

Distribution of genetic variation underlying adult migration timing in steelhead of the Columbia River basin

Erin Collins¹, John Hargrove², Thomas Delomas², and Shawn Narum¹

¹Columbia River Inter-Tribal Fish Commission

²Pacific States Marine Fisheries Commission

May 4, 2020

Abstract

Fish migrations are energetically costly, especially when moving between fresh and saltwater, but are a viable strategy for Pacific salmon and trout (*Oncorhynchus* spp.) due to the advantageous resources available at various life stages. Anadromous steelhead (*O. mykiss*) migrate vast distances and exhibit variation for migration phenotypes that have a genetic basis at candidate genes known as *greb1L* and *rock1*. We examined the distribution of genetic variation at 13 candidate markers spanning *greb1L*, intergenic, and *rock1* regions versus 246 neutral markers for 113 populations (n = 9,471) of steelhead from inland and coastal lineages in the Columbia River. Patterns of population structure with neutral markers reflected genetic similarity by geographic region as demonstrated in previous studies, but candidate markers clustered populations by predominate genetic variation associated with migration timing. Mature alleles for late migration had the highest frequency overall in steelhead populations throughout the Columbia River, with only 9 of 113 populations that had a higher frequency of premature alleles for early migration. While a single haplotype block was evident for the coastal lineage, we identified multiple haplotype blocks for the inland lineage. The inland lineage had one haplotype block that corresponded to candidate markers within the *greb1L* gene and immediately upstream in the intergenic region, and the second block only contained candidate markers from the intergenic region. Haplotype frequencies had similar patterns of geographic distribution as single markers, but there were distinct differences in frequency between the two haplotype blocks for the inland lineage. Redundancy analyses were used to model environmental effects on allelic frequencies of candidate markers and significant variables were migration distance, temperature, isothermality, and annual precipitation. This study improves our understanding of the spatial distribution of genetic variation underlying migration timing in steelhead as well as associated environmental factors and has direct conservation and management implications.

Title: Distribution of genetic variation underlying adult migration timing in steelhead of the Columbia River basin

Authors: Erin E. Collins¹, John S. Hargrove², Thomas A. Delomas², Shawn R. Narum¹

Affiliations: ¹ Columbia River Inter-Tribal Fish Commission

² Pacific States Marine Fisheries Commission, Eagle Fish Genetics Lab, 1800 Trout Road, Eagle, ID 83616

Contact information: Hagerman Genetics Lab

3059-F National Fish Hatchery Road Hagerman, ID 83332

Keywords: *Oncorhynchus* , anadromous, *greb1L* , population genetics, landscape genetics

Abstract

Fish migrations are energetically costly, especially when moving between fresh and saltwater, but are a viable strategy for Pacific salmon and trout (*Oncorhynchus* spp.) due to the advantageous resources available

at various life stages. Anadromous steelhead (*O. mykiss*) migrate vast distances and exhibit variation for migration phenotypes that have a genetic basis at candidate genes known as *greb1L* and *rock1*. We examined the distribution of genetic variation at 13 candidate markers spanning *greb1L*, intergenic, and *rock1* regions versus 246 neutral markers for 113 populations ($n = 9,471$) of steelhead from inland and coastal lineages in the Columbia River. Patterns of population structure with neutral markers reflected genetic similarity by geographic region as demonstrated in previous studies, but candidate markers clustered populations by predominate genetic variation associated with migration timing. Mature alleles for late migration had the highest frequency overall in steelhead populations throughout the Columbia River, with only 9 of 113 populations that had a higher frequency of premature alleles for early migration. While a single haplotype block was evident for the coastal lineage, we identified multiple haplotype blocks for the inland lineage. The inland lineage had one haplotype block that corresponded to candidate markers within the *greb1L* gene and immediately upstream in the intergenic region, and the second block only contained candidate markers from the intergenic region. Haplotype frequencies had similar patterns of geographic distribution as single markers, but there were distinct differences in frequency between the two haplotype blocks for the inland lineage. Redundancy analyses were used to model environmental effects on allelic frequencies of candidate markers and significant variables were migration distance, temperature, isothermality, and annual precipitation. This study improves our understanding of the spatial distribution of genetic variation underlying migration timing in steelhead as well as associated environmental factors and has direct conservation and management implications.

1 Introduction

Many animals undertake long-distance migration from their natal sites to capitalize on abundant resources that may increase survival, fecundity, and fitness (Dingle & Drake, 2007). Migrations offer temporal and spatial availability of resources, along with seasonal suitability of migratory corridors and natal areas (Edwards & Richardson, 2004; Forrest & Miller-Rushing, 2010). The migration of *Oncorhynchus* spp. (Pacific salmon and trout) is a critical cultural, economic, and ecological resource throughout their native range. Conservation of salmon and steelhead is based upon maintaining phenotypic and genetic variation of distinct populations and a principal focus involves conserving migration timings across large drainages such as the Columbia River basin. Many populations are managed according to degree of reproductive isolation and life history variation. Evolutionarily significant units (ESU) of Pacific salmon and trout are defined as a Distinct Population Segment (DPS) under the US Endangered Species Act (ESA) (Ryder, 1986; Waples, 1991) and each DPS is determined by whether it is sufficiently reproductively isolated and of evolutionary importance to the species (Waples, 1991). Since the late 1800s, wild Pacific salmon and trout have experienced a steady decline in abundance and range. The freshwater range of Pacific salmon and trout has shrunk to about 60% of the historical range (National Research Council, 1996; English et al., 2006). The decline has been initially attributed to overharvest, habitat degradation (logging, mining, agricultural practices), and other anthropogenic development, but modern anthropogenic activity including hydroelectric dams' disruption of migratory routes, climate change, and an ongoing decrease in suitable habitat have also contributed to decline (Chapman, 1986; Meehan, 1991; Crozier et al., 2008).

Steelhead (*O. mykiss*) may undertake long migrations (over a thousand kilometers) in early life stages and return to natal sites to spawn (Busby et al., 1996; Keefer et al., 2014). Migratory phenotypes in the Columbia River basin vary by genetic lineage that have been previously characterized as either coastal or inland (Utter et al., 1980; Busby et al., 1996; Quinn, 2018). The two genetic lineages are geographically separated; the coastal lineage inhabits streams west of the Cascade Mountains and the inland lineage inhabits streams east of the Cascades (Busby et al., 1996; Brannon et al., 2004). Out of 15 steelhead ESUs in the Columbia River basin, 11 are listed under the ESA (Waples et al., 2001); one steelhead ESU is endangered and ten are threatened (Quinn, 2018). According to the ESA, an estimated one-third of Pacific salmon and trout populations and all five DPS of steelhead in the Columbia River are listed as threatened or endangered (Gustafson et al., 2007). Steelhead have also been extirpated from the upper Snake River and Columbia River head-waters (Gustafson et al., 2007).

Populations of steelhead consist of individuals that spawn at similar times and are genetically similar at neutral genetic markers, but individuals within a population may display significant variation in when they enter freshwater or arrive at spawning grounds (Quinn, 2018). Steelhead spawn in the spring, but can begin migration as early as summer of the previous year before spawning or as late as winter/spring just before spawning (Quinn et al., 2015). Steelhead migration may be characterized as bimodal in some rivers (Leider et al., 1986; Hess et al., 2016), with migrations referred to as early migrating summer-run (premature) or late migrating winter-run (mature; Quinn et al., 2015). Steelhead that exhibit early migration enter freshwater before they are sexually mature, and then hold in freshwater for several months throughout the winter before maturing and spawning the following spring (Quinn et al., 2015; Quinn, 2018). Steelhead that exhibit late migrations become sexually mature in the ocean before migration into freshwater only weeks to a few months before spawning at natal sites in the spring (Quinn et al., 2015; Quinn, 2018). Significantly more stream-maturing steelhead populations have been extirpated than ocean-maturing steelhead populations (Gustafson et al., 2007).

Migration timing in Pacific salmon and trout has been demonstrated to be heritable (Quinn et al., 2000; Thériault et al., 2007; Carlson & Seamons, 2008; Quinn et al., 2015). Further, migration timing is associated with a genomic region of major effect in both steelhead and Chinook salmon (*O. tshawytscha*) (Hess et al., 2016; Prince et al., 2017; Micheletti et al., 2018a; Narum et al., 2018; Thompson et al., 2019). Restriction site associated DNA sequencing (RAD-seq) studies have revealed single-nucleotide polymorphisms (SNPs) within the *greb1L* gene region that are associated with steelhead migration timing (Hess et al., 2016; Prince et al., 2017). Additional candidate genes associated with migration timing have been identified using a genome resequencing technique (Pool-seq; Micheletti and Narum, 2018) that generates allele frequencies at the population level spanning nearly half the genome. Pooled-sequencing (Pool-seq) methods and PoolParty analysis pipeline (Micheletti & Narum, 2018) revealed further SNPs associated with migration timing and expanded the genomic region of discovered SNPs to three more candidate genes (*rock1*, *mib1*, *abhd3*, and intergenic region between *greb1L* and *rock1*) (Micheletti et al., 2018a). While this genomic region of major effect may have direct conservation applications such as refining conservation units and fisheries harvest (Waples & Lindley, 2018), further understanding is needed including inheritance patterns and linkage relationships among candidate markers, and the influence of landscape characters on the distribution and frequency of candidate markers.

The *greb1L* gene is broadly present and conserved in vertebrates and the function is believed to be similar to *greb1*, which has been shown to modulate estrogen receptors and augment the role of estrogen in transcription in humans (Mohammed et al., 2013). Markers shown to have non-conservative and non-synonymous mutations by Micheletti et al. (2018a) indicate that this genetic region is under selection and the markers in the intergenic region, upstream of *greb1L*, associated with migration timing could be promoters or enhancers and regulate expression of *greb1L* (Kilpinen et al., 2013). Recent studies suggest that *greb1L* plays a role in early and late migration phenotypes in steelhead and Chinook salmon (Hess et al., 2016; Prince et al., 2017; Micheletti et al., 2018a; Narum et al., 2018; Thompson et al., 2019). Migration to spawning grounds is intrinsically linked to sexual development and maturation in Pacific salmon and trout and these processes have been attributed to *greb1L* in chum salmon (*Oncorhynchus keta*) and other species (Ghosh et al., 2000; Rae et al., 2006; Pellegrini et al., 2012; Choi et al., 2014).

In this study, we examined the distribution of genetic variation for the candidate genomic region in steelhead and were able to expand upon the number of candidate markers associated with migration timing, the number of individuals sampled, and improved sampling coverage across the Columbia River basin compared to previous studies (Hess et al., 2016; Prince et al., 2017; Micheletti et al., 2018a). We used 13 candidate markers spanning *greb1L*, *rock1*, and the intergenic region to test combinations of markers that resulted in haplotypes representative of migration timing phenotypes. Four of the candidate markers were previously identified with RADseq and pooled sequencing methods (Hess et al., 2016; Micheletti et al., 2018a), and nine additional candidate markers were developed with pooled sequencing methods (Table 1). Sample collections were distributed across the Columbia River basin, allowing for comparisons of candidate allelic and haplotypic frequencies for migration timing in a variety of steelhead habitats to better understand the spatial distribution

of genetic variation underlying steelhead migration timing. Finally, we use landscape genetic analyses to expand upon the evaluation of environmental drivers of genetic variation identified by Micheletti et al. (2018b) for these candidate markers and for an expansion of collection sites. To distinguish between migration timing phenotypes and associated genetic variation, we use the terminology of ‘early’ and ‘late’ to refer to migration phenotypes and ‘premature’ and ‘mature’ to refer to genetic variation (alleles, genotypes, or haplotypes).

2 Methods

2.1 Sample Collection

Natural-origin steelhead were collected from populations of both the inland and coastal lineages across multiple years from 1996 to 2018 and exhibit a range of migration timing. Samples were collected with a variety of methods, such as electrofishing, smolt traps, and weirs. Non-lethal fin tissue samples and biologically relevant metrics were collected from both smolts and returning adults. Steelhead were collected from locations distributed throughout the Columbia River basin. Sample sizes and coordinates for each collection provided in Figure 1 and Appendix S1 in Table S1.

2.2 Molecular methods

DNA was extracted from tissue in accordance with a Chelex 100 method (Sigma-Aldrich, St Louis, MO). All specimens were genotyped with genotyping-in-thousand by sequencing method (GT-seq; Campbell et al., 2015). All samples and loci with $\geq 10\%$ missing genotypes were removed from further analyses for quality control purposes. Over the period that these individuals were genotyped, various genetic marker panel updates occurred, resulting in slight variances of the mix of putatively neutral and adaptive markers available (Appendix S1 Table S1). Samples were genotyped with GTseq panels ranging from 368-390 SNPs, and genotype data were retained when $>90\%$ loci successfully genotyped and had an estimated $<0.5\%$ genotyping error based on replicate genotyping.

2.3 Molecular analyses

Putatively neutral markers were assessed using a combination of multivariate methods to detect underlying population structure, which we expected to coincide with coastal and inland lineages described in previous studies. A principal component analysis (PCA) was plotted for all populations based on allele frequencies of putatively neutral markers determined to be without linkage disequilibrium (LD). The PCA of putatively neutral markers accompanies a discriminate analysis of principal components (DAPC) with the R package adegenet 2.1.0 to assign probability of individual membership to genetic groups, K , revealed with the putative neutral marker PCA (Jombart, 2008; Jombart & Ahmed, 2011). The DAPC recovers maximum genetic variation between groups, while minimizing genetic variation within groups (Jombart, 2008; Jombart & Ahmed, 2011). The ‘find.clusters’ adegenet function ran for 25 instances for $K=1$ through $K=10$. The Bayesian information criterion (BIC) was averaged and scaled by the standard deviation for each K value. The most appropriate number of genetic groups was determined with the greatest ΔK value as described in Evanno et al. (2005).

The distribution of genetic variation underlying adult migration timing in steelhead across the landscape was described by genotype frequencies. We examined 13 markers occurring on chromosome 28 within the *greb1L*, *rock1*, and intergenic region between *greb1L* and *rock1* that were previously shown to be strongly associated with adult migration timing (Hess et al., 2016; Micheletti et al., 2018a; Table 1). Initially the two most significant SNPs were retained from a previous RAD study (Hess et al. 2016), and the remaining 11 with the strongest association with migration timing from the whole genome resequencing conducted by Micheletti et al. (2018a). To reduce ascertainment bias, we examined genetic variation in this candidate region from several populations of *O. mykiss* in the region to design primers (Table 1). Premature, mature, and heterozygote genotypes for migration timing were established based on genotype association from previous studies (Hess et al., 2016; Micheletti et al., 2018a), as well as using reference Skamania broodstock genotyped, which is a hatchery-strain intensively selected for early migration and cultured since 1956 with steelhead from the Washougal and Klickitat Rivers (Chilcote et al., 1986). Premature, mature, and heterozygote migration

timing genotype proportions were assessed across all collection locations. A PCA of allele frequencies of adaptive markers was also conducted for all collection locations to assess genetic groupings based on migration timing.

We assessed linkage disequilibrium (LD) within the 13 candidate markers to identify haplotype blocks that would be informative for estimating frequencies of migration types. Haplotype blocks within the 13 candidate markers were defined with solid spine LD analysis in the Java Runtime Environment software, Haploview 4.2, across all collection locations (Barrett et al., 2005). A solid spine of LD was extended across a haploblock if D' , or a normalization of the coefficient of LD, exceeded 0.74. The same markers were assessed for LD in individuals from coastal and inland lineages (as delineated by DAPC) separately. The effect of population structure on the LD of the markers was corrected in the analysis with the LDcorSV 1.3.2 R package (Mangin et al., 2012; Appendix S1 Table S2). Variation of genotype proportions were also evaluated with various groupings of the candidate markers.

Redundancy analyses (RDA) were conducted for all Columbia River basin collections to model the degree to which the variation in environmental variables explained the variation in allele frequencies of candidate markers included in the haplotype blocks (Borcard et al., 1992; Kierepka & Latch, 2015). Redundancy analysis was performed on two sets of collections, all populations and each lineage (coastal vs. inland) using the R package Vegan 2.5-6 (Oksanen et al., 2019). We used environmental variables that were significantly associated with adaptive genetic variation in a previous study (Micheletti et al., 2018b; Table 2; Appendix S1 Table S3). When two highly correlated (>0.75 pairwise correlation; Asuero et al., 2006) environmental variables were identified, one was removed from further analyses and the variable kept was determined from biological relevance to salmonids according to previous studies (Olsen et al., 2011; Hecht et al., 2015; Micheletti et al., 2018b). One-way analysis of variance (ANOVA) with a Tukey's range test (Tukey, 1949) identified significant variability in salmonid habitat. Environmental variables were analyzed with the "envfit" PCA function of the vegan R package. The ANOVA test and PCA together determined significant environmental variables within and among *O. mykiss* habitats measured in this study. The final RDAs were run with significant environmental variables retained from permutation tests with 1000 permutations ($\alpha=0.05$). Frequency of alleles in the haplotype block associated with migration timing were correlated to environmental variables with RDA constraint scores. Constraint scores indicated the degree of correlation and whether there was a positive or negative relationship between environmental variables and allelic frequencies.

3 Results

After aligning markers in common for all samples, 246 neutral markers (Hess et al., 2016) and up to 13 candidate markers from chromosome 28 (Table 1) were included for further analyses. A total of 9,471 individuals from 113 populations met inclusion criteria ($>90\%$ loci successfully genotyped and had an estimated $<0.5\%$ genotyping error based on replicate genotyping) and were included in this study.

Population structure as visualized by the PCA of allelic frequencies of neutral markers indicated genetic divergence by geographic locations (Figure 2). DAPC and ΔK revealed two genetic groupings that coincided with coastal and inland localities (Appendix S1 Figure S1). Most coastal collections, except for Mill Creek and Indian Creek, exhibited non-overlapping allele frequencies relative to all inland collections. The Klickitat River which is located between coastal and inland populations formed a cluster intermediate of the two population types. Inland collections from the Yakima and Clearwater rivers clustered distinctly from others in study (Figure 2).

A second PCA produced using candidate markers separated individuals according to proportion of premature and mature migration genotypes (Figure 3). In contrast to results with neutral markers that separated individuals by sample location and population structure, the PCA with adaptive markers separated individuals by migration timing genotypes. Cluster membership delineated via DAPC assuming $K=2$ grouped individuals from 25 putatively coastal lineage collections together and grouped individuals from 90 putatively inland lineage collections together in a second cluster (Appendix S1 Figure S1).

Candidate markers were analyzed for all sampling locations in Haploview with solid spine and this resulted

in two haploblocks, one with markers 1-7 and another with markers 8-13 (Figure 4a). One haplotype block contained all markers within *greb1L* and another included all or the majority of markers located within the intergenic region upstream of *greb1L* and *rock1*. There was one marker located within *rock1*, but it did not demonstrate as strong of LD as other markers included in the second haplotype block. The intergenic haplotype block, containing markers 8-12, maintained high LD in both inland and coastal collections.

When haplotype blocks were examined separately for coastal (Figure 4b) and inland (Figure 4c) lineages, high LD was retained at markers 8-12 for both lineages. Additionally, minor allele frequencies (MAF) were lower for all inland markers except for candidate markers 8-12 (Appendix S1 Table S3; Figure S2). Variation in LD occurred among markers 1-7 and was stronger in the coastal lineage (Figure 4b-c). Elevated LD in the coastal lineage markers resulted in one haplotype block, spanning markers 1-12 (Figure 4b). The solid spine analysis revealed three haplotype blocks in the inland lineages which were split between markers two and three and markers seven and eight (Figure 4c). The haplotype block split between markers seven and eight observed in the inland lineage was the same position as the split in all collections (Figure 4a), indicating the split for all collections was influenced by the inland collections. Further, a greater divergence between average MAF values can be observed between markers seven and eight of the inland collections than in the coastal collections (Figure S2). Confidence intervals (0.95 upper, 0.7 lower; Gabriel et al. 2002) and the four gamete rule, which assumes recombination when all four possible haplotypes are detected at frequencies exceeding 0.01 (frequency > 0.02-0.03; Wang et al. 2002), were applied in further LD analyses. Variation in haplotype blocks was observed between analyses and the differences were the inclusion or exclusion of markers 1 and 13 and the split between markers 5 and 6 or between markers 7 and 8. The difference in the location of where haplotype blocks were split could be influenced by fixed alleles at markers 4, 6, and 7 in some collections (Appendix S1 Table S3). All Snake River collections were limited to markers 2, 3, 6, and 9 because these markers were developed earlier than the rest and were the only markers available for these collections. This resulted in limited data availability (4 instead of 13 candidate markers) for the farthest inland collections. Haploview analysis comparing lineages was done both with and without the individuals that were only genotyped at 4 of the 13 markers and both analyses yielded the same results.

We examined six different combinations of markers to determine which markers produce similar frequency results: a single marker (9), three markers (2,3,6), four markers (2,3,6,9), five markers (8-12), six markers (2-7), and 11 markers (2-12). This allowed for comparison across marker groups to determine if frequencies across different marker combinations were similar. In general, all six combinations of marker groups provided similar haplotype frequencies with differences in associated haplotypes only differing by 1-7% (Figure S3). The groups with the most similar haplotype frequencies were marker 9 alone and markers 8-12, followed by markers 2, 3, and 6 and markers 2-7, markers 2, 3, 6, and 9 and markers 2-12 have similar average genotype frequencies (Figure S3).

Average genotype proportions were mapped across all collection locations with markers 2, 3, 6, and 9 because all collections were genotyped at these markers (Figure 5). The most common genotype in individuals sampled was the mature genotype. The mature genotype was predominant throughout much of the range in the Columbia River, however many populations west of the Cascade Mountains and in the Salmon River have greater proportions of the premature genotype than other collections (Figure 5). However, only 9 of the 113 populations had a higher frequency of premature alleles for early migration.

To evaluate haplotype frequencies for a single haplotype block in as many locations as possible, we further scrutinized haplotypes for markers 2, 3, 6 across the landscape and found five unique haplotypes (Figure S4a). Haplotype frequencies for collections (Figure S4a) showed similar patterns of geographic distribution as the genotype frequencies (Figure 5), but with improved resolution for heterozygous haplotypes that were within a single haplotype block underlying *greb1L*. According to results of overall haplotype frequency (Figure S4a), the heterozygote haplotype 4 is present more frequently than the premature haplotype 5. Additionally, there is a distinct separation of heterozygote haplotypes between coastal (haplotypes 2 and 3) and inland (haplotype 4) collections (Figure S4a).

To model impacts of significant environmental variables on allelic frequencies of migration timing associated

markers, RDAs were done for all Columbia River basin collections and then separately for coastal and inland lineage collections. Significant environmental variables retained in the RDA for all collections were migration distance, minimum temperature of the warmest month, 20-year average August water temperature, annual mean temperature, isothermality, and annual precipitation (Figure 6). Annual precipitation had the greatest effect when all collections were analyzed together (Figure 6). Environmental variables retained in the coastal lineage RDA were average temperature of the coldest quarter and precipitation of the wettest month (Figure S5a). Environmental variables retained in the interior lineage RDA were 20-year average August water temperature and minimum temperature of the warmest month (Figure S5b).

4 Discussion

This study provides further insight to the spatial distribution of genetic variation underlying migration timing in a broad range of steelhead populations. Genetic relationships were characterized for neutral and candidate markers for 113 populations, supporting previous findings of population structure and demonstrated strong differences between major lineages. We determined linkage blocks for 13 candidate markers associated with migration timing and found different heterozygote haplotypes were found to be predominant in coastal versus inland lineages. Environmental drivers of candidate variation revealed the importance of temperature and precipitation to selection on variation for migration in this system. Overall, this study provides extensive geographic variation for candidate markers associated with migration timing that is expected to be important for conservation applications in this species (Waples & Lindley, 2018).

Patterns of genetic variation among steelhead populations were highly distinct between neutral and candidate markers. Neutral structure was consistent with previous studies with various marker types that largely correspond to geographic population structure and significant heterogeneity in environmental conditions at collections sites (Blankenship et al., 2011; Matala et al., 2014; Micheletti et al., 2018b). For example, the Clearwater River steelhead have consistently showed a distinct genetic signal from others in the Snake River basin regardless of marker type (Narum et al., 2008; Campbell et al., 2012; Matala et al., 2014; Micheletti et al., 2018b). Additionally, the neutral markers provided further resolution than previous studies for the inland lineage especially for populations in the Yakima River drainage that were distinct from the rest of the populations in the middle Columbia River. The distinct neutral patterns in the Clearwater and Yakima River drainages were likely due to different levels of genetic influence from hatchery programs (Blankenship et al., 2011). Populations in the Yakima River have likely been minimally influenced by hatchery programs as there has been no direct hatchery program for steelhead in this subbasin. Similarly, large stretches of the Clearwater River basin, including the Selway and Lochsa Rivers, are managed exclusively for wild fish (Nielsen et al., 2009; Campbell et al., 2012). Presence of dams and historical rapid population declines may also differentiate sub-basins of inland steelhead from one another (Blankenship et al., 2011; Matala et al., 2014). Finally, the Klickitat River is positioned in a geographically intermediate location at the eastern base of the Cascades and considered to be on the boundary of the coastal and inland lineages. The intermediate status of the Klickitat River collections was evident in the neutral PCA population structuring which is consistent with previous studies (e.g., Micheletti et al., 2018b). This intermediate signal was also observed in two other populations, Fifteenmile Creek and Mill Creek, which may indicate gene flow with steelhead in the Klickitat R. or admixture. In contrast to geographical patterns observed at neutral loci, the candidate PCA divided collections by their predominant adult migration timing. The Skamania stock was a useful reference for the extreme extent of fixed genetic variation for premature alleles due to artificial selection for early migration timing over several decades in this hatchery program. At the other end of the spectrum, the mature genotype was predominant in most collections, while the heterozygote collections were dispersed across the basin, but with divergent ratios of haplotypes between coastal and inland lineages. The presence of genetic variation for premature alleles in the inland lineage suggests that some populations of steelhead (i.e., those in the Salmon R. drainage) may exhibit phenotypic variation for early and late arrival timing to spawning grounds as shown by Micheletti et al. (2018b).

Haplotype blocks of markers with the greatest association with one another and with the migration timing phenotype improve ability to evaluate genetic variation associated with migration timing across the land-

scape. In addition to LD assessments, we evaluated differences between average genotype frequencies with fewer candidate markers. Marker 9 had the most similar average genotype frequencies to markers 8-12 for all genotypes and markers 8-12 had the greatest LD in all collections. This finding suggests that marker 9 could be useful under circumstances of limited genotyping abilities. This same marker was also helpful at distinguishing patterns in steelhead arrival timing to spawning grounds (Micheletti et al., 2018b). However, it is still beneficial to assess collections with entire haplotype blocks when possible, to generate numerous haplotype combinations instead of only three genotypes gained from a single marker.

We observed significant association between multiple environmental variables and candidate markers when examined across lineages, which was expected given that environmental conditions vary significantly across the Columbia River basin landscape. We found migration distances, temperature variables, and precipitation variables had the strongest association to adaptation for all collections which was consistent with previous landscape genomics analyses (Micheletti et al., 2018b). Migration distance traveled between the Pacific Ocean and spawning sites ranged from 60 to 1,400 km, presenting a vast difference between coastal and inland lineages in energetic allocation before spawning (Olsen et al., 2011; Hecht et al., 2015). Migration distance was not significantly associated with migration timing within each lineage suggesting that variation at candidate markers is not highly distinct among populations at small geographic scales. Significant association of temperature with candidate markers was not surprising since fish rely on environmental temperatures to regulate body temperatures and trigger migratory behavior (Jonsson 1991; Sykes et al., 2009), and extreme temperatures can inhibit cardiac and metabolic proficiencies (Chen et al., 2018). Further, genetic disparities in thermal tolerance when encountering temperature barriers has been found to contribute to local adaptation in salmonids (Eliason et al., 2011; Narum et al., 2013; Muñoz et al., 2015). Finally, the significance of precipitation with variation at candidate markers is expected to be important since precipitation conditions can impact survival and selection on genes associated with thermal tolerance when flow is low (Heath et al., 2002) and water temperatures are elevated (Narum et al., 2013). In contrast, when precipitation is high and stream flow is powerful, conditions may become energetically costly for migrating steelhead, but also provide cues for migration to spawning grounds (Keefer et al. 2014; 2018). Significantly associated environmental variables within each lineage were more limited than across lineages of steelhead, and largely reflected regional differences in precipitation within the coastal lineage and temperature within the inland lineage.

In this study, we assessed the spatial distribution of candidate haplotype frequencies because selective pressures on steelhead migration are disparate across the heterogeneous landscape. The coastal lineage contained steelhead maturing both in the ocean and streams, whereas inland lineage steelhead only matured in streams. Initial studies (Hess et al., 2016; Prince et al., 2017; Thompson et al., 2019) identified and associated *greb1L* genotypes with freshwater entry, while Micheletti et al. (2018a) revealed a greater *greb1L* association with arrival timing to spawning grounds. We also detected more than one genotype present in inland collections, further supporting an association with arrival timing to spawning grounds introduced by Micheletti et al. (2018a). Our study incorporated more collections and more candidate markers associated with migration timing than previous studies, which allowed us to determine haplotypes to describe the spatial pattern of mature and premature genotypes across the Columbia River basin in greater detail. Coastal collections exhibited greater genetic diversity at candidate markers, but greater influence of premature alleles from Skamania and other hatchery stocks. In the inland lineage, the mature genotype was detected at high frequency despite all inland steelhead maturing in freshwater, supporting findings by Micheletti et al. (2018a). Variation in the second haplotype block, which includes markers in the intergenic region, indicates that inland populations retain genetic variation that may allow for variable timing in arrival to spawning grounds. However, further studies are needed that dissect arrival phenotypes and the association at candidate markers at *greb1L* and *rock1*.

From a management perspective, detailing the distribution of migration run timing has direct conservation implications. Early migrating fish spend less time feeding in the nutrient rich ocean, resulting in less opportunities for growth and potential for decreased reproductive success. Further, more time in freshwater systems exposes early migrators to thermal stress, disease, and greater risk for impacts of climate change and selective fisheries (Quinn et al., 2015). Thus, steelhead with this early migration pattern have increased odds

of extirpation and may require greater conservation efforts (Prince et al., 2017). Previous findings (Micheletti et al., 2018a) were bolstered by this study that indicate greater genetic diversity at candidate genes for inland collections than previously understood. Effective conservation efforts to maintain or increase genetic variation underlying migration timing is expected to provide broader life history diversity for populations to endure stochastic environments. Thus, maintenance of genetic diversity associated with migration timing across the Columbia River basin may be a key to promote resilient steelhead populations that are able to recover from anthropogenic impacts.

Data Accessibility: Genotype data are available in Dryad at doi:10.5061/dryad.jh9w0vt80.

Competing Interests: None declared

Acknowledgements: Thanks to all tribes and agencies that provided samples, laboratory staff involved in sample processing (CRITFC, IDFG), Funding from Bonneville Power Administration grant number 2008-907-00.

Author Contributions: SRN designed and directed the study. EEC analyzed the data. All authors interpreted the results and wrote the manuscript.

Literature Cited

- Asuero, A. G., Sayago, A., & Gonzalez, A. G. (2006). The correlation coefficient: An overview. *Critical Reviews in Analytical Chemistry*, **36**(1), 41-59.
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**(2), 263-265.
- Blankenship, S. M., Campbell, M. R., Hess, J. E., Hess, M. A., Kassler, T. W., Kozfkay, C. C., Matala, A. P., Narum, S. R., Paquin, M. M., Small, M. P., Stephenson, J. J., Warheit, K. I., & Moran, P. (2011). Major lineages and metapopulations in Columbia River *Oncorhynchus mykiss* are structured by dynamic landscape features and environments. *Transactions of the American Fisheries Society*, **140**, 665-684.
- Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, **73**(3), 1045-1055.
- Brannon, E. L., Powell, M. S., Quinn, T. P., & Talbot, A. (2004). Population structure of Columbia River Basin Chinook salmon and steelhead trout. *Reviews in Fisheries Science*, **12**(2-3), 99-232.
- Busby, P. J., Wainwright, T. C., Bryant, G. J., Lierheimer, L. J., Waples, R. S., Waknitz, F. W., & Lagomarsino, I. V. (1996). Status review of west coast steelhead from Washington, Idaho, Oregon, and California.
- Campbell, M. R., Kozfkay, C. C., Copeland, T., Schrader, W. C., Ackerman, M. W., & Narum, S. R. (2012). Estimating abundance and life history characteristics of threatened wild Snake River steelhead stocks by using genetic stock identification. *Transactions of the American Fisheries Society*, **141**(5), 1310-1327.
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, **15**(4), 855-867.
- Carlson, S. M., & Seamons, T. R. (2008). A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. *Evolutionary Applications*, **1**(2), 222-238.
- Chapman, D. W. (1986). Salmon and steelhead abundance in the Columbia River in the nineteenth century. *Transactions of the American Fisheries Society*, **115**(5), 662-670.
- Chen, Z., Farrell, A. P., Matala, A., Hoffman, N., & Narum, S. R. (2018). Physiological and genomic signatures of evolutionary thermal adaptation in redband trout from extreme climates. *Evolutionary Applications*, **11**(9), 1686-1699.

Chilcote, M. W., Leider, S. A., & Loch, J. J. (1986). Differential reproductive success of hatchery and wild summer-run steelhead under natural conditions. *Transactions of the American Fisheries Society* , **115**(5) , 726-735.

Choi, Y. J., Kim, N. N., Shin, H. S., & Choi, C. Y. (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-Australasian Journal of Animal Sciences* , **27**(4) , 479.

Crozier, L. G., Hendry, A. P., Lawson, P. W., Quinn, T. P., Mantua, N. J., Battin, J., Shaw, R.G., & Huey, R. B. (2008). Potential responses to climate change in organisms with complex life histories: evolution and plasticity in Pacific salmon. *Evolutionary Applications* ,**1**(2) , 252-270.

Dingle H., Drake V. A. (2007). What is migration? *BioScience* ,**57** , 113–121.

Edwards M., & Richardson A. J. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*, **430**(7002) , 881.

Eliason, E. J., Clark, T. D., Hague, M. J., Hanson, L. M., Gallagher, Z. S., Jeffries, K. M., & Farrell, A. P. (2011). Differences in thermal tolerance among sockeye salmon populations. *Science* ,**332** , 109–112. <https://doi.org/10.1126/science.1199158>

English, K. K., Peacock, D., & Spilsted, B. (2006). North and central coast core stock assessment program for salmon. Prepared for Pacific Salmon Foundation and Fisheries and Oceans Canada by LGL Limited Environmental Research Associates and Fisheries and Oceans Canada, Sidney, B.C.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**(8) , 2611-2620.

Forrest J., & Miller-Rushing A. J. (2010). Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philosophical Transactions of the Royal Society B* , **365** , 3101-3112.

Gabriel, S. B, Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., Defelice, M., Lochner, A., Faggart, M. (2002). The structure of haplotype blocks in the human genome. *Science* , **296** , 2225–2229. doi: 10.1126/science.1069424.

Ghosh, M. G., Thompson, D. A., & Weigel, R. J. (2000). PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer Research* , **60**(22) , 6367–6375.

Gustafson, R. G., Waples, R. S., Myers, J. M., Weitkamp, L. A., Bryant, G. J., Johnson, O. W., & Hard, J. J. (2007). Pacific salmon extinctions: quantifying lost and remaining diversity. *Conservation Biology*, **21**(4) , 1009-1020.

Heath, D. D., Busch, C., Kelly, J., & Atagi, D. Y. (2002). Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). *Molecular Ecology* ,**11**(2) , 197-214.

Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology*,**24**(22) , 5573-5595.

Hess, J. E., Zandt, J. S., Matala, A. R., & Narum, S. R. (2016). Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society B: Biological Sciences*,**283**(1830) , 20153064.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* , **24** , 1403-1405. doi:10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*,**27**(21) , 3070-3071. doi:10.1093/bioinformatics/btr521

- Jonsson, N. (1991). Influence of water flow, water temperature and light on fish migration in rivers. *Nordic Journal of Freshwater Research* , **66**(1991) , 20-35.
- Keefer, M. L., & Caudill, C. C. (2014). Homing and straying by anadromous salmonids: A review of mechanisms and rates. *Reviews in Fish Biology and Fisheries* , **24** , 333–368. doi.org/10.1007/s11160-013-9334-6
- Keefer, M. L., Jepson, M. A., Clabough, T. S., Johnson, E. L., Narum, S. R., Hess, J. E., & Caudill, C. C. (2018). Sea-to-sea survival of late-run adult steelhead (*Oncorhynchus mykiss*) from the Columbia and snake rivers. *Canadian Journal of Fisheries and Aquatic Sciences* , **75**(3) , 331–341. doi.org/10.1139/cjfas-2016-0430
- Kierepka, E. M., & Latch, E. K. (2015). Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology Resources*, **15**(3) , 512-525.
- Kilpinen, H., Waszak, S. M., Gschwind, A. R., Raghav, S. K., Witwicki, R. M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N. I., Yurovsky, A., Lappalainen, T., Romano-Palumbo, L., Planchon, A., Bielser, D., Bryois, J., Padiou, I., Udin, G., Thurnheer, S., Hacker, D., Core, L. J., Lis, J. T., Hernandez, N., Reymond, A., Deplancke, B., & Dermitzakis, E. T. (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* , **342**(6159) , 744-747.
- Leider, S. A., Chilcote, M. W., & Loch, J. J. (1986). Comparative life history characteristics of hatchery and wild steelhead trout (*Salmo gairdneri*) of summer and winter races in the Kalama River, Washington. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**(7) , 1398-1409.
- Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., & Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* , **108**(3) , 285-291.
- Matala, A.P., Ackerman, M. W., Campbell, M. R., & Narum, S. R. (2014). Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes. *Evolutionary Applications*, **7**(6) , 682-701.
- Meehan, W. R. (1991). Influences of forest and rangeland management on salmonid fishes and their habitats. *American Fisheries Society Special Publication*, **19** , Evans City, PA.
- Micheletti, S. J., Hess, J. E., Zendt, J. S., & Narum, S. R. (2018a) Selection at a genomic region of major effect is responsible for evolution of complex life histories in anadromous steelhead. *BMC Evolutionary Biology*, **18**(1) , 140.
- Micheletti, S. J., Matala, A. R., Matala, A. P., & Narum, S. R. (2018b). Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). *Molecular Ecology*, **27**(1) , 128-145.
- Micheletti, S. J., & Narum, S. R. (2018). Utility of pooled sequencing for association mapping in nonmodel organisms. *Molecular Ecology Resources* , **18**(4) , 825-837.
- Mohammed, H., D'Santos, C., Serandour, A. A., Ali, H. R., Brown, G. D., Atkins, A., Rueda, O. M., Holmes, K. A., Theodorou, V., Robinson, J. L. L., Zwart, W., Saadi, A., Ross-Innes, C. S., Chin, S. F., Menon, S., Stingl, J., Palmieri, C., Caldas, C., Carroll, J. S. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Reports*, **3**(2) , 342-349.
- Munoz, N. J., Farrell, A. P., Heath, J. W., & Neff, B. D. (2015). Adaptive potential of a Pacific salmon challenged by climate change. *Nature Climate Change* , **5** , 1–5.
- Narum, S. R., Banks, M., Beacham, T. D., Bellinger, M. R., Campbell, M. R., Dekoning, J., Elz, A., Guthrie Iii, C. M., Kozfkay, C., Miller, K. M. and Moran, P. (2008). Differentiating salmon populations at broad

and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* , **17**(15) , 3464-3477.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* , **22**(11) , 2841-2847.

Narum, S. R., Di Genova, A., Micheletti, S. J., & Maass, A. (2018). Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proceedings of the Royal Society B: Biological Sciences* , **285**(1883) , 20180935.

National Research Council. (1996). Upstream: salmon and society in the Pacific Northwest. National Academy Press, Washington, D.C.

Nielsen, J. L., Byrne, A., Graziano, S. L., & Kozfkay, C. C. (2009). Steelhead genetic diversity at multiple spatial scales in a managed basin: Snake River, Idaho. *North American Journal of Fisheries Management* , **29** , 680–701. DOI.org/10.1577/M08-105.1

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>

Olsen, J. B., Crane, P. A., Flannery, B. G., Dunmall, K., Templin, W. D., & Wenburg, J. K. (2011). Comparative landscape genetic analysis of three Pacific salmon species from subarctic North America. *Conservation Genetics* , **12**(1) , 223-241.

Pellegrini, C., Gori, I., Achdari, C., Hornung, D., Chardonnens, E., Wunder, D., Fiche, M., & Canny, G. O. (2012). The expression of estrogen receptors as well as GREB1, c-MYC, and cyclin D1, estrogen-regulated genes implicated in proliferation, is increased in peritoneal endometriosis. *Fertility and Sterility* , **98** , 1200–1208.

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., Hotaling, T. J., Spidle, A. P., & Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances* , **3**(8) , e1603198.

Quinn, T. P., Unwin, M. J., & Kinnison, M. T. (2000). Evolution of Temporal Isolation in the Wild: Genetic Divergence in Timing of Migration and Breeding by Introduced Chinook Salmon Populations. *Evolution* , **54**(4) , 1372. [https://doi.org/10.1554/0014820\(2000\)054\[1372:eotiit\]2.0.co;2](https://doi.org/10.1554/0014820(2000)054[1372:eotiit]2.0.co;2)

Quinn, T. P., McGinnity, P., & Reed, T. E. (2015). The paradox of “premature migration” by adult anadromous salmonid fishes: patterns and hypotheses. *Canadian Journal of Fisheries and Aquatic Sciences*, **73**(7) , 1015-1030.

Quinn, T. P. (2018). The Behavior and Ecology of Pacific Salmon and Trout. 2nd edition American Fisheries Society University of Washington Press.

Rae, J. M., Johnson, M. D., Cordero, K. E., Scheys, J. O., Larios, J. M., Gottardis, M. M., Pienta, K. J., & Lippman, M. E. (2006). GREB1 is a novel androgen-regulated gene required for prostate cancer growth. *The Prostate*, **66**(8) , 886–894.

Ryder, O. A. (1986). Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*, **1** , 9-10.

Sykes, G. E., Johnson, C. J., & Shrimpton, J. M. (2009). Temperature and flow effects on migration timing of Chinook salmon smolts. *Transactions of the American Fisheries Society* , **138**(6) , 1252-1265.

Theriault, V., Garant, D., Bernatchez, L., & Dodson, J. J. (2007). Heritability of life-history tactics and genetic correlation with body size in a natural population of brook charr (*Salvelinus fontinalis*). *Journal of Evolutionary Biology* , **20**(6) , 2266–2277.

Thompson, T. Q., Bellinger, M. R., O'Rourke, S. M., Prince, D. J., Stevenson, A. E., Rodrigues, A. T., Sloat, M. R., Speller, C. F., Yang, D. Y., Butler, V. L., Banks, M. A., Miller, M. R. (2019). Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. *Proceedings of the National Academy of Sciences* ,**116**(1) , 177–186.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99-114.

Utter, F. M., Campton, D., Grant, S., Milner, G., Seeb, J., & Wishard, L. (1980). Population structures of indigenous salmonid species of the Pacific Northwest. *Salmonid ecosystems of the North Pacific*. Oregon State University Press, Corvallis, 285-304.

Wang, N., Akey, J. M., Zhang, K., Chakraborty, R., Jin, L. (2002). Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *American Journal of Human Genetics* , **71** , 1227–1234. doi: 10.1086/344398

Waples, R. S. (1991). Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. *Marine Fisheries Review*, **53**(3) , 11-22.

Waples, R. S., Gustafson, R. G., Weitkamp, L. A., Myers, J. M., Jjohnson, O. W., Busby, P. J., Hard, J. J., Bryant, G. J., Waknitz, F. W., Nelly, K., Teel, D., Grant, W. S., Winans, G. A., Phelps, S., Marshall, A., & Baker, B. M. (2001). Characterizing diversity in salmon from the Pacific Northwest. *Journal of Fish Biology*, **59** , 1-41.

Waples, R. S., & Lindley, S. T. (2018). Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evolutionary Applications* , **11**(9) , 1518-1526.

Tables

Table 1. Migration timing associated candidate marker information. The number corresponds to the SNP order. SNP names, chromosome number, position, gene, primers, probes, and orientation are also listed.

Table 2. Notation and descriptions, units, resolution, variable class, source, and if the variable was retained in the model are listed for all environmental variables assessed with the RDA models.

Figures

Figure 1. Collection sites numbered according to Table S1.

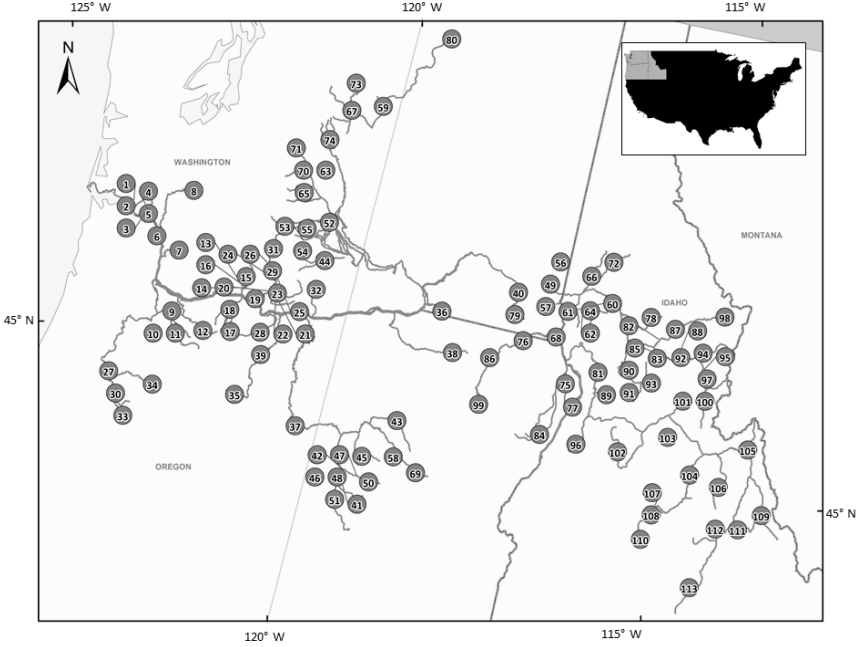


Figure 2. Neutral marker PCA plot for all populations. See Table S1 for collection names.

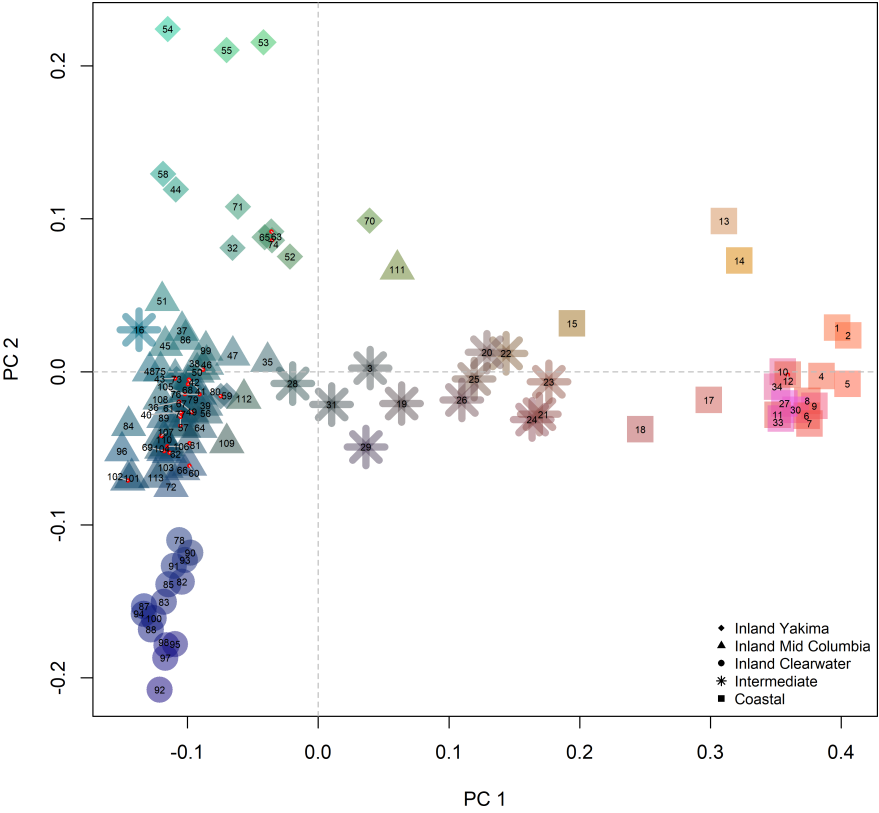


Figure 1:

Figure 3. PCA of adaptive markers (2,3,6,9) for all populations identified by predominate genotype (prema-
ture, mature, heterozygous). See Table S1 for collection names.

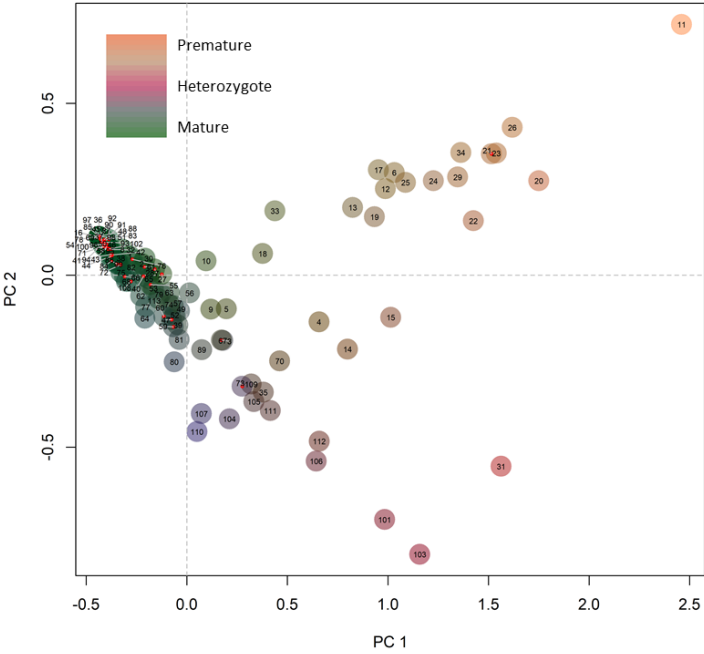
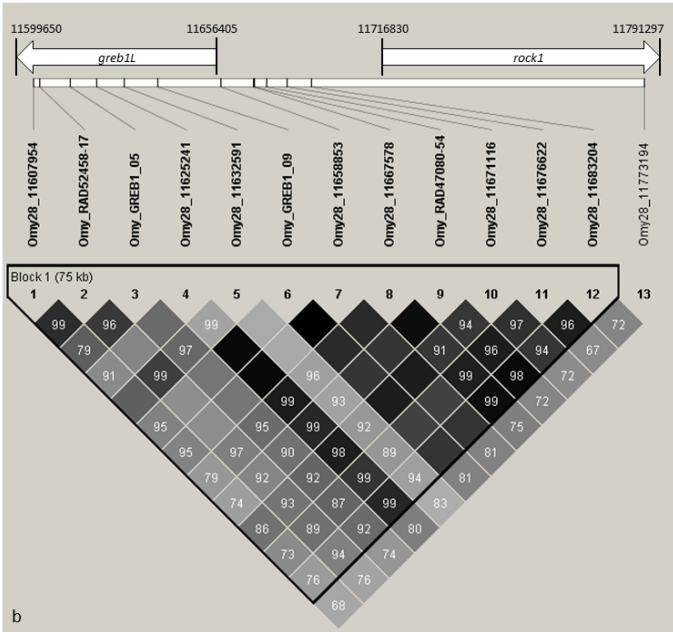
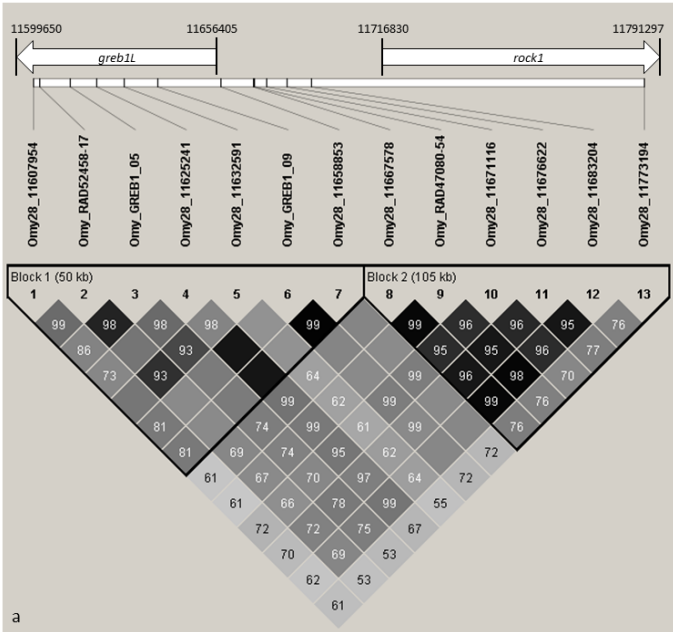


Figure 2:

Figure 4. R^2 values defined by LDcorSV and blocks defined with solid spine LD in Haploview for all
populations (a) genotyped at 13 markers. Coastal (b) and inland (c) populations (based on DAPC) R^2
values of 13 markers and blocks defined with solid spine LD in Haploview.



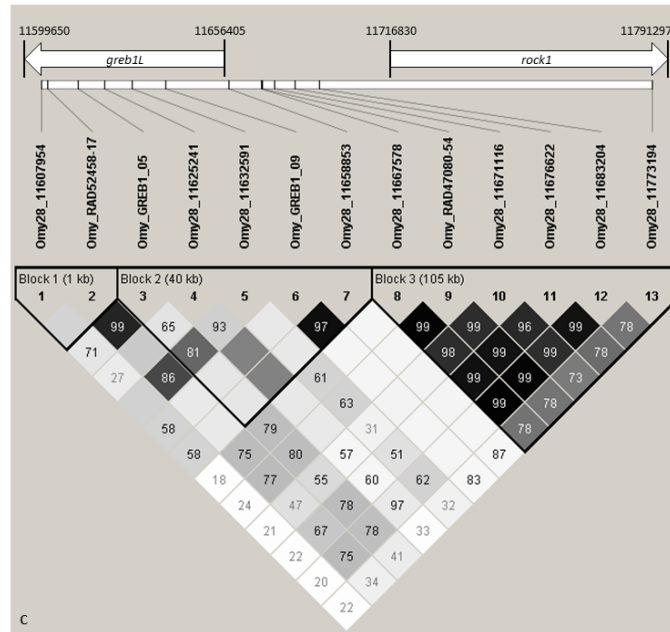


Figure 5. The map demonstrates the proportions of individuals at each collection location with the five haplotypes from markers 2, 3, 6, and 9. These 4 markers were evaluated to include all collections. Populations with 100 individuals or greater were reduced to 100 with the same haplotype proportions to keep the circles on the map as visible as possible. See Table S1 for collection names and exact genotype proportions.

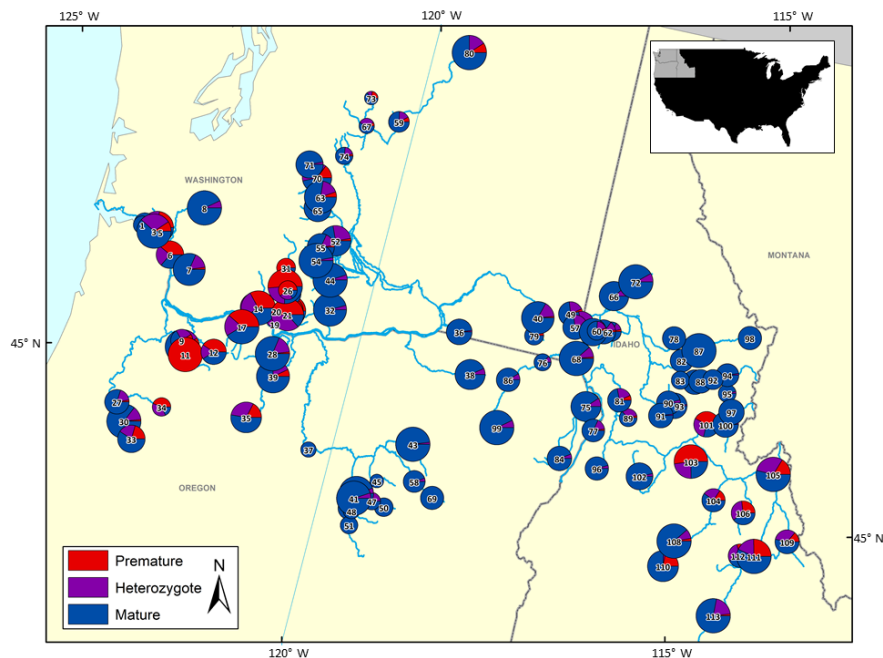
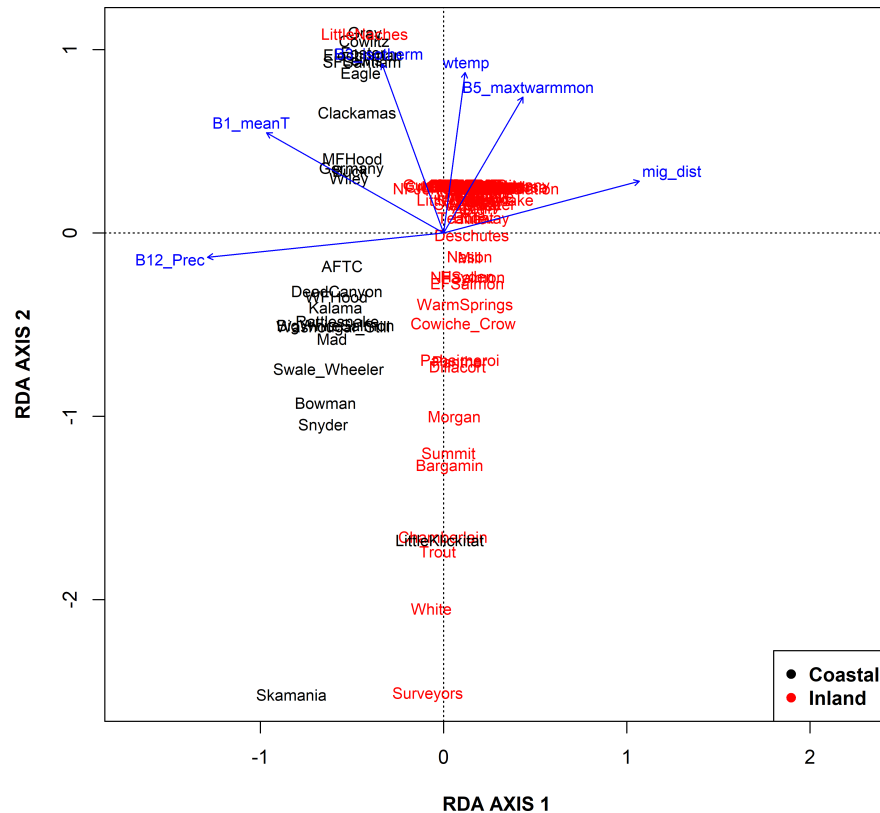


Figure 6. RDA of all collections in Columbia River basin to model the degree to which the variation in environmental variables explains the variation in allele frequencies for candidate markers for all collections

in the *greb1L* haplotype block (2,3,6). The populations are represented by text and colored black or red in accordance with their lineage determined by DAPC in adegenet. The arrows spatially denote a significant influence of environmental variables and the length of the arrow indicates the extent of the effect. Environmental variables retained were migration distance, minimum temperature warmest month, August water temperature over a 20-year average, annual mean temperature, isothermality, and annual precipitation.



Hosted file

Tables_Figures.docx available at <https://authorea.com/users/316699/articles/446837-distribution-of-genetic-variation-underlying-adult-migration-timing-in-steelhead-of-the-columbia-river-basin>

Hosted file

Appendix_S1_materials.docx available at <https://authorea.com/users/316699/articles/446837-distribution-of-genetic-variation-underlying-adult-migration-timing-in-steelhead-of-the-columbia-river-basin>