2009-2010 Annual Report

INFLUENCE OF ENVIRONMENT AND LANDSCAPE ON SALMONID GENETICS

Prepared by:

Shawn Narum
Andrew Matala
Nate Campbell

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: 7/1/09-6/30/10
Submitted: July 15, 2010

ABSTRACT

Environmental and landscape features can greatly contribute to population structure, life history diversification, and adaptation of salmonids. This 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 (i.e., smoltification and thermal tolerance); and 2) Controlled Experiments – laboratory/hatchery experiments with controlled environmental variables to validate phenotypic response of fish with given genotypes. In Objective 1, we identify candidate SNP markers under selection with the intent of validating these putative candidate markers under controlled circumstances in Objective 2. In the first year of this study, work was only completed to address Objective 1 while tasks under Objective 2 will be carried out in future years of the study, pending approval from BPA. In the first year of this project, goals were achieved for Objective 1 with candidate markers identified for thermal adaptation and anadromy in natural populations of *O. mykiss*.

For thermal adaptation, we used a limited genome scan approach to test for candidate markers under selection in populations occurring in desert and montane streams. An environmental approach to identifying outlier loci, Spatial Analysis Method (SAM), and linear regression of minor allele frequency with environmental variables revealed six candidate markers ($p < 0.01$). Putatively neutral markers identified high genetic differentiation among desert populations relative to montane sites, likely due to intermittent flows in desert streams. Additionally, populations exhibited a highly significant pattern of isolation-by-temperature ($p < 0.0001$) and those adapted to the same environment had similar allele frequencies across candidate markers, indicating selection for differing climates. These results imply that many genes are involved in adaptation of redband trout to differing environments

For smoltification/anadromy, we used a limited genome scan approach to identify candidate genetic markers associated with anadromy in 10 populations of *O. mykiss* from the Klickitat River, Washington USA. From an initial panel of 96 single nucleotide polymorphism (SNP) markers, we identified 3 SNPs that were significantly associated with anadromy ($p < 0.01$) after accounting for underlying population structure and selective environmental conditions. Univariate logistic regression of allele frequencies and residency/anadromy were also significant ($p < 0.05$) and thus three SNPs were considered candidate markers associated with anadromy (Omy_IL6-320, Omy_LDHB-2_i6, and Omy_ndk-152). A multivariate logistic model was developed from allele frequencies of these three markers to predict the potential of anadromy in natural populations. This model was applied to eight additional populations of *O. mykiss* to evaluate the utility of this model. Results of this study indicate that these markers are strong candidates associated with anadromy in *O. mykiss* of the Klickitat River, but further testing is needed to evaluate this association across a broader distribution of this species' range. Common garden experiments may also help clarify the association of genotype of these candidate markers with smoltification phenotypes of individual fish.

Table of Contents

List of Figures

List of Tables

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 (IDFG), Joe Zendt (YKFP), and field assistants. We are grateful for extensive contributions in the genetics laboratory from Lori Maxwell. David Graves provided maps for figures.

Introduction

Environmental and landscape features can greatly contribute to the population structure, life history diversification, and local adaptation of organisms in aquatic habitats (reviewed in Storfer et al. 2006). Geographic barriers to dispersal include recent events that may have been human induced (e.g., dams) as well as ancient events such as glaciations and formation of mountain chains (e.g., Castric et al. 2001). However, other environmental characteristics such as elevation, temperature, forest cover, and precipitation may influence distribution, adaptation, and gene flow of species (Dionne et al. 2008; Narum et al. 2008). For example, the geographic distributions of species ranges' are often determined by thermal tolerance (Brannon et al. 2004) and may necessitate adaptations for survival in extreme environments (Rodnick et al. 2004).

In this study, we plan to screen a suite of approximately 100 SNP markers (e.g., Campbell et al. 2009) in natural populations of steelhead and Chinook salmon for which we have previous information regarding traits of interest (i.e., thermal tolerance and smoltification). Markers within the suite will be split into categories of putative candidate or neutral loci for analyses of local adaptation and gene flow, respectively.

Screening with many genetic markers provides the opportunity to investigate local adaptation in natural populations and identify candidate genes under selection (Beaumont and Nichols 1996; Beaumont and Balding 2004; Excoffier et al. 2009). This has become a commonly employed approach in ecological and population genetics studies to detect outlier loci that are putatively under selection (e.g., Vasemagi and Primmer 2005; Nosil et al. 2008). Additionally, correlation methods can be highly informative to identify markers in coding and cis-regulatory regions of known functional genes that are associated with specific selective pressures or phenotypes (Lyman and Mackay 1998; Chase et al. 2009; Torgerson et al. 2009). With increasing genomic information available for non-model organisms, single nucleotide polymorphisms (SNPs) have begun to see increased use as genetic markers for population genetic studies (e.g., Morin et al. 2004). These sequence polymorphisms are densely scattered throughout the genome of most organisms, and are commonly observed in both coding and non-coding regions of functional genes making them ideal markers to study adaptive molecular variation (e.g., Akey et al. 2002). In a large suite of unlinked SNPs that are distributed across the genome (e.g., Campbell et al. 2009), it is possible to utilize both functionally neutral and adaptive markers within a single study. This combination of information provides a powerful approach to study questions in ecological genetics since both demographic processes (i.e., gene flow and genetic drift) and local adaptation (i.e., selection) may be inferred.

While candidate markers under selection can be used to address local adaptation in natural populations, the inclusion of neutral markers also provides the opportunity to evaluate gene flow among populations in relationship to geological or environmental barriers. The study will also provide resources for evaluating the maintenance of biologically relevant genetic diversity in hatcheries. A variety of statistical models have been developed to address specific questions related to genetic structure due to environment and landscape features (reviewed in Manel et al. 2003; Storfer et al. 2006). For example, ordination models with canonical correspondence analysis have been used as an alternative to Mantel tests to simultaneously evaluate drainage, altitude, and human

impacts to genetic diversity of salmonid fishes (Angers et al. 1999; Costello et al. 2003). Since many environmental features are inter-correlated, multivariate modeling of parameters related to genetic structure can be employed with tools such as GESTE (Foll and Gaggiotti 2006). Recent applications of interpolation models that utilize multivariate analyses such as principal components analysis (PCA) have also demonstrated that habitat and landscape features can identify and predict spatial patterns associated with restricted gene flow (Piertney et al. 1998). When PCA results are interpolated and overlaid with GIS data, synthesis maps can identify genetic patterns related to landscape (e.g., Narum et al. 2008). In this study, we plan to apply these approaches to better understand environmental genetics of steelhead and Chinook salmon in the Columbia River.

Report Structure

This report is divided into two sections, one for each of the traits of interest in the first year of study. The first section reports on candidate markers related to thermal adaptation and the second section on candidate markers associated with anadromy.

Section 1: Thermal adaptation

Introduction

In nature, widely distributed species often occupy environments outside of their optimal preference, causing selective pressure for physiological and genetic adaptation. This is especially true for species that are heavily impacted by human disturbances that may expose them to more extreme environments than normally observed, and adaptation to differing habitat may be further reinforced by limited gene flow due to habitat fragmentation, dispersal barriers, or relocation to new areas. Under these circumstances, selective pressure has typically been shown to be divergent (e.g., Hendry *et al.* 2008) but balancing selection may also play a role to maintain genetic variation (e.g., MHC genes and immune response; Piertney and Oliver 2006). While selection and subsequent genetic adaptation has been well demonstrated under laboratory conditions for model organisms such as *Drosophila spp*. (e.g., Hoffman *et al.* 2003) and zebrafish (e.g., Neuhauss *et al.* 1999), adaptation in natural populations of non-model species is difficult to study. This is due to several confounding factors such as the inability to isolate variables in complex environments, uncontrolled mating (predetermined parental crosses are not possible), generation length, and limited genome information (i.e., annotated sequence, quantitative trait loci, and linkage maps).

Genome scans with many genetic markers provide the opportunity to investigate local adaptation in natural populations and identify candidate genes under selection. This has become a commonly employed approach in ecological and population genetics studies to detect outlier loci that are putatively under selection (e.g., Vasemagi & Primmer 2005; Nosil *et al.* 2008). A variety of tests have been developed to identify "general" outlier loci from a neutral distribution (Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al.* 2009), and additional tests are also available to identify loci associated with specific environmental variables (e.g., Joost et al. 2008). Additionally, correlation methods can be highly informative to identify markers in coding and cis-regulatory regions of known functional genes that are associated with specific selective pressures or phenotypes (Lyman & Mackay 1998; Umina *et al.* 2005; Chase *et al.* 2009; Torgerson *et al.* 2009). With increasing genomic information available for non-model organisms, single nucleotide polymorphisms (SNPs) have begun to see increased use as genetic markers for population genetic studies (e.g., Luikart *et al.* 2003; Morin *et al.* 2004). These sequence polymorphisms are dense throughout the genome of most organisms, and are commonly observed in both coding and non-coding regions of functional genes making them ideal markers to study adaptive molecular variation (e.g., Akey *et al.* 2002). In a large suite of unlinked SNPs that are distributed across the genome, it is possible to utilize both functionally neutral and adaptive markers within a single study. This combination of information provides a powerful approach to study questions in ecological genetics since both demographic processes (i.e., gene flow and genetic drift) and local adaptation (i.e., selection) may be inferred. In this study, we employed a combination of these approaches to test for neutral and adaptive genetic differences in redband trout (*Oncorhynchus mykiss gairdneri*) inhabiting different thermal environments.

Climate has been shown to influence a variety of traits in fish including thermal tolerance, growth, development, and disease resistance (Crozier et al. 2008). Cellular response to thermal stress and adaption to extreme temperatures has been shown to be polygenic in fish, with genes involved from many different biological pathways. Under heat stress, a wide variety of genes are differentially expressed including those related to immune response, signal transduction, protein processing, response to stress, and metabolism (Kassahn *et al.* 2007). Adaptation to cold stress has also been shown to be necessary for organisms occupying very cold temperatures (Ciardiello *et al.* 2000). Water temperature can affect growth and development rates in fish, with warmer water resulting in accelerated rates unless temperature exceeds optimal levels and causes stress (e.g., Beer and Anderson 2001). Studies have shown that fry emergence timing is associated with water temperature and often matches local conditions through either spawning date or embryo development (e.g., Brannon et al. 1987; Beacham and Murray 1990). Warm or cold water temperatures may also cause higher pathogen population growth rates, and increased likelihood of disease in fish (Holt 1989; Marcogliese 2001). Resistance to diseases can increase in fish populations following exposure to various pathogens (Zinn et al. 1977), thus resulting in adaptation to local pathogens. Therefore, extensive selective pressure can occur for several traits as a result of differing climates throughout a species' range.

Redband trout occupy a wide range of habitats including desert and montane streams, with significant differences in habitat characteristics such as elevation, gradient, substrate, shading, and temperature (Meyer et al. 2010). Physiological differences have also been observed in redband trout occupying desert streams (Gamperl et al. 2002). These studies indicate the potential for local adaptation of redband trout to differing habitats across environmental gradients. In this study, individual redband trout were sampled across several tributaries representing desert and montane streams and screened with 96 SNP markers from functional genes of diverse biological pathways. We tested predictions that variable climates would result in divergent selection and local adaptation of redband trout to each environment.

Methods

Collection Sites

A total of 499 individual redband trout were sampled across 12 tributaries from southwest Idaho, USA representing populations that were pre-classified as desert or montane streams (six sites each) as determined by geographic location (Figure 1). Individual fish were sampled by electro-fishing in multiple reaches within a site to avoid sampling related individuals. Fish were a mix of age classes as determined by length frequency and sampled during the summer seasons of 2002-2005. A non-lethal fin clip was collected from each fish as sample tissue and immediately preserved in 100% non-denatured ethanol.

A total of nine environmental variables were recorded in order to characterize habitat of each collection site (Table 1). Latitude and longitude coordinates were recorded for each collection area with a field GPS instrument and used to estimate elevation, temperature, and precipitation (Table 1). Elevation was determined from a U.S. Geological Survey 10-meter digital elevation model (DEM). Annual average

maximum air temperatures (hereafter referred to as "air temperature") and annual total precipitation were simulated at 800-meter cell resolution from a model based on climate normals from a 30 year period (1971-2000) in PRISM (Parameter-elevation Regressions on Independent Slopes Model; http://www.prism.oregonstate.edu/) of the Oregon Climate Service. Daily stream temperatures were collected from each site from May to October of 2009 in order to validate the use of daily and long-term air temperatures for analysis. Stream temperature data was unavailable from Fawn Cr. as thermal logger devices at that site were tampered with and destroyed. Additional habitat data was collected for each site as described in Meyer et al. (2010) for six variables (stream gradient, specific water conductivity, stream width, percent fine sediment, percent gravel, and percent cobble/boulder). Briefly, stream gradient was determined by elevational increments bounding each site by approximately 1 km, specific conductivity (µS/cm) was measured with a conductivity meter (accurate to ± 2%), stream width was estimated by averaging 10 transects from each site, and percent substrate composition was visually estimated within 1 m of transects at each site. All nine environmental variables were tested for differences between desert and montane collections with ANOVA (df = 11).

SNP Genotyping & Descriptive Statistics:

Tissue samples from each individual were processed with Qiagen DNeasy® kits to extract DNA from fin clips stored in 100% ethanol. Isolated DNA from each sample was genotyped for 96 SNP markers (see Supplemental Table S1) with Taqman chemistry (Applied Biosystems) and Fluidigm 96.96 dynamic array chips (reaction volumes of ~7nL) for SNP genotyping. Since genotyping in nL reaction volumes reduces the average starting copy number to a range where genotyping accuracy becomes less reliable (Campbell & Narum 2009a), a pre-amplification protocol was used to increase the number of starting copies. Pre-amplification occurred in 7µl reactions with 2µl of genomic DNA and 5µl of PCR cocktail (3.5µl of Qiagen Multiplex Mastermix, 0.875µl of 96 pooled primer sets at 0.36uM, and 0.625µl water) under the following thermal cycling program: initial denature at 95°C for 15 minutes, 14 cycles of 95°C for 15 seconds and 60°C for 4 minutes, hold at 4°C. Immediately after cycling, 133µL of nuclease free H_2O or TE buffer was added to each PCR reaction and stored at 4°C.

Pre-amplified template DNA was then genotyped with Fluidigm 96.96 dynamic array chips that included a three step process: 1) SNP assays (Taqman primers/probes) and DNA samples were mixed according to manufacturers protocols and loaded onto the chip with a Fluidigm IFC Controller instrument, 2) target SNPs were amplified for 50 cycles on a Eppendorf thermal cycler specially formatted for the Fluidigm 96.96 chip, and 3) chips were scanned with a Fluidigm EP-1 instrument to detect fluorescently labeled allele-specific probes. Genotypes for each assay were auto-scored with Fluidgim SNP Analysis v.2.1.1 software and verified by eye with scoring guides provided by an assay database and a heterozygous indicator sample for each SNP. Since the SNP markers used in this study were ascertained from a broad panel of samples including related populations from the Columbia River (e.g., Campbell et al. 2009), ascertainment bias should be limited. Any potential ascertainment bias should effect populations in this study equally since none were directly included in SNP discovery.

Tests for linkage disequilibrium (LD) between all pairs of loci were also performed using the MCMC approximation of the exact test in GENEPOP v. 3.3 (Raymond &

Rousset 1995). Because multiple comparisons were involved, correction against Type I error was made in both tests with the B-Y FDR method (False Discovery Rate; Benjamini and Yekutieli 2001) that provides increased power relative to the Bonferroni method (Narum 2006).

A total of 96 SNPs were screened in this study, but 20 markers were dropped from further statistical analyses (see Supplemental Table S1 for list). Five SNPs were dropped from analyses due to poor amplification and low quality genotype plots. Three SNPs in the panel were used to detect potential hybrids of *O. mykiss* and *O. clarki* (cutthroat trout), but dropped from further analysis since hybrids were not identified in this study. Eight pairs of SNPs were known to be physically linked on the same gene and only the most informative of each pair was included in analyses. An additional four pairs of SNPs had highly significant LD ($p < 0.0001$), and the least informative of each pair was dropped from further analyses. A total of 76 remaining SNPs were included in subsequent statistical tests (Supplemental Table S1).

Deviation from Hardy-Weinberg equilibrium (HWE) was evaluated at each locus and population with the Markov Chain Monte Carlo (MCMC) approximation of Fisher's exact test implemented in GENEPOP v. 3.3 (1000 batches with 1000 iterations; Raymond & Rousset 1995). Minor allele frequency (MAF), unbiased heterozygosity (H_E), and global F_{ST} were estimated for each SNP in each collection with GENEPOP. Effective population size (N_e) was estimated for each collection with the program LDNE (Waples and Do 2008). To determine confidence interval (CI) for N_e, we used the option of "jackknife CIs" which corrects for bias in parametric CIs when LDNE estimates are based on many loci (Waples 2006). We present results from setting the minimum frequency of alleles included in the analysis (P_{crit}) to a value of 0.02, which is a middle-range default value of the program. The P_{crit} value chosen was a compromise between a higher value which will lower the upward bias of N_e and a lower value which increases the precision of the estimate (Waples & Do 2008). Differences in H_E and N_e were tested with ANOVA (df = 11).

We investigated patterns of deviation from neutral expectations among the 76 SNPs with an outlier approach based on simulation methods initially proposed by Beaumont & Nichols (1996). This approach was implemented in LOSITAN (Antao *et al.* 2008) that simulates a distribution of F_{ST} values under neutral expectations to identify candidates for positive and balancing selection from a plot of average locus heterozygosity versus F_{ST} (Beaumont & Nichols 1996; Beaumont & Balding 2004). Simulations were run to independently generate a distribution of F_{ST}, based on 50,000 replicates, for 76 SNPs under an infinite alleles mutation model. The simulation results were then plotted to represent the median, and the 95% and 99% quantiles. Loci lying outside these quantiles were outliers putatively under directional or balancing selection, respectively. Simulations were done iteratively to avoid an upward bias in quantile ranges (potentially resulting in Type I error for balancing selection) by removing outlier loci above the 95% and 99% quantiles in the initial runs as implemented in LOSITAN. Alternative outlier tests (e.g., Beaumont & Balding 2004; Foll & Gaggiotti 2008, Excoffier *et al.* 2009) were not included in this study since these methods are unable to identify associations with specific variables and the general candidate markers identified by these methods are often discordant (e.g., Akey et al. 2009; Hermisson et al. 2009).

Tests for Association of SNP Markers and Environment

Since tests for outlier loci can often provide false positives or false negatives (e.g., Akey *et al.* 2009), a three step process was used to further identify candidate markers in this study. These steps included a spatial analysis method (SAM) to identify initial candidate markers associated with specific environmental variables, followed by univariate linear regression of SNP allele frequencies and environmental variables to identify secondary candidate markers, and finally a verification step to correct for underlying population structure. Details regarding each of these steps are provided below.

In order to identify SNP markers that were associated with environmental variables, we used the SAM program developed by Joost et al. (2008). This approach utilizes general linear models and logistic regression to identify significant associations of habitat characteristics with presence/absence of alleles at genetic markers across all individuals in the study. In the current data set, one allele for each SNP was recoded as present ("1") or absent ("0") for all individual fish. Missing habitat and genotype data were recoded with the non-numerical designation of "NaN". Due to the potential of this test to identify false positive associations, a conservative alpha of 0.01 with Bonferroni correction (final critical value of 0.00146) were used to reduce false positives in the Wald test (as recommended by Joost et al. 2008). Markers with *p*-values below the critical value were considered initial candidates for association with respective environmental variables.

To further reduce false positives and provide a more conservative test for association of markers at the population rather than individual level (SAM), allele frequencies of markers that were considered initial candidates from SAM were evaluated for correlation with each environmental variable for the 12 collection sites. We expect that this population level approach is more conservative because the regression analyses with MAF and temperature only include 12 data points (one per population) as opposed to 499 data points (one per individual) and thus the chance of detecting a significant result is lower. Univariate linear regression analysis with the least squares method was used to determine the relationship between MAF of each SNP and the environmental variable. A critical level of 0.01 was used to identify significant associations. Initial candidate markers from SAM that were also significant in linear regression tests were considered to be secondary candidate markers. All other markers were considered as putatively neutral for further statistical analyses.

It is well known that underlying population structure can bias association tests and result in false positive loci that are not actually associated with variables of interest (Lander & Schork 1994; Pritchard & Rosenberg 1999; Excoffier et al. 2009). In order to limit association bias due to underlying population structure in secondary candidate markers in this study, analyses with STRUCTURE v.2.3.2 (Pritchard *et al.* 2000a; Hubisz *et al.* 2009) and STRAT v.1.1 (Pritchard *et al.* 2000b) were implemented as suggested by Pritchard and Rosenberg (1999). In a recent review by Zhang *et al.* (2008), this STRAT approach has been shown to account for population structure in association studies equally well as other leading methods such as principal components analysis (i.e., Price *et al.* 2006). We followed the procedure for running STRAT with the following steps: 1) candidate loci were identified with regression analyses, 2) candidate loci and any other significantly linked markers (LD with $p < 0.0059$) were removed from the data set, 3)

remaining loci (putatively neutral) were used to run STRUCTURE and select the most likely number of distinct populations ("k") from 10 iterations for each potential k value ranging from 1-10 (50,000 burnin followed by 100,000 MCMC repetitions), and 4) ancestry coefficients (mean Q values from CLUMPP; Jakobsson & Rosenberg 2007) and environmental variables for each individual were included in STRAT to test candidate markers for significant association with environmental variables.

Population structure and adaptation

In order to infer the degree of relatedness between sample collections, pairwise chord distances (Cavalli-Sforza & Edwards 1967) were calculated between all population samples with the POPULATIONS software package (Langella 2001). These genetic distances were then used to construct neighbor-joining trees of sample populations for 1,000 bootstrap replicates. A consensus dendrogram was then constructed with the program TREEVIEW (Page 1996). Two dendrograms were constructed, one to represent genetic relationships due to demographic processes with putatively neutral markers, and another to evaluate populations potentially under thermal adaptation with candidate SNPs.

To test for differentiation among populations, pairwise values of the variance in allele frequencies among population samples (F_{ST}; Weir & Cockerham 1984) were estimated in GENEPOP with putatively neutral SNPs and separately with candidate temperature SNPs. Approximations of exact tests with MCMC were performed in GENEPOP (1,000 batches and 1,000 iterations) to determine significance of pairwise genic differentiation between all collections. Significance levels were adjusted for multiple tests with B-Y FDR (Benjamini & Yekutieli 2001; Narum 2006). Differences in pairwise F_{ST} matrices were evaluated to determine if specific populations had higher adaptive divergence than neutral divergence.

Patterns of isolation-by-distance among sites were evaluated with Mantel tests using fluvial stream distance and pairwise F_{ST} from SNPs determined to be either putatively neutral or candidate markers. The regression of the pairwise $F_{ST}/(1- F_{ST})$ on geographic distance was used to determine significance of Mantel tests in GENEPOP (Raymond & Rousset 1995). The procedure was modified to test a model of isolation-by-temperature by substituting pairwise differences in temperature versus panels of putatively neutral and temperature associated SNPs and for each matrix in the Mantel tests. Thus Mantel tests were able to capitalize on the cumulative effect of all candidate markers (multi-locus F_{ST}), rather than single loci, to estimate adaptive divergence and evaluate signals of local adaptation due to temperature differences among collections.

Results

Habitat differences and descriptive statistics

Tests of ANOVA revealed significant differences between desert and montane habitats for the following three environmental variables: width ($p = 0.011$), precipitation ($p = 0.00001$), and temperature ($p = 0.002$). The other six environmental variables were not significantly different between the two climates [gradient ($p = 0.170$), specific conductivity ($p = 0.067$), fines ($p = 0.463$), gravel ($p = 0.910$), cobble/boulder ($p = 0.727$), and elevation ($p = 0.879$)]. Daily stream and air temperatures for each site were

significantly correlated (all p-values < 1.0 x 10^{-8}), with a range of r^2 values from 0.24-0.72 across sites (Supplemental Figure S1).

Of 912 tests for deviation from HWE (76 SNPs in 12 populations), there were 10 significant results (BY-FDR critical value of 0.007) that included 9 heterozygote deficits and 1 heterozygote excess. Deviations were stochastically distributed across loci and populations with 2 deficits in Big Jacks Cr. (Omy_97865-196 and Omy_aromat-280), 1 deficit in Fawn Cr. (Omy_Ots249-227), 3 deficits and 1 excess in Duncan Cr. (Omy_aldB-165, Omy_aromat-280, Omy_Ots208-138, Omy_nkef-308, respectively), 1 deficit in Whiskey Cr. (Omy_121713-115), 2 deficits in Johnson Cr. (Omy_121713-115, Omy_LDHB-2_i6), and none in other populations or loci.

Descriptive statistics such as H_E and N_e were estimated with 65 putatively neutral SNPs for each collection and ranged from 0.141-0.180 and 18.3-406.3, respectively (Table 2). Sites with the lowest estimates of N_e were Duncan Cr. (N_e = 18.3) and Fawn Cr. (N_e = 26.1) while the highest estimates were from Bennett Cr. (N_e = 350.9) and Deer Cr. (N_e = 406.3). No significant differences were found among desert and montane collections for H_E (p = 0.53) or N_e (p = 0.38). However, pairwise exact tests of allele frequencies among collections were all significantly differentiated from one another with or without correction for multiple tests (all p-values were < 1.0 x 10^{-4}).

Simulation results from LOSITAN identified a total of eight loci that fell outside the 95% quantile for neutral expectations, with five candidates for divergent selection and three for balancing selection. The five candidates for divergent selection were Omy_97856-196, Omy_aspAT-413, Omy_arp-630, Omy_tlr5-205, and Omy_nkef-308 and the three for balancing selection were Omy_hsp47-86, Omy_hsc71-80, and Omy_NAKATPa3-50. At the more stringent 99% quantile, only two markers were considered candidates, one each for positive (Omy_aspAT-413) and balancing (Omy_hsc71-80) selection.

Tests for association of SNP markers and environment

Analysis with SAM provided a total of 24 markers that were significant for one or more environmental variable for a total of 61 significant associations (Supplemental Table S2). Each of the nine environmental variables were significant with at least one marker (Supplemental Table S2), but temperature had the most number of associated markers (17), followed by precipitation (13), specific conductivity (13), elevation (11), fines (3), gradient (1), width (1), gravel (1), and cobble/boulder (1). The 24 markers that were significantly associated with environmental variables were considered initial candidate markers.

Linear regression of population allele frequencies and environmental variables was used as a secondary criterion to screen the initial 24 candidate markers identified from SAM. Of the 61 initial associations, only seven had significant correlation between the specified environmental variable and MAF. Of these seven significant results, five were markers associated with temperature, one marker with precipitation, and one marker with specific conductivity. The five loci with allele frequencies significantly correlated with temperature (Figure 2) were Omy_aldB-165 (p = 0.0051), Omy_gdh-271 (p = 0.0005), Omy_Ogo4-212 (p = 0.0008), Omy_stat3-273 (p = 0.0089), and Omy_tlr5-205 (p = 0.0062). One of these five markers (Omy_Ogo4-212) was also significantly correlated with specific conductivity (p = 0.005). An additional marker was associated

with precipitation (Omy_hsf2-146; $p = 0.007$; Figure 2). In total, 7 of the 61 significant associations identified by SAM were also significantly correlated with linear regression of population allele frequencies. Therefore, the 6 SNPs that accounted for these 7 significant associations were considered secondary candidate markers for further verification.

In the third and final verification step for candidate markers, correction for underlying genetic structure was applied as described in the methods section (STRUCTURE/STRAT) with the following panels of markers. A total of 65 putatively neutral SNPs were used to determine population structure in STRUCTURE, removing the six candidate SNPs and five markers linked to them (Omy_113490-15, Omy_aromat-28, Omy_IL17-185, Omy_Ots249-22, Omy_rapd-167). Results from STRUCTURE provided support for k=5 as determined by criteria in Evanno *et al.* (2005). Mean ancestry coefficient (Q) and environmental variables (temperature, precipitation, and specific conductivity) for each individual were included as input for STRAT analysis to correct for population structure of the six candidate markers. All six candidate markers remained highly significant after correction for genetic structure (all p-values < 0.0001) and therefore were considered validated candidate SNPs associated with either temperature (Omy_aldB-165, Omy_gdh-271, Omy_Ogo4-212, Omy_stat3-273), temperature and specific conductivity (Omy_Ogo4-212), or precipitation (Omy_hsf2-146). Thus, panels of 65 putatively neutral markers and 5 temperature associated markers were used in further analyses. Single loci associated with either precipitation or specific conductivity remained verified candidates, but were not tested further.

Population structure and adaptation

Patterns of genetic structure and adaptation were observed when suites of 65 putatively neutral or 5 temperature associated markers were examined separately. With the panel of 65 putatively neutral SNP markers, all collections were genetically differentiated from one-another as pairwise estimates of F_{ST} were highly significant ($p <$ 0.0001) among all population combinations (Table 3, lower matrix). However, desert sites on average had higher pairwise F_{ST} values than montane sites. Results from STRUCTURE analysis with the 65 putatively neutral SNPs further revealed that collections from desert locations were more genetically distinct than those from montane sites. With the panel of 5 temperature candidate SNPs, the highest divergence was observed between desert and montane comparisons (all highly significant), and least among montane vs. montane collections (7 of 15 significant), while all 15 desert vs. desert comparisons remained significant (Table 3, upper matrix). The two matrices were highly correlated (Mantel $p = 0.001$), but 11 of 65 comparisons had at least two fold greater F_{ST} for the candidate than the neutral markers. These 11 comparisons primarily involved one desert population (Little Jacks Cr.) and one montane population (Fawn Cr.; Table 3) that represented the two extreme ends of temperature variation of the 12 collections (Table 1) at 16.8°C and 12.2°C, respectively. Little Jacks Cr. had ≥ two fold higher candidate F_{ST} in 4 montane and 1 desert populations, and Fawn Cr. had ≥ two fold higher candidate F_{ST} in 4 desert and 2 montane populations (Table 3) as would be expected if adaptive divergence was greater than neutral divergence between populations occupying differing climates.

Neighbor-joining dendrograms with each suite of markers provided differing perspectives regarding genetic similarity of populations included in this study. With the panel of 65 putatively neutral markers, there was no clear evidence for genetic structure of tributary collections within the same river system except the two sites in the Boise River and three sites from the Bruneau River (Figure 3a). Further, bootstrap support was below 50% for all clusters (except the two Boise River sites) indicating genetic similarity among collections and limited divergence. Conversely, clear patterns of climate adaptation were observed with the five temperature associated markers (Figure 3b). Clustering patterns from the neighbor-joining tree with the five temperature associated SNPs indicated each site may be adapted to one of three climates (cool, intermediate, or warm; Figure 3b, Table 2). Of note, one collection (McMullen Cr.) that was originally classified as a desert population clustered with cool populations, and another desert collection appeared intermediate (Deer Cr.). All collections originally considered in the montane regime clustered as cool or intermediate in the neighbor-joining tree. When reclassified into climate categories, air temperature was significant among groups, but differences in N_e and H_E were not significant (Table 2).

Results with the five temperature associated SNPs also revealed that cluster patterns in the neighbor-joining tree were consistent with results from pairwise F_{ST} with these five markers (Table 3), as both analyses demonstrated genetic similarity of sites within each group of cool, intermediate, or warm adapted populations. Population pairs within cool and intermediate groups were not well differentiated (15 of 28 comparisons were not significant), while the warm adapted populations remained significantly different from one another and cool/intermediate collections in all but one pairwise comparison. Of the warm adapted populations, Bennett Cr. had the least differentiation from intermediate sites. Results in STRUCTURE with the five temperature associated SNPs had k = 2, with three desert populations (Big Jacks Cr., Little Jacks Cr. and Duncan Cr.) containing a majority of membership of one group, and all other populations having majority membership in the other group.

There was not significant support for an isolation-by-distance model (Figure 4a) among all collections with 65 putatively neutral SNPs ($r^2 = 0.02$, Mantel test $p = 0.49$). However, when desert populations were removed, montane populations had a significant isolation-by-distance relationship ($r^2 = 0.25$, Mantel test $p = 0.012$). Isolation-by-distance was not significant with the 5 temperature candidate markers ($r^2 = 0.006$, Mantel test $p = 0.14$). These weak relationships of isolation-by-distance were in contrast to the strong isolation-by-temperature relationship observed among collections with the 5 candidate SNPs associated with temperature (Figure 4b; $r^2 = 0.69$, Mantel test $p < 0.00001$). The isolation-by-temperature relationship remained significant with all 18 temperature SNPs that were initially identified by SAM ($r^2 = 0.53$, Mantel test $p = 0.008$), but not with the remaining 58 SNPs ($r^2 = 0.01$, Mantel test $p = 0.14$). Single marker relationships with other environmental variables (precipitation and specific conductivity) were not explored due to limited ability to detect population differentiation with one locus.

Discussion

In this study, we demonstrated that environmental factors may limit gene flow and also act as a driving force for local adaptation of redband trout in various climates.

Previous studies have established temperature as an environmental variable that shapes genetic structure in natural populations of salmonids (Dionne *et al.* 2008; Narum *et al.* 2008). Here we found clear evidence for genetic divergence and limited gene flow in desert populations, but redband trout from cooler sites were genetically similar to one-another. Climate of desert streams may limit gene flow of redband trout directly through barriers to dispersal such as intermittent stream flow in summer seasons and disconnect among stream systems (Zoellick 1999; Meyer *et al.* 2010). While an isolation-by-distance relationship with neutral markers was not significant across all desert and montane populations, this was likely due to the high genetic differentiation of isolated desert populations in close proximity to one-another. In fact, montane populations fit an isolation-by-distance model when desert collections were removed from the analysis with neutral markers. Despite higher potential for gene flow in montane populations and evidence for stronger drift in desert populations, neither genetic diversity nor effective sizes were significantly different among these groups. Overall, neutral genetic variation appears to be influenced by a variety of factors including barriers to dispersal (i.e., disconnected desert streams), localized spawning, and small population size which are known to influence genetic structure in this species (e.g., Taylor 1995; Heath et al. 2002).

Conversely, divergence at candidate loci may be reinforced by local adaptation to environmental conditions (i.e., isolation-by-adaptation; Nosil *et al.* 2008). In the current study, a pattern of isolation-by-temperature was evident for loci associated with temperature but not for neutral markers, suggesting that the multi-locus panel of candidate markers was successful at identifying adaptive divergence among collections in differing climates. Further, greater levels of adaptive than neutral divergence were apparent in specific populations of redband trout, particularly among those at temperature extremes from other populations in the study. In addition to markers associated with temperature, two other climate related variables (precipitation and specific conductivity) were found to be significantly associated with two SNP markers. The strong association of candidate loci with temperature, precipitation, and specific conductivity across populations provided support for selection gradients in redband trout related to their environment. These results indicated that selection is acting to influence allele frequencies at specific genes and redband trout have likely adapted to differing climates throughout their range.

While a single non-synonymous SNP can be responsible for genetic adaptation to a particular environment (e.g., Hoekstra *et al.* 2006), many traits are polygenic. Selection for traits such as thermal tolerance in salmonids may encompass a wide variety of biological pathways such as development rate (Hendry *et al.* 1998), immune response to specific pathogens (Dionne *et al.* 2007), metabolism (Rodnick *et al.* 2004), and stress response (Fowler *et al.* 2009). Our results suggest that numerous candidate genes are involved in climate adapation in redband trout, consistent with studies in fish indicating that temperature can affect expression of a variety of genes (Kassahn *et al.* 2007) and life stages (Fowler *et al.* 2009). The evidence for adaptive divergence with multiple candidate SNP markers suggests that many genes are involved in selection and adaptation of redband trout to differing environments.

Markers identified as candidates for selection in this study will require further validation given certain caveats of our approach. Our study included only a limited number of markers due to the lack of genome information for this species, and thus the

markers identified as candidates for selection may be part of a selective sweep that would only be identified with a denser scan. A more thorough genome scan would also be likely to produce further candidates than those observed here. Another caveat is that correlation tests of MAF with temperature were based primarily on air instead of water temperatures due to the need for long term datasets to characterize selective environments over many generations. Daily air temperature was shown to be significantly correlated with daily stream temperatures at all sites in 2009 (Supplemental Figure S1), but long term stream data would likely improve regression analysis for redband trout. However, long term stream temperature data sets (~20-30 years) that better capture oscillations in climate are rare and we have demonstrated that air temperature may serve as an adequate proxy when water temperature data is limited. This relationship is consistent with studies that have specifically examined correlation of air and water temperatures (e.g., Morrill *et al.* 2005).

Adequate characterization of environmental variables is an important but challenging task for genetic association studies in natural populations. Variability in natural environments is inherent as seasonal and annual changes occur with climate related factors (e.g., temperature and precipitation), and human related disturbances create difficulty in evaluating habitat over long term periods. In some cases, inter-related environmental variables may benefit studies since one or more factors may be identified as significant when long term variation cannot be fully realized. This complexity in characterizing natural environments also reinforces the need for validation of genetic and phenotypic associations under controlled environments.

Candidate markers that were associated with environmental variables are reasonable targets for further study, however, the significant signal may also be due to linkage with genes or markers not included in this study (e.g., Charlesworth *et al.* 1997). We identified six SNPs as highly significant candidates for selection related to environment in populations of redband trout. These six SNPs were located in cis-regulatory (non-coding) regions of genes, and thus may be under regulatory control or closely linked with variation in coding regions. Putative gene function for the six candidate markers (Table 4) suggest that they may play roles in local adaptation to climate in traits such as thermal tolerance and disease resistance. However, these may simply be genes closely linked to the actual regions of the genome controlling for these traits. Only one (Omy_tlr5-205) of the six candidate markers identified in this study was significant with the F_{ST} outlier approach (Beaumont & Nichols 1996) and highlights the potential for false negatives with general F_{ST}-based approaches. Additional loci identified as candidates with the F_{ST} outlier approach may be under climate independent selection and thus would not be expected to be correlated with variables tested in this study, or they may be false positives. Further, habitat characteristics that are related to climate may co-vary and it was not possible in this study to identify or predict specific phenotypic traits with genotypic data at candidate markers. However, common garden studies under controlled environments may assist with determination of candidate markers and their association with specific traits such as thermal tolerance, disease resistance, growth, and stress response. More extensive sequencing efforts (i.e., pyro and next generation sequencing) and gene expression of candidate regions may also help to elucidate the specific genes and their functional role in adaptation to climate.

Despite these challenges, genome-wide association studies with dense SNP markers offer the potential to predict phenotypes and "genetic merit" of individuals (Lee *et al.* 2007; Wray *et al.* 2009) or populations (de Roos *et al.* 2009). While these types of methods have been advocated to predict genetic risk in humans (e.g., Morrison *et al.* 2007) and to enhance artificial selection programs (e.g., Lande & Thompson 1990), they may also have the potential to identify adaptive variation in natural animal populations for ecological and conservation applications. Our study identified six candidate markers in redband trout that were significantly associated with environment, and further evaluations may lead to a larger number of markers to make robust predictions regarding this species' ability to adapt to climate change. Previous studies suggest that climate change may greatly impact seasonal cues in nature (e.g., Bradshaw & Holzapfel 2008) and cause shifts in species distributions and migration patterns (Berthold *et al.* 1992; Bradshaw *et al.* 2004; Hari *et al.* 2006). However, locally adapted reaction norms may be sufficiently plastic to allow for adaptation to different environments if phenotypic and genetic variation exists (e.g., Jensen et al. 2008). Thus, monitoring of candidate gene allele frequencies along genetic clines may prove to be effective for quantifying the influence of climate change on natural populations (e.g., Umina et al. 2005). Further, the potential to predict genetic adaptability of individuals and populations to changing climate conditions may have profound implications for many species that face extensive anthropogenic disturbances, but more advanced models are needed that address this issue.

In this study, candidate markers provided the opportunity to classify populations of redband trout according to local adaptation rather than general location. Three categories (cool, intermediate, and warm) were observed for redband trout that did not always correspond to their initial climate classification (desert or montane) based on stream location. This was particularly true for sites initially designated as desert populations and may reflect the variability of a stream within the same corridor since headwaters may be located in cooler environments than lower segments. Further, upwelling from groundwater sources are common and may create thermal refuges for fish and other aquatic organisms (e.g., Ebersole et al. 2001; Tate et al. 2007). Therefore, characterization of local adaptation with genetic markers may be an effective tool to determine biological differences (i.e., adaptive variation) for conservation of redband trout and other species.

References

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

Airaksinen S, Rabergh CMI, Sistonen L, Nikinmaa M (1998) Effects of heat shock and hypoxia on protein sysnthesis in rainbow trout (Oncorhynchus mykiss) cells. The Journal of Experimental Biology, 201, 2543-2551.

Akey JM (2009) Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Research, 19, 711-722.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MC (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Research, 12, 1805-1814.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics, 9, 323.

Beacham TD, Murray CB (1990) Temperature, egg size, and development of embryos and alevins of five speciesof Pacific salmon: a comparative analysis. Transactions of the American Fisheries Society, 119, 927–945.

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

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology, 13, 969-980.

Beer WN, Anderson JJ (2001) Effect of spawning day and temperature on salmon emergence: interpretations of a growth model for Methow River Chinook. Canadian Journal of Fisheries and Aquatic Sciences, 58, 943–949.

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

Berthold P, Helbig AJ, Mohr G et al. (1992) Rapid microevolution of migratory behaviour in a wild bird species. Nature, 360, 668–670.

Bilodeau AL, Waldbieser GC (2005) Activation of TLR3 and TLR5 in channel catfish exposed to virulent Edwardsiella ictaluri. Developmental and Comparative Immunology, 29, 713–721.

Bradshaw WE, CM Holzapfel (2008) Genetic response to rapid climate change: it's seasonal timing that matters. Molecular Ecology, 17, 157-166.

Bradshaw WE, Zani PA, Holzapfel CM (2004) Adaptation to temperate climates. Evolution, 58, 1748–1762.

Brannon EL (1987) Mechanisms stabilizing salmonid fry emergence timing. Canadian Special Publication of Fisheries and Aquatic Sciences, 96, 120–124.

Buckley BA, Somero GN (2009) cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish Trematomus bernacchii to alter gene expression in response to heat stress. Polar Biology, 32, 403-415.

Campbell NR, Narum SR (2009a) Quantitative PCR assessment of microsatellite and SNP genotyping with variable quality DNA extracts. Conservation Genetics, 10, 779-784.

Campbell NR, Narum SR (2009b) Identification and characterization of heat shock response related single nucleotide polymorphisms in O. mykiss and O. tschawytscha. Molecular Ecology Resources, 9, 1460-1559.

Campbell NR, Overturf K, Narum SR (2009) Characterization of 22 novel single nucleotide polymorphism markers in steelhead and rainbow trout. Molecular Ecology, Resources 9, 318-322.

Cavalli-Sforza LL, Edwards WWF (1967) Phylogenetic analysis; models and estimation procedures. Evolution, 32, 550-570.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical Research, 70, 155-174.

Chase K, Jones P, Martin A, Ostrander EA, Lark KG (2009) Genetic mapping of fixed phenotypes: disease frequency as a breed characteristic. Journal of Heredity, 100, S37-S41.

Ciardiello MA, Camardella L, Carratore V, di Prisco G (2000) L-Glutamate dehydrogenase from the Antarctic fish Chaenocephalus aceratus: Primary structure, function and thermodynamic characterization: relationship with cold adaptation. Biochimica et Biophysica Acta – Protein Structure and Molecular Enzymology, 1543, 11-23.

Crozier LG, Hendry AP, Lawson PW, Quinn TP, Mantua NJ, Battin J, Shaw RG, Huey RB (2008) Potential responses to climate change in organisms with complex life histories: evolution and plasticity in Pacific salmon. Evolutionary Applications, 1, 252-270.

De Roos APW, Hayes BJ, Goddard ME (2009)Reliability of genomic predictions across multiple populations. Genetics, 183, 1545-1553.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (2007) Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution, 61, 2154-2167.

Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. Molecular Ecology, 17, 2382-2396.

Ebersole JL, Liss WJ, Frissell CA (2001) Relationship between stream temperature, thermal refugia and rainbow trout Oncorhynchus mykiss abundance in arid-land streams in the northwestern United States. Ecology of Freshwater Fish, 10, 1-10.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14, 2611-20.

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

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

Fowler SL, Hamilton D, Currie S (2009) A comparison of the heat shock response in juvenile and adult rainbow trout (Oncorhynchus mykiss) – implications for increased thermal sensitivity with age. Canadian Journal of Fisheries and Aquatic Sciences, 66, 91-100.

Gamperl AK, Rodnick KJ, Faust HA, Venn EC, Bennett MT, Crawshaw LI, Keeley ER, Powell MS, Li HW (2002) Metabolism, swimming performance, and tissue biochemistry of high desert redband trout (Oncorhynchus mykiss ssp.): Evidence for phenotypic differences in physiological function. Physiological and Biochemical Zoology, 75, 413–431.

Hari RE, Livingstone DM, Siber R, Burkhardt-Holm P, Guttinger H (2006) Consequences of climatic change for water temperature and brown trout populations in Alpine rivers and streams. Global Change Biology, 12, 10-26.

Heath DD, Busch C, Kelly J, Atagi DY (2002) Temporal change in genetic structure and effective population size in steelhead trout (Oncorhynchus mykiss). Molecular Ecology, 11, 197-214.

Hendry AP, Hensleigh JE, Reisenbichler RR (1998) Incubation temperature, developmental biology, and the divergence of sockeye salmon (Oncorhynchus nerka)

within Lake Washington. Canadian Journal of Fisheries and Aquatic Sciences, 55, 1387-1394.

Hendry AP, Farrugia TJ, Kinnison MT (2008) Human influences on rates of phenotypic change in wild animal populations. Molecular Ecology, 17, 20-29.

Hermisson J (2009) Who believes in whole-genome scans for selection? Heredity, 103, 283-284.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. Science, 313, 101–104.

Holt RA, A Amandi, JS Rohovec, JL Fryer (1989) Relation of water temperature to bacterial cold-water disease in coho salmon, Chinook salmon, and rainbow trout. Journal of Aquatic Animal Health, 1, 94-101.

Hoffmann AA, Hallas RJ, Dean JA, Schiffer M (2003) Low potential for climatic stress adaptation in a rainforest Drosophila species. Science, 301, 100-102.

Huber M, Guderley H (1993) The effect of thermal acclimation and exercise upon the binding of glycolytic enzymes in muscle of the goldfish Carassius auratus. Journal of Experimental Biology, 175, 195-209.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources, 9, 1322 – 1332.

Jensen LF, Hansen MM, Pertoldi C, Holdensgaard G, Mensberg KLD, Loeschck V (2008) Local adaptation in brown trout early life-history traits: implications for climate change adaptability. Proceeding of the Royal Society of London Series B, 275, 2859-2868.

Kassahn KS, Crozier RH, Ward AC, Stone G, Caley MJ (2007) From transcriptome to biological function: environmental stress in an ectothermic vertebrate, the coral reef fish Pomacentrus moluccensis. BMC Genomics, doi:10.1186/1471-2164-8-358

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.

Joost S, Kalbermatten M, Bonin A (2008) Spatial Analysis Method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection, Molecular Ecology Resources, 8, 957–960.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 124, 743–756.

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Langella O (2001) Populations 1.2.24: population genetic structure (individuals or populations distances, phylogenetic trees). Available: http://www.pge.cnrs.gif.fr/bioinfo/populations/.

Lee SH, van der Werf JHJ, Hayes BJ, Goddard ME, Visscher PM (2008) Predicting unobserved phenotypes for complex traits from whole-genome SNP Data. PLoS Genetics, 4: e1000231. doi:10.1371/journal.pgen.1000231

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.

Lyman RF, Mackay TFC (1998) Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in Drosophila melanogaster: the delta-hairless gene region. Genetics, 149, 983-998.

Marcogliese DJ (2001) Implications of climate change for parasitism of animals in the aquatic environment. Canadian Journal of Zoology, 79, 1331–1352.

Meyer KA, Lamansky J, Schill DJ (2010) Biotic and abiotic factors related to redband trout occurrence and abundance in desert and montane streams. Western North American Naturalist, 70, 67-76.

Morrill JC, Bales RC, Conklin MH (2005) Estimating stream temperature from air temperature: implications for future water quality. Journal of Environmental Engineering, 131, 139-146.

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

Morrison AC, et al. (2007) Prediction of coronary heart disease risk using a genetic risk score: the atherosclerosis risk in communities study. American Journal of Epidemiology, 166, 28–35.

Narum SR (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics, 7, 783-787.

Narum SR, Zendt J, Graves D, Sharp B (2008) Influence of landscape on resident and anadromous life history types of Oncorhynchus mykiss. Canadian Journal of Fisheries and Aquatic Sciences, 65, 1013-1023.

Neuhauss SCF, Biehlmaier O, Seeliger MW, Das T, Kohler K, Harris WA, Baier H (1999) Genetic disorders of vision revealed by a behavioral screen of 400 essential loci in zebrafish. Journal of Neuroscience, 19, 8603-8615.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. Evolution, 62, 316-336.

Olsen JB, Bentzen P, Seeb JE (1998) Characterization of seven microsatellite loci derived from pink salmon. Molecular Ecology, 7, 1083-1090.

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences, 12, 357-358.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. Heredity, 96, 7–21.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38, 904–909.

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. American Journal of Human Genetics, 65, 220-228.

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity, 86, 248-249.

Rodnick KJ, Gamperl AK, Lizars KR, Bennett MT, Rausch RN, Keeley ER (2004) Thermal tolerance and metabolic physiology among redband trout populations in south-eastern Oregon. Journal of Fish Biology, 64, 310-335.

Sprowles AE, Stephens MR, Clipperton NW, and May BP (2006) Fishing for SNPs: A targeted locus approach for single nucleotide polymorphism discovery in rainbow trout. Transactions of the American Fisheries Society, 135, 1698-1721.

Tate KW, Lancaster DL, DF Lile (2007) Assessment of thermal stratification within stream pools as a mechanism to provide refugia for native trout in hot, arid rangelands. Environmental Monitoring and Assessment, 124, 289-300.

Taylor EB (1995) Genetic variation at minisatellite DNA loci among North Pacific populations of steelhead trout and rainbow trout (Oncorhynchus mykiss). Journal of Heredity, 86, 354–363.

Torgerson DG, et al. (2009) Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. PLoS Genetics, doi:10.1371/journal.pgen.1000592

Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA (2005) A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. Science, 308, 691-693.

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Molecular Ecology, 14, 3623-2642.

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics, 7, 167-184.

Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources, 8, 753-756.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution, 38, L1358-1370.

Wray NR, Goddard ME, Visscher PM (2007) Prediction of individual genetic risk from genome-wide association studies. Genome Research, 17, 1520-1528.

Xue Y, et al. (2009) Population differentiation as an indicator of recent positive selection in humans: an empirical evaluation. Genetics, 183, 1065-1077.

Zhang F, Wang Y, Deng H (2008) Comparison of population-based association study methods correcting for population stratification. PLoS One, doi:10.1371/journal.pone.0003392

Zinn JL, Johnson KA, Sanders JE, Fryer JL (1977) Susceptibility of salmonid species and hatchery strains of Chinook salmon (Oncorhynchus tshawytscha) to infections by Ceratomyxa shasta. Journal of the Fisheries Research Board of Canada, 34, 933–936.

Zoellick BW (1999) Stream temperatures and the elevational distribution of redband trout in southwestern Idaho. Great Basin Naturalist, 59, 136-143.

Table 1. Sampling locations and environmental variables for redband trout collected in drainages of southwest Idaho, USA.

Map No.	Region	Collection	Drainage	Latitude / Longitude	n	Mean Annual Max. Air Temp (°C)	Elevation (m)	Gradient (%)	Specific Conductivity (µS/cm)	Width (m)	Fines (%)	Gravel (%)	Cobble/ Boulder (%)	Mean Annual Precipitation (cm)
1	Desert	Deer Cr.	Bruneau R.	42.043633 / -115.09903	54	13.8	1842	1.9	71.0	1.3	5.0	25.0	57.0	45.2
2	Desert	Big Jacks Cr.	Bruneau R.	42.532970 / -116.18385	30	14.5	1704	3.2	150.0	1.6	5.0	26.0	54.0	38.1
3	Desert	Little Jacks Cr.	Bruneau R.	42.728700 / -116.10516	60	16.8	1081	1.4	124.0	2.3	9.0	25.0	31.0	26.9
4	Desert	Duncan Cr.	Bruneau R.	42.492565 / -116.06401	73	15.0	1638	1.8	134.0	2.4	25.0	38.0	36.0	31.2
5	Desert	McMullen Cr.	Snake R.	42.329260 / -114.38737	28	14.2	1486	0.3	108.0	1.8	7.0	56.0	31.0	32.4
6	Desert	Bennett Cr.	Snake R.	43.224200 / -115.50603	30	14.6	1406	1.1	95.0	2.0	24.0	22.0	5.0	54.9
7	Montane	Johnson Cr.	Boise R.	43.936886 / -115.28058	59	13.4	1716	2.1	n/a	n/a	n/a	n/a	n/a	82.7
8	Montane	upp Manns Cr.	Weiser R.	44.574870 / -116.95055	31	12.1	1502	1.5	120.0	4.7	0.0	27.0	37.0	87.4
9	Montane	Fawn Cr.	Payette R.	44.382336 / -116.05894	30	12.2	1596	0.6	25.0	3.2	17.0	54.0	6.0	73.8
10	Montane	L.Weiser Cr.	Weiser R.	44.525684 / -116.24255	29	13.3	1377	3.5	65.0	8.4	3.7	7.9	78.9	69.2
11	Montane	Whiskey Jack Cr.	Boise R.	43.649063 / -115.35925	53	13.0	1714	3.7	58.0	4.1	15.0	39.0	33.0	85.5
12	Montane	Keithley Cr.	Weiser R.	44.553380 / -116.88535	22	13.0	1370	5.5	96.0	3.4	7.0	27.0	48.0	87.8

n/a = data not available

Table 2. Reclassification of sites by climate categories, followed by unbiased heterozygosity (H_E) and estimate of effective size (N_e) with 95% CI. * indicates significant value in ANOVA (air temperature $p = 0.009$).

Map No.	Climate Category	Collection	Region	Annual Avg. Max. Air Temp (°C)*	H_E	Ne	95% CI
5	Cool	McMullen Cr.	Desert	14.2	0.173	31.5	19.1-65.6
8	Cool	upper Manns Cr.	Montane	12.1	0.156	89.8	37.5-infinite
9	Cool	Fawn Cr.	Montane	12.2	0.141	26.1	15.6-53.3
11	Cool	Whiskey Jack Cr.	Montane	13.0	0.165	63.6	39.0-131
Average of "cool" populations				12.9	0.159	52.8	---
1	Intermediate	Deer Cr.	Desert	13.8	0.151	406.3	91.4-infinite
7	Intermediate	Johnson Cr.	Montane	13.4	0.166	228.5	90.7-infinite
10	Intermediate	L. Weiser Cr.	Montane	13.3	0.180	99.4	40.1-infinite
12	Intermediate	Keithley Cr.	Montane	13.0	0.167	49.5	22.6-871.8
Average of "intermediate" populations				13.4	0.166	195.9	---
2	Warm	Big Jacks Cr.	Desert	14.5	0.176	34.8	21.3-71.7
3	Warm	Little Jacks Cr.	Desert	16.8	0.146	133	61.1-2426.1
4	Warm	Duncan Cr.	Desert	15.0	0.153	18.3	14.1-23.6
6	Warm	Bennett Cr.	Desert	14.6	0.148	350.9	44.2-infinite
Average of "warm" populations				15.2	0.156	134.3	---

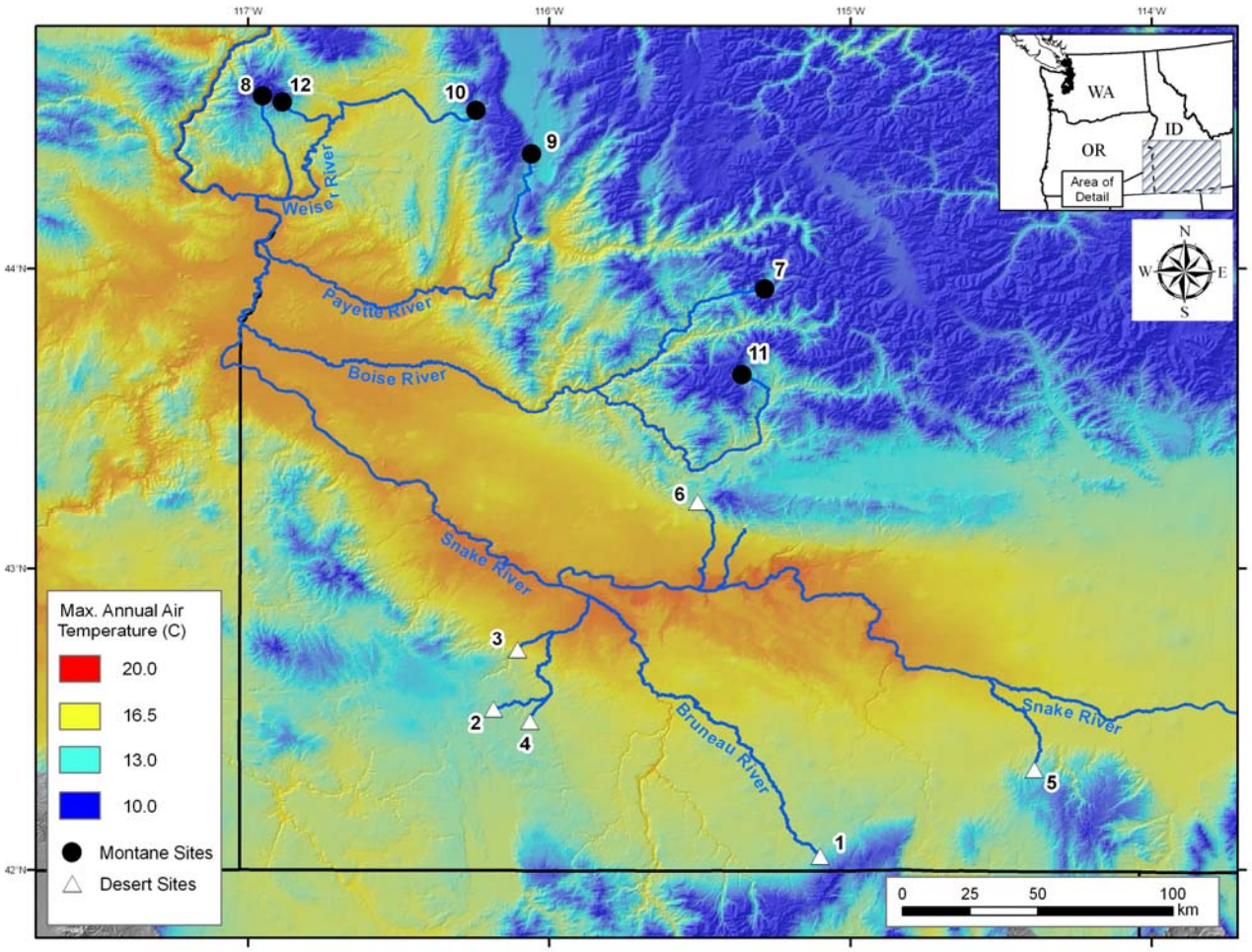
Table 3. Pairwise F_{ST} with 65 putatively neutral markers below diagonal, and 5 temperature associated SNPs above the diagonal. All pairwise comparisons were significant at corrected critical value of 0.01 except those labeled with "NS" (not significant). Differences in upper versus lower F_{ST} matrices that were ≥ 2.0 are indicated by "2X". The first six collections were from desert regions, and the last six collections were from montane regions.

	Deer Cr. 1	Big Jacks Cr. 2	Little Jacks Cr. 3	Duncan Cr. 4	McMullen Cr. 5	Bennett Cr. 6	Johnson Cr. 7	up Manns Cr. 8	Fawn Cr. 9	L.Weiser Cr. 10	Whiskey Jack Cr. 11	Keithley Cr. 12
Deer Cr.	---	0.149	0.387 2X	0.143	0.050	0.041	-0.006 NS	0.055	0.120	0.000 NS	0.027	0.006 NS
Big Jacks Cr.	0.132	---	0.166	0.087 2X	0.061	0.062	0.130	0.096	0.279 2X	0.126	0.103	0.114
Little Jacks Cr.	0.193	0.104	---	0.153	0.319	0.312	0.382 2X	0.362	0.504 2X	0.357 2X	0.315	0.339
Duncan Cr.	0.144	0.042	0.166	---	0.150	0.099	0.147	0.150	0.298 2X	0.129	0.104	0.112
McMullen Cr.	0.113	0.153	0.236	0.187	---	0.058	0.031	0.006 NS	0.089	0.022 NS	0.018 NS	0.019 NS
Bennett Cr.	0.090	0.123	0.221	0.157	0.060	---	0.028 NS	0.053	0.208 2X	0.049	0.073	0.049
Johnson Cr.	0.060	0.083	0.155	0.110	0.078	0.084	---	0.032 NS	0.100	-0.003 NS	0.024	0.001 NS
upper Manns Cr.	0.068	0.090	0.188	0.091	0.067	0.060	0.045	---	0.085 2X	0.025 NS	0.027 NS	0.015
Fawn Cr.	0.125	0.085	0.189	0.103	0.090	0.067	0.079	0.043	---	0.074	0.099	0.078 2X
L.Weiser Cr.	0.050	0.064	0.165	0.077	0.059	0.055	0.037	0.025	0.044	---	0.003 NS	-0.017 NS
Whiskey Jack Cr.	0.095	0.093	0.187	0.122	0.090	0.092	0.049	0.065	0.068	0.070	---	-0.005 NS
Keithley Cr.	0.078	0.061	0.173	0.096	0.074	0.030	0.046	0.031	0.035	0.033	0.061	---

Table 4. Putative gene function of candidate markers associated with temperature, precipitation, and conductivity.

SNP Marker	Gene	Putative Function	Reference
Omy_gdh-271 (temperature)	Glutamate dehydrogenase	An enzyme present in mitochondria of eukaryotes that converts glutamate to α-Ketoglutarate, for urea synthesis. Relevant Finding: Cold adaptation in Antarctic fish	Ciardiello *et al.* 2000
Omy_stat3-273 (temperature)	Signal Transducers and Activator of Transcription	This protein transmits signals that help control the body's response to bacteria and fungi, and regulation of inflammation. Relevant Finding: Response to heat stress	Buckley and Somero 2009
Omy_aldB-165 (temperature)	Aldolase B	A glycolytic enzyme found in the liver and kidney. Relevant Finding: Thermal acclimation & stress	Huber and Guderley 1993
Omy_tlr5-205 (temperature)	Toll-like receptor 5	This gene is a member of the Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of innate immunity. Relevant Finding: Immune response	Bilodeau and Waldbieser 2005
Omy_Ogo4-212 (temperature and conductivity)	unknown	unknown; no BLAST similarity	Olsen *et al.* 1998
Omy_hsf2-146 (precipitation)	Heat shock factor 2	Regulates the synthesis of heat-shock proteins during differentiation and development. Relevant Finding: Responsive to erythrocyte differentiation	Airaksinen *et al.* 1998

Figure 1. Map of the study area in southwest Idaho, with inset of Pacific Northwest USA. Sites where redband trout (*O. mykiss*) were collected are numbered and correspond to stream locations in Table 1. Sites classified as desert or montane streams are indicated by either white triangles or black dots, respectively.



.

Figure 2. Correlation of minor allele frequency (MAF) for five candidate temperature markers and one candidate precipitation marker.

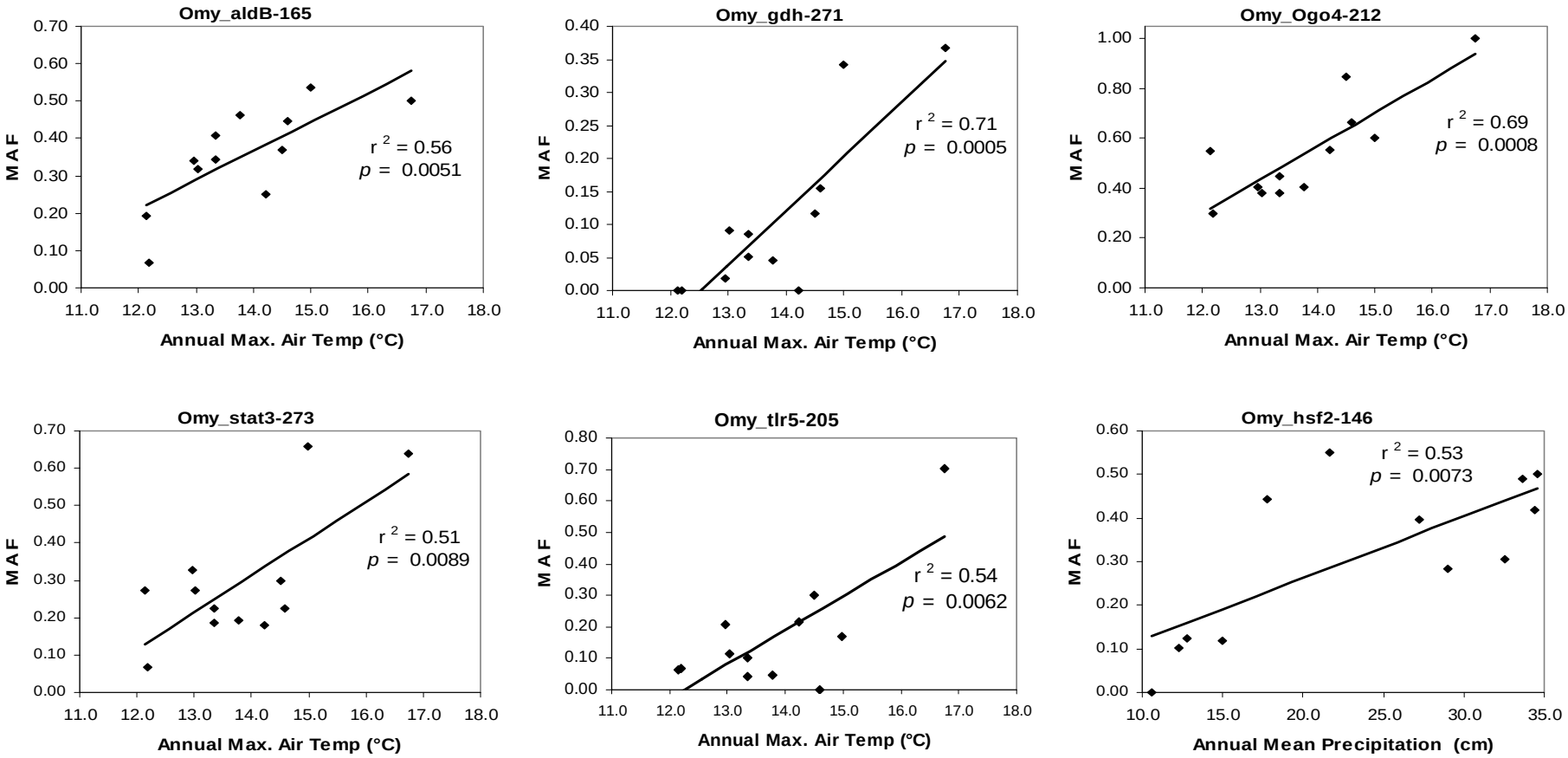
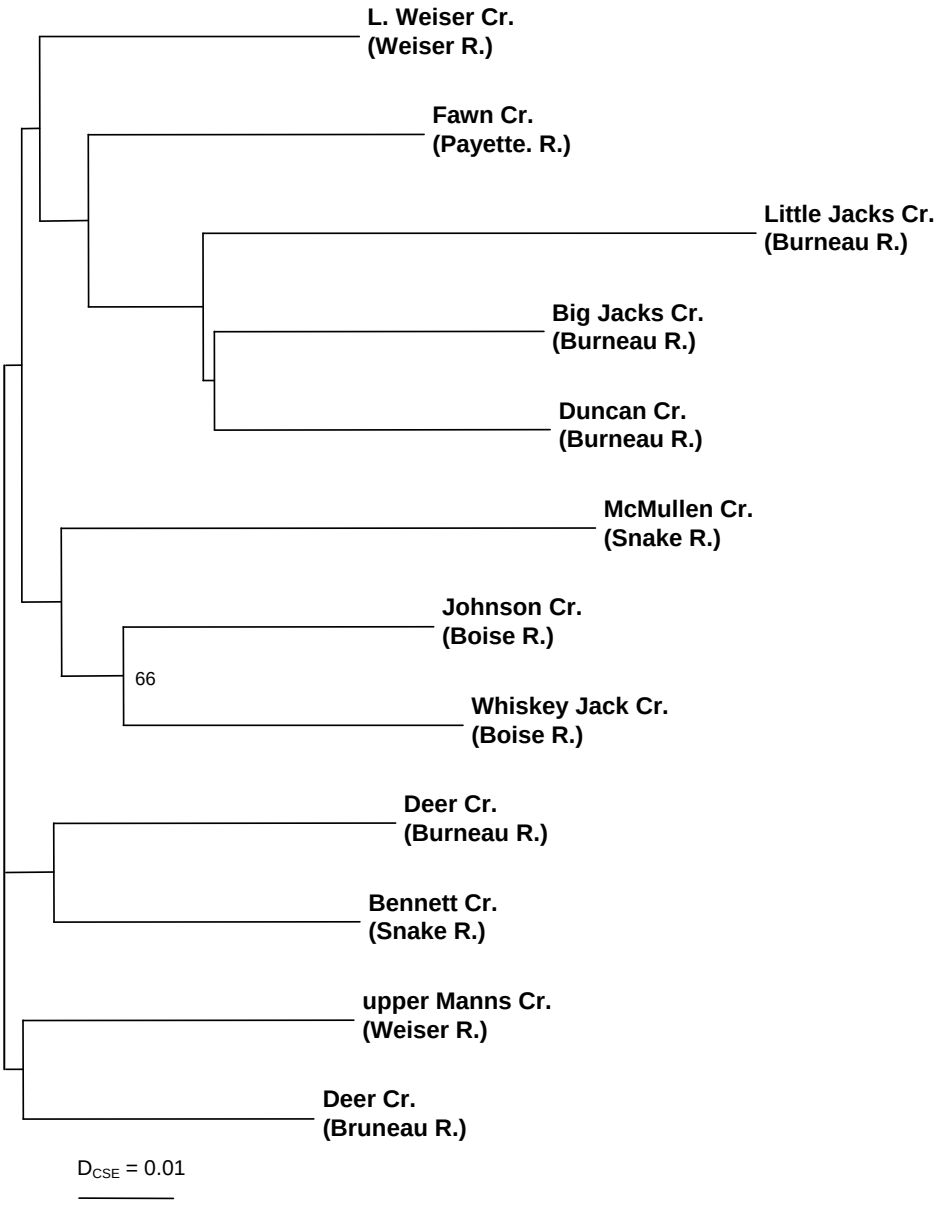


Figure 3. Neighbor-joining diagram of 12 collections of redband trout (*O. mykiss*) as determined by a) panel of 65 putatively neutral SNPs, and b) panel of 5 candidate markers associated with temperature. Arrows depict cool, intermediate, or warm adapted populations as categorized by clustering patterns. Chord distance (Cavalli-Sforza and Edwards 1967) is given at the lower left, and bootstrap values from 1,000 iterations that were greater than 50% are shown.

3a)



3b)

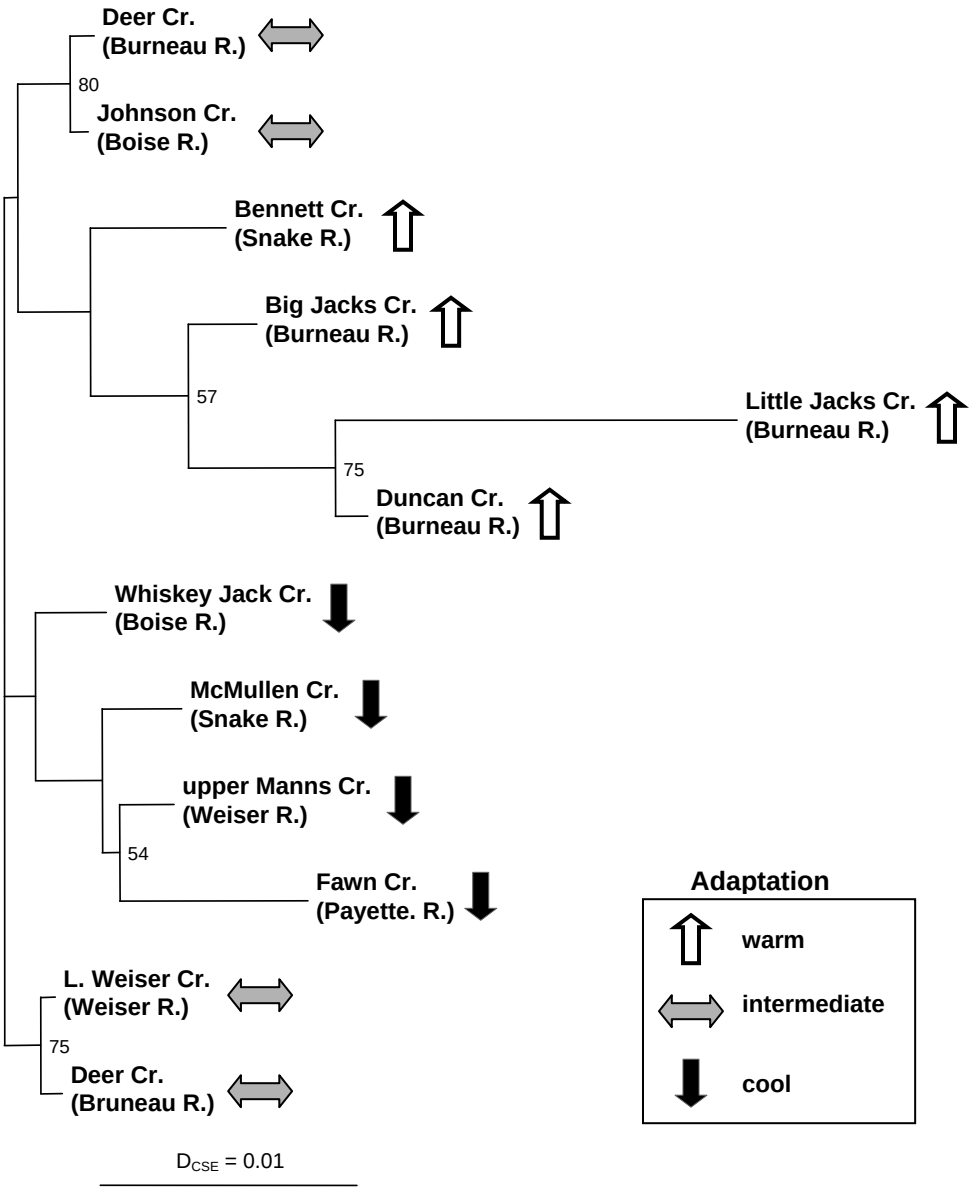
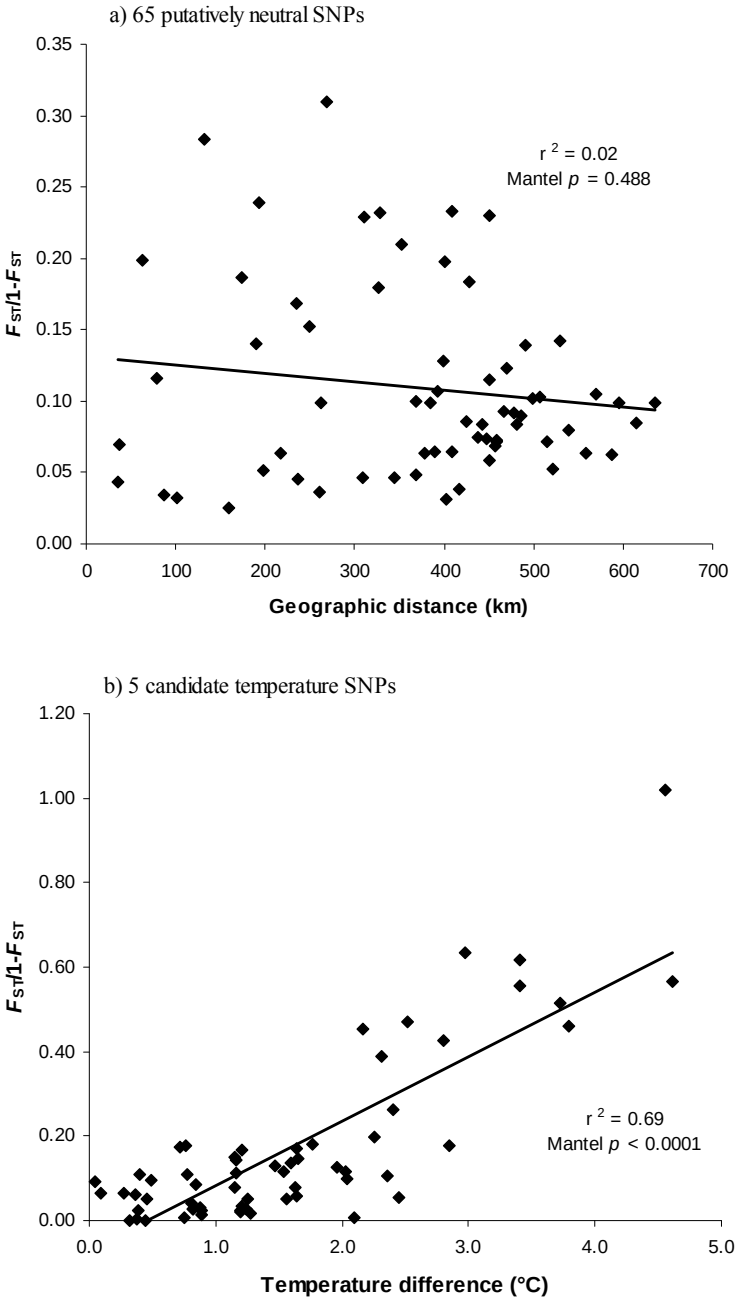

Deer Cr.
(Burneau R.)
80
Johnson Cr.
(Boise R.)
Bennett Cr.
(Snake R.)
Big Jacks Cr.
(Burneau R.)
57
Little Jacks Cr.
(Burneau R.)
75
Duncan Cr.
(Burneau R.)
Whiskey Jack Cr.
(Boise R.)
McMullen Cr.
(Snake R.)
upper Manns Cr.
(Weiser R.)
54
Fawn Cr.
(Payette. R.)
L. Weiser Cr.
(Weiser R.)
75
Deer Cr.
(Bruneau R.)
D_{CSE} = 0.01
Adaptation
warm
intermediate
cool

Figure 4. Patterns of a) isolation-by-distance (65 neutral SNPs), and b) isolation-by-temperature (5 candidate SNPs). Pairwise Fst values are means over loci included in each panel. Geographic distance equals fluvial distance among sites, and temperature equals the difference in annual maximum air temperature of sites.



Supplemental Table S1. List of 96 SNP assays genotyped in *O. mykiss gairdneri*.

	Marker Name	Gene name:	Removed	Assay Reference:
1	Omy_myclarp404-111	acidic ribosomal phosphoprotein	*	Unpublished N. Campbell[1]
2	Omy_myclgh1043-156	growth hormone 1	*	Unpublished N. Campbell[1]
3	Omy_Omyclmk436-96	map kinase 4	*	Unpublished N. Campbell[1]
4	Omy_113490-159	unknown gene (expressed)		Unpublished C. Garza[2]
5	Omy_114315-438	unknown gene (expressed)		Unpublished C. Garza[2]
6	Omy_121006-131	immediate early response protein 5		Unpublished C. Garza[2]
7	Omy_121713-115	similar to vertibrate polymerase		Unpublished C. Garza[2]
8	Omy_123044-128	unknown gene (expressed)		Unpublished C. Garza[2]
9	Omy_123048-119	40S ribosomal protein S26		Unpublished C. Garza[2]
10	Omy_126278-43	Hsp70 binding protein 1	**	Unpublished C. Garza[2]
11	Omy_127236-583	A312aR protein		Unpublished C. Garza[2]
12	Omy_128693-455	Stathmin-like 4	***	Unpublished C. Garza[2]
13	Omy_130295-98	type II keratin K8b		Unpublished C. Garza[2]
14	Omy_130524-160	Growth arrest and DNA-damage-inducible protein		Unpublished C. Garza[2]
15	Omy_187760-385	unknown gene (expressed)		Unpublished C. Garza[2]
16	Omy_95489-239	unknown gene (expressed)		Unpublished C. Garza[2]
17	Omy_96222-125	elongation factor 1-alpha		Unpublished C. Garza[2]
18	Omy_97077-73	Anti-proliferative cofactor		Unpublished C. Garza[2]
19	Omy_97660-230	unknown gene (expressed)		Unpublished C. Garza[2]
20	Omy_97865-196	60S ribosomal protein L37a		Unpublished C. Garza[2]
21	Omy_97954-618	unknown gene (expressed)		Unpublished C. Garza[2]
22	Omy_aldB-165	aldolase B		Campbell *et al.* (2009)
23	Omy_aldB-414	aldolase B	***	Campbell *et al.* (2009)
24	Omy_ALDOA_1	aldolase A		Aguilar and Garza (2008)
25	Omy_aromat-280	aromatase		Unpublished J. DeKoning[3]
26	Omy_arp-630	acidic ribosomal phosphoprotein		Campbell *et al.* (2009)
27	Omy_aspAT-123	aspartate aminotransferase	***	Campbell *et al.* (2009)
28	Omy_aspAT-413	aspartate aminotransferase		Campbell *et al.* (2009)
29	Omy_b1-266	BAC clone sequence		Sprowles *et al.* (2006)
30	Omy_b9-164	BAC clone sequence		Sprowles *et al.* (2006)
31	Omy_BAC-B4-126	from BAC clone	***	Unpublished S. Young[4]
32	Omy_BAC-B4-324	from BAC clone	***	Unpublished S. Young[4]
33	Omy_cd28-130	CD28 like protein		Unpublished J. DeKoning[3]
34	Omy_cd59-206	CD59 like protein 1		Unpublished J. DeKoning[3]
35	Omy_cd59b-112	CD59 like protein 2		Unpublished J. DeKoning[3]
36	Omy_colla1-525	alpha 1 type 1 collagen		Unpublished J. DeKoning[3]
37	Omy_cox1-221	cyclooxygenase 1		Campbell *et al.* (2009)
38	Omy_cox2-335	cyclooxygenase 2		Unpublished J. DeKoning[3]
39	Omy_crb-106	carbonyl reductase - intron 3	***	Sprowles *et al.* (2006)
40	Omy_CRBF1-1	carbonyl reductase - intron 1		Aguilar and Garza (2008)
41	Omy_cxcr-169	CXC chemokine receptor		Unpublished J. DeKoning[3]
42	Omy_e1-147	BAC clone sequence		Sprowles *et al.* (2006)
43	Omy_g1-103	glucose-6 phosphate dehydrogenase		Sprowles *et al.* (2006)
44	Omy_g12-82	cDNA clone 1RT158N23_B_G12		Unpublished J. DeKoning[3]
45	Omy_gdh-271	glutamate dehydrogenase		Campbell *et al.* (2009)
46	Omy_GH1P1-2	growth hormone 1 (promoter)		Aguilar and Garza (2008)
47	Omy_gh-334	growth hormone 1		Campbell *et al.* (2009)
48	Omy_gh-475	growth hormone 1		Campbell *et al.* (2009)
49	Omy_gsdf-291	gonadal-somal derived growth factor		Unpublished J. DeKoning[3]
50	Omy_hsc713-453	heat shock cognate 71	***	Campbell & Narum (2009b)
51	Omy_hsc715-80	heat shock cognate 71		Campbell & Narum (2009b)

52	Omy_hsf1b-241	heat shock factor 1b		Campbell & Narum (2009b)
53	Omy_hsf2-146	heat shock factor 2		Campbell & Narum (2009b)
54	Omy_hsp47-86	heat shock protein 47		Campbell & Narum (2009b)
55	Omy_hsp70aPro-329	heat shock protein 70a (promoter)		Campbell & Narum (2009b)
56	Omy_hsp90BA-193	heat shock protein 90BA	***	Campbell & Narum (2009b)
57	Omy_hsp90BA-229	heat shock protein 90BA		Campbell & Narum (2009b)
58	Omy_IL17-185	interleukin 17		Unpublished J. DeKoning[3]
59	Omy_IL1b-163	interleukin 1b		Unpublished J. DeKoning[3]
60	Omy_IL6-320	interleukin 6		Unpublished J. DeKoning[3]
61	Omy_inos-97	inducible nitric oxide synthase		Unpublished J. DeKoning[3]
62	Omy_ldh-156	lactate dehydrogenase	***	Sprowles *et al.* (2006)
63	Omy_LDHB-1_i2	lactate dehydrogenase		Aguilar & Garza (2008)
64	Omy_LDH-B2	lactate dehydrogenase	**	Brunelli *et al.* (2008)
65	Omy_LDHB-2_e5	lactate dehydrogenase		Aguilar & Garza (2008)
66	Omy_LDHB-2_i6	lactate dehydrogenase		Aguilar & Garza (2008)
67	Omy_mapK3-103	map kinase 4		Unpublished N. Campbell[1]
68	Omy_mcsf-268	macrophage colony-stimulating factor receptor		Unpublished J. DeKoning[3]
69	Omy_mcsf-371	macrophage colony-stimulating factor receptor		Unpublished J. DeKoning[3]
70	Omy_MYC_2	c-Myc (3' UTR)		Aguilar & Garza (2008)
71	Omy_myo1a-264	myostatin 1a	**	Campbell *et al.* (2009)
72	Omy_myo1b-111	myostatin 1b	**	Campbell *et al.* (2009)
73	Omy_myoD-178	myosin D		Campbell *et al.* (2009)
74	Omy_nach-200	voltage-gated sodium channel alpha type IV		Unpublished J. DeKoning[3]
75	Omy_NaKATPa3-50	Na/K ATPase alpha 3		Campbell *et al.* (2009)
76	Omy_NIT_R_1	novel immune type receptor 2	**	Aguilar & Garza (2008)
77	Omy_nkef-241	natural killer efficiency factor	***	Campbell *et al.* (2009)
78	Omy_nkef-308	natural killer efficiency factor		Campbell *et al.* (2009)
79	Omy_nramp-146	natural resistance associated macrophage protein		Campbell *et al.* (2009)
80	Omy_Ogo4-212	microsatellite Ogo4		Campbell *et al.* (2009)
81	Omy_OmyP9-180	P9-B genomic sequence		Sprowles *et al.* (2006)
82	Omy_Ots208-138	microsatellite Ots208		Campbell *et al.* (2009)
83	Omy_Ots249-227	microsatellite Ots249		Campbell *et al.* (2009)
84	Omy_oxct-85	3-oxoacid CoA transferase 1a		Unpublished J. DeKoning[3]
85	Omy_PEPA-i6	nonspecific dipeptidase (intron 6)		Aguilar & Garza (2008)
86	Omy_R0917-230	recombination activating gene	***	Sprowles *et al.* (2006)
87	Omy_R1175-137	recombination activating gene		Sprowles *et al.* (2006)
88	Omy_rapd-132	RAPD (OPC-02) marker sequence		Sprowles *et al.* (2006)
89	Omy_rapd-167	RAPD (OPC-02) marker sequence		Sprowles *et al.* (2006)
90	Omy_sSOD-1	superoxide dismutase 1		Brunelli *et al.* (2008)
91	Omy_star-206	steroidogenic acute regulatory protein		Unpublished J. DeKoning[3]
92	Omy_stat3-273	signal transducer/activator of transcription		Unpublished J. DeKoning[3]
93	Omy_tgfb-207	TGF-beta gene	***	Unpublished J. DeKoning[3]
94	Omy_tlr3-377	toll like receptor 3		Unpublished J. DeKoning[3]
95	Omy_tlr5-205	toll like receptor 5		Unpublished J. DeKoning[3]
96	Omy_u07-79-166	unknown gene (from BAC/cDNA clone)		Unpublished S. Young[4]

* hybrid marker

** poor genotype data

*** linked marker

[1] contact: camn@critfc.org

[2] contact: carlos.garza@noaa.gov

[3] contact: dekoning@comcast.net

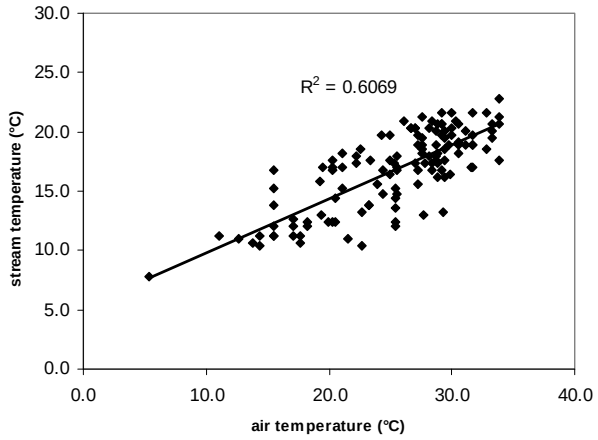
[4] contact: youngsfy@dfw.wa.gov

Supplemental Table S2. List of SNP markers and significantly associated environmental variables identified with SAM analyses.

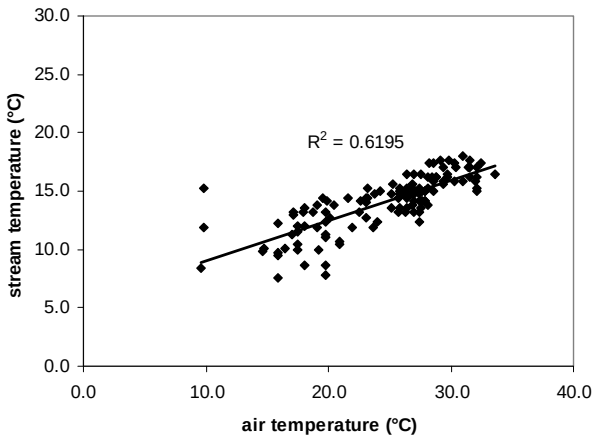
SNP Marker	Associated Environmental Variables				
Omy_113490-159	Conductivity	Fines			
Omy_130295-98	Air Temp.	Elevation			
Omy_130524-160	Precipitation	Elevation	Conductivity		
Omy_97660-230	Precipitation	Air Temp.	Conductivity		
Omy_aldB-165	Air Temp.				
Omy_arp-630	Precipitation	Air Temp.	Conductivity	Cobble/Boulder	
Omy_aspAT-413	Precipitation	Air Temp.	Conductivity		
Omy_cd59-206	Precipitation	Air Temp.	Elevation	Conductivity	
Omy_colla1-525	Precipitation	Conductivity	Fines		
Omy_cox1-221	Air Temp.				
Omy_cox2-335	Precipitation	Air Temp.	Elevation	Conductivity	
Omy_g1-103	Elevation				
Omy_g12-82	Precipitation	Air Temp.	Elevation	Gravel	
Omy_gdh-271	Air Temp.				
Omy_hsf1b-241	Width				
Omy_hsf2-146	Precipitation	Air Temp.	Elevation	Gradient	Conductivity
Omy_IL17-185	Elevation				
Omy_mcsf-268	Air Temp.	Elevation			
Omy_mcsf-371	Air Temp.	Elevation			
Omy_nkef-308	Air Temp.	Conductivity			
Omy_Ogo4-212	Precipitation	Air Temp.	Conductivity		
Omy_stat3-273	Precipitation	Air Temp.	Conductivity	Fines	
Omy_tlr5-205	Precipitation	Air Temp.	Elevation		
Omy_u07-79-166	Precipitation	Air Temp.	Elevation	Conductivity	

Supplemental Figure S1. Correlation of daily air and stream temperatures for each site from May through October, 2009. All relationships were significant with $p < 1.0 \times 10^{-8}$. Stream temperatures for Fawn Cr. were not available.

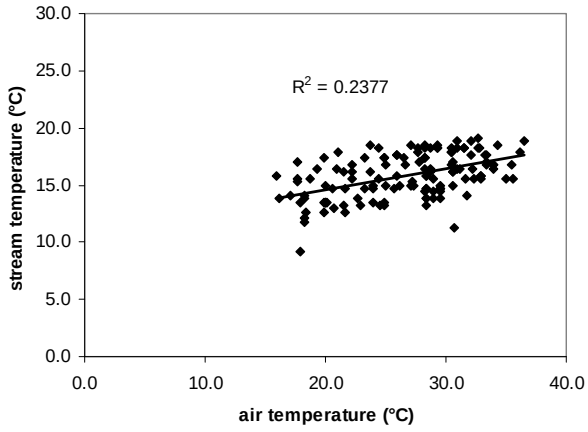
Deer Cr.



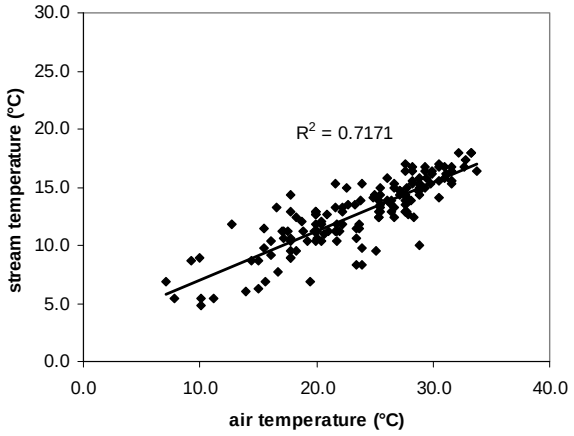
Johnson Cr.



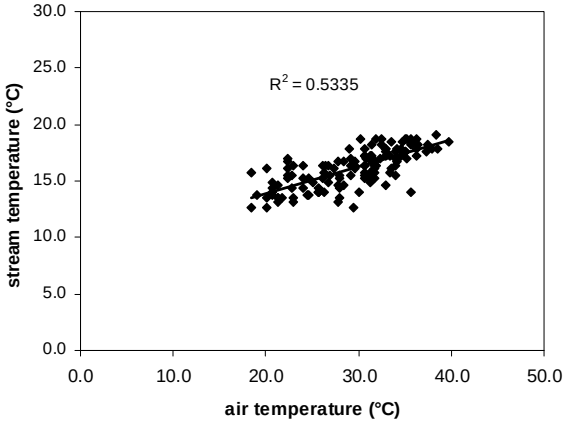
Big Jacks Cr.



upper Manns Cr.

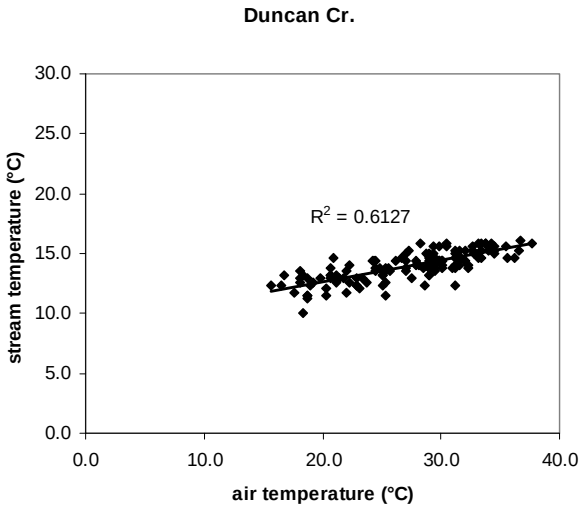


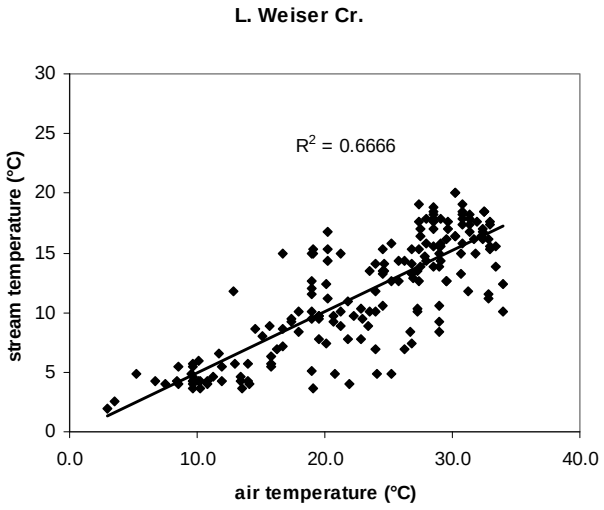
Little Jacks Cr.

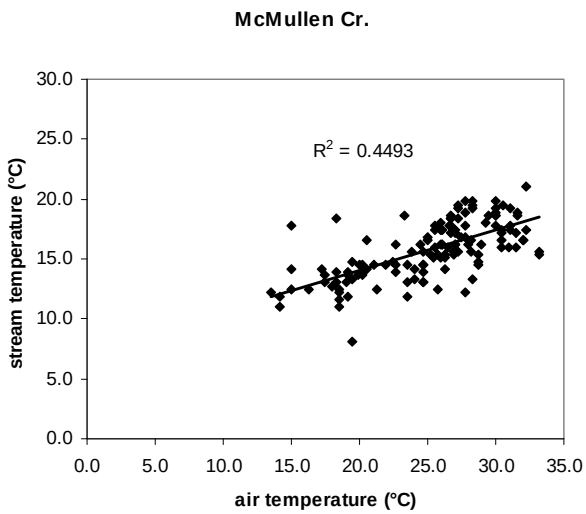


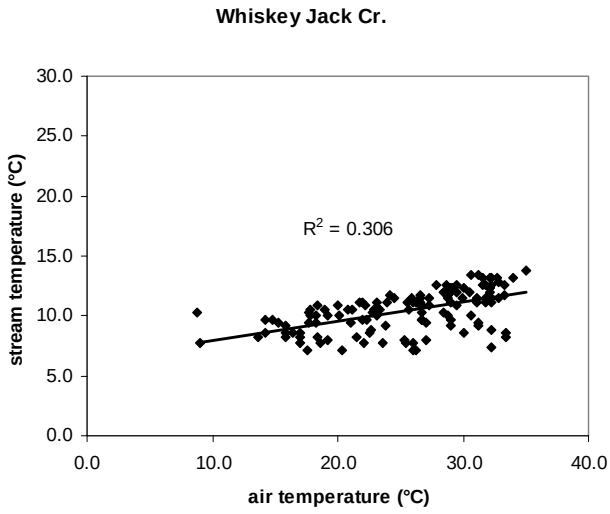
Fawn Cr.

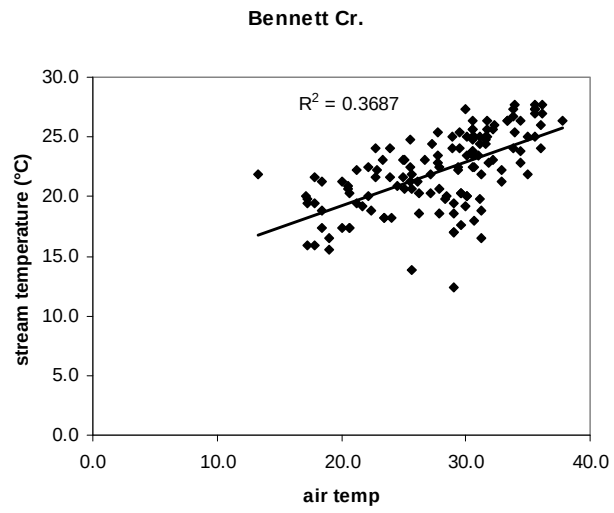
NOT AVAILABLE

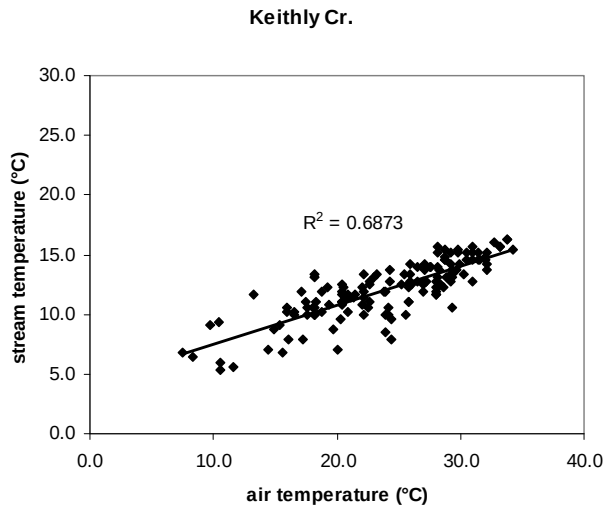

Duncan Cr.
stream temperature (°C)
air temperature (°C)
R2 = 0.6127
30.0
25.0
20.0
15.0
10.0
5.0
0.0
0.0
10.0
20.0
30.0
40.0


L. Weiser Cr.
stream temperature (°C)
air temperature (°C)
R2 = 0.6666
30
25
20
15
10
5
0
0.0
10.0
20.0
30.0
40.0


McMullen Cr.
stream temperature (°C)
air temperature (°C)
R2 = 0.4493
30.0
25.0
20.0
15.0
10.0
5.0
0.0
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0


Whiskey Jack Cr.
stream temperature (°C)
air temperature (°C)
R2 = 0.306
30.0
25.0
20.0
15.0
10.0
5.0
0.0
0.0
10.0
20.0
30.0
40.0


Bennett Cr.
stream temperature (°C)
air temp
R2 = 0.3687
30.0
25.0
20.0
15.0
10.0
5.0
0.0
0.0
10.0
20.0
30.0
40.0


Keithly Cr.
stream temperature (°C)
air temperature (°C)
R2 = 0.6873
30.0
25.0
20.0
15.0
10.0
5.0
0.0
0.0
10.0
20.0
30.0
40.0

Section 2: Anadromy and Smoltification

Introduction

Both environmental and genetic factors determine if individual *O. mykiss* remain as resident rainbow trout, or undergo the necessary physiological changes (smoltification) to prepare for anadromy (Shapovalov and Taft 1954; Ricker 1972; Randall et al. 1987; Peven et al. 1994). While some of the environmental factors that contribute to smoltification such as photoperiod and temperature have been evaluated (Hoar 1976), the genetic mechanisms that contribute to migratory behavior are not well known. Recent studies have confirmed that genetic factors do play a role in smoltification (Thrower et al. 2004; Nichols et al. 2008), and quantitative trait loci (QTL) have been identified that are associated with phenotypic traits of smolts such as silvery appearance and body shape (Nichols et al. 2008). However, further research is needed to extend these studies in order to apply this information toward monitoring and recovery of natural populations of *O. mykiss*.

Smoltification is an important trait of interest since resident rainbow trout and anadromous steelhead life history types (*O. mykiss*) may give rise to one another (e.g., Zimmerman and Reeves 2000; Thrower et al. 2004). Therefore, resident rainbow trout are a potential contributing source in steelhead population abundance, and may be influential in the recovery of ESA listed steelhead stocks; current recovery plans do not typically include resident rainbow trout despite this potential. Although genetic studies have indicated that resident and anadromous life history types commonly interbreed when found in sympatry (e.g., Docker and Heath 2003; Olsen et al. 2005), the ecotypes may persist as genetically distinct breeding populations in geographically proximate locations with differing environments (Narum et al. 2004; Narum et al. 2008).

In this study, we evaluated a panel of 96 single nucleotide polymorphism (SNP) markers from diverse functional genes to test for association with anadromy in populations with distinct ecotypes of *O. mykiss* (i.e., resident or anadromous). We discuss significant candidate markers and their putative gene function, along with a simple multivariate logistic model to identify the potential of individual markers for predicting ecotypes in the Klickitat River, Washington USA, and possibly across broader geographic regions.

Methods

Tissue Samples and Populations

Ten populations of *O. mykiss* from the Klickitat River (Figure 1) that have been previously characterized as either primarily resident or anadromous (Narum et al. 2007; Narum et al. 2008) were used to test for markers associated with anadromy. Briefly, resident or anadromous status of each population was determined by genetic stock identification of adult steelhead returning to the sub-basin over multiple years. Populations with high proportions of steelhead were considered primarily anadromous, and those with zero or low proportion were considered primarily resident (see Narum et al. 2007 for details). These ecotype classifications were further validated in landscape genetics analyses with 13 microsatellite loci that provide evidence for the role that environmental conditions play in structuring resident and anadromous populations in the Klickitat River (Narum et al. 2008). The same fish tissues from Narum et al. (2008) were used to test a panel of SNP markers for association with anadromy in the Klickitat River. For the current study, tests for anadromy were focused on 10 populations in the middle section of the Klickitat River sub-basin that included 6 resident and 4 anadromous populations from a similar geographic area (Figure 1). Eight other collections from the Klickitat River were treated as unknowns and used to test a

simple predictive model for ecotype distinction (Figure 1). However, all collections had *a-priori* information regarding primary life history ecotype from previous studies (Narum et al. 2007; Narum et al. 2008).

SNP Genotyping & Descriptive Statistics:

Tissue samples from each individual were processed with Qiagen DNeasy® kits to extract DNA from fin clips stored in 100% ethanol. Isolated DNA from each sample was genotyped for 96 SNP markers (see Table 1) with Taqman chemistry (Applied Biosystems) and Fluidigm 96.96 dynamic array chips (reaction volumes of ~7nL) for SNP genotyping. Since genotyping in nL reaction volumes reduces the average starting copy number to a range where genotyping accuracy becomes less reliable (Campbell & Narum 2008), a pre-amplification protocol was used to increase the number of starting copies. Pre-amplification occurred in 7µl reactions with 2µl of genomic DNA and 5µl of PCR cocktail (3.5µl of Qiagen Multiplex Mastermix, 0.875µl of 96 pooled primer sets at 0.36uM, and 0.625µl water) under the following thermal cycling program: initial denature at 95°C for 15 minutes, 14 cycles of 95°C for 15 seconds and 60°C for 4 minutes, hold at 4°C. Immediately after cycling, 133µL of nuclease free H_2O or TE buffer was added to each PCR reaction and stored at 4°C.

Pre-amplified template DNA was then genotyped with Fluidigm 96.96 dynamic array chips that included a three step process: 1) SNP assays (Taqman primers/probes) and DNA samples were mixed according to manufacturers protocols and loaded onto the chip with a Fluidigm IFC Controller instrument, 2) target SNPs were amplified for 50 cycles on a Eppendorf thermal cycler specially formatted for the Fluidigm 96.96 chip, and 3) chips were scanned with a Fluidigm EP-1 instrument to detect fluorescently labeled allele-specific probes. Genotypes for each assay were auto-scored with Fluidgim SNP Analysis v.2.1.1 software and verified by visual inspection with scoring guides provided by an assay database and a heterozygous indicator sample for each SNP. Since the SNP markers used in this study were ascertained from a broad panel of samples including related populations from the Columbia River (e.g., Campbell et al. 2008), ascertainment bias should be limited. Any potential ascertainment bias should effect populations in this study equally since none were directly included in SNP discovery.

Tests for linkage disequilibrium (LD) between all pairs of loci were performed using the MCMC approximation of the exact test in GENEPOP v. 3.3 (Raymond & Rousset 1995). Because multiple comparisons were involved and power was high with these tests, a Bonferroni correction was used to protect against Type I error.

A total of 96 SNPs were screened in this study, but 11 markers were dropped from further statistical analyses (see Table 1 for list). Two SNPs in the panel were used to detect potential hybrids of *O. mykiss* and *O. clarki* (cutthroat trout), but dropped from further analysis since hybrids were not identified in this study. One SNP was a sex marker (Brunelli et al. 2008) and only used to test for sex-linkage with other loci. Eight pairs of SNPs had highly significant LD ($p < 0.0001$), and the least informative of each pair was dropped from further analyses. A total of 85 remaining SNPs were included in subsequent statistical tests (Table 1).

Deviation from Hardy-Weinberg equilibrium (HWE) was evaluated at each locus and population with the Markov Chain Monte Carlo (MCMC) approximation of Fisher's exact test implemented in GENEPOP v. 3.3 (1000 batches with 1000 iterations; Raymond & Rousset 1995). Minor allele frequency (MAF) was estimated for each SNP in each collection with GENEPOP.

Tests for SNP Markers Associated with Anadromy

Since tests for outlier loci can often provide false positives or false negatives (Akey 2009; Hermisson 2009), a three step process was used to identify candidate markers for anadromy in this study: step 1) perform tests to identify initial candidate markers after correcting for underlying population structure, step 2) perform outlier tests to eliminate loci that were associated with selective environmental variables (not necessarily anadromy), and step 3) perform univariate logistic regression of SNP allele frequencies on anadromy. Only those markers that were statistically significant for all three criteria were considered candidate makers associated with anadromy. Additional markers were identified that may be under selective pressure in differing environments; these were not evaluated further in tests of anadromy.

In order to limit association bias due to underlying population structure in tests for candidates (step 1), analyses with STRUCTURE v.2.3.2 (Pritchard *et al.* 2000a; Hubisz *et al.* 2009) and STRAT v.1.1 (Pritchard *et al.* 2000b) were implemented as suggested by Pritchard and Rosenberg (1999). In a recent review by Zhang *et al.* (2008), this approach has been shown to account for genetic structure in association studies equally well as other methods such as principal components analysis (i.e., Price *et al.* 2006). We followed the procedure for running STRAT with the following steps: 1) neutral microsatellite loci from Narum et al. (2008a) were used to run STRUCTURE and select the most likely number of distinct populations ("k") from 10 iterations for each potential k value ranging from 1-10 (50,000 burnin followed by 100,000 MCMC repetitions), 2) ancestry coefficients (mean Q values from CLUMPP; Jakobsson & Rosenberg 2007) for each individual were used to correct for underlying population structure, 3) multilocus genotypes from 85 SNP markers and primary life history of each individual were included in STRAT to test candidate markers for significant association with anadromy.

In order to identify SNP markers that were associated with environmental variables (step 2), we used the Spatial Analysis Method (SAM) developed by Joost et al. (2008). This approach utilizes general linear models and logistic regression to identify significant associations of habitat characteristics with presence/absence of alleles at genetic markers across all individuals in the study. In the current data set, one allele for each SNP was recoded as present ("1") or absent ("0") for all individual fish. Missing genotype data were recoded with the non-numerical designation of "NaN". Six environmental variables were included as described in Narum et al. (2008a): elevation (m), mean daily high temperature (C), mean annual precipitation (cm), stream gradient-below (3km downstream of each site), stream gradient-general (500m below to 500m above each site), and upstream distance (km) from the mouth of the Klickitat River. Due to the high power of the SAM test to identify associations, a conservative alpha of 0.01 with Bonferroni correction (final critical value of 0.00196) were used to reduce false positives in the Wald test (as recommended by Joost et al. 2008). Markers with *p*-values below the critical value were considered candidates for association with respective environmental variables.

The third step was initiated to further reduce the risk of false positives and to provide a more stringent test for association of markers. All markers still considered as candidates after accounting for association bias using STRUCTURE/STRAT that were also not shown to be associated with environmental variables, were finally evaluated with univariate logistic regression at the 10 collection sites. Minor allele frequency (MAF) for each marker was plotted against resident or anadromous life history and fitted to the following logistic regression curve using KaleidaGraph v4.0 (Synergy Software): $y = a + (b - a) / (1 + (x / c)^d)$, where a = y max, b = y min, c = x at 50% y, and d = slope. Candidate markers that passed the first two criteria and were also found to be significant in logistic regression tests ($p < 0.05$)

were considered to be candidate markers for anadromy. This multivariate model was tested on all 18 collections identified in Figure 1, with 10 of the collections used to identify candidate markers associated with anadromy and 8 collections treated as unknown ecotype (but all had *a-priori* information regarding primary ecotype).

Results

Of 850 tests for deviation from HWE (85 SNPs in 10 populations), there were 10 significant results that included 5 heterozygote deficits and 5 heterozygote excess. Deviations were distributed randomly across loci and populations, with 1 excess in Dead Canyon Cr. (Omy_gluR-79), 1 excess in lower White Cr. (Omy_Ots208-138), 1 deficit in lower Summit Cr. (Omy_121713-115), 1 deficit in lower Trout Cr. (Omy_96222-125), 2 deficits in Tepee Cr. (Omy_Ots208-138 and Omy_tlr5-205), 1 excess in upper White Cr. (Omy_Ots208-138), 1 deficit in upper Summit Cr. (Omy_cox1-221), and 2 excess in Brush Cr. (Omy_113490-159 and Omy_hsp47-86). Since no consistent trends in deviations were observed among loci or populations, the markers and populations did not appear to violate null assumptions of HWE. None of the 85 SNPs showed significant linkage disequilibrium with sex.

In the intial step to identify candidate markers, analyses with STRUCTURE and STRAT were applied using microsatellite (neutral) markers to identify underlying population structure, and then all SNP markers were tested for association with anadromy. Results from STRUCTURE provided support for k = 6 as determined by criteria in Evanno *et al.* (2005). Mean ancestry coefficient (Q) at k = 6 was included as input for STRAT analysis to correct for population structure, and results indicated there were 12 SNPs significantly association with anadromy (Table 2). These 12 markers were considered initial candidate markers.

Analysis with SAM provided a total of 19 markers that were significant for one or more environmental variable for a total of 42 significant associations (Table 3). Each of the six environmental variables were significant with at least one marker (Table 3), but temperature had the greatest number of associated markers (11), followed by elevation (10), upstream distance (8), precipitation and stream gradient-below (5 each), and stream gradient-general (3). Of the initial 12 candidate markers identified in STRAT, 6 of these also showed a significant association with environmental variables and thus were not considered candidates strictly for anadromy. The remaining 6 SNPs (Omy_130524-160, Omy_CD28-130, Omy_hsp90BA-229, Omy_IL6-320, Omy_LDHB-2_i6, and Omy_ndk-152) remained under consideration as candidate markers for anadromy, and were tested with the third and final verification step. Of these, three markers had allele frequencies significantly correlated with anadromy in logistic regression (Figure 1). Thus, three markers (Omy_IL6-320, Omy_LDHB-2_i6, and Omy_ndk-152) were verified under our protocol as candidate markers associated with anadromy in *O. mykiss* of the Klickitat River.

A simple multiple regression model was developed from allele frequencies of the three validated candidate markers to predict anadromy. The resulting model was:

Logit P = -27.686 + (53.667 * MAF Omy_ LDHB-2_i6) + (71.727 * MAF Omy_ndk-152) + (52.691 * MAF Omy_IL6-320)

where MAF equals the minor allele frequency of the given marker. The model was also applied to eight additional collections of *O. mykiss* from the upper and lower regions of the Klickitat River. Results of the model were observed to be consistent with previous determinations of residency or anadromy (Narum et al. 2007; Narum et al. 2008), but two populations appeared intermediate between ecotype categories (Snyder Cr. and upper Klickitat River; Table 4).

Discussion

Smoltification is a complex trait with many potential contributing factors. In concert these factors dictate whether an individual fish remains in freshwater as a resident or undergoes anadromous migration to seawater, which are alternate fates with profound influence on survival and productivity. Environmental cues such as photoperiod and temperature have been shown to play a key role in smoltification (Zaugg and Wagner 1973; Hoar 1976; Peven 1994), indicating phenotypic plasticity in this trait. However, breeding studies of resident and anadromous ecotypes of *O. mykiss* have also shown a substantial genetic basis in plasticity among family lines (Thrower et al. 2004). The genetic basis of smoltification in *O. mykiss* has been further evaluated with evidence for at least one major QTL (quantitative trait locus) region, and multiple linkage groups involved in smoltification (Nichols et al. 2008).

In this study, we identified three SNPs that were associated with anadromy, and additional markers that may be under selection in *O. mykiss* from differing environments. The three markers associated with anadromy were not significantly linked, and were from genes with varying functions including interleukin 6 (Omy_IL6-320), lactate dehydrogenase (Omy_ LDHB-2_i6), and nucleoside diphosphate kinase (Omy_ndk-152). Putative gene function for these three genes were intriguing since nucleoside diphosphate kinase and lactate dehydrogenase are metabolic genes, and interleukin 6 is an immune response gene. Metabolic genes in particular could influence migration and sexual maturity in fish (e.g., Leonard and McCormick 1999), and interleukin 6 is a mediator of inflammatory response in muscle (Iliev et al. 2007). However, it is possible that these may not be the actual genes responsible for variation in anadromy observed in this study, but rather markers closely linked with other genes. This highlights the need for further verification of these candidate markers in common garden studies to test genotype and phenotypic response, and the need for linkage mapping with fine scale QTL analyses.

Interestingly, our study included markers from genes that had *a-priori* potential to be associated with anadromy, but were not significant candidates in any phase of our analyses for candidate markers. This included markers from Na^+/K^+-ATPase alpha 3 and p53 tumor suppessor genes that are believed to be involved in osmoregulation and smoltification, respectively (Lee et al. 1998; Nichols et al. 2008). However, only the alpha 1a and 1b isoforms of Na^+/K^+-ATPase are found to be differentially regulated in rainbow trout in salinity challenges (Richards et al. 2003), and thus the alpha 3 isoform may not be directly involved in osmoregulation or the process of smoltification. Similarly, Nichols et al. (2008) had identified p53 as a gene in the major QTL region for smoltification (i.e., linkage group OC20 in Nichols et al. 2008), but we found no evidence for association with anadromy for this marker. If p53 is indeed involved in smoltification, the non-significant finding in this study may be due to the genome position (3' UTR) of the SNP marker relative to coding regions. However, other studies (M. McPhee, unpublished data) have noted lack of association with anadromy with several markers from the major QTL region for smoltification identified by Nichols et al. (2008). It also possible that the evolutionary advantage for retaining allelic variation at these loci is strong enough that directional selection does not occur in resident populations. Retention of life history plasticity and the ability to smolt carries across generations and therefore the genetic component involved in osmoregulation must also carry or remain latent.

Several additional SNP markers in this study showed significant association with environmental variables regardless of ecotype, but these genes could play an indirect role in smoltification. In particular, several SNPs appeared to be possible candidates related to

temperature (Table 3) which is expected to be a major environmental factor involved in smoltification and anadromy (e.g., Peven 1994). In fact, landscape genetics analyses indicate that temperature is the primary environmental factor involved in structuring resident and anadromous populations of *O. mykiss* in the Klickitat River (Narum et al. 2008). Environmental influence plays a large role in smoltification, and may also regulate gene expression of loci contributing to this trait.

The multivariate model developed in this study appears to be a useful tool for determining the potential for anadromy in populations of *O. mykiss* in the Klickitat River. Results from the model accurately predicted primary ecotypes for the additional populations, and identified two collections with intermediate levels of anadromy (Snyder Cr. and the upper mainstem Klickitat River; Table 4). These two intermediate collection sites have been previously noted as either populations that appear to receive small reproductive contributions from anadromous fish, or were collected in areas with high potential for anadromy (Narum et al. 2008). While the model was useful in the Klickitat River sub-basin, further study is necessary to validate this finding and determine if it applies on a broader geographic scale. This includes evaluation of these candidate SNPs in differing ecotypes of *O. mykiss* throughout the Columbia River Basin and other parts of the species range. Since *O. mykiss* from the Klickitat River appear to be in a transition zone between genetically distinct coastal and inland lineages of the Columbia River (Blankenship et al. *submitted*), these candidate loci will need extensive testing in collections comprised of a single lineage (interior or coastal). Common garden studies with resident and anadromous strains of fish reared under controlled environmental conditions would also allow for testing of individual smoltification phenotypes and genotypes at candidate markers.

References

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

Akey, J.M. 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Research 19:711-722.

Brunelli, J.P., K.J. Wertzler, K. Sundin, and G.H. Thorgaard. 2008. Y-specific sequences and polymorphisms in rainbow trout and Chinook salmon. Genome 51:739-748.

Campbell, N.R., and S.R. Narum. 2009a. Quantitative PCR assessment of microsatellite and SNP genotyping with variable quality DNA extracts. Conservation Genetics 10:779-784.

Campbell, N.R., and S.R. Narum. 2009b. Identification and characterization of heat shock response related single nucleotide polymorphisms in *O. mykiss* and *O. tschawytscha*. Molecular Ecology Resources 9:1460-1559.

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

Docker, M.F., and D.D. Heath. 2003. Genetic comparison between sympatric anadromous steelhead and freshwater resident rainbow trout in British Columbia, Canada. Conservation Genetics 4:227–231.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-20.

Hermisson, J. 2009. Who believes in whole-genome scans for selection? Heredity 103:283-284.

Hoar, W.S. 1976. Smolt transformation: evolution, behavior, and physiology. Journal of the Fisheries Research Board of Canada 33:1234-1252.

Hubisz, M.J., D. Falush, M. Stephens, and J.K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322-1332.

Iliev, D.B., B. Castellana, S. MacKenzie, J.V. Planas, and F.W. Goetz. 2007. Cloning and expression analysis of an IL-6 homolog in rainbow trout (*Oncorhynchus mykiss*). Molecular Immunology 44:1803-1807.

Jakobsson, M., N.A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806.

Joost, S., M. Kalbermatten, and A. Bonin. 2008. Spatial Analysis Method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection, Molecular Ecology Resources 8:957–960.

Lee, T.H., J.C. Tsai, M.J. Fang, M.J. Yu, and P.P. Hwang. 1998. Isoform expression of Na+-K+-ATPase a-subunit in gills of the teleost *Oreochromis mossambicus*. American Journal of Physiology 275:R926-R932.

Leonard, J.B.K., and S.D. McCormick. 1999. The effect of migration distance and timing on metabolic enzyme activity in an anadromous Clupeid, the American shad (*Alosa sapidissima*). Fish Physiology and Biochemistry 20:163-179.

Narum, S.R., C. Contor, A. Talbot, and M. Powell. 2004. Genetic divergence of sympatric resident and anadromous forms of *Oncorhynchus mykiss* in the Walla Walla River and Columbia River Basin, USA. Journal of Fish Biology 65:471-488.

Narum, S.R., J.S. Zendt, C. Frederiksen, and W.R. Sharp. 2007. Substructure of native steelhead in the Klickitat River and genetic ID of spawning populations. Pages 294-295 *in* R. F. Carline and C. LoSapio, editors. Wild Trout IX: Sustaining wild trout in a changing world. Wild Trout Symposium, Bozeman, Montana.

Narum, S. R., J. Zendt, D. Graves, and B. Sharp. 2008. Influence of landscape on resident and anadromous life history types of *Oncorhynchus mykiss*. Canadian Journal of Fisheries and Aquatic Sciences 65:1013-1023.

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

Olsen, J.B., K. Wuttig, D. Fleming, E.J. Kretschmer, and J.K. Wenburg. 2005. Evidence of partial anadromy and resident-form dispersal bias on a fine scale in populations of *Oncorhynchus mykiss*. Conservation Genetics 7:613-619.

Peven, C.M., R.R. Whitney, and K.R. Williams. 1994. Age and length of steelhead smolts from the mid-Columbia River basin, Washington. North American Journal of Fisheries Management 14:77-86.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38:904–909.

Pritchard, J.K., and N.A. Rosenberg. 1999. Use of unlinked genetic markers to detect population stratification in association studies. American Journal of Human Genetics 65:220-228.

Pritchard, J., M. Stephens, P. Donnelly. 2000a. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pritchard, J.K., M. Stephens, N.A. Rosenberg, and P. Donnelly. 2000b. Association mapping in structured populations. American Journal of Human Genetics 67:170-181.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Randall, R.G., M.C. Healy, and J.B. Dempson. 1987. Variability in length of freshwater residence of salmon, trout, and char. American Fisheries Society Symposium 1:27-41.

Richards, J.G., J.W. Semple, J.S. Bystriansky, and P.M. Schulte. 2003. Na/K-ATPase a-isoform switching in gills of rainbow trout (*Oncorhynchus mykiss*) during salinity transfer. Journal of Experimental Biology 206:4475-4486.

Ricker, W. E. 1972. Hereditary and environmental factors affecting certain salmonid populations. In: R. C. Simon and P. A. Larkin (editors), The stock concept in Pacific salmon, p. 19-160. H.R. MacMillan Lectures in Fisheries, Univ. of British Columbia, Vancouver, B.C.

Shapovalov, L., and A. C. Taft. 1954. The life histories of the steelhead rainbow trout (*Salmo gairdneri gairdneri*) and silver salmon (*Oncorhynchus kisutch*) with special reference to Waddell Creek, California, and recommendations regarding their management. DFG Bulletin No. 98.

Sprowles, A.E., M.R. Stephens, N.W. Clipperton, and B.P. May. 2006. Fishing for SNPs: A targeted locus approach for single nucleotide polymorphism discovery in rainbow trout. Transactions of the American Fisheries Society 135:1698-1721.

Thrower, F. P., J. J. Hard and J. E. Joyce. 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.

Zaugg, W.S., and H.H. Wagner. 1973. Gill ATPase activity related to parr-smolt transformation and migration in steelhead trout (*Salmo gairdneri*): influence of photoperiod and temperature. Comparative Biochemistry and Physiology B 45:955–965.

Zhang, F., Y. Wang, H. Deng. 2008. Comparison of population-based association study methods correcting for population stratification. PLoS One doi:10.1371/journal.pone.0003392

Zimmerman, C.E., and G.H. Reeves. 2000. Population structure of sympatric anadromous and non-anadromous *Oncorhynchus mykiss*: evidence from spawning surveys and otolith microchemistry. Canadian Journal of Fisheries and Aquatic Sciences 57:2152-2162.

Table 1. List of 96 SNP markers assayed in *O. mykiss* from the Klickitat River.

	SNP Marker Name	Putative Gene Name	Removed	Reference
1	Omy_SEXY1	Sex determination assay	*	Unpublished M. Campbell[1]
2	Omy_myclarp404-111	acidic ribosomal phosphoprotein	**	Unpublished N. Campbell [2]
3	Omy_Omyclmk436-96	map kinase 4	**	Unpublished N. Campbell [2]
4	Omy_113490-159	unknown gene (expressed)		Unpublished C. Garza[3]
5	Omy_114315-438	unknown gene (expressed)		Unpublished C. Garza[3]
6	Omy_121006-131	immediate early response protein 5		Unpublished C. Garza[3]
7	Omy_121713-115	similar to vertibrate polymerase		Unpublished C. Garza[3]
8	Omy_123044-128	unknown gene (expressed)		Unpublished C. Garza[3]
9	Omy_123048-119	40S ribosomal protein S26		Unpublished C. Garza[3]
10	Omy_127236-583	A312aR protein		Unpublished C. Garza[3]
11	Omy_128693-455	Stathmin-like 4		Unpublished C. Garza[3]
12	Omy_130295-98	type II keratin K8b	***	Unpublished C. Garza[3]
13	Omy_130524-160	Growth arrest & DNA-damage-inducible prot. alpha		Unpublished C. Garza[3]
14	Omy_187760-385	unknown gene (expressed)		Unpublished C. Garza[3]
15	Omy_95489-239	unknown gene (expressed)		Unpublished C. Garza[3]
16	Omy_96222-125	elongation factor 1-alpha		Unpublished C. Garza[3]
17	Omy_97077-73	Anti-proliferative cofactor		Unpublished C. Garza[3]
18	Omy_97660-230	unknown gene (expressed)		Unpublished C. Garza[3]
19	Omy_97865-196	60S riposomal protein L37a		Unpublished C. Garza[3]
20	Omy_97954-618	unknown gene (expressed)		Unpublished C. Garza[3]
21	Omy_aldB-165	aldolase B	***	Campbell et al. (2009)
22	Omy_aldB-414	aldolase B		Campbell et al. (2009)
23	Omy_ALDOA_1	aldolase A		Aguilar and Garza (2008)
24	Omy_aromat-280	aromatase		Unpublished J. Dekoning[4]
25	Omy_arp-630	acidic ribosomal phosphoprotein		Campbell et al. (2009)
26	Omy_aspAT-123	aspartate aminotransferase		Campbell et al. (2009)
27	Omy_aspAT-413	aspartate aminotransferase		Campbell et al. (2009)
28	Omy_b1-266	BAC clone sequence		Sprowles et al. (2006)
29	Omy_b9-164	BAC clone sequence		Sprowles et al. (2006)
30	Omy_BAC-B4-126	from BAC clone	***	Unpublished S. Young[5]
31	Omy_BAC-B4-324	from BAC clone		Unpublished S. Young[5]
32	Omy_cd28-130	CD28 like protein		Unpublished J. Dekoning[4]
33	Omy_cd59-206	CD59 like protein 1		Unpublished J. Dekoning[4]
34	Omy_cd59b-112	CD59 like protein 2		Unpublished J. Dekoning[4]
35	Omy_colla1-525	alpha 1 type 1 collagen		Unpublished J. Dekoning[4]
36	Omy_cox1-221	cyclooxygenase 1		Campbell et al. (2009)
37	Omy_cox2-335	cyclooxygenase 2		Unpublished J. Dekoning[4]
38	Omy_crb-106	carbonyl reductase - intron 3		Sprowles et al. (2006)
39	Omy_CRBF1-1	carbonyl reductase - intron 1	***	Aguilar and Garza (2008)
40	Omy_cxcr-169	CXC chemokine receptor		Unpublished J. Dekoning[4]
41	Omy_dacd1-131	defender against cell death 1	***	Unpublished N. Campbell [2]
42	Omy_e1-147	BAC clone sequence		Sprowles et al. (2006)
43	Omy_g1-103	glucose-6 phosphate dehydrogenase		Sprowles et al. (2006)
44	Omy_g12-82	cDNA clone 1RT158N23_B_G12		Unpublished J. Dekoning[4]
45	Omy_gadd45-332	gadd45 beta ii		Unpublished N. Campbell [2]
46	Omy_gdh-271	glutamate dehydrogenase		Campbell et al. (2009)
47	Omy_gh-334	growth hormone 1		Unpublished N. Campbell [2]
48	Omy_gh-475	growth hormone 1		Unpublished N. Campbell [2]
49	Omy_GHSR-121	growth hormone secretegogue receptor		Unpublished N. Campbell [2]
50	Omy_gluR-79	glucocorticoid receptor		Unpublished N. Campbell [2]
51	Omy_hsc715-80	heat shock cognate 71		Campbell and Narum (2009b)

52	Omy_hsf1b-241	heat shock factor 1b		Campbell and Narum (2009b)
53	Omy_hsf2-146	heat shock factor 2		Campbell and Narum (2009b)
54	Omy_hsp47-86	heat shock protein 47		Campbell and Narum (2009b)
55	Omy_hsp70aPro-329	heat shock protein 70a (promoter)		Campbell and Narum (2009b)
56	Omy_hsp90BA-193	heat shock protein 90BA	***	Campbell and Narum (2009b)
57	Omy_hsp90BA-229	heat shock protein 90BA		Campbell and Narum (2009b)
58	Omy_IL17-185	interleukin 17		Unpublished J. Dekoning[4]
59	Omy_IL1b-163	interleukin 1b		Unpublished J. Dekoning[4]
60	Omy_IL6-320	interleukin 6		Unpublished J. Dekoning[4]
61	Omy_inos-97	inducible nitric oxide synthase		Unpublished J. Dekoning[4]
62	Omy_LDHB-1_i2	lactate dehydrogenase		Aguilar and Garza (2008)
63	Omy_LDHB-2_e5	lactate dehydrogenase		Aguilar and Garza (2008)
64	Omy_LDHB-2_i6	lactate dehydrogenase		Aguilar and Garza (2008)
65	Omy_mapK3-103	map kinase 4		Unpublished N. Campbell [2]
66	Omy_mcsf-268	macrophage colony-stimulating factor receptor		Unpublished J. Dekoning[4]
67	Omy_mcsf-371	macrophage colony-stimulating factor receptor		Unpublished J. Dekoning[4]
68	Omy_metA-161	metallothionein A gene		Unpublished N. Campbell [2]
69	Omy_metB-138	metallothionein B gene		Unpublished N. Campbell [2]
70	Omy_myoD-178	myosin D		Campbell et al. (2009)
71	Omy_nach-200	voltage-gated sodium channel alpha type IV		Unpublished J. Dekoning[4]
72	Omy_NaKATPa3-50	Na/K ATPase alpha 3		Campbell et al. (2009)
73	Omy_ndk-152	nucleoside diphosphate kinase		Unpublished N. Campbell [2]
74	Omy_nkef-241	natural killer efficiency factor	***	Campbell et al. (2009)
75	Omy_nkef-308	natural killer efficiency factor		Campbell et al. (2009)
76	Omy_nramp-146	natural resistance associated macrophage protein		Campbell et al. (2009)
77	Omy_Ogo4-212	microsatellite Ogo4		Campbell et al. (2009)
78	Omy_OmyP9-180	P9-B genomic sequence		Sprowles et al. (2006)
79	Omy_Ots208-138	microsatellite Ots208		Campbell et al. (2009)
80	Omy_Ots249-227	microsatellite Ots249		Campbell et al. (2009)
81	Omy_oxct-85	3-oxoacid CoA transferase 1a		Unpublished J. Dekoning[4]
82	Omy_p53-262	Tumor protein p53		Unpublished N. Campbell [2]
83	Omy_pad-196	peptidylarginine deiminase		Unpublished N. Campbell [2]
84	Omy_PEPA-i6	nonspecific dipeptidase (intron 6)		Aguilar and Garza (2008)
85	Omy_R0917-230	recombination activating gene		Sprowles et al. (2006)
86	Omy_R1175-137	recombination activating gene		Sprowles et al. (2006)
87	Omy_rapd-132	RAPD (OPC-02) marker sequence		Sprowles et al. (2006)
88	Omy_rapd-167	RAPD (OPC-02) marker sequence		Sprowles et al. (2006)
89	Omy_SECC22b-88	Vesicle-trafficking protein SEC22b-B		Unpublished N. Campbell [2]
90	Omy_sSOD-1	superoxide dismutase 1		Brunelli et al. (2008)
91	Omy_star-206	steroidogenic acute regulatory protein		Unpublished J. Dekoning[4]
92	Omy_stat3-273	signal transducer/activator of transcription		Unpublished J. Dekoning[4]
93	Omy_tgfb-207	TGF-beta gene	***	Unpublished J. Dekoning[4]
94	Omy_tlr3-377	toll like receptor 3		Unpublished J. Dekoning[4]
95	Omy_tlr5-205	toll like receptor 5		Unpublished J. Dekoning[4]
96	Omy_u07-79-166	unknown gene (from BAC/cDNA clone)		Unpublished S. Young[5]

* sex marker

** hybrid identification marker (*O. clarki* and *O. mykiss*)

*** linked marker

[1] contact: matthew.campbell@idfg.idaho.gov

[2] contact: camn@critfc.org

[3] contact: carlos.garza@noaa.gov

[4] contact: dekoning@comcast.net

[5] contact: youngsfy@dfw.wa.gov

Table 2. Initial list of candidate markers for anadromy identified with STRAT (Pritchard and Rosenberg 1999). SNP number matches Table 1.

No.	SNP Marker	Putative Gene Name	Significance
7	Omy_121713-115	similar to vertibrate polymerase	**
13	Omy_130524-160	type II keratin K8b	**
32	Omy_cd28-130	cell death 28 like protein	**
33	Omy_cd59-206	cell death 59 like protein 1	**
49	Omy_GHSR-121	growth hormone secretegogue receptor	***
57	Omy_hsp90BA-229	heat shock protein 90BA	**
60	Omy_IL6-320	interleukin 6	***
64	Omy_LDHB-2_i6	lactate dehydrogenase	***
68	Omy_metA-161	metallothionein A gene	**
69	Omy_metB-138	metallothionein B gene	**
73	Omy_ndk-152	nucleoside diphosphate kinase	**
90	Omy_sSOD-1	superoxide dismutase 1	**

* $p < 0.01$

*** $p < 0.001$

Table 3. List of SNP markers and significantly associated environmental variables identified with SAM analyses.

SNP Marker	Associated Environmental Variables			
Omy_121713-115	Elevation			
Omy_aromat-280	GenGradient			
Omy_arp-630	Temperature	Elevation		
Omy_aspAT-123	GenGradient			
Omy_b1-266	Temperature	Elevation	UpDistance	
Omy_cd59-206	Temperature	Elevation	UpDistance	Precipitation
Omy_e1-147	Temperature			
Omy_g1-103	Temperature	UpDistance	Precipitation	
Omy_gh-475	BelGradient			
Omy_GHSR-121	Temperature	UpDistance		
Omy_IL17-185	GenGradient			
Omy_IL1b-163	Temperature	Elevation	UpDistance	Precipitation
Omy_mcsf-268	Temperature	Elevation	UpDistance	
Omy_metA-161	UpDistance	Precipitation		
Omy_metB-138	Temperature	UpDistance		
Omy_OmyP9-180	Elevation			
Omy_Ots208-138	Temperature	Elevation		
Omy_sSOD-1	Elevation	Precipitation		
Omy_tlr3-377	Temperature	Elevation		

Table 4. Comparison of ecotype designation from *a-priori* classification versus prediction with multivariate logistic regression based on allele frequency of three candidate SNP markers.

Map Code	Population	a-priori ecotype**	model prediction
1	Dead Canyon Cr.	anadromous	anadromous
2	lower White Cr.	anadromous	anadromous
3	lower Trout Cr.	anadromous	anadromous
4	lower Summit Cr.	anadromous	anadromous
5	Teepee Cr.	resident	resident
6	upper White Cr.	resident	resident
7	Brush Cr.	resident	resident
8	upper Summit Cr.	resident	resident
9	upper Trout Cr.	resident	resident
10	Surveyors Cr.	resident	resident
11	Snyder Cr.*	resident	intermediate
12	Swale Cr.*	anadromous	anadromous
13	lower Little Klickitat R.*	anadromous	anadromous
14	Bowman Cr.*	anadromous	anadromous
15	upper Little Klickitat R.*	resident	resident
16	upper mainstem Klickitat R.*	resident	intermediate
17	Piscoe Cr.*	resident	resident
18	Diamond Fork*	resident	resident

* populations below line treated as unknown ecotype in application of the model to test prediction
** as determined in previous studies (Narum et al. 2007; Narum et al. 2008)

Figure 1. Map of the Klickitat River sub-basin with sample collection sites for *O. mykiss*. The dashed oval identifies the 10 collections treated as known ecotypes identified in previous studies (Narum et al. 2007; Narum et al. 2008), with sites outside the oval treated as unknown ecotype. Symbols for each site represent primary ecotype for each collection with resident (circles), anadromous (squares), or unknown (triangle).

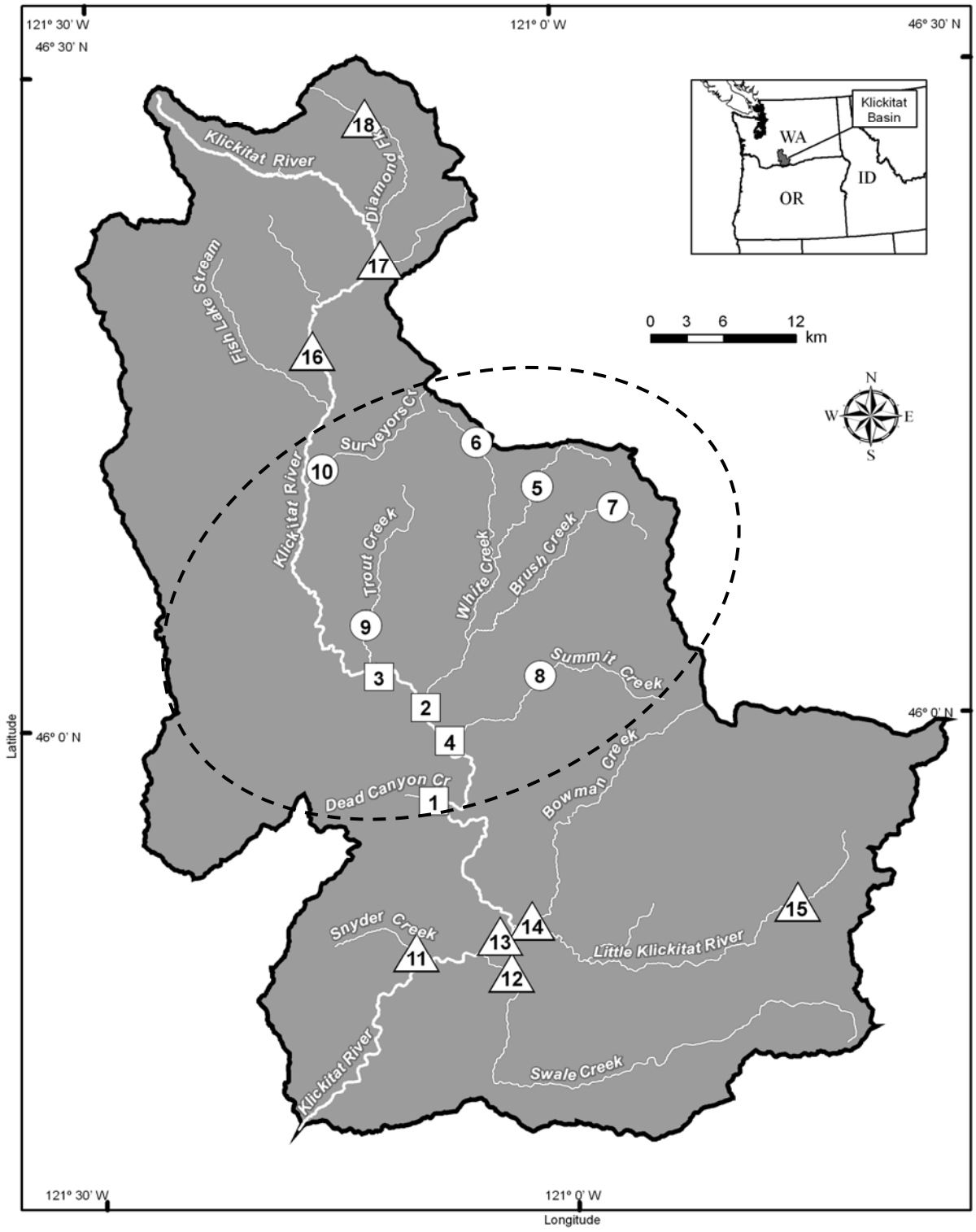


Figure 2. Logistic regression of minor allele frequency (MAF) of three candidate markers on life history ecotype. Six resident populations were coded "0" and four anadromous were coded "1".