2012-2013 Annual Report

INFLUENCE OF ENVIRONMENT AND LANDSCAPE ON SALMONID GENETICS

Prepared by:

Shawn Narum
Nate Campbell
Ben Hecht
Jon Hess

Columbia River Inter-Tribal Fish Commission
Hagerman Fish Culture Experiment Station

Prepared for:
U.S. Department of Energy
Bonneville Power Administration
Division of Fish and Wildlife
P.O. Box 3621
Portland, OR 97283-3621
Project Number: 2009-005-00
Contract Number: 43268

Reporting Period: 7/1/12-6/30/13

Submitted: 6/30/13

ABSTRACT/SUMMARY

Environmental and landscape features can greatly contribute to population structure, life history diversification, and adaptation of salmonids. This ongoing project combines two studies from the Fish & Wildlife Program Accords with the following objectives: 1) Environment & Landscape Genetics – Evaluate genetic structure of natural populations of salmonids relative to their environment and identify candidate markers associated with traits that are related to adaptation of steelhead and Chinook salmon populations; and 2) Controlled Experiments – laboratory/hatchery experiments with controlled environmental variables to validate phenotypic response of fish with given genotypes.

During the performance period of July 1, 2012 to June 30, 2013, work was completed that addresses components of both Objective 1 and Objective 2. For Objective 1, an extensive genetic baseline with 192 SNPs was created for steelhead throughout the Columbia River Basin, with ongoing analyses of landscape genetics. For Objective 2, empirical work was done to further advance our understanding multiple traits related to recovery of salmonids in the Columbia River. Investigations during the past performance period focused on the genomic basis for disease resistance in *O. mykiss* (Section 1), and run-timing in Chinook salmon (Section 2). Further, progress was also made towards developing projects to investigate the genomic basis for age-at-maturity in Chinook salmon and thermal tolerance in *O. mykiss* (Section 3).

Table of Contents

Acknowledgements

The authors of this report would like to thank the following individuals and organizations for providing time and expertise toward this project: The Bonneville Power Administration for providing funding for this research project and to its staff for their assistance. Tissue samples were collected and provided by Steve Elle, Kevin Meyer, Chris Sullivan of IDFG, Joe Zendt and field assistants of YKFP, and Scott Lapatra of Clear Springs Foods. We are grateful for extensive contributions in the genetics laboratory from Vanessa Morman, Stephanie Harmon, Lori Maxwell, Amanda Matala, Megan Moore, Nick Hoffman, and Jeff Stephenson.

Report Structure

This report is divided into three sections. The first section reports on disease resistance, the second section on run-timing, and the third section provides brief summaries for ongoing and future work on several traits of interest including heritability and genetic basis of age-at-maturity and ongoing work for thermal adaptation.

SECTION 1: Genomics of disease resistance traits in *O. mykiss*

Introduction

The practice of marker assisted selection (MAS) utilizes sets of genetic markers linked to desirable traits in order to guide breeding decisions for commercial production of agricultural species. In fact, MAS has proven to be a useful tool for the development of agricultural animal and plant strains with dramatically improved trait characteristics (i.e. Han et al. 1997; Singh et al. 2001; Seraj et al. 2005). Global aquaculture of rainbow trout (*Oncorhynchus mykiss*) is extensive for both commercial and sport fishing purposes, but use of MAS to improve traits such as disease resistance is currently lacking. In order for MAS to become useful for rainbow trout trait selection, closely associated genetic markers must first be identified from a dense panel of markers throughout the genome. To this point, association and QTL mapping approaches in this species have relied on relatively small numbers of genetic markers. These approaches have either failed to identify highly associated loci (Overturf et al. 2010) or identified significant associations within very large blocks of LD (Rodriquez et al. 2004; Barroso et al. 2008). However, recent advancements in high-throughput sequencing technology have made genotyping-by-sequencing methods possible (Narum et al. 2013), bypassing the need for expensive a-priori marker identification (i.e. Hess et al. 2013; Keller et al. 2013). Methods such as restriction site associated sequencing (RAD) have enabled genome wide association studies (GWAS) with large numbers of SNPs for species with limited genomic resources (i.e. Hecht et al. 2013).

Worldwide cultivation of rainbow trout is extensive with an estimated 354,000 metric tons produced in 2012 (Tveterås 2013). These fish are widely reared an artificial hatchery environments where well oxygenated cool water is available (12-15 ˚C), grow quickly on pelleted fish meal based diets, and respond well to handling. However, as with other aquaculture species, rainbow trout are susceptible to disease and outbreaks can result in significant losses (LaPatra et al. 2001a). As such, precautions are taken to ameliorate disease outbreaks through better management practices such as using disinfecting foot baths, iodine treatment of eggs, and limiting reused water. However, disease outbreaks can still occur through various sources such as transmission from wild sources or asymptomatic infection within the hatchery population. Other approaches such as immune boosting feed formulations (i.e. Navarre and Halver 1989 & Brunt et al. 2007) and vaccination against common pathogens have also been explored (Alvarez et al. 2008; LaFrenz et al. 2003; LaPatra et al. 2001b) although these approaches also increase production costs.

Another approach to develop disease resistant rainbow trout strains is through selective breeding. Resistance to disease is a difficult trait to phenotype and methods for quantification of disease resistance have used offspring mortality rates following pathogen exposure to measure parent breeding potential (i.e. Henryon et al. 2005). Selection of broodstock fish is then based on mortality rates rather than phenotype for selective breeding purposes. Over generations, gene variants within the cultivated fish strain that are beneficial for disease resistance are enhanced while those associated with disease manifestation are diminished. This method has been used to create rainbow trout strains with improved resistance to a few common pathogens. However,

selected strains remain susceptible to other diseases and the selection process may reduce negatively affect favorable production traits such as growth rate (Henryon et al. 2002).

In this study RAD sequencing was used to identify genetic markers within a population of cultured rainbow trout and identify significant associations with disease resistance. A set of disease resistance associated genetic markers was then analyzed further to assess their ability to discriminate susceptible from resistant fish segregating within the farm population.

Methods

Disease Challenge and Sample collection

Samples were collected from disease challenged fish and their parents in brood years 2008 and 2010 from the Clear Springs Foods Inc. research facility in Buhl, Idaho. Selected fish from the previous generation were artificially spawned to produce 2,500 fertilized eggs from 130 families each over the course of 13 weeks. Fin tissue samples from each parent fish were collected at the time of spawning. The offspring were grown to ~1 gram (62 days post-fertilization) and 100 fish per family are randomly selected for disease challenge. Fifty fish from each family were infected with CWD by injection of 10 uL of a 0.2 OD suspension of *F. psychrophilum* while the remaining fish were infected with IHNV by immersion into a volume of water 10x the total body weight of the fish in grams containing 10,000 plaque forming units of IHNV per mL for 1 hour. Following exposure, the fish were moved to 19L tanks and monitored for a period of 3 weeks with mortality recorded daily. Fin tissue samples were collected from mortalities during the 3 week monitoring period and survivor samples were taken at the conclusion of the challenge. The percentage of mortality for each family with each pathogen was recorded.

Mortality rates for each of the disease challenged families were examined and a broad range of mortality was targeted for each pathogen for inclusion in RAD sequencing. The mortality rates for CWD ranged from 4% to 96% and included 20 families. For IHNV family mortality range was 4% to 92% and also included 20 families. A total of 456 samples, including both parents and disease challenged offspring, were selected for RAD sequencing and included families from two brood years (2008 & 2010). Each parent with equal numbers of offspring mortalities ($N = 5$) and survivors ($N = 5$) were chosen for sequencing (Table 1).

RAD library preparation

Fin tissue samples from individuals selected for RAD genotyping were used for DNA using Qiagen DNeasy 96 kits. Quantification of extract DNA was done using Invitrogen Quant-It pico green reagent and a Perkin Elmer Victor V fluorimeter. Of the 456 samples chosen for inclusion in RAD library preparation 27 had insufficient DNA concentration after extraction and quantitation. DNA extracts from the remaining 429 samples were digested, barcoded, and sonicated before pooling into 10 libraries of between 40 and 48 samples each. Samples were normalized to 5 ng/uL and 500 ng of each sample was digested with Sbf1-HF restriction enzyme in NEBuffer 4 (New England Biolabs (NEB)). Barcoded adapters were then ligated onto the cut ends of the restriction sites using T4 DNA ligase (NEB) and the samples were then pooled into libraries of 48 individuals each. The remaining steps of library preparation were carried out as described in Miller *et al. 2012* and Hecht *et al. 2013*. The concentration of a 1:1000 dilution of each completed library was determined by quantitative PCR (qPCR) using Life Technologies

PowerSYBR reagent and Kappa biosystems Illumina library DNA standards run on an Applied Biosystems 7900 instrument. Library concentration ranged from 6.5 nM to 71 nM after the addition of the P2 adapter and 15 cycles of PCR amplification. The concentration of each library was normalized to 5nM and sequenced on an Illumina HiSeq 2000 instrument.

Genotyping pipeline

Raw sequencing data included 101 bases per read and averaged 151 million reads per library with a range of 143 – 245M. The sequences were quality filtered, trimmed, and split into individual files based on barcode sequence using a custom perl script described in Miller *et al 2012*. This process also strips the 6 base barcode and 6 base partial *Sbf1* site located on the 5' end of each sequence while also trimming another 30 bases off of the lower quality 3' end of each sequence. After quality filtering of the sequencing reads roughly 90% of the sequencing reads were retained. Of the reads passing quality filtering, roughly 70% began with a valid barcode and partial *Sbf1* sequence. The average number of reads per individual was 2.1M but with a standard deviation of 0.86M illustrating the read count variation between individuals.

Three sets of parents ($N = 6$) with high read counts were normalized to 2.4M reads each and used for sequence alignment and SNP identification. These sequences were then combined and collapsed into unique sequences for alignment to one another using the program Novoalign (Novocraft, Selangor, Malaysia). The remaining steps for identifying SNPs within the RAD sequences and genotyping individuals were done according to the methods detailed in Miller *et al.* 2012 and Hecht *et al.* 2013. In brief, a custom perl script was used to identify and output allele sequences for loci containing a single SNP site from the alignment data. Parameters for identification of SNPs within alignments of RAD sequences were set such that only those containing a single SNP variation which occurred more than 5 times within an alignment were collected. A total of 5,647 putative SNP loci were identified within the sequencing data for these 6 individuals. Allele counts from each putative SNP locus were used to generate genotypes within all individuals in the study. A minimum read depth of 10 counts per locus was required in order for genotyping and genotypes were attempted in all individuals regardless of read count at this stage. Allele ratios were then used for determining genotypes for each RAD locus for each individual. Genotypes were scored using the following ratios [Allele 1 Homozygote > 7/1 > Heterozygote > 1/7 > Allele 2 Homozygote].

Data Analysis

Genotype data were refined by eliminating individuals and SNP loci with more than 20% missing genotypes. Observed and expected heterozygosity (H_{obs} & H_{exp}) was calculated for each locus and differences in these values of greater than 0.4 were excluded from the data set. Further, Hardy-Weinberg equilibrium was tested for each locus, and any loci deviating from equilibrium were removed to reduce the inclusion of paralogous sequence variants (PSVs). After filtering, a final data set containing 384 individuals genotyped at 4,661 SNP loci were used to generate an input file for the program TASSEL (Bradbury *et al.* 2007) to determine association between genotype and resistance to either CWD or IHNV. Of the 384 individuals genotyped, 159 were challenged with CWD and 153 were challenged for IHNV (Table 1). The remaining 72 fish were parents of the disease challenged fish but had no associated phenotype since they were not directly challenged.

The TASSEL program uses either a general linear model (GLM) which may include covariates or a mixed linear model (MLM) which includes a kinship matrix in addition to any covariates to determine association between traits and phenotype. In our case, we described disease resistance as a simple binary trait where mortalities were given a value of 1 and survivors a value of 2. Several types of covariates were generated for inclusion in both the GLM and MLM analysis including STRUCTURE Q-coordinates (Pritchard & Wen 2003), factorial coordinates (GENETIX: Belkhir et al. 2004), and principal components values (GENALEX 6: Peakall & Smouse 2006). Covariate data was generated using 1,300 SNP loci with 95% genotype frequency or greater and 20% or higher minor allele frequency. A kinship matrix was generated using the EMMA algorithm in the program GAPIT (Lipka *et al. 2012*) for inclusion in the MLM analysis using the same loci used for the covariate data. In order to discern true associations from those attributable to genetic background and family effects it was appropriate to also include kinship and covariate data. This study was designed as to reduce family effects by including equal numbers of mortalities and survivors from each family cross. However, since the data was collected from an inbred hatchery strain it was possible that more cryptic relatedness between families could yield false positive associations. To account for this possibility a pairwise kinship matrix was included in TASSEL's mixed linear model. We determined which and how many covariates for our final data set using various iterations of the GLM and MLM analysis in TASSEL and subsequent evaluation of QQ-plots. The final analysis used six principal components as covariates along with the kinship matrix in the MLM. Statistically significant loci were identified by applying a BY-FDR correction for multiple tests (False Discovery Rate of Benjamini & Yekutieli 2001).

Following the identification of SNP loci associated with disease resistance, the TASSEL program was used to generate genetically evaluated breeding values (GEBV) for each of the fish in the study. This analysis uses best linear unbiased predictor (BLUP) values for each locus to determine GEBV. For comparison of GEBV data from IHNV and CWD each value of GEBV was converted to a number between 0 and 1 representing the upper and lower bounds of the distribution.

For mapping significantly associated loci to linkage groups we used RAD sequences generated from Miller *et al.* 2012 to match identical loci in our data which used the same restriction enzyme (*Sbf1*). We then filtered the matching loci to those containing map positions ($N = 721$). Linkage groups for significant loci with direct matches to mapped loci from Miller *et al.* were recorded ($N = 4$). Genotypes for the mapped loci and the significantly associated loci were used for pairwise linkage disequilibrium (LD) tests using GENEPOP (Raymond and Rousset 1995). All samples were treated as a single population for the LD test using the standard settings (dememorisation 10,000; batches 100; iterations per batch 5,000). Output data was filtered for pairwise comparisons containing significantly associated loci and perl was used to add linkage groups and map positions to the data. A Bonferroni correction (α=0.01) for significance was used to filter p-values ($p\text{-}value < 0.00001339$) for each pairwise comparison. Linkage group was determined by examination of the map position of all loci determined to be in statistically significant LD. In some cases the linkage group was ambiguous and LG was recorded as not determined ("ND"; Table 2). Linkage group names are reported according to both Miller et al. 2012 and Palti et al. 2011. Linkage group number and chromosome number were resolved by Phillips et al. 2006 and are reported as such in the Palti et al. 2011 genetic map. This map was

also used to determine proper chromosome numbers in previous QTL studies by matching common microsatellite markers.

Results

In order to account for kinship in the GWAS model, a subset of 1,300 SNP genotypes were used to produce a kinship matrix that accurately reconstructed all known 39 families included in the analysis and also revealed other relatedness patterns among family groups (Figure 1a). The same subset of SNP genotypes were also used to generate principal components values using the program GenAlEx. This analysis also illustrated the ability of the genotype data to accurately cluster individuals into known families (Figure 1b).

Association analysis using the program TASSEL was conducted using several covariates in both the general and mixed linear models. Most of the significantly associated loci remained largely the same regardless of the number or type of covariates used. Q-values were quickly eliminated as a viable option for this study due to the genetic homogeneity of the hatchery strain and inability to identify the number of distinct founding populations. Factorial coordinates (FCs) and principal coordinates (PCs) used as covariates produced nearly identical results. However, since both PCs and FCs provided similar data, PCs were chosen as they are the most commonly used of the two in association mapping. After examination of QQ plots using 3 and 6 PCs, a combination of six PCs as covariates was chosen along with the kinship matrix in the MLM analysis. Statistically significant loci were identified using BY-FDR corrections for multiple tests (False Discovery Rate of Benjamini & Yekutieli 2001) which produced a significance cutoff of $p < 1.073\text{x}10^{-5}$ at $\alpha = 0.05$. Figure 2 shows the negative log of each locus's p-value along with three increasingly stringent significance values (markers are not ordered by genetic position). In total there were 31 loci significantly associated with either IHNV or CWD survival identified (Table 2).

To assess the value of these markers for prediction of survival to disease exposure we generated genetically evaluated breeding values (GEBV). We generated GEBV for all the individuals in our study using the markers identified as significantly associated with resistance to each disease. The distribution of GEBVs between known mortalities and survivors (Figure 3) was examined by ANOVA and found to be highly significant between the survivor and mortality groups for both diseases (IHNV: $p\text{-}value = 1.9\text{x}10^{-17}$; CWD: $p\text{-}value = 9.5\text{x}10^{-14}$).

Disease resistance associated loci were mapped to linkage groups by LD analysis by matching RAD loci from a previously generated genetic map (Miller et al. 2012). Using a custom perl script to identify common allele sequences we were able to identify 721 mapped RAD loci contained in our genotype data. A filtered set of genotype data that contained only mapped and significantly associated RAD loci was then used for LD analysis using the program GENEPOP. Linkage group was determined on a per locus basis considering mapped loci with significant Bonferroni corrected p-values (α=0.01; $p\text{-}value < 0.00001339$). For most of the disease resistance associated loci a confident linkage group assignment could be made ($N = 25$; Table 2). However, there were 6 loci with statistically significant LD with roughly equal numbers of mapped loci representing several linkage groups and therefore could not be determined. Of the disease resistance loci we could map, several linkage groups were observed multiple times. For instance, chromosome 6 was represented in 8 of the 19 IHNV associated RAD loci and 2 of the

12 CWD associated loci indicating that this genomic region plays an important role in disease resistance.

Discussion

Association analysis uncovered 31 SNP loci significantly associated with resistance to either IHNV or CWD in rainbow trout. Moreover, when these associated loci were analyzed collectively, they demonstrate potential to predict an individual's resistance to disease. However, unlike QTL analysis which is generally done using the offspring of a known F2 cross, association studies simply use populations of "affected" and "unaffected" individuals which are not necessarily related. Therefore, association studies require a closer proximity between the trait gene and a linked genetic marker due to linkage decay. Since rainbow trout have a genome size of about 3B bases and utilized a restriction enzyme with an 8 base recognition sequence, our RAD loci were expected to average 65K bases apart. However, since the RAD sequences also must contain a single bi-allelic SNP site, true marker density averages only one marker every 650K bases. It is therefore probable that there are genomic regions associated with disease resistance that were not detected due to a lack of LD with a nearby marker. Indeed, the disease associated markers identified in this study do not fully differentiate susceptible from resistant fish, as illustrated by the significant overlap in GEBV between individuals in the mortality and survivor groups. Although, aside from genetic causes, other factors such as environment, experimental variation, and possible epigenetic effects may also play a role in determining disease resistance.

Selective breeding of this strain of rainbow trout was first implemented for resistance to CWD in 2001 and IHNV in 2000. Since broodstock fish are bred as two year olds, 3-4 generations of selective breeding had taken place before samples were taken for this project in 2008 & 2010. Under conditions of strong selection, alleles of large effect could have been pushed to near fixation in the interval between starting selective breeding and collection of samples. If this is the case, then genomic regions with the strongest contribution to disease resistance would not have been detected in the current study. However, significant variation in family performance when exposed to each pathogen was still observed in these years and care was taken to select samples from families with a wide range of mortality rates. Moreover, broodstock fish were selected based on family performance in a combination of third-use water hardiness, development, and growth, as well as resistance to IHNV and CWD. We suspect this combinatorial approach softens the selective pressure for disease resistance allowing the detection of associated genetic markers even after several generations.

False positive associations resulting from enriched traits within certain populations and families often confound genetic association studies (Pritchard et al. 2000). For this reason great care is taken to discern true genetic associations from those resulting simply from genetic similarity between affected individuals. In this study this concern was addressed using three approaches. First, equal numbers of mortalities and survivors were from each family for RAD sequencing. By doing this both the control group and test group had matching genetic backgrounds. Secondly, a kinship matrix was generated to account for relatedness with the best fit linear regression in order to discern true associations from those resulting from family effects. Lastly, PCA coordinates were used as a measure of the genetic background of each individual fish in the study. Similar to the kinship values, these are subtracted from the linear model in order to test if the trait is more closely correlated with genotype or genetic background. Using these filters a

relatively small number of markers were identified associated with disease resistance (*N* = 31; IHNV = 19; CWD = 12) representing only a few linkage groups, providing confidence that the identified markers are indeed associated with resistance to these two diseases. Further, this study identified significant associations to 3 of 6 chromosomes previously reported to contain QTL for IHNV resistance in rainbow trout outcrosses (Rodruiquez et al 2004; Barroso el al 2008). However, these studies did not implicate the most commonly associated chromosome (Chr6) as containing QTL for IHNV resistance. Though, chromosome 6 may not have played a role in IHNV resistance in Yellowstone cutthroat trout (*O. clarkii bouvieri*) or steelhead trout (anadromous form of *O. mykiss*) outcrosses in those studies.

Unlike other agriculture species, rainbow trout lack the genomic resources that would make it possible to more finely map associated markers and identify specific genes involved in disease resistance. Indeed, chromosome positions for each of our SNP sites would make association plots (Figure 2) more meaningful and allow for easier identification of associated genomic regions. However, marker assisted selection (MAS) for enrichment of the disease resistance trait is still possible although not as powerful. In the absence of genomic context, MAS for this species would rely on enrichment of resistance associated alleles but would require ongoing GWAS testing to ensure recombination doesn't alter association with the causative gene variant. Future studies with finer scale mapping of the genetic role in disease resistance could certainly make both trait prediction as well as MAS more accurate.

This study provides an example of how RAD sequencing can be used to generate thousands of SNP markers and conduct GWAS in species with limited genomic resources. The ability to identify genetic markers associated with physical traits makes applied genetic techniques such as MAS possible within these species as well. In this study RAD sequencing was used to identify and genotype 4,661 SNP markers segregating within a hatchery population of rainbow trout. Further, trait data was utilized to isolate 31 markers associated with survival following exposure to IHNV or CWD. Using genetically evaluated breeding values (GEBV) generated using genotypes at resistance associate loci only; this work demonstrates the ability to use these markers to predict phenotype. Lastly, 25 of our 31 resistance associated loci were successfully mapped to linkage groups by matching RAD loci to a set of previously mapped loci by LD analysis. These markers offer a foundation for further study of the underlying genomic regions related to resistance to IHNV and CWD in rainbow trout and incorporation of marker assisted selection to aquaculture programs for this species.

References

Alvarez B, Alvarez J, Menendez A, Guijarro J. (2008) A mutant in one of two *exbD* loci of a TonB system in *Flavobacterium psychrophilum* shows attenuated virulence and confers protection against cold water disease. *Microbiology*, **154**:1144-1151.

Barroso RM, Wheeler PA, LaPatra SE, Drew RE, Thorgaard GH. (2008) QTL for IHNV resistance and growth in a rainbow (*Oncorhynchus mykiss*) x Yellowstone cutthroat (*Oncorhynchus clarki bouvieri*) trout cross. *Aquaculture* **277**:156-163.

Belkhir K, Borsa P, Chikhi L, Raufaste M, Bonhomme F. (2004) GENETIX 4.05, logiciel sous Windows^{TM} pour la Ge´ne´tique des Populations. Montpellier, France, Laboratoire Ge´nome, Populations, Interactions, CNRS UMR 5171, Universite´ de Montpellier II.

Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Annals of Statistics,* **29**, 1165-1188.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633-2635.

Brunt J, Newaj-Fyzul A, Austin B. (2007) The development of probiotics for the control of multiple bacterial diseases of rainbow trout, Oncorhynchus mykiss (Walbaum). *Journal of Fish Diseases*, **30**:573-579.

Han F, Romagosa I, Ullrich SE, Jones BL, Hayes PM, Wesenberg DM. (1997) Molecular marker-assisted selection for malting quality traits in barley. *Molecular Breeding*, **3**:427-437.

Hecht BC, Campbell NR, Holecek DE, Narum SR. (2013) Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow trout and steelhead trout. *Molecular Ecology*, **22**:3061-3076.

Henryon M, Jokumsen A, Berg P, Lund I, Pedersen P, Olesen N, SlierendrechtW. (2002) Genetic variation for growth rate, feed conversion efficiency, and disease resistance exists within a farmed population of rainbow trout. *Aquaculture* **209**, 59–76.

Henryon M, Berg P, Olesen NJ, Kjaer TE, Slierendracht WJ, Jokumsen A, Lund I. (2005) Selective breeding provides an approach to increase resistance of rainbow trout (*Oncorhynchus mykiss*) to the disease, enteric redmouth disease, rainbow trout fry syndrome, and viral haemorrhagic septicaemia. *Aquaculture*, **250**:621-636.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR. (2013) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, **22**:2898-2916.

Keller I, Wagner CE, Gueter L, Mwaiko S, Selz OM, Sivasundar A, Wittwer S, Seehausen O. (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, **22**:2848-2863.

LaFrenz BR, LaPatra SE, Jones GR, Cain KD. (2003) Passive immunization of rainbow trout, *Oncorhynchus mykiss* (Walbaum), against *Flavobacterium psychrophilum*, the causative agent of bacterial coldwater disease and rainbow trout fry syndrome. *Journal of Fish Diseases*, **26**:377-384.

LaPatra SE, Corbeil S, Jones GR, Shewmaker WD, Lorenzen N, Anderson ED, Kurath G. (2001b) Protection of rainbow trout against infectious hematopoietic necrosis virus four days after specific or semi-specific DNA vaccination. *Vaccine*, **19**:4011-4019.

LaPatra SE, Batts WN, Overturf K, Jones GR, Shewmaker WD, Winton JR. (2001a) Negligible risk associated with the movement of processed rainbow trout from an infectious hematopoietic necrosis virus (IHNV) endemic area. *Journal of Fish Diseases*, **24**:399– 408.

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. (2012) GAPIT: Genome Association and Prediction Integrated Tool. *Bioinformatics*, 28(18):2397-2399.

Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad III CE, Palti Y, Doe CQ, Thorgaard GH. (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology*, **21**:237-249.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe A. (2013) Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology, **22**:2841-2847.

Navarre O, Halver JE. (1989) Disease resistance and humoral antibody production in rainbow trout fed high levels of vitamin C. *Aquaculture*, 79(1–4), 207–221.

Overturf K, LaPatra S, Towner R, Campbell N, Narum S. (2010) Relationships between growth and disease resistance in rainbow trout, Oncorhynchus mykiss (Walbaum). *Journal of Fish Diseases*, **33**:321-329.

Palti Y, Genet C, Luo MC, Charlet A, Gao G, Hu Y, Cantano-Sanchez C, Tabet-Canale K, Krieg F, Yao J, Vallejo RL, Rexroad III CE (2011) A first generation integrated map of the rainbow trout genome. BMC Genomics, 12:180-188.

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

Phillips RB, Nichols KM, DeKoning JJ, Morasch MR, Keatley KA, Rexroad III C, Gahr SA, Danzmann RG, Drew RE, Thorgaard GH (2006) Assignment of rainbow trout linkage groups to specific chromosomes. *Genetics*, **174**:1661-1670.

Pritchard JK, Stephens M, Rosenberg NA, Donnely P (2000) Association mapping in structured populations. *American Journal of Human Genetics*, **67**:170-181.

Pritchard JK, Wen W (2003) *Documentation for STRUCTURE software: Version 2*. Available from http://pritch.bsd.uchicago.edu.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity,* **86**, 248-249.
Rodriguez MF, LaPatra S, Williams S, Famula T, May B. (2004) Genetic markers associated with resistance to infectious hematopoitic necrosis virus in rainbow trout and steelhead trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture*, **241**:93-115.

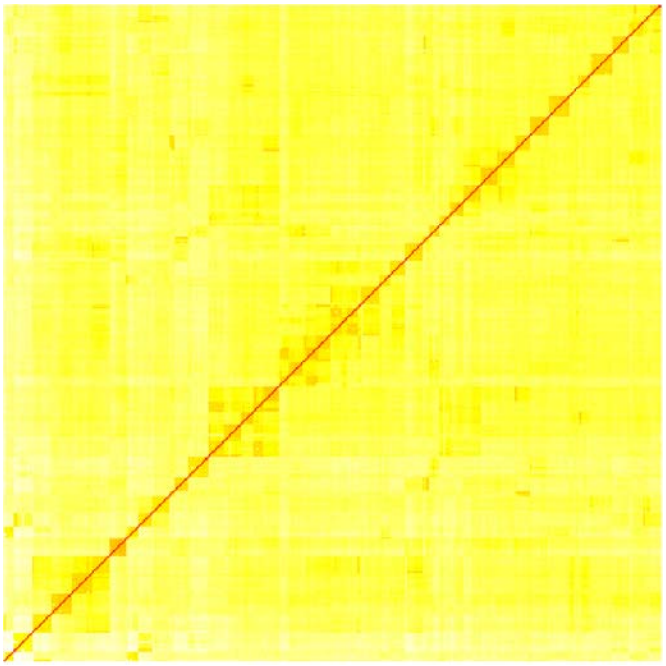
Serraj R, Hash CT, Rizvi SMH, Sharma A, Yadav RS, Bidinger FR. (2005) Recent advances in marker-assisted selection for drought tolerance in perl millet. *Plant Production Science* **8**:334-337.

Singh S, Sidhu JS, Huang N, Vikal Y, Li Z, Brar DS, Dhaliwal HS, Khush GS. (2001) Pyramiding three bacterial blight resistance genes (xa5, xa13 and Xa21) using marker assisted selection in to indica rice cultivar PR106. *Theoretical Applied Genetics* **102**:1011-1015.

Tveterås R., 2013. PRODUCTION: Global Fish Review. *The Global Aquaculture Advocate*, **Jan/Feb**, 10-11.

Figure 1: a) Heat map of pairwise kinship among individuals included in the study. Red squares indicate an individual's relatedness to itself and orange blocks indicate genetically reconstructed family groups. b) Principal coordinates plot of all individuals. Family groups are indicated by marker color/shape.

a.)



b.)

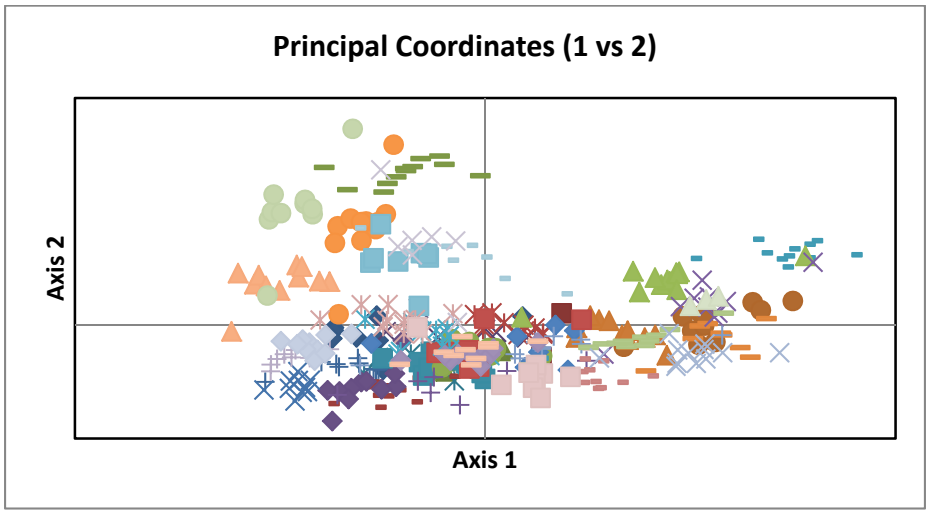
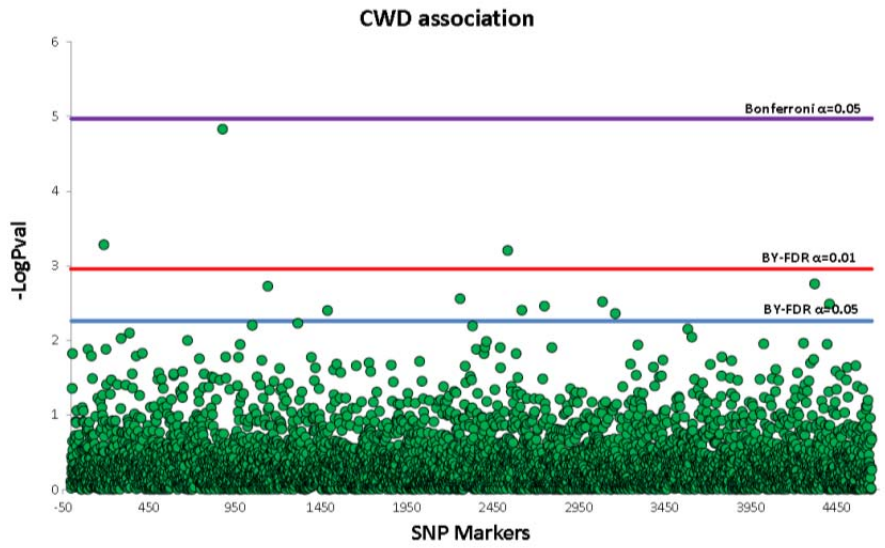


Figure 2: Manhattan plots of significance values for each marker associated with resistance to each disease , A) cold water disease (CWD), and B) Infectious hematopoietic necrosis virus (IHNV).. Horizontal lines in each figure represent increasingly stringent critical values. Markers are not ordered by genetic position.

A.



B.

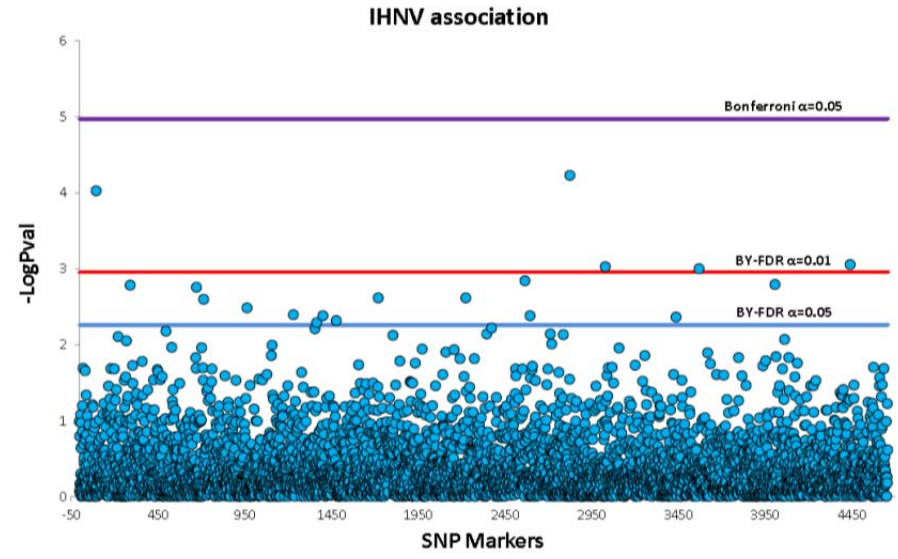
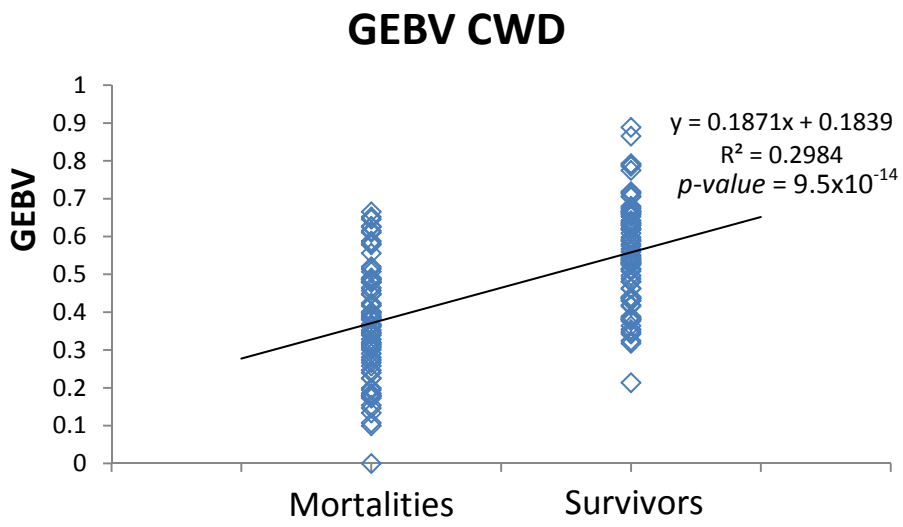


Figure 3: Distribution of genetically evaluated breeding values (GEBVs) among disease challenged fish using significantly associated markers for A) cold water disease (CWD) and B) Infectious hematopoietic necrosis virus (IHNV).

A.



B.

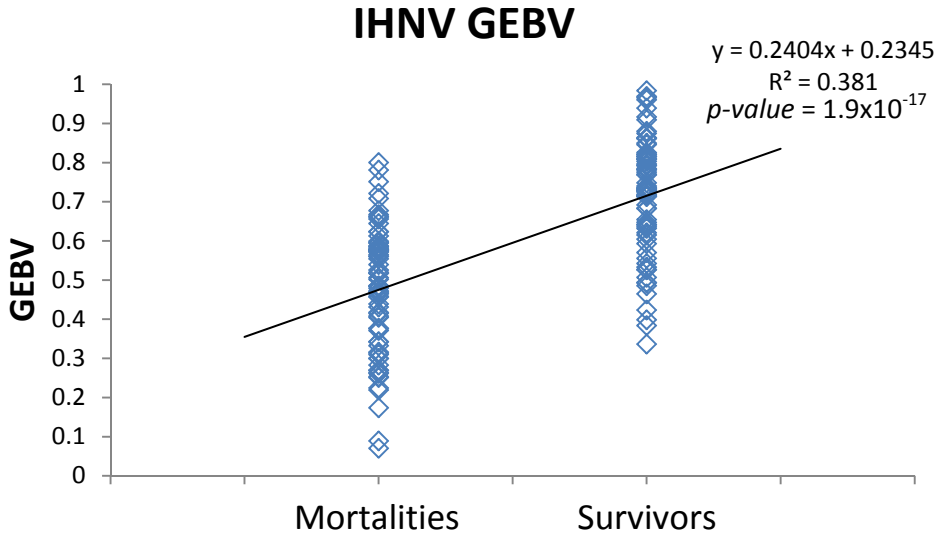


Table 1: Families evaluated for disease resistance and chosen for inclusion in association testing.

Family ID	Disease Challenge	Year Class	Mortalities (Genotyped/Target)	Survivors (Genotyped/Target)	Mortality rate
08-130	CWD	2008	2/2	3/5	0.04
10-125	CWD	2010	0/3	3/5	0.06
08-121	CWD	2008	2/4	5/5	0.08
08-092	CWD	2008	5/5	5/5	0.12
08-073	CWD	2008	5/5	4/5	0.13
10-060	CWD	2010	5/5	5/5	0.14
10-078	CWD	2010	3/5	5/5	0.16
08-105	CWD	2008	5/5	5/5	0.28
10-028	CWD	2010	5/5	5/5	0.36
10-127	CWD	2010	5/5	0/5	0.61
10-101	CWD	2010	4/5	4/5	0.64
08-025	CWD	2008	3/5	5/5	0.65
10-094	CWD	2010	3/5	3/5	0.69
08-129	CWD	2008	1/5	4/5	0.70
08-082	CWD	2008	5/5	5/5	0.79
10-119	CWD	2010	5/5	4/5	0.80
08-075	CWD	2008	4/5	5/5	0.86
10-076	CWD	2010	4/5	5/5	0.89
10-068	CWD	2010	4/5	5/5	0.91
08-108	CWD	2008	5/5	4/4	0.91
10-059	IHNV	2010	1/2	5/5	0.04
10-053	IHNV	2010	3/5	5/5	0.10
08-113	IHNV	2008	0/2	5/5	0.17
10-015	IHNV	2010	4/5	5/5	0.26
10-007	IHNV	2010	5/5	3/5	0.27
08-094	IHNV	2008	5/5	5/5	0.27
08-117	IHNV	2008	5/5	5/5	0.28
10-001	IHNV	2010	5/5	5/5	0.29
08-008	IHNV	2008	3/5	4/5	0.33
08-043	IHNV	2008	1/5	1/5	0.33
08-110	IHNV	2008	5/5	5/5	0.33
10-050	IHNV	2010	5/5	5/5	0.75
08-093	IHNV	2008	5/5	4/5	0.80
10-067	IHNV	2010	5/5	5/5	0.83
10-023	IHNV	2010	5/5	4/4	0.91
08-107	IHNV	2008	5/5	3/4	0.91
10-021	IHNV	2010	5/5	4/4	0.92
08-039	IHNV	2008	4/5	1/3	0.94
08-018	IHNV	2008	1/5	0/3	0.94
10-124	IHNV	2010	5/5	2/2	0.96

Table 2: RAD markers showing significant association with disease resistance to cold water disease (CWD) or infectious hematopoietic necrosis virus (IHNV). Linkage group (LG) names are as reported by Miller et al. 2012 / Palti et al. 2011. Marker sequences follow the SbfI recognition sequence "CCTGCAGG". Best Linear Unbiased Predictor (BLUP) values are given for each locus.

Marker	**Disease**	***p-value***	***BLUP***	**LG**	**Marker sequence**
R45138	CWD	1.49E-05	1.78E-01	WS08/Omy16	TGGTCGCAAGGGGAAACATAGCCGCCATAGGCATCCTAAGCCTTTTAGGGC[T/A]GCAAAAC
R46743	CWD	5.31E-04	9.13E-03	WS08/Omy16	TTCATTCTCAACAACATCCAT[G/A]GAGTATGACCTTACCTCACAGAGGAAGCGGCGCAGGT
R26956	CWD	6.33E-04	1.23E-01	WS23/Omy17	GGAGGAAGAGAGGATGTGGGGGAGGAAGAGAGGATGTGGG[G/A]GAGGTTGGAAACATGATT
R46637	CWD	1.77E-03	-5.24E-02	ND	CCTTGGTAGTGGGCCGTGTCGGTGGCACTG[T/C]ATTATCCTCAAAGCTGGTAAAGAAGGTG
R08795	CWD	1.91E-03	1.42E-01	WS24/Omy25	CCACTTTCTGTCGCTCTCTTTGTCTCAGTCATTCCCTTTCTT[G/T]TGATTTCTTTCCCTTC
R13883	CWD	2.79E-03	-1.43E-01	WS13/Omy06	TGCAGTTCTGAAGGTTGATTCAGCAGTATTTGGGGCATCGAAGTGATGGGACAT[A/G]TGAC
R49259	CWD	3.06E-03	1.49E-01	ND	AAATCACGCACAGAACAAGC[A/T]GTACGGCTGGTGGTATTGTCATGCTTGAAGGTCATGTC
R34531	CWD	3.29E-03	-5.92E-02	WS13/Omy06*	TGATGCCCATGGCCTTGATCTTGGCCGA[T/C]TCGCTGCCCAGCTCGTTCAACTGCTGTTTA
R40622	CWD	3.52E-03	7.22E-02	WS01/Omy04	AGAGTAGA[C/T]CTCCAGGAACAGGTTAGGCCACCACTGACAAAGAACATCCAGCATTAGAG
R46597	CWD	3.99E-03	7.56E-02	WS23/Omy17	[C/T]GGTTTAGGACCTTTGTAATGACTTTCATTTAGCTAGCGCAAGTATTTGGTTCTGGGTT
R20199	CWD	4.03E-03	1.14E-01	WS01/Omy04	AAGGGTAACACATTTCCTGGGTTATTCAACTGGCCCCTCCGTCG[C/T]GACGTCCTCTGAAT
R53165	CWD	4.45E-03	-2.64E-02	ND	TCTGTGACATCAACATGCTGTTTTCCTCAGTAGGGTCATGCAAAAACAAATG[T/G]CACCTT
R21407	IHNV	5.98E-05	5.24E-02	WS13/Omy06	GAGTGAGAATGGAGGACAAGAAGGTAGATG[G/C]GGTAAGGCCATGGCCAGTCCCAACCACC
R14353	IHNV	9.52E-05	-1.33E-01	WS13/Omy06	AGTACATAACATGAC[C/T]GTACATATTTAATATGCTATTCAAGTGTTTTAAGACACCAATA
R24813	IHNV	8.95E-04	7.86E-02	WS13/Omy06*	TTTGACTCCTGATGAATGCCCCGTCTGCGGAATCTCTCTTGCTCGCTCTCTC[A/G]GCAATC
R53135	IHNV	9.53E-04	-1.23E-01	ND	CACATTGCCTGGTGGCGTTTATTAATGTTTACATACTGTACCGTTT[A/C]TTTTTTTTTAAA
R03138	IHNV	1.01E-03	9.38E-03	WS13/Omy06*	TGCCCGGCCCGCCACAAGGAGTCGCTTGA[A/G]CGAGTAAAGTCCCCCTGGTTGTGATACAG
R39227	IHNV	1.45E-03	1.49E-01	WS07/Omy11	C[C/A]TCCAGCTGCCTGCCTCACAGGCCAAATATGCTATTTAGAAAATTGGGATATAAGAGA
R53350	IHNV	1.63E-03	-1.53E-02	WS13/Omy06	CCAAACTCTCCCCTAACCCGGACGACGCTGGGAAAATTGTA[C/T]GCCGCCCTATGGGTCTC
R52586	IHNV	1.66E-03	5.55E-02	WS28/Omy26	TAGACTGTAC[T/C]AAGAGCTGAATCACAACTCCTGTATGGATAACACCACACTGGTCAAGT
R46528	IHNV	1.78E-03	-6.56E-02	WS24/Omy25	AGTGAGTCACTG[C/A]TTTGATGTTTGCAGAGCACGACAGAGTGTTGTGCAGGGTCATTGCA
R48860	IHNV	2.44E-03	1.37E-01	WS14/Omy10	CGCTCGCCC[T/C]GCCACAAGGAGTCGTTAGAACGCGATGAGCCAAGTAAAGCCCCCCAAGG
R33079	IHNV	2.46E-03	-1.16E-02	WS13/Omy06	CCATGGGCAACCTCACCCCTCTCCATAC[C/G]CCTACTCCTCCACAGTCCCCTCAGAGCTCA
R52674	IHNV	2.54E-03	-6.30E-02	WS13/Omy06	GGACCCACCAATCCTGAAGAAGTTAAAGACTCCAGCCATTCAAGC[C/A]ATAGACTGGCTGG
R19689	IHNV	3.29E-03	1.14E-01	WS03/Omy05	ATCATCTGAGACCAGCCACCCAGACAGCTGTGGGTTTGTGGGTTTGCACAACCAAATA[A/T]
R04597	IHNV	4.07E-03	-7.38E-02	WS03/Omy05	CTCATCCTGACATGACCCTGCAGCTTGACAATGTCACCAGCCATACT[G/A]CTCGTTCTGAG

R14972	IHNV	4.19E-03	1.80E-02	WS13/Omy06	CTACCACTCTTTCCCCAAGTAAGTA[T/C]CCGGGTGAATGGTTAAACAACTAGCCTAAACAG
R18850	IHNV	4.19E-03	-8.30E-02	ND	TCTGGTGTCAGAATGCGCA[C/A]GGAGCTCCTCAGTGGATTACTTCTGGTACAGAGAGACAC
R45284	IHNV	4.37E-03	7.09E-02	WS29/Omy13	CTACTATTAAAGTATCACTGCTTTGGTGTGTAATTCCAAT[G/A]CCCCTTTATCTTTTAAGT
R05136	IHNV	4.87E-03	-2.22E-02	WS14/Omy10*	ACGGG[A/T]ATCACAATAGGGAGGAGGATGTCTTCCCTGTAATGCACAGTGTTAAGATTGCC
R52799	IHNV	5.16E-03	-1.76E-01	ND	GCTACAGACAGAC[T/G]ATCACTCTATCATGGTAGGATGGATGGAGGGAGGGAGCACATCAC

* Directly matches mapped RAD locus from Miller et al. 2012.

SECTION 2: Genomic basis for adult run-timing in Chinook salmon

To investigate adult run timing in Chinook salmon, samples were provided from the Klickitat River from fish that display a range of run timing. These samples were RAD sequenced to genotype thousands of SNP markers to identify markers that are associated with run timing in this species. We will use the following two primary methods for identification of SNP associations with the run timing trait: mixed linear model for genome-wide association mapping (e.g. Zhang et al. 2010) and mapping by admixture disequilibrium (e.g. Montana and Pritchard 2004). The latter method can be applied to populations that are products of relatively recent admixture of two major lineages, which is the case for the Klickitat River spring-run Chinook salmon population making it an appealing system for study. In the case of the Klickitat River Chinook salmon admixed population, the two parent lineages were the interior stream-type lineage (native spring-run population) and the interior ocean-type lineage which is an exogenous stock with a summer-/fall-run life history. This admixture was likely human-mediated through hatchery practices (Hess et al. 2011).

To facilitate these analyses, we obtained a mapping family consisting of samples of 220 eyed eggs from a single parent pair that was spawned at Klickitat Hatchery in 2011 in order to create a linkage map. RAD sequence data has been generated for the linkage map and analyses will proceed in the upcoming performance period. The total raw number of SNPs that were identified using the single parent pair was 11,325, and after quality filtering, there were 6,771 SNPs that had complete genotypes for 70% of the mapping family.

Our goal for these analyses was to obtain 2.5 million sequence reads per individual, because this threshold was determined to yield approximately 70% of genotypes across a set of quality filtered SNP loci. At this threshold of 2.5 million reads per individual, to date our RAD sequence dataset includes 208 individuals for the linkage mapping family, 122 individuals with run-timing phenotypes for association mapping analyses, and 434 individuals from reference populations that represent the parent lineages involved in the Klickitat River spring Chinook salmon admixture (Table 1). Analyses to generate a linkage map and complete association mapping will continue in the upcoming year of this project.

Table 1. Sample sizes obtained with adequate numbers of RAD sequences per individual.

Collection	Objective	Lineage	Subbasin	Run type	#>2.5M
Single Parent Pair Cross	Linkage mapping family	ST	Klickitat River	Spring	208
Klickitat River Lyle Falls	Association mapping	ST	Klickitat River	Spring	122
McKenzie Hatchery	Reference collection/Admixture Disequilibrium Mapping	LC	Willamette River	Spring	39
Cowlitz River	Reference collection/Admixture Disequilibrium Mapping	LC	Cowlitz River	Fall	34
Spring Creek Hatchery	Reference collection/Admixture Disequilibrium Mapping	LC	White Salmon River	Fall	48
Wells Hatchery	Reference collection/Admixture Disequilibrium Mapping	OT	Upper Columbia River	Summer	41
Priest Rapids Hatchery	Reference collection/Admixture Disequilibrium Mapping	OT	Upper Columbia River	Fall	46
Lyons Ferry Hatchery	Reference collection/Admixture Disequilibrium Mapping	OT	Snake River	Fall	48
Wenatchee River	Reference collection/Admixture Disequilibrium Mapping	ST	Upper Columbia River	Spring	72
Rapid River Hatchery	Reference collection/Admixture Disequilibrium Mapping	ST	Snake River	Spring	44
Sawtooth Hatchery	Reference collection/Admixture Disequilibrium Mapping	ST	Upper Salmon River	Spring/Summer	22
Pahsimeroi Hatchery	Reference collection/Admixture Disequilibrium Mapping	ST	Upper Salmon River	Spring/Summer	40
Total					764

Note: Lineage refers to the three major genetic lineages in the Columbia River Basin (Narum et al. 2010): lower Columbia-type (LC), interior stream-type (ST), and interior ocean-type (OT). The number of individuals with greater than 2.5 million reads is indicated (#>2.5M).

References

Hess, J.E., A.P. Matala, J.S. Zendt, C.R. Frederiksen, B. Sharp, and S.R. Narum. 2011. Introgressive hybridization among major Columbia River Chinook salmon (Oncorhynchus tshawytscha) lineages within the Klickitat River due to hatchery practices. Canadian Journal of Fisheries and Aquatic Sciences 68:1876-1891.

Montana, G. and J.K. Pritchard. 2004. Statistical tests for admixture mapping with case-control and cases-only data. American Journal of Human Genetics 75:771-789.

Narum, S.R., J.E. Hess, A.P. Matala. 2010. Examining Genetic Lineages of Chinook Salmon in the Columbia River Basin. Transactions of the American Fisheries Society 139:1465-1477.

Zhang ZW, Ersoz E, Lai CQ, *et al.* (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* **42**, 355-U118.

SECTION 3: Ongoing/Future Studies

As an ongoing Accords project, preparations are underway for further evaluation of multiple traits such as age-at-maturity (jacking) in Chinook salmon and thermal adaptation in *O. mykiss*. More details are provided below regarding plans to investigate each of these traits in the upcoming performance period (July 1, 2013 – June 30, 2014).

Heritability and Genomic Basis for Age-at-maturity

To investigate age-at-maturity (jacking) in Chinook salmon, we are utilizing known pedigrees to evaluate heritability of jacking and the genomic basis for age-at-maturity. Migratory Chinook salmon typically reach maturity between the ages of three to five years. The three year old age class is almost exclusively comprised of males called "Jacks", which exhibit an alternative reproductive strategy. These males are smaller and less competitive breeders than older males and generally spawn with older females in a "sneaker" reproductive tactic. Overall the "Jack" life history comprises a greater proportion of the returning males in hatchery derived Chinook salmon than natural Chinook salmon in Johnson Creek, ID. The propensity to "Jack" is strongly associated with juvenile growth and rearing environment, where those males that enjoy a productive environment with plentiful resource will grow faster than they might in the wild leading to an increased likelihood to "Jack". For this study we used multi-generational pedigrees to estimate the heritability of "Jacking" in both hatchery and wild origin Johnson Creek male Chinook salmon, to determine the extent of the genetic contribution to this life history tactic.

Additionally, the known pedigree population of Chinook salmon from Johnson Creek provide the opportunity to investigate the genomic basis for age-at-maturity. Age at reproductive maturity in Chinook salmon is closely linked to overall body size at maturity and reproductive success, wherein older and larger returning Chinook salmon are capable of outcompeting younger and smaller conspecifics for reproductive opportunities. Additionally, larger fish produce more gametes, and thus are capable of producing more offspring. It is known that age at maturity is in part determined by a heritable genetic component in addition to environmental effects, and age at maturity is therefore susceptible to natural and artificial selection pressures. Many Chinook salmon hatchery programs report a shift in the age structure of their hatchery returns, with a larger proportion of younger and smaller fish returning, when compared to wild populations. It is possible that hatchery selection regimes are unintentionally selecting for a younger age at maturity in their programs. Additionally, selection in fisheries for larger fish could also drive selection toward younger and smaller returning adult salmon. Ultimately selection in the direction of smaller and younger maturing salmon could result in the decline of recruitment for each generation, as fewer and fewer offspring are generated in each cohort.

Advancing our understanding of the genetic basis of age at maturity, would potentially allow hatchery programs to screen their brood stock for genetic markers that would favor later age at

maturity. Current high-throughput sequencing and genetic marker discovery technologies in addition to advances in statistical approaches make it possible to identify regions of the genome associated with complex life history traits. Using these approaches we outline a project to sequence more than 400 Chinook salmon returns from Johnson Creek, a tributary of the Salmon River in Central Idaho, for more than 5,000 SNP markers distributed throughout the genome. The 400 target samples are subdivided into classes representing their origin (hatchery or wild), their sex (male or female) and their age at return (3, 4, or 5 years of age) in addition to their year of birth (Table 1). By identifying these categories, we can test different hypotheses about the effects and interactions of sex, origin, and cohort environment on the age at maturity, all factors suspected of contributing to the overall trait variation. The primary aim of this study, however, is to identify genetic markers closely linked to variation in age at maturity. Achieving this goal, we would contribute substantially to the early phases of dissecting the genetic basis of age at maturity in Chinook salmon, and potentially other salmonid fishes.

Table 1 A summary of the 408 Chinook salmon samples that will be used for SNP genotyping and genetic association mapping from Johnson Creek, ID. Samples are categorized based on their origin, either hatchery derived or wild (no hatchery influence), sex, age at maturity (3, 4, or 5 years of age), and Brood Year (BY, representing the year they were born).

	Hatchery						Wild						
	Male			Female			Male			Female			
BY/Age	3	4	5	3	4	5	3	4	5	3	4	5	Total
1998	15	8	31		9	10	10	9	18	1	7	10	128
2003	17	8	6	1	9	10	3	8	18		8	17	105
2004	17	8	1	1	9	6	22	11	4		10	10	99
2005	17	8	1		9	1	14	9	2		9	6	76
Total	66	32	39	2	36	27	49	37	42	1	34	43	408

Thermal adaptation in O. mykiss

For thermal tolerance, further testing is ongoing to determine the relationship between allelic association of genes with thermal tolerance and gene expression data. This study will use a combination of RAD-seq and RNA-seq to determine associations of fish from desert and montane streams. Additionally, *O. mykiss* populations throughout the Columbia River Basin will be RAD sequenced to allow the opportunity to characterize genetic adaptation across a broader range of the species.