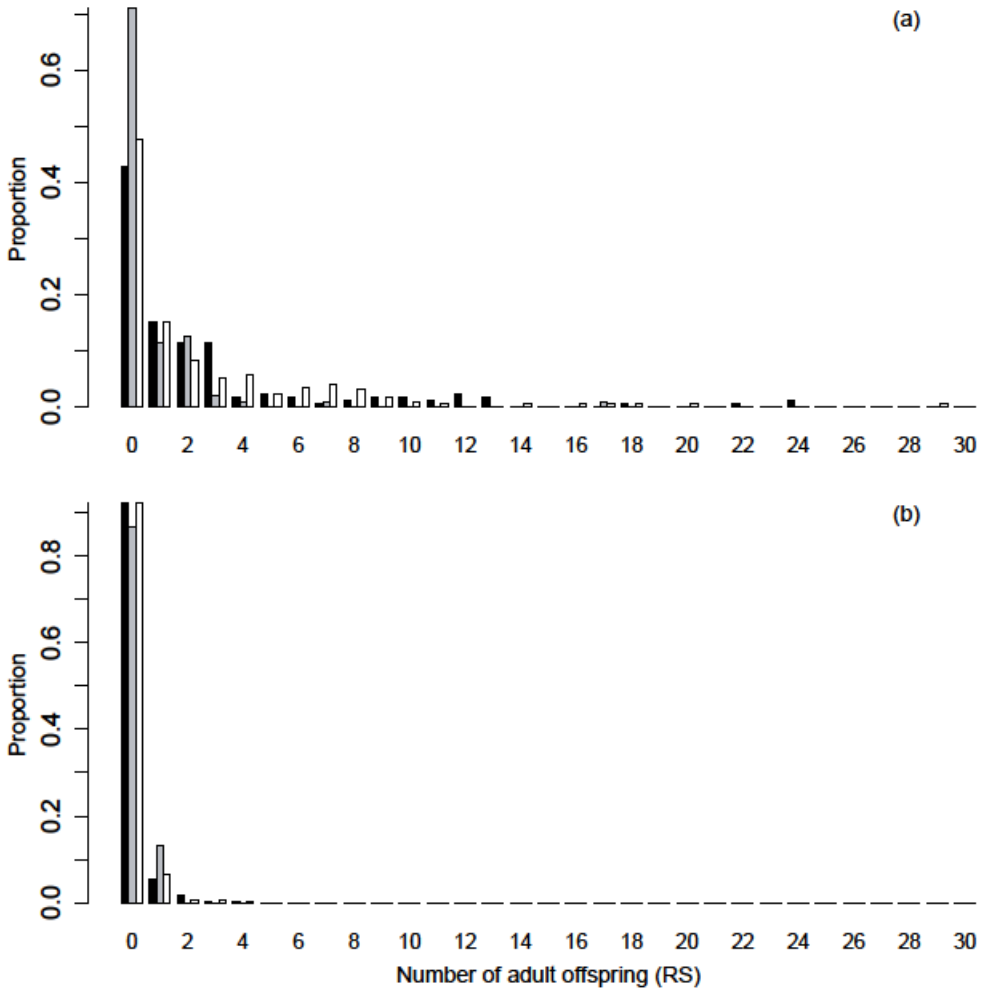


Figure 1.2 - Relationship between date of return (a, b) or fork length (c, d) and reproductive success. Plot (a) and (c) are for the 2006 parental cohort, and plot (b) and (d) are for the 2007 parental cohort. Observed data are shown as thick lines for 3-year-old males, thick dashed lines for 2-year-old males and thin lines for females. Note the different scales on *y* axes.

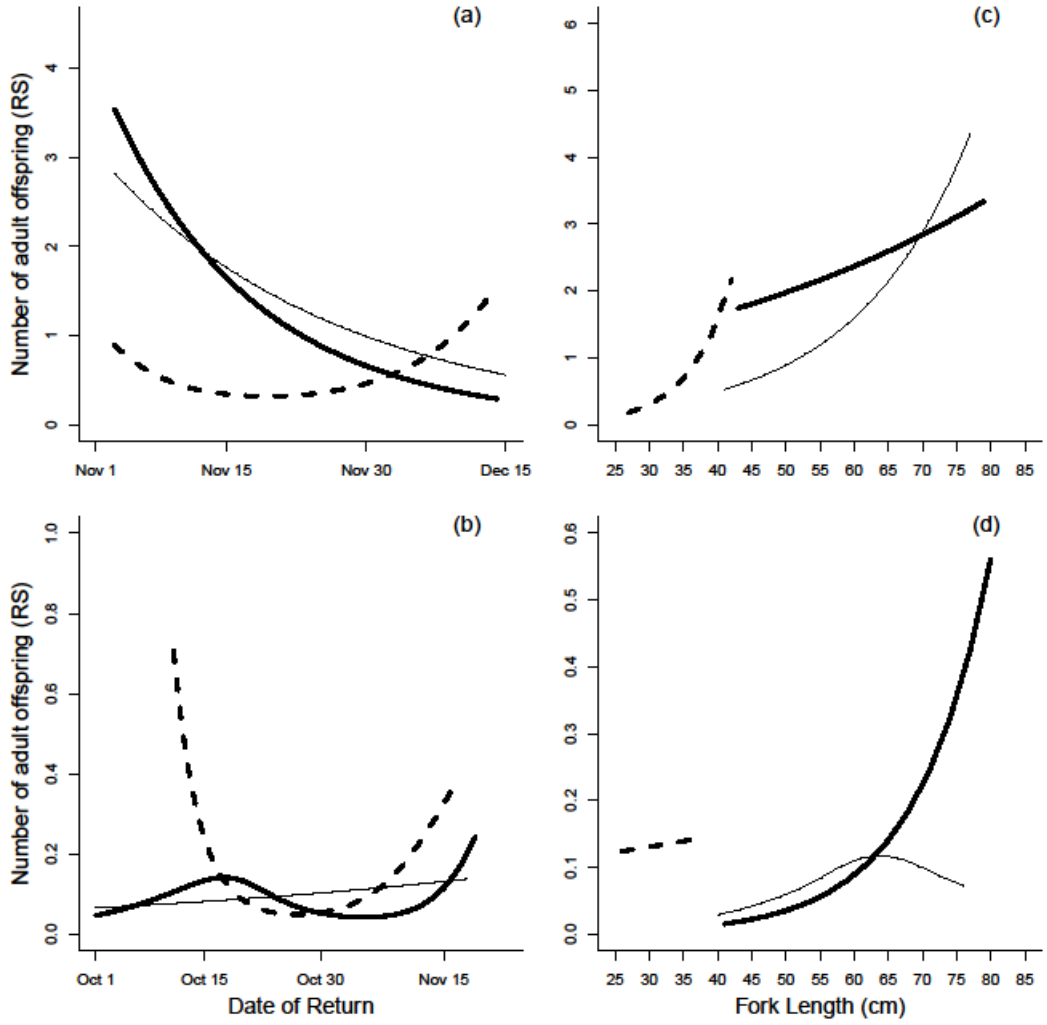
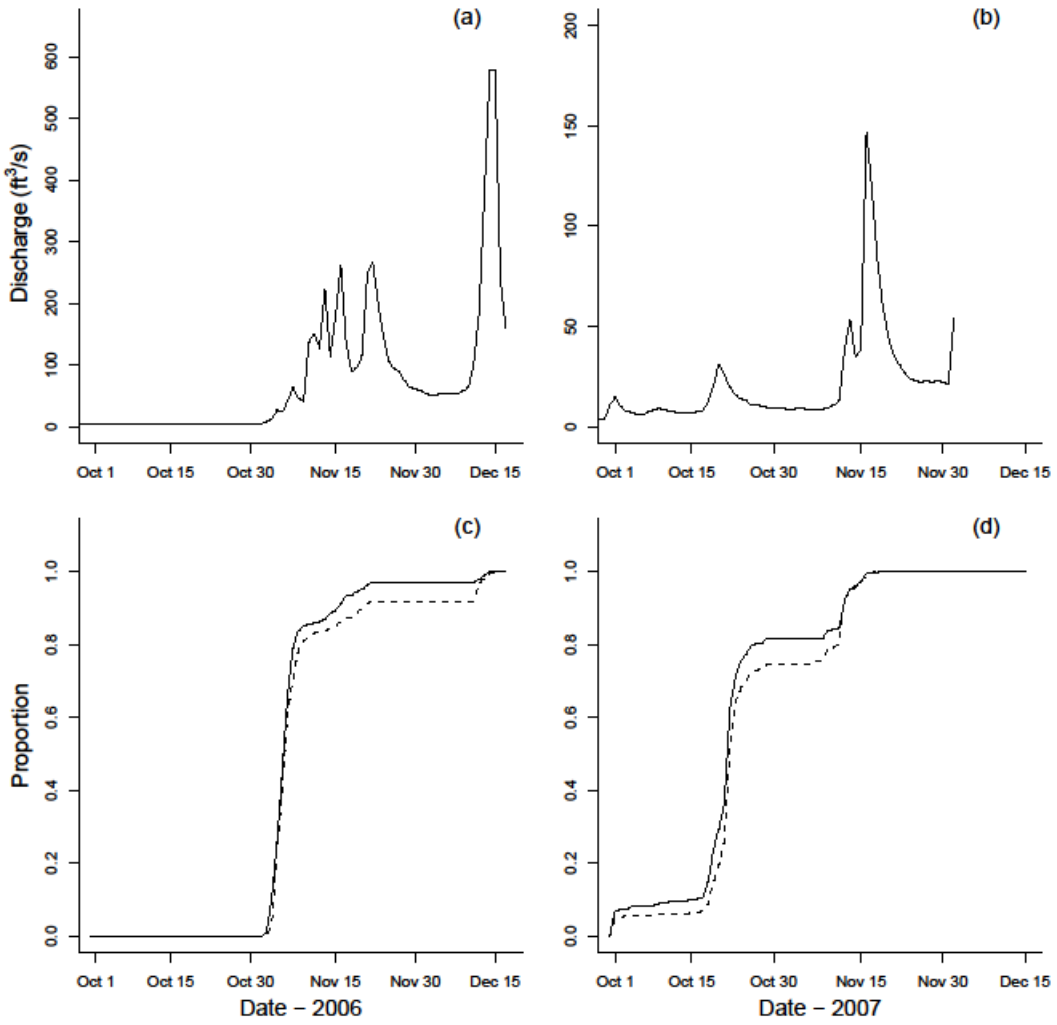


Figure 1.3 - The amount of water discharge (a, b) and the proportion of arrived spawners (c, d). Plot (a) and (c) are for 2006, and plot (b) and (d) are for 2007. For the proportion of arrived spawners, observed data are shown as solid lines for males and dashed lines for females. In 2007, discharge data was unavailable from December 3rd to December 15th. Note the different scales on stream discharge.



Tables

Table 1.1 - Summary of phenotypic information and reproductive success. Number of individuals (N), sex ratio, ratio of 2-year-old to 3-year-old males, date of return, body length, reproductive success (RS). One standard deviation is shown in parentheses.

Year	Sex	Age	N	Sex Ratio$_{male/female}$	Ratio$_{2\text{-year-old/3-year-old male}}$	Mean calendar day (days)	Mean length (cm)	Mean RS	Variance RS
2006	Male	2	104	1.17	0.63	310 (6.20)	34.15 (3.60)	0.69	3.75
	Male	3	166			312 (7.93)	66.57 (7.51)	2.69	20.18
	Female	3	231		NA	313 (10.58)	65.48 (5.40)	2.31	14.98
2007	Male	2	30	0.83	0.06	299 (10.26)	31.56 (3.70)	0.13	0.12
	Male	3	504			296 (11.19)	58.88 (7.66)	0.11	0.17
	Female	3	643		NA	299 (11.23)	59.20 (5.50)	0.10	0.15

Table 1.2 - Summary of the parentage analysis. Number of offspring that were assigned to both parents, to a father only, to a mother only, or to no parents for the 2006 and 2007 cohort.

Year	Both Parents	Father Only	Mother Only	No Parents	Total
2006	470	47	33	430	980
2007	46	14	5	186	251

Table 1.3 - Selection gradients for body length and date of return. Coefficients of the best models of relative fitness for the 2006 and 2007 parental cohort. Models were chosen based on the Akaike information criterion (AIC) score. Because of the log-link used with the zero-inflated model, these coefficients are in log space and quadratic terms are not transformed (doubled). One standard deviation is shown in parentheses.

Year	Sex	Age	N	Intercept	Date	Length	Date^2	Length^2	Date*Length
2006	Male	2	104	-0.57* (0.29)	-0.77 (0.53)	-0.29 (0.32)	0.77*** (0.20)	0.27 (0.20)	-3.23*** (0.92)
	Male	3	166	-0.03 (0.24)	-0.50** (0.16)	NA	NA	NA	NA
	Female	3	231	0.34 (0.21)	-0.10 (0.36)	0.30* (0.12)	-0.12 (0.14)	NA	-0.14 (0.15)
2007	Male	2	30	0.14 (0.50)	NA	-0.12 (0.58)	NA	NA	NA
	Male	3	504	0.01 (0.25)	-0.25 (0.28)	0.83*** (0.24)	NA	-0.12 (0.15)	0.19 (0.18)
	Female	3	643	2.12** (0.76)	0.32 (0.25)	-0.12 (0.27)	-0.53* (0.23)	-0.63** (0.23)	-0.09 (0.30)

*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$

Supplementary Material

S1.1 - Changes in male to female sex ratio. Plot (a) is for the 2006 parental cohort, and plot (b) is for the 2007 parental cohort. Note the different scales on *y* axes.

S1.2 - Microsatellite loci used for parentage analysis of coho salmon. T_a = annealing temperature. Repeat units and allele ranges are given in base pairs (bp).

S1.3 - Population genetic data at 11 microsatellite loci in coho salmon samples. N = number of genotyped inidividuals, N_A = number of alleles, H_O = observed heterozygosity, H_E = expected heterozygosity, P_{HWE} = probability of Hardy-Weinberg equilibrium (significant results in bold), F_{IS} = inbreeding coefficient.

References

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart, 2010 Genomics and the future of conservation genetics. Nat. Rev. Genet. 11: 697-709.

Anderson, J. H., P. L. Faulds, W. I. Atlas, G. R. Pess, and T. P. Quinn, 2010 Selection on breeding date and body size in colonizing coho salmon, *Oncorhynchus* kisutch. Mol. Ecol. 19: 2562-2573.

Andersson, M., 1994 *Sexual selection*. Princeton University Press, Princeston.

Araki, H., and M. S. Blouin, 2005 Unbiased estimation of relative reproductive success of different groups: evaluation and correction of bias caused by parentage assignment errors. Mol. Ecol. 14: 4097-4109.

Beacham, T. D., and C. B. Murray, 1993 Fecundity and egg size variation in North American Pacific salmon (*Oncorhynchus*). J. Fish Biol. 42: 485-508.

Bell, G., 2010 Fluctuating selection: the perpetual renewal of adaptation in variable environments. Philos. Trans. R. Soc. London Ser. B-Biological Sci. 365: 87-97.

Van den Berghe, E. P., and M. R. Gross, 1984 Female Size and Nest Depth in Coho Salmon (*Oncorhynchus* kisutch). Can. J. Fish. Aquat. Sci. 41: 204-206.

Van den Berghe, E. P., and M. R. Gross, 1986 Length of breeding life of coho salmon(*Oncorhynchus* kisutch). Can. J. Zool. 64: 1482-1486.

Van den Berghe, E. P., and M. R. Gross, 1989 Natural selection resulting from female breeding competition in a Pacific salmon (coho: *Oncorhynchus* kisutch). Evolution (N. Y). 43: 125-140.

Blanchfield, P. J., and M. S. Ridgway, 2005 The relative influence of breeding competition and habitat quality on female reproductive success in lacustrine brook trout (Salvelinus fontinalis). Can. J. Fish. Aquat. Sci. 2694-2705.

Brodie, E. D., A. J. Moore, and F. J. Janzen, 1995 Visualizing and quantifying natural selection. Trends Ecol. Evol. 10: 313-318.

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

Conner, J. K., 2001 How strong is natural selection? Trends Ecol. Evol. 16: 215-217.

Crandall, K. A., and O. R. P. Bininda-Emonds, 2000 Considering evolutionary processes in conservation biology. Trends Ecol. Evol. 15: 290-295.

Crozier, L. G., A. P. Hendry, P. W. Lawson, T. P. Quinn, N. J. Mantua *et al.*, 2008 Potential responses to climate change in organisms with complex life histories: evolution and plasticity in Pacific salmon. Evol. Appl. 1: 252-270.

Dickerson, B. R., T. P. Quinn, and M. F. Willson, 2002 Body size, arrival date, and reproductive success of pink salmon, *Oncorhynchus* gorbuscha. Ethol. Ecol. Evol. 14: 29-44.

Essington, T. E., T. P. Quinn, and V. E. Ewert, 2000 Intra- and inter-specific competition and the reproductive success of sympatric Pacific salmon. Can. J. Fish. Aquat. Sci. 57: 205-213.

Fisher, R. A., 1930 *The genetical theory of natural selection*. Clarendon Press, Oxford.

Fleming, I. A., 1998 Pattern and variability in the breeding system of Atlantic salmon (*Salmo* salar), with comparisons to other salmonids. Can. J. Fish. Aquat. Sci. 55: 59-76.

Fleming, I. A., 1996 Reproductive strategies of Atlantic salmon: ecology and evolution. Rev. Fish Biol. Fish. 6: 379-416.

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

Foote, C. J., 1990 An experimental comparison of male and female spawning territoriality in a Pacific salmon. Behaviour 115: 283-314.

Foote, C. J., G. S. Brown, and C. C. Wood, 1997 Spawning success of males using alternative mating tactics in sockeye salmon, *Oncorhynchus* nerka. Can. J. Fish. Aquat. Sci. 54: 1785-1795.

Ford, M. J., J. J. Hard, B. Boelts, E. Lahood, and J. Miller, 2008 Estimates of natural selection in a salmon population in captive and natural environments. Conserv. Biol. 22: 783-794.

Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen, and E. B. Taylor, 2011 Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity (Edinb). 106: 404-420.

Fukushima, M., T. J. Quinn, and W. W. Smoker, 1998 Estimation of eggs lost from superimposed pink salmon (*Oncorhynchus* gorbuscha) redds. Can. J. Fish. Aquat. Sci. 55: 618-625.

Garant, D., J. J. Dodson, and L. Bernatchez, 2001 A genetic evaluation of mating system and determinants of individual reproductive success in Atlantic salmon (*Salmo* salar L.). J. Hered. 92: 137-145.

Goddard, M. E., and B. J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381-391.

Gross, M. R., 1985 Disruptive selection for alternative life histories in salmon. Nature 313: 47-48.

Gross, M. R., 1991 Salmon breeding behavior and life history evolution in changing environments. Ecology 72: 1180-1186.

Gustafson, R. G., R. S. Waples, J. M. Myers, L. A. Weitkamp, G. J. Bryant *et al.*, 2007 Pacific salmon extinctions: quantifying lost and remaining diversity. Conserv. Biol. 21: 1009-1020.

Healey, M. C., 1986 Optimum size and age at maturity in Pacific salmon and effects of size-selective fisheries. Can. Spec. Publ. Fish. Aquat. Sci. 89: 39-52.

Hendry, A. P., B. H. Letcher, and G. Gries, 2003 Estimating natural selection acting on stream-dwelling Atlantic salmon: implications for the restoration of extirpated populations. Conserv. Biol. 17: 795-805.

Hilborn, R., and M. Mangel, 1997 *The ecological detective*. Princeton University Press, Princeston.

Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang *et al.*, 2001 Strength and tempo of directional selection in the wild. Proc. Natl. Acad. Sci. U. S. A. 98: 9157-9160.

Holtby, L. B., and M. C. Healey, 1986 Selection for adult size in female coho salmon (*Oncorhynchus* kisutch). Can. J. Fish. Aquat. Sci. 43: 1946-1959.

Holtby, L. B., and M. C. Healey, 1990 Sex-specific life history tactics and risk-taking in coho salmon. Ecology 71: 678-690.

Jones, O. R., and J. Wang, 2009 COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol. Ecol. Resour. 10: 551-555.

Kimura, M., 1984 *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

Kingsolver, J. G., and S. E. Diamond, 2011 Phenotypic selection in natural populations: what limits directional selection? Am. Nat. 177: 346-357.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri *et al.*, 2001 The strength of phenotypic selection in natural populations. Am. Nat. 157: 245-261.

Kingsolver, J. G., and D. W. Pfennig, 2007 Patterns and power of phenotypic selection in nature. Bioscience 57: 561-572.

Kinsel, C., P. Hanratty, M. Zimmerman, B. Glaser, S. Gray *et al.*, 2009 Intensively Monitored Watersheds: 2008 Fish Population Studies in the Hood Canal and Lower Columbia Stream Complexes. Popul. Stud. (NY).

Kitano, S., 1996 Size-related factors causing individual variation in seasonal reproductive success of fluvial male Dolly Varden (Salvelinus malma). Ecol. Freshw. Fish 5: 59-67.

Koseki, Y., and I. A. Fleming, 2007 Large-scale frequency dynamics of alternative male phenotypes in natural populations of coho salmon (*Oncorhynchus* kisutch): patterns, processes, and implications. Can. J. Fish. Aquat. Sci. 64: 743-753.

Lande, R., and S. J. Arnold, 1983 The measurement of selection on correlated characters. Evolution (N. Y). 37: 1210-1226.

Lapointe, M., B. Eaton, S. Driscoll, and C. Latulippe, 2000 Modelling the probability of salmonid egg pocket scour due to floods. Can. J. Fish. Aquat. Sci. 57: 1120-1130.

Morbey, Y., 2000 Protandry in Pacific salmon. Can. J. Fish. Aquat. Sci. 57: 1252-1257.

Morbey, Y. E., and P. A. Abrams, 2004 The interaction between reproductive lifespan and protandry in seasonal breeders. J. Evol. Biol. 17: 768-778.

Morita, K., and M. Fukuwaka, 2006 Does size matter most? The effect of growth history on probabilistic reaction norm for salmon maturation. Evolution (N. Y). 60: 1516-1521.

Moritz, C., 2002 Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst. Biol. 51: 238-254.

National Research Council, 1995 *Science and the endangered species act*. National Academies Press.

Olsen, E. M., and L. A. Vøllestad, 2003 Microgeographical variation in brown trout reproductive traits: possible effects of biotic interactions. Oikos 100: 483-492.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley, 2004 MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4: 535-538.

Orr, H. A., 1998 Using quantitative trait locus data. Genetics 149: 2099-2104.

Peakall, R., and P. E. Smouse, 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288-295.

Quinn, T. P., 2005 *The behavior and ecology of Pacific salmon and trout.* University of Washington Press, Seattle.

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

Quinn, T. P., A. P. Hendry, and G. B. Buck, 2001 Balancing natural and sexual selection in sockeye salmon: interactions between body size, reproductive opportunity and vulnerability to predation by bears. Evol. Ecol. Res. 3: 917-937.

Raymond, M., and F. Rousset, 1995 GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. J. Hered. 86: 248-249.

Riester, M., P. F. Stadler, and K. Klemm, 2009 FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics 25: 2134-2139.

Sandercock, F. K., 1991 *Life history of coho salmon (Oncorhynchus kisutch).* University of British Columbia Press, Vancouver, Canada.

Seamons, T. R., P. Bentzen, and T. P. Quinn, 2007 DNA parentage analysis reveals inter-annual variation in selection: results from 19 consecutive brood years in steelhead trout. Evol. Ecol. Res. 9: 409-431.

Seamons, T. R., P. Bentzen, and T. P. Quinn, 2004 The effects of adult length and arrival date on individual reproductive success in wild steelhead trout (*Oncorhynchus* mykiss). Can. J. Fish. Aquat. Sci. 61: 193-204.

Serbezov, D., L. Bernatchez, E. M. Olsen, and L. A. Vøllestad, 2010 Mating patterns and determinants of individual reproductive success in brown trout (*Salmo* trutta) revealed by parentage analysis of an entire stream living population. Mol. Ecol. 19: 3193-3205.

Siepielski, A. M., J. D. DiBattista, and S. M. Carlson, 2009 It's about time: the temporal dynamics of phenotypic selection in the wild. Ecol. Lett. 12: 1261-1276.

Siepielski, A. M., J. D. DiBattista, J. A. Evans, and S. M. Carlson, 2010 Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. Proc. R. Soc. B Biol. Sci. 278: 1572-1580.

Simpson, G. G., 1953 The Baldwin effect. Evolution (N. Y). 7: 110-117.

Steen, R. P., and T. P. Quinn, 1999 Egg burial depth by sockeye salmon (*Oncorhynchus* nerka): implications for survival of embryos and natural selection on female body size. Can. J. Zool. 77: 836-841.

Stockwell, C. A., A. P. Hendry, and M. T. Kinnison, 2003 Contemporary evolution meets conservation biology. Trends Ecol. Evol. 18: 94-101.

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

Theriault, V., L. Bernatchez, and J. J. Dodson, 2007 Mating system and individual reproductive success of sympatric anadromous and resident brook charr, Salvelinus fontinalis, under natural conditions. Behav. Ecol. Sociobiol. 62: 51-65.

USGS USGS Water Data for Washington.

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

Wang, J., and A. W. Santure, 2009 Parentage and Sibship Inference From Multilocus Genotype Data Under Polygamy. Genetics 181: 1579-1594.

Waples, R. S., 1994 Genetic considerations in recovery efforts for Pacific salmon. Conserv. Biol. 8: 884-886.

Waples, R. S., and A. P. Hendry, 2008 Special Issue: Evolutionary perspectives on salmonid conservation and management. Evol. Appl. 1: 183-188.

Weitkamp, L. A., T. C. Wainwright, G. J. Bryant, G. B. Milner, D. J. Teel *et al.*, 1995 Status review of coho salmon from Washington, Oregon, and California:.

Wood, S., 2001 mgcv: GAMs and generalized ridge regression for R. R News 20-25.

Zar, J. H., 1999 *Biostatistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey.

Zeileis, A., C. Kleiber, and S. Jackman, 2008 Regression models for count data in R. J. Stat. Softw. 27: 1-25.

Chapter 2: Comparative mapping between coho salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event

Miyako Kodama[1], Marine S O Brieuc[1], Robert H Devlin[2], Jeffrey J Hard[3] and Kerry A Naish[1]

[1] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98105, USA

[2] Fisheries and Oceans Canada, West Vancouver, British Columbia, V7K 1N6, Canada

[3] National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA 98112, USA

*This chapter has been published as Comparative mapping between coho salmon (*Oncorhynchus* kisutch) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event, Kodama M, Brieuc MSO, Devlin RH, Hard JJ and Naish KA, G3: Genes | Genomes | Genetics 4 (9), 1717-1730,

Abstract

Whole genome duplication has been implicated in evolutionary innovation and rapid diversification. In salmonid fishes, however, whole genome duplication significantly pre-dates major transitions across the family, and re-diploidization has been a gradual process between genomes that have remained essentially collinear. Nevertheless, pairs of duplicated chromosome arms have diverged at different rates from each other, suggesting that the retention of duplicated regions through occasional pairing between homeologous chromosomes may have played an evolutionary role across species pairs. Extensive chromosomal arm rearrangements have been a key mechanism involved in re-dipliodization of the salmonid genome and so we investigated their influence on degree of differentiation between homeologs across salmon species. We derived a linkage map for coho salmon and performed comparative mapping across syntenic arms within the genus *Oncorhynchus*, and with the genus *Salmo*, to determine the phylogenetic relationship between chromosome arrangements and the retention of undifferentiated duplicated regions. A 6596.7cM female coho salmon map, comprising 34 linkage groups with 7415 and 1266 non-duplicated and duplicated loci respectively, revealed uneven distribution of duplicated loci along and between chromosome arms. These duplicated regions were conserved across syntenic arms across *Oncorhynchus* species, and were identified in metacentric chromosomes likely formed ancestrally to the divergence of *Oncorhynchus* from *Salmo*. These findings support earlier studies, where observed pairings involved at least one metacentric chromosomes. Re-diploidization in salmon may have been prevented or retarded by the formation of metacentric chromosomes following the whole genome duplication event, and may explain lineage-specific innovations in salmon species if functional genes are found in these regions.

Introduction

Whole genome duplication (WGD) is a mutational mechanism that can serve as a primary driver of evolutionary novelty (Ohno 1970; Zhang 2003; Crow and Wagner 2006; Lynch 2007; Edger and Pires 2009). Changes in ploidy levels following WGD can lead to dramatic alterations at the cellular and phenotypic level (Mayfield-Jones *et al.* 2013), and provide additional genetic variation for mutation, drift and selection to act upon. These evolutionary processes can result in new adaptations and species diversification (Van de Peer *et al.* 2009; Storz *et al.* 2013). Genome sequencing projects are increasingly revealing that WGD is widespread in many key lineages, such as flowering plants and vertebrates, and represents an ongoing phenomenon in many species (Otto and Whitton 2000; Van de Peer *et al.* 2009). Understanding the processes governing the return to a diploid mode - diploidization - by comparing the genomes of species descended from a WGD event can provide insights into the event's role in evolutionary innovation and persistence of duplicated regions (Jaillon *et al.* 2009; Mayfield-Jones *et al.* 2013).

The stabilization of the duplicated genome through diploidization can be achieved by rearrangements (such as translocations, fissions, fusions and transpositions), gene loss, and sequence deletion and divergence (Hufton and Panopoulou 2009; Schubert and Lysak 2011). These processes tend to reduce the similarity of the duplicated ohnologs (Wolfe 2001), and the homeologous chromosomes resulting from WGD, but the exact mechanisms vary across lineages (Hufton and Panopoulou 2009). Whole genome duplication has been frequently implicated in evolutionary innovation in eukaryotic genomes of paleopolyploids (ancient polyploids; Ohno 1999; Lynch and Conery 2000; Jaillon *et al.* 2004; Cañestro *et al.* 2013), but evidence in plants suggests that the rate of diversification and extinction of neopolyploids can be lower than that of related diploid lineages (Mayrose *et al.* 2011). Increasing the number of studies on mesopolyploids - organisms in the intermediate process of diploidization (Mayfield-Jones *et al.* 2013) - will provide a clearer understanding of contribution of WGD events to evolutionary innovation.

Salmonid fishes are descended from a whole genome duplication event in an autotetraploid ancestor (Allendorf and Thorgaard 1984), distinct from the second round of duplication (2R) that occurred basal to the vertebrate tree and the third round (3R) early in the evolution of the teleosts, 225 to 333 Mya (Hurley *et al.* 2007; Postlethwait 2007; Santini *et al.* 2009; Near *et al.* 2012). This fourth round (4R) of duplication was recently estimated as occurring 88-103 million years ago (Macqueen and Johnston 2014; see also Crête-Lafrenière *et al.* 2012; Alexandrou *et al.* 2013; Berthelot *et al.* 2014). Although the genomes of these species are returning to a stable diploid state through chromosomal rearrangements and divergence of homeologous chromosomes, evidence of tetrasomic inheritance in males and extensive rearrangements among chromosomes has shown that restoration of diploidy is not yet complete (Wright *et al.* 1983; Allendorf and Thorgaard 1984; Allendorf and Danzmann 1997). Comparative genome sequencing between ohnologs in rainbow trout has revealed extensive collinearity between the duplicated chromosomes, loss of about half the protein-coding regions through pseudogenization, but retention of most of the duplicated miRNA genes (Berthelot *et al.* 2014).

The role of the WGD event in salmonid trait innovation and diversification is unclear. Recent evidence based on molecular clock estimates suggest that duplication is unlinked to a major transition in life history, anadromy (Alexandrou *et al.* 2013; Macqueen and Johnston 2014), and preceded rapid species diversification by several million years (Berthelot *et al.* 2014; Macqueen and Johnston 2014). Rather, both transitions appear to correspond with climate cooling (Macqueen and Johnston 2014). Re-diploidization has been subsequently characterized as a gradual process unlinked to significant genome rearrangements (Berthelot *et al.* 2014). However, it has also been argued that the duplication event might have provided the raw material for evolution to act upon, and that differential divergence of duplicated regions might have promoted speciation at varying time points (Macqueen and Johnston 2014). Large-scale genome characterization in the salmonids is increasingly revealing the location of genes or regions that may have played a role in adaptation and diversification (Davidson *et al.* 2010; Bourret *et al.* 2013; Larson *et al.* 2013). Therefore, it is important to combine these studies with an understanding of the mechanism and timing of divergence between homeologous chromosome arms across salmon lineages so that it will be possible to understand how duplication played a role in evolution of salmon, and whether key genomic regions might explain innovation across a subset of species.

It has been known for some time that one of the key mechanisms for diploidization within the subfamilies Coregoninae and Salmoninae (which includes *Salmo*, *Salvelinus*, and *Oncorhynchus*) has occurred through Robertsonian rearrangements of whole chromosome arms (Ohno 1999; Phillips and Ráb 2001). Most chromosome arms are syntenic between Salmoninae species, and the combined efforts of genome mapping, and karyotyping have permitted alignment of chromosome arms among several species within this subfamily (Danzmann *et al.* 2005; Phillips *et al.* 2009; Lubieniecki *et al.* 2010; Lien *et al.* 2011; Timusk *et al.* 2011; Guyomard *et al.* 2012; Naish *et al.* 2013; Phillips *et al.* 2013). Chromosome arm number is largely conserved (NF = approx. 100) but the numbers of chromosomes vary substantially across species as a result of the Robertsonian rearrangements (Phillips and Ráb 2001). The exception is Atlantic salmon, with reduced chromosome arm number compared to the other species (NF = 72-74). However, large syntentic blocks within the arms of this species correspond to whole arms in other members of the Salmoninae, making comparative studies feasible across this subfamily as a whole.

Comparative mapping between Chinook salmon (*O.tshawytscha*) and rainbow trout (*O.mykiss*) has revealed evidence for the retention of at least eight metacentric chromosomes and four acrocentric chromosomes that are ancestral to species divergence within the genus *Oncorhynchus* (Naish *et al.* 2013; Phillips *et al.* 2013). One of the eight metacentric chromosomes and one of the acrocentric chromosomes are also ancestral to the divergence between *Salmo* and *Oncorhynchus*. There is also further evidence for another ancestral metacentric and an ancestral acrocentric chromosome, but these have undergone subsequent rearrangements within at least one descendant species (Naish *et al.* 2013; Ostberg *et al.* 2013; Brieuc *et al.* 2014). High-density linkage maps have revealed that recently diverged or undifferentiated duplicated loci are not uniformly distributed among chromosomes (Atlantic salmon, Lien *et al.* 2011; Chinook salmon, Brieuc *et al.* 2014), and the biased distribution of duplicated loci along chromosome arms provides evidence that pairs of homeologous arms have diverged at different rates from each other (Brieuc *et al.* 2014). This finding confirms

observations from previous studies conducted with lower marker densities (e.g. Danzmann *et al.* 2005; Guyomard *et al.* 2012). Intriguingly, homeologous pairings have been observed to include at least one metacentric chromosome (Wright *et al.* 1983), and duplicated markers map to such chromosomes (Brieuc *et al.* 2014), supporting the view that metacentric chromosomes play an important role in homeologous pairing (Phillips *et al.* 2009; Brieuc *et al.* 2014). These observations raise the interesting possibility that the evolutionary timing of metacentric chromosome formation during re-diploidization following the WGD event might influence the retention, and hence the evolutionary role, of duplicated regions across species. Therefore, by comparing chromosomal arrangements and distribution of duplicated regions across salmon species, we aim to provide a contextual framework for the further investigation of loci involved in species diversification.

The development of a high density linkage map for a less-described salmon species will contribute further information on chromosome rearrangements that have already been defined in several salmon species, and enhance our understanding of the timing of these arrangements in a phylogenetic context. Examining the distribution of duplicated regions across individual chromosome arms in a second species - beyond Chinook salmon (Brieuc *et al.* 2014) - will also facilitate an understanding of the relationship between timing of metacentric chromosome formation and diversification between homeologs. Coho salmon (*Oncorhynchus kisutch*) is a species whose genome has not been extensively described to date. A low density linkage map of coho salmon has been constructed using microsatellites (McClelland and Naish 2008), but this map is not sufficiently resolved to study the consequences of WGD because there is a low number of duplicated loci mapped. A high density map in this species is feasible, given recently emerged sequencing technologies (e.g. Baird *et al.* 2008). Coho and Chinook salmon are sister species (Crête-Lafrenière *et al.* 2012); therefore, comparative mapping across coho and Chinook salmon, as well as more divergent species in the genus *Oncorhynchus* and *Salmo* will help validate the hypotheses and provide more robust evidence on the process of chromosomal evolution following WGD.

The aim of our research is to determine the relationship between chromosome arrangements and the retention of recently diverged or undifferentiated duplicated regions by deriving a linkage map for coho salmon and comparing this map with those of Chinook salmon, rainbow trout, and Atlantic salmon. We therefore constructed high-density linkage maps for coho salmon using restriction site associated DNA (RAD) sequencing (Baird *et al.* 2008). By achieving this objective we also produced a reference database of RAD markers that can be used for alignment of sequences generated in future work, and described in detail the properties of the coho linkage map. Coho salmon chromosome arms were identified by comparative mapping with Chinook salmon using markers in common between the species, and whole chromosome arm homologies were described across species to improve our current understanding of chromosome arm rearrangement within the genera *Oncorhynchus* and *Salmo*. Linkage groups representing homeologous chromosome arms in coho salmon were discovered using duplicated markers, and regions of duplicated markers were compared across species to determine the extent to which these regions were conserved across lineages. By identifying genomic regions that are in the process of diploidization, and linking these regions to chromosomal rearrangements, we aim to provide the basis for determining the role of duplication in

maintaining ongoing polymorphisms and explaining processes of diversification across Salmoninae species.

Materials and Methods

Justification and description of sample collection and experimental crosses

A two-step approach was used to develop genomic resources and construct linkage maps. First, RAD sequences from individuals sampled from multiple populations were used to construct a reference database for aligning loci across mapping families. This reference database was screened for errors, duplicated loci, and repeat regions following approaches described by (Brieuc *et al.* 2014), and loci were subsequently named to ensure consistency across mapping families. Second, specific cross types were used to perform the mapping: gynogenetic haploid crosses were used to map both duplicated and non-duplicated loci, and diploid crosses were used to construct sex-specific maps.

The reference database of RAD loci was constructed using sequences from 583 individuals representing four populations in the Pacific Northwest of the United States and Canada: 1) the Washington Department of Fish and Wildlife's (WDFW) Wallace River Hatchery (47º87'N, 121º71'W), 2) the Domsea broodstock population, which originated in 1973 and 1974 from Wallace River Hatchery, 3) Bingham Creek (47º15'N, 123º40'W), a tributary to the Satsop River in the Southwest Washington Coast/Lower Columbia ESU, and 4) Chehalis River located in British Columbia, Canada (49º29'N, 121º94'W).

An initial framework map was constructed using two haploid crosses (Haploid Family 1 and 2) comprising 64 and 62 individuals respectively. These types of crosses have the advantage of identifying duplicated loci, because these loci will appear as heterozygotes in the offspring if they are polymorphic, while non-duplicated loci will be homozygous. Haploid families were created at the University of Washington hatchery facility (47º65'N, 122º31'W), following the protocol of (Thorgaard *et al.* 1983). Embryos were collected before hatching and preserved in 100% ethanol.

Sex-specific maps were created using two F3 outbred diploid crosses and one outbred diploid cross. Specifically, F3 diploid crosses were created from a cultured line originally derived from an outbred cross between two populations in Washington State, USA (McClelland and Naish 2010). F0 Males were collected from Bingham Creek in Southwest Washington (47º15'N, 123º40'W). F0 Females were obtained from the Domsea broodstock farm. Two F3 crosses were established in December 2010 by mating two F2 full-sibs to create one family, and two F2 half-sibs to create the other. The two families comprised 55 and 67 offspring respectively (Diploid Family 1 and 2). An additional diploid outbred cross was created from an aquaculture population using coho salmon derived from the Chehalis River located in British Columbia, Canada (49º29'N, 121º94'W). Specifically, cultured individuals were repeatedly backcrossed with wild individuals from the Chehalis River for six generations, and diploid crosses were created in January 2011. One diploid family from these crosses, comprising 99 individuals was used for further analyses (Diploid Family 3).

DNA extraction, sequencing and amplification of sex-linked markers

Genomic DNA from the sampled individuals was extracted using the DNeasy extraction kit (QIAGEN, Valencia, CA, USA) following the manufacturer's procedures. The DNA was digested with SbfI, and a 6-nucleotide barcode was added to each sample for individual identification following protocols described in Baird *et al.* (2008). Between 24 to 36 individuals were pooled in a single library and sequenced with 100 bp single-read lengths using the Illumina HiSeq 2000 sequencer. The sequences were separated by individual using PROCESS_RADTAGS implemented in STACKS (Catchen *et al.* 2011; Catchen *et al.* 2013). Because the quality score of sequences decreased beyond 74 nucleotides, sequences were trimmed to 74 nucleotides in order to remove low quality sequences. A locus was defined as a 74-nucleotide RAD sequence for the purpose of this study.

Genetic sex was determined in the two diploid families (Diploid Family 1 and 2) using a Y-linked growth hormone pseudogene, (GH5 and GH6; Devlin *et al.* 2001) and sex-determining gene, sdY (sdY E2S1 and sdY E2AS4; Yano *et al.* 2012). Polymerase chain reactions were performed for each set of primers using a QIAGEN Multiplex PCR kit. Specifically, reaction mixtures consisted of 10-200ng genomic DNA, 1x QIAGEN Multiplex PCR Master Mix, 0.25 μM of GH5 and GH6, or 0.4μM of sdY E2S1 and sdY E2AS4, making up a total volume of 10μl. Cycling conditions consisted of a 15-min, initial activation step at 95°C, 30 cycles of 30-s denaturing step at 94°C, 90s annealing step at 60 °C and 60-s extension step at 72 °C, and a 10-min, final extension step at 72°C.

Reference database of RAD loci

RAD loci that are found within repeat regions, and loci containing repeat units, can confound the identification of unique loci. Therefore, a reference sequence database comprising a set of pre-screened RAD loci was first created from the survey of four populations following bioinformatic procedures fully described in Brieuc *et al.* (2014). This database served as a resource for aligning loci across studies. In brief, sequences from all 583 individuals sampled across the four populations described earlier were extracted using STACKS 0.9995 (Catchen *et al.* 2011). Both monomorphic and polymorphic loci that were sequenced with a depth greater than 5X in more than 496 individuals (85%) were retained in a temporary database and used for further screening.

Loci in the temporary database that corresponded to repeat regions and loci containing repeat units were removed using two alignment-based strategies, following the protocol of Brieuc *et al.* (2014). First, loci in the temporary database were aligned against themselves using BOWTIE (Langmead *et al.* 2009) by allowing a maximum of three nucleotide mismatches per locus. A locus that aligned to several loci, or a locus that did not align to itself was removed from the temporary database. Then, a BLAST search (Altschul *et al.* 1990) of the temporary database was conducted against itself. Loci that did not return a match, or loci where the best match were not themselves, were removed from the temporary database.

Using the updated temporary database of RAD loci, polymorphic duplicated loci were identified based on two haploid families. First, sequences from these haploid families were

aligned to the temporary database using BOWTIE, allowing a maximum of three nucleotide mismatches per locus. Sequences from the haploid individuals that aligned to more than one locus in the database could not be confidently relied upon in further analyses; they were thus identified as "blacklisted" loci and removed from the temporary database. Subsequently, polymorphic loci sequenced with a depth greater than 10X per haploid individual were identified using STACKS and retained for further screening. Among these polymorphic loci, a locus was identified as being putatively duplicated when more than one haploid offspring in a family was heterozygous at this particular locus (Brieuc *et al.* 2014). A final reference database comprising named duplicated and non-duplicated loci, as well as loci removed from the alignment-based screening steps and "blacklisted" loci, was created.

Genotyping of individuals in map crosses

Haploid individuals were genotyped at both non-duplicated and duplicated loci. Sequences from all haploid individuals were aligned to the non-duplicated and duplicated loci from the final reference database using BOWTIE by allowing a maximum of three nucleotide mismatches per locus. In haploids, we have shown reliable identification of single loci that have up to three SNPS - we have confirmed this result with genome mapping. To remain consistent, we used up to three mismatches so that we could differentiate between non-duplicated loci and duplicated loci. Both this study and a previous one (Brieuc *et al.* 2014) have shown that very reliable linkage results can be obtained in haploids using these criteria. Polymorphic loci sequenced with a depth greater than 10X per individual were identified using STACKS. Both non-duplicated and duplicated markers in the haploid families were used for mapping, described below and following the protocol of (Brieuc *et al.* 2014). Polymorphic duplicated loci were mapped when one of the paralogs was polymorphic (OPP - one paralog polymorphic, parental genotypes aa and ab, or aa and bc) or when both paralogs were polymorphic (BPP - parental genotypes ab and ac, or ab and cd; see table 1 in Brieuc *et al.* 2014).

Diploid individuals were only genotyped at non-duplicated loci. Sequences from all diploid individuals were aligned to the non-duplicated loci identified in the final reference database using BOWTIE by allowing a maximum of three nucleotide mismatches per locus. Subsequently, polymorphic loci were identified in each diploid family using STACKS, and genotypes at these loci were determined when alleles were sequenced with a depth greater than 10X per individual.

STACKS employs a maximum likelihood statistical model to identify sequence polymorphisms and determine individual genotypes (Catchen *et al.* 2011; Catchen *et al.* 2013). This approach can be biased towards heterozygous genotypes when sequence depths differ between the two alleles. To correct this bias against heterozygous genotypes, genotypes were corrected after running STACKS with the Python script developed by (Brieuc *et al.* 2014). Specifically, individuals were determined as heterozygotes at a locus if both alleles had a depth of more than two and the total read depth was 10X or greater.

Linkage mapping

Linkage maps in all haploid and diploid families were constructed using software for genetic mapping, ONEMAP 2.0-3 (Margarido *et al.* 2007), implemented in R version 3.0.2 (R Development Core Team 2013). Because coho salmon has 30 chromosome pairs (Phillips and Ráb 2001), each mapping family was expected to have at least 30 linkage groups. Linkage groups were named "Co", following the convention used in mapping studies in salmonids: this practice uses abbreviated common names for groups that are not yet anchored to chromosomes (e.g.Danzmann *et al.* 2005; Naish *et al.* 2013). RAD loci with 20% or less missing values among individuals within a family were employed for linkage analyses, and these loci were assigned to linkage groups in each family separately using a minimum log of odd ratio (LOD) score of 4.0 and a maximum recombination fraction of 0.25. The LOD score was subsequently increased by 1.0 until the number of linkage groups reached 30 or higher. An integrated haploid map was first constructed from the two haploid families using MERGEMAP (Wu *et al.* 2011) because genotypes at duplicated loci were only determined in these families. This integrated haploid map was later used to examine the distribution of duplicated loci across all linkage groups and identify linkage groups involved in recent or ongoing homeologous pairing.

Recombination rates in male salmonids tend to be smaller than those observed for females (Sakamoto *et al.* 2000; Ostberg *et al.* 2013), but these differences tend to decrease with high marker density and genome coverage (Rexroad *et al.* 2008; Lien *et al.* 2011). We used the female meiosis from the three diploid families to estimate marker order in these crosses. The data from all haploid and diploid female parents were then combined to calculate an integrated female haploid/diploid map using MERGEMAP.

Ordering markers in the diploid male map was computationally difficult, potentially due to reduced recombination and occasional tetrasomic inheritance in males (Allendorf and Danzmann 1997). Therefore, information from the integrated female map constructed with haploid and diploid mothers was used to infer the order and map the male meiosis in the three diploid families (Diploid Family 1, 2, 3). Polymorphic loci in common between the male parents and the integrated female map, as well as the Y-linked growth hormone pseudogene and sex-determining gene, sdY were grouped using a log of odd ratio (LOD) score of 4.0 and a maximum recombination fraction of 0.25 using ONEMAP. The LOD score was subsequently increased by 1.0 until the number of linkage groups reached 30 or higher. Grouped loci were then ordered based on the known order on the integrated female map using the make.seq and map functions implemented in ONEMAP. The position of a Y-linked growth hormone pseudogene and sex-determining gene, sdY on the male map was estimated in the two diploid families (Diploid Family 1 and 2) using the try.seq and map functions implemented in ONEMAP. The data from all diploid male parents were then combined to calculate an integrated male map using MERGEMAP.

Comparative mapping with Chinook salmon and comparison with other salmonid species
The reference database for coho salmon containing duplicated and non-duplicated RAD loci was aligned to the 54,937 filtered RAD loci identified in Chinook salmon (Brieuc *et al.* 2014) using BOWTIE, allowing no more than 3 nucleotide mismatches per locus. Homologies between Chinook and coho salmon were determined by examining the chromosomal arm locations of

shared loci between the two species. Putative centromere positions on coho linkage groups were estimated based on markers mapped in the gynogenetic diploid families in Chinook salmon (Brieuc *et al.* 2014). The order of mapped loci between the Chinook and coho salmon map was compared to determine if marker orders for chromosomes or chromosomal arms between the species were conserved. Finally, homologies identified between Chinook and coho salmon were used to infer homologies across coho salmon, rainbow trout and Atlantic salmon using molecular markers in common between published maps (Phillips *et al.* 2009; Lien *et al.* 2011; Miller *et al.* 2012; Brieuc *et al.* 2014).

Homeologous relationships and the distribution of duplicated loci across genomes
As we point out earlier, two categories of duplicated loci were identified in this study: where one of the paralogs was polymorphic (OPP) or both paralogs were polymorphic (BPP). Duplicated loci with both paralogs polymorphic (BPP) were used to infer homeologous linkage groups, since both paralogs could be mapped. The positions of duplicated loci were subsequently examined on the integrated haploid map to determine whether there was a bias in the distribution of these loci across linkage groups. A kernel smoothing approach using a sliding window of 2cM was used to determine whether there was a regional bias in distribution of these loci for each linkage group, following methods described in (Brieuc *et al.* 2014). Homeologous relationships detected in coho salmon were also compared to those identified in Chinook and Atlantic salmon (Lien *et al.* 2011; Brieuc *et al.* 2014).

Results

Reference database of RAD loci

A reference database comprising a unique set of RAD loci was created for the purpose of sequence alignment and identification of polymorphisms across individuals. A total of 70,037 loci were sequenced with a depth greater than 5X per individual in at least 496 individuals. These loci formed the temporary reference database of RAD loci, and they were retained for further screening. Sequence alignment using BOWTIE showed that 4075 loci did not align uniquely to themselves and likely corresponded with repeat regions; therefore these loci were removed from the temporary database. After performing the BLAST search of the temporary reference database against itself, 2085 loci did not return a match or the best match score was not the locus itself. It was possible that these loci contained repeat sequences; therefore, these loci were also removed from the temporary reference database. Sequences from the haploid individuals were aligned to the reference database using BOWTIE; 3706 loci from the haploid individuals aligned to several other loci, and these loci were thus blacklisted and removed from the reference database. Additionally, 7235 loci were identified as polymorphic duplicated loci in the haploid families. The final reference database comprising 52,936 non-duplicated loci and 7235 duplicated loci, as well as the 9866 loci that were removed by screening, are given in Supplementary Material S2.1.

Linkage mapping

An initial framework map was constructed with two haploid families. Haploid Family 1 and 2 had 3976 and 4048 biallelic polymorphic RAD loci respectively, comprising a total of

6652 unique RAD loci. Among these loci, a mixture of duplicated and non-duplicated loci (3897 loci in Haploid Family 1; 3996 loci in Haploid Family 2) were successfully assigned to 30 linkage groups with a LOD score of 5.0 to 7.0. The total map length for the Haploid Family 1 and 2 was 3040.1 cM and 3185.5 cM, respectively. The integrated haploid map had 5377 non-duplicated markers and 1266 duplicated markers with a total map length of 3602.6 cM (Supplementary Material S2.2).

Linkage analyses were conducted in the diploid families, following the construction of the integrated haploid map. Diploid Family 1, 2 and 3 had 1360, 1176 and 1931 biallelic non-duplicated loci that were polymorphic in each female parent, respectively. Among these loci, a set of loci (1214 loci in Diploid Family 1; 1138 loci in Diploid Family 2; 1765 loci in Diploid Family 3) were successfully assigned to 30 linkage groups with a LOD score of 4.0 to 8.0. The total map length for the Diploid Family 1, 2 and 3 was 3714.1cM, 3068.9cM and 5047.2cM, respectively. Although the Diploid Family 3 had the largest total map length, it also had the highest number of markers mapped. Since more recombination events are captured with more markers (Liu 1998), it is not surprising that the Diploid Family 3 had the largest total map length. Finally, data from the haploid parents and diploid female parents were combined; an integrated haploid/diploid female map measured 6596.7 cM, and it comprised 7415 non-duplicated markers and 1266 duplicated markers (Table 2.1; Figure 2.1; Supplementary Material S2.2).

The male meiosis was mapped using linkage analyses in the diploid families. Among the 8681 loci placed on the integrated haploid/diploid female map, Diploid Family 1, 2 and 3 had 846, 814 and 879 polymorphic loci in common for each male parent. Among these loci, a set of loci (792 loci in Diploid Family 1; 790 loci in Diploid Family 2; 851 loci in Diploid Family 3), as well as the Y-linked growth hormone pseudogene and sex-determining gene, sdY were successfully assigned to 30 linkage groups with a LOD score of 4.0 to 7.0. Both the growth hormone pseudogene and sdY mapped to the beginning of the linkage group, Co30. All linkage groups were successfully merged, except for Co22 which was split into two linkage groups (Co22_1 and Co22_2; Supplementary Material S2.3). The number of markers in common between the integrated male and female maps varied for each linkage group, ranging from 25 to 106 markers per linkage group (Supplementary Material S2.3). The male map had a total map length of 4141.76 cM (Supplementary Material S2.3).

The comparison between the male and female linkage groups reflected different recombination patterns between the sexes (Figure 2.2; Supplementary Material S2.4). Although telomeres were not mapped in males due to a lack of duplicated markers, many male linkage groups were expanded in size toward the terminal regions relative to the female, as seen by the increased distance in these regions reflecting more recombination events. Such patterns were particularly prominent for several linkage groups (Co02, Co04, Co05, Co08 - Co10, Co13 - Co15, Co17 - Co19, Co21 - Co29). Although qualitative, there was also evidence of suppressed recombination around the region containing the centromere in male compared to female integrated map for all linkage groups, as the male map had reduced distance in these regions compared to the female map.

Comparative mapping with Chinook salmon and comparisons with other salmonid species

We mapped 664 RAD loci in coho salmon that had been previously placed on the Chinook salmon map, which permitted the identification of homologous chromosomal arms between the two species (Table 2.1). On the basis of this comparison, we also identified the putative locations of centromeres within coho salmon linkage groups (Figure 2.1). Two homologous relationships between the species were inferred. An arm of the linkage group Co07a had 3 markers that mapped to Ots34 in Chinook salmon. On the other hand Co12b had 6 markers that mapped to Ots34 and 3 markers that mapped to Ots11p. Ots11p and Ots34 are likely involved in recent or ongoing homeologous pairing in Chinook salmon (Brieuc *et al.*, 2014), therefore it was not surprising that loci on Ots34 mapped to both homologous arms in coho salmon. In this case, we assumed that Co07a was homologous to Ots11p, and Co12b was homologous to Ots34 for reasons given in the discussion.

Comparative mapping with Chinook salmon permitted inference of the structure of coho linkage groups. Twenty linkage groups in coho salmon corresponded to putative bi-armed metacentric chromosomes, and 10 linkage groups corresponded to putative uni-armed acrocentric chromosomes (Figure 2.1). These inferred structures are in agreement with the known chromosome structures in coho salmon (Phillips and Ráb 2001). The short (p) arm of an acrocentric chromosome is usually uncharacterized in mapping studies because there are often insufficient markers describing this region (Brieuc *et al.* 2014). In this study, we identified the small arm for two putative acrocentric chromosomes (Co22, Co29) through comparative mapping with Chinook salmon.

Comparative mapping between the Chinook and coho salmon maps also provided information on chromosomal arrangements that are shared between the two species. Eighteen chromosomes are conserved between the species (Table 2.1); specifically, 9 metacentric chromosomes and 9 acrocentric chromosomes were conserved between the species. The remaining chromosome structures likely support independent Robertsonian rearrangements that occurred after descent from a common ancestor.

Five metacentric linkage groups in coho salmon (Co10 - Co14) consist of one acrocentric chromosome and one arm from a metacentric chromosome in Chinook salmon. Four metacentric linkage groups in coho salmon (Co15 - Co18) comprise arms that are found in two separate metacentric chromosome pairs in Chinook salmon. Two metacentric linkage groups (Co19, Co20) comprise two acrocentric chromosome pairs in Chinook salmon. Finally, one acrocentric linkage group (Co30) corresponds to an arm that is a part of a metacentric chromosome pair in Chinook salmon.

The orders of the RAD loci on the Chinook and coho salmon maps were compared across each linkage group or for each chromosome arm to determine whether any chromosomal inversions occurred following divergence between the species. There was a strong linear relationship among mapped loci for all the linkage groups or arms (Supplementary Material S2.5), suggesting that marker orders were conserved for all chromosomes or chromosomal arms. Such analyses provide additional evidence for the occurrence of centrometric inversion in Omy20 in rainbow trout following divergence between rainbow trout and Chinook/coho salmon,

and this chromosomal inversion may be exclusive to rainbow trout (Naish *et al.* 2013; Ostberg *et al.* 2013; Brieuc *et al.* 2014).

The homologies we observed between Chinook and coho salmon permitted alignment of coho linkage groups to those of rainbow trout and the Atlantic salmon, and the results are summarized in Table 2.1. Three acrocentric and 8 metacentric chromosomes were conserved among coho salmon, Chinook salmon and rainbow trout. Comparison between the *Oncorhynchus* species and the Atlantic salmon revealed that one metacentric and one acrocentric chromosomes were conserved across all compared species.

Homeologous relationships and the distribution of duplicated loci across linkage groups

The identification of linkage groups involved in homeologous pairing, as well as the localization of duplicated loci across individual linkage groups, was examined using the integrated haploid female map. A total of 1169 duplicated loci (1066 OPP and 103 BPP) were placed on this map. These loci were not distributed evenly among the linkage groups ($\chi 2$ test for uniform distribution across linkage groups, after correction for the number of markers per linkage group: pvalue ~ 0, df = 29); 87.0% of the duplicated loci were located on 16 linkage groups (Figure 2.3). There was a bias in distribution of these loci along the 16 linkage groups; duplicated loci were mostly found in the distal regions of all 16 linkage groups (Figure 2.4). Homeologies were identified between putative chromosome arms using marker pairs in which both paralogs were polymorphic (Table 2.2). All 8 homeologous arm pairs with a high retention of duplicated loci detected in coho salmon were also observed in Chinook salmon, and 4 homeologous arm pairs were conserved in Atlantic salmon (Table 2.2). All of these chromosome arms, likely involved in recent or ongoing homeologous pairing, involved at least one metacentric chromosome ancestral to the divergence between Pacific salmon species.

Discussion

Here, we aimed to examine the relationship between chromosomal evolution and retention of duplicated regions within the genus *Oncorhynchus*, and between this genus and *Salmo*, by deriving a linkage map for coho salmon, and comparing this map to that of Chinook salmon, rainbow trout and Atlantic salmon. Thirty linkage groups including 20 putative metacentric and 10 putative acrocentric chromosomes were described across two haploid and three diploid families. Chromosomal rearrangements were identified by comparing homologous arms between coho salmon, Chinook salmon, rainbow trout and Atlantic salmon. Results confirmed the conservation of at least one metacentric chromosome between *Oncorhynchus* and *Salmo* (Co09), seven metacentric chromosomes across the genus *Oncorhynchus* (Co01 - Co07; Naish *et al.* 2013; Phillips *et al.* 2013; Brieuc *et al.* 2014), and detected a polymorphism in another across coho and Chinook salmon and rainbow trout (Co14 and Co15). Another metacentric chromosome was detected as ancestral to coho and Chinook salmon only (Co08). The placement of 1166 duplicated loci on the consensus haploid map of 7415 markers revealed that these loci were not evenly distributed across all linkage groups, supporting an earlier finding in Chinook salmon (Brieuc *et al.* 2014); namely, that homeologous pairs diverged from each other at different rates following the whole genome duplication event. Regions of the genome with polymorphic duplicated markers were found on the same eight pairs of homologous

chromosome arms (16 arms in total) across coho and Chinook salmon. Each of the eight pairs of chromosomes likely involved in ongoing or recent homeologous pairing included at least one of the ancestral metacentric chromosomes that are conserved between the two species. The other chromosome arm involved in the pairing may be part of either an acrocentric chromosome or metacentric chromosome. The data suggest that Robertsonian rearrangements that result in metacentric chromosome formation prior to the diversification of homeologous pairs might partly explain the uneven retention of duplicated regions across the genome, at least within Pacific salmon.

The consensus male map, constructed with three diploid families, was significantly smaller (4141.7 cM) than the consensus female map (6596.7 cM) constructed with two haploid families and three diploid families. There are three main reasons that might explain the reduced map size in male compared to the female map. First, the difference could simply be a function of more markers being placed on the female map (8581 for the female and 2041 for the male), as map size tends to increase when more markers are added (Liu 1998). Second, some male linkage groups only represented a portion of those in female; for example, in three metacentric chromosomes (Co07, Co09, Co13) only one arm and a region containing the centromere were mapped, the duplicated markers were not mapped in male. Third, recombination in males is suppressed relative to females, and male maps in salmonids often tend to be smaller. Comparisons between the consensus male map and the consensus (haploid and diploid) female map indicated that recombination in the male was suppressed around the region containing the centromere in some linkage groups while recombination in female seemed suppressed toward telomeric regions.

Accurate identification of chromosome structure in coho salmon relied on aligning homologous chromosome arms with Chinook salmon. In addition, regions containing the centromere were inferred through comparative mapping between coho and Chinook salmon as no gynogenetic diploid families were employed for this study to identify the exact location of the centromere. Homology of Co07a and Co12b with Chinook salmon chromosome arms Ots11p and Ots34 was not completely resolved. These arms are homeologous to each other within both species, and markers on Ots34 mapped to both Co07a and Co12b in coho salmon. We inferred that Co07a was homologous to Ots11p, as the arm Co07a is part of a metacentric chromosome that is conserved in Chinook salmon (Ots11) and rainbow trout (Omy19, Table 2.1). In fact, (Naish *et al.* 2013) speculate that these are conserved across the genus because Chinook salmon and rainbow trout are distantly related. In contrast, the metacentric chromosome Co12 is not conserved in Chinook salmon and rainbow trout. We assumed these structures in the subsequent discussion.

Coho salmon map coverage, size and differences in sex-specific recombination

The coho salmon linkage map constructed in this study has 8681 markers, spanning all predicted 30 linkage groups. This coverage is comparable to recently published maps across a number of salmon species (Lien *et al.* 2011; Everett *et al.* 2012; Miller *et al.* 2012; Brieuc *et al.* 2014). We observed different map sizes in the consensus female maps; the map constructed with combined haploid and diploid families had a size of 6596.7 cM, which is significantly larger than the coho map created with haploids alone (3602.6 cM). There are several reasons that might

explain the differences such as nonrandom missing values (Jorgenson *et al.* 2005), genotyping errors (Hackett and Broadfoot 2003) and numbers of markers mapped. Potential bias against heterozygotes in RAD sequencing (Brieuc *et al.* 2014) may also partly explain the inflated map distances in the haploid/diploid map, especially since the size of the map created with only haploid families in this study was much smaller.

Male salmonids are the heterogametic sex (Allendorf and Thorgaard 1984). In this study, both the Y-linked growth hormone pseudogene and sex-determining gene, sdY mapped to the telomeric region on the acrocentric chromosome, Co30. This finding is in agreement with earlier findings (Phillips *et al.* 2005); that the sex chromosome in this species is acrocentric and the SEX locus was located on the telomeric region of this chromosome. Mapping has shown that the sex chromosome is not conserved across the species, and that a small male-specific region including the sex determining gene has been repeatedly transposed to different chromosomes in different salmon species (Phillips *et al.* 2001; Woram *et al.* 2003; Yano *et al.* 2012).

Our results showed that suppressed recombination around the region containing the centromere in male was widely apparent, while higher recombination was observed in telomeric regions for some male linkage groups relative to female. The results are in agreement with a number of studies performed on other salmonid species (Sakamoto *et al.* 2000; Moen *et al.* 2008; Rexroad *et al.* 2008; Lien *et al.* 2011). Male recombination rate in telomeric regions of certain rainbow trout linkage groups has been shown to be higher than that of female, but lower in centromeric regions (Sakamoto *et al.* 2000). Such different recombination patterns might be in part explained by occasional multivalent formation during male meiosis, in which crossovers between homeologous chromosomes are increased in the telomeric regions while crossovers between homologous chromosomes are hindered in the centromeric regions through structural constraints (Sakamoto *et al.* 2000). In our study, however, suppressed recombination around the centromere and increased recombination in telomeric regions were apparent for many male linkage groups including the ones not likely involved in homeologous pairing. Some studies have also found notable clustering of markers in centromeric regions for many male linkage groups (Nichols *et al.* 2003; Lien *et al.* 2011; Miller *et al.* 2012). In Atlantic salmon, high marker densities were also involved in regions close to the centromere for male linkage groups with a lower frequency of duplicated markers (Lien *et al.* 2011). Recombination rates are known to differ between the sexes in a wide range of species (Lenormand and Dutheil 2005); while homeologous chromosome pairing during male meiosis could certainly account for some of the differences, the origin of the sex differences observed in this study still remains unclear.

Comparative genome mapping

Comparative mapping provided insights into the process of chromosomal evolution occurring after the whole genome duplication event, and this is the first study that characterizes chromosomal evolution between coho and Chinook salmon. Nine metacentric and nine acrocentric chromosomes appear to be conserved between these two species. Among these conserved chromosomes, one metacentric (Co08 in coho and Ots12 in Chinook) and five acrocentric chromosomes are unique to coho and Chinook salmon, suggesting that the structure of these linkage groups is ancestral to the divergence of these species relative to rainbow trout

and Atlantic salmon. The remaining linkage groups are not conserved, reflecting the chromosomal rearrangements since coho and Chinook salmon diverged.

Syntenic relationships between the Chinook salmon and rainbow trout maps permitted comparisons across the genus *Oncorhynchus* (Naish *et al.* 2013; Phillips *et al.* 2013). There are four acrocentric chromosomes that are conserved across all three *Oncorhynchus* species (Co26 - Co29). Similarly, eight metacentric chromosomes are conserved among all three species (Co01 - Co07, Co09). The results support the hypothesis of (Naish *et al.* 2013) that the arm rearrangements that resulted in these metacentric chromosomes are ancestral to the divergence of the species and could be conserved across the genus *Oncorhynchus*. There is one interesting extension to these earlier observations. One metacentric chromosome in Chinook salmon, Ots08, sometimes occurs as a metacentric chromosome (Omy25p and q) or as two acrocentric arms (Omy25 and Omy 29) in rainbow trout (Figure 2.5). Here, we found that the homologous arms in coho salmon occur in two separate unrelated metacentric linkage groups (Co14a and Co15a; Figure 2.5). In Atlantic salmon, these two arms are fused together along with a third arm to form a large acrocentric chromosome (Figure 2.5). Robertsonian fusions are common, and can also form acrocentric chromosomes: this outcome is likely more frequent in Atlantic salmon than in Pacific salmon (Phillips and Ráb 2001). Taken together, the configurations of these particular chromosomes suggest they may have undergone recurrent fusions and fissions across species. One metacentric and one acrocentric chromosome, Co09 and Co26, respectively, appear to be conserved across coho salmon, Chinook salmon, rainbow trout and Atlantic salmon (Phillips *et al.* 2009; Brieuc *et al.* 2014). Our study supports the earlier findings that this metacentric chromosome is likely ancestral to the divergence of the two genera, *Salmo* and *Oncorhynchus*. It also appears that this chromosome is conserved in Arctic charr and Brook charr within the genus Salvelinus (Timusk *et al.* 2011), both of which share a more recent common ancestor with *Oncorhynchus*. In addition, earlier results provided evidence for the occurrence of centrometric inversion in Omy20 in rainbow trout following divergence between rainbow trout and Chinook/coho salmon, and this chromosomal inversion may be exclusive to rainbow trout (Naish *et al.* 2013; Ostberg *et al.* 2013; Brieuc *et al.* 2014). In the current study, marker orders are fully conserved between coho and Chinook salmon chromosome arms, Co25 and Ots25, respectively, further supporting this earlier observation (Supplementary Material S2.5).

Conservation of reduced divergence between homeologous chromosomes across species
While our results confirm that the divergence rates of homeologs following the whole genome duplication event have not been uniform (Brieuc *et al.* 2014), the key finding of this study is that the ancestral metacentric chromosomes retain recently diverged duplicates and are the ones likely involved in recent or ongoing homeologous pairing (Co01 - Co07, conserved among all three *Oncorhynchus* species; Co09, conserved among all four species). Such findings suggest that homeologies may be preferentially retained between larger metacentric chromosomes (Phillips *et al.* 2009), and the involvement of at least one metacentric chromosome provides the stability required for the formation of multivalents (Wright *et al.* 1980; Wright *et al.* 1983; Brieuc *et al.* 2014). These results support the hypothesis of Phillips *et al.* (2009), which suggested that diploidization of chromosomes not involved in homeologous parings may have occurred in the ancestral salmonid before the divergence between *Salmo* and *Oncorhynchus*. We speculate that this process also differed to some extent following the divergence of the two genera. Although the exact distribution of duplicated markers along chromosome arms in

Atlantic salmon has not been described (Lien *et al.* 2011), only four out of eight homeologous pairings appear to share polymorphic duplicated loci between the Atlantic salmon and Chinook and coho salmon grouping (Table 2.2).

The implications of our findings for species divergence within the subfamily Salmoninae will become clearer once we gain a greater understanding of the role of duplicated regions in evolution. If the duplicated regions we detected have genes that permit greater flexibility for adaptation by providing the opportunity to acquire additional or novel functions (Soltis and Soltis 2000; Koop and Davidson 2008; Van de Peer *et al.* 2009; Feldman *et al.* 2012; Alexandrou *et al.* 2013; Macqueen and Johnston 2014), then retention of particular duplicated regions within certain lineages may explain their subsequent innovation and diversification. On the other hand, the physical formation of metacentric chromosomes may inhibit diploidization because of ongoing recombination; such chromosomes may continue to exhibit tetrasomic inheritance, thus becoming "evolutionary dead ends". Whole genome sequencing of duplicated chromosome arms in rainbow points towards the fact that duplicated protein coding loci have simply become lost through gradual change (Berthelot *et al.* 2014); the conserved metacentric chromosomes may continue exhibiting tetrasomic inheritance and prevent functional divergence of protein coding regions. In this study, we provide preliminary evidence that the evolutionary timing of metacentric chromosome formation varied, which might have impacted the rate of diploidization across different lineages. As comparative genome sequencing of salmon species continues, comparing the rates of differentiation along certain chromosome arms between species and identifying the location of genes involved in diversification will provide insights into the role of WGD in salmon evolution. Here, we have identified chromosome arms of interest for further efforts addressing such questions.

Here, we developed an extensive set of genomic resources for coho salmon: a reference database of unique RAD loci, two types of consensus female linkage maps and a consensus male linkage map. A dense female map constructed in this study permitted alignment of linkage groups in this species with that of Chinook salmon, enabling interspecies comparisons with related salmon species. Syntenic relationships across multiple salmonid species identified in this study provided strong evidence for chromosomal rearrangements and conservation of metacentric and acrocentric chromosomes following the divergence between *Salmo* and *Oncorhynchus*. We have also identified linkage groups that have been recently or may be involved in ongoing homeologous pairing in coho salmon. Such pairings were conserved with related Pacific salmon species. Ancestral metacentric chromosomes appear to retain recently diverged duplicated regions and be potentially involved in homeologous pairings; such results indicate that diploidization may have been prevented or retarded in these ancestral metacentric chromosomes following the whole genome duplication event. The resources developed here will facilitate genome-wide studies in coho salmon, such as genome scans, QTL mapping and genome-wide association studies (Naish and Hard 2008), as well as provide resources for studies concerning ecology and evolution in related salmon species.

Acknowledgements

We thank Isadora Jimenez-Hidalgo for assistance with genomic data processing. Dan Drinan provided thoughtful discussions and helpful suggestions. Initial experimental lines were obtained from SweetSpring Salmon / Aquaseed Corporation, suppliers of the Domsea

broodstock, and we particularly thank Per Heggelund, Greg Hudson and Patty Munsell for their help. We also thank Linda Park and Jim Myers for help in setting up the initial crosses and David Rose for maintaining the lines. Finally, we are grateful to two anonymous referees and the Associate Editor for their very helpful comments on the first draft of the manuscript. Funding for this study was provided by NOAA Fisheries/Federal Columbia River Power System (FCRPS) Biological Opinion Remand Funds to KAN and JJH, School of Aquatic and Fishery Sciences and Graduate Opportunities and Minority Achievement Program (GO-MAP, University of Washington) award to MK.

Figures

Figure 2.1 - Graphical representation of 30 consensus linkage groups in haploid and diploid female coho salmon. Co01 to Co20 are metacentric, and Co21 to Co30 are acrocentric, inferred from comparative mapping. The size of linkage groups ranges from 127.18 to 295.21 centiMorgan (Kosambi), and each line corresponds to the location of one or more markers. The putative location of the centromere, estimated by comparative mapping with Chinook salmon, is represented in red.

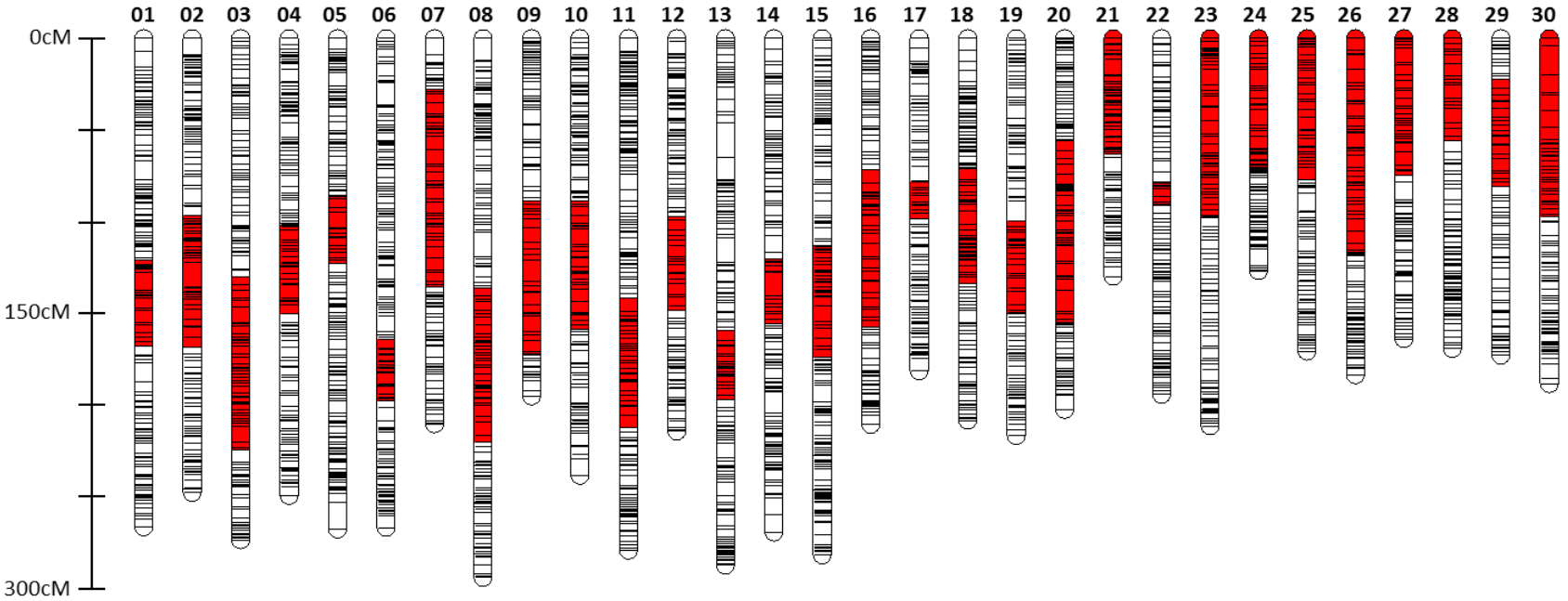


Figure 2.2 - Comparison of map distances between common markers mapped in male and female coho salmon. Linkage groups, Co04 (a), Co17 (b), Co18 (c) and Co26 (d) are given as an examples. The putative region containing the centromere is represented by the cross-hatched area. All linkage groups are presented in Supplementary Material S2.4.

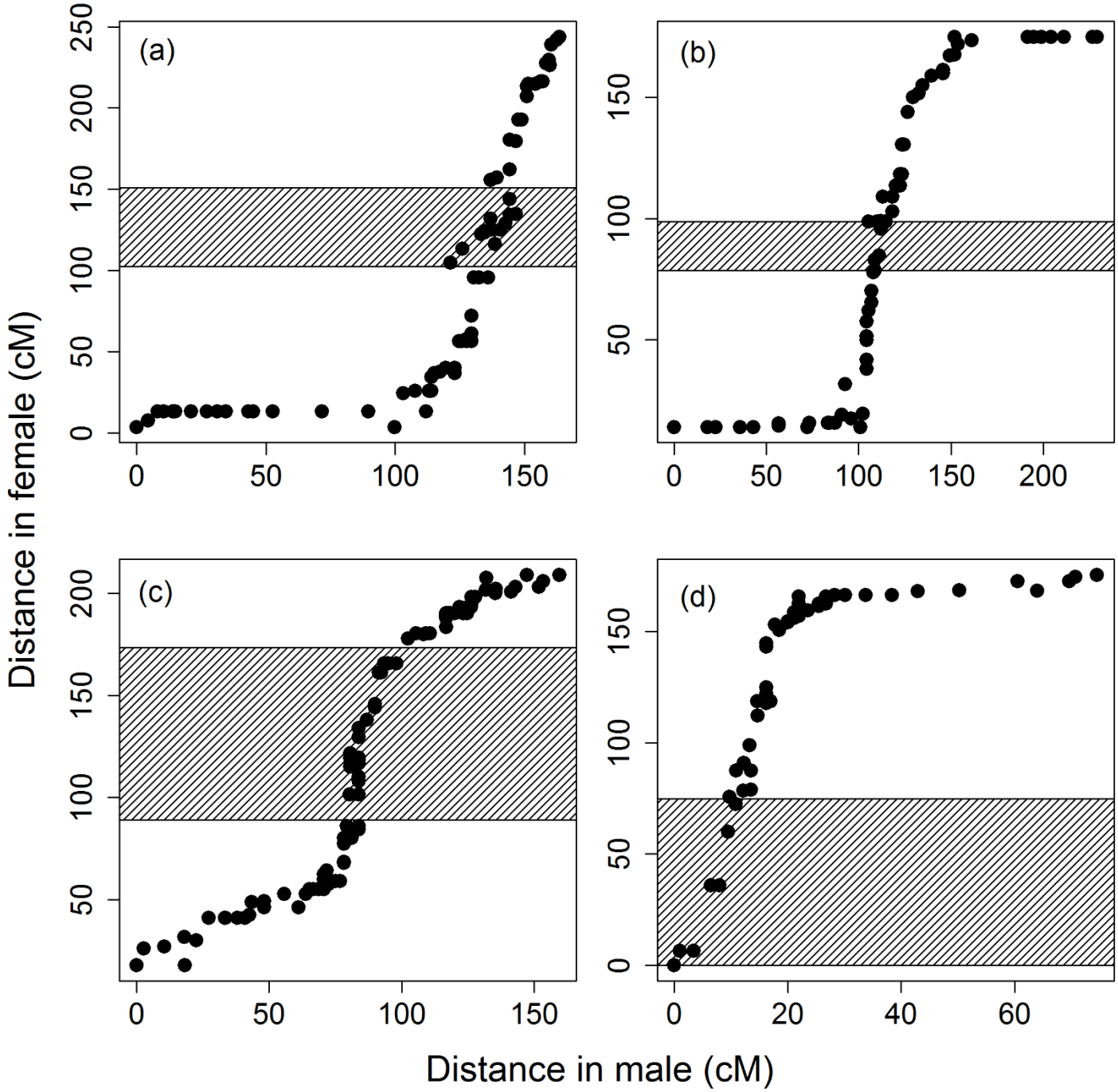


Figure 2.3 - Number of markers and distribution of duplicated markers across each coho salmon linkage groups. Non-duplicated loci are represented in light grey. Duplicated loci are represented in dark grey (loci with only one paralog polymorphic) or black (both paralogs polymorphic).

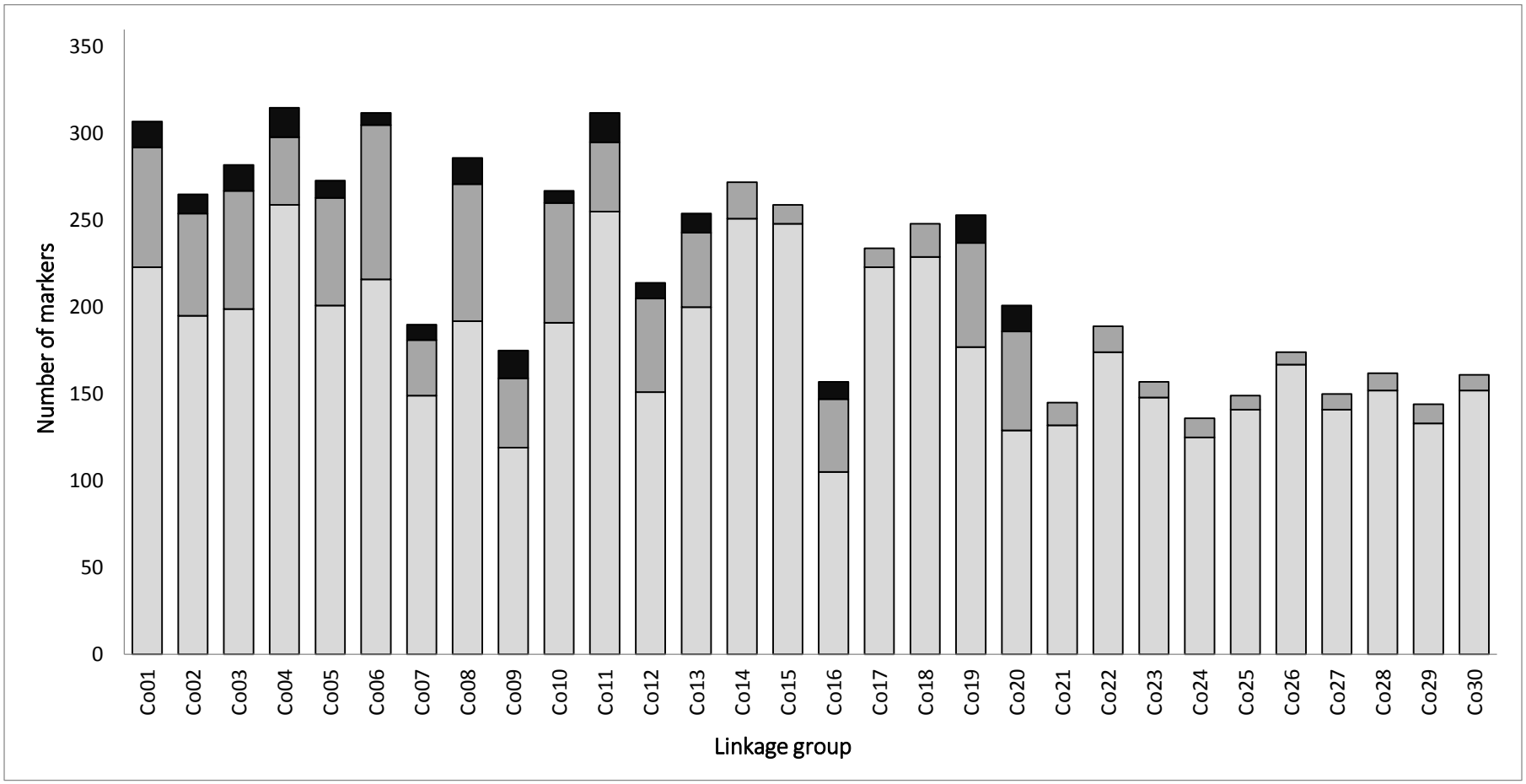


Figure 2.4 - Distribution of duplicated and non-duplicated loci along the 16 linkage groups with a high proportion of duplicated loci. Non-duplicated loci are represented in light grey. Duplicated loci are represented in dark grey (loci with one paralog polymorphic) or in charcoal (loci with both paralogs polymorphic). The putative region containing the centromere is represented by the cross-hatched area.

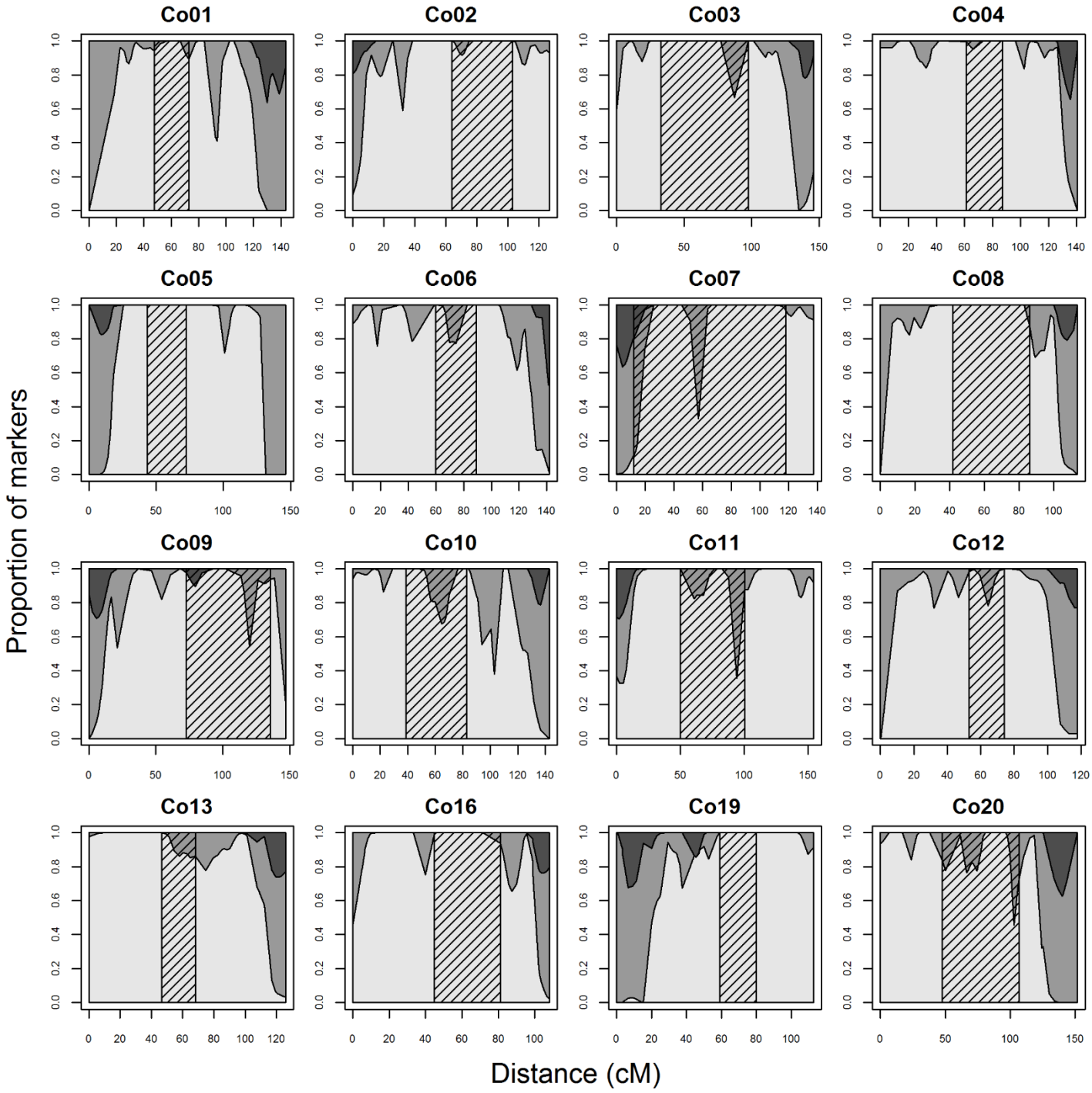
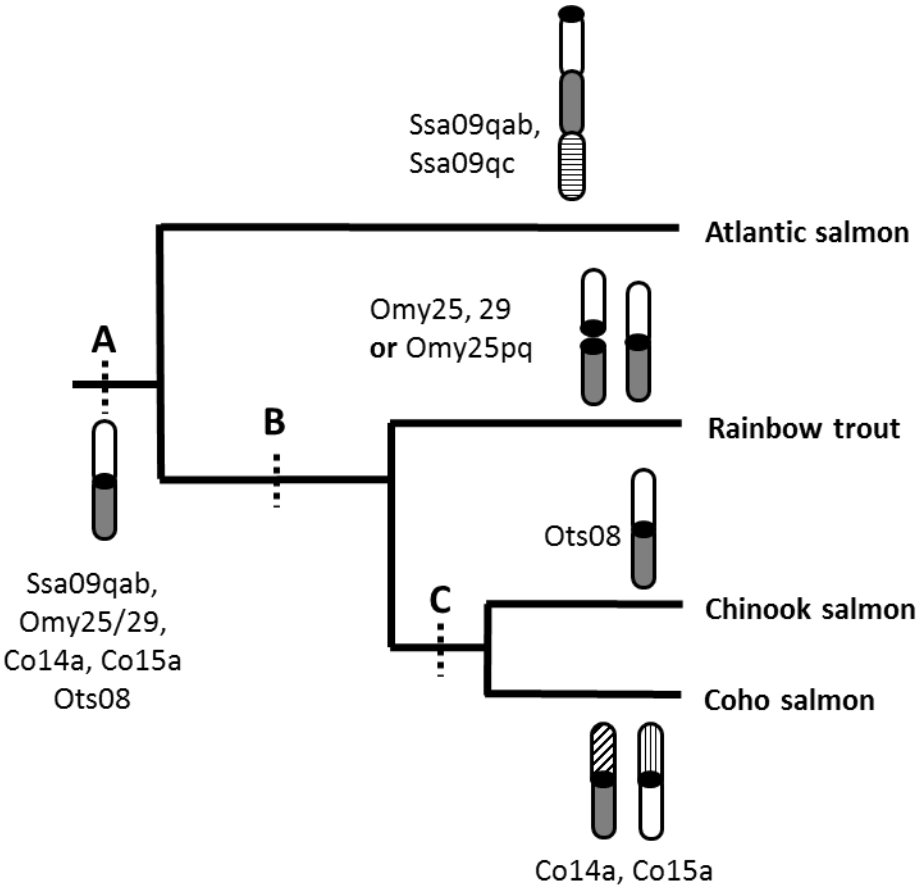


Figure 2.5 - Phylogenetic tree showing the orientation of homologous arms, Ssa09qab, Omy25/29, Co14a, Co15a and Ots08 in Atlantic salmon, rainbow trout, coho salmon and Chinook salmon, respectively. Chromosomal rearrangements and homeologous relationships conserved across species at phylogenetic nodes A, B, C are summarized in Table 2.1 and 2.2.



Tables

Table 2.1 - Description of the coho salmon consensus linkage map constructed with haploid and diploid female parents, and comparison with chromosome arms of Chinook salmon, rainbow trout and Atlantic salmon. Linkage groups (Co) were randomly assigned numbers, and arm names are given as "a" and "b". Homologous arms in Chinook salmon, rainbow trout and Atlantic salmon are based on chromosome names for each species (Ots, Omy and Ssa respectively), with known orientations (p is the short arm, q the long arm).
* and ** denote inferred relationship; there were no markers in common between Co07a and Ots11p(*), and there are markers in common between Co12b and Ots34/Ots11p (**). *** denotes a chromosomal arm that is composed entirely of ribosomal DNA. The final column designates chromosomal rearrangements conserved across species; letter corresponds to phylogenetic placement in Figure 2.5.
[1] indicates incompletely resolved relationships between Atlantic salmon and rainbow trout according to published studies.

Coho linkage group	Size (cM)	Number of markers	Coho linkage arms	Chinook chromosome (Philips *et al.*, 2013 ; Brieuc *et al.*, 2014)	Rainbow trout chromosome (Philips *et al.*, 2009; Brieuc *et al.*, 2014)	Atlantic salmon chromosome (Danzmann *et al.*, 2008; Philips *et al.*, 2009)	Chromosomal rearrangement conserved across species
Co01	267.51	393	Co01a	Ots02p	Omy17p	Ssa02q	B, metacentric
			Co01b	Ots02q	Omy17q	Ssa12qb[1]	
Co02	248.6	344	Co02a	Ots03p	Omy03p	Ssa02p	B, metacentric
			Co02b	Ots03q	Omy03q	Ssa25	
Co03	274.51	360	Co03a	Ots04p	Omy06p	Ssa24	B, metacentric
			Co03b	Ots04q	Omy06q	Ssa26	
Co04	250.26	395	Co04a	Ots06p	Omy01p	Ssa16qa	B, metacentric
			Co04b	Ots06q	Omy01q	Ssa18qa	
Co05	268.59	376	Co05a	Ots07p	Omy07p	Ssa17qb	B, metacentric
			Co05b	Ots07q	Omy07q	Ssa22	
Co06	267.8	415	Co06a	Ots09p	Omy12p	Ssa13qb	B, metacentric
			Co06b	Ots09q	Omy12q	Ssa03q	
Co07	211.09	241	Co07a	Ots11p*	Omy19p	Ssa04p	B, metacentric
			Co07b	Ots11q	Omy19q	Ssa01p	

Co08	295.21	382	Co08a_1	Ots12p	Omy11p	Ssa20qa	C, metacentric
			Co08a_2		Omy11q		
			Co08b	Ots12q	Omy26	Ssa11qb	
Co09	196.03	207	Co09a	Ots15p	Omy21p	Ssa07p	A, metacentric
			Co09b	Ots15q	Omy21q	Ssa07q	
Co10	239.27	346	Co10a	Ots01p	Omy04p	Ssa23	-
			Co10b	Ots27	Omy13q	Ssa06q	
Co11	280.26	419	Co11a	Ots01q	Omy23	Ssa01qa	-
			Co11b	Ots29	Omy15p	Ssa29	
Co12	215.12	286	Co12a	Ots05p	Omy08p	Ssa15qa	-
			Co12b	Ots34**	Omy10q	Ssa08q	
Co13	288.32	350	Co13a	Ots05q	Omy05q	Ssa10qa	-
			Co13b	Ots23	Omy02p	Ssa05q	
Co14	270.54	356	Co14a	Ots08p	Omy25p	Ssa09qa	-
			Co14b	Ots31	Omy14p	Ssa14qb	
Co15	282.89	355	Co15a	Ots08q	Omy25q (Omy29)	Ssa09qb	-
			Co15b	Ots13q	Omy27	Ssa20qb	
Co16	211.31	220	Co16a	Ots10p	Omy09p	Ssa18qb	-
			Co16b	Ots14p	Omy18p	Ssa16qb	
Co17	181.92	292	Co17a	Ots13p	Omy18q	Ssa27	-
			Co17b	Ots16q	Omy09q	Ssa15qb	
Co18	209.02	323	Co18a	Ots14q	Omy24	Ssa09qc	-
			Co18b	Ots16p	Omy11p	Ssa19qa	
Co19	217.65	315	Co19a	Ots17	Omy15q	Ssa17qa	-
			Co19b	Ots21	Omy14q	Ssa05p	
Co20	203.33	247	Co20a	Ots24	Omy16p	Ssa19qb	-
			Co20b	Ots32	Omy13p	Ssa12qa[1]	
Co21	130.25	184	Co21	Ots18	Omy04q	Ssa06p	C, acrocentric

Co22	194.97	243	Co22	Ots19	Omy02q	Ssa10qb	C, acrocentric
Co23	212.27	213	Co23	Ots20	Omy05p	Ssa01qb	C, acrocentric
Co24	127.18	187	Co24	Ots22	Omy16q	Ssa13qa	C, acrocentric
Co25	171.35	196	Co25_1	Ots25	Omy20p	Ssa08p***	A, acrocentric
			Co25_2		Omy20q	Ssa28	
Co26	184.21	225	Co26	Ots26	Omy22	Ssa21	A, acrocentric
Co27	164.56	198	Co27	Ots28	Omy28	Ssa03p	B, acrocentric
Co28	169.82	214	Co28	Ots30	Omy10p	Ssa04q	C, acrocentric
Co29	173.74	181	Co29	Ots33	OmySex	Ssa11qa	B, acrocentric
Co30	189.15	218	Co30	Ots10q	Omy08q	Ssa14qa	-
Total	6596.7	8681					

Table 2.2 - Homeologous chromosome arm pairs identified in coho salmon and the number of marker pairs supporting the homeologous relationship. Corresponding known homeologous relationships in Chinook salmon, rainbow trout and Atlantic salmon are shown. The final column designates known conservation of metacentric chromosome with high frequency of duplicated markers; letter corresponds to phylogenetic placement in Figure 2.5. Star (*) denotes possible earlier chromosomal arrangement (A) and subsequent rearrangement.

^: homeologous relationships with little or no support in Atlantic salmon (Lien *et al.* 2011)

+: we have corrected the homeologous relationship between Omy17q and Omy13p; evidence suggests that this relationship was incorrectly reported as being between Omy17p and Omy13p in previous studies

Homeology in coho salmon	Number of marker pairs supporting homeolog pairings	Homeology in chinook salmon (Brieuc *et al.*, 2014)	Homeology in rainbow trout (Phillips *et al.*, 2006)	Homeology in Atlantic salmon (Lien *et al.*, 2011)	Known phylogenetic placement of Metacentric arrangement
Co01b - Co20b	15	Ots02q - Ots32	Omy17q+- Omy13p	Ssa02q - Ssa12qa^	B
Co02a - Co13b	11	Ots03p - Ots23	Omy03p - Omy02p	Ssa02p - Ssa05q	B*
Co03b - Co08b	15	Ots04q - Ots12q	Omy06q - Omy26	Ssa26 - Ssa11qa^	B, C
Co04b - Co11a	17	Ots06q - Ots01q	Omy01q - Omy23	Ssa18qa - Ssa01qa^	B
Co05a - Co16b	10	Ots07p - Ots14p	Omy07p - Omy18p	Ssa17qa - Ssa16qb^	B
Co06b - Co10b	7	Ots09q - Ots27	Omy12q - Omy13q	Ssa03q - Ssa06p	B*
Co07a - Co12b	9	Ots11p - Ots34	Omy19p - Omy10q	Ssa04p - Ssa08q	B*
Co09a - Co19a	16	Ots15p - Ots17	Omy21p - Omy15q	Ssa07p - Ssa17qa	A

Supplementary Material

S2.1 - Reference database of RAD loci for coho salmon

S2.1.1 - Non-duplicated RAD loci in the reference database and corresponding sequences. Mapped markers have chromosome arm added to the name (CoXXXXX_CoXX[a or b]). Markers falling in the putative location of the centromere, estimated by comparative mapping with Chinook salmon, have the linkage group followed by u (CoXXXXX_CoXXu) where u designates arm placement unknown.

S2.1.2 - Duplicated RAD loci in the reference baseline and corresponding sequences. The letter D was added after the marker name. Mapped markers have chromosome arm added to the name (CoXXXXX_CoXX[a or b]). Markers falling in the putative location of the centromere, estimated by comparative mapping with Chinook salmon, have the linkage group followed by u (CoXXXXX_CoXXu).

S2.1.3 - RAD loci removed from the temporary reference database during screening and sequences.

S2.2 - Linkage maps

S2.2.1 - The consensus female map constructed with two haploid families; marker distances are shown in centiMorgan (Kosambi distance).

S2.2.2 - The consensus female map constructed with two haploid and three diploid families; details given are linkage groups, marker distances in centiMorgan (Kosambi), assignment of the marker to the Chinook salmon female map, Chinook salmon chromosomes and its position on the Chinook map.

S2.2.3 - The consensus male map constructed with three diploid families; details given are linkage groups, marker distances in centiMorgan (Kosambi) and corresponding marker position on the consensus female map constructed with haploid and diploid female parents.

S2.3 - Difference in sizes between integrated female and male coho salmon linkage groups. The map lengths are presented in centiMorgan (Kosambi).

S2.4 - Relationship among positions of mapped RAD loci in common between the sexes. Presumed regions containing the centromere are represented by the cross-hatched area. The consensus female map constructed with haploid and diploid families (S2.2.2) was used for this comparison.

S2.5 - Relationship among mapped RAD loci in coho and Chinook salmon for all the linkage groups or arms.

References

Alexandrou, M. A., B. A. Swartz, N. J. Matzke and T. H. Oakley, 2013 Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. Mol. Phylogen. Evol. 69**:** 514-523.

Allendorf, F. W., and R. G. Danzmann, 1997 Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. Genetics. 145**:** 1083-1092.

Allendorf, F. W., and G. H. Thorgaard, 1984 Tetraploidy and the evolution of salmonid fishes, pp. 1-27 in *Evolutionary Genetics of Fishes*, edited by B. J. Turner. Plenum Press, New York.

Altschul, S. F., W. R. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. 215**:** 403-410.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS. One. 3**:** e3376.

Berthelot, C., F. Brunet, D. Chalopin, A. Juanchich, M. Bernard *et al.*, 2014 The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5**:** 3657.

Bourret, V., M. P. Kent, C. R. Primmer, A. Vasemagi, S. Karlsson *et al.*, 2013 SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Mol. Ecol. 22**:** 532-551.

Brieuc, M. S. O., C. D. Waters, J. E. Seeb and K. A. Naish, 2014 A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral Whole Genome Duplication event. G3-Genes. Genom. Genet. 4**:** 447-460.

Cañestro, C., R. Albalat, M. Irimia and J. Garcia-Fernandez, 2013 Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. Semin. Cell Dev. Biol. 24**:** 83-94.

Catchen, J., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait, 2011 STACKS: building and genotyping loci de novo from short-read sequences. G3-Genes. Genom. Genet. 1**:** 171-182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores and W. A. Cresko, 2013 STACKS: an analysis tool set for population genomics. Mol. Ecol. 22**:** 3124-3140.

Crête-Lafrenière, A., L. K. Weir and L. Bernatchez, 2012 Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. PloS. One. 7**:** e46662.

Crow, K. D., and G. P. Wagner, 2006 What is the role of genome duplication in the evolution of complexity and diversity? Mol. Biol. Evol. 23**:** 887-892.

Danzmann, R. G., M. Cairney, W. S. Davidson, M. M. Ferguson, K. Gharbi *et al.*, 2005 A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily : Salmoninae). Genome. 48**:** 1037-1051.

Davidson, W. S., B. F. Koop, S. J. M. Jones, P. Iturra, R. Vidal *et al.*, 2010 Sequencing the genome of the Atlantic salmon (*Salmo salar*). Genome. Biol. 11.

Devlin, R. H., C. a. Biagi and D. E. Smailus, 2001 Genetic mapping of Y-chromosomal DNA markers in Pacific salmon. Genetica. 111**:** 43-58.

Edger, P. P., and J. C. Pires, 2009 Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome. Res. 17**:** 699-717.

Everett, M. V., M. R. Miller and J. E. Seeb, 2012 Meiotic maps of sockeye salmon derived from massively parallel DNA sequencing. BMC. Genomics. 13.

Feldman, M., A. A. Levy, T. Fahima and A. Korol, 2012 Genomic asymmetry in allopolyploid plants: wheat as a model. J. Exp. Bot. 63**:** 5045-5059.

Guyomard, R., M. Boussaha, F. Krieg, C. Hervet and E. Quillet, 2012 A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. BMC. Genet. 13**:** 15.

Hackett, C. A., and L. B. Broadfoot, 2003 Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity. 90**:** 33-38.

Hufton, A. L., and G. Panopoulou, 2009 Polyploidy and genome restructuring: a variety of outcomes. Curr. Opin. Genet. Dev. 19**:** 600-606.

Hurley, I. A., R. L. Mueller, K. A. Dunn, E. J. Schmidt, M. Friedman *et al.*, 2007 A new time-scale for ray-finned fish evolution. Proc. R. Soc. B. 274**:** 489-498.

Jaillon, O., J. M. Aury, F. Brunet, J. L. Petit, N. Stange-Thomann *et al.*, 2004 Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 431**:** 946-957.

Jaillon, O., J. M. Aury and P. Wincker, 2009 "Changing by doubling", the impact of whole genome duplications in the evolution of eukaryotes. C. R. Biol. 332**:** 241-253.

Jorgenson, E., H. Tang, M. Gadde, M. Province, M. Leppert *et al.*, 2005 Ethnicity and human genetic linkage maps. Am. J. Hum. Genet. 76**:** 276-290.

Koop, B. F., and W. S. Davidson, 2008 Genomics and the genome duplication in salmonids, pp. 77-86 in *Fisheries for Global Welfare and Environment, 5th World Fisheries Congress*.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome. Biol. 10**:** R25.

Larson, W. A., L. W. Seeb, M. V. Everett, R. K. Waples, W. D. Templin *et al.*, 2013 Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol. Appl. 7**:** 355-369.

Lenormand, T., and J. Dutheil, 2005 Recombination difference between sexes: A role for haploid selection. PLoS Biol. 3**:** 396-403.

Lien, S., L. Gidskehaug, T. Moen, B. J. Hayes, P. R. Berg *et al.*, 2011 A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. BMC. Genomics. 12**:** 615.

Liu, B. H., 1998 *Statistical Genomics: linkage, mapping, and QTL analysis*. CRC Press, New York.

Lubieniecki, K. P., S. L. Jones, E. A. Davidson, J. Park, B. F. Koop *et al.*, 2010 Comparative genomic analysis of Atlantic salmon, *Salmo salar*, from Europe and North America. BMC. Genet. 11**:** 105.

Lynch, M., 2007 *The Origins of Genome Architecture*. Sinauer Associates, Sunderland, MA.

Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science. 290**:** 1151-1155.

Macqueen, D. J., and I. A. Johnston, 2014 A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc. R. Soc. B. 281.

Margarido, G. R. A., A. P. Souza and A. A. F. Garcia, 2007 OneMap: software for genetic mapping in outcrossing species. Hereditas. 144**:** 78-79.

Mayfield-Jones, D., J. D. Washburn, T. Arias, P. P. Edger, J. C. Pires *et al.*, 2013 Watching the grin fade: Tracing the effects of polyploidy on different evolutionary time scales. Semin. Cell Dev. Biol. 24**:** 320-331.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker *et al.*, 2011 Recently formed polyploid plants diversify at lower rates. Science. 333**:** 1257-1257.

McClelland, E. K., and K. Naish, 2010 Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. Heredity. 105**:** 562-573.

McClelland, E. K., and K. A. Naish, 2008 A genetic linkage map for coho salmon (*Oncorhynchus kisutch*). Anim. Genet. 39**:** 169-179.

Miller, M. R., J. P. Brunelli, P. A. Wheeler, S. X. Liu, C. E. Rexroad *et al.*, 2012 A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol. Ecol. 21**:** 237-249.

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

Naish, K. A., and J. J. Hard, 2008 Bridging the gap between the phenotype and the genotype: linking genetic variation, selection, and adaptation in fishes. Fish. Fish. 9**:** 396-422.

Naish, K. A., R. B. Phillips, M. S. O. Brieuc, L. R. Newton, A. E. Elz *et al.*, 2013 Comparative genome mapping between Chinook salmon (*Oncorhynchus tshawytscha*) and rainbow trout (*O. mykiss*) based on homologous microsatellite loci. G3-Genes. Genom. Genet. 3**:** 2281-2288.

Near, T. J., R. I. Eytan, A. Dornburg, K. L. Kuhn, J. A. Moore *et al.*, 2012 Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. U.S.A. 109**:** 13698-13703.

Nichols, K. M., W. P. Young, R. G. Danzmann, B. D. Robison, C. Rexroad *et al.*, 2003 A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). Anim. Genet. 34**:** 102-115.

Ohno, S., 1970 *Evolution by gene duplication*. George Allen and Unwin, London.

Ohno, S., 1999 Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. Semin. Cell Dev. Biol. 10**:** 517-522.

Ostberg, C., L. Hauser, V. Pritchard, J. Garza and K. Naish, 2013 Chromosome rearrangements, recombination suppression, and limited segregation distortion in hybrids between Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and rainbow trout (*O. mykiss*). BMC. Genomics. 14**:** 570.

Otto, S. P., and J. Whitton, 2000 Polyploid incidence and evolution. Annu. Rev. Genet. 34**:** 401-437.

Phillips, R. B., K. A. Keatley, M. R. Morasch, A. B. Ventura, K. P. Lubieniecki *et al.*, 2009 Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). BMC. Genet. 10.

Phillips, R. B., N. R. Konkol, K. M. Reed and J. D. Stein, 2001 Chromosome painting supports lack of homology among sex chromosomes in *Oncorhynchus*, *Salmo*, and *Salvelinus* (Salmonidae). Genetica. 111**:** 119-123.

Phillips, R. B., M. R. Morasch, L. K. Park, K. A. Naish and R. H. Devlin, 2005 Identification of the sex chromosome pair in coho salmon (*Oncorhynchus kisutch*): lack of conservation of the sex linkage group with chinook salmon (*Oncorhynchus tshawytscha*). Cytogenet. Genome. Res. 111**:** 166-170.

Phillips, R. B., L. K. Park and K. A. Naish, 2013 Assignment of Chinook salmon (*Oncorhynchus tshawytscha*) linkage groups to specific chromosomes reveals a karyotype with multiple rearrangements of the chromosome arms of rainbow trout (*Oncorhynchus mykiss*). G3-Genes. Genom. Genet. 3**:** 2289-2295.

Phillips, R. B., and P. Ráb, 2001 Chromosome evolution in the Salmonidae (Pisces): an update. Biol. Rev. 76**:** 1-25.

Postlethwait, J. H., 2007 The zebrafish genome in context: ohnologs gone missing. J. Exp. Zool. B Mol. Dev. Evol. 308B**:** 563-577.

R Development Core Team, 2013 R: A language and environment for statistical ck omputing, pp. R Foundation for Statistical Computing. http://www.R-project.org, Vienna, Austria.

Rexroad, C. E., Y. Palti, S. A. Gahr and R. L. Vallejo, 2008 A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). BMC Genet. 9**:** 74.

Sakamoto, T., R. G. Danzmann, K. Gharbi, P. Howard, A. Ozaki *et al.*, 2000 A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterised by large sex-specific differences in recombination rates. Genetics. 155**:** 1331-1345.

Santini, F., L. J. Harmon, G. Carnevale and M. E. Alfaro, 2009 Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Biol. 9**:** 194.

Schubert, I., and M. A. Lysak, 2011 Interpretation of karyotype evolution should consider chromosome structural constraints. Trends Genet. 27**:** 207-216.

Soltis, P. S., and D. E. Soltis, 2000 The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. U.S.A. 97**:** 7051-7057.

Storz, J. F., J. C. Opazo and F. G. Hoffmann, 2013 Gene duplication, genome duplication, and the functional diversification of vertebrate globins. Mol. Phylogen. Evol. 66**:** 469-478.

Thorgaard, G. H., F. W. Allendorf and K. L. Knudsen, 1983 Gene-centromere mapping in rainbow trout: High interference over long map distances. Genetics. 103**:** 771-783.

Timusk, E. R., M. M. Ferguson, H. K. Moghadam, J. D. Norman, C. C. Wilson *et al.*, 2011 Genome evolution in the fish family salmonidae: generation of a brook charr genetic map and comparisons among charrs (Arctic charr and brook charr) with rainbow trout. BMC Genet. 12**:** 68.

Van de Peer, Y., S. Maere and A. Meyer, 2009 The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. 10**:** 725-732.

Wolfe, K. H., 2001 Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. 2**:** 333-341.

Woram, R. A., K. Gharbi, T. Sakamoto, B. Hoyheim, L. E. Holm *et al.*, 2003 Comparative genome analysis of the primary sex determining locus in five species of salmonid fish. Genome Res. 13**:** 272-280.

Wright, J. E., K. Johnson, A. Hollister and B. May, 1983 Meiotic models to explain classical linkage, pseudolinkage, and chromosome pairing in tetraploid derivative salmonid genomes. Isozymes. Curr. T. Biol. 10**:** 239-260.

Wright, J. E., B. May, M. Stoneking and G. M. Lee, 1980 Pseudolinkage of the duplicate loci for supernatant aspartate aminotransfease in brook trout, *Salvelinus fontinalis*. J. Hered. 71**:** 223-228.

Wu, Y., T. J. Close and S. Lonardi, 2011 Accurate construction of consensus genetic maps via integer linear programming. IEEE/ACM Trans. Comput. Biol. Bioinf. 8**:** 381-394.

Yano, A., R. Guyomard, B. Nicol, E. Jouanno, E. Quillet *et al.*, 2012 An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. Curr. Biol. 22**:** 1-6.

Zhang, J., 2003 Evolution by gene duplication: an update. Trends. Ecol. Evol. 18**:** 292-298.

Chapter 3: Genetic architecture of age at sexual maturity and growth-related traits in coho salmon and investigation of a trade-off between these traits

Abstract

Characterizing the genetic architecture of life history traits, such the number of loci linked with the traits, the distribution of their effects and the degree of correlation helps understand the causal basis of life history tradeoffs, and contributes to an accurate prediction of adaptive evolutionary change. In a number of salmonid species, faster growing fish in a given cohort mature earlier than conspecifics, and growth performance during spring and fall are important in determining age at sexual maturity. The aim of this study was to investigate the causal mechanisms underlying a tradeoff between age at sexual maturity and growth during these periods in coho salmon, by investigating the genetic architecture of age at sexual maturity and a suite of growth-related traits. We used a genome map based on 7415 non-duplicated RAD-sequenced tags to identify 37 Quantitative trait loci (QTL) linked to these traits in a study that spanned these decision periods. Several temporally expressed growth-related QTL mapped to the same position, suggesting that these regions affected growth across many months. We also found evidence for epistatic interactions between some growth traits, and that the effect of offspring sex on QTL expression differed. One genomic region was associated with age at sexual maturity and body length and weight measured during fall when maturation was initiated, indicating that the tradeoff between these traits may be due to genetic pleiotropy or co-localization of genes. Our results provide insights into the genetic mechanisms underlying a life history tradeoff in coho salmon, and have implications for understanding the genetic and evolutionary basis of life-history tradeoffs.

Introduction

Two fundamental components in life history theory are that natural selection acts to maximize fitness, and that tradeoffs between life history traits constrain their evolution (Roff 1992; Stearns 1992). A tradeoff exists when an increase in the value of one life history trait is linked to a decrease in another trait, thus balancing the fitness gains by one against reductions in the other (Roff 1992; Stearns 1992). Well-known examples of tradeoffs include survival versus reproduction, number versus quality of offspring and current versus future reproduction (Stearns 1989). Tradeoffs are typically described by negative phenotypic and genetic correlations between the two life history traits (Roff 2007; Fabian and Flatt 2012). Tradeoffs are of evolutionary importance because these values indicate that traits cannot evolve independently from each other, and that the adaptive evolution of these traits is highly constrained (Flatt and Heyland 2011; Fabian and Flatt 2012). The genetic basis of tradeoffs have been described by using quantitative genetic breeding designs to estimate the degree of correlation between traits, or by studying correlated responses to artificial selection (Roff 2007; Fabian & Flatt 2012; for examples, see Kim *et al.* 2010; Lee *et al.* 2013). These studies have helped predict the evolutionary trajectories of traits involved in tradeoffs. However, understanding the genetic architecture of life history traits such as the number of loci involved, the distribution of their effects and the degree of correlation between loci underlying different traits would provide

greater power in revealing the causal basis of tradeoffs. This information would contribute to more accurate predictions of adaptive evolutionary change.

Genomics has been increasingly used to study the genetic basis of life history variation (Roff 2007; Todesco *et al.* 2010; Bell and Aubin-Horth 2010; Olson *et al.* 2012), even in species or populations with little or no genetic information available. Such studies have been well-served by mapping of quantitative trait loci (QTL) underlying traits involved in tradeoffs (Norry *et al.* 2007; Haselhorst *et al.* 2011; Gnan *et al.* 2014; Dittmar *et al.* 2014). Such approaches could be particularly useful in determining whether the genetic basis of tradeoffs involves two or more tightly linked genes, or a gene with pleiotropic effects (Roff 2007; Flatt and Heyland 2011). For example, several studies in both plant and animal species have demonstrated that tradeoffs may be due to pleiotropy, epistatic interactions involving specific genomic regions, or environment- and population-specific QTL (e.g. Mitchell-Olds 1996; Shook & Johnson 1999; Nichols *et al.* 2008; Latta & Gardner 2009; Haselhorst *et al.* 2011; Gnan *et al.* 2014; Dittmar *et al.* 2014).

Age at sexual maturity is an important fitness-related trait in salmonids (Gross 1985; Fleming 1998; Kodama *et al.* 2012). This trait is linked to variation in growth. Male salmonids employ several alternative mating tactics to achieve fertilization success; smaller, younger-maturing males may adopt sneaking behavior to gain access to spawning females whereas large, late-maturating males may engage in fighting to outcompete smaller males to gain access to spawning females (Gross 1985; Holtby and Healey 1990; Fleming 1998). In female salmon, variability in size and age tends to be smaller than in males, but size is linked to reproductive output (van den Berghe and Gross 1984; Steen and Quinn 1999; Dickerson *et al.* 2002). Tradeoffs between age at sexual maturity and growth involve a shift in energy allocation from growth to reproduction (Roff 1992; Stearns 1992; Hendry and Stearns 2004). Maturation and growth are correlated in most salmonid species (Thorpe *et al.* 1983), where faster growing fish in a given cohort mature earlier than conspecifics (e.g. Hutchings & Jones 1998; Silverstein *et al.* 1998; Shearer & Swanson 2000; Morita & Fukuwaka 2006). Thorpe *et al.* (1998) have proposed a model where growth performance during fall one year prior to and spring six months prior to the maturation conversion determine the age of sexual maturity. Further, Shearer & Swanson (2000) observed that age at sexual maturity of male Chinook salmon was correlated with whole-body lipid levels measured during the fall when maturation was initiated, suggesting the added significance of this period. The model of Thorpe *et al.* (1998) assumes that age at sexual maturity is determined by a genotype-by-environment interaction during crucial "decision" periods. Specifically, maturation is triggered if growth performance in an individual fish surpasses a genetically determined threshold level during the critical periods. The model also incorporates the possibility that maturation is inhibited in response to poor growth performance during these time periods. The decision to mature occurs every year (Thorpe *et al.* 1998), resulting in variable age at maturity in many salmon species (e.g. Waples *et al.* 2001, 2009; Klemetsen *et al.* 2003).

The genetic basis of the tradeoff between age at sexual maturity and growth in salmon has been investigated through QTL mapping (Martyniuk *et al.* 2003; Moghadam *et al.* 2007; Küttner *et al.* 2011; Pedersen *et al.* 2013). Martyniuk *et al.* (2003) observed that QTL for male age at maturation in rainbow trout mapped to similar regions as those for body mass measured at

three time points, indicating that pleiotropy or linked genes may be responsible for the tradeoff. Similarly, QTL for adult maturation and body length or weight mapped to the same chromosomes in Arctic charr and Atlantic salmon (Küttner *et al.* 2011; Pedersen *et al.* 2013). These studies examined growth-related traits at one to four time periods. Of these, only Küttner *et al.* (2011) measured body weight during crucial decision periods suggested by Thorpe *et al.* (1998), and demonstrated that QTL linked with body weights had weak effects on maturation timing. A study that includes a suite of growth-related traits more frequently measured over the course of development will help assess the relationship between age at maturation and growth during specific decision periods within autumn and spring when tradeoffs might act. The current study is aimed at using frequent measures of growth that span the key maturation decision periods in a salmonid species with simple life history, coho salmon. This species is ideal to investigate the tradeoff between growth and age at sexual maturity because it is a short lived semelparous species, with most individuals maturing at three years, but a fraction of males (and occasionally, females) maturing as two year olds (Sandercock 1991; Quinn 2005). Given that age at sexual maturity is an observable binary trait in coho salmon, it is feasible to phenotype all individuals for this trait within two years. Information from this study will provide insights into how the maturation decision might be genetically influenced by temporal growth performance.

The aim of this study was to explore the genetic architecture of growth and age at sexual maturity, and investigate the genetic basis for a tradeoff between age at sexual maturity and growth during crucial decision periods in coho salmon. This aim will be achieved by detecting loci linked to age at maturity and to growth-related traits measured at different time points that span the key decision windows. Co-localization of these loci will be interpreted as evidence of a tradeoff (Roff 2007). A domesticated aquaculture strain that has been selectively bred for large size and early maturation for over 15 generations (Myers *et al.* 2001) was available for this study. Measurements in an F4 outbred cross using this strain were initiated during fall one year preceding the maturation conversion until the fall when maturation was initiated. Restriction site-associated DNA (RAD) sequencing (Baird *et al.* 2008) was performed on offspring, and sequences were aligned against a high-density linkage map for the species (Kodama *et al.* 2014) to locate QTL underlying the traits. Our results were aimed at providing a better understanding of the genetic mechanisms underlying a life history tradeoff in coho salmon, with the intention of improving understanding of life history evolution among salmonid species.

Materials and Methods

Experimental crosses

In December 2010, F3 crosses were created from outbred experimental lines of coho salmon established in 2001 (McClelland and Naish 2010). The F0 populations originated from Bingham Creek (47º15’N, 123º40’W), a tributary to the Satsop River that drains to the Southwest Washington coast, and from the Domsea aquaculture line. These populations differ at growth-related traits such as length, weight and daily growth rate (McClelland *et al.* 2005). The Domsea commercial line originated in 1973 and 1974 from the Washington Department of Fish and Wildlife’s (WDFW) Wallace River Hatchery (47º87’N, 121º71’W), and was selectively bred for early maturation and large size over 15 generations (Myers *et al.* 2001). Four families of F4

offspring were selected for QTL analyses (Family_1, Family_2, Family_3, Family_4). All individuals were marked with PIT tags (Biomark Inc., Boise, ID, USA) in August 2011 at their parr stage (the juvenile stage prior to physiological transformation for saltwater adaptation) for individual identification. Each family was placed in a 76cm diameter tank from August 2011 to August 2012 and then in a 180cm diameter tank from September 2012 to January 2013. Throughout the course of the experiment, individuals were fed to satiation and reared at a constant 10°C.

Phenotypic measures

Fall one year prior to and spring six months prior to the maturation conversion have been identified to be the critical seasons during which a decision to mature is taken (Thorpe *et al.* 1998). In coho salmon, early maturing fish in a cohort would reach reproductive stage in year 2. Therefore, these periods would occur at the end of the first year and during spring of the second year. To capture these important growth periods, fork length and body weight for Family_1 and Family_2 were measured every six weeks from August 2011 when fish were approximately 8 months of age, to November 2012 when fish were approximately 24 months of age, the period of first maturation. Each fish was lightly squeezed to ascertain production of milt or eggs, and thus maturation status, between September and December 2012. Mature fish were designated as 2-year-olds, and fish that did not mature by January 2013 were assumed to be later maturing 3-year-olds. To verify growth-related QTL in other crosses, fork length and body weight were measured every six weeks from August 2011 to February 2012 for Family_3 and Family_4, when fish were approximately 11 months of age and the experiment was terminated due to space constraints. Daily growth rates were calculated as the average daily change in body weight for all families.

DNA extraction and sequencing

Fin clips from all individuals were collected, and genomic DNA was extracted using the DNeasy extraction kit (QIAGEN, Valencia, CA, USA) following the manufacturer's procedures. DNA was digested with *Sbf*I, and a 6-nucleotide barcode was added to each sample for individual identification following protocols outlined in Baird *et al.* (2008). Thirty six individuals were pooled in a single library and sequenced with 100 bp single-read lengths using the Illumina HiSeq 2000 sequencer. The sequences were sorted, and the barcodes were used to separate individual sequences using the *process_radtags* function implemented in STACKS (Catchen *et al.* 2011, 2013). Because the quality score of sequences decreased beyond 74 nucleotides, sequences were trimmed to 74 nucleotides to remove low-quality sequences. A locus was defined as a 74-nucleotide RAD sequence for the rest of this study.

Genotyping

Sequences from all individuals were aligned to the reference database of RAD loci developed by Kodama *et al.* (2014) using BOWTIE by allowing a maximum of three nucleotide mismatches per locus. Polymorphic loci with 2 alleles were identified in each family using STACKS. Genotypes at these loci were determined when alleles were sequenced with a depth greater than 10X per individual. Due to the potential bias toward heterozygous genotypes,

genotypes were corrected after running STACKS with the Python script developed by Brieuc *et al.* (2014). Specifically, individuals were determined as heterozygotes at a locus if both alleles had a depth of more than two and the total read depth was 10X or greater.

Genetic sex was determined in all four families with a Y-linked growth hormone pseudogene (GH5 and GH6); (Devlin *et al.* 2001) and sex-determining gene, sdY (sdY E2S1 and sdY E2AS4); (Yano *et al.* 2013), as described in Kodama *et al.* (2014).

Statistical description of phenotypes

Since the results of QTL analyses are influenced by trait distribution and differences in traits between sexes, we conducted an initial exploration of the phenotype data. All lengths, weights and daily growth rates were normally distributed in each family, and age at sexual maturation had a binomial distribution for Family_1 and Family_2. Two-sample F tests were performed to determine whether variances were equal or unequal between females and males for all traits in each family, except for age at sexual maturity, which was dichotomous. To assess whether the phenotypic means for each traits differed significantly between females and males, independent two-sample t-tests were performed with or without equality of variances. Phenotypic correlations between trait pairs for lengths, weights and daily growth rates in male and female for each family were estimated using Pearson's correlations. Phenotypic correlations between age at sexual maturity and lengths, weights or daily growth rates were estimated using Point-Biserial correlations. All statistical tests were performed in R version 3.0.2 (R Development Core Team, 2010).

QTL analyses

Quantitative trait locus (QTL) analyses for all traits were performed using single, two- and multiple-dimensional QTL models in the R-based software package, R/qtl (Broman *et al.* 2003; Broman and Sen 2009). The female consensus linkage map for coho salmon developed by Kodama *et al.* (2014) was used as a framework for the analyses. Salmonids are known to have large differences in recombination rates between the sexes (Sakamoto *et al.* 2000), a result we also observed in coho salmon (Kodama *et al.* 2014); to account for such differences, QTL analyses were conducted separately for each F3 parent.

For normally distributed traits (weight, body length and daily growth rate), we used the multiple imputation approach to perform quantitative trait loci analyses (Sen and Churchill 2001; Broman *et al.* 2003; Broman and Sen 2009); missing genotypes given observed marker data were simulated with a step interval of 1 cM, with 216 draws per genotype and assuming a genotyping error rate of 0.01. For binary traits (age at maturity), we used the Haley-Knott regression (Broman and Sen 2009) to perform QTL analyses, which relies on genotype probabilities between observed marker data. These probabilities were calculated with a step size of 1 cM and a genotype error rate of 0.01.

Genotypic and phenotypic associations were examined through a series of computational steps. First, a genome-wide significance threshold ($\alpha = 0.05$) LOD score for each trait was determined by 1,000 permutations of genotypic and phenotypic data (Churchill and Doerge

1994; Doerge and Churchill 1996; Broman and Sen 2009). For age at sexual maturity, chromosome-specific significance threshold LOD scores were also estimated using 1,000 permutations of genotypic and phenotypic data. Second, single QTL analyses were conducted with the *scanone* function implemented in R/qtl. Sex was included as an additive or interactive covariate in the analyses for all traits (Sen and Churchill 2001; Broman and Sen 2009). Third, further testing was performed with QTL-linked markers as additive covariates in a single-QTL scan to detect additional loci with more modest effect; here, QTL that exceeded the 5% genome-wide and chromosome-specific significance thresholds for age at sexual maturity in the initial computation, or QTL that exceeded the 5% genome-wide significance threshold for all other traits, were used as an additive covariate in the model. This process was repeated until no new QTL were identified. Fourth, interactions between QTL were tested by performing a two-QTL genome scan to investigate the presence of epistatic interactions with the *scantwo* function. Fifth, significant QTL and QTL × QTL interactions detected by the single- and two-QTL analyses were fitted into a multiple QTL model with sex as a covariate. Improved estimates of the QTL locations were obtained with the *refineqtl* function implemented in R/qtl. Sixth, we used the improved estimates of the QTL locations to fit a multiple model containing all QTL effects and significant interactions. Any insignificant QTL or interaction terms at $\alpha = 0.05$ genome-wide thresholds, as well as at $\alpha = 0.05$ chromosome-specific thresholds for age at sexual maturity were removed, and the location of remaining QTL were refined in an iterative approach until only significant terms remained in the model. Finally, for all QTL significant at $P \leq 0.05$, the 95% Bayes credible interval (CI) was obtained using the *bayesint* function, and percentage of phenotypic variance explained (PVE) by each QTL was estimated using the *fitqtl* function.

Results

Phenotypic Trait measures and statistical description of traits

Two F4 families (Family_1 and Family_2) of 67 and 102 individuals, respectively, were measured over 16 months from August 2011 to November 2012 for 33 phenotypic traits associated with growth (Table 3.1): fork length at 11 time points ($Length_{23Aug11}$ to $Length_{28Nov12}$), body weight at the same 11 time points ($Weight_{23Aug11}$ to $Weight_{28Nov12}$), daily growth rate across 10 time intervals ($DGR_{23Aug11\_04Oct11}$ to $DGR_{15Oct12\_28Nov12}$) and age at sexual maturity at 22 to 24 months. Of the 67 F4 progeny from Family_1, 55.22% (n = 37) were genotypic males and 44.78% (n = 30) were genotypic females. Of the 102 F4 progeny from Family_2, 51.96% (n = 53) were genotypic males and 48.04% (n = 49) were genotypic females.

Two F4 families (Family_3 and Family_4) of 51 and 91 individuals, respectively, were measured over 7 months from August 2011 to February 2012 for 14 phenotypic traits associated with growth (Table 3.1): fork length at five time points ($Length_{23Aug11}$ to $Length_{07Feb12}$), body weight at five time points ($Weight_{23Aug11}$ to $Weight_{07Feb12}$) and daily growth rate across four time intervals ($DGR_{23Aug11\_04Oct11}$ to $DGR_{28Dec11\_07Feb12}$). Of the 51 F4 progeny from Family_3, 54.90% (n = 28) were genotypic males and 45.10% (n = 23) were genotypic females. Of the 91 F4 progeny from Family_4, 65.93% (n = 60) were genotypic males and 34.07% (n = 31) were genotypic females.

Variances between female and male offspring did not differ for most of the traits in Family_1, Family_2 and Family_3 (except for $DGR_{15Oct12\_28Nov12}$ in Family_1; $Length_{07Feb12}$, $Length_{28Nov12}$, $Weight_{23Aug11}$, $Weight_{07Feb12}$, $DGR_{04Oct11\_15Nov11}$, $DGR_{15Nov11\_28Dec11}$, $DGR_{28Dec11\_07Feb12}$, $DGR_{01May12\_12Jun12}$ in Family_2; $Weight_{15Nov11}$, Weight15Nov11, $DGR_{23Aug11\_04Oct11}$, $DGR_{28Dec11\_07Feb12}$ in Family_3; Supplementary Material Table S3.1). Furthermore, phenotypic means for all traits did not differ significantly between female and male offspring in these three families (Supplementary Material Table S3.1). In contrast, variances, as well as phenotypic means for all traits differed significantly between female and male offspring in Family_4 (Supplementary Material Table S3.1).

Almost all traits were correlated between each other in male and female offspring within all families ($P < 0.05$, 0.01 or 0.001; Supplementary Material Figure S3.1, see Table S3.2.a to Table S3.2.h for details). Specifically, almost all lengths, weights and daily growth rates were positively correlated with each other in both male and female progeny within all families (Supplementary Material Figure S3.1). On the other hand, age at sexual maturity was negatively correlated with almost all traits in both male and female progeny within Family_1 and Family_2 (Supplementary Material Figure S3.1).

Fork lengths and body weights steadily increased in both male and female progeny of all families over the course of experiment, and variance of these traits also increased over time (Supplementary Material Figure S3.2). Daily growth rates were relatively low from August 2011 to May 2012 when fish were 8 to 10 months of age in both male and female progeny of Family_1 and Family_2, and then increased from June 2012 to October 2012 prior to the maturation conversion (Supplementary Material Figure S3.2.c, S3.2.f, S3.2.i, S3.2.l). Interestingly, in average, early-maturing individuals (at age 2) were consistently larger in length and weight than late-maturing individuals for both male and female progeny for these two families (Supplementary Material Figure S3.3.a, S3.3.b, S3.3.d, S3.3.e). The pattern in changes for daily growth rate considerably differed between early- and late-maturing individuals; daily growth rate greatly increased from February 7, 2011 to July 24, 2011 for early-maturing individuals, whereas no such large increase was observed for late-maturing individuals (Supplementary Material Figure S3.3.c, S3.3.f). This observation was consistent between the two families. First signs of maturation in Family_1 and Family_2 were observed in October 2012. In Family_3, daily growth rates were low during the fall of 2011 in both sexes (Supplementary Material Figure S3.2.o, S3.2.r). Daily growth rates were fairly constant across the entire sampling period for both sexes in Family_4 (Supplementary Material Figure S3.2.u, Figure S3.2.x).

Genetic Markers

Family_1, Family_2, Family_3, and Family_4 had 1312, 2033, 1295 and 1950 biallelic loci, respectively, that were polymorphic in each female parent with known position on the coho linkage map constructed by Kodama *et al*. (2014). Family_1, Family_2, Family_3, and Family_4 had 1502, 1598, 1053 and 1630 biallelic loci, respectively, that were polymorphic in each male parent with known position on the coho linkage map. These loci were employed for further analyses.

QTL Analyses

QTL mapping analyses across all four families identified a total of 36 significant QTL at $\alpha = 0.05$ genome-wide thresholds for length, weight, daily growth rate or age at sexual maturity (Table 3.2). In addition, one significant QTL at a chromosome-specific threshold for age at sexual maturity was detected on Co10 (Table 3.2). QTL were not identified for daily growth rates at certain time intervals ($DGR_{28Dec11\_07Feb12}$, $DGR_{07Feb12\_01May12}$, $DGR_{12Jun12\_24Jul12}$, $DGR_{24Jul12\_11Sep12}$, and $DGR_{15Oct12\_28Nov12}$) and weight at one time point ($Weight_{07Feb12}$). For the rest of the traits, one or two QTL were detected for each trait (Table 3.2). The ability to detect QTL differed by family size and by sex. In total, only two QTL were found using the segregation information from the dams, while the rest were detected using the information from the sires (Table 3.2). Large family size did not always correspond to more power to detect QTL. Specifically, a total of one and two QTL were detected in Family_4 and Family_1, comprising 91 and 67 individuals, respectively. A total of five QTL was detected in Family_3 that contained 51 progeny. The largest family, Family_2 was the most informative with 29 detected QTL. Unique QTL were observed in each family, suggesting that different QTL might have segregated across families.

The 37 QTL that were identified mapped to nine unique regions of the coho genome (Figure 3.1; Table 3.2). Four of these regions on linkage groups Co07, Co21, Co23 and Co30 were associated with more than one QTL for various traits measured at different time points, suggesting that these regions may provide evidence for a genetic basis for phenotypic correlations among growth-related traits in this species. Interestingly, QTL on Co30 was associated with age at sexual maturity in Family_1, and co-localized with length and weight QTL measured in Family_2, when fish were maturing (October and November 2012; Figure 3.1). Although these QTL were observed across two different families, they overlapped with each other, suggesting that age at sexual maturity and growth during fall when maturation is initiated may be genetically correlated. We did not observe any overlaps between age at sexual maturity and growth during other critical decision periods (fall of 2011 and spring of 2012). Genomic regions on linkage groups Co01, Co03, Co05, Co10 and Co28 were only linked with one QTL underlying traits measured at different time points, indicating that these QTL may be only temporally expressed.

The effect of sex of offspring was significant for the majority of the traits (Table 3.2) although phenotypic means did not differ significantly between female and male offspring for almost all traits with mapped QTL. Sex × QTL interactions were also significant for numerous traits (Table 3.2), suggesting that QTL effects on these traits may differ between the sexes. QTL × QTL interactions were significant for $Length_{01May12}$, $Length_{12Jun12}$, $Weight_{01May12}$ and $Weight_{12Jun12}$ (Co07 × Co23; Table 3.2), indicating that the effect of one QTL depended on the genotype at the other QTL. QTL x QTL interactions were not observed between the phenotypic traits that were measured during the "decision periods" and age at sexual maturity (October 2011, November 2011, May 2012, June 2012).

Effect sizes of individual QTL varied from 12.93% to 33.78% for each trait (Table 3.2). Generally, a large proportion of phenotypic variance (PEV) was explained by QTL linked with

lengths measured when the fish were 8 months to 18 months of age ($Length_{23Aug11}$ to $Length_{12Jun12}$); all QTL except one explained more than 20% of the phenotypic variation for these traits. In contrast, a moderate proportion of variance (≤ 20%) was explained by QTL underlying lengths measured when fish were 19 to 24 months of age. The percentage of the phenotypic variation attributable to the QTL for weights and daily growth rates varied widely, ranging from 12.93% to 28.11%, and from 15.94% to 29.95%, respectively. Finally, age at sexual maturity QTL on Co10 and Co30 were of moderately large effect, explaning15.21% and 25.46% of the PVE of this trait, respectively. Across families, all QTL detected in Family_3 were of major effect, with PVE over 20%. PVE explained by QTL detected in Family_2 varied, ranging from 12.93% to 28.11%.

Discussion

The aim of this study was to reveal the genetic architecture of growth and age at sexual maturity, and to investigate the causal mechanisms for life history tradeoffs involving these traits, particularly those measured during the key decision windows for maturation. Here, we have located several QTL for age at sexual maturity and for length, weight and daily growth rate measured over 16 months of growth. Multiple QTL across different time periods localized to the same genomic region, suggesting that these regions affected growth across many months of development. We found that QTL × sex interactions were widespread, indicating that the effect of QTL on growth differed between the male and female offspring. There were also significant QTL × QTL interactions for a few traits, which means the expression of one QTL depended on that of the other QTL for these traits. Finally, our mapping efforts revealed the genetic basis of tradeoff involving age at sexual maturity and growth; a genomic region on a linkage group, Co30, was likely involved in a life history tradeoff between age at sexual maturity and growth during fall when maturation was initiated, although other relationships not detected by this study might also exist.

The distribution and effect sizes of the observed QTL indicate that growth and age at sexual maturation in coho salmon are controlled by a relatively low number of genomic regions that explain a fairly large proportion of the phenotypic variance observed (Figure 1; Table 2). These QTL may prove useful candidates in aquaculture breeding programs based on marker-assisted selection, providing accelerated genetic improvement for growth performance. It is important to note, however, that there are a few aspects of the analyses that limit the power of the study. The relatively small sample sizes (51 to 102 individuals per family) might lead to an overestimation of the QTL effect sizes (Beavis 1994, 1998), while underestimating the number of QTL with smaller effects (Mackay 2001; Doerge 2002; Mackay *et al.* 2009). Additionally, the power to detect QTL in outbred crosses is lower than inbred crosses, again restricting the study to loci of large effect (Lynch and Walsh 1998; Mackay 2001). Consequently, the genetic architecture of growth-related traits in this species may be more complex than described here, and further investigations will almost certainly reveal additional loci underlying the traits measured.

The co-localization of QTL underlying growth-related traits measured across different time periods suggest that same genes may be expressed over lifetime of individuals, or the traits

could be affected by closely linked loci. This result also provides evidence for a genetic basis for the extensive phenotypic correlations observed among the temporal measures. Our results are in agreement with McClelland & Naish (2010), confirming that numerous growth-related traits across several life stages are genetically correlated in coho salmon. In the present study, many QTL spanned relatively large segments of the linkage groups (up to ~202.76cM; Figure 3.1; Table 3.2), and most were detected with the segregation information from the sires (Table 3.2). Such results were expected because reduced recombination in male coho salmon (Kodama *et al.* 2014) provides a greater power to detect QTL, but there is a lower resolution of the QTL position (Hayes *et al.* 2006; see examples in Houston *et al.* 2009, Vasemägi *et al.* 2010). Consequently, it was not possible to distinguish between tightly linked loci and pleiotropy with the current study. Future efforts should be directed toward refinement of the QTL locations using the segregation information from dams to fully understand the genetic basis of phenotypic correlations in this species. Nevertheless, the findings depicted herein provide an important step toward identifying specific genomic regions that may explain phenotypic correlations in this species.

One key finding of this study is that the effects of observed QTL underlying numerous temporally measured traits varied widely between male and female offspring. This result implies that these QTL are expressed differently between the sexes. Such QTL × sex interactions affecting growth in salmonids have rarely been reported, except for Haidle *et al.* (2008), who observed several sex-specific QTL for early maturation and growth in rainbow trout (Haidle *et al.* 2008). Our study revealed that age at sexual maturity, as well as lengths, weights and daily growth rates over the course of development in coho salmon may be under sex-specific control, and it provides a genetic explanation for sexual dimorphism in growth patterns for salmonids (e.g. Gross 1985; Fleming 1998). Specifically, the expression of alternative mating tactics in male (Gross 1985; Holtby and Healey 1990; Fleming 1998) may be under separate genetic control, since such variation in age at maturity are typically at low frequency in females in wild populations. Our observations also highlight the importance of assessing the contribution of sex and its interaction with QTL when identifying QTL linked with growth-related traits in salmon.

This study also reports significant epistatic interactions between QTL linked with lengths and weights measured in spring prior to maturation conversion, in May and June of 2012 when fish were approximately 8 to 9 months of age. Such interactions were restricted to two linkage groups, Co07 and Co23. The effects of epistatic interactions on salmonid growth have been reported in a few studies (e.g. Nichols *et al.* 2008; Wringe *et al.* 2010). Nichols *et al.* (2008) observed significant epistatic interactions among QTL for smoltification and growth rate from February to June when fish were one year old in rainbow trout, and these effects were restricted to three linkage groups. Similarly, epistatic interactions among QTL for body weight observed by Wringe *et al.* (2010) were significant for a few linkage groups, although these effects were relatively weak in their study populations. Nevertheless, findings from these studies illuminate the importance of epistatic interactions among certain linkage groups on salmonid growth. Observed epistatic interactions may be attributed to the existence of co-adapted gene complexes that confer a selective advantage upon growth; if so, outbreeding could result in the disruption of these gene complexes, potentially producing a reduction in fitness, thus posing a genetic risk to wild populations.

Two QTL linked with age at sexual maturity have been characterized in this study (Co10 and Co30; Table 3.2). Interestingly, in Atlantic salmon, Pedersen *et al.* (2013) found a QTL linked with adult sexual maturation on Ssa23, which is homologous to Co10 in coho salmon (Co10 is a metacentric chromosome that represents a fusion of the Atlantic salmon chromosome arms, Ssa23 and Ssa06q; Kodama *et al.* 2014). QTL associated with adult maturation have also been mapped to a linkage group partially homologous to Co10 in rainbow trout (RT-24; Haidle *et al.* 2008). Both the present study and that on Arctic charr (Moghadam *et al.* 2007) detected QTL on the sex chromosome. However, homologies between chromosome arms in Arctic charr and coho salmon are currently unknown, therefore it is unclear whether these are homologous loci across the two species. Interestingly, the current study demonstrated that sex of offspring affected the expression of mapped QTL differently. The interaction between QTL on Co10 and sex was significant. QTL mapping was conducted separately on male and female progeny to confirm this finding, and results indicated that this QTL affects the age at sexual maturity in male and female progeny differently. In contrast, such QTL × sex interaction was not significant for QTL located on the sex chromosome Co30, implying that this QTL affects age at sexual maturity in both sexes equally. Such findings suggest that QTL on Co30 may have a potential for selection on age at sexual maturity in both sexes if targeted through marker-assisted selection. The results shed light on a genetic architecture underlying age at sexual maturity and demonstrated that the discovery of QTL linked with this trait should be performed while considering the effect of sex, as well as its interaction with QTL.

Significant negative phenotypic correlations between age at sexual maturity and growth at different time periods demonstrated that these traits were involved in a tradeoff over the course of development. However QTL for age at maturity and growth during fall one year prior to and spring six months prior to the maturation conversion, suggested by Thorpe *et al.* (1998), did not map to the same position; therefore, this study did not provide evidence for a genetic tradeoff between age at sexual maturity and growth during these critical decision periods. This result might have a biological explanation, but we cannot rule out the fact that the current study did not have sufficient power to detect this relationship. In contrast, QTL for body lengths and weights measured during fall when maturation was initiated (October and November 2012) overlapped with age at sexual maturity QTL on the sex chromosome, Co30; such results suggest that the observed tradeoff between these growth traits and age at sexual maturity may be due to genetic pleiotropy or co-localization of genes. However, the overlapped region spans over a relatively large distance, therefore this observation warrants fine-scale mapping of QTL in this region to distinguish between these two factors. Although not included in the model developed by Thorpe *et al.* (1998), Shearer & Swanson (2000) found that the rate of maturation of male Chinook salmon was correlated with whole-body lipid accumulation during fall when maturation was initiated, highlighting the importance of growth during this time period. The QTL mapping provides evidence for a genetic basis for this physiological observation, and indicates that this growth period may play an important role in determining age at sexual maturity and shaping the tradeoff.

In this study, we located numerous QTL linked with age at sexual maturity, as well as a suite of growth-related traits temporally measured over the critical decision periods in coho

salmon. The observed tradeoffs corresponding with maturation loci on Co30 may be due to the existence of pleiotropic effects or linkage disequilibrium, and future efforts should be directed towards fine-scale mapping of QTL in this region to distinguish between these factors. Under selection, traits involved in the observed tradeoff may respond in a correlated manner, therefore the process of adaptive evolution for these traits may be highly constrained. The QTL depicted herein may prove useful candidates in aquaculture breeding programs to provide accelerated genetic improvement for growth performance. Furthermore, findings from this study provide the foundation for future efforts in localizing candidate genes shaping life history trade-offs and exploring the adaptive significance of the tradeoff in natural populations.

Acknowledgements

We would like to thank Isadora Jimenez-Hidalgo for assistance with genomic data processing. Michael White, Anna Greenwood and Krista Nichols provided invaluable comments on the QTL analyses. We also appreciate Linda Park and Jim Myers for their help in establishing the initial crosses, as well as David Rose and Javier Cabrera for maintaining the lines. Initial experimental lines were obtained from SweetSpring Salmon/Aquaseed Corporation, suppliers of the Domsea broodstock and we would like to thank Per Heggelund, Greg Hudson and Patty Munsell for providing these lines. Funding for this study was provided by NOAA Fisheries/ Federal Columbia River Power System (FCRPS) Biological Opinion Remand Funds (to K.A.N. and J.J.H.), School of Aquatic and Fishery Sciences, and Graduate Opportunities and Minority Achievement Program (GO-MAP, University of Washington) award (to M.K.). The authors declare that they have no competing interests.

Figures

Figure 3.1 - Quantitative trait loci (shown as 95% Bayes credible intervals) for age at sexual maturity and growth-related traits. Only linkage groups with significant QTL are shown, and linkage groups are named accordingly to the previously published map (Kodama *et al.* 2014). QTL for age at sexual maturity, length, weight and daily growth rate are shown in dark red, green, blue and pink, respectively. Family name in which the corresponding QTL were detected are shown in parentheses.

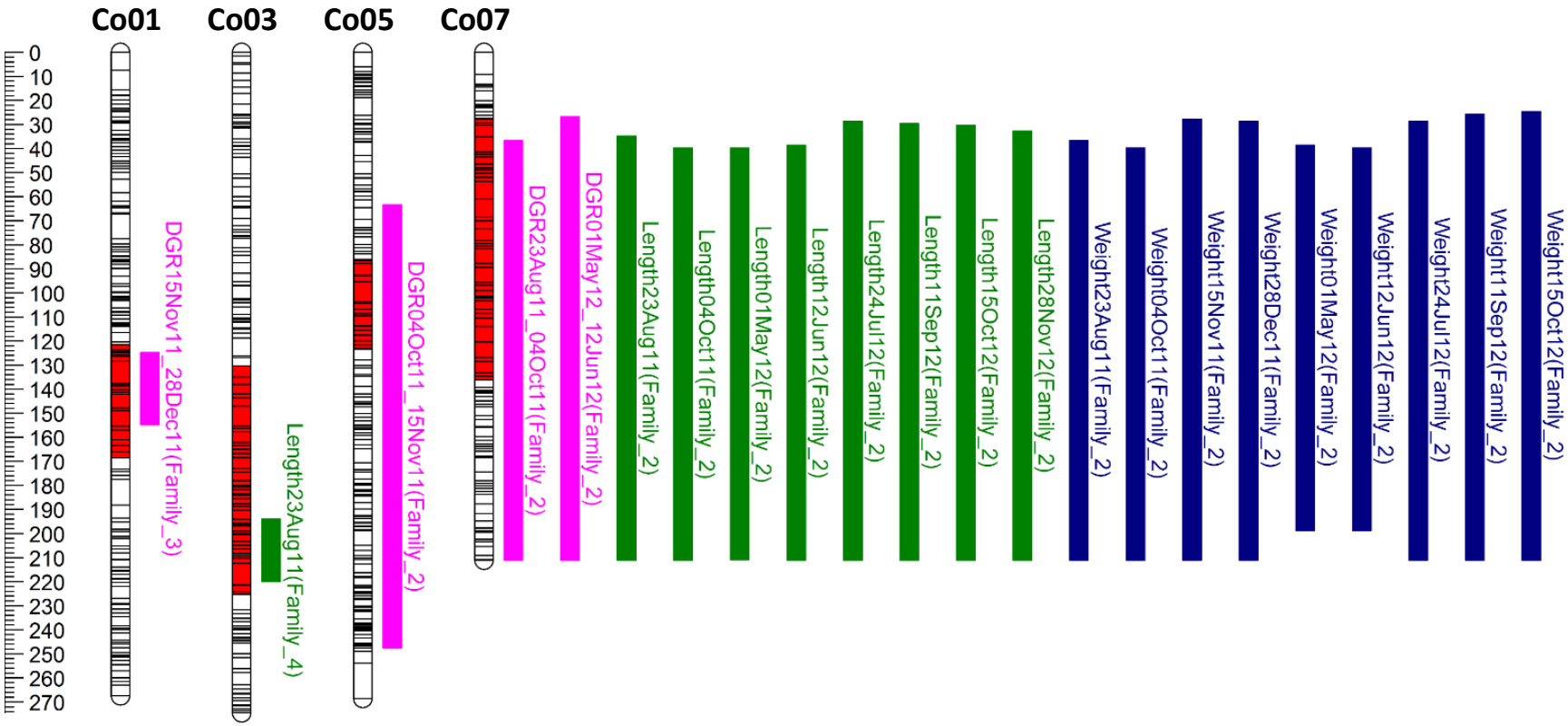
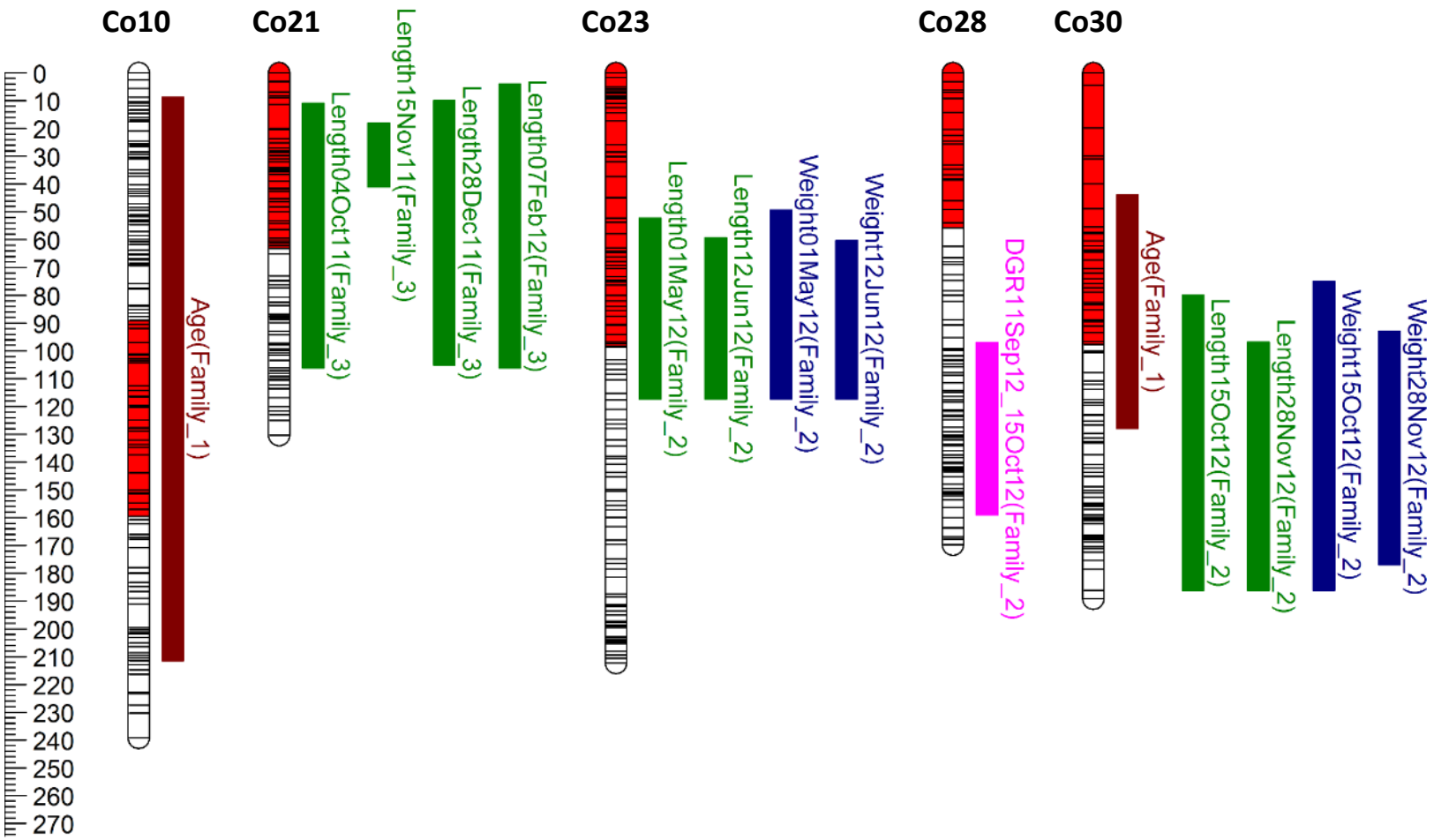


Figure 3.1 (Cont.)



Tables

Table 3.1 - Sampling date, the trait measured for QTL analysis and families in which measurements were taken.

Measurement Date	Measurements taken	Family
8/23/2011	Length, weight	Family_1, 2, 3, 4
10/4/2011	Length, weight	Family_1, 2, 3, 4
11/15/2011	Length, weight	Family_1, 2, 3, 4
12/28/2011	Length, weight	Family_1, 2, 3, 4
2/7/2012	Length, weight	Family_1, 2, 3, 4
5/1/2012	Length, weight	Family_1, 2
6/12/2012	Length, weight	Family_1, 2
7/24/2012	Length, weight	Family_1, 2
9/11/2012	Length, weight, maturity status	Family_1, 2
10/15/2012	Length, weight, maturity status	Family_1, 2
11/28/2012	Length, weight, maturity status	Family_1, 2

Table 3.2 - Summary of the QTL identified. Identification includes trait; family name in which the QTL was identified in (Family); the parent in which the QTL was detected (Source); linkage group; positions and markers to which QTL mapped for; 95% Bayes credible interval (Bayes CI 95%); max LOD score (LOD); percentage of variation explained (PVE); P value of the F-statistic. QTL are summarized in the order of age at sexual maturity, length, weight to daily growth rate, and from Family_1 to Family_4.

Trait	Family	Source	Linkage Group	Position (cM)	Closest Marker	Bayes CI (95%)	LOD	PVE	P value of the F-statistic
Age	Family_1	sire	Co10	8.66	Co51520_Co10a	8.66 - 211.42	3.39	15.21	<0.001
Age	Family_1	sire	Sex	NA	NA	NA	3.17	14.09	<0.001
Age	Family_1	sire	Co10 x Sex	NA	NA	NA	3.06	13.56	<0.001
Age	Family_1	sire	Co30	110.92	Co49383_Co30	51.92 - 132.92	5.25	25.46	<0.001
Length23Aug11	Family_2	sire	Co07	155.69	Co12537_Co07b	34.69 - 211.09	3.66	16.76	<0.001
Length23Aug11	Family_2	sire	Sex	NA	NA	NA	1.92	8.42	<0.05
Length23Aug11	Family_2	sire	Co07 x Sex	NA	NA	NA	1.8	7.87	<0.01
Length04Oct11	Family_2	sire	Co07	155.69	Co12537_Co07b	39.69 - 211.09	6.32	25.59	<0.001
Length04Oct11	Family_2	sire	Sex	NA	NA	NA	2.56	9.45	<0.01
Length04Oct11	Family_2	sire	Co07 x Sex	NA	NA	NA	2.23	8.56	<0.01
Length01May12	Family_2	sire	Co07	156.69	Co36884_Co07b	39.69 - 210.87	5.97	25.45	<0.001
Length01May12	Family_2	sire	Co23	115.33	Co43125_Co23	52.21 - 117.33	5.21	21.68	<0.001
Length01May12	Family_2	sire	Sex	NA	NA	NA	3.16	12.42	<0.01
Length01May12	Family_2	sire	Co07 x Co23	NA	NA	NA	1.16	4.28	<0.05
Length01May12	Family_2	sire	Co23 x Sex	NA	NA	NA	3.13	12.26	<0.001
Length12Jun12	Family_2	sire	Co07	156.69	Co36884_Co07b	38.69 - 211.09	6.18	26.29	<0.001
Length12Jun12	Family_2	sire	Co23	115.33	Co43125_Co23	59.33 - 117.33	5.63	23.55	<0.001
Length12Jun12	Family_2	sire	Sex	NA	NA	NA	2.39	9.11	<0.01
Length12Jun12	Family_2	sire	Co07 x Co23	NA	NA	NA	1.79	6.72	<0.01
Length12Jun12	Family_2	sire	Co23 x Sex	NA	NA	NA	2.29	8.7	<0.01
Length24Jul12	Family_2	sire	Co07	41.45	Co32046_Co07u	28.69 - 211.09	3.75	18.65	<0.001
Length24Jul12	Family_2	sire	Sex	NA	NA	NA	1.68	7.89	<0.05
Length24Jul12	Family_2	sire	Co07 x Sex	NA	NA	NA	1.04	6.52	<0.05

Length11Sep12	Family_2	sire	Co07	155.69	Co12537_Co07b	29.69 - 211.09	3.99	19.82	<0.001
Length11Sep12	Family_2	sire	Sex	NA	NA	NA	1.26	5.77	<0.1
Length11Sep12	Family_2	sire	Co07 x Sex	NA	NA	NA	1.17	5.37	<0.005
Length15Oct12	Family_2	sire	Co07	40.69	Co32046_Co07u	30.29 - 211.09	3.33	14.12	<0.001
Length15Oct12	Family_2	sire	Co30	137.92	Co13317_Co30	79.92 - 186.07	4.12	17.88	<0.001
Length28Nov12	Family_2	sire	Co07	183.61	Co35895_Co07b	32.69 - 211.09	3.32	13.25	<0.001
Length28Nov12	Family_2	sire	Co30	142.92	Co13317_Co30	96.92 - 186.07	5.05	21.14	<0.001
Weight23Aug11	Family_2	sire	Co07	210.95	Co69743_Co07b	36.69 - 211.09	4.02	18.35	<0.001
Weight23Aug11	Family_2	sire	Sex	NA	NA	NA	2.33	10.18	<0.01
Weight23Aug11	Family_2	sire	Co07 x Sex	NA	NA	NA	2.33	10.09	<0.01
Weight04Oct11	Family_2	sire	Co07	210.95	Co69743_Co07b	39.69 - 211.09	6.06	24.57	<0.001
Weight04Oct11	Family_2	sire	Sex	NA	NA	NA	2.71	10.15	<0.01
Weight04Oct11	Family_2	sire	Co07 x Sex	NA	NA	NA	2.29	8.46	<0.01
Weight15Nov11	Family_2	sire	Co07	191.69	Co26454_Co07b	27.69 - 211.09	3.13	13.16	<0.001
Weight15Nov11	Family_2	sire	Sex	NA	NA	NA	2.5	10.35	<0.01
Weight15Nov11	Family_2	sire	Co07 x Sex	NA	NA	NA	2.44	10.12	<0.01
Weight28Dec11	Family_2	sire	Co07	199.37	Co33887_Co07b	28.69 - 211.09	3.07	12.93	<0.01
Weight28Dec11	Family_2	sire	Sex	NA	NA	NA	2.4	9.97	<0.01
Weight28Dec11	Family_2	sire	Co07 x Sex	NA	NA	NA	2.37	9.84	<0.01
Weight01May12	Family_2	sire	Co07	188.69	Co03559_Co07b	38.69 - 198.69	5.16	23.42	<0.001
Weight01May12	Family_2	sire	Co23	81.33	Co38954_Co23	49.33 - 117.33	4.44	19.72	<0.001
Weight01May12	Family_2	sire	Sex	NA	NA	NA	2.08	8.63	<0.05
Weight01May12	Family_2	sire	Co07 x Co23	NA	NA	NA	1.34	5.43	<0.05
Weight01May12	Family_2	sire	Co23 x Sex	NA	NA	NA	2.05	8.48	<0.01
Weight12Jun12	Family_2	sire	Co07	156.69	Co36884_Co07b	39.69 - 198.69	6.6	28.11	<0.001
Weight12Jun12	Family_2	sire	Co23	115.27	Co43125_Co23	60.33 - 117.33	5.53	22.81	<0.001
Weight12Jun12	Family_2	sire	Sex	NA	NA	NA	2.24	8.34	<0.01
Weight12Jun12	Family_2	sire	Co07 x Co23	NA	NA	NA	1.73	6.4	<0.01
Weight12Jun12	Family_2	sire	Co23 x Sex	NA	NA	NA	2.13	7.96	<0.01
Weight24Jul12	Family_2	sire	Co07	211.09	Co39902_Co07b	28.69 - 211.09	3.81	18.97	<0.001
Weight24Jul12	Family_2	sire	Sex	NA	NA	NA	1.94	9.18	<0.05

Weight24Jul12	Family_2	sire	Co07 x Sex	NA	NA	NA	1.64	7.67	<0.01
Weight11Sep12	Family_2	sire	Co07	41.45	Co32046_Co07u	25.69 - 211.09	3.55	17.84	<0.001
Weight11Sep12	Family_2	sire	Sex	NA	NA	NA	1.49	7.01	<0.05
Weight11Sep12	Family_2	sire	Co07 x Sex	NA	NA	NA	1.24	5.83	<0.05
Weight15Oct12	Family_2	sire	Co07	41.45	Co32046_Co07u	24.69 - 211.09	3.05	13.49	<0.001
Weight15Oct12	Family_2	sire	Co30	141.92	Co13317_Co30	74.92 - 186.07	3.38	15.11	<0.001
Weight28Nov12	Family_2	sire	Co30	127.92	Co26282_Co30	92.92 - 176.92	3.94	22.49	<0.001
Weight28Nov12	Family_2	sire	Sex	NA	NA	NA	1.91	10.19	<0.05
Weight28Nov12	Family_2	sire	Co30 x Sex	NA	NA	NA	1.89	10.08	<0.01
DGR23Aug11_04Oct11	Family_2	sire	Co07	197.69	Co59243_Co07b	36.69 - 211.09	5.15	22.61	<0.001
DGR23Aug11_04Oct11	Family_2	sire	Sex	NA	NA	NA	2.11	8.548	<0.01
DGR23Aug11_04Oct11	Family_2	sire	Co07 x Sex	NA	NA	NA	1.33	5.27	<0.05
DGR04Oct11_15Nov11	Family_2	sire	Co05	246.31	Co41392_Co05b	63.41 - 247.56	3.7	15.94	<0.001
DGR04Oct11_15Nov11	Family_2	sire	Sex	NA	NA	NA	2.14	8.9	<0.01
DGR04Oct11_15Nov11	Family_2	sire	Co05 x Sex	NA	NA	NA	2.13	8.82	<0.01
DGR01May12_12Jun12	Family_2	sire	Co07	155.69	Co12537_Co07b	26.69 - 211.09	8.73	18.5	<0.001
DGR01May12_12Jun12	Family_2	sire	Sex	NA	NA	NA	4.73	10.02	<0.05
DGR01May12_12Jun12	Family_2	sire	Co07 x Sex	NA	NA	NA	7.09	7.51	<0.01
DGR11Sep12_15Oct12	Family_2	sire	Co28	114.34	Co26326_Co28	97.04 - 159.04	4.36	21.93	<0.001
DGR11Sep12_15Oct12	Family_2	sire	Sex	NA	NA	NA	4.26	21.4	<0.001
DGR11Sep12_15Oct12	Family_2	sire	Co28 x Sex	NA	NA	NA	4.25	21.29	<0.001
Length04Oct11	Family_3	sire	Co21	36.47	Co25552_Co21	11.04 - 106.07	3.47	27.26	<0.001
Length04Oct11	Family_3	sire	Sex	NA	NA	NA	0.91	6.13	<0.05
Length15Nov11	Family_3	sire	Co21	36.47	Co25552_Co21	18.04 - 41.04	4.59	33.78	<0.001
Length15Nov11	Family_3	sire	Sex	NA	NA	NA	2.03	13.22	<0.01
Length15Nov11	Family_3	sire	Co21 x Sex	NA	NA	NA	1.02	6.32	<0.01
Length28Dec11	Family_3	sire	Co21	36.47	Co25552_Co21	10.04 - 105.04	3.51	26.81	<0.001
Length28Dec11	Family_3	sire	Sex	NA	NA	NA	0.91	6.13	<0.05
Length07Feb12	Family_3	sire	Co21	36.47	Co25552_Co21	4.04 - 106.07	2.72	21.81	<0.001
DGR15Nov11_28Dec11	Family_3	dam	Co01	142.21	Co44266_Co01u	124.79 - 152.79	3.61	29.76	<0.001
Length23Aug11	Family_4	dam	Co03	212.31	Co58998_Co03u	194.04 - 220.04	4.91	23.48	<0.001

Length23Aug11	Family_4	dam	Sex	NA	NA	NA	4.15	19.42	<0.001
Length23Aug11	Family_4	dam	Co03 x Sex	NA	NA	NA	1.77	7.71	<0.01

Supplementary Material

Table S3.1 - Summary of the phenotypes measured in male and female offspring from each family. Mean and standard deviation (SD) for male and female, as well as p-values obtained from the F-tests and T-tests assuming equal or non-equal variance, are shown.

Table S3.2 - Phenotypic correlations among the measured traits. Upper triangular matrix shows phenotypic correlations, and lower triangular matrix shows corresponding p-values. P-values less than 0.05 are highlighted in red. The following tables show the degree of phenotypic correlations among traits for either female or male progeny in individual families:

Table S3.2.a - Female offspring in Family_1
Table S3.2.b - Male offspring in Family_1
Table S3.2.c - Female offspring in Family_2
Table S3.2.d - Male offspring in Family_2
Table S3.2.e - Female offspring in Family_3
Table S3.2.f - Male offspring in Family_3
Table S3.2.g - Female offspring in Family_4
Table S3.2.h - Male offspring in Family_4

Figure S3.1 - Correlation matrix showing the degree of correlations among measured traits for female and male offspring in each family. Higher positive correlations are depicted by narrower ellipses with darker blue shading, whereas lower positive correlations are depicted by broader ellipses with lighter blue shading. Higher negative correlation are depicted by narrower ellipses with darker red shading, whereas lower negative correlation are depicted by broader ellipses with lighter red shading. Correlations that are close to zero are shown as open, unfilled ellipses that are nearly circular. The following graphs illustrate the degree of phenotypic correlations among traits for either female or male progeny in individual families:

Figure S3.1.a - Female offspring in Family_1
Figure S3.1.b - Male offspring in Family_1
Figure S3.1.c - Female offspring in Family_2
Figure S3.1.d - Male offspring in Family_2
Figure S3.1.e - Female offspring in Family_3
Figure S3.1.f - Male offspring in Family_3
Figure S3.1.g - Female offspring in Family_4
Figure S3.1.h - Male offspring in Family_4

Figure S3.2 - Trend in average of length, weight and daily growth rate measured over time in female and male offspring for each family. Individual measures are depicted by open circular dots. Means are represented by black circular dots, and standard deviations are depicted by black bars. The following graphs illustrate individual trait changes for either female or male progeny in each family:

Figure S3.2.a - Length for female offspring in Family_1
Figure S3.2.b - Weight for female offspring in Family_1
Figure S3.2.c - Daily growth rate for female offspring in Family_1

Figure S3.2.d - Length male offspring in Family_1
Figure S3.2.e - Weight for male offspring in Family_1
Figure S3.2.f - Daily growth rate for male offspring in Family_1
Figure S3.2.g - Length female offspring in Family_2
Figure S3.2.h - Weight for female offspring in Family_2
Figure S3.2.i - Daily growth rate for female offspring in Family_2
Figure S3.2.j - Length male offspring in Family_2
Figure S3.2.k - Weight for male offspring in Family_2
Figure S3.2.l - Daily growth rate for male offspring in Family_2
Figure S3.2.m - Length female offspring in Family_3
Figure S3.2.n - Weight for female offspring in Family_3
Figure S3.2.o - Daily growth rate for female offspring in Family_3
Figure S3.2.p - Length male offspring in Family_3
Figure S3.2.q - Weight for male offspring in Family_3
Figure S3.2.r - Daily growth rate for male offspring in Family_3
Figure S3.2.s - Length female offspring in Family_4
Figure S3.2.t - Weight for female offspring in Family_4
Figure S3.2.u - Daily growth rate Female offspring in Family_4
Figure S3.2.v - Length male offspring in Family_4
Figure S3.2.w - Weight for male offspring in Family_4
Figure S3.2.x - Daily growth rate for male offspring in Family_4

Figure S3.3 - Trend in average values of length, weight and daily growth rate for early- and late-maturing female and male offspring for Family_1 and Family_2. Pink and blue lines represent early- and late-maturing individuals, respectively. Error bars represent standard deviation for each trait. The following graphs illustrate individual trait changes for either early- or late maturing female and male progeny in Family _1 and Family_2:

Figure S3.3.a - Length in Family_1
Figure S3.3.b - Weight in Family_1
Figure S3.3.c - Daily growth rate in Family_1
Figure S3.3.d - Length in Family_2
Figure S3.3.e - Weight in Family_2
Figure S3.3.f - Daily growth rate in Family_2

References

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

Beavis, W. D., 1998 QTL analyses: power, precision, and accuracy, pp. 145-162 in *Molecular Dissection of Complex Traits (ed. Paterson, A H)*, edited by C. Press. Boca Raton, FL.

Beavis, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies., pp. 250-266 in *Proceedings of the Annual Corn and Sorghum Industry Research Conference*, Chicago, IL, USA.

Bell, A. M., and N. Aubin-Horth, 2010 What can whole genome expression data tell us about the ecology and evolution of personality? Philos. Trans. R. Soc. B Biol. Sci. 365: 4001-4012.

Van den Berghe, E. P., and M. R. Gross, 1984 Female size and nest depth in coho salmon (*Oncorhynchus kisutch*). Can. J. Fish. Aquat. Sci. 41: 204-206.

Brieuc, M. S. O., C. D. Waters, J. E. Seeb, and K. A. Naish, 2014 A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication Event. G3-Genes. Genom. Genet. 4: 447-460.

Broman, K. W., and S. Sen, 2009 *A guide to QTL mapping with R/qtl*. Springer, New York.

Broman, K. W., H. Wu, S. Sen, and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889-890.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait, 2011 Stacks: building and genotyping Loci de novo from short-read sequences. G3-Genes. Genom. Genet. 1: 171-182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013 Stacks: an analysis tool set for population genomics. Mol. Ecol. 22: 3124-3140.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics. 971: 963-971.

Devlin, R. H., C. a Biagi, and D. E. Smailus, 2001 Genetic mapping of Y-chromosomal DNA markers in Pacific salmon. Genetica. 111: 43-58.

Dickerson, B. R., T. P. Quinn, and M. F. Willson, 2002 Body size, arrival date, and reproductive success of pink salmon, *Oncorhynchus gorbuscha*. Ethol. Ecol. Evol. 14: 29-44.

Dittmar, E. L., C. G. Oakley, J. Agren, and D. W. Schemske, 2014 Flowering time QTL in natural populations of *Arabidopsis thaliana* and implications for their adaptive value. Mol. Ecol. 23: 4291-4303.

Doerge, R. W., 2002 Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. 3: 43-52.

Doerge, R. W., and G. A. Churchill, 1996 Permutation tests for multiple loci affecting a quantitative character. Genetics. 142: 285-294.

Fabian, B., and T. Flatt, 2012 Life history evolution. Nat. Educ. Knowl. 3: 24.

Flatt, T., and A. Heyland, 2011 *Mechanisms of life history evolution: the genetics and physiology of life history traits and trade-offs*. Oxford University Press, Oxford, UK.

Fleming, I. A., 1998 Pattern and variability in the breeding system of Atlantic salmon (*Salmo salar*), with comparisons to other salmonids. Can. J. Fish. Aquat. Sci. 55: 59-76.

Gnan, S., A. Priest, and P. Kover, 2014 The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. Genetics. 1-34.

Gross, M. R., 1985 Disruptive selection for alternative life histories in salmon. Nature. 313: 47-48.

Haidle, L., J. E. Janssen, K. Gharbi, H. K. Moghadam, M. M. Ferguson *et al.*, 2008 Determination of quantitative trait loci (QTL) for early maturation in rainbow trout (*Oncorhynchus mykiss*). Mar. Biotechnol. 10: 579-592.

Haselhorst, M. S. H., C. E. Edwards, M. J. Rubin, and C. Weinig, 2011 Genetic architecture of life history traits and environment-specific trade-offs. Mol. Ecol. 20: 4042-4058.

Hayes, B. J., A. Gjuvsland, and S. Omholt, 2006 Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. Heredity. 97: 19-26.

Hendry, A. P., and S. C. Stearns, 2004 *Evolution illuminated: salmon and their relatives*. Oxford University Press, Oxford.

Holtby, L. B., and M. C. Healey, 1990 Sex-specific life history tactics and risk-taking in coho salmon. Ecology. 71: 678-690.

Houston, R. D., S. C. Bishop, A. Hamilton, D. R. Guy, A. E. Tinch *et al.*, 2009 Detection of QTL affecting harvest traits in a commercial Atlantic salmon population. Anim. Genet. 40: 753-755.

Hutchings, J. A., and M. E. B. Jones, 1998 Life history variation and growth rate thresholds for maturity in Atlantic salmon, *Salmo salar*. Can. J. Fish. Aquat. Sci. 55: 22-47.

Kim, S.-Y., J. C. Noguera, J. Morales, and A. Velando, 2010 Quantitative genetic evidence for trade-off between growth and resistance to oxidative stress in a wild bird. Evol. Ecol. 25: 461-472.

Klemetsen, A., P.-A. Amundsen, J. B. Dempson, B. Jonsson, N. Jonsson *et al.*, 2003 Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr *Salvelinus alpinus (L.)*: a review of aspects of their life histories. Ecol. Freshw. Fish 12: 1-59.

Kodama, M., M. S. O. Brieuc, R. H. Devlin, J. J. Hard, and K. A. Naish, 2014 Comparative mapping between coho salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. G3-Genes. Genom. Genet. 4: 1717-1730.

Kodama, M., J. J. Hard, and K. A. Naish, 2012 Temporal variation in selection on body length and date of return in a wild population of coho salmon, *Oncorhynchus kisutch*. BMC Evol. Biol. 12: 116.

Küttner, E., H. K. Moghadam, S. Skúlason, R. G. Danzmann, and M. M. Ferguson, 2011 Genetic architecture of body weight, condition factor and age of sexual maturation in Icelandic Arctic charr (*Salvelinus alpinus)*. Mol. Genet. Genomics 286: 67-79.

Latta, R. G., and K. M. Gardner, 2009 Natural selection on pleiotropic quantitative trait loci affecting a life-history trade-off in *Avena barbata*. Evolution. 63: 2153-2163.

Lee, W. S., P. Monaghan, and N. B. Metcalfe, 2013 Experimental demonstration of the growth rate-lifespan trade-off. Proc. R. Soc. B Biol. Sci. 280: 20122370.

Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, MA.

Mackay, T. F. C., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. 35: 303-339.

Mackay, T. F. C., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10: 565-577.

Martyniuk, C. J., G. M. L. Perry, H. K. Mogahadam, M. M. Ferguson, and R. G. Danzmann, 2003 The genetic architecture of correlations among growth-related traits and male age at maturation in rainbow trout. J. Fish Biol. 63: 746-764.

McClelland, E. K., J. M. Myers, J. J. Hard, L. K. Park, and K. A. Naish, 2005 Two generations of outbreeding in coho salmon (*Oncorhynchus kisutch*): effects on size and growth. Can. J. Fish. Aquat. Sci. 62: 2538-2547.

McClelland, E. K., and K. A. Naish, 2010 Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. Heredity. 105: 562-573.

Mitchell-Olds, T., 1996 Genetic constraints on life-history evolution: quantitative-trait loci influencing growth and flowering in *Arabidopsis thaliana*. Evolution. 50: 140-145.

Moghadam, H. K., J. Poissant, H. Fotherby, L. Haidle, M. M. Ferguson *et al.*, 2007 Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). Mol. Genet. Genomics. 277: 647-661.

Morita, K., and M. Fukuwaka, 2006 Does size matter most? The effect of growth history on probabilistic reaction norm for salmon maturation. Evolution. 60: 1516-1521.

Myers, J. M., P. O. Heggelund, G. Hudson, and R. N. Iwamoto, 2001 Genetics and broodstock management of coho salmon. Aquaculture 197: 43-62.

Nichols, K. M., A. F. Edo, P. A. Wheeler, and G. H. Thorgaard, 2008 The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. Genetics. 179: 1559-1575.

Norry, F. M., F. H. Gomez, and V. Loeschcke, 2007 Knockdown resistance to heat stress and slow recovery from chill coma are genetically associated in a quantitative trait locus region of chromosome 2 in *Drosophila melanogaster*. Mol. Ecol. 16: 3274-3284.

Olson, Å., A. Aerts, and F. Asiegbu, 2012 Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. New. Phytol. 194: 1001-1013.

Pedersen, S., P. R. Berg, M. Culling, R. G. Danzmann, B. Glebe *et al.*, 2013 Quantitative trait loci for precocious parr maturation, early smoltification, and adult maturation in double-backcrossed trans-Atlantic salmon (*Salmo salar*). Aquaculture. 410-411: 164-171.

Quinn, T. P., 2005 *The behavior and ecology of Pacific salmon and trout.* University of Washington Press, Seattle.

Roff, D. A., 2007 Contributions of genomics to life-history theory. Nat. Rev. Genet. 8: 116-125.

Roff, D. A., 1992 *The Evolution of life histories: theory and analysis*. Chapman and Hall, New York.

Sakamoto, T., R. G. Danzmann, K. Gharbi, P. Howard, A. Ozaki *et al.*, 2000 A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics. 155: 1331-45.

Sandercock, F. K., 1991 *Life history of coho salmon (Oncorhynchus kisutch).* University of British Columbia Press, Vancouver, Canada.

Sen, S., and G. A. Churchill, 2001 A statistical framework for quantitative trait mapping. Genetics. 159: 371-387.

Shearer, K. D., and P. Swanson, 2000 The effect of whole body lipid on early sexual maturation of 1+ age male chinook salmon (*Oncorhynchus tshawytscha*). Aquaculture. 190: 343-367.

Shook, D. R., and T. E. Johnson, 1999 Quantitative trait loci affecting survival and fertility-related traits in Caenorhabditis elegans show genotype-environment interactions, pleiotropy and epistasis. Genetics. 153: 1233-1243.

Silverstein, J. T., K. D. Shearer, W. W. Dickhoff, and E. M. Plisetskaya, 1998 Effects of growth and fatness on sexual development of chinook salmon (*Oncorhynchus tshawytscha*) parr. Can. J. Fish. Aquat. Sci. 2382: 2376-2382.

Stearns, S. C., 1992 *The evolution of life Histories*. Oxford University Press, New York.

Stearns, S. C., 1989 Trade-offs in life-history evolution. Funct. Ecol. 3: 259-268.

Steen, R. P., and T. P. Quinn, 1999 Egg burial depth by sockeye salmon (*Oncorhynchus nerka*): implications for survival of embryos and natural selection on female body size. Can. J. Zool. 77: 836-841.

Thorpe, J. E., M. Mangel, N. B. Metcalfe, and F. A. Huntingford, 1998 Modelling the proximate basis of salmonid life-history variation, with application to Atlantic salmon, *Salmo salar L*. Evol. Ecol. 12:581-599

Thorpe, J. E., R. I. G. Morgan, C. Talbot, and M. S. Miles, 1983 Inheritance of developmental rates in Atlantic salmon, *Salmo salar L*. Aquaculture. 33: 119-128.

Todesco, M., S. Balasubramanian, T. T. Hu, M. B. Traw, M. Horton *et al.*, 2010 Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. Nature. 465: 632-636.

Vasemägi, A., R. Gross, D. Palm, T. Paaver, and C. R. Primmer, 2010 Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. BMC. Genomics. 11: 156.

Waples, R., T. Beechie, and G. R. Pess, 2009 Evolutionary history, habitat disturbance regimes, and anthropogenic changes: what do these mean for resilience of Pacific Salmon populations? Ecol. Soc. 14: 3.

Waples, R. S., R. G. Gustafson, L. A. Weitkamp, J. M. Myers, O. W. Johnson *et al.*, 2001 Characterizing diversity in salmon from the Pacific Northwest. J. Fish Biol. 59: 1-41.

Wringe, B. F., R. H. Devlin, M. M. Ferguson, H. K. Moghadam, D. Sakhrani *et al.*, 2010 Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). BMC Genet. 11: 63.

Yano, A., B. Nicol, E. Jouanno, E. Quillet, A. Fostier *et al.*, 2013 The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. Evol. Appl. 6: 486-496.

Chapter 4: Population genomic analyses of domesticated coho salmon (*Oncorhynchus kisutch*) reveals loci involved in driving adaptive phenotypic differentiation following intensive selective breeding

Abstract

Characterizing the genomic basis of domestication selection helps understand the mechanism of adaptive evolution and the extent to which selection causes divergence between domesticated and wild animals. The aims of this study were to identify loci that were under the direct influence of domestication selection and to investigate how adaptive phenotypic responses were determined at specific sites in the genome. To achieve these aims, we surveyed loci and genomic regions for evidence of elevated divergence by directly comparing the domesticated line of coho salmon to its hatchery and wild source populations, and investigated whether there was evidence for selection at genomic regions linked with selectively bred traits during domestication. Our results revealed the extreme divergence in the domesticated population compared to the hatchery and the wild populations, and multiple analyses detected 7 genomic regions exhibiting high differentiation; of these, strong signatures of selection were observed on a region linked with age at sexual maturity, as well as two regions linked with growth that span the key maturation decision periods. Together, these results may be attributable to strong selection for rapid growth and early maturity in aquaculture, and provided insights into how and to what extent adaptive evolutionary changes result from selective breeding in a captive environment.

Introduction

Domestication leads to morphological, physiological, behavioral and reproductive changes, distinguishing domesticated individuals from their wild counterparts (Darwin 1859). Since Darwin, there has been an increasing interest in studying domesticated animal and crop species as a means of understanding the nature of selection. Domestication of plant and animal species related to food production began as early as in the late Pleistocene (Larson *et al.* 2014). In contrast, domestication of aquatic species such as salmon and trout has started more recently in the early 1900s (Gjøen and Bentsen 1997; Hutchings and Fraser 2008). In most cases, deliberate, continuous selection focusing on the development or enhancement of desired traits widely altered the phenotypic characteristics of these species (FAO 2007). Altogether, domesticated populations are excellent models to study the direct and indirect results of concerted selection on specific phenotypes, and to demonstrate how phenotypic responses are determined at specific sites in the genome, thereby enhancing the understanding of how inheritance facilitates or constrains adaptation.

Domestication is often characterized by genome-wide loss of genetic diversity, which initially results from small initial populations relative to their wild conspecifics (Eyre-Walker *et al.* 1998); subsequently, intense artificial selection targeting desirable traits causes genetic bottlenecks in each generation, further reducing genetic diversity throughout the genome (Doebley *et al.* 2006). Genes affecting desirable traits are more strongly influenced by selective breeding, thus being particularly prone to a drastic loss of genetic diversity compared to the rest

of the genome (Wright *et al.* 2005; Doebley *et al.* 2006); this outcome is because individuals with favorable alleles contribute more progeny to each subsequent generation, whereas individuals with other alleles are deliberately removed from the population (Wright *et al.* 2005; Doebley *et al.* 2006). Such "domestication genes" or candidate genes that fall within genomic regions that have been targeted by selection during domestication are increasingly identified in a number of animal and crop species with a large amount of genetic resources (Akey *et al.* 2010; Rubin *et al.* 2010; Rubin *et al.* 2012; Larson and Burger 2013; Olsen and Wendel 2013; Carneiro *et al.* 2014).

Although genome-wide polymorphism data are required to link genomic regions to desired traits and to find regions or genes implicated in domestication, obtaining sufficient markers has been a limiting step in such endeavors for non-model organisms (Slate *et al.* 2009; Davey *et al.* 2011; Barrett and Hoekstra 2011). However, the recent advent of Genotype-by-Sequencing technologies has made it possible to obtain genome-wide sequence data in almost any organisms of interest, even for species or populations with little or no genetic information available (Baird *et al.* 2008; Hohenlohe, Phillips, *et al.* 2010; Davey *et al.* 2011; Ekblom and Galindo 2011). Using such data, the patterns of genome-wide polymorphisms can be surveyed using samples from multiple populations to identify genomic regions that may have been influenced by certain selective pressures. This "genome-scan" approach reveals genomic regions of increased differentiation, indicating that these regions may have been subjected to selection (Naish and Hard 2008; Hohenlohe, Bassham, *et al.* 2010). Nevertheless, the function of these regions needs to be confirmed to demonstrate that the increased differentiation resulted from adaptive phenotypic changes via selection (Luikart *et al.* 2003). To achieve this goal, information on the genetic architecture of a given trait is essential, such as genes or loci that shape traits under selection. Combined, these analyses provide insights into how phenotypic differentiation resulting from domestication selection is determined at specific sites in the genome (e.g. Wilkinson *et al.* 2013; Carneiro *et al.* 2014; Sun *et al.* 2014).

The farming of salmon and trout species have rapidly developed since mid-1900s, and breeding programs are globally implemented to support the world-wide production of domesticated salmon (Gjøen and Bentsen 1997; Skaala *et al.* 2004; Ferguson *et al.* 2007). Selection for increased growth rate and body weight, as well as for early age at sexual maturity has resulted in rapid phenotypic differentiation (5-15% selection response per generation; Gjøen and Bentsen 1997; Thodesen *et al.* 1999; Myers *et al.* 2001; Martinez *et al.* 2006). Such rapid responses are likely to be determined by genomic divergence of farmed stocks compared to their wild origin (Karlsson *et al.* 2011), and there has been an increasing interest in linking genetic divergence with phenotypic differentiation to identify targets of selection in the genome. To date, a number of studies have compared farmed and wild salmon populations and identified genomic regions of high divergence that may have been influenced by domestication selection (Bourret *et al.* 2011; Karlsson *et al.* 2011; Vasemägi *et al.* 2012; Martinez *et al.* 2013; Mäkinen *et al.* 2014). However, studies that have attempted to link such genomic regions with phenotypic differentiation are scarce (e.g. Martinez *et al.* 2013), as prior knowledge on the genetic architecture of selectively bred traits is essential to achieve this goal. Quantitative trait loci (QTL) linked with selectively bred traits in an aquaculture strain of coho salmon have been

recently identified (Kodama *et al.* in prep). This information provides an opportunity to reveal loci that are associated with phenotypic differentiation that have resulted from selective breeding in a captive environment. Such knowledge is essential to improve our understanding of the mechanism and the extent to which selection causes divergence between domesticated and source populations.

The overall aim of this study was to identify the loci that played a role in driving adaptive phenotypic evolution in response to selective breeding. To achieve this aim, we surveyed loci and genomic regions for evidence of elevated divergence in a domesticated line of coho salmon compared to its hatchery and wild source populations. Specifically, a domesticated aquaculture strain that has been selectively bred for large size and early maturation (Myers *et al.* 2001) was available for this study, providing an exciting opportunity to learn how deliberate selection on particular phenotypes promotes adaptation. Genomic regions of elevated divergence were compared to QTL linked with selectively bred traits that have been mapped in an experimental outbred cross using the same aquaculture strain. Specifically, we examined divergence at genomic regions linked to age at sexual maturity, as well as weights, body lengths and daily growth rates at multiple time points (Kodama *et al.* in prep) and searched for evidence of selection at these QTL-linked loci.

Materials and Methods

Study system and sample collection

The data comprised coho salmon collected from three distinct populations from Washington State (USA): namely, the South Fork Skykomish River wild population, Wallace River Hatchery and Domsea broodstock. The wild population of the South Fork Skykomish River, a tributary to the Snohomish River that drains into Puget Sound, was used to found the Washington Department of Fish and Wildlife's (WDFW) Wallace River Hatchery (47º87'N, 121º71'W) in the early-1900s. Subsequently, Wallace River Hatchery coho was used to found a Domsea aquaculture line in 1971 and 1972. The Domsea broodstock is a commercial line, and it has been selectively bred for early maturation and large size since 1977 (Myers *et al.* 2001). Although Wallace River Hatchery coho had been segregated from the wild population since its founding, gene flow between the hatchery and wild individuals from the South Fork Skykomish River has been intentionally encouraged since 2005 (Washington Department of Fish and Wildlife 2013).

A total of 118 individuals were collected from the Domsea aquaculture line in 2006. Ninety eight individuals were collected from Wallace River Hatchery in 2002, prior to the integration of wild individuals from the South Fork Skykomish River. Ninety five individuals were collected from the wild population of the South Fork Skykomish River in November 2012 prior to spawning; these wild fish would have been subjected to two to three generations of gene flow from the hatchery population. Altogether, the data set represents temporal samples that have been under different selective pressures, and allows the identification of genomic regions that might have responded to domestication selection. Fin clips were stored in 100% ethanol.

DNA extraction, sequencing and genotyping

Genomic DNA from each fin clip was extracted using the DNeasy extraction kit (QIAGEN, Valencia, CA, USA) following the manufacturer's procedures. DNA was digested with the restriction enzyme *Sbf*I. Following this step, a unique, 6-nucleotide barcode was added to each sample for individual identification. Thirty six individuals were pooled in a single library and sequenced using the Illumina HiSeq 2000 sequencer, producing 100 bp single-read lengths.

DNA sequences were processed, and the barcodes were used to separate individual sequences using the *process_radtags* function implemented in STACKS (v. 1.09; Catchen *et al.* 2011; Catchen *et al.* 2013). Sequences from all individuals were trimmed to 74 nucleotides to remove low-quality sequences and aligned to the reference database for coho salmon, comprising 48,528 non-duplicated RAD loci developed by Kodama *et al.* (2014), using BOWTIE (Langmead *et al.* 2009). A maximum of three nucleotide mismatches per locus were allowed during alignment. Subsequently, polymorphic loci were identified using STACKS, and genotypes at these loci were determined when alleles were sequenced with a depth greater than 10X per individual. Due to the potential bias against heterozygous genotypes, genotypes were corrected after running STACKS using the Python script developed by Brieuc *et al.* (2014); specifically, individuals were determined as heterozygotes at each locus if both alleles had a depth of more than two and the total read depth was 10X or greater. Loci that aligned to the coho salmon linkage map (Kodama *et al.* 2014) and contained two alleles were retained for further analyses. Among these loci, loci with a minor allele frequency less than 5% across all populations were removed. Subsequently, individuals with more than 50% missing genotypes were removed.

Standard population genetic and population structure analyses

Observed and expected heterozygosity for each locus in each population was calculated using the R-package Adegenet v. 1.4-2 (Jombart 2008). GenePop v. 4.2 was used to perform exact tests for deviations from Hardy-Weinberg equilibrium and to calculate population pairwise F_{ST} values (Weir and Cockerham 1984; Raymond and Rousset 1995). Significance tests for population pairwise F_{ST} values were performed in GenePop using the default parameters.

A Principal Components Analysis (PCA) was conducted in Adegenet v. 1.4-2 (Jombart 2008) using all individuals to investigate the relationship among the three populations. Missing genotypes were first replaced by the scaled and centered allele frequencies using the *scaleGen* function, and PCA was performed using the *dudi.pca* function. Bayesian clustering in BAPS (Corander and Marttinen 2006) was also performed to examine the number of groups (K) present within the three populations. BAPS employs a stochastic optimization procedure to infer the optimal number of groups, and K ranging from 2 to 20 was used. Results from both PCA and BAPS showed that individuals from Wallace River Hatchery and the South Fork Skykomish River form one cluster. Because of low DNA quality and large amount of missing genotypes, fish from Wallace River Hatchery were removed from downstream analyses.

Relatedness within each population

Results from PCA and BAPS indicated that groups of closely related individuals may be present within the Domsea population. Because oversampling of related individuals is known to cause bias regarding higher F_{ST} values (Gross *et al.* 2011), which may further affect the detection of outlier loci and genomic regions exhibiting high divergence, related individuals were identified in all three populations. To examine if the data provide enough power to separate different relationships, we first simulated four predefined relationships of relatedness within the Domsea population, 0 (unrelated), 0.125 (double first cousins), 0.25 (halfsibs) and 0.5 (fullsibs) using the observed allele frequencies in Coancestry v1.0.1.5 (Wang 2011). Three sets of loci with two alleles with frequencies ranging from 0.05-0.95, 0.35-0.65 and 0.225-0.725 were used to perform the simulations and to identify which set of markers are the most powerful for distinguishing different relationships. A simulation for each set of loci was conducted with and without loci that mapped to the same position because relatedness estimators do not account for linkage between markers (Santure *et al.* 2010). An error rate of 0 and missing value of 0, as well as an error rate of 0.01 and missing value of 0.01 were used for each run to compare the results and examine the effect of these factors on the relatedness estimates. Simulation results indicated that the loci with two alleles with frequencies ranging from 0.225-0.725 (1930 loci in total) led to best separation of these four relationships. With this set of loci, results were similar between the runs performed with different error rates and missing values, as well as with and without linkage. Therefore, these loci were used to obtain an empirical estimate for pairwise relatedness between individuals within each population using Coancestry v1.0.1.5 (Wang 2011). An error rate of 0.01 and missing value of 0.01 were used to perform this run. Subsequently, these loci were used to perform Bayesian clusters in BAPS using Domsea population alone; observed discrete clusters were compared to the pairwise relatedness estimates to determine if these clusters correspond to the groups of closely related individuals. To examine the influence of closely related individuals on population genetic statistics, overall population pairwise F_{ST}, as well as observed and expected heterozygosities were calculated using a subset of Domsea individuals equally representing each cluster observed in BAPS.

Detection of outlier loci, genomic regions of high divergence and comparison with previous QTL study

The program Arlequin ver. 3.5 (Excoffier and Lischer 2010) was used to identify outlier loci consistent with domestication selection in the aquaculture line. Pairwise comparisons of F_{ST} values were performed between Domsea and South Fork Skykomish wild coho, using 20,000 simulations with 100 demes. Loci significant at 95% level after Bonferroni correction were determined as outliers. Additionally, Bayescan ver. 2.1 (Foll and Gaggiotti 2008) was used to identify outlier loci with default parameters. A locus that had its q-value less than 0.05 was considered to be under selection. Although it is recommended to have more than 6 populations to have a relaistic true-positive rate in detecting loci under directional selection (Foll and Gaggiotti 2008), we used Bayescan because of its lower type I and type II errors compared to other programs (Narum and Hess 2011).

The Kernel-smoothing sliding window approach (Hohenlohe, Bassham, *et al.* 2010; Brieuc *et al.* 2015) was used to identify genomic regions of high divergence and examine whether the outlier loci identified by Arlequin or Bayescan were located within or near such genomic regions. Moving weighted averages of F_{ST} values were calculated at 3cM bandwidth intervals across each linkage group. A 95% confidence interval for F_{ST} values was determined using a bootstrap approach, in which F_{ST} values for all mapped loci were bootstrapped 100,000 within each window. Genomic regions where the Kernel-smoothed moving average of F_{ST} surpassed the 95% confidence interval in each population pairwise comparison were identified as regions of increased divergence.

To investigate how closely related individuals in the Domsea population influence the detection of highly differentiated genomic regions, outlier loci and regions of high divergence were also identified using a subset of Domsea individuals with lower pairwise relatedness values. Specifically, Arlequin, Bayescan and sliding window analyses were performed on these individuals against South Fork Skykomish wild coho, and results were compared to those obtained using all Domsea individuals.

Results from Arlequin, Bayescan and the kernel-smoothing analyses were compared to the QTL linked with selectively-bred traits to identify which genomic regions may have been subjected to domestication selection and played a role in driving adaptive evolution. QTL were identified in the experimental outbred crosses established from an F4 mapping population based on the Domsea broodstock and a distantly related wild population (Kodama *et al.* in prep). Specifically, an experimental outbred line was created with females from the Bingham Creek (47º15'N, 123º40'W), a tributary to the Satsop River that drains to the Southwest Washington coast, and males from the Domsea aquaculture line. These populations exhibited differentiation at growth-related traits such as length, weight and daily growth rate (McClelland *et al.* 2005). QTL linked with growth were identified, namely, length and weight temporally measured when fish were 8 to 24 months of age, daily growth rate across 10 time intervals and age at sexual maturity (Kodama *et al.* in prep). The comparisons were used to determine whether the loci linked to mapped QTL are the same loci or in the nearby region that showed increased differentiation, and thus may have responded to domestication selection.

Effective population size

Effective population size (N_e) was estimated using *N_eEstimator* v. 2.01 (Do *et al.* 2014) with the linkage disequilibrium (LD) method (Hill 1981; Waples 2006). This method estimates N_e based on the Burrow's composite measure of disequilibrium calculated for each locus pair; random mating between individuals was assumed and alleles with frequency below 0.05 were excluded. Because the LD method assumes that loci are physically unlinked, calculations between pairs of loci on the same chromosome were removed (Larson *et al.* 2013). Effective population size estimates and 95% confidence intervals were obtained for each population (Do *et al.* 2014).

Results

DNA sequencing and genotyping

A total of 118, 98, and 95 individuals from the Domsea aquaculture line, Wallace River Hatchery and South Fork Skykomish River wild coho, respectively, were sequenced. A total of 10612 loci with two alleles were identified; of these, 4925 loci were mapped to the coho salmon linkage map developed by Kodama *et al.* (2014), and 157 loci had a minor allele frequency less than 5% across all populations. These loci were removed, and 4768 loci were retained for further analyses. Among the 311 sequenced individuals, 118 individuals from Domsea, 37 individuals from Wallace River Hatchery, and 94 individuals from South Fork Skykomish River were sequenced at more than 50% of the retained loci. DNA degradation for Wallace River Hatchery samples was widely apparent, and a high number of missing genotypes was observed for this population.

Standard population genetic and population structure analyses

Overall observed/expected heterozygosity (Ho/He) was 0.266/0.254 for Domsea, 0.425/0.337 for Wallace River Hatchery, and 0.336/0.319 for South Fork Skykomish River (Supplementary Material S4.1). Among 14304 tests, significant deviation from Hardy-Weinberg equilibrium at $\alpha = 0.01$ was observed for 668 tests (4.67% of population/locus combinations; Supplementary Material S4.1). Overall population pairwise F_{ST} was 0.1353 for Domsea vs. Wallace River Hatchery, 0.1197 for Domsea vs. SF Skykomish River, and 0.0116 for Wallace River Hatchery vs. SF Skykomish River, all of which were highly significant at $\alpha < 0.001$.

A Principal Components Analysis (PCA) revealed two distinct clusters, clearly separating Domsea broodstock from Wallace River Hatchery and South Fork Skykomish wild coho (Figure 4.1). Bayesian clusters of individuals in BAPS were slightly different from the clustering observed in PCA. Bayesian clustering in BAPS found four discrete clusters; three clusters represented individuals from Domsea broodstock, whereas one cluster represented Wallace River Hatchery and South Fork Skykomish wild coho (Figure 4.2).

Relatedness within each population

Individual pairwise relatedness estimates obtained from Coancestry indicated that some individuals may be closely related. Specifically, this program produced 6903 pairwise relatedness estimates among Domsea individuals. These estimates ranged from -0.260 to 0.642, and the median was 0.024. Bayesian clusters in BAPS using Domsea individuals alone revealed 11 discrete groups, ranging from 4 to 32 individuals per group (Figure 4.3; Supplementary Material S4.2); the average relatedness for each group ranged from 0.376 to 0.506, except for one cluster with average relatedness of 0.095 (Figure 4.3). These results indicate that clusters observed within the Domsea broodstock may have been caused by the presence of closely related individuals. Fifty one Domsea individuals including all individuals from Cluster 6 (average relatedness = 0.095; Figure 4.3), as well as 2-3 individuals from each cluster was used to calculate the overall population pairwise F_{ST}, as well as observed/expected heterozygosities; these estimates were similar to those obtained using all Domsea individuals. Specifically, the

overall population pairwise F_{ST} was 0.1200 (compared to 0.1197) when the subset of the Domsea population was compared to the South Fork Skykomish River wild population, and this F_{ST} value was highly significant at $\alpha < 0.001$. Overall observed and expected heterozygosity (Ho/He) was 0.262 and 0.251 respectively. Downstream analyses were performed on all Domsea individuals, as well as on a subset of Domsea individuals to account for relatedness.

Coancestry produced 4371 pairwise relatedness estimates among individuals within the South Fork Skykomish wild population. The median of these estimates was 0.054, and estimates ranged from -0.187 to 0.397, indicating that related individuals are unlikely to be present in this population (Figure 4.3).

Detection of outlier loci, genomic region under selection and comparison with previous QTL study

Genomic regions exhibiting an increased level of divergence compared to neutral expectations were detected using three approaches. The pairwise population divergence approach of Arlequin detected 8 loci that were significant outliers at $\alpha = 0.05$ after Bonferroni correction; these outlier loci were distributed over 5 linkage groups (Supplementary Material S4.3). The Bayesian approach implemented in Bayescan detected 13 loci that were significant outliers at q-value < 0.05, and these loci were distributed over 9 linkage groups (Supplementary Material S4.4). The sliding window analysis revealed 25 genomic regions of high divergence distributed over 17 linkage groups. Of these, 7 regions contained outlier loci detected by Arlequin or Bayescan; evidence consistent with putative domestication selection on Co02, Co17 and Co18 was particularly well supported where outlier loci detected by both Arlequin and Bayescan overlapped with the sliding window analyses (Figure 4.4). When accounting for relatedness, a region located on Co21 did not show any signs of increased differentiation (Supplementary Material S4.5). Fewer loci were identified as outliers when accounting for relatedness (Supplementary Material S4.5). This result might reflect lower power because the data set was reduced. On the other hand, the use of related individuals can inflate F_{IS} within a population, and thus impact F_{ST} values (see discussion).

The comparison of outlier loci and genomic regions of high divergence with QTL linked with growth (Kodama *et al.* in prep) revealed evidence consistent with selection on these loci in the Domsea broodstock. In total, 3 QTL were found near the genomic regions of high divergence located on Co05, Co10 and Co21 (Figure 4.5). Interestingly, genomic regions exhibiting high levels of differentiation on Co10 were found in close proximity to the QTL linked with age at sexual maturity, indicating that these regions might have responded to selective breeding favoring early maturity, a trait that was selected on in this line. Additionally, QTL linked with daily growth rate at one time interval (when fish were 8 to 10 months of age) was located near the outlier region detected in the Domsea broodstock on Co05 (Figure 4.5). Finally, QTL linked with lengths that were expressed throughout the experimental duration (when fish were 8 to 14 months of age) were found on Co21 near the genomic region of high divergence and outliers, indicating the importance of this region during domestication; however, this observation warrants further investigation because no signs of increased differentiation for this region or the markers was observed when accounting for relatedness in the Domsea population (Figure 4.5).

Effective population size

Effective size, $N_{e,}$ for the Domsea broodstock and the South Fork Skykomish wild coho differed significantly. The estimate for Domsea was 23 (95% CI = 23 - 23), whereas N_e for the South Fork Skykomish wild was 2422 (95% CI = 2243 - 2632).

Discussion

This study aimed to identify the loci that played a role in driving adaptive phenotypic evolution to the domesticated culture environment in response to selective breeding by directly comparing the domesticated line of coho salmon to its hatchery and wild source populations. The identification of putative signatures of selection on traits was based on the placement of QTL on the linkage map, sliding window analyses and outlier detection using two different statistical methods. Results revealed regions in the coho salmon genome linked to traits that may have diverged through selective breeding on growth-related traits in the domesticated population. Here, we have demonstrated that a few regions with trait-linked loci have diverged likely due to selective breeding that favored accelerated growth and early maturity.

Our results revealed the extreme divergence in the Domsea broodstock compared to the Wallace River hatchery and the South Fork Skykomish wild populations. Divergence of the Domsea broodstock may have initially occurred when the founding population was derived from the Wallace River hatchery and subjected to mass selection for growth (Novotny 1975); subsequently, intense selection for early maturity and faster growth imposed on this population have likely caused inbreeding and genetic bottlenecks. Combined, these selection schemes may not only explain the genetic differentiation observed for this population, but also explains its extremely small effective population size. Despite the fact that hatchery-reared salmon may also be under selection caused by captive rearing, our results indicated that the hatchery and wild populations were genetically similar to each other. Intentional promotion of gene flow from the wild to hatchery populations has likely reduced genetic differentiation between these two populations; this supports the idea that gene flow is important in reducing differentiation, even under selection (e.g. Ford 2002; Waters *et al.* in prep).

There were a discrepancy in results obtained by the sliding window analyses and outlier detection using Arlequin or Bayescan when accounting for relatedness. Among the 7 genomic regions of high divergence supported by these analyses using the all Domsea individuals, the sliding window analyses identified the six genomic regions of high divergence when accounting for relatedness. Of those, Arlequin or Bayescan did not detect outliers for Co05, Co08, Co17 and Co18 when accounting for relatedness. Such results may be due to reduced power since less than a half of the Domsea samples (50 vs. 118 individuals) were included in the analyses when accounting for relatedness. These observations indicate that evaluations using more Domsea individuals may resolve this uncertainty and further elucidate which regions are indeed exhibiting differentiation at the population level.

Strong signatures of selection on a region linked with age at sexual maturity on Co10 indicate that this trait may be shaped by a few loci of relatively large effect. Wild coho salmon is known to have a simple age structure, in which most individuals mature at three years with a

fraction of males (and occasionally, females) maturing as two year olds. Mapping of QTL in Kodama *et al.* (in prep) revealed that age at sexual maturity in this species may be predominately affected by a few loci of moderate to large effect. The outlier loci on Co10 overlapped with one such locus; furthermore, strong signatures of selection were also observed on regions linked with growth that span the key maturation decision periods on Co05 and Co07 (Kodama *et al.* in prep). Together, these results may be attributable to strong selection for rapid growth and early maturity in aquaculture, and such results contribute to growing evidence that a few loci of large effect may be associated with response to selection in key phenotypes correlated with fitness.

Although not verified by the mapped QTL (Kodama *et al.* in prep), outlier regions detected on Co02, Co17 and Co18 could also contain adaptively significant variation. Mäkinen *et al.* (2014) demonstrated that quantitative trait architecture, the strength of selection, and the duration of domestication all contribute to a probability of detecting selection in a captively bred population; despite intensive selection, their results showed that the chance of detecting selection was low when a trait was coded by a higher number of QTL. Experimental limitations such as relatively small family sizes used in Kodama *et al.* (in prep) indicate that QTL with small effect may have remained undetected, and detected outlier regions may be linked with such QTL. However, given the strong signatures of selection observed on these regions, a more plausible possibility is that these regions are associated with traits that were not accounted for in Kodama *et al.* (in prep) and correlated with growth and maturity, and future efforts should be directed towards confirming the function of these regions and identifying their adaptive significance.

The distribution of outlier regions was not uniform among the linkage groups; in fact, all linkage groups harboring outlier regions are metacentric chromosomes (Kodama *et al.* 2014), and four of them are likely involved in recent or ongoing homeologous pairing (Co02, Co05, Co08, Co10; Kodama *et al.* 2014). Telomeric regions of chromosomal arms involved in homeologous pairing mainly consist of duplicated markers (Kodama *et al.* 2014); because duplicated markers have not been genotyped in this study, these chromosomal arms are highly under represented here. Nevertheless, it was intriguing to have observed a bias towards chromosomes involved in homeologous pairing. Loci that are involved in homeologous pairing undergo tetrasomic inheritance in salmon; such loci have larger effective population sizes compared to those that undergo disomic inheritance (Allendorf *et al.* 2015), and the effectiveness of selection relative to drift may be high on these regions (Charlesworth 2009; Parisod *et al.* 2010). Combined with genic redundancy between homeologs, these regions may have provided novel opportunities for adaptive phenotypic evolution (Kodama et al. 2014), thus responding to selection observed in the present study.

Identifying genomic regions that are involved in driving adaptive phenotypic evolution remains as a key interest in evolutionary biology because such knowledge provides insights into how phenotypic differentiation resulting from certain selective pressures is determined at specific sites in the genome. Here, we surveyed loci and genomic regions for evidence of elevated divergence in a domesticated line of coho salmon compared to its hatchery and wild source populations. We found several regions exhibiting high differentiation, some of which were linked with commercially important traits targeted during selective breeding. In particular,

we observed a strong footprint of selection on a region linked with growth and age at sexual maturity, and these regions have likely differentiated in response to selective breeding targeting rapid growth and early maturity. Our results demonstrated how and to what extent adaptive evolutionary changes result from selective breeding in a captive environment, and revealed loci that may have played a role in adaptive phenotypic differentiation driven by domestication selection.

Acknowledgements

We would like to thank Isadora Jimenez-Hidalgo for assistance with genomic library construction. Charlie Waters, Marine Brieuc and Dan Drinan provided helpful suggestions and invaluable comments on the population genomic analyses. We would also like to thank Jim Myers and Edward Eleazor for providing samples. Funding for this study was provided by NOAA Fisheries/ Federal Columbia River Power System (FCRPS) Biological Opinion Remand Funds (to K.A.N. and J.J.H.), School of Aquatic and Fishery Sciences, and Graduate Opportunities and Minority Achievement Program (GO-MAP, University of Washington) award (to M.K.).

Figures

Figure 4.1 - Principal Components Analysis (PCA) of all individuals based on the 4768 mapped loci. Brown, yellow and blue dots represent Domsea, Wallace River Hatchery and South Fork Skykomish wild coho individuals, respectively. Lines connect individuals with their respective population mean. The scree plot shows the eigenvalues of the plotted axes in black.

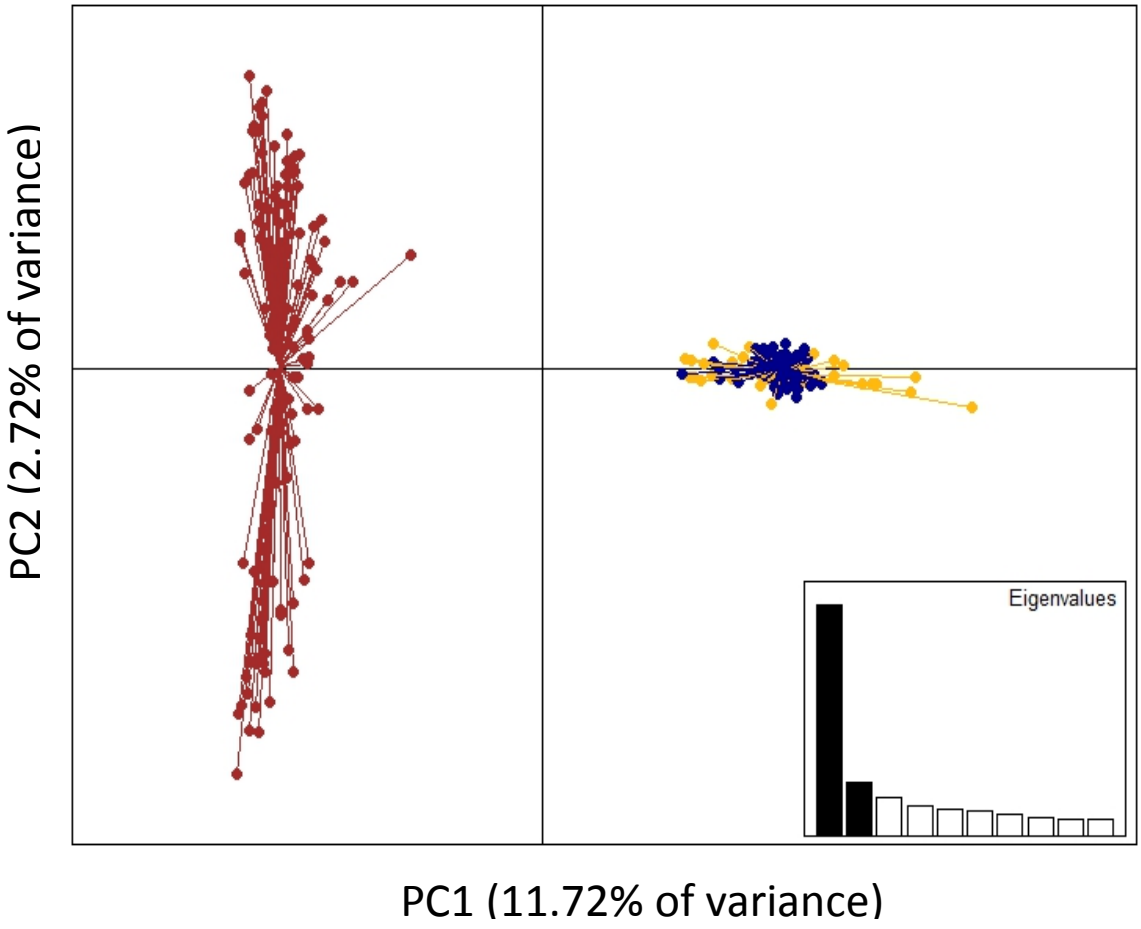


Figure 4.2 - Assignment of all individuals to clusters identified using Bayesian clustering in BAPS, based on the 4768 mapped loci. Predefined populations are indicated by the black vertical lines.

Figure 4.3 - Relatedness measures obtained by Coancestry. The x-axis represents relatedness estimates, and the y-axis represents each Domsea cluster observed in BAPS (Domsea 1 – Domsea 11), as well as the whole Domsea (Domsea) and South Fork Skykomish data set (South Fork). Individual pairwise relatedness estimates are depicted by open circular dots. Means are represented by black circular dots, and standard deviations are depicted by black bars.

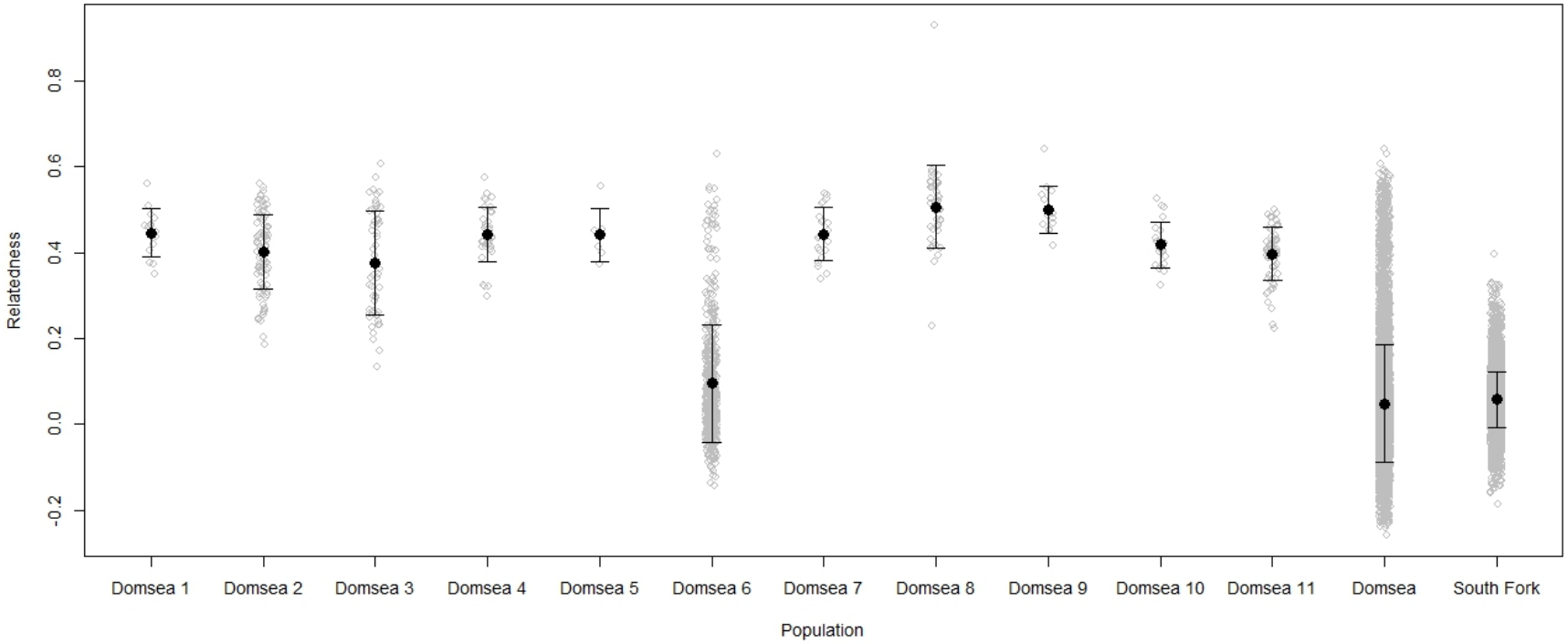


Figure 4.4 - Examples of outliers and genomic regions of high divergence detected in the pairwise comparison between the Domsea broodstock and the South Fork Skykomish wild. The y-axes represent F_{ST} and the x-axes represent the distance along each linkage group. The black lines represent the kernel smoothed moving average of F_{ST} values and the 95% confidence interval are shown by the gray shaded area. The blue triangles represent the outlier loci detected by Bayescan at q-value < 0.05. The red dots represent the outlier loci detected by Arlequin at $\alpha = 0.05$ after Bonferroni correction.

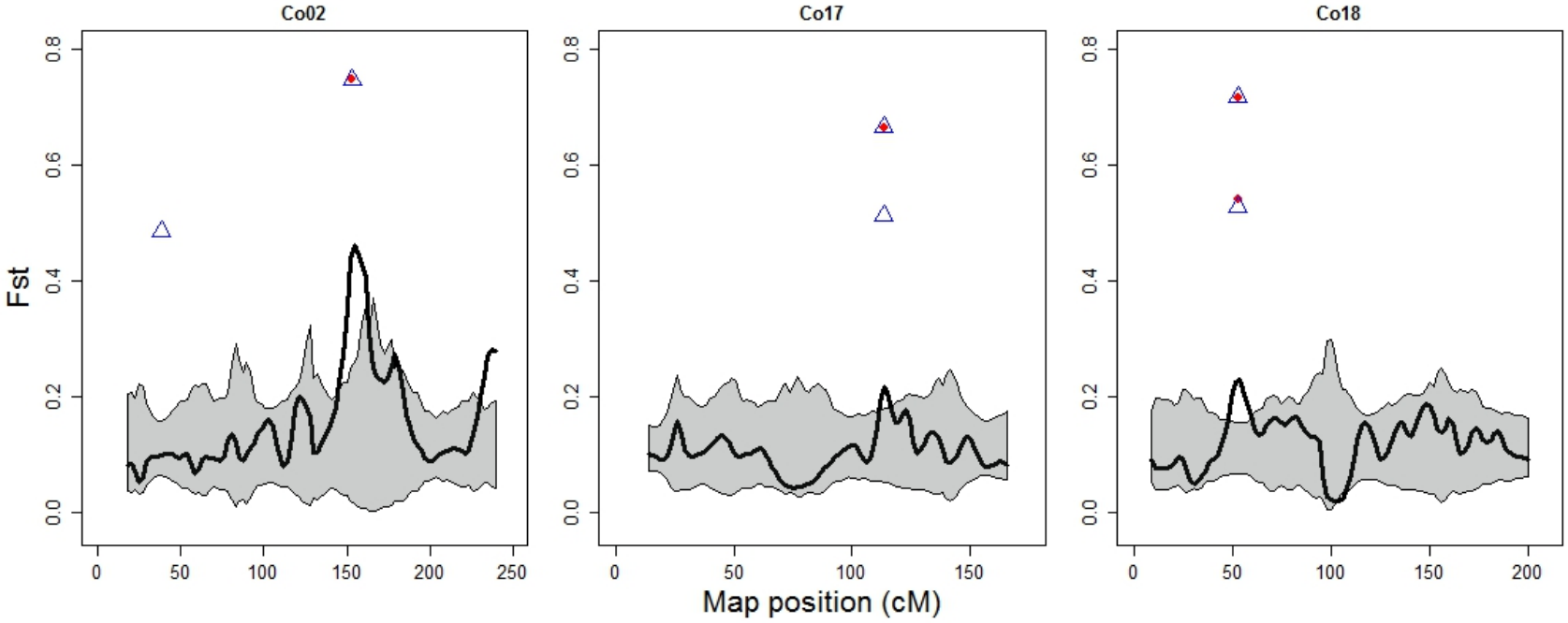
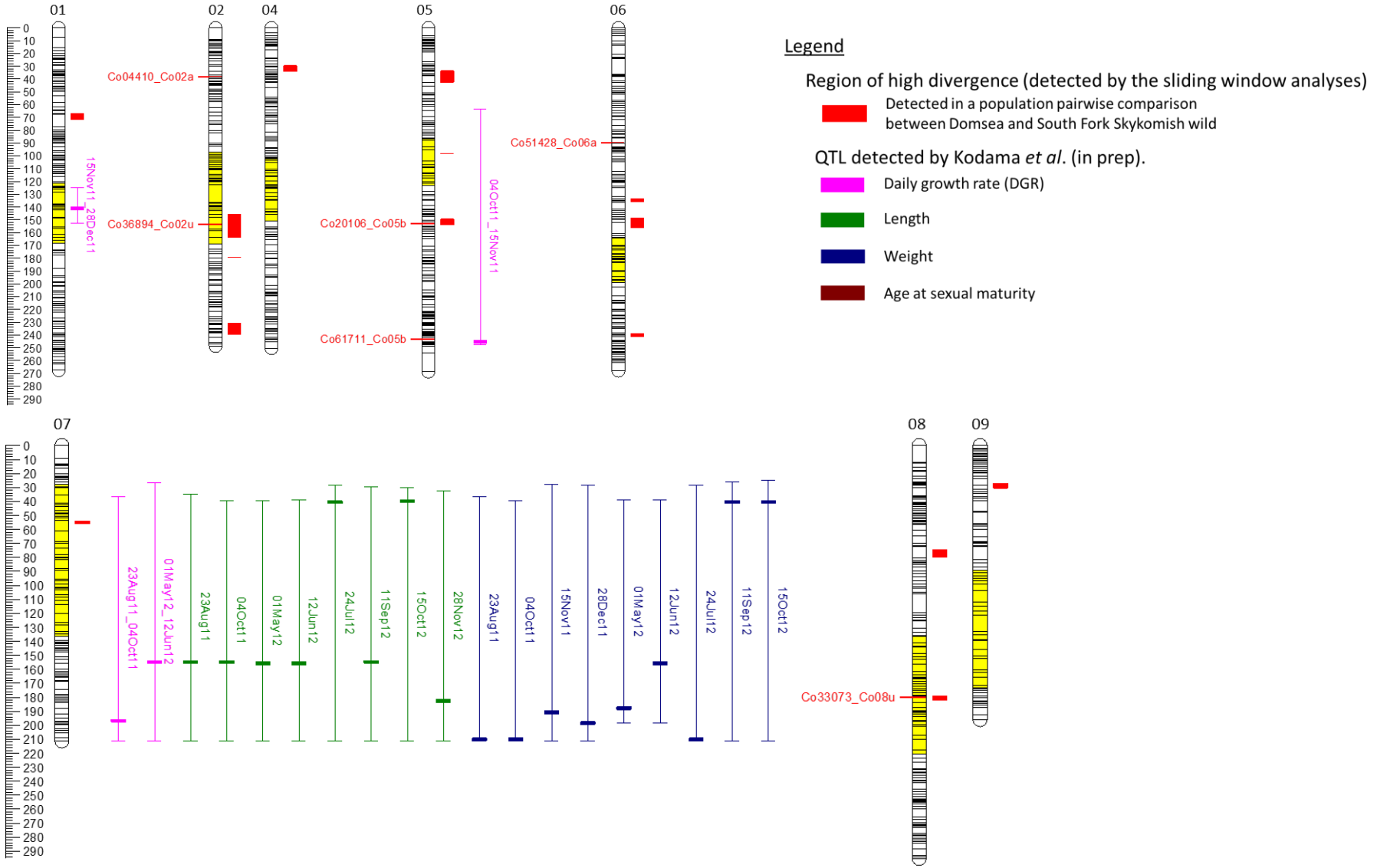
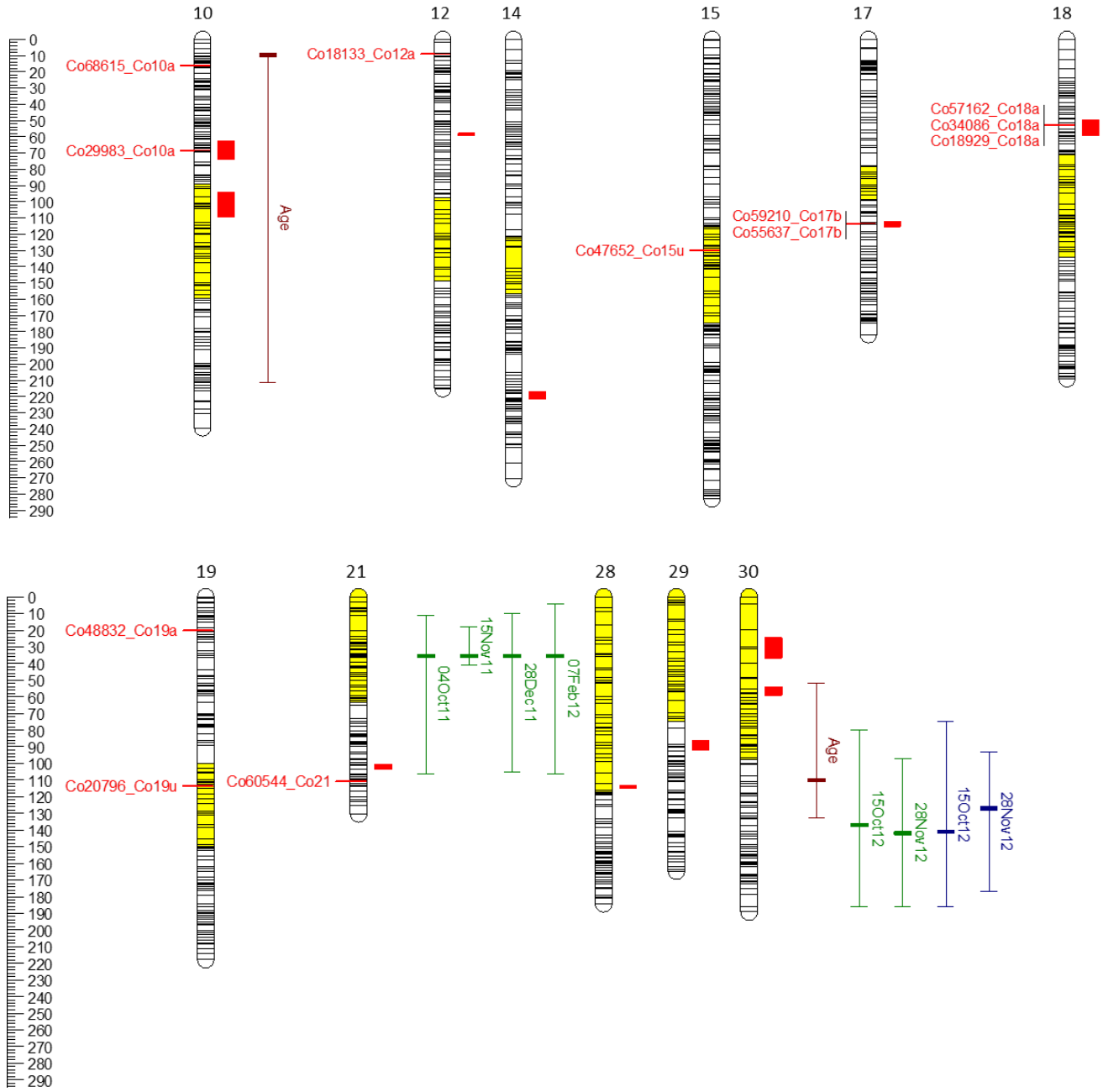


Figure 4.5 - Location of the outlier loci and genomic regions exhibiting high divergence, along with the growth-related QTL identified by Kodama *et al.* (in prep). Only linkage groups harboring outliers or regions of high divergence are shown. The putative location of the centromere is represented in yellow. Genomic regions of high divergence and QTL are shown on the right side of each linkage group. The marker associated with each QTL is shown by a horizontal bar. Outlier loci detected by Arlequin or Bayescan that are nearby or overlap with the genomic regions of high divergence are shown on the left side of each linkage group in red.




10
12
14
15
17
18
Co68615_Co10a
Co29983_Co10a
Age
Co18133_Co12a
Co47652_Co15u
Co59210_Co17b
Co55637_Co17b
Co57162_Co18a
Co34086_Co18a
Co18929_Co18a
19
21
28
29
30
Co48832_Co19a
Co20796_Co19u
Co60544_Co21
04Oct11
15Nov11
28Dec11
07Feb12
Age
15Oct12
28Nov12
15Oct12
28Nov12

Supplementary Material

S4.1 - Description of samples collected from Domsea broodstock, Wallace River Hatchery and South Fork Skykomish wild population. Year in which samples were collected, the number of samples collected, the number of samples that were genotyped at >50% of 4768 mapped loci, overall observed (Ho) and expected (He) heterozygosity and the number of loci that deviated from Hardy-Weinberg equilibrium at $\alpha = 0.01$.

S4.2 - Assignment of individuals from the Domsea population to groups identified in BAPS. BAPS runs were performed using loci with two alleles at moderate frequencies (allele 1 and allele 2 frequencies ranging from 22.5% to 77.5%; 1930 loci in total).

S4.3 - Outlier loci detected by Arlequin (at $\alpha < 0.05$ after Bonferroni correction) in pairwise comparisons between Domsea broodstock and South Fork Skykomish wild coho.**

S4.4 - Outlier loci detected by Bayescan (at q-value < 0.05) in pairwise comparisons between Domsea broodstock and South Fork Skykomish wild coho.

S4.5 - Comparisons between results of three tests for genomic regions of high divergence, performed using Domsea individuals or a subset of Domsea individuals (subset) against South Fork Skykomish wild population. Location of genomic regions exhibiting differentiation were identified by the sliding window analyses using the 4768 mapped loci. Outlier genomic regions with one locus of high divergence detected by the sliding window analyses are marked by a starting point. Outlier regions that were unique to one pairwise comparison are labeled as "X," while those that were common across both pairwise comparisons are labeled with corresponding map positions. Number of loci under selection detected by Arlequin or Bayescan that are within or nearby the outlier regions are also shown.

References

Akey, J. M., A. L. Ruhe, D. T. Akey, A. K. Wong, C. F. Connelly *et al.*, 2010 Tracking footprints of artificial selection in the dog genome. Proc. Natl. Acad. Sci. U.S.A. 107: 1160-1165.

Allendorf, F. W., S. Bassham, W. A. Cresko, M. T. Limborg, L. W. Seeb *et al.*, 2015 Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. J. Hered. 1-11.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS. One. 3: e3376.

Barrett, R. D. H., and H. E. Hoekstra, 2011 Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. 12: 767-780.

Bourret, V., P. T. O'Reilly, J. W. Carr, P. R. Berg, and L. Bernatchez, 2011 Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo sala*r) population following introgression by farmed escapees. Heredity. 106: 500-510.

Brieuc, M. S. O., K. Ono, D. Drinan, and K. A. Naish, 2015 Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). Mol. Ecol. 24: 2729-2746.

Brieuc, M. S. O., C. D. Waters, J. E. Seeb, and K. A. Naish, 2014 A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication Event. G3-Genes. Genom. Genet. 4: 447-460.

Carneiro, M., C. J. Rubin, F. D. Palma, and F. W. Albert, 2014 Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science. 345: 30-32.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait, 2011 Stacks: building and genotyping Loci de novo from short-read sequences. G3-Genes. Genom. Genet. 1: 171-182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013 Stacks: an analysis tool set for population genomics. Mol. Ecol. 22: 3124-3140.

Charlesworth, B., 2009 Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10: 195-205.

Corander, J., and P. Marttinen, 2006 Bayesian identification of admixture events using multilocus molecular markers. Mol. Ecol. 15: 2833-43.

Darwin, C., 1859 *Origins of species by means of natural selection, or the preservation of favoured races in the struggle for life*. Murray, London.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen *et al.*, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12: 499-510.

Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett *et al.*, 2014 NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Resour. 14: 209-214.

Doebley, J. F., B. S. Gaut, and B. D. Smith, 2006 The molecular genetics of crop domestication. Cell. 127: 1309-1321.

Ekblom, R., and J. Galindo, 2011 Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity. 107: 1-15.

Excoffier, L., and H. E. L. Lischer, 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10: 564-567.

Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut, 1998 Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. U.S.A. 95: 4441-4446.

FAO, 2007 *The state of the world's animal genetic resources for food and agriculture*.

Ferguson, A., I. Fleming, K. Hindar, Ø. Skaala, P. McGinnity *et al.*, 2007 Farm escapes, pp. 357-398 in *The Atlantic Salmon: Genetics, Conservation and Management.*, Blackwell, Oxford.

Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 180: 977-993.

Ford, M. J., 2002 Selection in captivity during supportive breeding may reduce fitness in the wild. Conserv. Biol. 16: 815-825.

Gjøen, H. M., and H. B. Bentsen, 1997 Past, present, and future of genetic improvement in salmon aquaculture. ICES. J. Mar. Sci. 54: 1009-1014.

Gross, A., A. Tönjes, P. Kovacs, K. R. Veeramah, P. Ahnert *et al.*, 2011 Population-genetic comparison of the Sorbian isolate population in Germany with the German KORA population using genome-wide SNP arrays. BMC. Genet. 12: 67.

Hill, W. G., 1981 Estimation of effective population size from data on linkage disequilibrium. Genet. Res. 38: 209-216.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6: e1000862.

Hohenlohe, P. A., P. C. Phillips, and W. A. Cresko, 2010 Using population genomics to detect selection in natural populations: key concepts and methodological considerations. Int. J. Plant. Sci. 171: 1059-1071.

Hutchings, J. A., and D. J. Fraser, 2008 The nature of fisheries- and farming-induced evolution. Mol. Ecol. 17: 294-313.

Jombart, T., 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24: 1403-1405.

Karlsson, S., T. Moen, S. Lien, K. A. Glover, and K. Hindar, 2011 Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol. Ecol. Resour. 11: 247-253.

Kodama, M., M. S. O. Brieuc, R. H. Devlin, J. J. Hard, and K. A. Naish, 2014 Comparative mapping between coho salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. G3-Genes. Genom. Genet. 4: 1717-1730.

Kodama, M., J. J. Hard, and K. A. Naish, In prep. Genetic architecture of age at sexual maturity and growth-related traits in coho salmon and investigation of a trade-off between these traits.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome. Biol. 10: R25.

Larson, G., and J. Burger, 2013 A population genetics view of animal domestication. Trends. Genet. 29: 197-205.

Larson, G., D. R. Piperno, R. G. Allaby, M. D. Purugganan, L. Andersson *et al.*, 2014 Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. U.S.A. 111: 6139-6146.

Larson, W. A., L. W. Seeb, M. V Everett, R. K. Waples, W. D. Templin *et al.*, 2013 Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol. Appl. 7: 355-369.

Li, Y.-H., J. C. Reif, S. A. Jackson, Y.-S. Ma, R.-Z. Chang *et al.*, 2014 Detecting SNPs underlying domestication-related traits in soybean. BMC. Plant. Biol. 14: 251.

Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet, 2003 The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. 4: 981-994.

Mäkinen, H., A. Vasemägi, P. McGinnity, T. F. Cross, and C. R. Primmer, 2014 Population genomic analyses of early-phase Atlantic Salmon (*Salmo salar*) domestication/captive breeding. Evol. Appl. 1-15.

Martinez, V., P. Dettleff, P. Lopez, G. Fernandez, A. Jedlicki *et al.*, 2013 Assessing footprints of selection in commercial Atlantic salmon populations using microsatellite data. Anim. Genet. 44: 223-226.

Martinez, V., A. Kause, E. Mäntysaari, and A. Mäki-Tanila, 2006 The use of alternative breeding schemes to enhance genetic improvement in rainbow trout: II. Two-stage selection. Aquaculture. 254: 195-202.

McClelland, E. K., J. M. Myers, J. J. Hard, L. K. Park, and K. A. Naish, 2005 Two generations of outbreeding in coho salmon (*Oncorhynchus kisutch*): effects on size and growth. Can. J. Fish. Aquat. Sci. 62: 2538-2547.

Myers, J. M., P. O. Heggelund, G. Hudson, and R. N. Iwamoto, 2001 Genetics and broodstock management of coho salmon. Aquaculture. 197: 43-62.

Naish, K. A., and J. J. Hard, 2008 Bridging the gap between the genotype and the phenotype: linking genetic variation, selection and adaptation in fishes. Fish. Fish. 9: 396-422.

Narum, S. R., and J. E. Hess, 2011 Comparison of Fst outlier tests for SNP loci under selection. Mol. Ecol. Resour. 11: 184-194.

Novotny, A. J., 1975 Net-pen culture of Pacific salmon in marine waters. Mar. Fish. Rev. 37: 36-47.

Olsen, K. M., and J. F. Wendel, 2013 A bountiful harvest: genomic insights into crop domestication phenotypes. Annu. Rev. Plant Biol. 64: 47-70.

Parisod, C., R. Holderegger, and C. Brochmann, 2010 Evolutionary consequences of autopolyploidy. New Phytol. 5-17.

Raymond, M., and F. Rousset, 1995 GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. J. Hered. 86: 248-249.

Rubin, C.-J., H.-J. Megens, A. Martinez Barrio, K. Maqbool, S. Sayyab *et al.*, 2012 Strong signatures of selection in the domestic pig genome. Proc. Natl. Acad. Sci. U.S.A. 109: 19529-19536.

Rubin, C. J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood *et al.*, 2010 Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 464: 587-591.

Santure, A. W., J. Stapley, A. D. Ball, T. R. Birkhead, T. Burke *et al.*, 2010 On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Mol. Ecol. 19: 1439-1451.

Skaala, Ø., B. Høyheim, K. Glover, and G. Dahle, 2004 Microsatellite analysis in domesticated and wild Atlantic salmon (*Salmo salar L.*): allelic diversity and identification of individuals. Aquaculture. 240: 131-143.

Slate, J., J. Gratten, D. Beraldi, J. Stapley, M. Hale *et al.*, 2009 Gene mapping in the wild with SNPs: guidelines and future directions. Genetica. 136: 97-107.

Sun, L., S. Liu, R. Wang, Y. Jiang, Y. Zhang *et al.*, 2014 Identification and analysis of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (*Ictalurus punctatus*). PLoS One 9: e109666.

Thodesen, J., B. Grisdale-helland, J. Helland, and B. Gjerde, 1999 Feed intake, growth and feed utilization of offspring from wild and selected Atlantic salmon (*Salmo salar*). Aquaculture. 180: 237-246.

Vasemägi, A., J. Nilsson, P. McGinnity, T. Cross, P. O'Reilly *et al.*, 2012 Screen for footprints of selection during domestication/captive breeding of Atlantic salmon. Comp. Funct. Genomics. 2012: 628204.

Wang, J., 2011 COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Mol. Ecol. Resour. 11: 141-145.

Waples, R. S., 2006 A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conserv. Genet. 7: 167-184.

Washington Department of Fish and Wildlife, 2013 Hatchery and genetic management plan (Wallace River coho hatchery program).

Waters, C. D., J. J. Hard, M. S. O. Brieuc, D. E. Fast, K. I. Warheit *et al.*, In prep. Managed gene flow reduces genetic risks associated with captive rearing.

Weir, B. S., and C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution. 38: 1358-1370.

Wilkinson, S., Z. H. Lu, H.-J. Megens, A. L. Archibald, C. Haley *et al.*, 2013 Signatures of diversifying selection in European pig breeds. PLoS. Genet. 9: e1003453.

Wright, S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley *et al.*, 2005 The effects of artificial selection on the maize genome. Science. 308: 1310-1314.