COMPARATIVE GENOMICS OF RAINBOW TROUT (*ONCORHYNCHUS MYKISS)*: ARE GENES ASSOCIATED WITH MIGRATION CONSERVED AMONG POPULATONS?

by

CATHERINE IRENE CLARE

Bachelor of Science, 2017
Saginaw Valley State University
University Center, MI

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 2021

ACKNOWLEDGEMENTS

I am both grateful and blessed to have the support of so many people during my time at TCU. First and foremost, I would like to thank my advisor, Dr. Matt Hale for his constant efforts in pushing me to become a better researcher and scientist. His encouragement and expertise in both designing and executing this project have been invaluable. I would also like to thank my committee members, Dr. Marlo Jeffries and Dr. Krista Nichols, whose feedback and guidance have aided in the development of both my project and my confidence. Various others within the Biology Department at TCU have been wonderful examples to me in balancing the stress of education and research. Additionally, I would like to thank my parents, Michael Clare and Leslie Clare, for always believing in me. And finally, I am grateful for the patience, kindness, and support of Justin Lewis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

The movement of animals to exploit seasonally available resources plays an important role in the development, maintenance, and overall health of ecosystems. This process, known as migration, is characterized by an onset of physiological and behavioral changes, and is regulated by a combination of genetic factors and environmental cues (Åkesson and Hedenström 2007; Liedvogel et al. 2011). Although migration is found in many taxa, some species demonstrate a partially migratory life history, consisting of individuals that undergo the migratory process, and also those that remain as non-migratory residents (Chapman et al. 2011). The disparity between these two ecotypes can lead to ecological and adaptive differences that are exclusive to each life history, and selection pressures that are unique to their needs (Winker 2010; Chapman et al. 2011; Gómez-Bahamón et al. 2020). These exclusive events can subsequently lead to evolutionary differences over time, including speciation caused by a behavioral barrier to gene flow (Winker 2010). Human influence in the form of dam construction, climate change, and habitat loss presents additional stressors on protecting and conserving species with divergent life histories and exacerbates the barriers between migratory and resident organisms. This highlights the need to understand the genetic diversity and phenotypic variability within migratory species, as anthropogenic influences are causing many to decline much quicker than their resident counterparts (Wilcove and Wikelski 2008; Liedvogel et al. 2011; Singh and Milner-Gulland 2011).

Migration is a quantitative trait, and consequently, the decision to migrate likely involves many genes of small effect, making it difficult to identify individual genes of interest that determine migratory behavior. Studies are further restricted by the impracticalities of raising migratory organisms in a laboratory, where they are unable to complete their life cycles. Nonetheless, heritability studies have shown that both the timing and propensity to migrate have

additive genetic components; however, the identification of the underlying genes controlling migratory tendency remains elusive (Liedvogel et al. 2011; Lugo Ramos et al. 2017; Merlin and Liedvogel 2019). The advent of next-generation sequencing has presented great opportunity to explore the relationship between genetics and migration, using techniques such as WGS, genome-wide association studies (GWAS), and restriction-association DNA sequencing (RAD-seq). As these technologies become increasingly affordable, it becomes more attainable to identify genetic differentiation and associate it with physiological or morphological differences in migratory or partially migratory species.

Salmonid species such as salmon, trout, and char are an economically important group of fishes that are valued widely for both sportfishing and human consumption, and play an important role in native subsistence fishing. Many traits and phenotypes in salmonids have been found to vary within and between populations of the same species, suggesting high levels of adaptive genetic variation. Because many species within this family are partially anadromous, breeding isolation between landlocked and anadromous populations can contribute to differences in gene flow and significant genetic variation present within their geographical distribution (see Liedvogel et al. 2011; Hardesty-Moore et al. 2018 for review). However, the separation of gene flow can also be location dependent, as salmonids tend to return to their natal stream to breed, preventing them from sharing alleles with different populations, unless they stray (Quinn 1993; Keefer and Caudill 2014). The intraspecific variation present within the family Salmonidae, as well as their relative abundance, their importance, and the observable physical changes present in the individuals undergoing the migratory process makes them an excellent model for studying the genetic basis of migration.

One of the most well-studied of these species is the rainbow trout (*Oncorhynchus mykiss)*, which is present in two distinct ecotypes— the rainbow (resident) and steelhead (anadromous)

trout. Heritability studies of *O. mykiss* have found evidence of a genetic component to migration (Thrower et al. 2004; Hecht et al. 2015), and quantitative trait loci (QTL) studies have found associations linked to multiple smoltification related traits in the species (Nichols et al. 2008; Le Bras et al. 2011; Hecht et al. 2015). The linkage mapping of *O. mykiss* from these QTL studies has provided a framework for whole-genome analyses, which was later enhanced by the use of RAD-seq. RAD tags collected were able to provide evidence that the genetic propensity of migratory traits can vary by population, but the coarse methodology of RAD-seq excludes regions of the genome that are not covered under the RAD-loci produced by the *SbfI* restriction enzyme (Hohenlohe et al. 2010; Hale et al. 2013). Therefore, these findings warrant further investigation using fine-scale methodology, such as whole-genome sequencing (WGS).

Historically, access to genomic resources has precluded identification and understanding of the genes and molecular pathways involved in the evolution of migratory (or resident) phenotypes. However, the 2019 publication of the *O. mykiss* genome (Pearse et al. 2019) provides an opportunity to implement fine-scale methodology and use whole-genome techniques to explore individual genes, mutations, and allele differentiation that contribute to migratory tendency in the rainbow trout. Whole-genome sequencing data can provide a detailed examination of the genetic basis of migration and migration-related traits in different populations subjected to various environmental pressures. Although the QTL and RAD-seq work have provided evidence of population differentiation between the two *O. mykiss* ecotypes (Nichols et al. 2008; Hale et al. 2013; Hecht et al. 2013), specific single-nucleotide polymorphisms (SNPs) have yet to be discovered that repeatedly distinguish migratory populations from resident populations.

The Sashin Creek system, located in Southeastern Alaska, provides a unique opportunity to compare and evaluate differences between anadromous and resident rainbow trout. This site contains a resident population formed by transplanting anadromous steelhead trout above two barrier waterfalls into a freshwater lake (Sashin Lake) nearly 100 years ago (Figure 1). These barriers allow smolted (adolescent silvered migratory fish) individuals to leave the lake but prevent migratory steelhead from returning to spawn, effectively forming a resident population with no subsequent genetic input from migratory individuals. Located below the barrier waterfalls in Sashin Creek is the ancestral population that still has access to the Pacific Ocean and consists largely of migratory *O. mykiss.* Migratory and resident *O. mykiss* in this system have evolved broad genetic differences that appear to be associated with traits connected to migration

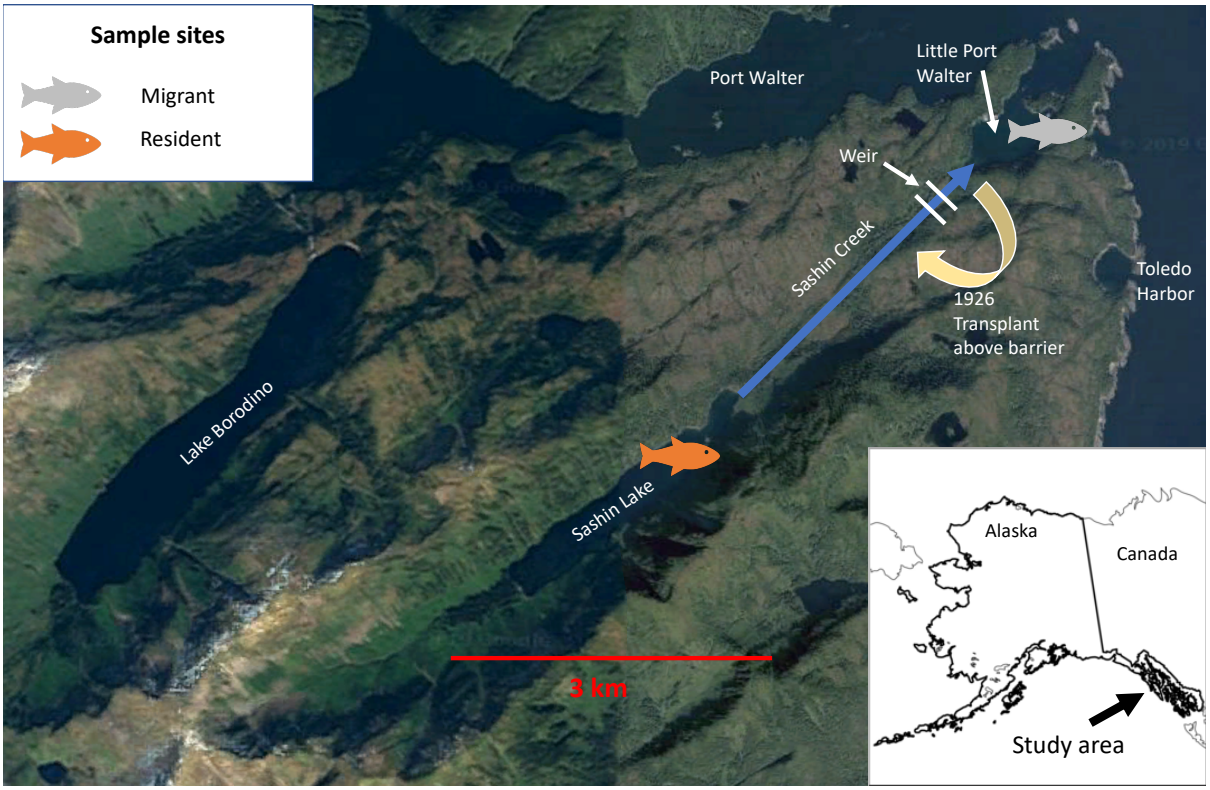

Figure 1 - Map of Sashin Creek and surrounding areas of Southeastern Alaska, USA. Study sites are indicated by fish symbols: migrant (grey) and resident (orange). Location of 1926 transplantation around the weir is shown to indicate the introduction of resident trout to the inland lakes (labeled). Blue arrow indicates the flow of Sashin Creek. Satellite image produced with Google Earth. The inset map presents the state of Alaska, with the study area identified.

and smoltification (Thrower and Joyce 2004; Hecht et al. 2015; Hale et al. 2016; Weinstein et al. 2019). However, the previous studies in this system are coarse in their genomic methods and have not yet utilized a fine-scale genome-wide approach to assess the alleles and genes associated with a tendency to migrate. Additionally, in-depth analyses regarding associations between migratory behavior and genetics have been largely limited to Sashin Creek, making it impossible to test if these same genetic associations are also important in genetic control of migratory populations from other locations. Moreover, the Sashin Creek site has been subjected to an unnatural selection event (i.e., the movement of migratory steelhead to Sashin Lake) which makes it unclear the extent to which the allele differentiation at this site is representative of natural populations.

To address these questions, we have used pooled sequencing (Pool-seq) to evaluate the shared genetic component in life-history type amongst geographically separated populations of *O. mykiss*. These study sites include the well-studied population of Sashin Creek, AK (Thrower et al. 2004; Hale et al. 2013; Weinstein et al. 2019) and a site in Little Sheep Creek, OR which has previously been investigated and compared to Sashin Creek using RAD-seq (Hale et al. 2013). The Little Sheep Creek (LSC), site located in Oregon, consists of an inland stream where a migrant population returns yearly to breed, and a resident population that remains in the surrounding stream year-round (Berntson et al. 2011). We then used the Pool-seq information to assess population differentiation and nucleotide diversity, and identify individual alleles that may contribute or correlate with the migratory life history type in both Sashin and LSC. Statistical tests were then used to identify locations within the genome where selection is acting on migratory individuals and to determine the type of selection event occurring. Our final objective was to compare the two populations to determine the extent to which specific alleles, genes, and chromosomal regions associated with resident and migratory phenotypes are shared between

multiple populations. Together, these results contribute to a body of information dedicated to understanding the shared and population-specific genetic control of migratory behavior in *O. mykiss*.

Methods

Sampling & DNA Extraction

Rainbow and steelhead trout (*Oncorhynchus mykiss*) were sampled from two populations: one in Sashin Creek, Alaska and one in Little Sheep Creek, Oregon. Migratory phenotype was determined by using physical traits, including silvery coloration, body shape, and the production of gametes, as sexual maturity precludes anadromy in rainbow trout (more sampling details can be found in Thrower et al. 2004 and Berntson et al. 2011). In the Sashin Creek system, samples were collected in the form of fin clips, and LSC samples included both fin clips and operculum samples. In the Sashin system, all residents were sampled from Sashin Lake and thus, were separated from migrants by two barrier waterfalls. All migrants sampled were adults returning to spawn downstream of the waterfalls. This location separation minimizes admixture between the two ecotypes (Thrower et al. 2004). In the LSC system, migrants and residents coexist in the same water system, so accurate observational phenotype determination was critical in classifying life history type. To minimize the risk of premature juveniles being sampled, only adult fish with fork lengths (tip of snout to the fork of the caudal fin) greater than 150 mm and that were expressing gametes were considered residents (Berntson et al. 2011). To ensure only wild fish, and not hatchery fish, were collected for sampling, only fish with the full adipose fin were collected, as it is standard practice to remove the adipose fin in hatchery fish before release. At sampling, the fork length and sex of the fish were recorded, and the caudal fin or operculum sample was removed and stored in 95% ethanol. Fish were immediately released after

processing. Samples from the Sashin system were collected in 2017 and 2018, and samples from the LSC system were collected in 2011.

Pooled Sequencing Library Preparation

DNA was extracted from 174 fish - 40 Sashin Creek migrants, 40 Sashin Creek residents, 48 LSC migrants, and 46 LSC residents using DNeasy Tissue Extraction kits (QIAGEN Corporation, Valencia, California). Each pool consisted of an equal number of male and female individuals. Pooled sequencing was used to provide a cost and time efficient way of sequencing many individuals with high depth of coverage. The utility of Pool-Seq methods in non-model organisms has been discussed in detail in Micheletti & Narum (2018). Pool-Seq involves combining DNA from multiple samples of the same group into a singular pool and then sequencing the pooled mixture as a representation of the population-level allele frequencies (Schlötterer et al. 2014). To ensure adequate representation of each individual in the pool, each DNA sample was normalized to a standard of 300 µg/L before homogenization and then confirmed via Nanodrop. Four DNA pools were prepared in total, one for each of the corresponding groups – Sashin migrants (SM), Sashin Residents (SR), Little Sheep Creek migrants (LSM), Little Sheep Creek residents (LSR). Pools were then sent to the NorthWest Genomics Center at the University of Washington (Sashin pools) and Novogene (LSC pools) for paired-end 150 base-pair sequencing on Illumina NovoSeq.

Pool-Seq Alignment & Filtering

Sequences were analyzed following the *PoolParty* pipeline (https://github.com/StevenMicheletti/poolparty), as detailed in Micheletti and Narum (2018) with some custom modifications. Sequences were first quality filtered to remove scaffolds with poor quality bases (Q values < 20) or that measured < 50 base pairs in length. Quality filtered

sequences were then aligned to the rainbow trout genome (Omyk_1.0; GCA_002163495.1) using *bwa mem* with default settings (Li 2013). Duplicate sequences were removed using *samblaster* (Faust and Hall 2014), and *samtools* (Li 2011) was used to compile alignments and filter the reads to a minimum length of 36 and a quality score > 30 (Kofler et al. 2016). Alignments were then converted to the SAMtools *mpileup* file format, which summarizes total coverage information for each pool (Li et al. 2009) and then combined to a synchronized (sync) file to be used in *Popoolation2* (Kofler, Pandey, et al. 2011) and *PoPoolations* (Kofler, Orozco-terWengel, et al. 2011). Each of the four pools was then run through filtering criteria, and positions were retained for use in statistical calculations if they met the minimum threshold of 25 X depth of coverage, and maximum threshold of 250 X depth of coverage to reduce the chances of paralogs and overrepresentation, respectively.

Bioinformatics & Statistical Calculations

Using the *Popoolation2* pipeline (Kofler, Pandey, et al. 2011), Fixation index (F_{ST}) values were calculated to determine regions with high differentiation of resident and anadromous populations in both Sashin and LSC. Individual F_{ST} was calculated for each position within the genome, with a filtering criterion of an 8 read minimum in each position (min-count). This value was first calculated for each SNP individually to find SNPs with complete fixation ($F_{ST} = 1$) between pooled samples from the same site, and to identify SNPs with the most differentiation in each population. A mean F_{ST} approach was then used in both populations over a 500 base pair region with a step size of 500 using *Popoolation2* to identify genomic regions with high average differentiation (Kofler et al. 2011; Micheletti and Narum 2018). These values were filtered using a min-count of 6. To evaluate nucleotide diversity in each of the populations, a Tajima's D statistic was calculated over a 10,000 base pair window, with a step size of 10,000 for

anadromous and resident Sashin and LSC pools using the variance program in *PoPoolations* (Kofler, Orozco-terWengel, et al. 2011) with a min-count of 2 and min-qual of 20.

Pooled Sequencing Validation

Due to multiple historic whole-genome duplications, PCR and Sanger sequencing results were unable to be used for validation by confirming SNP presence on the primer products using SNaPshot or sequencing methods. We confirmed that 4 fixed values ($F_{ST} = 1$) tested via PCR and Sanger sequencing were in duplicated regions of the genome. Therefore, the primer products were likely from paralogous regions. The Pool-seq results were instead validated using genotype data from 30,644 RAD-tags previously gathered from the Sashin population (Hale et al. 2013). RAD-tags were aligned to the *O. mykiss* genome, and scaffolds with a singular match (>95% similarity) were used to calculate estimated allele frequencies in 25 locations of high F_{ST} (> 0.50). Allele frequencies were then estimated from genotype frequencies using the Hardy Weinberg equation ($p^2 + 2pq + q^2 = 1$ and $p + q = 1$). We used regression analysis with an ANOVA to confirm the association between Pool-seq and RAD-seq allele frequencies. Outliers were then detected by calculating the mean residual value for each group ± the standard deviation.

Results

The Illumina sequencing of the pooled samples returned a variable depth between 15 X and 250 X, with an average coverage depth of 29.661 reads for Little Sheep Creek (LSM and LSR) and 115.110 reads for Sashin (SM and SR) populations. Total number of quality filtered reads, as well as average coverage depth are reported for each population – SM, SR, LSM, and LSR in Table 1. A comparison of differentiation between resident and anadromous individuals was done

Table 1 – Sequencing coverage of paired-end Illumina sequencing of rainbow and steelhead O. mykiss from Sashin Creek, Alaska and Little Sheep Creek, Oregon. Total number of Quality Filtered (QF) reads (between 15X and 250X coverage, quality filter of 20, and scaffold length minimum of 50) is reported for each of the 4 populations, as well as the average depth of coverage, with the published O. mykiss genome size of 2.1 GB

Location	Ecotype	Total # QF Reads	Avg. Coverage Depth
Sashin	Rainbow	812,041,814	116.006
Sashin	Steelhead	799,458,233	114.208
Little Sheep	Rainbow	193,742,578	27.678
Little Sheep	Steelhead	221,513,116	31.645

by calculating the F_{ST} value for each SNP in Sashin and LSC populations (Figure 2). In the Sashin population, 41 loci contained fixed differences ($F_{ST} = 1$) between migrant and resident individuals (mean $F_{ST} = 0.059 \pm 0.076$). There were no fixed differences identified in the LSC population (mean $F_{ST} = 0.029 \pm 0.039$). The top 1% F_{ST} quantile began at 0.355 and 0.186 and the median F_{ST} was 0.032 and 0.014 for Sashin and LSC, respectively. Sashin Creek contained a greater density of polymorphic sites (F_{ST} density) throughout the genome, likely due to both a higher depth of sequencing coverage and reduced gene flow between the two ecotypes. Regions of high F_{ST} density in both populations were identified at 81.6 Mb in chromosome 1, between 9 and 10 Mb in chromosome 2, 22 – 25 Mb in chromosome 14, 8-9 Mb in chromosome 18, and positions 44 – 46 Mb in chromosome 22. Shared regions of low F_{ST} density were also identified and noted if outside of regions of low genome coverage, as they could indicate regions of high genetic conservation within the species.

Genome regions that showed a high individual F_{ST} in both Sashin and LSC were further investigated for the presence of protein-coding genes that might be the subject of selection (Table 2). To limit the total number of SNPs considered, only those with a minimum F_{ST} statistic of 0.5 for Sashin, and 0.25 for LSC were investigated. Of these, only one SNP, located on

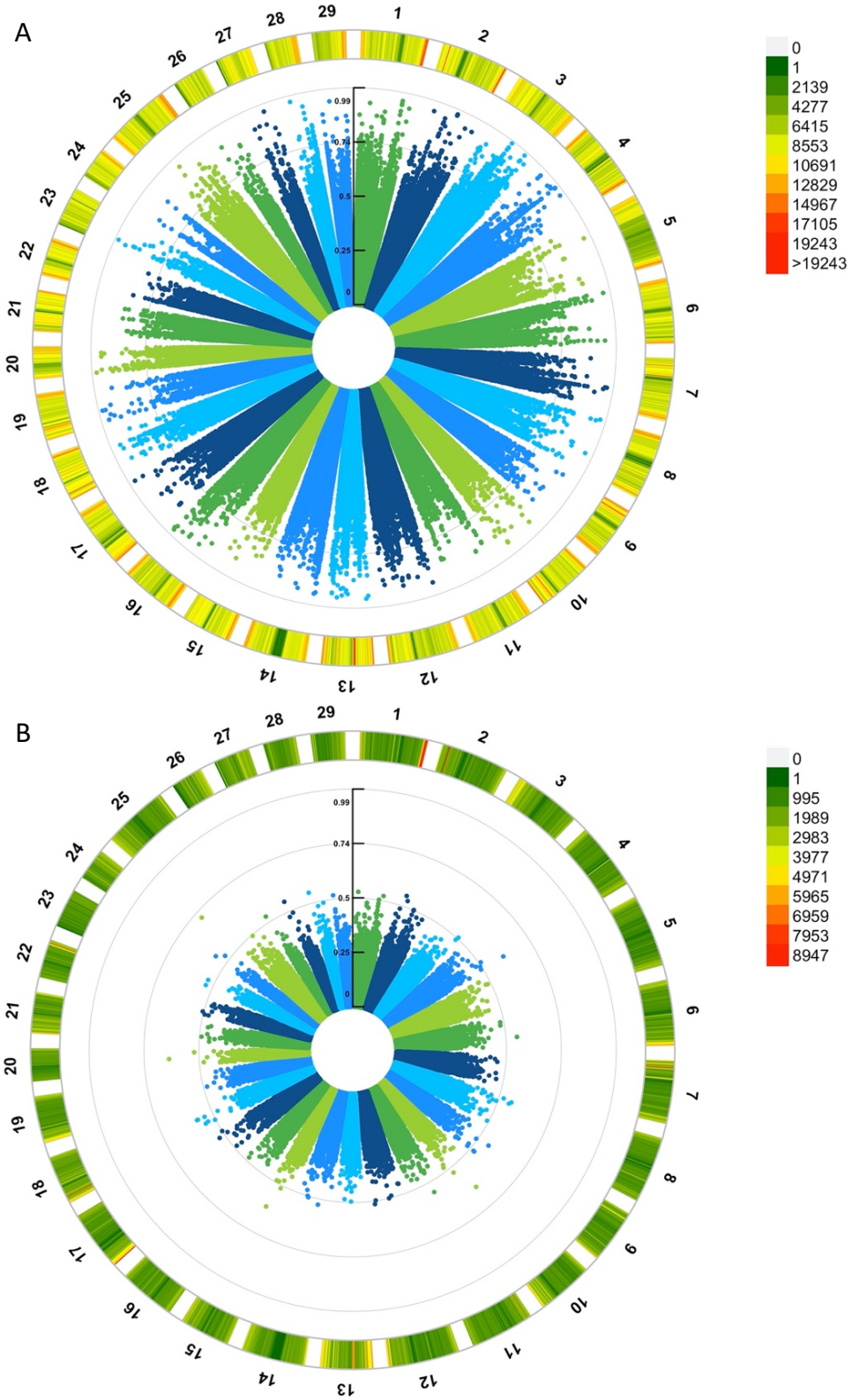

Figure 2 - Individual fixation index (F_{ST}) values for resident vs. anadromous *O. mykiss* of A) Sashin Creek and B) Little Sheep Creek on a circular Manhattan plot. In each plot, the inner circle represents each individual F_{ST} value, and the outer ring displays F_{ST} density per 1 Mb window size at each of the 29 chromosomes (green-low, red- high).

Table 2 – Chromosome and base pair locations of shared elevated F_{ST} between anadromous and resident populations. F_{ST} values are listed for both Little Sheep Creek (LSC) and Sashin (Sash) populations. If the polymorphism was found within a known gene, the gene name is given.

CHR	BP	LSC F_{ST}	Sash F_{ST}	Gene (if present)
1	13331	0.297	0.509	phosphatidylinositol-specific phospholipase C, X domain containing 1
1	69898903	0.291	0.560	-
3	74564628	0.316	0.684	-
4	34759942	0.275	0.930	leptin receptor
5	80722208	0.278	0.520	DnaJ homolog, subfamily C, member 6
6	20756423	0.334	0.550	-
6	20755871	0.263	0.597	-
7	21949457	0.313	0.563	-
8	17055130	0.416	0.605	tumor protein D53 homolog
13	37769240	0.255	0.601	cAMP-specific 3',5'-cyclic phosphodiesterase 4B
19	19454334	0.256	1	catenin alpha-2
20	20645201	0.280	0.558	transformation/transcription domain-associated protein
22	47499934	0.268	0.519	-
25	51690773	0.335	0.503	RNA polymerase II associated protein 1
25	66224036	0.289	0.510	-
26	14291145	0.333	0.545	-

chromosome 25, was within the protein-coding region of a gene. A mean F_{ST} statistic was calculated between residents and migrants in Sashin and LSC over 500 base-pair windows across the genome to allow for a more direct comparison between groups (Figure 3). This statistic, used in conjunction with individual F_{ST} and F_{ST} density, allowed for the identification of regions or areas that might be of greater interest in both populations. Tajima's D statistic was then calculated over a 10,000 base pair window across each of the 4 populations (Figure 4) to determine the type of selection event occurring at that position.

The Tajima's D statistic is calculated by comparing the number of expected values to the number of observed values over a given window. A Tajima's D value of zero indicates neutral selection pressures, a negative value indicates purifying selection, and a positive value indicates balancing selection. Any value that is greater than two or less than negative two is deemed

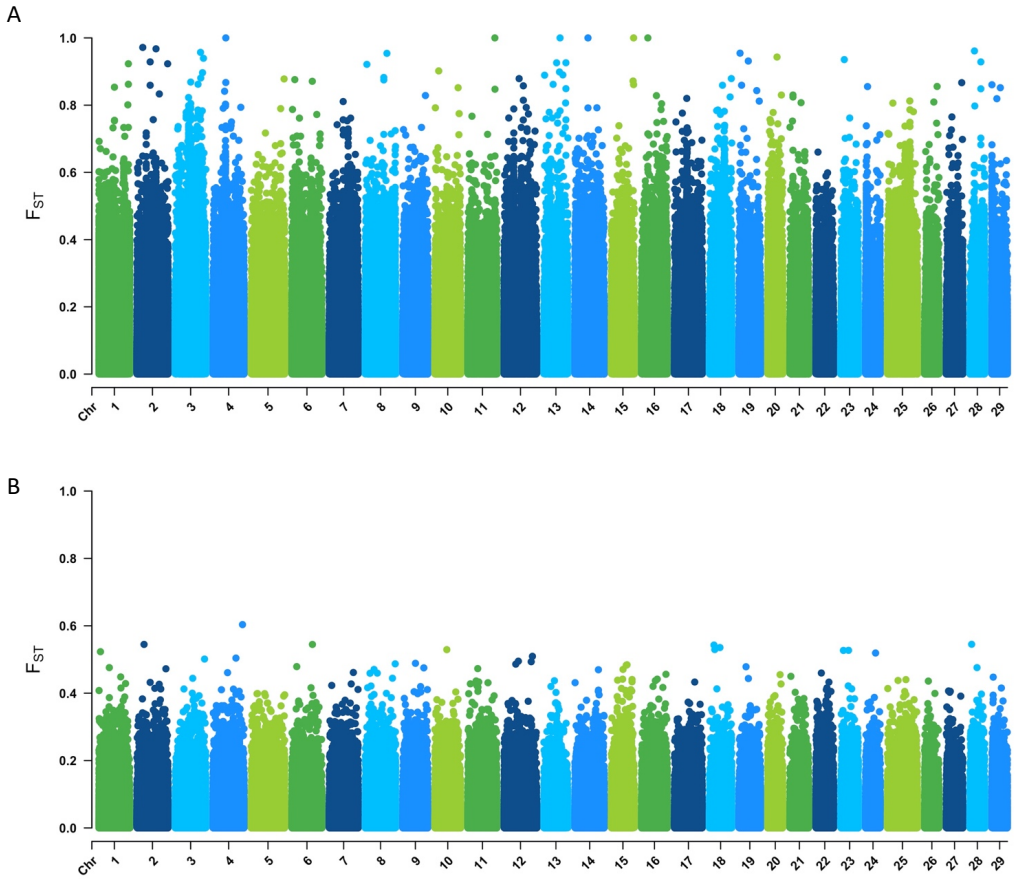

Figure 3 – Mean estimated fixation index (F_{ST}) between residents and migrants of A) Sashin Creek and B) Little Sheep Creek calculated over a 500 base pair window.

significant (Tajima 1989), and we used this threshold to identify 112 significant regions in the LSR population. In the SR population, we identified 24,192 regions using the same parameter. Between the two resident groups, 94 regions shared a significant value over the same window, and all were negative. In the migrant groups, we identified 222 significant Tajima's D values in the LSM pool, and 8853 in the SM pool. Of these, 151 were shared in the same window in both pools, and also were all negative. There were additionally 21 locations where the Tajima's D value was below negative two in all four populations, leaving 73 regions where there is evidence of a selection event occurring in the resident populations, but not the migrant populations, and 130 regions where there is evidence of selection acting upon migrants, but not residents.

We also identified areas where there were significant Tajima's D values in adjacent shared windows (+/- 10,000 base pairs), which indicates a large selective sweep that potentially spans more than 20,000 base pairs. Any genes in these regions were noted for the protein function and related pathways. We additionally used this data to identify trends in both resident

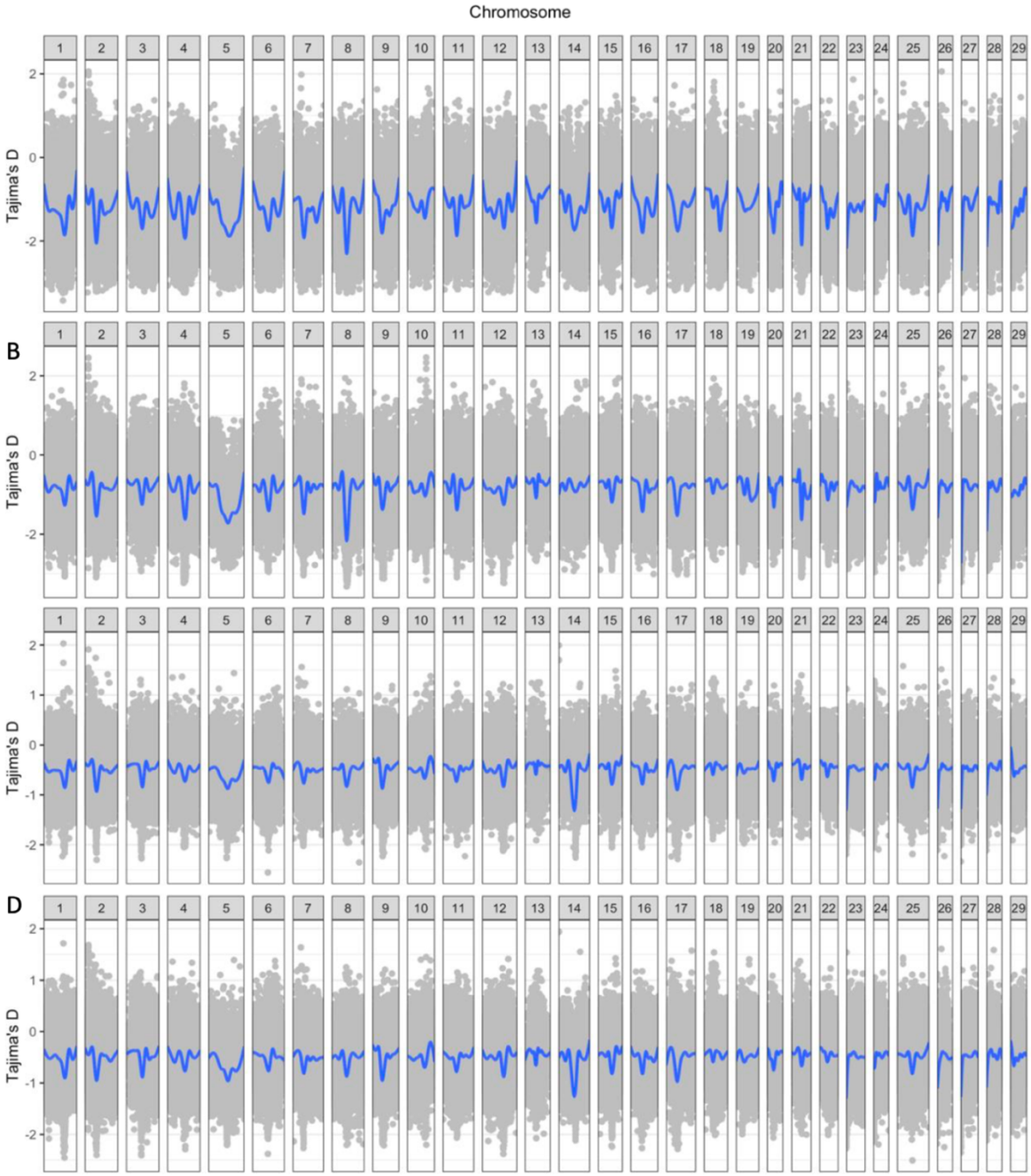

Figure 4- Tajima's D estimated statistic for A) Sashin residents, B) Sashin migrants, C) Little Sheep Creek residents, and D) Little Sheep Creek migrants plotted in grey. Tajima's D was estimated over a 10Kb sliding window, across the genome, and the smoothed average is shown by the blue line in each population.

and migrant phenotypes where large regions of significant Tajima's D values were found close together, indicating a shared selection event acting on the phenotype. For the resident populations, these large regions of selection were located near Chr2 – 29-30 Mb, Chr3:38-39 Mb, Chr5:46-47 Mb, Chr9:26-27 Mb, Chr12:53-55 Mb, Chr14:43-44 Mb, Chr15:36-37 Mb, and Chr17:28-30 Mb. In the migrant pools, we identified regions near Chr1:50-52, Chr2:28-30, Chr5:47-48, Chr9:26-27 Mb, Chr15:35-37 Mb, Chr17:28-30 Mb, and Chr27: 0.3– 1 Mb.

Validation of Pooled - Sequencing

Regression analysis confirmed equal values between RAD-seq allele frequency and Pool-seq allele frequency Sashin with an R^2 value of 0.887. An analysis of variance (ANOVA) achieved a p-value of 2.213E-12. In the resident population, the mean residual value was -3.937E-05, with a standard deviation of 0.115, validating 25/25 RAD-tags lying within 2 standard deviations of this mean residual value. The migrant population had a mean residual value of 2.566E-6, and a standard deviation of 0.0968, validating 24/25 migrant RAD-tags using our sequenced libraries (Figure 5).

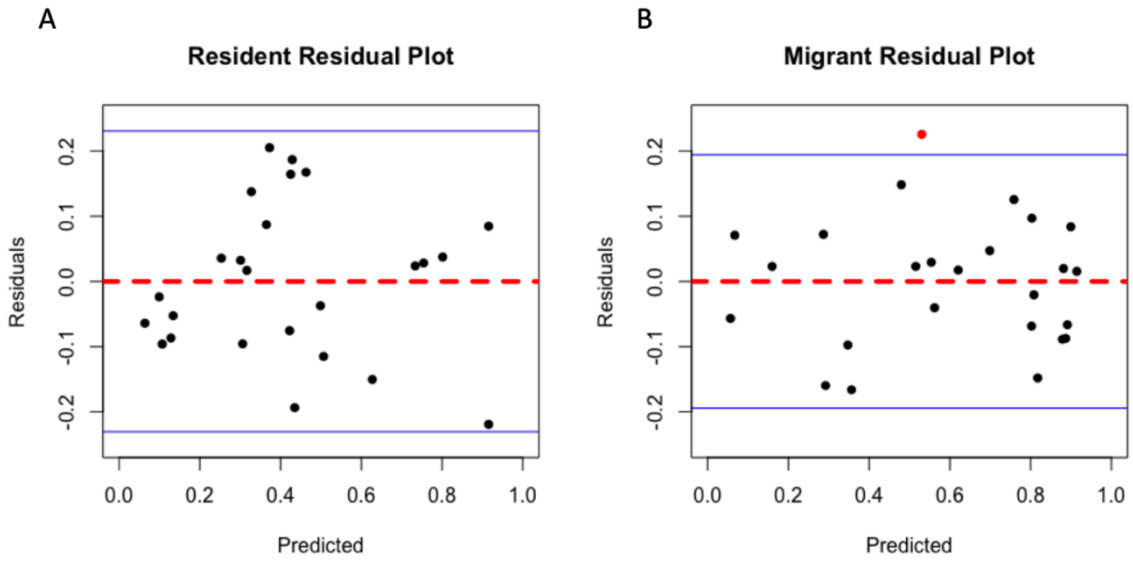

Figure 5 - Residual plots from the linear regression of A) Resident and B) Migrant allele frequencies vs. allele frequencies calculated from RAD-tag genotype frequencies (Hale et al. 2013). The red dashed line indicates the optimal value 0 (no deviation), the solid blue line shows the value of 2 times the standard deviation in each group. Any outliers are shown in red.

Discussion

The Sashin Creek system has become a model population for studying the genetic basis of anadromy in rainbow trout. Although associations between anadromy and specific genes of interest have previously been uncovered, these findings have only minimally been compared to other populations of rainbow trout (but see Hecht et al. 2013; Arostegui et al. 2019). Our use of WGS in both Sashin and LSC implements a fine-scale approach to associating genetics and migratory behavior, while creating an opportunity to assess the previous findings for relevance in other geographical areas (Thrower and Joyce 2004; Hale et al. 2013; Hecht et al. 2015; Weinstein et al. 2019). Our findings indicate that population- specific effects play a major role in the regulation of migratory tendency in *O. mykiss*. We were able to identify shared regions where selection events are occurring, both between and among the two locations and their respective ecotypes. Additionally, we identified regions of selection that were acting on either Sashin or LSC independently, indicating that geographical location, and not migratory tendency, has a larger effect on genetic diversity in *O. mykiss*.

Population Differentiation in Sashin and Little Sheep Creek

Due to the isolation and subsequent lack of gene flow between phenotypes in Sashin, we saw a higher degree of genetic differentiation between ecotypes in Sashin than in LSC. The F_{ST} values of LSC were far lower than those in Sashin, suggesting that migrant and resident individuals can and indeed do interbreed in LSC. The degree of differentiation between sites aligns closely with the RAD-seq results of Hale et al (2013), which used the same two study sites. Enhancing these findings with WGS allowed a more thorough genetic understanding of the degree of differentiation within these populations, and further suggest that the high degree of differentiation between life histories in Sashin Creek – and the lack of differentiation in LSC – is due to genome-wide genetic effects. The findings of Hale et al (2013) implicated two regions

with relatively high F_{ST} (> 0.20) located on chromosomes 2 and 7 in Sashin, but we did not see isolated peaks in those regions due to the increased density of F_{ST} values when using WGS. Since we used a cutoff F_{ST} value of 0.5, any values implicating differentiation to this extent would have been excluded as a non-outlier region. Another prior study of the Sashin population used a common garden experimental design and genotype information from the 57k rainbow trout SNP-chip and identified elevated F_{ST} in chromosomes 17 and 26 between resident and anadromous crosses, but the highest of these values was also below our threshold (Palti et al. 2015; Weinstein et al. 2019). This does not indicate that these regions are without importance in the question of migratory behavior, but just that these positions were not considered outliers in the scope of our study. Instead, the course methods of this prior research have identified individual regions that may be contributing to migratory behavior through genes of small effect, which are difficult to associate to phenotype using WGS methods.

We identified 42 positions where fixed F_{ST} values were present in the Sashin populations (F_{ST} = 1), of which 20 were located within protein-coding gene regions. Four of the 42 fixed positions were found within a 6000 base-pair region, located on the gene *leptin receptor-like* (*lepr*). Leptin hormones modulate appetite and energy use in mammals, and there is increasing evidence that the anorexigenic function is retained in salmonids (Murashita et al. 2008; Gong et al. 2013). Alterations in leptin receptor function or binding ability between the receptor and leptin could be advantageous during extended periods with low food availability, and recent research suggests that this hormone plays a role in migratory success (Fuentes et al. 2012; Choi et al. 2014). Four more fixed positions were found within a 3000 base-pair region located within the gene *myosin heavy chain, fast skeletal muscle-like*, two of which were found in the exon region. Myosin heavy chain proteins facilitate the conversion of ATP to ADP, functionally converting chemical energy into mechanical energy. Evidence of environmentally dependent

selection of myosin- related genes in poikilotherms suggests that temperature affects stability and functionality of this protein product (Watabe et al. 1997; Iwami et al. 2002). Differences in genetic code between phenotypes could be reflective of a selective sweep aligning with these adaptations.

Our investigation and comparison of the top F_{ST} values from Sashin and LSC uncovered 16 SNPs with elevated differentiation in both populations. Although eight SNPs were located within protein-coding genes, only a singular value was found within a protein-coding exon region, on *pap1 RNA polymerase II associated protein 1*. However, this polymorphism was both synonymous and reversed in each population, with the resident genotype in Sashin corresponding with the migrant genotype in LSC. Although this SNP likely has no effect on migratory tendency, it provides insight into the fragile network of regulatory mechanisms present when dealing with many genes of small effect. Of the remaining 15 SNPs with shared elevated F_{ST}, six positions were found to have the same distribution in both populations, with four of these in gene regions. Although these four SNPs were not found within known protein-coding regions, they could have regulatory effects on gene expression and therefore consequential effects on protein products. The genes that these SNPs were located within showed a wide range of functions, including metabolism & protein synthesis, which could have downstream effects on either the smoltification process or regulation of processes required for migratory (or resident) behavior (Stefansson et al. 2012; McCormick 2013). Specifically, the gene *phosphatidylinositol-specific phospholipase C, X domain containing 1* (*plcxd1*) encodes for phospholipase C involved in hydrolyzing phospholipids into fatty acids and other lipophilic molecules in humans (reviewed by Fukami 2002). Evidence of metabolic influence on migration is vast, indicating that standard metabolic rate may be a required threshold for smoltification in female *O. mykiss*, as it is associated with dominance rank, food acquisition, and energy use (Sloat and Reeves 2014).

Although the exact ontology pathway is unknown, smolting salmonids alter their metabolic state in preparation for movement into seawater (Stefansson et al. 2012; McCormick 2013). This is generally accompanied by a decrease in fatty stores throughout the body, causing the change in condition factor that occurs in migratory salmonids before they leave freshwater (McCormick and Saunders 1987; Sheridan 1989).

We identified a second association between anadromy and leptin within the gene *leptin receptor*, which contained elevated F_{ST} on the same SNP in both Sashin and LSC. This further suggests that leptin could be important in regulating migratory behavior. Both genes encode for proteins that function as receptors for the adipocyte hormone leptin, and evidence of its role in regulating hunger and homeostasis has been documented in the Arctic charr (*Salvelinus alpinus*) and chum salmon (*Oncorhynchus keta*; Jørgensen et al. 2013; Choi et al. 2014). It is known that there is a significant hormonal cascade that occurs during the onset of smoltification (McCormick 2009), but the role of leptin and its influence on migratory behavior is poorly understood. However, female smolted chum salmon exposed to freshwater produced a significant increase in both the activity of leptin and expression of related mRNA transcripts, indicating the role of leptin in both sexual maturation and the migratory process (Choi et al. 2014). In the rainbow trout, leptin has been associated with the management and utilization of energy stores (Johansson et al. 2016). Leptin manages energy metabolism and body weight, both of which have been shown to positively correlate with migratory behavior through QTL studies (Nichols et al. 2008; Hecht et al. 2015). Although the SNPs in the leptin-related genes and *plcx1* were not found in protein coding regions of the *O. mykiss* genome, the regulation of both genes and their expression could have unknown effects on modulating metabolic and hormone processes.

Nucleotide Diversity in Sashin & Little Sheep Creek

The measures of nucleotide diversity indicated strong departures from neutrality for SR, indicating that purifying selection likely plays a role in the structure and function of the genomes in this group. The negative Tajima's D values indicate a lower allelic diversity in certain regions of the genome, which corresponds to the increase in rare alleles we would expect to see in SR due to the bottlenecking event after the transplant in 1926 (Thrower and Joyce 2004). The result of our assessments of Tajima's D were less varied than previous studies (Hale et al. 2013; Weinstein et al. 2019). However, the positive values seen throughout the majority of the genome in the SNP-chip assessment of Weinstein et al. (2019) were likely due to the experimental design of genotyping the offspring of inbred progeny. Additionally, our assessment of nucleotide diversity included the entire genome, which was not able to be assessed in either prior study and provides better resolution of both the type and location of selection events. We then expect that other non-isolated sites would have Tajima's D values comparable to the neutrality observed in LSC, indicating a limitation of use of the Sashin Creek site.

Both Sashin populations (SM & SR) contained multiple genomic regions where the Tajima's D value was below -3, indicating strong purifying selection resulting in the reduction of genetic variation. Because more neutral values were detected in both LSC pools (LSM & LSR), the smooth-line trends were observed to be more similar based on geographic location and not ecotype. In populations where breeding sites are shared between ecotypes, such as LSC, phenotypic plasticity allows individuals to retain and adapt to their particular life history type (Dodson et al. 2013; Doctor et al. 2014). We then assume that the disparity in Tajima's D values between the two locations is due to a combination of both stronger selection on the resident phenotype to not produce migratory traits, and a bottlenecking event in the Sashin system (Thrower et al. 2004).

A comparison of smooth-line trends in Tajima's D values between all 4 pools was used to identify regions where differences in selection were present between populations, and two large visual differences were identified between LSC and Sashin on chromosomes 8 and 12. Although closer inspection showed that the negative trend on chromosome 12 was due to lack of sequencing coverage in the area, chromosome 8 demonstrated purifying selection occurring near the centromere, resulting in a large portion of potentially linked SNPs being inherited together in both Sashin populations (Figure 6). This region contained 76 genes connected to various physiological and biochemical processes. Prior studies at Sashin have not identified any significant genetic differentiation in this chromosomal region, but QTL studies in other populations have identified a few traits that are localized to chromosome 8, including body morphology, and increased growth and weight during the spring season (Nichols et al. 2008). Although no previous association has been found in Sashin, it is possible that selection on this region is connected to growth and weight that is not mirrored in LSC. The Sashin Creek system is higher in latitude than LSC and therefore sensitivity to photoperiod and a need for rapid growth in the short Alaskan summer may explain why selection operating in this region of the

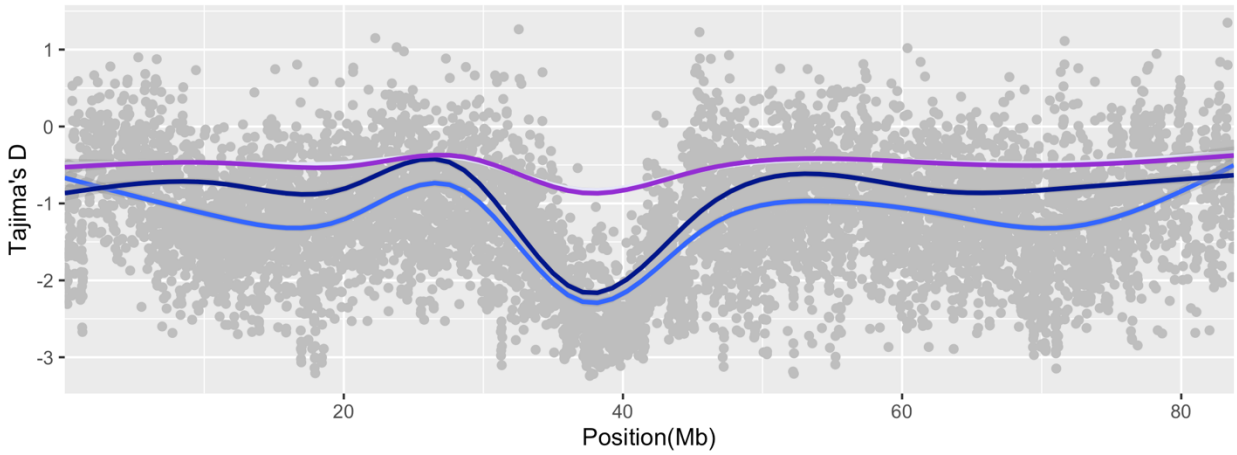

Figure 6 – Whole – chromosome Tajima's D values for chromosome 8. Each point represents a singular Tajima's D value over a 10k window for the Sashin Resident population. Each line represents the smoothed line average Tajima's D for Sashin Rainbow (blue), Sashin Migrant (navy), Little Sheep Rainbow (lavender), and Little Sheep Steelhead (violet).

genome is stronger in Sashin than in LSC. Indeed, RNA-seq experiments using Sashin Creek samples find consistent differential expression of several genes involved in photoperiod recognition and phototransduction (McKinney et al. 2015; Hale et al. 2016). Although it is currently unknown if these results are also found in other southern migratory populations, the differences in selection pressures observed through the Tajima's D statistic provide further evidence of the adaptation of *O. mykiss* to specific geographical areas.

Nucleotide diversity between resident & migrant pools

Both anadromous and resident rainbow trout breed annually in their natal freshwater streams across the Pacific Northwest. The isolation of different breeding sites of *O. mykiss* creates a gradient by which certain groups are isolated from each other, and therefore are not mixing or sharing many alleles. However, it is not uncommon for migrating salmonids to return to the wrong stream (reviewed by Quinn 1993), and there are many locations where steelhead and rainbow trout share a breeding site. Consequently, it is more likely for selection events within migratory species to be present widely across the population, while in landlocked resident populations we would expect to see a higher rate of isolated evolution of characteristics that are not shared in other geographic locations due to bottlenecking (reviewed by Keefer and Caudill 2014). For this reason, we identified regions of convergent selection by first looking at similarities and differences between the resident populations (SR & LSR), as they are more isolated from each other, and experience less gene flow than migrants.

Comparing the resident populations, there is a clear deviation between Sashin and LSC Tajima's D values, corresponding to the differences in life histories between these groups. We then looked at regions of the genome that showed significant negative Tajima's D values in both resident pools (SR and LSR), that were not mirrored in migrant pools (Table 3). The identified genes were not characteristically related to residency but could be important in pathways that

Table 3 – Locations & values of significant ($\theta < -2$) Tajima's D values that are shared among resident pools, but not found to be significant in migrant pools ($\theta > -2$). Tajima's D values were calculated over a 10 Kb window. Regions were included if a window larger than 20 Kb showed greater than 50% significant coverage, and the value was not found to be significant in either migrant population. All genes located within each region are included, as well as the function of the protein product.

CHR	Mb	Gene	Abb.	Protein Function	Tajima's D – Sashin	Tajima's D - LSC
3	38.86-38.88	protein tyrosine phosphatase receptor type Ub	ptprub	Regulation of cell growth	-2.540 - -2.549	-2.019 - -2.119
3	39.00 – 39.02	-	-	-	-2.620 - -2.749	-2.002 - -2.022
3	39.12-39.15	-	-	-	-2.455 - -2.458	-2.045 - -2.141
3	39.39 – 39.41	zinc finger 3	hivep3	Transcription factor	-2.338 - -2.766	-2.050 - -2.080
5	47.91-47.94	arf-GAP with coiled-coil, Ankyrin repeat and PH domain-containing protein 2	acap2	Endocytosis & cell signaling	-2.476 - -2.740	-2.080 - -2.122
12	53.82-53.84	14-3-3G1 protein	-	Cell regulation	-2.063 - -3.046	-2.086 - -2.117
		scavenger receptor cysteine-rich domain-containing group B protein	ssc4d	Scavenger receptor activity		
15	36.06-36.09	Mab-21 Domain Containing 2	mb21d2	unknown	-2.701 - -3.037	-2.003 - -2.051
		fibroblast growth factor 12	fgf12	Cell survival		
17	29.79-29.82	histone deacetylase 7	hdac7	Histone activation	-2.034 - -2.450	-2.054 - -2.094

regulate body changes or metabolic processes required for life in freshwater. For example, regulation of cell and tissue growth is important in homeostasis (McCormick and Saunders 1987; McCormick 2009), and transcription and signaling factors could have a great effect on the proteins the individual produces and the way it responds to intracellular signals (Bastow et al. 2004; Baerwald et al. 2016). It is also understood that the smoltification process is largely controlled by hormones, and alterations to these pathways by gene ubiquitination or transcription factors could affect the ability of certain hormones to bind, resulting in a cascading effect on physiological changes (McCormick 2009; Stefansson et al. 2012). None of the genes located within regions of significant Tajima's D had previously been found to be associated with migratory or resident tendency in *O. mykiss*, but *fibroblast growth factor 12* (*fgf12*) and *histone deacetylase 7* (*hdac7*) are connected in function to two genes (*fgfrl1* and *hdac11*) previously shown by Hale et al. (2016) to be differentially expressed between smolts and juvenile residents. These parallels could be indicative of the importance of cell signaling roles of fibroblast growth factors and potential histone modification caused by histone deacetylases. This could implicate epigenetic regulation or gene - environment interactions as significant contributing factors in the determination of life history type.

Convergent selection in migrants was identified by isolating regions of significant Tajima's D in Sashin and LSC migrant pools (SM and LSM) that were not mirrored in resident pools (Table 4). As with the resident pools, all identified significant regions contained a negative Tajima's D in both populations, indicating purifying selection events occurring at these loci, resulting in the removal of deleterious alleles that are disadvantageous for the migratory phenotype (Tajima 1989). The nine genes identified encode for proteins with functions varying from protein interactions and binding, to ion-exchange capacity. A few of the genes identified

Table 4 - Locations & values of significant ($\theta < -2$) Tajima's D values that are shared among migrant pools, but not found to be significant in resident pools ($\theta > -2$). Tajima's D values were calculated over a 10 Kb window. Regions were included if a window larger than 20 Kb showed greater than 50% significant coverage, and the value was not found to be significant in either resident population. All genes located within each region are included, as well as the function of the protein product.

CHR	BP	Gene	Abb.	Protein Function Pathway	Tajima's D – Sashin	Tajima's D - LSC
2	30.10-30.13	ankyrin repeat domain-containing protein 26	ankrd26	Protein-protein interactions	-2.370 - -2.882	-2.035 - -2.063
2	30.70-30.75	BR serine/threonine kinase 2a	brsk2	Polarization of neurons and axonogenesis, cell cycle progress and insulin secretion	-2.435 - -2.897	-2.007 - -2.044
5	47.36-47.39	adhesion G-protein coupled receptor D2	agrd2	Receptor & transmembrane signaling activity	-2.261 - -2.676	-2.027 - -2.035
9	26.11-26.17	ethanolamine-phosphate phospho-lyase	etnppl	Metabolism & glycerophospholipid biosynthesis	-2.339 - -2.620	-2.078 - -2.159
		collagen alpha-1(XXV) chain*	col25a1	Integrin pathway & extracellular matrix degradation		
9	29.36-29.38	serine/threonine-protein kinase WNK2	wnk2	Molecule transfer & transferase activity	-2.633 - -3.016	-2.077 - -2.101
23	0.43-0.46	stromal cell-derived factor 1	sdf1 (cxcl12)	receptor binding & chemokine activity	-2.047 - -2.874	-2.002 - -2.053
27	0.47-0.50	-	-	-	-2.471 - -2.586	-2.229 - -2.261
28	0.42-0.45	Krueppel-like factor 2*	klf2	Transcription factor	-2.367 - -2.906	-2.001 - -2.119
		epidermal growth factor receptor substrate 15-like 1	eps15l1	Calcium ion bonding		

can be directly associated with migratory tendency. For example, the standard metabolic rate is often different between anadromous and resident salmonids (Sloat and Reeves 2014), so the role of *ethanolamine-phosphate phospho-lyase* (*etnppl*) as a metabolic agent could be related to these adaptations and subsequent conversion of metabolic products. Additionally, the gene *BR serine/threonine kinase 2a* (*brsk2*) is known to be related to neurological development and function as well as insulin secretion in humans, which could have an important role in both neurological dynamics of migratory tendency and endocrine regulation. Insulin is known to be a causative hormone in the smoltification process (McCormick 2009), elevating before other hormones and then binding to hepatocytes to initiate the parr-to-smolt transformation (Gutiérrez and Plisetskaya 1991). Moreover, as insulin is also important with respect to metabolism, it is thought to be critical in regulating the changes necessary for both smoltification and migration (Gutiérrez and Plisetskaya 1991; McCormick 2009).

Finally, we identified the gene *epidermal growth factor receptor substrate 15-like 1* (*eps1511*), which is related to a pathway involved in calcium ion binding in humans. Ion exchange capacity is critical in the evolution of migratory tendency, as the ability to change from a freshwater to saltwater environment involves differing regulatory mechanisms of ion management (Stefansson et al. 2012; McCormick 2013). To survive in freshwater, resident individuals are required to actively uptake ions (Na^{2+}, Cl^{-}, Ca^{2+}) to compensate for passive ion loss (Stefansson et al. 2012; McCormick 2013). However, in saltwater the mechanism is reversed, and individuals are required to actively dispel those same ions to remain in homeostasis. Therefore, any mechanism that improves the function or transcription of genes involved in this process could be advantageous depending on the specific SNP change and the life history type of the individual.

Limitations & Scope of Findings

Although there are many benefits to utilizing whole-genome sequencing of *O. mykiss* to study the genetics of migration, limitations within the methodology restrict the extent to which we can validate the importance of both genes and genome regions identified. Importantly, this study only looked at two populations, and more populations would provide further support of divergence or convergence for migratory tendency. Additionally, in Sashin Creek there are physical barriers between life history types, whereas in LSC, both phenotypes exist in tandem. Therefore, it is possible that some of the differences between life history types in Sashin are due to drift or selection not operating on genes or regions of the genome important in migratory behavior. In LSC, however, it is possible that differences could be due to population-specific effects from the environment, heritability, or gene/environment interactions.

Evidence of population-specific control of migration has previously been observed in *O. mykiss*, specifically in regard to the chromosomal inversion located on chromosome 5. Pearse et. al 2014 first identified this inversion as a region strongly associated with anadromy in multiple California and Oregon trout populations (Pearse et al. 2014). Upon the completion of the *O. mykiss* genome, the inversion was annotated and classified as following a latitudinal cline, thought to be due at least in part to environmentally-dependent selection (Pearse et al. 2014; Pearse et al. 2019). Although the inversion was found to have minimal influence on migration in the Sashin population (Weinstein et al. 2019), there is evidence that the inversion is associated with migration in adfluvial (but not anadromous) life history types in Alaskan resident rainbow trout (Arostegui et al. 2019). However, using Pool-seq prevents us from viewing the genetic effect of the inversion on our samples as we cannot view individual genotypes, and therefore we are unable to comment on its relevance in migratory behavior in our populations. This last point is another limitation to Pool-seq designs; although Pool-seq allows for a population-level survey

of genetic diversity, it prevents the determination of individual genotypes. Thus, it is impossible to investigate additive effects, recombination, or to determine the link between genotype and phenotype.

Given the limitations of Pool-seq, our data are still useful in adding to the growing body of evidence that suggests that migration is controlled in both a population-specific way and by many alleles of small effect. Our understanding of the association between genotype and phenotype is growing, but in salmonids we rely heavily upon molecular pathways that have been developed in other organisms, such as mice or humans, which do not necessarily have the same function in *O. mykiss*. Lastly, the data generated here highlight some areas of the rainbow trout genome that may be important in understanding the link between genes and propensity to migrate, or parts of the genome that are differentially expressed between migrants and residents. Although we were unable to find localized results that implicate a singular region with a large effect, our findings instead highlight areas for future studies that focus on using more populations for gene by gene, gene by environment, and epigenetic control, to gain a deeper understanding of the intra-genetic control of this species.

Conclusion

In this study we successfully utilized Pool-seq to evaluate differentiation and nucleotide diversity located within two separated populations of migratory and resident rainbow trout. Our results indicate that the development of their respective ecotype is likely governed by many genes of small effect, and that these alleles may be important in regulating gene expression. In addition, by comparing our results and the findings of prior studies with a population in Little Sheep Creek, Oregon, we strongly suggest that the genetic basis of migration is controlled in a population-specific manner, as only limited sharing of alleles exists between Sashin and LSC. However, we did identify eight shared regions of the genome which suggested purifying

selection in both resident populations, and eight other regions of the genome responding to selection in both populations of migratory steelhead trout. Although there is still much that is not understood about the biological regulation of migratory behavior, particularly in non-model species, our work contributes to a body of information uncovering the genetic control of the migratory process. To effectively evaluate this process, future research efforts should focus their efforts beyond the site in Sashin Creek, Alaska, as population-specific effects likely play a role in regulating migratory behavior.

REFERENCES

Åkesson S, Hedenström A. 2007. How migrants get there: migratory performance and orientation. BioScience. 57(2):123–133.

Arostegui MC, Quinn TP, Seeb LW, Seeb JE, McKinney GJ. 2019. Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. Mol Ecol. 28(6):1412–1427.

Baerwald MR, Meek MH, Stephens MR, Nagarajan RP, Goodbla AM, Tomalty KMH, Thorgaard GH, May B, Nichols KM. 2016. Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout. Mol Ecol. 25(8):1785–1800. doi:10.1111/mec.13231.

Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C. 2004. Vernalization requires epigenetic silencing of FLC by histone methylation. Nature. 427(6970):164–167.

Berntson EA, Carmichael RW, Flesher MW, Ward EJ, Moran P. 2011. Diminished reproductive success of steelhead from a hatchery supplementation program (Little Sheep Creek, Imnaha Basin, Oregon). Trans Am Fish Soc. 140(3):685–698.

Chapman BB, Brönmark C, Nilsson J-Å, Hansson L-A. 2011. The ecology and evolution of partial migration. Oikos. 120(12):1764–1775.

Choi YJ, Kim NN, Shin HS, Choi CY. 2014. The expression of leptin, estrogen receptors, and vitellogenin mRNAs in migrating female chum salmon, *Oncorhynchus keta*: the effects of hypo-osmotic environmental changes. Asian-Australas J Anim Sci. 27(4):479–487.

Doctor K, Berejikian B, Hard JJ, VanDoornik D. 2014. Growth-mediated life history traits of steelhead reveal phenotypic divergence and plastic response to temperature. Trans Am Fish Soc. 143(2):317–333.

Dodson JJ, Aubin-Horth N, Thériault V, Páez DJ. 2013. The evolutionary ecology of alternative migratory tactics in salmonid fishes. Biol Rev. 88(3):602–625.

Faust GG, Hall IM. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 30(17):2503–2505.

Fuentes EN, Kling P, Einarsdottir IE, Alvarez M, Valdés JA, Molina A, Björnsson BT. 2012. Plasma leptin and growth hormone levels in the fine flounder (Paralichthys adspersus) increase gradually during fasting and decline rapidly after refeeding. Gen Comp Endocrinol. 177(1):120–127.

Fukami K. 2002. Structure, Regulation, and Function of Phospholipase C Isozymes. J Biochem (Tokyo). 131(3):293–299.

Gómez-Bahamón V, Márquez R, Jahn AE, Miyaki CY, Tuero DT, Laverde-R O, Restrepo S, Cadena CD. 2020. Speciation associated with shifts in migratory behavior in an avian radiation. Curr Biol. 30(7):1312-1321.e6.

Gong N, Einarsdottir IE, Johansson M, Björnsson BT. 2013. Alternative splice variants of the rainbow trout leptin receptor encode multiple circulating leptin-binding proteins. Endocrinology. 154(7):2331–2340.

Gutiérrez J, Plisetskaya EM. 1991. Insulin binding to liver plasma membranes of coho salmon during smoltification. Gen Comp Endocrinol. 82(3):466–475.

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*). Comp Biochem Physiol Part D Genomics Proteomics. 20:136–150.

Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM. 2013. Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. G3 GenesGenomesGenetics. 3(8):1273–1285.

Hardesty-Moore M, Deinet S, Freeman R, Titcomb GC, Dillon EM, Stears K, Klope M, Bui A, Orr D, Young HS, et al. 2018. Migration in the Anthropocene: how collective navigation, environmental system and taxonomy shape the vulnerability of migratory species. Philos Trans R Soc Lond B Biol Sci. 373(1746).

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):3061–3076.

Hecht BC, Hard JJ, Thrower FP, Nichols KM. 2015. Quantitative genetics of migration-related traits in rainbow and steelhead trout. G3 GenesGenomesGenetics. 5(5):873–889.

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 Genet. 6(2):e1000862.

Iwami Y, Ojima T, Inoue A, Nishita K. 2002. Primary structure of myosin heavy chain from fast skeletal muscle of chum salmon *Oncorhynchus keta*. Comp Biochem Physiol B Biochem Mol Biol. 133(2):257–267.

Johansson M, Morgenroth D, Einarsdottir IE, Gong N, Björnsson BT. 2016. Energy stores, lipid mobilization and leptin endocrinology of rainbow trout. J Comp Physiol [B]. 186(6):759–773.

Jørgensen EH, Martinsen M, Strøm V, Hansen KER, Ravuri CS, Gong N, Jobling M. 2013. Long-term fasting in the anadromous Arctic charr is associated with downregulation of metabolic enzyme activity and upregulation of leptin A1 and SOCS expression in the liver. J Exp Biol. 216(17):3222–3230.

Keefer ML, Caudill CC. 2014. Homing and straying by anadromous salmonids: a review of mechanisms and rates. Rev Fish Biol Fish. 24(1):333–368.

Kofler R, Langmüller AM, Nouhaud P, Otte KA, Schlötterer C. 2016. Suitability of different mapping algorithms for genome-wide polymorphism scans with pool-seq data. GenesGenomesGenetics. 6(11):3507–3515.

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. Kayser M, editor. PLoS ONE. 6(1):e15925.

Kofler R, Pandey RV, Schloetterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 27(24):3435–3436.

Le Bras Y, Dechamp N, Krieg F, Filangi O, Guyomard R, Boussaha M, Bovenhuis H, Pottinger TG, Prunet P, Le Roy P, et al. 2011. Detection of QTL with effects on osmoregulation capacities in the rainbow trout (*Oncorhynchus mykiss*). BMC Genet. 12(1):46.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21):2987–2993.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078–2079.

Liedvogel M, Åkesson S, Bensch S. 2011. The genetics of migration on the move. Trends Ecol Evol. 26(11):561–569.

Lugo Ramos JS, Delmore KE, Liedvogel M. 2017. Candidate genes for migration do not distinguish migratory and non-migratory birds. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 203(6):383–397.

McCormick SD. 2009. Evolution of the hormonal control of animal performance: Insights from the seaward migration of salmon. Integr Comp Biol. 49(4):408–422.

McCormick SD. 2013. Smolt Physiology and Endocrinology. Fish Physiol. 32:199–251.

McCormick SD, Saunders RL. 1987. Preparatory Physiological Adaptations for Marine Life of Salmonids: Osmoregulation, Growth, and Metabolism. Am Fish Soc Symp. 1:211–229.

McKinney GJ, Hale MC, Goetz G, Gribskov M, Thrower FP, Nichols KM. 2015. Ontogenetic changes in embryonic and brain gene expression in progeny produced from migratory and resident Oncorhynchus mykiss. Mol Ecol. 24(8):1792–1809.

Merlin C, Liedvogel M. 2019. The genetics and epigenetics of animal migration and orientation: birds, butterflies and beyond. J Exp Biol. 222(Suppl 1).

Micheletti SJ, Narum SR. 2018. Utility of pooled sequencing for association mapping in nonmodel organisms. Mol Ecol Resour. 18(4):825–837.

Murashita K, Uji S, Yamamoto T, Rønnestad I, Kurokawa T. 2008. Production of recombinant leptin and its effects on food intake in rainbow trout (*Oncorhynchus mykiss*). Comp Biochem Physiol B Biochem Mol Biol. 150(4):377–384.

Nichols KM, Edo AF, Wheeler PA, Thorgaard GH. 2008. The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. Genetics. 179(3):1559–1575.

Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, Rexroad CE, Moen T. 2015. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. Mol Ecol Resour. 15(3):662–672.

Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, Anderson EC, Rundio DE, Williams TH, Naish KA, et al. 2019. Sex-dependent dominance maintains migration supergene in rainbow trout. Nat Ecol Evol. 3(12):1731–1742.

Pearse DE, Miller MR, Abadía-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 R Soc B Biol Sci. 281(1783).

Quinn TP. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fish Res. 18(1):29–44.

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nat Rev Genet. 15(11):749–763.

Sheridan MA. 1989. Alterations in lipid metabolism accompanying smoltification and seawater adaptation of salmonid fish. Aquac. 82(1):191–203.

Singh NJ, Milner-Gulland EJ. 2011. Conserving a moving target: planning protection for a migratory species as its distribution changes. J Appl Ecol. 48(1):35–46.

Sloat MR, Reeves GH. 2014 Jan 6. Individual condition, standard metabolic rate, and rearing temperature influence steelhead and rainbow trout (*Oncorhynchus mykiss*) life histories. Can J Fish Aquat Sci. 71(4):491-501.

Stefansson SO, Haugland M, Björnsson BT, McCormick SD, Holm M, Ebbesson LOE, Holst JChr, Nilsen TO. 2012. Growth, osmoregulation and endocrine changes in wild Atlantic salmon smolts and post-smolts during marine migration. Aquac. 362–363:127–136.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123(3):585–595.

Thrower F, Guthrie C, Nielsen J, Joyce J. 2004. A comparison of genetic variation between an anadromous steelhead, *Oncorhynchus mykiss*, population and seven derived populations sequestered in freshwater for 70 Years. Environ Biol Fishes. 69(1):111–125.

Thrower FP, 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. Propagated Fish in Resource Management. Vol. 44. Bethesda: Amer Fisheries Soc. p. 485–496.

Watabe S, Hirayama Y, Nakaya M, Kakinuma M, Kikuchi K, Guo X-F, Kanoh S, Chaen S, Ooi T. 1997. Carp expresses fast skeletal myosin isoforms with altered motor functions and structural stabilities to compensate for changes in environmental temperature. J Therm Biol. 22(6):375–390.

Weinstein SY, Thrower FP, Nichols KM, Hale MC. 2019. A large-scale chromosomal inversion is not associated with life history development in rainbow trout from Southeast Alaska. PloS One. 14(9):e0223018.

Wilcove DS, Wikelski M. 2008. Going, Going, Gone: Is Animal Migration Disappearing. PLOS Biol. 6(7):e188.

Winker K. 2010. On the origin of species through heteropatric differentiation: a review and a model of speciation in migratory animals. Ornithol Monogr. 69:1–30.

Yin, L. et al. 2021. rMVP: A Memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. GPB.

VITA

PERSONAL BACKGROUND

Catherine Irene Clare, born 27 November, 1994 in Saginaw, Michigan to Michael P. Clare and Leslie S. Clare.

Graduated from Valley Lutheran High School in Saginaw, Michigan in June of 2012.

EDUCATION

M.S. Biology, Texas Christian University

B.S. Biology, Saginaw Valley State University

AWARDS AND FUNDING

Science & Engineering Research Center (SERC) Grant – Texas Christian University - *$1600*

Adkins Fellowship – Texas Christian University - *$3600*

TEACHING EXPERIENCE

Teaching Assistantship, Texas Christian University

- Introductory Biology II Laboratory – Spring 2020 and Spring 2021
- Contemporary Issues in Biology Laboratory – Fall 2019 and Fall 2020

Laboratory Technician, Oakland Community College – Fall 2018 and Spring 2019

ABSTRACT

COMPARATIVE GENOMICS OF RAINBOW TROUT (*ONCORHYNCHUS MYKISS)*: ARE GENES ASSOCIATED WITH MIGRATION CONSERVED AMONG POPULATONS?

By Catherine Irene Clare, M.S. 2021

Department of Biology

Texas Christian University

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

The rainbow trout, *Oncorhynchus mykiss*, is a partially migratory organism used to study the genetic control of migration. Much of this research has taken place at a unique site in Sashin Creek, Alaska, where the resident and migrant *O. mykiss* populations are isolated from each other. However, it is unknown the extent to which findings here are shared with other populations. Here we used pooled sequencing to gather genomic data from 174 fish in two locations – Sashin Creek, Alaska and Little Sheep Creek, Oregon. Four sequenced pools were developed based on phenotype and population. We then measured differentiation between the populations to identify regions that may be associated with the resident or migratory phenotype. Our findings identified specific genes and chromosomal regions that may be important in the regulation of migratory tendency in this species, and indicate that there are population-specific controls that regulate migratory behavior.