SHOULD I STAY OR SHOULD I GO? ANALYZING THE GENETIC BASIS OF MIGRATION-RELATED TRAITS IN RAINBOW TROUT (*ONCORHYNCHUS MYKISS*).

by

SPENCER YAEL WEINSTEIN

Bachelor of Arts, 2016

Bachelor of Science, 2016

University of Massachusetts, Amherst

Amherst, Massachusetts

Submitted to the Graduate Faculty

of the College of Science and Engineering

Texas Christian University

in partial fulfillment of the requirements

for the degree of

Master of Science

May 2018

This page is intentionally left blank.

ACKNOWLEDGEMENTS

There are people both within and outside of TCU without whom my project would not have been possible. First and foremost, I have to thank my advisor, Dr. Matt Hale. Among other things, I appreciate the balance that you struck between answering my questions and pushing me to answer them on my own. I've definitely learned a lot in my time working with you. Thank you as well to my other two committee members, Dr. Amanda Hale and Dr. Dean Williams. Dean helped answer some data and genetic analysis questions throughout my project, and Amanda provided comments and suggestions on my written thesis.

My project overall was possible because of the years of research that Frank Thrower and Krista Nichols have done in the Sashin system; they, among others, have really set the stage to allow for the research presented within. Garrett McKinney provided invaluable assistance with many of the statistical programs that I used in the analysis of my data, and I am grateful for his help as well.

Finally, I have to thank my family and friends who have supported me and helped keep me sane. You've made Texas feel a little less far away.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Migration, the movement of individuals between different habitats, is an evolutionary life history strategy that is widespread across animal taxa (Dingle and Drake 2007). Unlike dispersal, which is a permanent movement to a novel habitat (Howard 1960, Brönmark et al. 2013), migration occurs seasonally with individuals moving for increased access to resources before returning to their natal habitats to breed (Greenwood 1980, Brönmark et al. 2013). Taxa- and population-specific migration can take one of many forms (Dingle and Drake 2007), including occurring uniformly within species or with populations exhibiting partial migration, a phenomenon seen in birds and fish in which some individuals in the population migrate and others remain as residents of their natal areas (Jonsson and Jonsson 1993, Chapman et al. 2011, Dodson et al. 2013, Brönmark et al. 2013). Overall, migration may be considered an adaptive evolutionary strategy, allowing individuals to take advantage of new resources thereby maximizing their individual fitness (Brönmark et al. 2013).

In looking at the variety of migratory life histories that species and populations can adopt, members of the Salmonidae family – including salmon, trout, and char – are a prime example. Some species, such as pink salmon (*Oncorhynchus gorbuscha*) and chum salmon (*Oncorhynchus keta*) are obligate migrants, with each individual undergoing migration between freshwater and the ocean. Others, including brook trout (*Salvelinus fontinalis*), are potadromous, where migration occurs entirely in freshwater, including between rivers and lakes. Still more salmonid species exhibit partial migration, as seen in many populations of rainbow trout (*Oncorhynchus mykiss*). Within these partially migratory populations, residents and migrants spend the first two years of their lives in freshwater (Sogard et al. 2009) prior to the divergence between life histories. Resident rainbow trout go on to reach sexual maturity while spending their entire lives in freshwater (Jonsson and Jonsson 1993, Sogard et al. 2009). In contrast, migrants are

anadromous and migrate to the marine ecosystem at approximately two years of age (Sogard et al. 2009), where they spend 2-5 years before reaching sexual maturity and returning to their natal streams to spawn (Gross et al. 1988).

Within partially migratory salmonid populations, both genetics and the environment are known to play a role in ultimately determining which individuals migrate and which remain resident (Jonsson and Jonsson 1993, Björnsson et al. 2011, Dodson et al. 2013). While much information has been gathered on the ecological and environmental variables connected to migration, the genetic basis of migration remains, overall, poorly understood (Liedvogel et al. 2011). Environmental cues, including changes in resource availability (Jonsson and Jonsson 1993) are early triggers of the migration process. As the main goal of migration is increased resource access, abundant resources in the freshwater system will decrease the likelihood of anadromy; given the risks that migration carries, plentiful food in the freshwater environment may cause individuals to remain resident even if they are genetically able to migrate (Jonsson and Jonsson 1993).

In preparation for migration, anadromous individuals undergo a complex physiological process called smoltification, which requires substantial behavioral, morphological, biochemical, and physiological changes within the individual (Jonsson and Jonsson 1993, McCormick et al. 2009, Ebbesson et al. 2011, McCormick 2013, Hecht et al. 2014). Individuals experience alterations in coloration, increases in growth, thyroid, and corticosteroid hormones, increases in lipid metabolism and protein production, and an increase in length relative to weight, which causes smolts to have a lower condition factor – a weight-to-length ratio – than residents of the same age (McCormick 2009, McCormick 2013). While resident and migrant rainbow trout are indistinguishable prior to the migrants undergoing smoltification, following those processes they are, unlike many species that exhibit partial migration, phenotypically distinguishable: residents

exhibit a darker coloration with distinguishable parr marks and a characteristic pink stripe along the body, contrasted to the silvery hue that migrants develop (Figure 1; Jonsson and Jonsson 1993, McCormick 2013, Dodson et al. 2013). Notably, the climax of smoltification is the development of the ability to osmoregulate in seawater and the corresponding ability of the migratory individual to survive in the marine environment (Jonsson and Jonsson 1993).

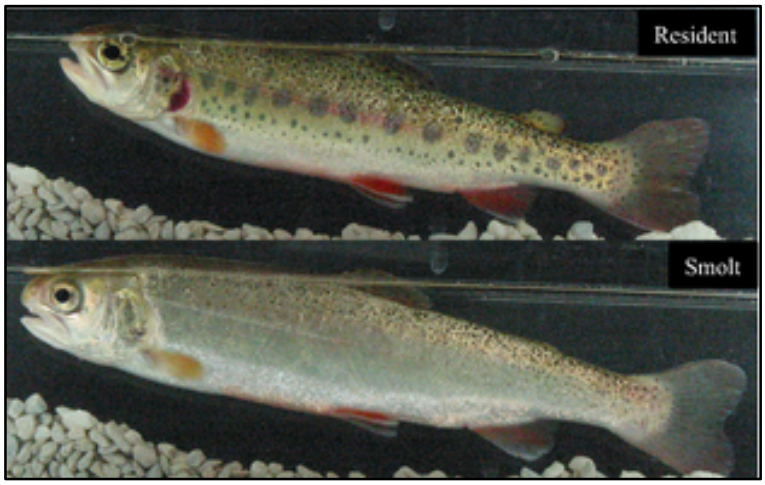

Figure 1: *Phenotypic differences between resident and smolt O. mykiss following smoltification. Both fish are of the same age. Figure taken from Baerwald et al. (2015), see references for a complete citation. Permission was obtained from Wiley.*

Smoltification, and therefore migration, is a life history tradeoff. Migration is energy intensive and mortality rates are high at its onset, but surviving individuals often have access to better food resources than are available in the freshwater system (Jonsson and Jonsson 1993). Increased resource access allows for increased fecundity within migrant individuals, meaning that migrants will ultimately have the capacity to produce more offspring than their resident counterparts (Brönmark et al. 2013). Residents reach sexual maturity at a younger age and are iteroparous, or able to reproduce multiple times throughout their life, but are smaller and experience lower lifetime fecundity than do migrants (Fleming and Reynolds 2004). Migrants delay sexual maturation and reproduction for two to five years after entering the marine environment but, while semelparous, return larger and more fecund than residents (Fleming and Reynolds 2004, Kendall et al. 2014). This increase in fecundity, combined with a higher

energetic cost of egg production as compared to sperm production, causes, in general, a female bias in the proportion of out-migrating smolts (Jonsson and Jonsson 1993, Ohms et al. 2014).

In studying the genetic basis of migration, the rainbow trout population of Sashin Creek and Sashin Lake on Baranof Island in southeastern Alaska has become a model population (Thrower and Joyce 2004, Thrower et al. 2004). Here, a largely migratory rainbow trout population is found naturally in the creek, while the lake, which is isolated upstream of the creek by two barrier waterfalls, was stocked in 1926 with juveniles from the creek (Thrower and Joyce 2004, Thrower et al. 2004). Since then, strong selective pressures against the production of migrants have resulted in the development of a largely resident lake population (Figure 2;

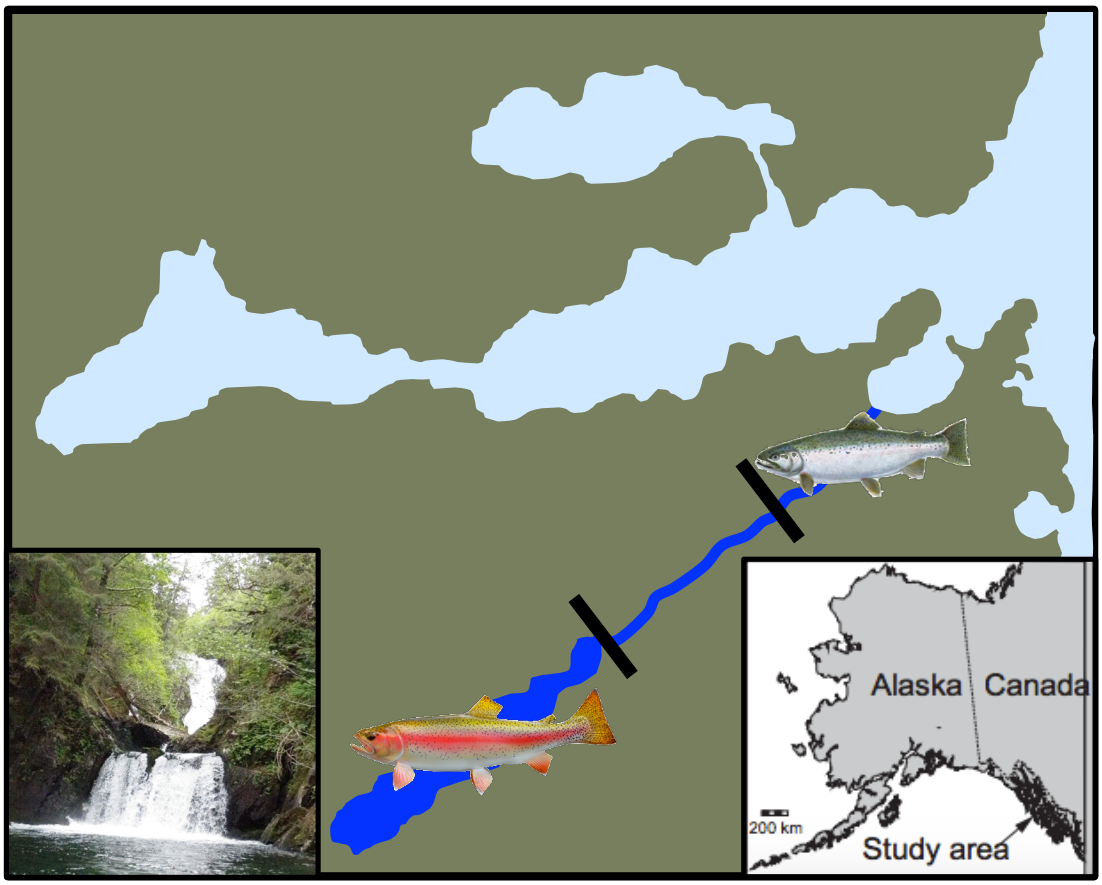

Figure 2: *Zoom-in of Sashin Creek and Lake, located in southeastern Alaska. The dark blue portion of the map represents the study system, with the barrier waterfalls and predominant life history of rainbow trout in each part of the system indicated. Insets show an overall map of the area and one of the barrier waterfalls.*

Thrower and Joyce 2004). Any migrant produced from resident parents in the lake can move downstream over the barrier waterfalls, but is unable to return to the lake to spawn, so these individuals represent lost genetic information to the lake environment.

Within Sashin as a whole, the development of the migratory phenotype is highly heritable (Thrower et al. 2004, Hecht et al. 2015), with offspring typically following the same life history trajectory as their parents. It is important to note, however, that environmental influences still play a role here, and, as such, some individuals develop into the alternative phenotype from their parents. Given the high heritability of migration in the Sashin system (Thrower and Joyce 2004), multiple studies have sought to locate regions of the genome associated with the genetic basis of life history (Hecht et al. 2012, Hale et al. 2013). Early studies identified quantitative trait loci (QTL), or regions of the genome that are associated with phenotypic variation; these QTL studies in the Sashin population have identified several areas of the genome associated with migration-related traits (Hecht et al. 2012). Two main QTL hotspots linked to several migration-related traits have been identified, one on chromosome 12 and one on chromosome 14 (Hecht et al. 2012). Additionally, a QTL hotspot on chromosome 5 linked to development rate and life history determination has been identified in multiple populations as well (Nichols et al. 2007, Miller et al. 2012, Hale et al. 2014, Pearse et al. 2014). This suggests that these chromosomes contain a gene or, more likely, a series of genes, crucial to the development of the migratory phenotype (Hecht et al. 2012). Due to the nature of the QTL analysis, the identified regions were not very specific and spanned all or most of their associated chromosome; while the researchers were able to map QTL back to the rainbow trout genome, there was a lack of precision and therefore an inability to identify specific genes involved in smoltification. To get at more precise areas of the genome associated with migration, a 2013 study looked at single nucleotide polymorphisms (SNPs) that were significant between resident and migrant rainbow trout (Hale et

al. 2013). Of the 7,921 SNPs identified, and 1,998 that were mapped to the genome, significant SNPs were identified on nearly every chromosome (Hale et al. 2013), suggesting that migration is controlled by a complex, polygenic pattern of inheritance. Given the complexity of smoltification – successful smoltification necessitates a suite of biological, physiological, morphological, and behavioral changes in the individual (Nichols et al. 2008, McCormick et al. 2009) – we expect to see this polygenic control of migration.

While we know that both genetics and the environment are important in life history determination, in none of these previous studies have the potential effects of environmental variation been separated out to ensure that the variation between residents and migrants was solely a result of underlying genetic factors. In previous studies, all residents have been sampled from Sashin Lake and all migrants have been sampled from Sashin Creek, and there are different environmental influences acting upon both parts of the system. Given the role of the environment in phenotypic development, the designs of these studies have made it impossible to isolate the influence of genetics on life history. To separate this genetic effect, my project minimizes environmental variation to the maximum extent possible to identify specific genes and areas within the genome that are associated with the development of the migratory phenotype. As an individual's life history cannot be determined until approximately age two, and migratory populations are declining across the continental United States due to climatic changes and anthropogenic disturbances, we aim to use these data to develop a predictive model that can be of use to fisheries managers working to restore and maintain migratory populations. Specifically, my objectives are to:

1. Use data from individuals raised in a common garden environment to identify polymorphic markers (SNPs) associated with the differentiation between resident and migrant phenotypes in *O. mykiss*;

2. Link changes between life history types to specific genes in the genome, and;
3. Use markers identified in Objective 1 as associated with life history to develop a model to predict an individual's ultimate phenotype prior to life history determination.

Methods

Experimental design and sample collection

In 2010, sexually mature resident fish were taken from Sashin Lake, and returning migrants were collected as they entered Sashin Creek. These fish were artificially spawned, and the offspring of the resident-by-resident (RxR; parents from the lake) and the anadromous-by-anadromous (AxA; parents from the creek) crosses were raised in a common environment. Given the different environmental pressures operating on the creek and lake systems, the common garden environmental design seeks to remove any life history effects originating from environmental factors, thereby attributing all subsequent variation to underlying genetics. Only one family of each cross type was used as parents to generate the F_1 samples that were used in this study.

In 2012, when the F_1 generation was 2 years old, fin clips were taken from 327 individuals from the RxR family and 389 individuals from the AxA family. Life history was determined through multiple characteristics: smolts were identified based on skin silvering and a condition factor (weight/length^3 x 100,000) < 1.0, and residents were identified based on gamete expression, a condition factor > 1.0, and a lack of smoltification characteristics. Individuals that appeared phenotypically in between the residents and migrants and that exhibited an uncertain life history were classified as indeterminates.

Objective 1: Use data from individuals raised in a common garden environment to identify polymorphic markers (SNPs) associated with the differentiation between resident and migrant phenotypes in O. mykiss.

Genetic methods and analysis

DNA was extracted from the sampled individuals following a Qiagen DNeasy Blood and Tissue DNA extraction protocol (Qiagen, Hilden, Germany) and diluted to a standard concentration of 50 ng/μL. Given the known sex bias in life history, known resident and migrant individuals were sexed at the OmyY1 locus, a marker on the male-specific portion of the Y chromosome (Brunelli et al. 2008). The sexing protocol was as follows: 94°C for 3 minutes, 30 cycles of 94°C for 50 seconds, 60°C for 50 seconds, and 72°C for one minute, and 72°C for 2 minutes. Each well of the PCR plate contained 0.1 μL each of the forward and reverse OmyY1 primers, 5 μL of GoTaq, 2.8 μL of nano-pure H_2O, and 2 μL of DNA for the sample being sexed. Sex was determined by running 4 μL of PCR product on a 1.5% agarose gel stained with Gel Red and viewed under UV light.

We genotyped 192 samples on an Affymetrix SNP chip (Neogen GeneSeek Operations, Lincoln, Nebraska). These individuals were selected based on life history, condition factor, sex, and familial origin; the total sample was comprised of male and female residents and migrants from both the RxR and the AxA families (Table 1). Condition factor (weight/length^3 x 100,000) was the quantitative factor which we used to select specific individuals within the sex, life history, and family categories. As condition factor decreases during smoltification, we selected residents with the highest condition factors and smolts with the lowest condition factors; all chosen samples were genotyped at 57,501 known polymorphic locations across the genome. Using the flanking sequences associated with each polymorphic marker, SNPs were mapped to the most recent edition of the rainbow trout genome using the default parameters in Bowtie2. We

aimed for an even number of individuals in each family, and within families, an even number of each sex and life history. However, given the known sex bias in life history, in some categories there simply were not enough individuals to make an even distribution possible.

Identification of statistically significant markers

Upon receiving the genotypic data from the Affymetrix SNP chip, the data for each family were separated and treated as separate in analysis. For each family, we assigned a binary code to residents and migrants and used TASSEL (Bradbury et al. 2007) to identify SNPs significantly associated with phenotype. Fixed markers, and those for which a p-value could otherwise not be calculated, were removed from analysis, as were markers that could not be mapped back to the genome. Using these filtered data, we performed an FDR correction on the p-value to minimize our chance of Type I error using the 'p.adjust' function in the 'stats' package in R; 0.01 was used as our level of significance.

Sliding window analysis

After identifying individual markers significantly associated with life history, we performed a sliding window analysis to identify regions of the genome potentially under selection. The sliding window analysis was carried out within the R package 'WindowScanR'; we performed this analysis using raw p-values, observed (H_o) and expected (H_S) heterozygosities (a measure of genetic variation), and Tajima's D (measuring relative allele frequencies per marker within each family) calculated through TASSEL, and with F_{ST} calculated using the R package 'hierfstat'. As F_{ST} is a measure of population differentiation, and the parents of the individuals sampled in this study originated in two distinct natural environments, we treated each family as a population for this sliding window analysis. For all statistics excluding Tajima's D,

the sliding window analysis was performed using a window size of 1,000,000 base pairs and a step size of 500,000 base pairs. In a genomic context, this means that, moving sequentially along each chromosome, we took a 1,000,000 base pair segment of the genome and calculated the average of every individual marker within that region for the statistic in question; that segment was then moved along each chromosome in increments of 500,000 base pairs. For Tajima's D, the process was the same but each window contained 50 markers with a step size of 25 markers.

Table 1: *Numbers of individuals from each family, of both sexes and life histories, that were genotyped on the Affymetrix SNP chip.*

	Migrant x Migrant	**Resident x Resident**
Male Smolts	38	16
Female Smolts	37	18
Male Matures	38	24
Female Residents	11	10
Total No. Genotyped	**124**	**68**

Objective 2: Link changes between life history types to specific genes in the genome.

Functional analysis

We identified genes encompassing one or more significant markers by performing a BLAT analysis, mapping the annotated transcriptome to the *O. mykiss* genome. The start and end position of each mapped gene were then overlaid with the positional information for each SNP to identify overlap. For this analysis, we analyzed our families both separately and together to see if any genes or regions of the genome appeared to be associated with life history regardless of familial origin. A major limitation of this analysis was the inability to map every transcriptome contig back to the genome, due to a combination of a relatively recent salmonid whole genome

duplication (causing tetraploid regions of the *O. mykiss* genome; Lien et al. 2016) and possibly incomplete gene annotations. We used a stringent e-value cut-off of $1e^{-10}$, potentially filtering out contigs that are interesting or important but that did not match exactly back to the genome.

Objective 3: Use markers identified in Objective 1 as associated with life history to develop a model to predict an individual's ultimate phenotype prior to life history determination.

Predictive modeling

By age 2, most individuals had differentiated into either residents or migrants, and were classified accordingly. At that time, however, some individuals displayed a phenotype intermediate to that of either life history and, as such, were termed 'indeterminates'. These individuals retained some of the resident-specific parr marks and a faint pink stripe, but had also begun to display skin silvering. While it is thought that these indeterminates will migrate at age-3, their ultimate life history remains unknown. To develop our predictive model, we genotyped indeterminate individuals from the RxR family at markers with the smallest p-values – these individuals are all full-siblings of the RxR offspring genotyped on the Affymetrix SNP chip.

The first step toward developing our predictive model was to optimize our primers, each amplifying a statistically significant SNP, via PCR to identify appropriate annealing temperatures (Table 2). Following optimization, each indeterminate sample was genotyped at each marker for its respective family. This consisted of a four-step process: 1) the initial PCR reaction to amplify the DNA; 2) cleaning the PCR product to remove primer dimer; 3) a SNaPshot reaction to add a fluorescent base at the polymorphic position correspondent to the individual's genotype, and; 4) running the SNaPshot product on an Applied Biosystems 3130xl Genetic Analyzer to visualize the genotypes.

In the initial PCR, we added to each well of the PCR plate 2 µL 5x GoTaq Flexi Buffer, 0.8 µL magnesium chloride, 1 µL DNTPs (diluted to 2 mM), 0.1 µL Taq, 1 µL of DNA from the given indeterminate individual, the appropriate amount of 10x forward and reverse primer for the given marker (Table 2), and nano-pure H_2O to bring the total volume in each well to 10 µL. The amplification process consisted of a three-minute denaturation period at 94°C, followed by 40 cycles of 94°C for 30 seconds, the appropriate annealing temperature for each marker for 30 seconds, and 72°C for 30 seconds, and finally a constant temperature of 72°C for five minutes. Following this initial PCR, the 5 µL of the product was taken and cleaned with a mixture of 0.7 µL NEB2 buffer, 0.01 µL exo, 0.2 µL rsap, and 1 µL H_2O. This reaction was then run on the thermal cycler at 37°C for 45 minutes and 80°C for 20 minutes to inactivate the enzyme. Once the PCR product was cleaned, we performed a SNaPshot reaction consisting of 2 µL of the cleaned PCR product, 4 µL of 2x forward primer, 1 µL of the SNaPshot mix (Applied Biosystems, Inc., Foster City, California), and 3 µL of nano-pure H_2O. The reaction was amplified using the following protocol: 95°C for 3 minutes, and 40 cycles of 95°C for 10 seconds, 50°C for 5 seconds, and 60°C for 30 seconds. Finally, we combined 3 µL of the post-SNaPshot PCR product with 12 µL of Super-Di Formamide, denatured it at 95°C for 3 minutes, and genotyped the samples on an Applied Biosystems 3130xl Genetic Analyzer (Applied Biosystems, Inc., Foster City, California). The genotypes for each individual at each locus were determined using GeneMapper 4.1 (Applied Biosystems, Inc., Foster City, California).

Once we had compiled the multi-locus genotypes for each indeterminate individual, we used Dataiku's Collaborative Data Science Platform (Dataiku, New York, New York) to develop a model to predict the individual's ultimate phenotype. We first compiled a matrix containing the phenotype and genotype at each of the ten selected markers for the RxR offspring genotyped for Objective 1, and used those data as our 'training' and 'test' sets in determining a best-fit model.

We then applied the model to our indeterminate samples which were genotyped at the same markers, and predicted the probability of each indeterminate developing into either life history.

Table 2: *Markers for each family at which indeterminate samples were successfully genotyped.*

Marker	Family	Annealing Temperature (°C)	Amount of F/R Primer (µL)
Affx-88936379	RxR	58	0.5
Affx-88952521	RxR	58	0.5
Affx-88907930	RxR	58	1
Affx-88933446	RxR	58	1
Affx-88939951	RxR	58	1
Affx-88954026	RxR	58	1
Affx-88959912	RxR	58	1
Affx-88960642	RxR	58	0.5
Affx-88923392	RxR	61	0.1
Affx-88925168	RxR	61	0.1

Results

Objective 1: Use data from individuals raised in a common garden environment to identify polymorphic markers (SNPs) associated with the differentiation between resident and migrant phenotypes in O. mykiss.

Identification of statistically significant markers

Of the 192 individuals genotyped on the Affymetrix SNP chip, 183 were successfully genotyped; the remaining 9 were left out of further analyses. Sixty-four of these individuals belonged to the RxR family, and the remaining 119 belonged to the AxA family. In identifying markers associated with life history differentiation, we treated both families separately. For both families, we first removed from further analysis markers that could not be mapped back to a specific chromosome or position of the *O. mykiss* genome; this left us with 55,657 markers that could be mapped to a genome position. For several markers for either family, we were unable to calculate a p-value and concluded that those markers were fixed within that family. In total, there were 32,274 markers for the AxA family and 22,777 markers for the RxR family for which we were able to calculate a p-value. Following the FDR correction on the p-value, 4,994 markers (~15%) were significantly associated with life history within the AxA family, and 436 markers (~2%) were significantly associated with life history within the RxR family. In the AxA family, significant markers were located on all 29 *O. mykiss* chromosomes, while in the RxR family there were significant markers on every chromosome except for chromosome 18.

Sliding window analysis

We performed sliding window averages for F_{ST} (Figure 3), H_O and H_S (Figure 4), Tajima's D (Figure 5), and raw p-values generated from the life history determination analysis (Figure 6). Sliding window analyses for all metrics except for F_{ST} were performed independently for each family. One F_{ST} analysis was performed, with each family treated as a separate population; this allowed us to evaluate population differentiation. F_{ST} values (Figure 3) at individual markers ranged from zero to 0.89, while the average F_{ST} within each sliding window ranged from zero to 0.53. The overall mean F_{ST} was 0.086. Regions of elevated F_{ST}, indicating

divergence between the two families, were seen on chromosomes 7, 10, 17, and 26. Observed and expected heterozygosities were 0.181 (H_O) and 0.171 (H_S) for the AxA family and 0.148 (H_O) and 0.137 (H_S) for the RxR family. Along with an overall elevated H_O as compared to H_S, there were individual points along several chromosomes (Figure 4) where H_O was greater, indicating more genetic variation than expected at those points. We saw an overall positive trend in Tajima's D (Figure 5), suggesting either balancing selection or a recent population expansion. One region spanning most of chromosome 5 stood out across metrics (Figures 3-5) when looking at both the AxA family and the RxR family, and was characterized by low H_O and H_S (suggesting reduced genetic variation), low F_{ST} (suggesting minimal differentiation between our two families), and negative Tajima's D (this can indicate either purifying selection or a recent population bottleneck – here, we think this region is indicative of purifying selection).

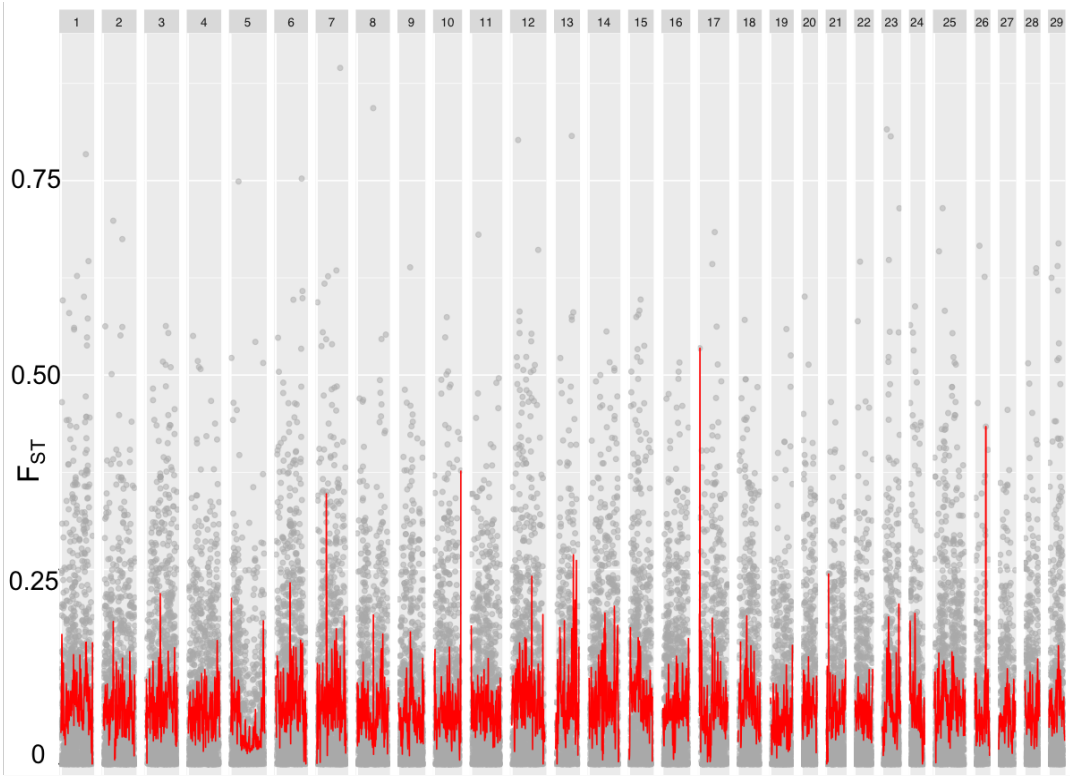

Figure 3: *Sliding window average of F_{ST} for markers that were mapped to the O. mykiss genome. In the comparison, all individuals from the AxA family were used as one population, and all individuals from the RxR family were used as the other population. Grey points represent F_{ST} values for individual markers along each chromosome. The red line is the sliding window average along each chromosome. Each bar represents a chromosome (numbered at top).*

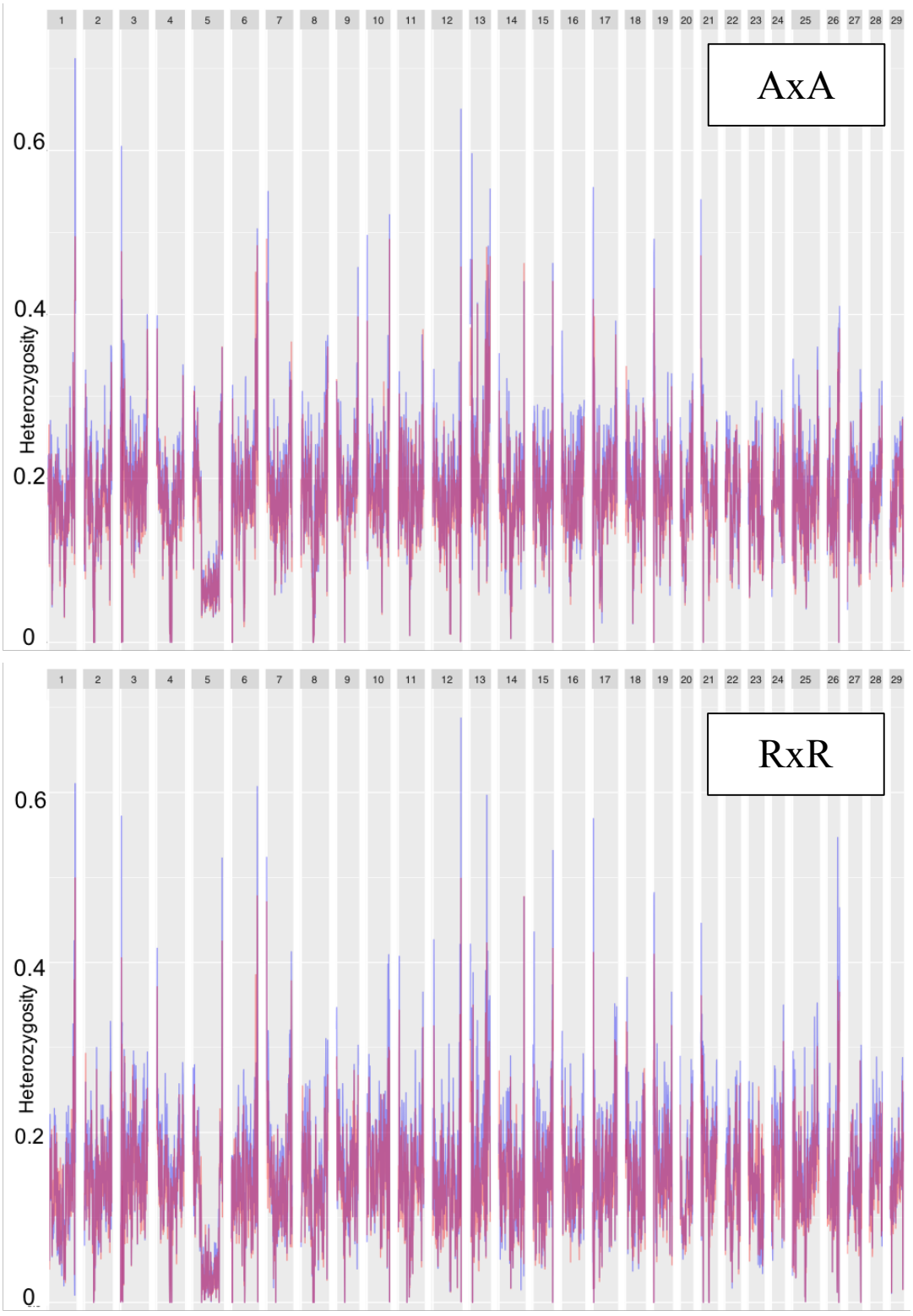

Figure 4: *Sliding window average for observed (blue) and expected (red) heterozygosities for markers from the AxA family (top) and the RxR family (bottom) that were mapped to the O. mykiss genome. Each bar represents a chromosome (numbered at top).*

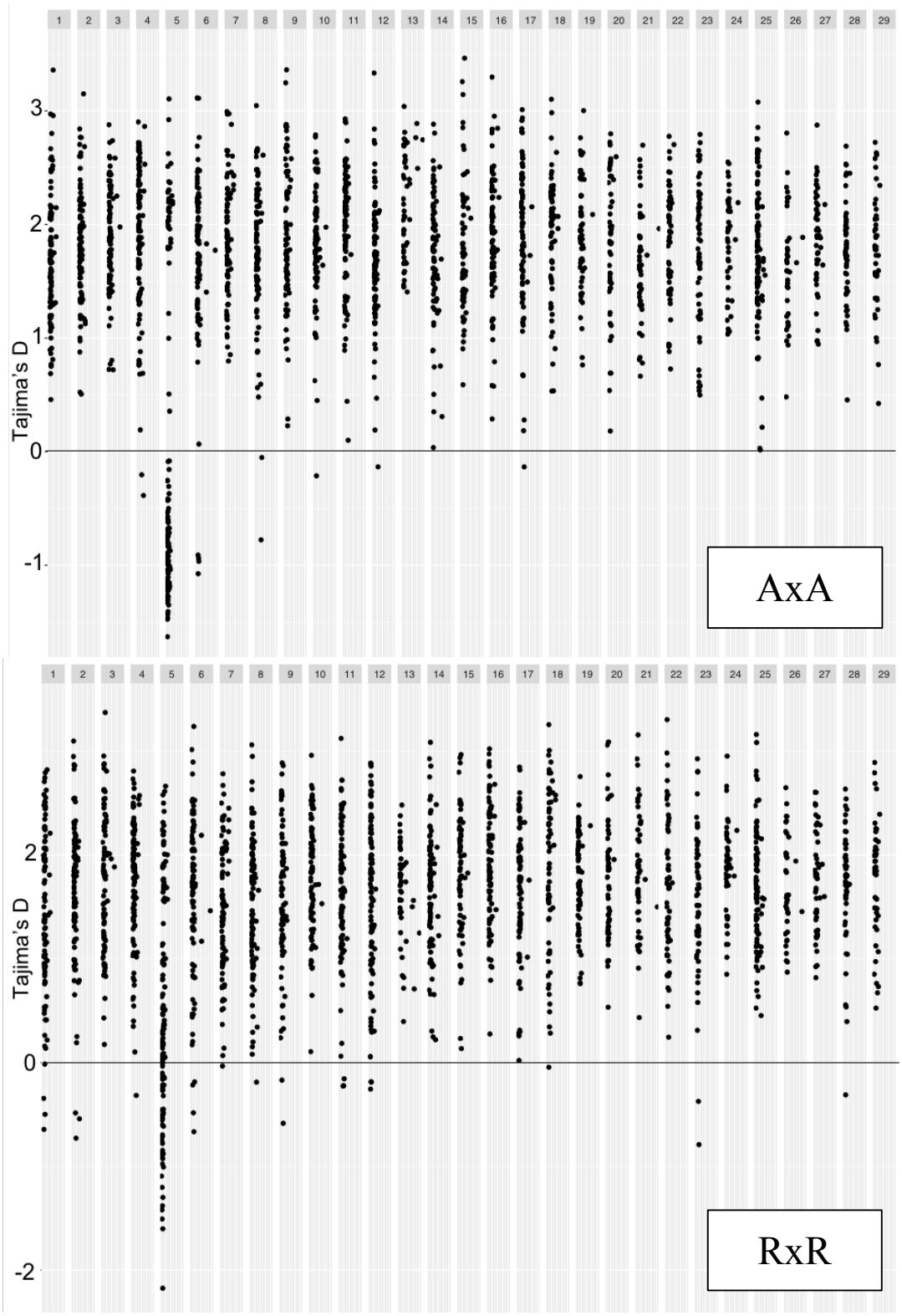

Figure 5: *Sliding window average for Tajima's D within the AxA family (top) and the RxR family (bottom). Each bar represents a chromosome (numbered at top).*

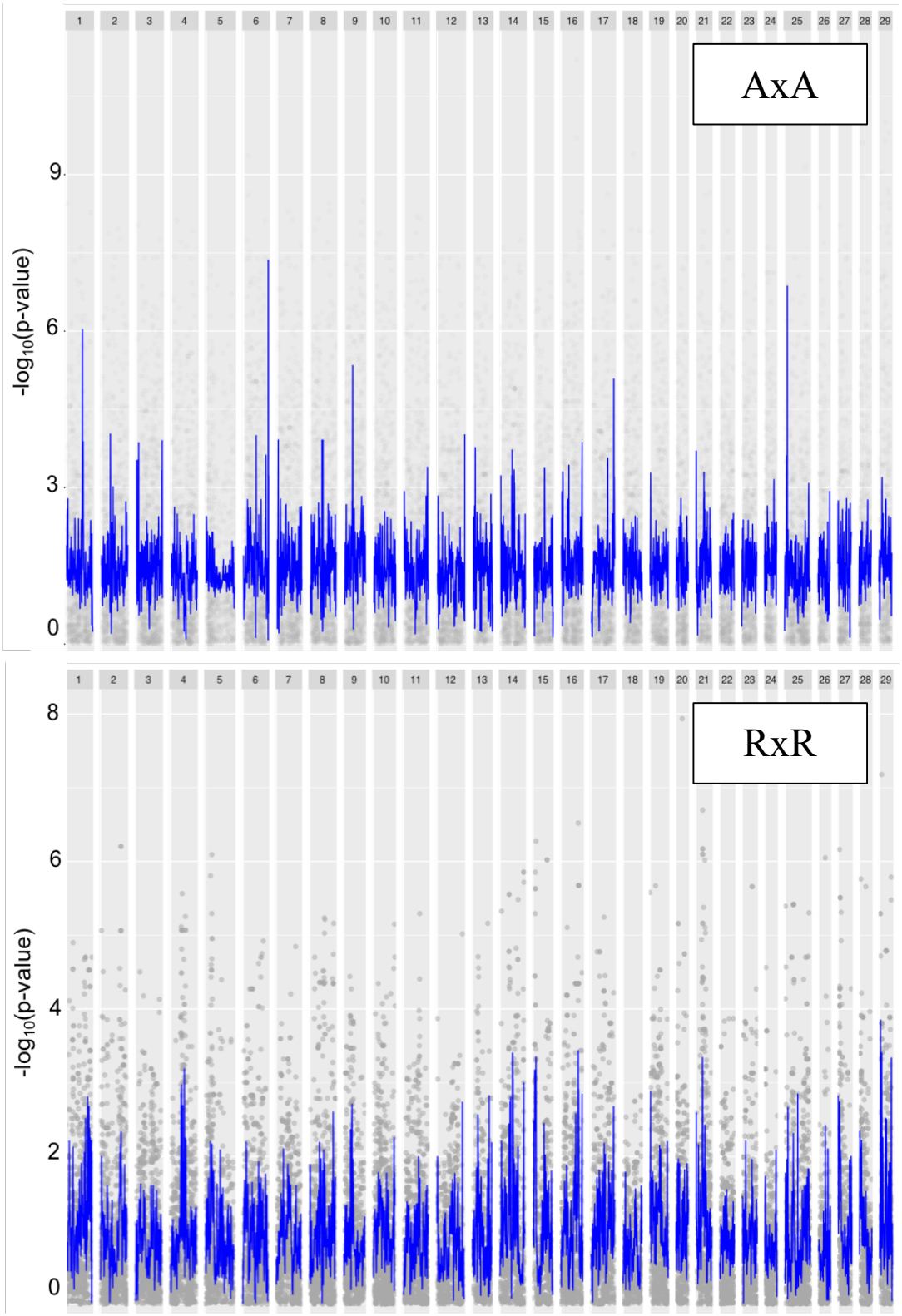

Figure 6: *Sliding window average of raw p-values for markers from the AxA family (top) and the RxR family (bottom) that were mapped to the O. mykiss genome. Grey points represent individual markers along each chromosome. The blue line is the sliding window average along each chromosome. Each bar represents a chromosome (numbered at top).*

Objective 2: Link changes between life history types to specific genes in the genome.

Functional analysis

For the functional analysis, we mapped the annotated *O. mykiss* transcriptome to the genome and identified genes encompassing significant markers from Objective 1 for both families separately and together. We identified give genes on three chromosomes (Table 3) containing markers associated with smoltification in both families, though none of these genes have functions that are known to be related to smoltification. We identified, in the AxA family, significant markers associated with 29 genes whose functions are connected to pathways previously hypothesized to be important in smoltification; one or more significant markers were located within the start and end position of each gene. We grouped these genes into five categories (Table 4): apoptosis (we expect that cells are dying throughout smoltification), phototransduction (the eye must change to allow the individual to see in salt water), lipid and fatty acid metabolism (metabolism changes during smoltification, with smolts undergoing a period of fasting), growth and development (condition factor changes to adapt the body to the marine environment), and ion homeostasis (the individual must maintain homeostasis in salt water). No markers tied to smoltification-related genes were found in the RxR family.

Table 3: *Regions of the genome that contain significant markers for both families.*

Chromosome	Start Position	End Position	Annotation
omy02	65249829	65709098	si:dkey-153m14.1
omy02	66720536	67198313	basic helix-loop-helix family, member e41
omy16	55765846	55782186	interleukin 17 receptor E-like
omy16	55805949	55837272	intraflagellar transport 122 homolog
omy29	7014767	7044810	NSA2 ribosome biogenesis homolog

Table 4: *Genes in migrant family with functions related to smoltification and that encompass one or more significant SNPs.*

	Chromosome	Start Position	End Position	Annotation
Apoptosis	omy06	82137635	82138023	death-associated protein kinase 2
	omy07	78259944	78263272	programmed cell death 2-like
	omy09	54514509	54517736	caspase 9
Phototransduction	omy04	679023	1282575	stimulated by retinoic acid 6
	omy16	1946000	1980894	opsin 5
	omy27	1547273	1576062	S-antigen; retina and pineal gland (arrestin)
Lipid/Fatty Acid Metabolism	omy02	53713259	53730095	cholesteryl ester transfer protein, plasma
	omy08	9245714	9294768	apolipoprotein B
	omy09	51478701	51489578	proteolipid protein 2
	omy16	52253283	52264527	lipopolysaccharide binding protein
Growth and Development	omy01	14034893	14075728	Janus kinase 2
	omy01	61470543	61579640	thyroid adenoma associated
	omy02	74185429	74189997	cortactin binding protein 2
	omy04	679023	1282575	stimulated by retinoic acid 6
	omy07	11474375	11476826	cathepsin S
	omy08	76874702	76882803	purine nucleoside phosphorylase
	omy14	59706925	59715742	cathepsin O
	omy21	44434836	44435808	methyltransferase like 12
	omy22	28173844	28550792	fibroblast growth factor 14
	omy22	32457242	32466188	myosin, light chain 1, alkali; skeletal, fast
	omy28	11228455	11268694	epidermal growth factor receptor
Ion Homeostasis	omy01	52900204	52981711	SPARC related modular calcium binding 2
	omy01	55613945	55619319	transmembrane protein 72
	omy07	72345554	72864226	solute carrier family 39, member 8
	omy09	44791396	44824431	catenin, beta interacting protein 1
	omy22	2791347	3084253	sodium leak channel, non-selective
	omy22	47354055	47378208	transmembrane protein 237
	omy23	9020635	9022818	calcium homeostasis modulator 3
	omy28	5952093	6379074	solute carrier family 26, member 5

Objective 3: Use markers identified in Objective 1 as associated with life history to develop a model to predict an individual's ultimate phenotype prior to life history determination.

Predictive Modeling

We successfully optimized ten markers (Table 2) and used them to develop a predictive model for 44 indeterminate individuals from the RxR family. The life histories and genotypes of our known individuals at these markers were used to develop the predictive model. To determine the modeling method most applicable to the data, we used the 'Insights' prediction model template in the Dataiku software; the Gradient Boosted Trees model, a prediction model based on the construction of decision trees, best fit our data. We trained the model using the genotypes and life histories of the 64 RxR offspring that were genotyped on the Affymetrix SNP chip, and tested the model on the data for our 44 indeterminates. Using the model, we predicted 24 indeterminates to ultimately develop into residents, and 20 to ultimately develop into migrants.

Discussion

Objective 1: Use data from individuals raised in a common garden environment to identify polymorphic markers (SNPs) associated with the differentiation between resident and migrant phenotypes in O. mykiss.

Identification of statistically significant markers

We identified markers significantly associated with life history that were, for both familial crosses, scattered throughout the genome, providing further support to the idea that life history determination is controlled via a complex, polygenic pattern of inheritance (Hale et al. 2013). Smoltification requires modification of an individual's behavior, morphology, and physiology, and these complex processes require the action of many genes across the genome. The presence of statistically significant markers on nearly every *O. mykiss* chromosome, both

here and in previous studies (Hale et al. 2013), suggests that life history and the development of the migratory phenotype are controlled by many markers of small effect, rather than by one 'master switch' or by few genes of large effect.

This trend of an array of markers tied to life history determination holds constant across studies and populations, though there appears to be population-level specificity in terms of the precise genomic regions underlying these influential markers (Hecht et al. 2013). These population-level differences may be explained, at least in part, by the repeated evolution of partial migration. Partial migration has been shown to have evolved repeatedly in rainbow trout, with residents and migrants in a given population more closely related to one other than either life history is to the same life history in another population (Docker and Heath 2001). Consequently, this may increase the likelihood of population-level variation in the genetic basis of life history.

The combination of different methodologies and different markers used across studies allows more for the comparison of trends than for direct marker- or position-level comparisons. When comparing the data presented here to previous studies from Sashin (Hale et al. 2013), we see here greater marker density both overall and when looking solely at markers significantly associated with life history. In this study, the combination of a new version of the genome to map our sequences back to and a newer technique allowed us to successfully genotype individuals at more SNPs and to map more of those SNPs to a unique chromosomal position. Additionally, our question focuses on the genetic basis of smoltification, whereas Hale et al. (2013) focused on the genetic basis of successful migration. At the onset of migration, mortality rates are high (Jonsson and Jonsson 1993), individuals face increased predation risk and are exposed to new pathogens. Due to these high mortality rates, undergoing smoltification does not guarantee that an individual will successfully migrate, which is classified by the individual

returning to freshwater to spawn. Therefore, our dataset consists of genotypic data for individuals that both survived and perished during migration; this is a larger subset of individuals than was used in 2013, further contributing to the differences in marker density seen here.

When comparing our two families, we saw greater genetic variation within the AxA family, as evidenced by the identification of 4,994 statistically significant markers (~15% of total mapped markers), as compared with the 436 for the RxR family (~2% of total mapped markers). This trend is as expected given the parental origins of the individuals sampled here; the AxA parents were sampled from Sashin Creek, and the RxR parents were sampled from Sashin Lake. As Sashin Lake is isolated by two barrier waterfalls, new individuals – and, consequently, sources of genetic variation – are unable to move into the lake. Therefore, the genetic variation in the lake is constrained to that of the individuals already present. Additionally, as the lake population was founded about 90 years ago by taking approximately 60 juvenile fish from the creek and relocating them into the lake, any offspring produced from lake parents essentially represent a limited sample of the genetic variation within the creek, barring any mutations that occurred only in the lake post-separation. In addition to the genetic variation present in the creek, migrant individuals that have strayed into Sashin from other populations, along with individuals migrating downstream from the lake (given the occurrence of lake-specific polymorphisms), provide new genetic material and therefore increased variation to the creek population.

Sliding window analysis

Performing sliding window analyses for various population genetic metrics allowed us to visualize regions of the genome potentially under selection. The metrics we focused on were the previously calculated p-values associated with life-history development, variations between observed and expected heterozygosities, F_{ST}, and Tajima's D. Though in identifying significant

markers we corrected for multiple tests and a potentially inflated type I error rate through use of an FDR correction on the p-value, we chose to perform our sliding window analysis on the raw p-values from the life history development analysis. The FDR correction minimizes variance between the p-values for any given marker, and in looking for these regions of the genome under selection, we wanted to maintain any natural variation between markers. The parameters for the sliding window analysis – namely window size and step size – were chosen by trial and error. We wanted parameters that would smooth out some of the background noise from all of the individual markers, without minimizing it so much that we were unable to discern patterns and trends. Overall, performing sliding window analyses allowed us to observe patterns within the genome, expanding our focus from individual markers under selection.

For all statistics, we calculated overall genome-wide averages and we identified regions of the genome potentially under selection. We saw peaks for the raw p-value sliding window on several chromosomes, most notably when looking at the AxA family (Figure 6). These peaks were located on chromosomes 1, 6, 9, 17, and 25, indicating increased differentiation between the residents and migrants of that family within those regions. None of these peaks were identified by Hale et al. (2013) when they looked at the genetic basis of successful migration in Sashin, though Hecht et al. (2012) identified a QTL on chromosome 1 associated with condition factor, and two QTLs on chromosome 6 associated with body shape.

In combining our families to measure F_{ST}, we calculated an average F_{ST} of 0.086. This is comparable to previous F_{ST} estimates in Sashin (F_{ST} = 0.069; Hale et al. 2013) but is higher than that of other rainbow trout populations (in Little Sheep Creek, Oregon, F_{ST} = 0.003; Hale et al. 2013). One reason for this is the nature of the Sashin system: in many other populations, including in Little Sheep Creek, residents and migrants coexist in the same space, whereas in Sashin the two life histories are largely separated by the two barrier waterfalls. Consequently,

such separation would lead to greater differentiation between our two families as there would be if the population were continuous.

Average H_O and H_S were, respectively, 0.181 and 0.171 for the AxA family, and 0.148 and 0.137 for the RxR family. The observed heterozygosities calculated here are lower than those previously calculated in Sashin, where the average H_O for the migrant population (sampled in the creek) and for the resident population (sampled in the lake) were both above 0.2 (Hale et al. 2013). This indicates reduced genetic variation within our samples, almost certainly as a result of analyzing two full sibling families which inherently have less variation than the natural population. Within both families analyzed here, however, H_O was greater than H_S, indicating that more genetic variation was observed within our samples than was expected. When looking at Tajima's D for both the AxA and RxR families, we see an overall positive trend across the genome. A positive Tajima's D can be indicative of one of two things – either balancing selection or a recent population expansion. Currently, we are unsure which explanation is most responsible for the trends seen here.

There was one region spanning most of chromosome 5 that stood out across metrics. This region was characterized by little differentiation between the residents and migrants within the AxA family (Figure 6); low H_O and H_S for both families (suggesting low genetic variation within this region; Figure 4); low F_{ST} between the RxR and AxA families (suggesting minimal population differentiation; Figure 3); and negative Tajima's D (Figure 5), indicating an excess of rare polymorphisms and possibly signaling purifying selection (the purging of deleterious alleles) or genetic hitchhiking (where alleles at a locus not under selection change in frequency because that locus is in linkage disequilibrium with one that is under selection). While not measured here, previous work in Sashin has shown, in the migrant population, reduced

nucleotide diversity on chromosome 5. Together, these data suggest that this region is being actively conserved.

Previous research across multiple rainbow trout populations (Nichols et al. 2007, Miller et al. 2012, Hale et al. 2014, Pearse et al. 2014) have identified a major development rate QTL and polymorphic markers strongly associated with life history determination spanning this region on chromosome 5. These QTLs are important in regulating the amount of time it takes for an embryo to hatch (Nichols et al. 2007), which in turn has an influence on an individual's ultimate life history (Nichols et al. 2008; Hale et al. 2014). Somewhat counterintuitively, a faster development rate is associated with residency in the Sashin system, while a slower development rate is associated with smoltification (Hale et al. 2014). In Sashin, the development rate QTL is hypothesized to be a cold water adaptation (Hale et al. 2014); cold water systems, such as Sashin Lake, have shorter growing seasons than their warm water counterparts, so embryos must develop faster for a chance at survival. We have not measured development rate in these fish, so any link that we have made between the data presented here and development rate is based on other research. However, in Sashin there is a difference in development rate between cross type with progeny from resident families developing significantly faster than progeny from anadromous families (Hale et al. 2014). As there is a strong heritable component to smoltification in this population, this suggests that RxR individuals will develop faster and are more likely to be resident (Hale et al. 2014).

It is important to note that, due to limitations on marker density (Hecht et al. 2012), the development rate QTL has not been identified or mapped in the Sashin population. However, there is evidence in Sashin that polymorphisms in genes associated with the development rate QTL – specifically in *Clock1a*, *Myd118-1*, and *GHR1* – are associated with time to hatch and, consequently, life history (Hale et al. 2014). These three candidate genes are all known to be

located on chromosome 5 and, therefore, links between genotypic variation in these genes and development rate in Sashin confirms the importance of this region of the genome for the Sashin Creek system. Together, the previously published data combined with the data presented here indicate that there are genes or polymorphic markers on chromosome 5 that are tied to development and life history determination across rainbow trout populations.

Within multiple *O. mykiss* crosses, this region on chromosome 5 has been shown to control over 20% of the observed variation within development rate (Nichols et al. 2007, Miller et al. 2012); it is also seen to be in strong linkage disequilibrium across populations (Miller et al. 2012, Pearse et al. 2016), suggesting that the region is crucial in influencing development rate and life history. Regions in strong linkage disequilibrium have a reduced recombination rate, meaning that the region is highly conserved and possibly indicating that there are beneficial alleles within that region that are consistently inherited as one "unit" – in this case, some of those alleles are linked to development rate and are directly selected upon. We have not directly measured linkage disequilibrium and recombination rate in Sashin, but all metrics – decreased H_O and H_S, negative Tajima's D, and low F_{ST} – suggest a reduced recombination rate, with combinations of alleles being actively conserved. These alleles are most likely associated with development rate, which has a known link with migration and life history.

The presence of this highly conserved region across rainbow trout populations necessitates comparisons to the three-spined stickleback, where independent freshwater populations have undergone parallel evolution – standing genetic variation caused low frequency mutations associated with freshwater adaptation to go to fixation when marine populations were isolated in freshwater lakes following the last ice age (Hohenlohe et al. 2010). In rainbow trout, this conserved region could also be indicative of parallel evolution, where mutations associated with development rate are selected for when one life history is selected against. This region may

also be an adaptation to development in cold water (Miller et al. 2012), and have evolved to maximize resources in systems with shorter growing seasons. Finally, independent selection may be occurring, where allelic diversity is reduced in this region because of the presence of genes associated with development rate and, therefore, life history. The situation with rainbow trout is different from that of stickleback in that offspring of both life histories are still produced from parents of either life history, and our data currently do not support one hypothesis over another, but future research may provide further insight into this.

Objective 2: Link changes between life history types to specific genes in the genome.

Functional analysis

The 28 genes with predicted smoltification-related functions that encompassed one or more significant markers were assigned to one of five groups. These groups - apoptosis, phototransduction, lipid and fatty acid metabolism, growth and development, and ion homeostasis – are all functionally tied to the smoltification process. Five genes on three chromosomes contained significant markers within each family (Table 4), though none of these genes have functions known to be related to smoltification. However, this suggests that these regions of the genome may be important in life history determination regardless of an individual's familial origin.

No gene identified through this analysis overlapped with the QTL hotspots previously identified on chromosomes 12 or 14 (Hecht et al. 2012), or with the time-to-hatch QTL on chromosome 5. Salmonids have experienced a recent whole genome duplication event approximately 80 million years ago (Lien et al. 2016), and retain functional gene duplicates across much of the genome. When mapping the annotated transcriptome to the *O. mykiss* genome, we removed any genes that did not confidently map to one distinct position in the

genome. Of the 22,415 contigs in the transcriptome, just under 10,000 mapped to one chromosome, and 4,598 of those were annotated. It is likely that contigs mapping to genes containing significant markers were removed due to the stringent filtering used in this analysis, and that mapping a broader subset of contigs in the transcriptome would allow us to identify these genes. Additionally, we have no evidence here that the genes containing significant markers are directly involved in the smoltification process, only that their functions have known ties to functions important to smoltification.

Other work in Sashin looking at patterns of gene expression has identified many genes known to be involved in the smoltification process (Hecht et al. 2014, Hale et al. 2016); these include genes that regulate thyroid hormone, growth hormone, and cortisol hormone. Smoltification is a complex process, and the lack of candidate genes that we find here containing SNPs significantly associated with life history does not mean that those candidate genes are not important in the smoltification process. The SNPs analyzed here were scattered throughout the genome and may not have overlapped with any of these genes of known importance. Alternatively, these genes may be differentially expressed within our population and therefore be biologically significant, but not significant on a statistical level. Different genes are also expressed at different times throughout the smoltification process, and so a lack of differential expression may be indicative of the exact moment in time when our fish were sampled. As we do not have gene expression data for the samples analyzed in this study, we are unable to assess patterns of gene expression within the individuals studied here and confirm how those data interact with the statistical results presented within.

Along with gene expression work, there has also been epigenetic work done in other rainbow trout crosses that identified, when comparing residents and migrants, several differentially methylated DNA regions (Baerwald et al. 2015). Some of these regions overlapped

with QTL previously identified in the same cross (Nichols et al. 2008), suggesting that genetics and epigenetics may work together in determining and distinguishing life history types (Baerwald et al. 2015). While we do not have epigenetic data for the Sashin population, there may well be regions of the genome that exhibit differential methylation patterns as well, and that are important in and play into the differentiation between life histories.

Objective 3: Use markers identified in Objective 1 as associated with life history to develop a model to predict an individual's ultimate phenotype prior to life history determination.

Predictive Modeling

Given that indeterminate individuals exhibit a mix of resident and migrant phenotypes, it is generally thought that these individuals will undergo smoltification at age-3, rather than age-2, thus ultimately developing into migrants (Hecht et al. 2015). Of the 44 indeterminates whose life histories were predicted here, roughly half were predicted as either life history, with slightly more (54.5%, n = 24) predicted to remain resident. All 44 individuals were offspring of the RxR cross, and given the known heritability of life history (Thrower et al. 2004, Hecht et al. 2015), it is possible that cross type influenced the high predicted rate of residency. However, as some indeterminates are predicted to migrate and these individuals are adopting the alternative life history as their parents, we see that life history is not completely heritable and is additionally influenced by external factors.

The main reasoning behind developing this model was for its potential utility in predicting the life history of an individual before the onset of the smoltification process. We do not have data on the ultimate life history trajectory of any indeterminates whose life history was predicted here, and in order to be successful, this model must be ground-truthed by comparing it to samples of known life history both in Sashin and in other populations. Given the population

specific nature of the heritability of smoltification in rainbow trout, we are unsure whether this model will be applicable in other populations. It is possible, however, that across populations similar regions of the genome can be targeted and observed to predict an individual's life history. Even if the model can only be used within Sashin, where fisheries managers may be able to use this model to target the precise genomic markers identified here, it will still allow us to make predictions on ultimate life histories and help to maintain the anadromous steelhead population in Sashin Creek.

Limitations

There are both benefits and downsides to the common garden experimental design employed here. We know that both genetics and the environment are involved in life history determination, and until now, the environment has been left in as a confounding variable in all studies performed in Sashin (Hecht et al. 2012, Hale et al. 2013, Hecht et al. 2015). By raising the offspring of the RxR and AxA crosses in a shared environment, the common garden experiment removes, to the maximum extent possible, environmental variation from the analysis of life history determination; any differences that we see between residents and migrants should be a result of underlying genetic variation. Consequently, one of the greatest benefits from this experimental setup is that we are able to assess genetic variation and the influences of genetics on life history determination in the absence of environmental factors, which have the known potential to override genetics in determining life history. However, given that we know the importance of the environment in life history determination, removing the environment as a variable as was done here may mean that our results are not applicable to wild rainbow trout populations. Additionally, another limitation is that every individual analyzed here is the product

of meioses within four breeding individuals. As we are analyzing two sets of full siblings, we expect to see reduced genetic variation within our samples compared to the wild population.

Work on sockeye salmon (*O. nerka*; Nichols et al. 2016), which also exhibit partial migration, and rainbow trout (Docker and Heath 2001) has shown that life history differentiates individually in different populations; residents and migrants in any given system are more closely related to each other than to individuals of the same life history type from a different population (Docker and Heath 2001, Nichols et al. 2016). It is possible, therefore, that our results, in particular the predictive model, may not be applicable to other rainbow trout populations. There is no way to assess this without testing our model on individuals from populations other than Sashin, and on wild individuals from Sashin. Continued work may include obtaining samples from other populations and assessing whether our model is applicable to the natural environment.

Conclusions

In this study we identified markers significantly associated with life history in the offspring of a cross between two resident parents and a cross between two migrant parents. As has previously been demonstrated in Sashin, these markers were distributed throughout the genome, suggesting that life history determination is controlled via a complex, polygenic pattern of inheritance (Hale et al. 2013). We identified a region on chromosome 5 that has been previously found to be important in development rate, and from that in life history differentiation. We also developed a model to predict life history in individuals before they undergo smoltification or reach sexual maturity, which has the potential use and application to fisheries managers. Migratory rainbow trout populations are decreasing in number throughout

much of their range, due to factors such as climate change and anthropogenic activities, and we hope that this model can be used as a starting point in conserving migratory populations.

REFERENCES

Björnsson BT, Stefansson SO, and McCormick SD. 2011. Environmental endocrinology of salmon smoltification. Gen Comp Endocrinol 170(2):290-8.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-35.

Brönmark C, Hulthén K, Nilsson P, Skov C, Hansson L, Brodersen J, and Chapman B. 2013. There and back again: migration in freshwater fishes. Can. J. Zool. 92(6):467-79.

Brunelli JP, Wertzler KJ, Sundin K, and Thorgaard GH. 2008. Y-specific sequences and polymorphisms in rainbow trout and Chinook salmon. Genome 51:739-48.

Chapman BB, Brönmark C, Nilsson J, and Hansson L. 2011. Partial migration: An introduction. Oikos 120(12):1761-3.

Dingle H and Drake VA. 2007. What is migration? Bioscience 57(2):113-21.

Docker MF and Heath DD. 2003. Genetic comparison between sympatric anadromous steelhead and freshwater resident rainbow trout in British Columbia, Canada. Conserv Genet 4(2):227-31.

Dodson JJ, Aubin-Horth N, Thériault V, and Páez DJ. 2013. The evolutionary ecology of alternative migratory tactics in salmonid fishes. Biological Reviews 88(3):602-25.

Ebbesson L, Nilsen T, Helvik J, Tronci V, and Stefansson S. 2011. Corticotropin-Releasing factor neurogenesis during midlife development in salmon: genetic, environmental and thyroid hormone regulation. J Neuroendocrinol 23(8):733-41.

Fleming IA, and Reynolds JD. 2004. Salmonid breeding systems. *In* Evolution illuminated: salmon and their relatives. *Edited by* A.P. Hendry and S.C. Stearns. Oxford University Press, Oxford. pp. 264-94.

Greenwood P. 1980. Mating systems, philopatry and dispersal in birds and mammals. Anim Behav 28(NOV):1140-62.

Gross M, Coleman R, and McDowall R. 1988. Aquatic productivity and the evolution of diadromous fish migration. Science 239(4845):1291-3.

Hale MC, Colletti JA, Gahr SA, Scardina J, Thrower FP, Harmon M, Carter M, Phillips RB, Thorgaard GH, Rexroad III CE. 2014. Mapping and expression of candidate genes for development rate in rainbow trout (*Oncorhynchus mykiss*). J Hered 105(4):506-20.

Hale MC, McKinney GJ, Thrower FP, Nichols KM. 2016. RNA-seq reveals differential gene expression in the brains of juvenile resident and migratory smolt rainbow trout (*Oncorhynchus mykiss*). Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 20:136-50.

Hale MC, Thrower FP, Berntson EA, Miller MR, and Nichols KM. 2013. Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. G3: Genes| Genomes| Genetics 3(8):1273-85.

Hecht BC, Campbell NR, Holecek DE, Narum SR. 2013. Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. Mol Ecol 22(11):3061-76.

Hecht BC, Hard JJ, Thrower FP, and Nichols KM. 2015. Quantitative genetics of migration-related traits in rainbow and steelhead trout. G3: Genes| Genomes| Genetics 5(5):873-89.

Hecht BC, Thrower FP, Hale MC, Miller MR, and Nichols KM. 2012. Genetic architecture of migration-related traits in rainbow and steelhead trout, *Oncorhynchus mykiss*. G3: Genes| Genomes| Genetics 2(9):1113-27.

Hecht BC, Valle ME, Thrower FP, and Nichols KM. 2014. Divergence in expression of candidate genes for the smoltification process between juvenile resident rainbow and anadromous steelhead trout. Marine Biotechnology 16(6):638-56.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics 6(2):e1000862.

Howard WE. 1960. Innate and environmental dispersal of individual vertebrates. Am Midl Nat 63(1):152-61.

Jonsson B and Jonsson N. 1993. Partial migration - niche shift versus sexual-maturation in fishes. Rev Fish Biol Fish 3(4):348-65.

Kendall NW, McMillan JR, Sloat MR, Buehrens TW, Quinn TP, Pess GR, Kuzishchin KV, McClure MM, and Zabel RW. 2014. Anadromy and residency in steelhead and rainbow trout (*Oncorhynchus mykiss*): A review of the processes and patterns. Can J Fish Aquat Sci 72(3):319-42.

Liedvogel M, Åkesson S, and Bensch S. 2011. The genetics of migration on the move. Trends in Ecology & Evolution 26(11):561-9.

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature 533(7602):200.

McCormick SD. 2009. Evolution of the hormonal control of animal performance: Insights from the seaward migration of salmon. Integrative and Comparative Biology 49(4):408-22.

McCormick SD. 2013. Smolt physiology and endocrinology. Fish Physiology 32:199-251.

McCormick SD, Lerner DT, Monette MY, Nieves-Puigdoller K, Kelly JT and Björnsson BT. 2009. Taking it with you when you go: How perturbations to the freshwater environment, including temperature, dams, and contaminants, affect marine survival of salmon. American Fisheries Society Symposium. 195.

Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE, Palti Y, Doe CQ, Thorgaard GH. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol Ecol 21(2):237-49.

Nichols KM, Broman KW, Sundin K, Young JM, Wheeler PA, Thorgaard GH. 2007. Quantitative trait loci x maternal cytoplasmic environment interaction for development rate in Oncorhynchus mykiss. Genetics 175(1):335-47.

Nichols KM, Edo AF, Wheeler PA, and Thorgaard GH. 2008. The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. Genetics 179(3):1559-75.

Nichols KM, Kozfkay CC, Narum SR. 2016. Genomic signatures among *Oncorhynchus nerka* ecotypes to inform conservation and management of endangered sockeye salmon. Evolutionary Applications 9(10):1285-300.

Ohms HA, Sloat MR, Reeves GH, Jordan CE, and Dunham J.B. 2014. Influence of sex, migration distance, and latitude on life history expression in steelhead and rainbow trout (*Oncorhynchus mykiss*). Can J Fish Aquat Sci 71(1):70-80.

Pearse DE, Miller MR, Abadia-Cardoso A, Garza JC. 2014. Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. Proc Biol Sci 281(1783):20140012.

Sogard SM, Williams TH, and Fish H. 2009. Seasonal patterns of abundance, growth, and site fidelity of juvenile steelhead in a small coastal California stream. Trans Am Fish Soc 138(3):549-63.

Thrower FP and Joyce JE. 2004. Effects of 70 years of freshwater residency on survival, growth, early maturation, and smolting in a stock of anadromous rainbow trout from southeast Alaska. American Fisheries Society Symposium. 485 p.

Thrower FP, Hard JJ, and Joyce JE. 2004. Genetic architecture of growth and early life-history transitions in anadromous and derived freshwater populations of steelhead. J Fish Biol 65(s1):286-307.

VITA

PERSONAL BACKGROUND

Spencer Yael Weinstein, born 28 May 1994 in Red Bank, New Jersey to Robin P. and Michael A. Weinstein.

Graduated from Lenape High School in Medford, New Jersey in June 2012.

EDUCATION

M.S. Biology, Texas Christian University	2018
B.S. Natural Resources Conservation, University of Massachusetts Amherst	2016
B.A. Geography, University of Massachusetts Amherst	2016

EXPERIENCE

Teaching Assistantship, Texas Christian University	2016 – 2018
Undergraduate Researcher, University of Massachusetts Amherst	2014 – 2016
REU – Virginia Institute of Marine Science	2015
REU – South Dakota State University	2014

ABSTRACT

SHOULD I STAY OR SHOULD I GO? ANALYZING THE GENETIC BASIS OF MIGRATION-RELATED TRAITS IN RAINBOW TROUT (*ONCORHYNCHUS MYKISS*).

By Spencer Yael Weinstein, M.S. 2018

Department of Biology

Texas Christian University

Thesis Advisor: Dr. Matthew C. Hale, Assistant Professor of Biology

Rainbow trout (*Oncorhynchus mykiss*) exhibit partial migration, and are well-suited to studying the genetic basis of migration due to phenotypic differences between residents and migrants. We used data from a common garden experiment to identify SNPs associated with migration in the F_1 generation of two familial crosses. We genotyped 192 individuals at 57,501 known polymorphic positions across the genome, and performed sliding window analyses for various metrics to identify regions of the genome potentially under selection. We identified 4,994 and 436 significant SNPs in each of our families, suggesting that life history is controlled via a complex, polygenic pattern of inheritance. Within one family, we located significant markers associated with 29 genes with functions connected to smoltification-related pathways. Chromosome 5 stood out across sliding window analyses, indicating that this region is under selection. We also developed a model to predict life history, which we hope will have conservation implications.