2011-2012 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: 7/1/11-6/30/12
Submitted: 6/30/12

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, 2011 to June 30, 2012, 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 thermal tolerance in *O. mykiss* (Section 1) and smoltification in *O. mykiss* (Section 2). However, progress was also made towards developing projects to investigate disease resistance in *O. mykiss*, adult run-timing in Chinook salmon, and further work on thermal tolerance in *O. mykiss* (Section 3).

Table of Contents

ABSTRACT/SUMMARY **2**
Acknowledgements **4**
Report Structure 4
SECTION 1: Thermal adaptation of *O. mykiss* 5
Introduction 5
Methods 6
Results 7
Discussion 9
References 11
SECTION 2: Genetic basis of smoltification in *O. mykiss* 18
Introduction 18
Materials and Methods 20
Results 24
Discussion 28
References 31
SECTION 3: Ongoing/Future Studies 44
References 45

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, Chris Frederiksen and field assistants of YKFP, Scott Blankenship of WDFW, Scott Lapatra of Clear Springs Foods, and Dennis Scarnecchia of U. of Idaho. 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 thermal adaptation, the second section on smoltification, and the third section provides brief summaries for ongoing and future work on several traits of interest including disease resistance, run-timing, heritability of age-at-maturity, and ongoing work for thermal adaptation.

SECTION 1: Thermal adaptation of *O. mykiss*

Introduction

Environmental temperatures have extensive biological implications for all organisms, but ectotherms in aquatic systems are particularly affected by thermal profiles and climate regimes. This is because aquatic ectotherms primarily exchange heat with their environment through conduction and convection, and their body temperature closely follows the temperature of surrounding water. Thus environmental temperatures greatly influence the distribution, physiology, and behavior of aquatic organisms and local adaptation to thermal systems is the norm (e.g., Schluter 2000). Further, many aquatic organisms have developed capacities for thermal acclimation that provide greater tolerance to chronic exposure to stressful temperatures (Tomanek and Somero 1999; Hoffman et al. 2003; Sinclair and Roberts 2005).

The processes of thermal acclimation and adaptation are often interdependent and distinct strategies enable species to occur over broad geographic ranges with highly variable climate regimes. Empirical studies have typically shown that populations at thermal extremes consist of specialists, and populations elsewhere contain generalists that function over a wider range of temperatures (reviewed in Pörtner 2010). However, temperatures may affect organisms differently, with shifts in reaction norms that cause changes in niche width, mean performance, and/or optimal temperature (Knies et al. 2006). In the case of thermal adaptation within a species that occurs over a geographical range with variable climate, reaction norms for performance could include multiple changes such as horizontal shifts related to environmental temperature along with reduction of niche width (Fig. 1). Further, initial exposure to extreme temperatures often causes greater tolerance of subsequent thermal extremes ("acclimatization", "preconditioning", or "hardening"; Beitinger and Bennett 2000; Hoffman et al. 2003; Sinclair et al. 2003; Sinclair and Roberts 2005; Pörtner 2010). For this study we define thermal adaptation as *evolution* of a population to an altered reaction norm for temperature, whereas thermal acclimation/acclimatization is any *phenotypic response* of individuals to environmental temperature that alters performance, and plausibly changes fitness (following Angiletta 2009). Therefore, acclimation responses of individuals can produce fitness advantages that produce evolutionary change in a population over generations (e.g., Pörtner 2010).

Anthropogenic induced climate change (IPCC 2007) has caused concerns that some fishes may go extinct or need to relocate due to limitations related to thermal tolerance (e.g., Perry et al. 2005; Pörtner and Knust 2007). However, thermal acclimation and adaptation of fishes from variable environments has not been well studied and thus limits the ability to predict the adaptive potential of natural populations under scenarios of climate change. Current molecular and genomic tools provide the opportunity to investigate the heat shock response of fishes from varying thermal regimes and link that information to adaptive regions of the genome that are under selection. Specifically, heat shock proteins (Hsps) have been demonstrated to be induced and act as molecular chaperones in all organisms under a variety of stressors including heat stress (Feder and Hoffman 1999; Basu et al. 2002; Sorensen et al. 2003) and thus quantitative gene expression can provide a basis for the physiological response of individuals under experimental thermal conditions. Next-generation sequencing technologies allow for

dense genome wide association mapping of quantitative and binary phenotypic traits (i.e., heat shock response and survival/mortality, respectively). These approaches offer the power of genome wide association in combination with the physiological response of heat stress to identify patterns of both acclimation and adaptation in natural populations from varying thermal environments.

In this study we investigate patterns of thermal acclimation and adaptation in populations of redband trout (*O. mykiss gairdneri*) from differing aquatic environments. Redband trout are common in the interior region of the Pacific Northwest of the United States and occupy streams in both desert and montane climates (Meyer et al. 2010). Previous studies of this species have demonstrated local adaptation of populations to warm and cool thermal regimes (Narum et al. 2010). We test hypotheses related to the heat shock response and identify genomic regions associated with survival to elevated temperatures over a six week thermal stress experiment within and among strains from warm (Little Jacks Cr.) and cooler (Keithley Cr.) environments, plus their F1 crosses. Specific hypotheses include the following expectations: 1) acclimatization of the heat shock response over time in all strains, 2) adaptive heat shock response in the warm population relative to those from cooler environments, and 3) adaptive signatures of selection at genomic regions relevant to thermal stress.

Methods

Fish Populations:

Gametes and fry were collected from two locations (Little Jacks Cr. and Keithley Cr., Idaho USA) intended to represent warm and cool adapted populations based on previous studies (Narum et al. 2010). Gametes were fertilized to produce progeny of a pure warm adapted strain, pure cool adapted strain, and their F1 crosses. Fry were reared in constant 15°C spring water until they reached an average weight of 5 grams, and then each strain was divided into treatment and control groups. Three replicate tanks were used to estimate survival for all treatment and control groups. Fish in recirculating treatment tanks experienced diel temperature cycles over six weeks that reached a maximum of 28.5°C in the afternoon and a minimum of 17.0°C at night (mean temperature gradient of ~1.5°C per hour), while fish in control tanks were held at a constant temperature of 15°C (Fig. S1).

Gene expression:

Gill and liver tissues were sampled from fish euthanized with MS-222 at the time of peak temperatures from both treatment and control tanks on four separate days during the six week thermal stress experiment (Day1, Day3, Day7, Day28). Three fish were netted from each replicate tank for a total of nine samples representing each strain for each control and treatment group on each of the four sampling days. Tissues were stored in RNA-later and frozen at -80°C. Total RNA was isolated with RNeasy kits (Qiagen) and cDNA was made with kits from Ambion. Quantitative real-time polymerase chain (qRT-PCR) reactions with Syber Green assays (LifeTechnologies) were designed for six heat shock genes (Table S1) with sequences from Campbell and Narum (2009). All qRT-PCR reactions were completed with a standard curve and beta-actin as a reference gene. Expression values for each individual were normalized to beta-actin, and mean expression was calculated for each strain for control and treatment groups on each of the

four sampling days. The mean expression of unstressed control fish was subtracted from mean expression of treatment fish of each strain in order to correct for constitutive expression and any potentially confounding signals of expression (e.g., handling stress). Significance between strains and days were tested with ANOVA.

Genome wide association:
Fin tissue was collected from any fish mortalities by date during the course of the six week thermal stress experiment, and from all surviving fish at the end of the experiment. Tissues were preserved by dry storage on Whatman paper until DNA extractions were completed with DNeasy kits (Qiagen). Template DNA from all individuals was quantified with a spectrophotometer (Victor xx) and normalized. Samples were prepped for library construction with restriction-site associated DNA (RAD) protocols (Baird et al. 2008; Miller et al. 2011). Briefley, DNA was digested with Sbf1 and subsequently ligated with both a barcode adapter and an Illumina sequencing adapter. Barcoding adapters allowed multiple individuals (16-48) to be pooled in single libraries for sequencing on an Illumina HiSeq 2000 instrument with single-end 100 reads (U. of Oregon). Samples were sequenced to reach a minimum target of 1.5 million reads per individual and data was analyzed following the pipeline by Miller et al. (2011). Briefly, sequence reads of 100bp were trimmed to 85bp to remove errors on the terminal ends, and reads from 10 individuals were used for initial identification of SNPs and reads from all other individuals were subsequently aligned with Novoalign (Novocraft). Genotypes for each SNP were scored using criteria that required a minimum of five reads to call homozygotes, and a ratio of 9:1 alternative nucleotides to call a heterozygote. Another 188 SNPs from ESTs were genotyped with Taqman assays run on 96.96 integrated fluidic chips from Fluidigm with standard protocols as described in Narum et al. (2010). In total, 12,684 SNPs were genotyped across samples and tests for quality control and genome wide association were done in PLINK (Purcell et al. 2007). Two separate strategies were used for genome wide association tests with survival/mortality: first testing for association overall three populations taking stratification into account, and then testing for association within each of the three strains (Little Jacks Cr.., Keithley Cr., and F1 crosses). In order to reduce false positives, BY-FDR corrections for multiple tests were used to identify significant associations.

Results

A strong signal for acclimatization of the heat shock response was observed in all strains of redband trout as measured by quantitative polymerase chain reaction (qPCR) in six heat shock genes. In five of the six heat shock genes (Fig. 2a-f), mRNA expression was highly induced at Day 1 but was significantly ($p \leq 0.05$) reduced in the remainder of the six week thermal stress experiment (peak temperatures reaching ~28°C daily). By Day 3, levels of expression for hsp70, hsp90, hsp47, hsp27 (Fig. 2a-d) were significantly lower than Day1 ($p \leq 0.05$) and more similar to control fish that were held at a constant temperature of 15°C. In contrast to the strongly induced upregulation observed in the four hsp genes, the two transcription factors hsf1 and hsf2 (Fig. 2e-f) were downregulated and the signal of acclimatization was either weak (hsf2) or absent (hsf1). In the case of

hsf2, expression was not significantly different than Day1 until Day 7 or Day 28, depending on the strain.

In addition to acclimatization, patterns of adaptation were also observed in the heat shock response between the strains of redband trout from warm (Little Jacks Cr.) or cooler (Keithley Cr.) streams as well as their F1 crosses. In most cases, the relative expression of heat shock genes in fish from the desert population of Little Jacks Cr. was more similar to control (unstressed) fish and significantly different ($p \leq 0.05$) from those from the montane population of Keithley Cr. (Fig. 2a-f). The F1 crosses often had levels of expression that were intermediate to the pure strains, and were significantly different ($p \leq 0.05$) from at least one of the pure strains at Day1 in four of the genes (hsp70, hsp47, hsf1, hsf2). These patterns of adaptation were most prevalent at first exposure to high temperatures on Day1, but hsp47 (Fig. 2c) and hsf2 (Fig. 2f) showed consistently significant differences among strains over the course of the experiment even after acclimatization. This was not the case for hsp70 (Fig. 2a) and hsf1(Fig. 2e) as relative expression between strains was only significant on Day1, but not at Day3, Day7, or Day28. Other spurious differences in gene expression among strains were observed in hsp90 (Fig. 2b) and hsp27, but no clear patterns were evident.

Tissue samples were collected from individual fish to test for genome wide association with survival and mortality over the course of the six week thermal stress experiment. In general, mortality was highest during the first 24 hours of exposure to high temperatures. Interestingly, redband trout from the desert population of Little Jacks Cr. had higher mortality ($p = 0.056$) in the first 24 hours than those from the montane population of Keithley Cr., but montane fish had highest mortality over the remainder of the six week experiment (no significant differences among strains; Fig. 3). Mortality of F1crosses was not significantly different from either of the pure strains despite higher rates in the first 24 hours and lower rates in the remainder of the experiment than Keithley Cr. fish (Fig. 3).

Tests for genome wide association with survival/mortality revealed regions of the genome that are under selection for thermal stress. Greater than 10,000 SNPs were genotyped in individuals with restriction-site associated DNA (RAD) sequenced tags, and tests for stratified allelic association across all strains identified 17 SNPs that were highly significant after corrections for multiple tests to avoid false positives (Table 1). The majority of these associated SNPs were from unknown or unannotated genes, but two were from 3' UTR regions of known ESTs of hsp47 (involved in heat shock response) and Na–K-ATPase-α3 (involved in osmoregulation and oxygen transport in gills). Additionally, peaks of genomic association in LD blocks were identified on chromosomes x, y, z (Fig. 4) and are strong candidate regions for further studies to pinpoint polymorphisms responsible for adaptation to thermal stress.

Association of SNPs within each strain followed an expected pattern of adaptation among strains where there was much greater standing variation for selection to act upon in montane fish from Keithley Cr. and F1 crosses than in the strain fro Little Jacks Cr. There were 9 SNPs significantly associated with survival/mortality in the Keithley Cr. strain and 14 in the F1 crosses as opposed to none in the desert strain from Little Jacks Cr. (Table S2). Of the 9 significantly associated SNPs from Keithley Cr., many were from unknown genes but hsp47 was identified as one of the most significantly associated

genes (p = 9.86 e-09). Two other heat shock genes were associated at lower significance levels (hsf2 and hsc71) along with other candidates such as Na–K-ATPase-α3 (Table S2).

Discussion

When individuals in natural populations are exposed to thermal stress there are limited possibilities for response including individual efforts to avoid the stressor (movement, behavior, physiological changes), the population may adapt over generations to tolerate the stressor, or the population may be extirpated. The current study in redband trout demonstrates that these fish undergo acclimatization, but also have an adapted heat shock response in a portion of their range in desert environments.

Acclimatization of the heat shock response occurred in all strains of redband trout under repeated thermal stress over a period of six weeks. The acclimatization response was rapid, occurring within Day3 of exposure to elevated temperatures and continuing through the remainder of the study. However, studies of the heat shock response in model species such as *Drosphila* (Sorensen et al. 1999) and other aquatic ectotherms (Tomenek and Somero 1999; Podrabsky and Somero 2004; Portner et al. 2010) have shown that acclimatization can occur more quickly and it is possible that finer scale temporal sampling would identify a similarly timed response in redband trout. Acclimatization and reduction of the heat shock response has been shown to occur because the benefit of increased stress resistance becomes hampered by physiological costs of producing Hsps such as high energy demands, impaired growth, and reduced fitness (Sorensen et al. 2003).

In addition to acclimatization, results from our study demonstrate that an adaptive heat shock response has evolved in redband trout in desert environments that may reduce energetic demands of Hsp production and costs to development and fitness. High induction of Hsps in the montane strain appeared to have improved short term survival during the first exposure to high temperatures, but demands of Hsp production may have led to poorer long term survival. In contrast, the desert strain had significantly lower heat shock response than the montane fish and F1 crosses, suggesting that these desert fish have adapted alternative mechanisms to deal with thermal stress with less physiological costs. While low Hsp response in the desert strain was likely responsible for relatively high mortality at first exposure to elevated temperatures, stochastic temperature profiles from nature could not be exactly replicated in a laboratory environment and an abrupt shift to experimental temperatures may have exceeded upper thermal tolerance limits of all strains. In particular, fish in the laboratory were reared at a constant temperature near their optimum (15°C) for several weeks before being moved directly into diel temperature regimes that increased approximately 1.5°C per hour to a peak of nearly 28°C (Figure S1). While this was an effective study design to induce Hsps in controlled environments to demonstrate differences among strains, more gradual seasonal changes are encountered in nature even in desert environments.

Similar adaptive heat shock response has been observed in both laboratory experiments (Bettencourt et al. 1999; Sørensen et al. 1999; Lansing et al. 2000) and studies of natural populations (Sørensen et al. 2001; Fangue et al. 2006) of thermal stress, with populations from warmer environments expressing lower levels of Hsps than those from cooler environments due to evolution of alternative mechanisms to deal with stress

response and costs related to Hsp expression (Sørensen et al. 1999; Sørensen & Loeschcke 2002b). The phenomenon of adaptive heat shock response observed in redband trout and other organisms indicates that physiological costs to expressing Hsps are common in nature and organisms evolve additional mechanisms to lower these costs. It is yet unknown if parallel solutions among organisms are made to reduce these costs, but oxygen transport across gill membranes is expected to be critical in aquatic organisms due to oxygen limitations (Portner 2010), and one such gene (*Na–K-ATPase*) involved in osmoregulation has been determined to be a candidate in both European flounder (*Platichthys flesus*; Larsen et al. 2007) and our current study of redband trout.

Results from genome wide association tests validated the adaptive heat shock response and found significant association at a SNP in the 3'UTR of the hsp47 gene, but also identified multiple other genomic regions that were associated with survival under thermal stress. These genomic regions are strong candidates for alternative mechanisms regulating the response to thermal stress. Previous studies in marine fishes have shown that oxygen delivery is limiting in climate related stressors (Portner and Knust 2007), and thus genes involved in oxygen transport are expected to play a significant role. Additionally, we expect that metabolic and immune pathways could be involved given the energy demands and potential for disease under thermal stress.

Climate induced thermal stress is expected to increase in future generations and may have significant impacts on marine (Perry et al. 2005) and freshwater fishes (Ficke et al. 2007). However, freshwater fishes typically have less potential to alter their distribution due to natural and anthropogenic barriers (e.g., waterfalls and dams, respectively) and may be forced to adapt or be extirpated. Therefore, studies that investigate the adaptive responses of thermal stress are critical for understanding the potential for adaptation in species such as redband trout that currently occupy broad geographic regions with different climates. Our results identify genomic regions under selection for thermal stress and these may be utilized to screen broadly across the species range to help predict the potential for adaptation to various scenarios of climate change. Our study indicates that populations from warm environments may already have an adaptive heat shock response and populations from cool climates may retain suitable genetic diversity to adapt to changing climate. However, the largest conservation concern is expected to be for populations that are adapted to very cold environments and no longer retain the genetic variation to adapt to shifting climates.

References

Schluter D (2000) *The Ecology of Adaptive Radiation* (Oxford University Press, Oxford).

Tomanek L, Somero GN (1999) Evolutionary and acclimation-induced variation in the heat-shock responses of congeneric marine snails (genus *Tegula*) from different thermal habitats: implications for limits of thermotolerance and biogeography. *J Exp Biol* 202: 2925-2936.

Hoffmann AA, Sørensen JG, Loeschcke V (2003) Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *J Thermal Biol* 28:175–216.

Sinclair BJ, Roberts SP (2005) Acclimation, shock and hardening in the cold. *J Therm Biol* 30:557-562.

Pörtner HO (2010) Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating climate-related stressor effects in marine ecosystems. *J Exp Biol* 213:881-893.

Knies JL, Izem R, Supler KL, Kingsolver JG, Burch CL (2006) The genetic basis of thermal reaction norm evolution in the lab and natural phage populations. *PLoS Biol* 4:1-8.

Beitinger TL, Bennett WA (2000) Quantification of the role of acclimation temperature in temperature tolerance of fishes. *Env Biol Fish* 58:277-288.

Sinclair BJ, Vernon P, Klok CJ, Chown SL (2003) Insects at low temperatures: an ecological perspective. *Trends Ecol Evol* 18:257-262.

Angilleta MJ (2009) *Thermal Adaptation: a Theoretical and Empirical Synthesis* (Oxford University Press, Oxford).

Intergovernmental Panel on Climate Change (2007) *Climate Change 2007: Synthesis Report, Forth Assessment Report of the Intergovernmental Panel on Climate Change*.

Perry AL, Low PJ, Ellis JR, Reynolds JD (2005) Climate change and distribution shifts in marine fishes. *Science* 308:1912-1915.

Pörtner HO, Knust R (2007) Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science* 315, 95-97.

Feder ME, Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Phyiol* 61:243–282.

Basu N, et al. (2002) Heat shock protein genes and their functional significance in fish. *Gene* 295:173–183.

Sørensen JG, Kristensen TN, Loeschcke V (2003) The evolutionary and ecological role of heat shock proteins. *Ecol Lett* 6, 1025–1037.

Meyer KA, Lamansky J, Schill DJ (2010) Biotic and abiotic factors related to redband trout occurrence and abundance in desert and montane streams. *West N Am Naturalist* 70**:**67-76.

Narum SR., Campbell NR, Kozfkay CC, Meyer KA (2010) Adaptation of redband trout in desert and montane environments. *Mol Ecol* 19:4622-4637.

Campbell NR, Narum SR (2009) Identification and characterization of heat shock response related single nucleotide polymorphisms in *O. mykiss* and *O. tschawytscha*. *Mol Ecol Res* 9:1460-1559.

Baird NA, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376.

Miller MR, et al. (2011) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Mol Ecol* doi:10.1111/j.1365-294X.2011.05305.x

Purcell S, et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559-575.

Sørensen JG, Michalak P, Justesen J, Loeschcke V (1999) Expression of the heat-shock protein HSP70 in *Drosophila buzzatii* lines selected for thermal resistance. *Hereditas* 131:155–164.

Podrabsky JE, Somero GN (2004) Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus*. *J Exp Biol* 207:2237-2254.

Bettencourt BR, Feder FE, Cavicchi S (1999). Experimental evolution of Hsp70 expression and thermotolerance in *Drosophila melanogaster*. *Evolution* 53:484–492.

Lansing E, Justesen J, Loeschcke V (2000) Variation in the expression of Hsp70, the major heat-shock protein, and thermotolerance in larval and adult selection lines of *Drosophila melanogaster*. *J Thermal Biol* 25:443–450.

Sørensen JG, Dahlgaard J, Loeschcke V (2001) Genetic variation in thermal tolerance among natural populations of *Drosophila buzzatii*: down regulation of Hsp70 expression and variation in heat stress resistance traits. *Funct Ecol* 15:289–296.

Fangue NA, Nann AF, Hofmeister M, Schulte1 PM (2006) Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, *Fundulus heteroclitus*. *J Exp Biol* 209:2859-2872.

Sørensen JG, Loeschcke V (2002) Natural adaptation to environmental stress via physiological clock-regulation of stress resistance in *Drosophila*. *Ecol Lett* 5:16–19.

Larsen PF, Schulte PM, Nielsen EE (2011) Gene expression analysis for the identification of selection and local adaptation in fishes. *J Fish Bio* 78:1-22.

Ficke AD, Myrick CA, Hansen LJ (2007) Potential impacts of global climate change of freshwater fisheries. *Rev Fish Biol Fisheries* 17:581-613.

Figure 1. Reaction norms for performance may shift (width, mean, and/or peak) to become locally adapted to environmental conditions such as temperature. In the example shown here, the mean reaction norm for the species has shifted from the standard condition (black line) to either cool (blue line) or warm (red line) environments, and niche width has narrowed.

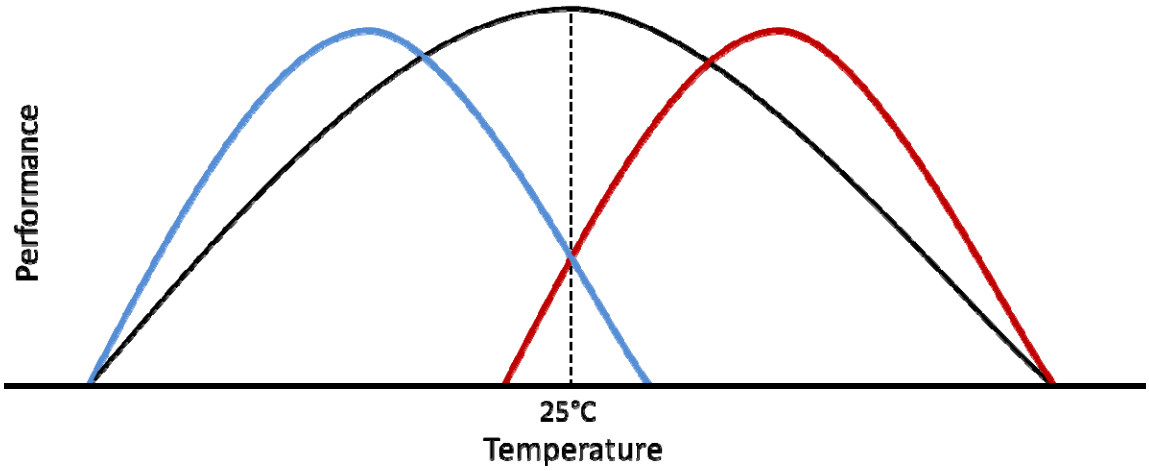


Figure 2. Mean relative gene expression of six heat shock genes for three redband trout strains reared under chronic diel thermal stress, a) heat shock protein 70, b) heat shock protein 47, c) heat shock protein 90, d) heat shock protein 27, e) heat shock factor 1b, f) heat shock factor 2. Mean values ± SE are shown for each of four collection dates for fish from Little Jacks Cr. (LJ, warm adapted), Keithley Cr. (K, cool adapted), and their F1 cross (LJ x K). Values shown are the difference between fish reared at treatment and control temperatures such that control values are equal to zero, and treatment values that are closer to zero are more similar to controls. Significant differences between strains at each time period are shown by letters above the bars.

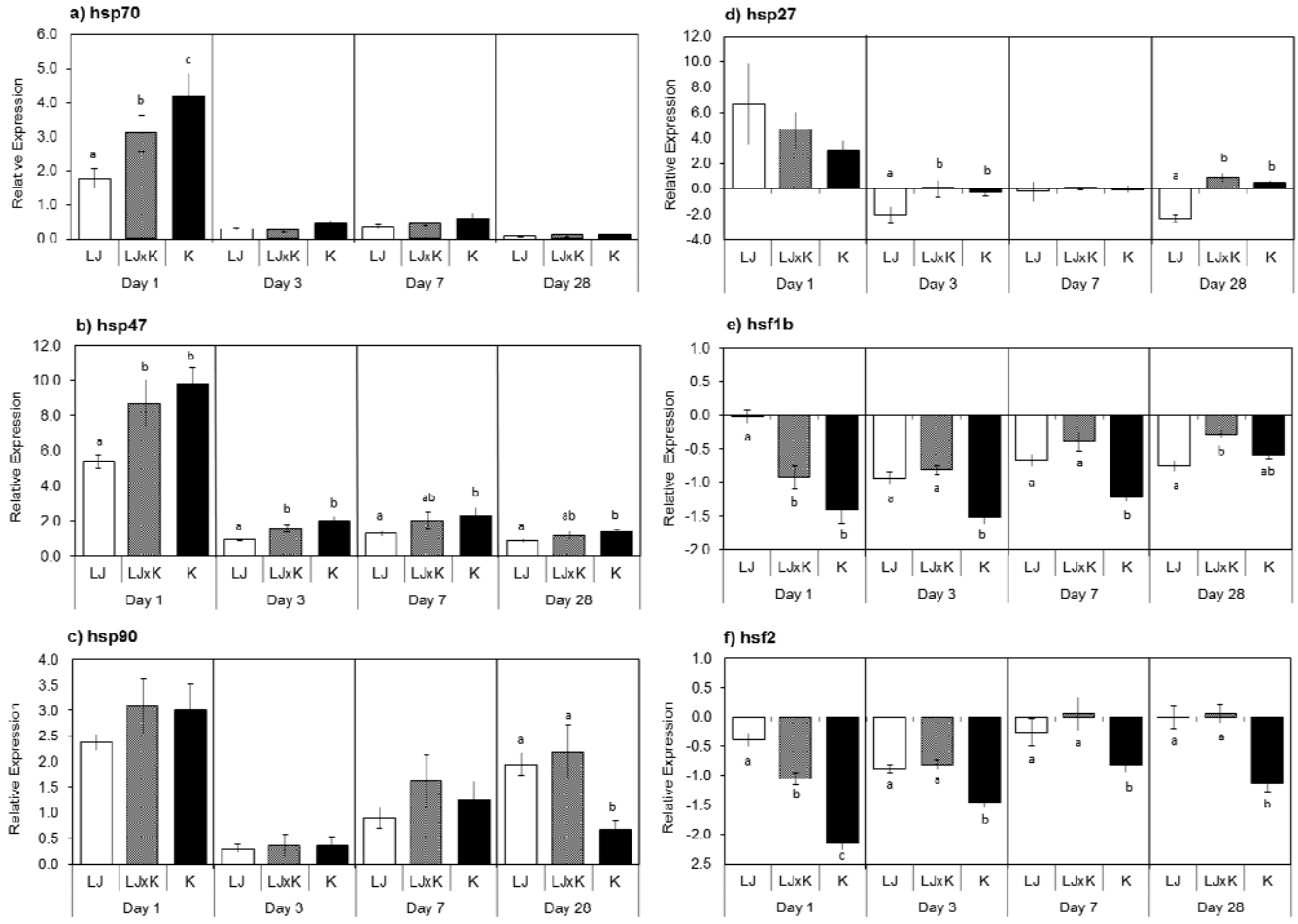
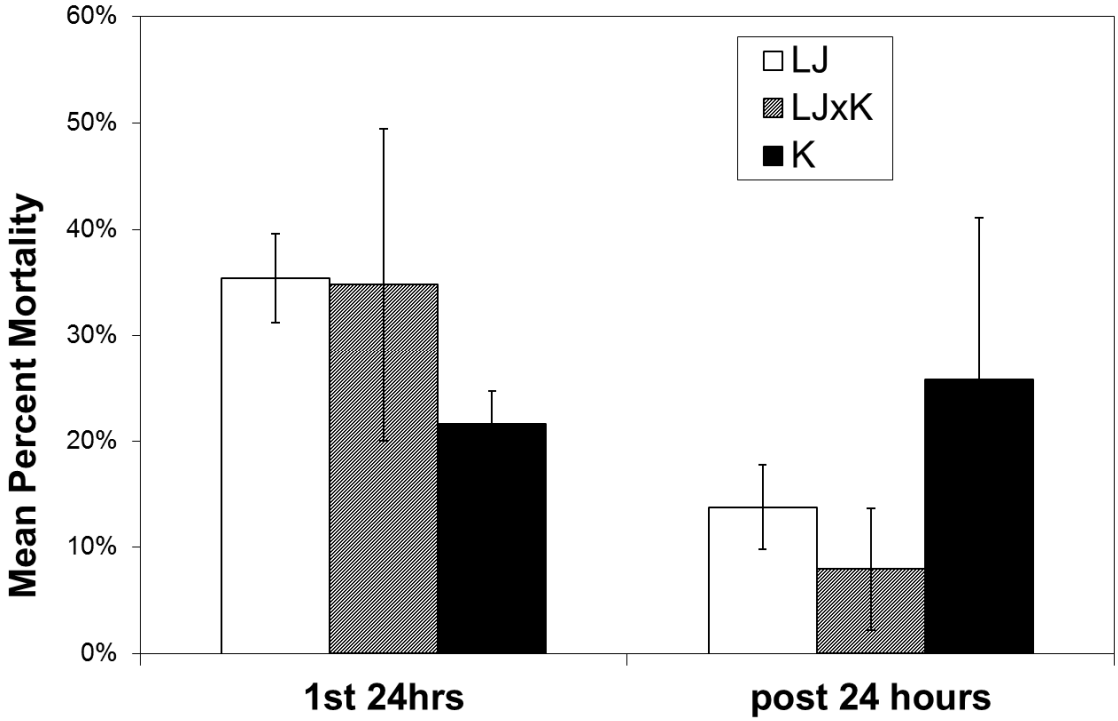
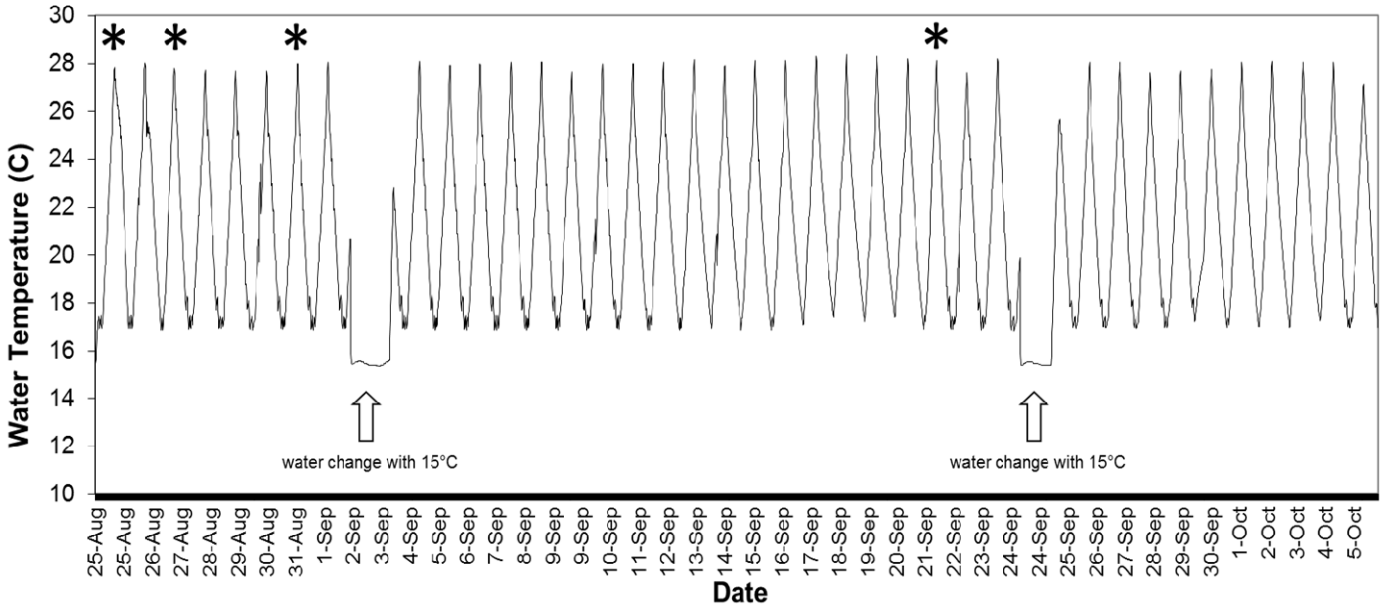


Figure 3. Mean percent mortality (± SE) for three strains of redband trout reared under chronic temperature stress. Results are shown for initial mortality after first 24 hours of exposure to diel thermal stress, and for the remaining six weeks of the experiment (post 24 hours). Strains are from Little Jacks Cr. (LJ, warm adapted), Keithley Cr. (K, cool adapted), and their F1 cross (LJ x K).



Supplemental Figure 1. Diel water temperatures of treatment tanks for a six week thermal stress experiment of redband trout. Control tanks were maintained at a constant 15°C throughout the experiment (not shown). Asterisks indicate the four dates when samples (liver and gill) were collected from fish for gene expression analyses. Two arrows show dates when water was exchanged with fresh water.



SECTION 2: Genetic basis of smoltification in *O. mykiss*

Introduction

Migration is a complex suite of physiological, morphological, and behavioral traits that act in concert with environmental cues to move animals over great distances (Dingle 2006; Dingle& Drake 2007). Migratory species throughout the globe provide tremendous ecological benefits and services by shuttling resources from nutrient rich to nutrient poor habitats. Perhaps one of the greatest examples of this resource shuttling can be seen amongst the salmonid family of fishes (salmon, trout, and charr) wherein anadromous migratory individuals transfer resources from the nutrient rich ocean back to their nutrient poor natal freshwater rivers and streams providing benefits and services in complex ways to entire ecosystems, from insects, to bears and trees (Quinn 2005; Wilcove& Wikelski 2008).

Within the family of salmonid fishes are species that consist entirely of anadromous migratory life-history types, species that remain entirely "resident" in natal freshwater habitats, and species that exhibit both anadromous and resident life-history tactics in sympatry (Quinn& Myers 2004). All salmonid fishes hatch and rear in freshwater and after a period of juvenile growth those that are destined to make the journey to sea undergo a complex transformation called smoltification. Smoltification prepares anadromous juveniles physiologically, morphologically, and behaviorally for their journey in the ocean, and is triggered by environmental queues including changes in body growth, photoperiod, water temperature, and water flow (Zaugg& McLain 1972; Hoar 1976; Wedemeyer *et al.* 1980; Quinn 2005). After a period of growth and maturation at sea, anadromous salmonids find their way home to their natal streams to spawn. Alternatively resident salmonids hold in freshwater, never undergo the process of smoltification, and reach sexual maturity without a marine migration.

Those fishes that undergo smoltification are called "smolts" and can be qualitatively and quantitatively differentiated from their resident or pre-smolt conspecifics allowing for the relative determination of an individual's life-history (McLeese *et al.* 1994; Beeman *et al.* 1995; Haner *et al.* 1995; Ando *et al.* 2005). During smoltification smolts morph to a more slender and fusiform body shape with decreased body condition (Hoar 1976; Beeman *et al.* 1995) and experience a drastic shift in body coloration, morphing from a colorful banded body pattern to a bright reflective silver scheme with a dark blue-green backs characteristic of other fishes that dwell in pelagic marine environments (Haner *et al.* 1995; Quinn 2005). Additionally as smolts begin to change they experience an increase in Na^+,K^+-ATPase activity level in their gills, allowing them to effectively osmoregulate once they reach the saline environment of the ocean (Zaugg& McLain 1972; Schrock *et al.* 1994). The culmination of the smoltification process is a behavioral change, wherein smolts shift from holding in their natal freshwater habitats to actively migrating downstream to the ocean (Hoar 1958; Groot 1972; Quinn 2005).

Migratory salmonid populations provide a tremendous economic and cultural benefit to humans (Quinn 2005), however throughout their range they suffer from overexploitation (www.fao.org/docrep/015/i2389e/i2389e.pdf) with some populations extirpated and many others listed under the Endangered Species Act as being threatened or endangered for extinction (Gustafson *et al.* 2007). Salmonids are not the only migratory species to suffer this fate as many of the globe's migratory species have either disappeared or their populations are in steep decline, largely due to human activities (Wilcove& Wikelski 2008). Despite the tremendous benefits migratory species provide to ecological and economic

systems, and given their rapid global decline, we know relatively little of the genetic basis of this complex life history in any species (Dingle 2006; Liedvogel *et al.* 2011).

In large part the lack of knowledge can be attributed to difficulties in designing robust genetic experiments in migratory species. Difficulties can stem from a lack of genetic resources, as many of these species are not model organisms, or difficulty in quantifying the migratory life-history trait and behavior (Liedvogel *et al.* 2011). Current high throughput genotyping methods, such as restriction-site associated DNA (RAD) tag sequencing using next-generation sequencing platforms (Miller *et al.* 2007; Baird *et al.* 2008), allow for the interrogation of hundreds to thousands of SNPs randomly distributed throughout the genome at a relatively low cost. These platforms make genetic and genomic tools available to non-model studies and can aid researchers in revealing the genomic regions and even genes contributing to trait variation in the wild. These methods include the construction of high density genetic linkage maps (Slate *et al.* 2009; Amores *et al.* 2011), population genomics (Luikart *et al.* 2003; Storz 2005; Hohenlohe *et al.* 2010), and association and quantitative trait loci (QTL) mapping (Johnston *et al.* 2011). As many migratory species face similar physiological and environmental challenges in carrying out their life-history it is purported that several of the associated traits have arisen through convergent evolutionary mechanisms. Therefore understanding the genetic basis of migration in one species could provide insight into the genetic basis of migration in other taxa (Dingle 2006) and ultimately assist in the protection and management of such species.

Rainbow trout (*Oncorhynchus mykiss*) are a salmonid species which exhibit tremendous life-history variation, including variation in the propensity to migrate. Throughout their native range rainbow trout populations exhibit either a wholly freshwater "resident" tactic in which the fish are called rainbow trout, an anadromous form called "steelhead", or populations with both rainbow and steelhead trout living in sympatry (Behnke 2002). Additionally rainbow and steelhead trout can give rise to one another in wild and experimental populations (Zimmerman& Reeves 2000; Thrower *et al.* 2004). It has been shown that migration is in part directed by an underlying heritable genetic component (Johnsson *et al.* 1994; Thrower *et al.* 2004) in addition to environment factors (Narum *et al.* 2008). Previous studies investigating the genetic basis of anadromy in rainbow and steelhead trout have identified several candidate markers (Limborg *et al.* 2012; Martinez *et al.* 2011; Narum *et al.* 2011) or quantitative trait loci (QTL) (Nichols *et al.* 2008; Hecht *et al.* in review) associated with the variable life-history. Some of the QTL regions that have been discovered are shared among studies, whereas others are unique to the mapping populations used, suggesting the potential for independent parallel mechanisms of migration arising within this species (Hecht *et al.* in review). While QTL mapping in experimental crosses are a powerful tool for detecting genetic loci contributing to trait variation (Doerge 2002), it is desirable in studies of evolutionary genetics to understand the genetic architecture of phenotypic variation in natural populations (Kruuk *et al.* 2008). Here we utilized RAD-tag sequencing to genotype thousands of polymorphic SNP loci throughout the genome in two wild populations of sympatric rainbow and steelhead trout. We aim to test the hypothesis that conserved genetic loci are associated with the propensity to migrate across geographic regions. Additionally we aim to determine whether a population that has been sequestered for more than fifty years following an anthropogenic barrier to migration still contains a genetic basis to the trait. We test these hypotheses using a genome-wide association study (GWAS) approach. Moreover, in the absence of a reference genome in any salmonid species, we use RAD-tag based SNP linkage maps from previous studies (Miller *et al.* 2012; Hecht *et al.* in review), to map SNPs to regions of the genome.

Materials and Methods

Genetic Samples

In order to detect loci associated with the propensity to migrate two populations of *O. mykiss* were sampled from the Pacific Northwest of the United States, each known to currently or historically harbor steelhead and rainbow trout living in sympatry. One sample was collected from an upper Yakima River population in the state of Washington, USA which drains into the Columbia River and maintains its migratory connection to the Pacific Ocean through fish passages in downstream dams. The other population is from Upper Mann Creek, a tributary of the Snake River in Idaho, USA which before the completion of the Brownlee Dam in 1958 had migratory access to the Pacific Ocean but has since been sequestered upstream of the dam (Holecek *et al.* 2012) (Fig. 1).

Tissue samples were collected from juvenile Yakima River fish in the spring of 2008 from the North and Middle forks as well as the main-stem of the Teanaway River (Fig. 1). Fish from these sites in the Yakima River were collected with standard electrofishing procedures. Both putative migratory smolts and non-migratory residents were collected and marked using passive integrated transponder (PIT) tags and released after a fin clip was collected. These same river reaches were surveyed during the summer and fall of the same year with the same methods. If a PIT tagged fish that was recovered displayed signs of spawning activity by exuding gametes, it was classified as a non-migratory resident fish having reached sexual maturity in its natal freshwater habitat (Quinn 2005) and assigned a value of "1" for the binary life history trait "SMOLT". If a PIT tag was detected at a downstream receiver array in the Yakima River or one of four Columbia River dams, the fish was classified as a migratory steelhead smolt, having out-migrated from its natal stream, a behavior characteristic of smolting salmonids (Groot 1972), and was assigned a value of "2" for the binary trait SMOLT. Only PIT tagged samples that could be assigned as residents or smolts based on the above criteria were used for further analyses from this collection.

Upper Mann Creek samples were subsampled and described in full in a previous study (Holecek *et al.* 2012). Briefly, scale samples were collected in the spring of 2009, between March 18th and June 3rd, the period at which steelhead smolts in this population would be expected to out-migrate to the ocean (Holecek *et al.* 2012). Samples were collected in a rotary screw trap and euthanized with a lethal dose (150 mg/L) of tricaine methanesulfonate (MS-222). At this time fish were measured for fork and total body length (mm), and scale samples were taken for both the estimation of age and for the extraction of for genetic analysis. Samples were categorized as putative residents or smolts based on overall body coloration and morphology following the methods of Negus (2003). Putative residents were those that had retained parr marks and had dark and/or colorful bodies and were assigned a value of "1" for the binary trait SMOLT. Putative smolts had a predominantly silver body color and lacked distinguished parr marks (Holecek *et al.* 2012) and were assigned a value of "2" for SMOLT. Silvering of juvenile salmonids is tightly linked to the smoltification process, and is a key determinant of migratory propensity (Haner *et al.* 1995; Ando *et al.* 2005). At the time of sampling a segment of the first gill arch was also collected in order to measure gill Na^+,K^+-ATPase activity ($\mu mol\ Pi \cdot h^{-1} \cdot mg\ protein^{-1}$) as further described in Holecek *et al.* (2012). Increased gill Na^+,K^+-ATPase activity is expected to be higher in smolting salmonids compared to pre-smolt or non-migratory resident fish and is thus a quantitative measure of the seawater readiness of an individual (Zaugg& McLain 1972; Folmar& Dickhoff 1980).

DNA was extracted from dried fin or scale tissue samples using a DNeasy Blood and Tissue Kit with additional RNase A treatment (Qiagen, Valencia, CA, U.S.A.) following the manufacturer's recommended protocols. Extracted genomic DNA was quantified using

Quantit PicoGreen dsDNA Assay Kits (Invitrogen, Grand Island, NY, U.S.A.) and a Victor2 microplate fluorometer (Perkin Elmer, Waltham, MA, U.S.A.).

Molecular Biology

Restriction-site Associated DNA (RAD) (Miller *et al.* 2007) libraries prepared for Illumina sequencing were produced using a protocol modified from one previously published (Miller *et al.* 2012). The protocol was modified to allow for the use of 150 ng of total gDNA, which was necessary when working with scale tissue samples that yield low quantities, by bringing all of the *SbfI* digested and P1 RAD barcode adapted gDNA forward for each sample when pooling samples into a library. In this manner multiple aliquots of each library were produced for shearing in a Bioruptor 300 sonication device (Diagenode, Denville, NJ, U.S.A.). Here we pooled 38 or 39 uniquely barcoded samples per library, in a total of five libraries. Each barcode sequence contained six bases that differed from any other barcode within the same library by at least two bases. The library aliquots were combined after sonication and concentrated using a MinElute PCR Purification kit (Qiagen, Valencia, CA, U.S.A.) in preparation for agarose gel size selection. The remainder of the library preparation follows the methods outlined in Miller *et al.* (2012). Prior to sequencing, RAD libraries were quantified using real time PCR and an Illumina Library Quantification Kit following the manufacturer's instructions (Kapa Biosystems Inc, Woburn, MA, U.S.A.) on an ABI 7900HT Sequence Detection System (Life Technologies, Grand Island, NY, U.S.A.). Libraries were sequenced on an Illumina HiSeq2000 sequencer (Illumina Inc., San Diego, CA, U.S.A.) at a single read length of 100 bp.

Among migratory salmonid fishes, females have a higher tendency to undergo the smoltification process and migrate to sea than males (Dellefors& Faremo 1988; Jonsson *et al.* 1998). Sex was therefore scored, so that it could be used as a covariate in GWAS models to account for the effect of sex on the binary life history trait SMOLT or ATPase activity. Males were scored as a "1" and females were scored as a "2". Samples that could not be definitively sexed based on the expression of gametes at the time of collection were sexed using a rainbow trout genetic sex marker, OmyY1, described elsewhere (Brunelli *et al.* 2008). This marker has been shown to be between 94-98% accurate at calling the actual sex of Columbia and Snake River rainbow trout (Brunelli *et al.* 2008) and is thus an acceptable proxy for phenotypic sex. For a small number of individuals with limited DNA for which we could neither determine the sex phenotypically or using the genetic marker OmyY1, sex was imputed based on the average sex of the life history trait SMOLT from which it was assigned. In this manner the individual for whom sex was imputed is non-informative for the covariate, but still informative in GWAS models for its life-history classification and/or for its ATPase activity.

Bioinformatics and Genotyping

Genotyping and SNP discovery was performed using Perl scripts and a bioinformatics pipeline provided and detailed in Miller *et al.* (2012) which incorporates the alignment tool Novoalign (Novocraft, Selangor, Malaysia). Briefly, 100 bp reads were trimmed from the 3' end to 71bp to remove the portion of the read that is most prone to sequencing error and to reduce the probability of observing multiple SNPs in a single read which can increase the number of false SNPs discovered. Trimmed reads were filtered using quality scores to eliminate poor quality reads and those that contained one or more ambiguous base calls. From the 5' end of the read the six base barcode sequence and partial *SbfI* site sequence were also removed after reads had been separated based on their unique barcode sequence.

SNPs were discovered using filtered reads from ten individuals which had an overall quality filtered sequence depth greater than three million reads. Selecting ten individuals allowed for the discovery of alleles with a frequency of 0.05 or greater, which was our target minor allele frequency (MAF) cut off. Three million randomly selected reads were collected from each of the ten selected samples to assure that each sample was contributing an equal proportion of the overall variation in sequence. Six of the ten individuals were from the Upper Yakima population, with three being smolts, and three being residents. One smolt and resident pair were taken from each of the three sampling locations (North Fork Teanaway, Middle Fork Teanaway, and main stem Teanaway) in the Upper Yakima. Reads from these six samples were pooled as if from a single individual representing the Upper Yakima population into a single FASTA file. Four samples were taken from Upper Mann Creek, with two being putative smolts, and the other two being putative resident fish. These four samples were also combined into a single FASTA file to represent the Upper Mann Creek population. Reads within these two FASTA files were collapsed such that each unique sequence is represented only once in the file and each header line contains the population name, a unique identification number for the read, and the number of read occurrences for that particular sequence. These two FASTA files were then combined to yield a single FASTA file containing all the unique reads from each population. The reads from this concatenated FASTA file containing all possible sequences from ten individuals were then used to make the SNP index, or a catalogue of all unique RAD tag sequences as described further in Miller *et al.* (2012) using the alignment program Novoalign (Novocraft, Selangor, Malaysia). The resulting alignment represents a map of the pairwise alignment scores for sequences within (internal alignments) and between (external alignments) the two populations. This map was used to group distinct loci with the following criteria: a sequence read had to occur more than six times, unless it had a perfect external alignment with a sequence that occurred more than six times. A sequence had to have at least one external alignment, to verify that the sequence occurred in both populations. No more than one SNP locus per RAD tag sequence and no more than two alleles per SNP locus was allowed. Once an index of loci was generated, each individual was aligned against the index according to methods described in Miller *et al.* (2012) to determine the presence and quantity of each allele sequence within an individual. Genotypes were only scored if the sum of both alleles' read counts were equal to or greater than five. If the sum of the read counts was less than five the genotype was scored as missing. Heterozygous genotypes were assigned if the ratio of reads between allele 1 and allele 2 (A1/A2) was between (10/1) to (1/10) at a locus. If the ratio was greater than (10/1) the genotype was scored as being homozygous for the A1 allele. If the ratio was below (1/10) the genotype was scored as being homozygous for the A2 allele.

The common ancestor of salmonid fishes experienced a whole-genome duplication event. As a result of this duplication, modern species experience residual tetrasomic inheritance at some loci (Sanchez *et al.* 2009). These duplicated loci can manifest themselves as false-positive SNPs, or paralogous sequence variants (PSVs) (Miller *et al.* 2012). In order to detect false-positive SNPs, we aligned RAD-tag sequences from four unique doubled haploid (DH) clonal lines of rainbow trout, which are homozygous at all loci. If one of the DH lines is heterozygous at a locus, we removed the locus from further analysis as being a putative PSV. The DH lines included Clearwater (CL), Oregon State University (OSU), Swanson (SW), and Whale Rock Reservoir (WR) lines and are further detailed elsewhere (Young *et al.* 1996; Nichols *et al.* 2007; Miller *et al.* 2012). The RAD-tag sequences for the DH lines were generated from previous studies and can be downloaded online from the NCBI Sequence Read Archive under project accession SRP008051 (Miller *et al.* 2012) and SRA052219 (Hecht *et al.* in review).

Polymorphic *SbfI* RAD-tag loci from this study were aligned against those discovered in two RAD-tag based rainbow trout linkage maps (Miller *et al.* 2012; Hecht *et al.* in review) to assist in the placement of our loci into one of the 29 rainbow trout linkage groups. Miller *et al.* (2012), identified 40,641 monomorphic and polymorphic RAD-tag sites, and named them R00001 to R40641. Additionally Hecht *et al.* (in review) have identified 4,980 unique polymorphic RAD-tag SNPs which were added to the 40,641 from Miller *et al.* (2012) and named R40642-R45621. We aligned our polymorphic RAD-tag loci against this composite database of 45,621 RAD loci to identify exact sequence matches, so that homologous SNPs could be identified between studies and our unique loci could be named following the nomenclature set forth in Miller *et al*. (2012). Having the ability to assign even a proportion of the SNPs to a linkage group in the absence of a completed salmonid genome assembly, would allow for a comparison to other studies which have identified regions of the genome under selection (Martinez *et al.* 2011), or QTL for smoltification related traits (Nichols *et al.* 2008; Hecht *et al.* in review).

GWAS for the propensity to migrate and ATPase Activity

Sex was tested for its effect on the binary life-history classification and ATPase Activity using an analysis of variance (ANOVA) in the statistical computing environment R (http://www.R-project.org/). A significant sex effect would qualify its use as a cofactor in statistical models described below. In order to test for genome-wide associations between RAD-tag SNP genotypes and the propensity to migrate, we employ a unified mixed linear model (MLM) approach, also known as a "*Q-K*" model, which simultaneously accounts for both population structure (*Q*) and cryptic familial relatedness (*K*) among samples (Balding 2006; Yu *et al.* 2006). It is known that both population structure and familial relatedness can cause spurious genotype-phenotype associations, and should be accounted for in tests of genetic association (Balding 2006). Population structure was estimated by analyzing a subset of 1000 RAD-tag SNPs with 100% genotyping frequency and a global MAF ≥ 0.1 using the program Structure version 2.3.3 (Pritchard *et al.* 2000). Structure was run to determine the most likely number of distinct populations (*k*) from 10 iterations for each potential *k*-value from 1 to 10 using a burn-in period of 10,000 followed by 100,000 MCMC repetitions. Once a single *k* value was determined based on criteria set forth by Evanno *et al.* (2005), three additional iterations (for a total of four) using this value of *k* were run, and the average coefficient of ancestry for each individual across the *k* clusters and four runs was calculated by the program CLUMPP version 1.1.2 (Jakobsson& Rosenberg 2007). Coefficients of ancestry were then used as cofactors in tests of association to account for underlying population structure (*Q*). Familial relatedness was estimated by calculating a kinship matrix using the complete set of RAD-tag SNPs across all fish samples used for GWAS. The R package GAPIT (http://www.maizegenetics.net/gapit) was used to generate the matrix using the EMMA algorithm (Kang *et al.* 2008).

Association tests were implemented with the compressed mixed linear model (MLM) approach (Yu *et al.* 2006; Zhang *et al.* 2010) and a general linear model (GLM) executed in the software program TASSEL version 3.0.121 (Bradbury *et al.* 2007), which employs the EMMA (Kang *et al.* 2008) and P3D (Zhang *et al.* 2010) algorithms to speed computation. Association tests were run with the following approaches: A) a global approach where 1) a GLM model for all samples was run using only sex (if a significant contributor to the variation) and population structure (*k* -1 coefficients of ancestry (*Q*)) as cofactors for SNP effects; and 2) a MLM where sex (if a significant contributor to the variation), population structure (*k* -1 coefficients of ancestry (*Q*)) and a genetic covariance matrix (kinship matrix, *K*) were used as cofactors for SNP effects; or B) a population level

approach where each sampling population (Yakima River or Upper Mann Creek) was run in a separate analysis using the GLM or MLM as outlined above. Significant associations were determined using a B-Y FDR p-value adjusted for multiple testing (Benjamini& Yekutieli 2001) at the $\alpha = 0.05$ level. This multiple test correction has been shown to be less conservative than a Bonferroni adjustment (Narum 2006), however, with a phenotype as complex as the propensity to migrate we feel it is more appropriate to identify plausible associations than to hold the analysis so stringent that only the largest effects can be detected. ATPase activity was tested using a GLM and MLM as outlined above for the Upper Mann Creek population only, as these were the only samples analyzed for this trait.

Prior to analyses, SNP markers were filtered for minor allele frequencies (MAF) of ≥ 0.05, and a minimum genotyping frequency of 80%. Markers were also tested for deviations from Hardy Weinberg Equilibrium (HWE) expectations for each sampling population individually to curtail Wahlund effects. Significance for HWE deviations was determined at the $\alpha = 0.05$ level adjusted for multiple testing using a Bonferroni correction. Loci that deviated significantly from HWE expectations were discounted in association tests to avoid false positives. Within the individual sampling location analyses (i.e. Yakima only or Upper Mann Creek only), markers were only pruned if they did not meet filtering criteria when measured within the population under scrutiny. In an effort to annotate the RAD-tag sequences of all of the loci found to be significant in all of the GWAS models outlined above, we performed a BLAST against the NCBI RefSeq collection using the blastn feature of the program Blast2GO (http://www.blast2go.com). In total, 67 bases of the 100bp RAD-tag sequence were blasted including the complete 8 base *SbfI* site on the 5' end of the sequence.

Results

Genetic Samples

A total of 132 samples that were collected from multiple sites of the upper Yakima River basin, were sequenced with the RAD protocol. However, five of these samples produced low yield RAD-tag sequence data, and were pruned from all analyses, leaving 127 samples from the upper Yakima River basin. Of those remaining 127 samples, 98 were classified as non-migratory resident rainbow trout exuding gametes at the time of collection and 29 were classified as migratory steelhead smolts having been surveyed by PIT tag at downstream dams. Of the non-migratory resident fish which could be definitively sexed based on the expression of gametes, 12 were female and 86 were male. The immature migratory smolts were sexed using the genetic sex marker OmyY1 (Brunelli *et al.* 2008). Using this marker we concluded that 14 of the smolts were female, 8 were male, and 7 were unknown (Table 1).

A total of 57 fish were sequenced from the Upper Mann Creek sample, with two samples removed from all downstream analysis due to low yield sequence data. Of the remaining 55 samples, 27 were categorized as putative non-migratory residents with an average fork length of 118mm (range 69-186mm), and an average scale age of 1.7 years (range 1-3 years). Twenty-eight samples were categorized as putative migratory smolts with an average fork length of 161mm (range 130-240mm) and an average scale age of 2.4 years (range 2-4 years). All of the Upper Mann Creek samples, except for two mature resident males, were sexed using the genetic sex marker OmyY1 (Brunelli *et al.* 2008). Based on these results, 16 of the resident fish were called as females, 10 as males and 1 was unknown. Of the smolts, 21 were called as females, 4 as males, and 3 were unknown (Table 1). In samples for which the sex could not be determined by methods outlined above, sex was imputed based on the average sex of the assigned life-history class of the individual. For

smolts the average sex was female, and unknown smolt samples (n=10) were assigned the sex of female, while for unknown resident fish the average sex was male, and unknown resident fish (n=1) were assigned the sex of male.

Sex was tested for its effects on both the binary life history trait and ATPase activity in an ANOVA model. Sex was determined to contribute significantly to the overall binary life history classification ($F = 72.216$, $P < 0.0001$), and was used as a cofactor in GWAS models for this trait. For ATPase activity there was no significant effect of sex detected ($F = 0.997$, $P = 0.322$) and sex was excluded as a cofactor in these models. When partitioning ATPase activity by life-history classification we see only a marginally significant difference ($t = -2.07$, $df = 53$, $P = 0.04$) between the mean activity levels for smolts (mean = 2.745, std.dev = 1.218, n = 28) and residents (mean = 2.199, std.dev = 0.634, n = 27).

Molecular Biology

Five RAD-tag libraries were prepared for 100bp single read sequencing. The average number of raw sequence reads across these five libraries was 184,231,878 (range = 165,592,275 – 204,779,007) with an average number of quality filtered reads per library of 125,033,942 (range = 107,332,212 – 137,514,197). The average number of quality filtered reads per individual was 2,935,039 (range = 1,007,399 – 6,052,847) when not considering those seven individuals pruned as discussed above. All raw Illumina sequence data was submitted to the NCBI SRA database under project accession SRA055295.
In total 12,073 polymorphic RAD-tag loci were detected between the 10 individuals and two populations used in the alignment index. Of those loci, 8,219 aligned perfectly to loci previously discovered (Miller *et al.* 2012; Hecht *et al.* in review) and were named accordingly (Table S1). Of those with perfect alignment, 4,563 were mapped to one of 29 linkage groups in Miller *et al.* (2012), and 415 were mapped in Hecht *et al.* (in review) for a total of 4,920 (58 of which were shared between the two linkage maps) of 12,073 polymorphic loci which have been previously mapped to a rainbow trout linkage group. Additionally 3,854 unique loci were discovered here, and named R45622 – R49475. Of the 12,073 loci in this study, 877 (~8%) were found to be heterozygous in at least one of the four DH fish controls. While it is possible that some of the loci in these DH samples were heterozygous not because they were truly PSVs or tetrasomically inherited, but could instead be the result of residual parental nuclear DNA, barcode jumping, or sequencing error (Miller *et al.* 2012) we opted here to err on the side of caution and remove all of these loci from further analyses leaving us with 11,196 loci.

GWAS for the propensity to migrate and ATPase activity

Coefficients of ancestry (Q) were estimated in order to account for underlying population structure in our sample using the programs STRUCTURE and CLUMPP. The results of this analysis provided support for $k = 4$ clusters (Table S2). The Yakima River basin samples were clustered into two groups, with the main stem Teanaway samples clustering into a single cluster, and the North and Middle Fork Teanaway samples clustering into another with some evidence of admixture from the main stem Teanaway population. The Upper Mann Creek samples also clustered into two groups with evidence of admixture between the two clusters. Clustering in neither population was associated with life history classification. It has been shown that the Upper Mann Creek population is a pure redband trout stock, native to the upper Columbia and Snake River systems (Kozfkay *et al.* 2011), and multiple small drainages which feed into Upper Mann Creek are likely to be the source of this structure.

Global SMOLT GWAS

For global analyses 517, loci were found to deviate from HWE expectations in the Yakima River sampling population. In Upper Mann Creek, 240 loci were found to deviate at the same significance level. While some loci may have deviated from HWE due to Wahlund effects, we conservatively discounted all SNPs with HWE deviations since they could be PSVs, genotyping errors, or sequencing errors. Markers were pruned that had a global MAF < 0.05 and a genotype frequency of < 80% across the samples. After filtering, the global SNP dataset consisted of 5,019 polymorphic RAD-tag SNPs for use in the global association tests.

The Global GLM identified 58 significant loci (range $P = 2.63E^{-6} – 5.34E^{-3}$), 12 of which are mapped to 10 different linkage groups in the RAD linkage maps (Table 2). These 58 loci had a mean MAF of 0.2 (range 0.05 - 0.48) and mean genotype frequency of 94.4% (range 81.3-100%). The global MLM identified 219 significant loci (range $P = 6.69E^{-8} – 5.44E^{-3}$), 49 of which are mapped to 19 different linkage groups from the RAD linkage maps (Table 2). These 219 significant markers had a mean MAF of 0.22 (range 0.05 – 0.5) and mean genotype frequency of 94.8% (range 80.2-100%). Additionally 10 of the 59 significant loci identified in the GLM overlapped with those identified in the MLM.

Yakima population SMOLT GWAS

When analyzing the data set within the Yakima sampling population, 6,858 loci were removed for not meeting the filtering criteria leaving 4,338 loci for analyses. When performing the GLM 107 SNP markers were found to be significantly associated with the binary trait (range $P = 1.83E^{-9} – 5.47E^{-3}$), including 22 markers which were assigned to 13 different linkage groups in the RAD linkage maps (Table 2). The mean MAF of those significant SNPs was 0.21 (range 0.05 – 0.49) and mean genotype frequency was 94.6% (range 80.3-100%). In performing the MLM in the Yakima samples, 50 loci were significantly associated with the binary trait (range $P = 4.28E^{-7} – 5.52E^{-3}$). Of those 50 markers, 11 were assigned to eight linkage groups in the RAD linkage maps (Table 2). Thirty-four of these 50 markers also overlapped with the significant markers found in the Yakima only GLM. The mean MAF of these 50 loci was 0.19 (range 0.05 – 0.49) with a mean genotype frequency of 95.1% (range 80.3-100%).

Upper Mann Creek SMOLT GWAS

When analyzing the data set within the Upper Mann Creek sampling population alone, 5,157 loci were removed for not meeting the filtering criteria, leaving 6,039 markers for analyses. When analyzing the binary trait within this population the GLM identified 22 significant loci (range $P = 5.34E^{-5} – 5.32E^{-3}$), seven of which were assigned to six linkage groups in the RAD linkage maps (Table 2). The mean MAF of these 22 loci was 0.22 (range 0.05 – 0.45) with an average genotype frequency of 98.4% (range 87.3-100%). The MLM model identified 138 significant loci (range $P = 1.4E^{-5} – 5.37E^{-3}$), 18 of which are assigned to 14 linkage groups in the RAD linkage maps (Table 2). Only three of the loci identified in this model were also identified in the GLM. The mean MAF of the 138 significant loci was 0.22 (range 0.05 – 0.49) with an average genotype frequency of 97.4% (range 81.8-100%). Across all six of the models analyzed for the binary trait, 504 unique loci have been found to be significantly associated with the propensity to migrate in at least one of the models (Table S3). Of those 504, 93 have been assigned to 27 of the 29 linkage groups from the RAD linkage maps (Fig. 2), with only Omy5 and Omy26 not represented among the mapped loci (Table 2). Three loci (R12248, R44513, and R46239) were significantly associated in four of

the six models, including both GLM and MLM models from the global analysis and the Yakima population analysis. One of these markers, R12248, has been assigned to linkage group Omy12. Ten loci (R00232, R01282, R15864, R16102, R29254, R32458, R35562, R40048, R47293, and R48642) were significant in three of the six models, with each model represented among these ten except for the MLM in the Upper Mann Creek population, which did not detect any of these loci to be significantly associated with the binary trait. Four of these ten have been mapped including R29254 to Omy2, R35562 to Omy10, R32458 to Omy11, and R00232 to Omy12. Sixty-one markers were found to be significant in at least two of the six models, with 15 of those markers being mapped in one of the RAD linkage maps. The remaining 430 significant markers were only detected as significant in one of the six models (Table S3).

The linkage group with the most significant loci detected across the binary analyses was Omy11, having 14 total instances (Table 2; Fig. 2). Ten of these 14 instances are unique loci, with three loci being detected in more than one test, including locus R32458, which was detected in three tests overall. The loci detected in the most tests genome-wide included R12248, R44513, and R46239, with R12248 being assigned to linkage group Omy12 (Fig. 2). Each of these three loci was detected in four of the binary trait models including both GLM and MLM from the global and Yakima-wise tests (Table S3).

Upper Mann Creek ATPase GWAS

Association tests for ATPase activity were carried out on the Upper Mann Creek sampling population in the same manner as described above for the binary trait analysis. Of the 6,039 markers analyzed, 49 were significant (range $P = 1.07E^{-4} – 5.31E^{-3}$) in the GLM model, ten of which are assigned to nine linkage groups in the RAD linkage maps (Table 2). The mean MAF of these loci was 0.19 (range 0.05 – 0.46) with an average genotyping frequency of 97.6% (range 81.8-100%). The MLM model identified ten significant markers (range $P = 1.52E^{-3} – 5.14E^{-3}$), three of which were assigned to three linkage groups from the RAD linkage map of Miller *et al.* (2012) (Table 2). The mean MAF of these ten loci was 0.22 (range 0 .07 – 0.40) with an average genotyping frequency of 99% (range 94.5-100%). All ten of these loci were also significant in the GLM with three of them assigned to three different linkage groups in the RAD linkage map of Miller *et al.* (2012) (Table 2). In total 49 markers were detected for association with ATPase activity, three of which were also detected in at least one of the binary analyses. Of these 49 loci, 10 were assigned to a linkage group in the RAD linkage maps (Table S3).

BLAST of significant loci

Performing a BLAST analysis with an eValue detection threshold of $1E^{-3}$ of all of the 550 significant loci detected between the eight GWAS analyses identified 280 alignment hits. The mean minimum eValue of these 280 hits was $8.93E^{-6}$ (range = $7.35E^{-26}$ - $3.81E^{-4}$) (Table S4). With the relatively short sequence length of the RAD-tag loci, we expect many of these hits with higher eValues to be less than reliable and suggest caution interpreting these results.

Comparison across the GWAS models

When comparing the results across the GWAS models we were interested in whether the same loci, or even regions of the genome based on mapped marker assignment, would be detected across the Yakima and Upper Mann Creek population for the SMOLT analysis. While several loci between these populations co-localize based on mapped markers, many do not (Fig. 2). Additionally very few loci detected in the ATPase analysis

co-localize with loci detected in the SMOLT analysis. A total of 550 significant loci were detected across all the GWAS models (Fig. 3).

Discussion

In this study, we utilized a genotyping-by-sequencing technique to interrogate thousands of RAD-tag SNPs in two wild populations of resident and anadromous rainbow and steelhead trout (*O. mykiss*) from the Pacific Northwest of the United States to investigate the genetic basis of migration in this species. One of the populations studied maintains its connection to the Pacific Ocean, allowing for the completion of anadromous migrations, while the other population has been cut-off from its access to the ocean for more than 50 years but still produces individuals that exhibit juvenile morphological characteristics of smolts (Holecek *et al.* 2012). Performing GWAS on thousands of RAD-tag based SNPs, we detected hundreds of loci significantly associated with the propensity to migrate both in individual population analyses and joint global analyses. Additionally we screened a subset of our sample for associations with ATPase activity and identified several loci associated with this trait.

In the absence of a reference genome sequence we utilize two RAD-tag based genetic linkage maps (Miller *et al.* 2012; Hecht *et al.* in review) to assign 68% of the loci discovered here to loci previously discovered. This suggests that RAD-tag based genotyping studies can be compared with a relatively high degree of overlap between studies within this species, making RAD-tag based genotyping approaches suitable to comparative analyses. Additionally 37% of the total polymorphic loci detected here, had previously been integrated into one of the two genetic linkage maps. Using these linkage maps we were able to assign 18% of the significant loci in the binary life-history trait analyses and 20% of the significant loci in the ATPase activity level analyses to linkage groups, which allow us to compare our GWAS results with previous results from QTL or genome scan studies. Additionally by combining our data with these previous studies we highlight that in any species where a RAD-tag based linkage map can be produced in either wild (e.g. Amores *et al.* (2011)) or experimental (e.g. Miller *et al.* (2012)) crosses, a substantial proportion of the RAD-tag markers mapped may be informative in other populations and crosses and thus available for GWAS or population based genome scans.

Results of the binary life-history analysis have identified several genetic regions significantly associated with this trait. Our results corroborate previous studies aimed at dissecting and detecting regions of the genome associated with migration related traits in this species, wherein each study using either a genome scan for signatures of selection (Martinez *et al.* 2011), or QTL mapping (Nichols *et al.* 2008; Hecht *et al.* in review), have identified loci across the genome on several different linkage groups. Indeed, migration is a complex life history trait and we do not expect a single locus to underlie this adaptive tactic. We do however aim to identify genes of large effect both within lineages of rainbow and steelhead trout or among lineages throughout their native range.

Based on the results here and in previous findings there are certainly linkage groups that have been implicated in multiple studies across multiple populations and families in varied geographic regions. For example a genome scan in a population of rainbow trout that had been transplanted above a barrier waterfall in California, when compared to the source population which retains access to the sea, show signatures of divergent selection at loci that map to linkage group Omy10 among several others (Martinez *et al.* 2011). This same linkage group harbors several QTL for migration related traits detected in a cross between a historic resident rainbow trout strain from California and a migratory steelhead strain from

Idaho, and has been suggested to be a master genetic switch for the migratory life history in this species (Nichols *et al.* 2008). Here we also detect loci to be associated with the binary life history classification on this linkage group, suggesting that there may be a conserved mechanism between these populations of rainbow and steelhead trout in this genomic region.

Two linkage groups in particular stand out in binary trait analysis here. Linkage group Omy11, which contained the highest number of significant associations across all the models based on mapped loci, however no mapped loci were significant in the Upper Mann Creek analysis on this linkage group. This linkage group is known to harbor several migration related QTL for osmoregulatory ability and body shape morphology (Le Bras *et al.* 2011; Hecht *et al.* in review). Another study (Martinez *et al.* 2011) has also identified two loci from this group that display evidence of differential selection between above barrier and below barrier populations of rainbow and steelhead trout in California. The other notable linkage group in this study is Omy12, wherein nine total associations were found between the binary trait and mapped loci across the GWAS models. Of those associations one locus was detected in four models, and another was detected in three, providing strong support for these associations. Three independent QTL analyses in divergent populations of rainbow and steelhead trout have identified several QTL on this linkage group, including QTL for the binary life history trait, and multiple additional QTL for smoltification related traits (Wringe *et al.* 2010; Le Bras *et al.* 2011; Hecht *et al.* in review). These results suggest that conserved genetic mechanisms associated with smoltification and migration related traits map to this linkage group. When comparing the binary trait GWAS results between the two populations, we were interested in determining whether the same regions are conserved among these relatively closely related trout populations, or whether loci associated with migration persisted within Upper Mann Creek given its 50 years of isolation. Our results indicate that, among the mapped loci, a small proportion are significant in both Yakima and Upper Mann Creek populations. Some overlap was detected on linkage groups Omy1, 12, 15, 20 and 27 providing suggestive evidence of conserved mechanisms; however the data seem to suggest that most associations (e.g. Omy2 and Omy11) occur in one population or the other, but only rarely in both. It should be noted that only ~20% of the significant loci have been assigned to a linkage map, so among the 80% of unassigned loci, it is possible that more loci overlap. However based on the single marker results, without information on the relationship between those markers, only a small portion of the loci are significant in both populations and these regions may be areas of major effect for migration (Fig. 3).

Significant loci that did not overlap between the two populations may indicate that some were false positives or have minor effect for this trait, but this result may also be due to the two different methods employed in each population to determine the binary life history. In the Yakima population, putative smolts were detected based on downstream migratory behavior, whereas in Upper Mann Creek putative smolts were detected based on overall body morphology. Both Hecht *et al.* (in review) and Nichols *et al.* (2008) also used skin coloration and morphology to assign individuals to a life history class, and the QTL detected in both their studies corroborate significant loci detected in the Yakima River and the Upper Mann Creek population analyses here. It is also possible that these two populations in the Columbia River basin persist under very different selection pressures, and thus different genetic mechanisms determine the overall life history in each population.

Overall, there are several loci associated with the propensity to migrate in the Upper Mann Creek population suggesting that despite the 50 years of sequestration, there is still genetic variability that contributes to this life history tactic. Other studies have also documented similar results in populations sequestered by anthropogenic barriers to migration (Thrower *et al.* 2008; Holecek *et al.* 2012;) or years of freshwater domestication (Pascual *et*

al. 2001). It should be recognized however that the propensity to migrate and ability to complete anadromous migration may involve different complex traits. So while this particular population seems to harbor variation for the propensity to migrate, further research is needed to determine the ultimate effects of sequestration on this population's ability to not only produce smolts, but to produce smolts that will successfully return. Additionally we have identified several associations with ATPase activity in this population, which suggests that variation in this trait also persists, despite a potentially strong selection pressure against it in a sequestered freshwater environment, where increased osmoregulatory ability may have little benefit to an individual and may even yield a cost (Aykanat *et al.* 2011).

Here we have demonstrated the utility of RAD-tag based genotyping methods by performing GWAS in wild populations of rainbow and steelhead trout, to identify hundreds of loci associated with the propensity to migrate. Many of those loci align perfectly with loci detected in linkage mapping studies, which allowed the relative assignment of several loci to genetic linkage groups. Additionally we have added to a growing knowledge base on the genetic architecture of migration in rainbow and steelhead trout and determined that there is strong evidence to suggest that complex genetic mechanisms between populations may contribute significantly to shaping this trait, with some parallel mechanisms between populations. This has tremendous implications in migratory fish conservation and management, wherein locally adapted populations should be treated as individual units of conservation. These findings could transcend across migratory species globally where species adapt to dramatic shifts in their environment to survive migrations, such that selection has delicately shaped life-histories.

References

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188**, 799-U779.

Ando D, Kitamura T, Mizuno S (2005) Quantitative analysis of body silvering during smoltification in masu salmon using chromameter. *North American Journal of Aquaculture* **67**, 160-166.

Aykanat T, Thrower FP, Heath DD (2011) Rapid evolution of osmoregulatory function by modification of gene transcription in steelhead trout. *Genetica* **139**, 233-242.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Balding DJ (2006) A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* **7**, 781-791.

Beeman JW, Rondorf DW, Tilson ME, Venditti DA (1995) A Nonlethal Measure of Smelt Status of Juvenile Steelhead Based on Body Morphology. *Transactions of the American Fisheries Society* **124**, 764-769.

Behnke RJ (2002) *Trout and Salmon of North America.* The Free Press, New York, NY.

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

Bradbury PJ, Zhang Z, Kroon DE*, et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.

Brunelli JP, Wertzler KJ, Sundin K, Thorgaard GH (2008) Y-specific sequences and polymorphisms in rainbow trout and Chinook salmon. *Genome* **51**, 739-748.

Dellefors C, Faremo U (1988) Early sexual-maturation in males of wild sea trout, Salmo trutta L., inhibits smoltification. *Journal of Fish Biology* **33**, 741-749.

Dingle H (2006) Animal migration: is there a common migratory syndrome? *Journal of Ornithology* **147**, 212-220.

Dingle H, Drake VA (2007) What is migration? *Bioscience* **57**, 113-121.

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* **3**, 43-52.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Folmar LC, Dickhoff WW (1980) The parr-smolt transformation (smoltification) and seawater adaptation in salmonids - A review of selected literature. *Aquaculture* **21**, 1-37.

Groot C (1972) Migration of yearling sockeye salmon (Oncorhynchus nerka) as determined by time-lapse photography of sonar observations. *Journal of the Fisheries Research Board of Canada* **29**, 1431-&.

Gustafson RG, Waples RS, Myers JM*, et al.* (2007) Pacific salmon extinctions: quantifying lost and remaining diversity. *Conservation Biology* **21**, 1009-1020.

Haner PV, Faler JC, Schrock RM, Rondorf DW, Maule AG (1995) Skin Reflectance as a Nonlethal Measure of Smoltification for Juvenile Salmonids. *North American Journal of Fisheries Management* **15**, 814-822.

Hecht BC, Thrower FP, Hale MC, Miller MR, Nichols KM (in review) The genetic architecture of migration related traits in rainbow and steelhead trout, *Oncorhynchus mykiss*.

Hoar WS (1958) The evolution of migratory behavior among juvenile salmon of the genus Oncorhynchus. *Journal of the Fisheries Research Board of Canada* **15**, 391-428.

Hoar WS (1976) Smolt transformation - evolution, behavior, and physiology. *Journal of the Fisheries Research Board of Canada* **33**, 1233-1252.
Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**, e1000862.
Holecek DE, Scarnecchia DL, Miller SE (2012) Smoltification in an impounded, adfluvial redband trout population upstream from an impassable dam: does it persist? *Transactions of the American Fisheries Society* **141**, 68-75.
Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.
Johnsson JI, Clarke WC, Blackburn J (1994) Hybridization with domesticated rainbow trout reduces seasonal-variation in seawater adaptability of steelhead trout (Oncorhynchus mykiss). *Aquaculture* **121**, 73-77.
Johnston SE, McEwan JC, Pickering NK*, et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* **20**, 2555-2566.
Jonsson N, Jonsson B, Hansen LP (1998) Long-term study of the ecology of wild Atlantic salmon smolts in a small Norwegian river. *Journal of Fish Biology* **52**, 638-650.
Kang HM, Zaitlen NA, Wade CM*, et al.* (2008) Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709-1723.
Kozfkay CC, Campbell MR, Meyer KA, Schill DJ (2011) Influences of habitat and hybridization on the genetic structure of redband trout in the Upper Snake River basin, Idaho. *Transactions of the American Fisheries Society* **140**, 282-295.
Kruuk LEB, Slate J, Wilson AJ (2008) New answers for old questions: the evolutionary quantitative genetics of wild animal populations. In: *Annual Review of Ecology Evolution and Systematics*, pp. 525-548. Annual Reviews, Palo Alto.
Le Bras Y, Dechamp N, Krieg F*, et al.* (2011) Detection of QTL with effects on osmoregulation capacities in the rainbow trout (Oncorhynchus mykiss). *BMC Genetics* **12**.
Liedvogel M, Akesson S, Bensch S (2011) The genetics of migration on the move. *Trends in Ecology & Evolution* **26**, 561-569.
Limborg MT, Blankenship SM, Young SF*, et al.* (2012) Signatures of natural selection among lineages and habitats in Oncorhynchus mykiss. *Ecology and Evolution* **2**, 1-18.
Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* **4**, 981-994.
Martinez A, Garza JC, Pearse DE (2011) A Microsatellite Genome Screen Identifies Chromosomal Regions under Differential Selection in Steelhead and Rainbow Trout. *Transactions of the American Fisheries Society* **140**, 829-842.
McLeese JM, Johnsson J, Huntley FM, Clarke WC, Weisbart M (1994) Seasonal changes in osmoregulation, cortison, and cortisol receptor activity in the gills of parr/smolt of steelhead trout and steelhead-rainbow trout hybrids, Onchrohynchus mykiss. *General and Comparative Endocrinology* **93**, 103-113.
Miller MR, Brunelli JP, Wheeler PA*, et al.* (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology* **21**, 237-249.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**, 240-248.
Narum SR (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics* **7**, 783-787.
Narum SR, Zendt JS, Frederiksen C*, et al.* (2011) Candidate Genetic Markers Associated with Anadromy in Oncorhynchus mykiss of the Klickitat River. *Transactions of the American Fisheries Society* **140**, 843-854.
Narum SR, Zendt JS, Graves D, Sharp WR (2008) Influence of landscape on resident and anadromous life history types of Oncorhynchus mykiss. *Canadian Journal of Fisheries and Aquatic Sciences* **65**, 1013-1023.
Negus MT (2003) Determination of smoltification status in juvenile migratory rainbow trout and Chinook salmon in Minnesota. *North American Journal of Fisheries Management* **23**, 913-927.
Nichols KM, Broman KW, Sundin K*, et al.* (2007) Quantitative trait loci x maternal cytoplasmic environment interaction for development rate in Oncorhynchus mykiss. *Genetics* **175**, 335-347.
Nichols KM, Edo AF, Wheeler PA, Thorgaard GH (2008) The genetic basis of smoltification-related traits in Oncorhynchus mykiss. *Genetics* **179**, 1559-1575.
Pascual M, Bentzen P, Rossi CR*, et al.* (2001) First documented case of anadromy in a population of introduced rainbow trout in Patagonia, Argentina. *Transactions of the American Fisheries Society* **130**, 53-67.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
Quinn TP (2005) *The Behavior and Ecology of Pacific Salmon and Trout* American Fisheries Society, Bethesda, MD.
Quinn TP, Myers KW (2004) Anadromy and the marine migrations of Pacific salmon and trout: Rounsefell revisited. *Reviews in Fish Biology and Fisheries* **14**, 421-442.
Sanchez CC, Smith TPL, Wiedmann RT*, et al.* (2009) Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics* **10**.
Schrock RM, Beeman JW, Rondorf DW, Haner PV (1994) A microassay for gill sodium, potassium activated ATPase in juvenile Pacific salmonids. *Transactions of the American Fisheries Society* **123**, 223-229.
Slate J, Gratten J, Beraldi D*, et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97-107.
Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.
Thrower FP, Hard JJ, Joyce JE (2004) Genetic architecture of growth and early life-history transitions in anadromous and derived freshwater populations of steelhead. *Journal of Fish Biology* **65**, 286-307.
Thrower FP, Joyce JE, Celewycz AG, Malecha PW (2008) The potential importance of reservoirs in the western United States for the recovery of endangered populations of anadromous steelhead. In: *Balancing Fisheries Management and Water Uses for Impounded River Systems* (eds. Allen MS, Sammons S, Maceina MJ), pp. 309-324. American Fisheries Society, Bethesda, MD.
Wedemeyer GA, Saunders RL, Clarke WC (1980) Environmental factors affecting smoltification and early marine survival of anadromous salmonids. *Marine Fisheries Review* **42**, 1-14.

Wilcove DS, Wikelski M (2008) Going, going, gone: Is animal migration disappearing? *Plos Biology* **6**, 1361-1364.
Wringe BF, Devlin RH, Ferguson MM*, et al.* (2010) Growth-related quantitative trait loci in domestic and wild rainbow trout (Oncorhynchus mykiss). *BMC Genetics* **11**, 63.
Young WP, Wheeler PA, Fields RD, Thorgaard GH (1996) DNA fingerprinting confirms isogenicity of androgenetically derived rainbow trout lines. *Journal of Heredity* **87**, 77-81.
Yu JM, Pressoir G, Briggs WH*, et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**, 203-208.
Zaugg WS, McLain LR (1972) Changes in gill adenosine-triphosphatase activity associated with parr-smolt transformation in steelhead trout, coho, and spring chinook salmon. *Journal of the Fisheries Research Board of Canada* **29**, 167-&.
Zhang ZW, Ersoz E, Lai CQ*, et al.* (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* **42**, 355-U118.
Zimmerman CE, Reeves GH (2000) Population structure of sympatric anadromous and nonanadromous Oncorhynchus mykiss: evidence from spawning surveys and otolith microchemistry. *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 2152-2162.

Table 1. Summary of samples collected in sampling population YAK (Yakima River basin, WA) and UMC (Upper Mann Creek, ID). Sex (F = female, M = male, UNK = unknown), SMOLT (life-history classification, R (resident), S (smolt)), ATPase activity measured (y/n)

Population	Sex	SMOLT	ATPase	Count
YAK	F	R	no	12
YAK	M	R	no	86
YAK	UNK	-	-	-
YAK	F	S	no	14
YAK	M	S	no	8
YAK	UNK	S	no	7
UMC	F	R	yes	16
UMC	M	R	yes	10
UMC	UNK	R	yes	1
UMC	F	S	yes	21
UMC	M	S	yes	4
UMC	UNK	S	yes	3

Table 2. Number of mapped loci for each linkage group detected as significant in each GWAS model. Linkage group assignment based on linkage maps of Hecht et al. (in review) and Miller et al. (2012).

Linkage group	SMOLT Global		SMOLT Yakima		SMOLT UMC		ATPase UMC		SMOLT Total	ATPase Total	Grand Total
	GLM	MLM	GLM	MLM	GLM	MLM	GLM	MLM			
OmySex		2							2	0	2
Omy1	1	1	1	2		1			6	0	6
Omy2	2	4			2	1			9	0	9
Omy3		4	1	1					6	0	6
Omy4		2				1			3	0	3
Omy5							1	1	0	2	2
Omy6					1	2			3	0	3
Omy7		1				1	1		2	1	3
Omy8		5	2				1	1	7	2	9
Omy9						1	1	1	1	2	3
Omy10	1	2			1				4	0	4
Omy11	1	8	3	2					14	0	14
Omy12	2	3	2	1		1			9	0	9
Omy13		2	1	1					4	0	4
Omy14						1			1	0	1
Omy15		2	1		1				4	0	4
Omy16	1		3	2			2		6	2	8
Omy17		1	1				1		2	1	3
Omy18	1		4	1					6	0	6
Omy19	1		1						2	0	2
Omy20	1	1	1		1				4	0	4
Omy21		1				1			2	0	2
Omy22		1			1	2	1		4	1	5
Omy23		2				2	1		4	1	5
Omy24						1			1	0	1
Omy25	1	6					1		7	1	8
Omy26									0	0	0
Omy27		1	1	1		2			5	0	5
Omy28						1			1	0	1
TOTAL	12	49	22	11	7	18	10	3	119	13	132

Figure 1. Map of Pacific Northwest of the United States of America, showing the states of Idaho and Washington. Highlighted are the sampling locations where rainbow and steelhead trout were collected. Locations include the Teanaway River complex (North Fork, Middle Fork and main stem Teanaway) in Washington and Mann Creek in Idaho.

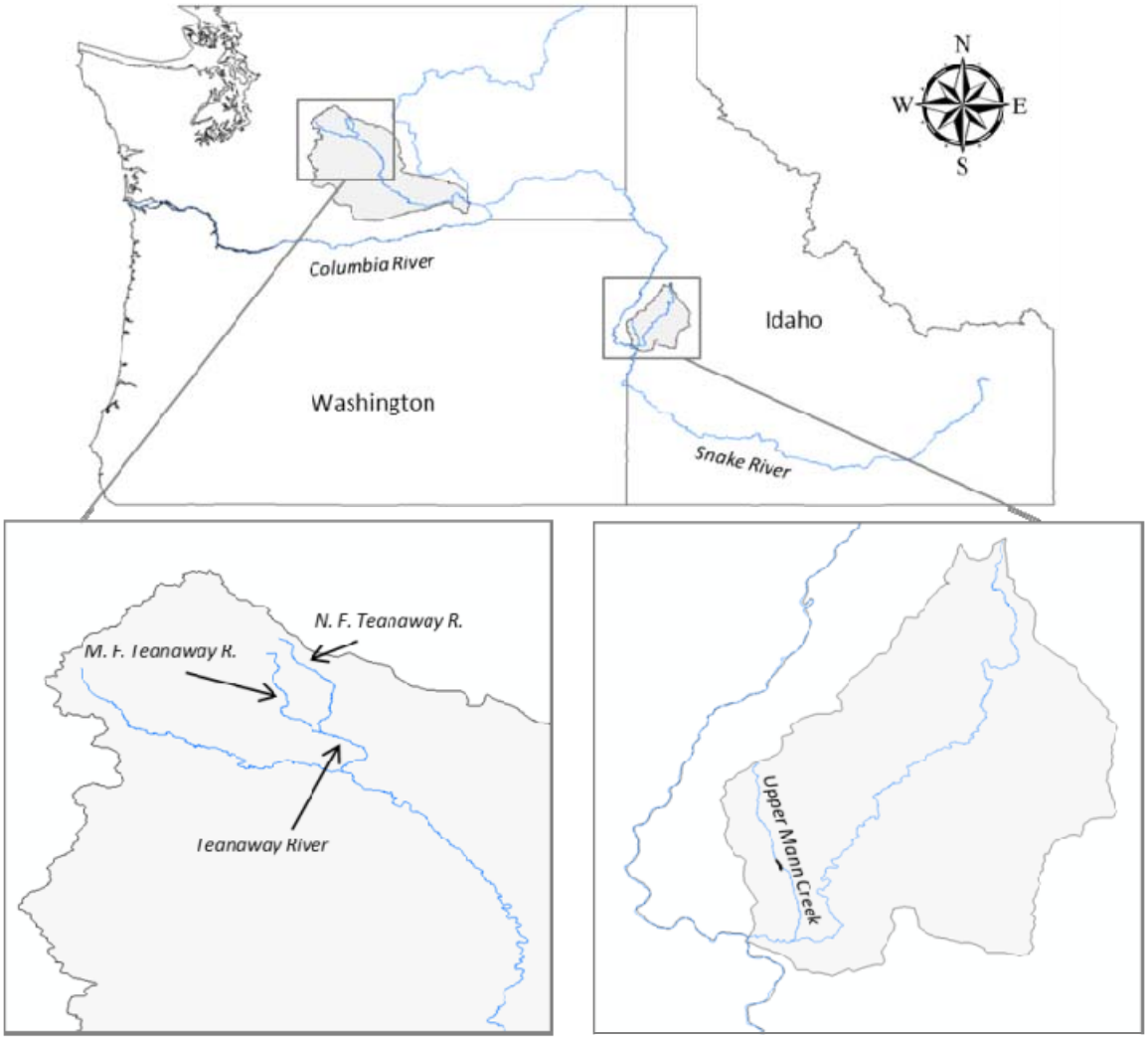
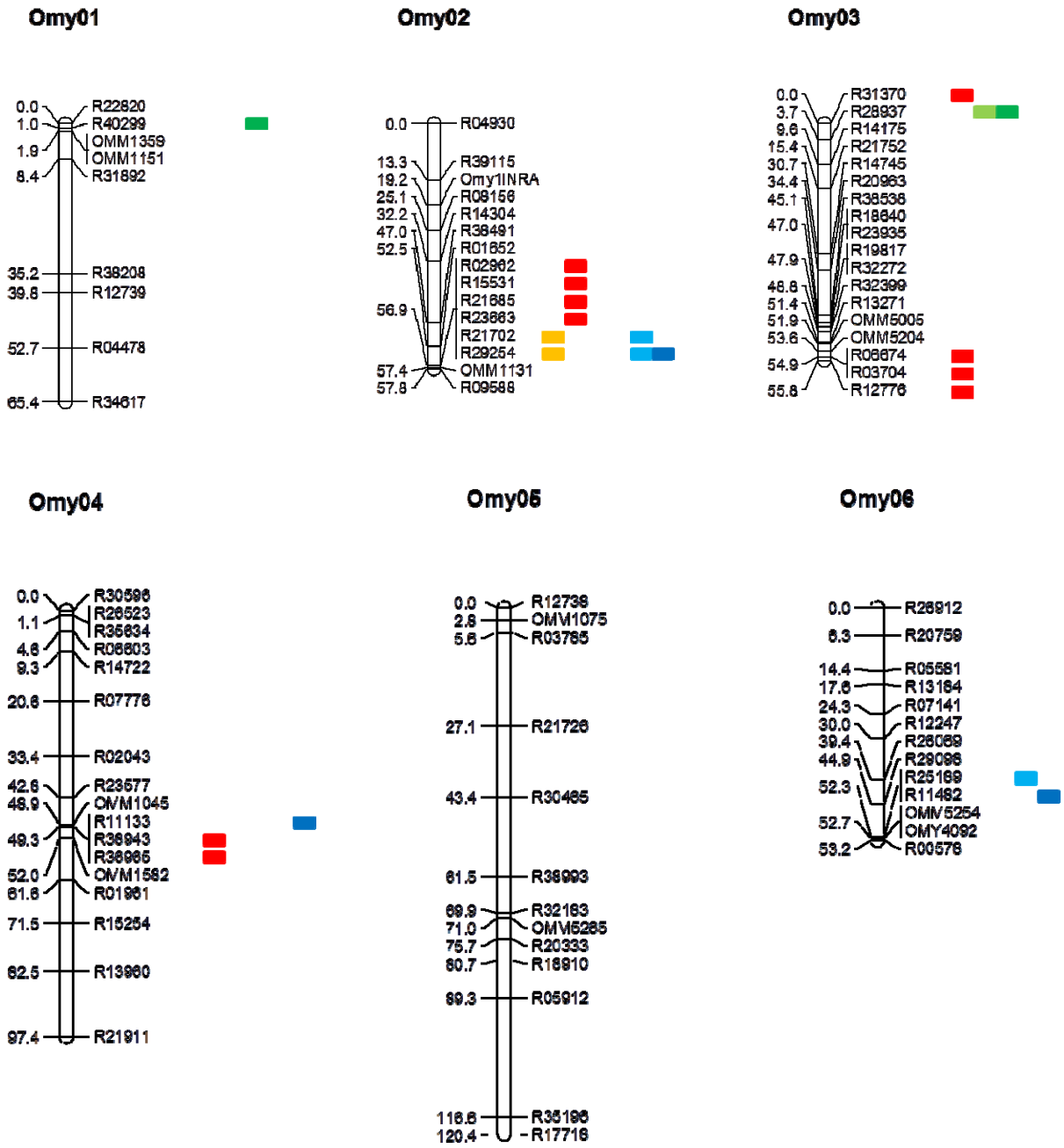
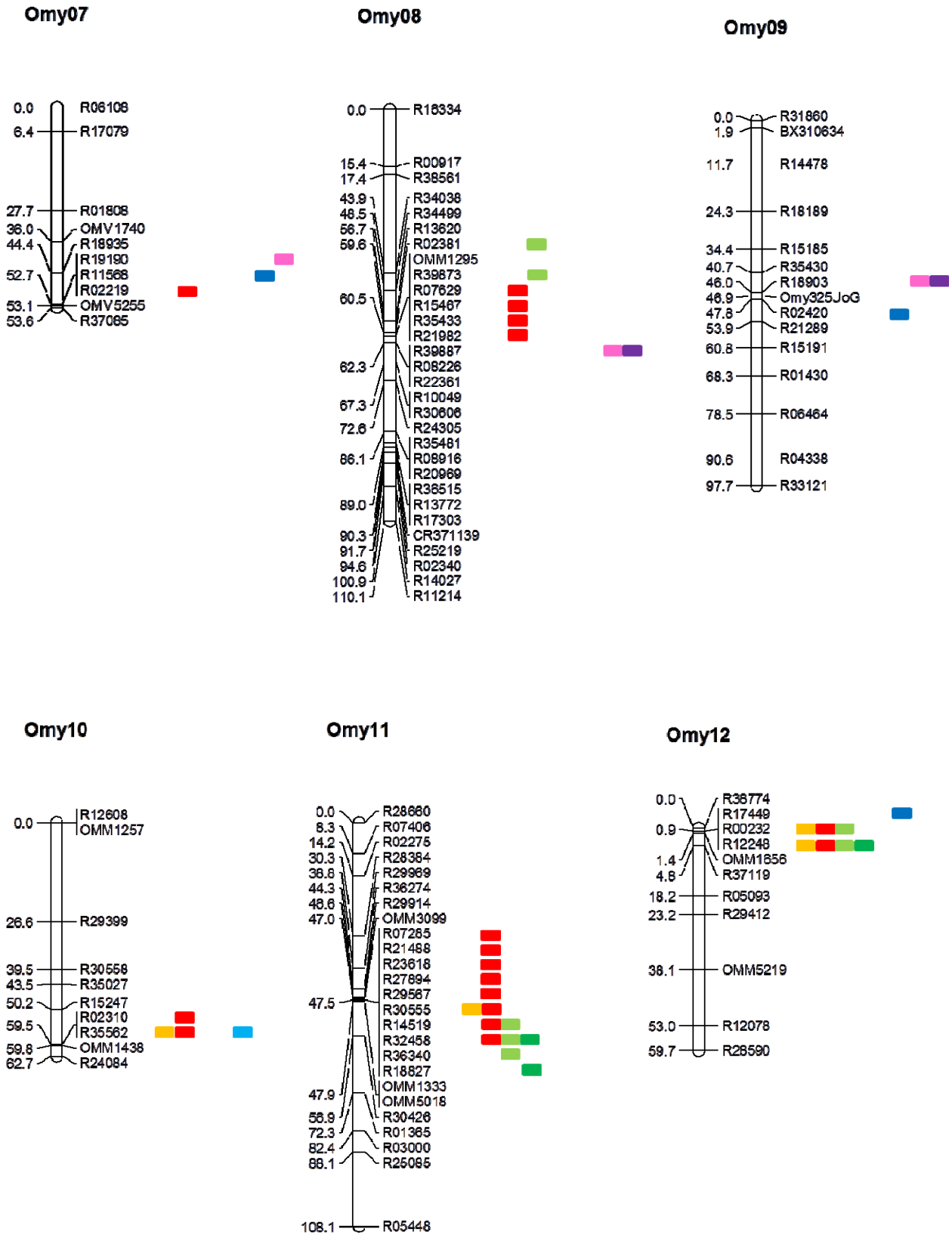
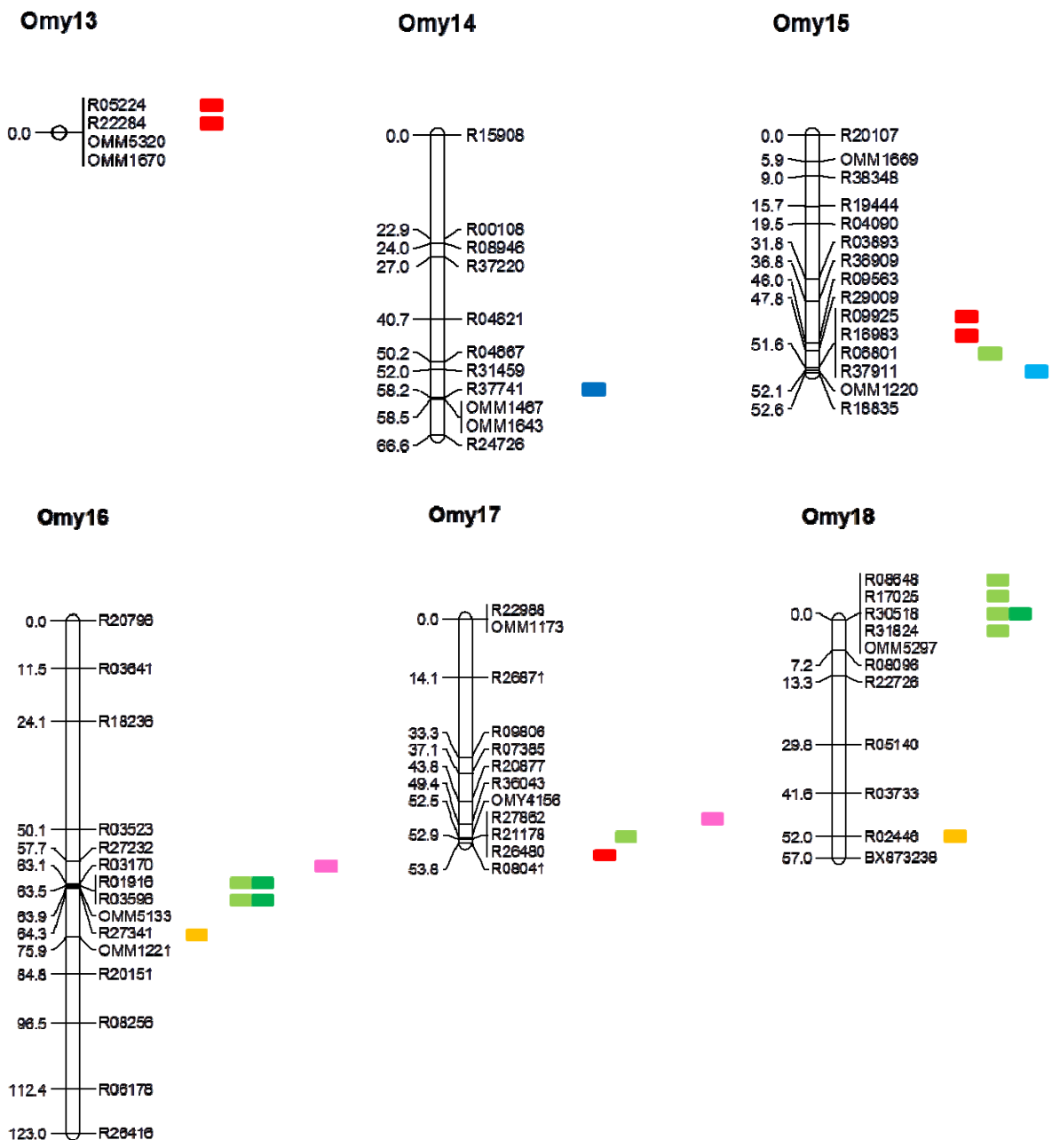
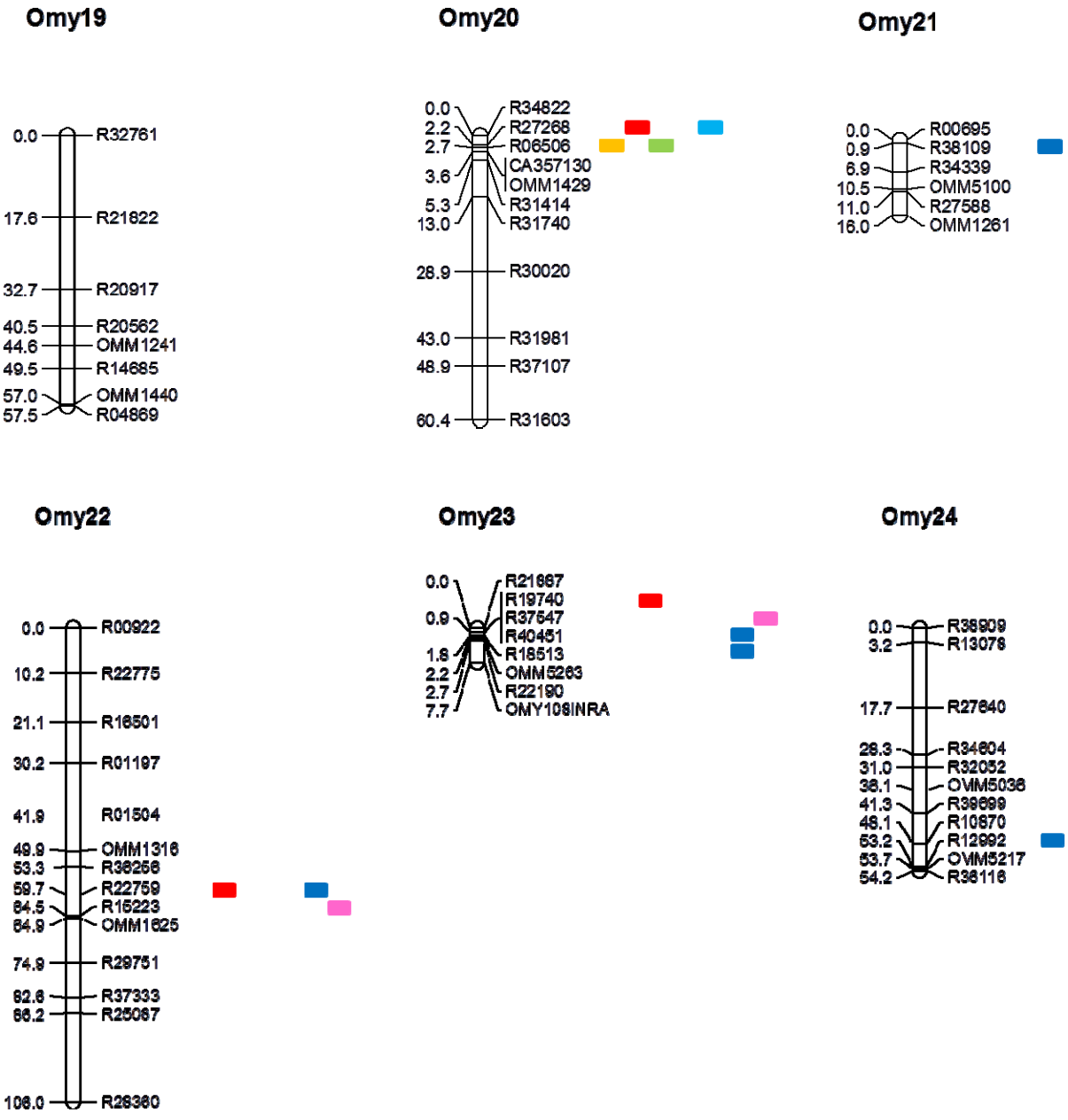


Figure 2. Rainbow trout RAD-tag genetic linkage map modified from Miller *et al.* (2012). Significant loci detected from the eight different association tests performed here, with each test represented by a different colored pill. Only those significant RAD-tag loci (n = 88) that aligned perfectly to Miller *et al.* (2012) and that were mapped are represented here.




Omy07
0.0 R06108
6.4 R17079
27.7 R01808
36.0 OMV1740
44.4 R18935
R19190
52.7 R11568
R02219
53.1 OMV5255
53.6 R37085
Omy08
0.0 R16334
15.4 R00917
17.4 R38561
43.9 R34038
48.5 R34499
56.7 R13620
59.6 R02381
OMM1295
R39873
60.5 R07629
R15467
R35433
R21982
62.3 R39887
R08226
R22361
67.3 R10049
R30606
72.6 R24305
R35481
86.1 R08916
R20969
R36515
89.0 R13772
R17303
90.3 CR371139
91.7 R25219
94.6 R02340
100.9 R14027
110.1 R11214
Omy09
0.0 R31860
1.9 BX310634
11.7 R14478
24.3 R18189
34.4 R15185
40.7 R35430
46.0 R18903
46.9 Omy325JoG
47.8 R02420
53.9 R21289
60.8 R15191
68.3 R01430
78.5 R06464
90.6 R04338
97.7 R33121
Omy10
0.0 R12608
OMM1257
26.6 R29399
39.5 R30558
43.5 R35027
50.2 R15247
R02310
59.5 R35562
59.8 OMM1438
62.7 R24084
Omy11
0.0 R28660
8.3 R07406
14.2 R02275
30.3 R28384
38.6 R29969
44.3 R36274
46.6 R29914
47.0 OMM3099
R07265
R21488
R23618
R27894
R29567
47.5 R30555
R14519
R32458
R36340
R18827
OMM1333
47.9 OMM5018
55.9 R30426
72.3 R01365
82.4 R03000
88.1 R25085
108.1 R05448
Omy12
0.0 R36774
R17449
0.9 R00232
R12248
1.4 OMM1656
4.8 R37119
18.2 R05093
23.2 R29412
38.1 OMM5219
53.0 R12078
59.7 R28590


Omy13
0.0
R05224
R22284
OMM5320
OMM1670
Omy14
0.0
R15908
22.9
R00108
24.0
R08946
27.0
R37220
40.7
R04821
50.2
R04667
52.0
R31459
58.2
R37741
58.5
OMM1467
OMM1643
66.6
R24726
Omy15
0.0
R20107
5.9
OMM1669
9.0
R38348
15.7
R19444
19.5
R04090
31.8
R03893
36.8
R36909
46.0
R09563
47.8
R29009
R09925
R16983
51.6
R06801
R37911
52.1
OMM1220
52.6
R16835
Omy16
0.0
R20796
11.5
R03641
24.1
R18236
50.1
R03523
57.7
R27232
63.1
R03170
R01916
63.5
R03596
63.9
OMM5133
64.3
R27341
75.9
OMM1221
84.6
R20151
96.5
R08256
112.4
R06178
123.0
R28416
Omy17
0.0
R22986
OMM1173
14.1
R26871
33.3
R09806
37.1
R07385
43.8
R20877
49.4
R36043
52.5
OMY4156
R27862
52.9
R21178
R26480
53.8
R08041
Omy18
R08648
R17025
0.0
R30518
R31824
OMM5297
7.2
R08098
13.3
R22726
29.8
R05140
41.6
R03733
52.0
R02446
57.0
BX873238


Omy19
0.0 R32761
17.6 R21622
32.7 R20917
40.5 R20562
44.6 OMM1241
49.5 R14685
57.0 OMM1440
57.5 R04869
Omy20
0.0 R34822
2.2 R27268
2.7 R06506
3.6 CA357130
OMM1429
5.3 R31414
13.0 R31740
28.9 R30020
43.0 R31981
48.9 R37107
60.4 R31603
Omy21
0.0 R00695
0.9 R38109
6.9 R34339
10.5 OMM5100
11.0 R27588
16.0 OMM1261
Omy22
0.0 R00922
10.2 R22775
21.1 R16501
30.2 R01197
41.9 R01504
49.9 OMM1316
53.3 R36256
59.7 R22759
64.5 R15223
64.9 OMM1625
74.9 R29751
82.6 R37333
86.2 R25067
106.0 R28360
Omy23
0.0 R21667
0.9 R19740
R37547
R40451
1.8 R18513
2.2 OMM5263
2.7 R22190
7.7 OMY108INRA
Omy24
0.0 R38909
3.2 R13078
17.7 R27640
28.3 R34604
31.0 R32052
36.1 OMM5036
41.3 R39699
48.1 R10870
53.2 R12992
53.7 OMM5217
54.2 R36116

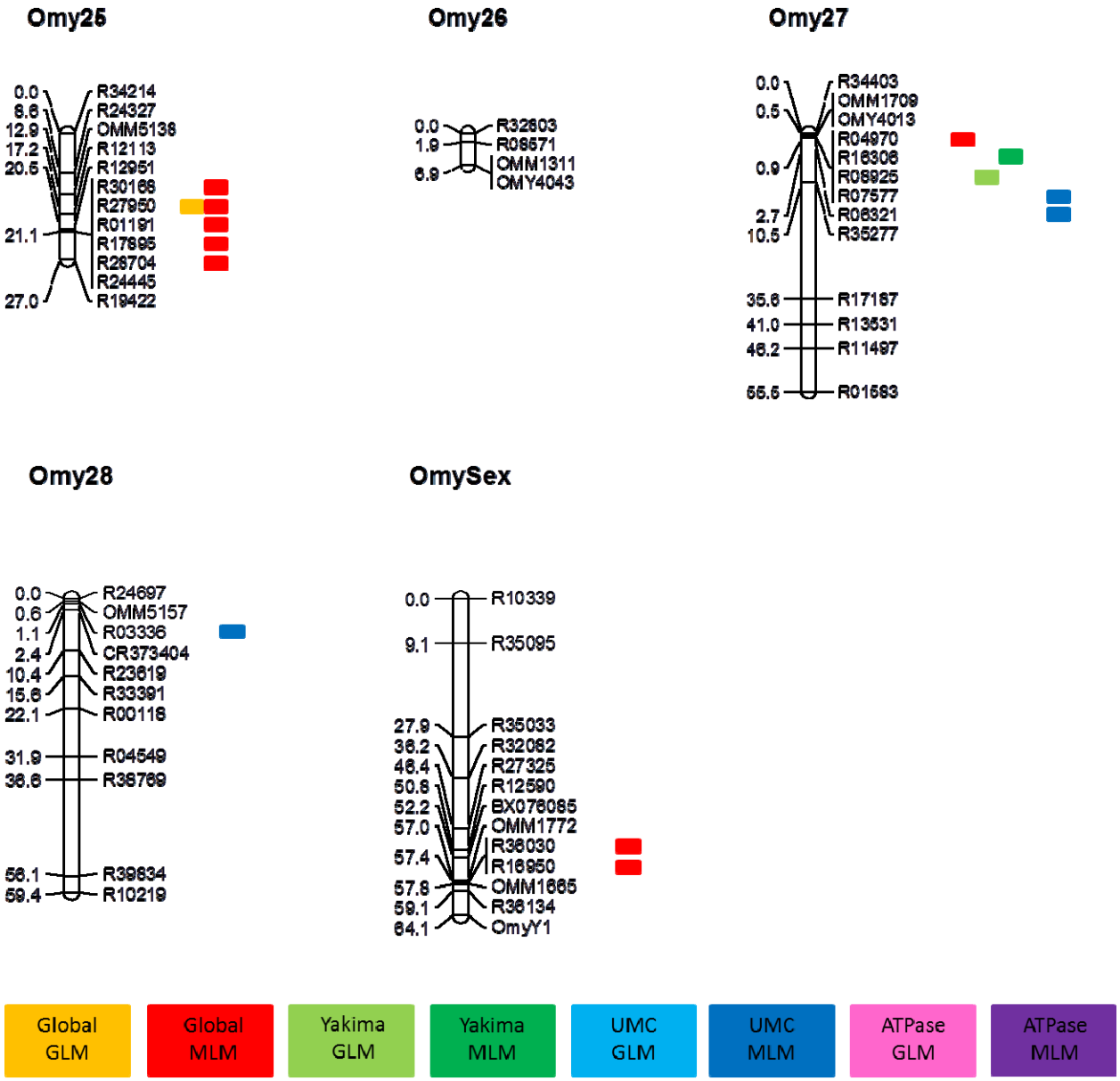
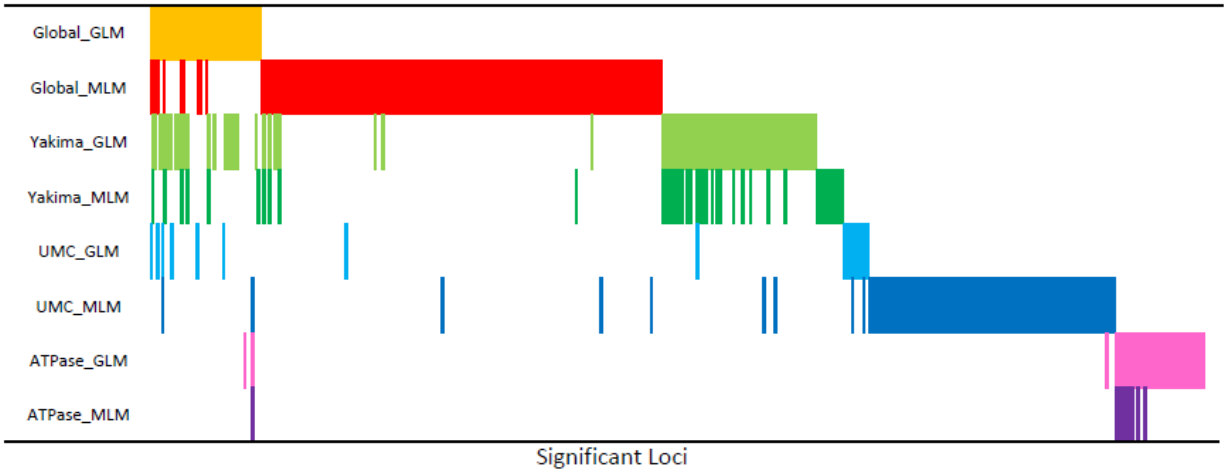

Omy25
0.0 R34214
8.6 R24327
12.9 OMM5138
17.2 R12113
20.5 R12951
R30166
R27950
21.1 R01191
R17895
R28704
R24445
27.0 R19422
Omy26
0.0 R32603
1.9 R08571
OMM1311
6.9 OMY4043
Omy27
0.0 R34403
0.5 OMM1709
OMY4013
R04970
R16306
0.9 R08925
R07577
2.7 R06321
10.5 R35277
35.6 R17167
41.0 R13531
46.2 R11497
55.5 R01583
Omy28
0.0 R24697
0.6 OMM5157
1.1 R03336
2.4 CR373404
10.4 R23619
15.6 R33391
22.1 R00118
31.9 R04549
36.6 R36769
56.1 R39834
59.4 R10219
OmySex
0.0 R10339
9.1 R35095
27.9 R35033
36.2 R32062
46.4 R27325
50.8 R12590
52.2 BX076085
57.0 OMM1772
R36030
57.4 R16950
57.8 OMM1665
59.1 R36134
64.1 OmyY1
Global GLM
Global MLM
Yakima GLM
Yakima MLM
UMC GLM
UMC MLM
ATPase GLM
ATPase MLM

Figure 3. Overlap of significant loci detected in eight different GWAS models. Each row represents a different model, and each column represents one of 550 significant loci detected in at least one model. Data is sorted by marker p-value of model Global_GLM first, then by Global_MLM, then by Yakima_GLM, etc. Colored cells represent a significant association at the column marker for the row model.



SECTION 3: Ongoing/Future Studies

As an ongoing Accords project, preparations are underway for further evaluation of multiple traits such as disease resistance in *O. mykiss*, adult run-timing in Chinook salmon, age-at-maturity (jacking) in Chinook salmon, and thermal adaptation. More details are provided below regarding plans to investigate each of these traits in the upcoming performance period (July 1, 2012 – June 30, 2013).

For disease resistance, we coordinated with existing collaborators at USDA and Clear Springs Research to investigate markers associated with infectious haematopoietic Necrosis virus (IHNV) and bacterial water disease (CWD) as previous studies have not yet identified strong genetic correlations (Overturf et al. 2010). Controlled experiments were designed and completed at the Clear Springs facility to challenge *O. mykiss* from known family groups to both IHNV and CWD. Samples were collected from male and female parents and their offspring, and survival of offspring from each family were identified following direct exposure to the two diseases. In the 2011-2012 performance period, individuals from family groups displaying high and low levels of survival to the diseases were sequenced with RAD-seq protocols (Miller et al. 2007) and this extensive data set will be analyzed in the upcoming year to identify SNPs that are associated with resistance to IHNV and CWD.

To investigate adult run timing in Chinook salmon, samples were provided from the Klickitat R. from fish that display a range of run timing. Fish from this drainage are particularly well suited for this study since this population of spring run Chinook salmon has been introgressed with fall run Chinook salmon through hatchery practices (Hess et al. 2011). These samples will be RAD sequenced to genotype thousands of SNP markers to identify markers that are associated with run timing in this species. To facilitate this analysis, samples from mapping families were collected from Klickitat Hatchery in order to create a linkage map. RAD-seq data has been generated for the linkage map and analyses will proceed in the upcoming performance period.

To investigate age-at-maturity (jacking) in Chinook salmon, we are planning to utilize information from multiple generation pedigrees to determine the heritability of this trait in hatchery and natural populations.

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.

References

Hess, J.E., and S.R. **Narum**. 2011. SNP loci correlated with run-timing in adult Chinook salmon from the Columbia River Basin. Transactions of the American Fisheries Society 140:855-864.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**, 240-248.

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