Received: 14 July 2017 | Revised: 13 October 2017 | Accepted: 16 October 2017

DOI: 10.1111/mec.14407

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*)

Steven J. Micheletti | Amanda R. Matala | Andrew P. Matala | Shawn R. Narum

Columbia River Inter-Tribal Fish Commission, Hagerman Fish Culture Experiment Station, Hagerman, ID, USA

Correspondence
Steven J. Micheletti, Columbia River Inter-Tribal Fish Commission, Hagerman Fish Culture Experiment Station, Hagerman, ID, USA.
Email: mics@critfc.org

Abstract

Organisms typically show evidence of adaptation to features within their local environment. However, many species undergo long-distance dispersal or migration across larger geographic regions that consist of highly heterogeneous habitats. Therefore, selection may influence adaptive genetic variation associated with landscape features at residing sites and along migration routes in migratory species. We tested for genomic adaptation to landscape features at natal spawning sites and along migration paths to the ocean of anadromous steelhead trout (*Oncorhynchus mykiss)* in the Columbia River Basin. Results from multivariate ordination, gene–environment association and outlier analyses using 24,526 single nucleotide polymorphisms (SNPs) provided evidence that adaptive allele frequencies were more commonly associated with landscape features along migration paths than features at natal sites (91.8% vs. 8.2% of adaptive loci, respectively). Among the 45 landscape variables tested, migration distance to the ocean and mean annual precipitation along migration paths were significantly associated with adaptive genetic variation in three distinct genetic groups. Additionally, variables such as minimum migration water temperature and mean migration slope were significant only in inland stocks of steelhead that migrate up to 1,200 km farther than those near the coast, indicating regional differences in migratory selective pressures. This study provides novel approaches for investigating migratory corridors and some of the first evidence that environment along migration paths can lead to substantial divergent selection. Consequently, our approach to understand genetic adaptation to migration conditions can be applied to other migratory species when migration or dispersal paths are generally known.

KEYWORDS
adaptation, fish, landscape genetics, migration, population genetics – empirical, RADseq

1 | INTRODUCTION

Landscape genomics can reveal how climatic and topological variables influence population structure and local adaptation across diverse taxa (Manel et al., 2010; Sork et al., 2013; Storfer et al., 2007). Populations distributed across heterogeneous landscapes are expected to experience selective pressures from abiotic factors that will shape adaptive variation. However, many broadly distributed species exhibit complex patterns of dispersal and migration across their distribution and will travel through suboptimal habitat to reach preferred residing areas for reproduction, juvenile rearing or overwintering (Dingle & Drake, 2007; Tews et al., 2004). Thus, understanding genetic adaptation to the environment requires information about where organisms reside, the paths they may travel and the

duration of travel. While previous studies have used landscape genomic tools to understand how landscape variables at residing sites are associated with adaptive genomic variation (e.g., Cornelis et al., 2010; Thomas, 2010), fewer have investigated landscape variables along migration paths (e.g., Bourret, Dionne, Kent, Lien, & Bernatchez, 2013; Hecht, Matala, Hess, & Narum, 2015) and none have explicitly compared patterns of divergent selection between landscape variables experienced along migration paths to those at residing sites.

Tracking complex movement patterns is difficult for many species; however, anadromous, philopatric fishes have predictable migration paths from the ocean to their freshwater natal spawning and rearing sites (Quinn, 2011; Shafer, Northrup, Wikelski, Wittemyer, & Wolf, 2016). Accordingly, anadromous fishes provide an ideal system for understanding adaptation along migration paths in addition to natal habitat. Studies of adaptation among anadromous salmonids often focus on landscape features in the natal environment while incorporating limited information about migration paths. For instance, Hecht et al. (2015) found that adaptive genetic variation in Chinook salmon (*Oncorhynchus tshawytscha)* across their North American range was primarily associated with precipitation and isothermality at natal sites as well as migration distance to the ocean. Similarly, a study of Atlantic salmon (*Salmo salar*; Bourret et al., 2013) used single nucleotide polymorphisms (SNPs) to discover loci strongly correlated with variables such as summer air temperature and geological characteristics of entire watersheds that potentially drive homing. While these studies have successfully identified natal habitat characteristics that influence adaptive genetic variation and included migration distance as a general parameter, anadromous fish will undoubtedly be impacted by specific habitat characteristics during migration. However, approaches to investigate selection of specific variables through migratory corridors are currently lacking or not well documented.

Steelhead trout (*Oncorhynchus mykiss*), like many other salmonids, migrate from the ocean to freshwater to spawn and return to their birthing sites (Kendall et al., 2014). In the Columbia River Basin, steelhead comprise two distinct phylogenetic lineages: the interior lineage east of the Cascades and the coastal lineage west of the Cascades (Brannon, Powell, Quinn, & Talbot, 2004; Quinn, 2011). Across these major lineages, steelhead exhibit multiple life history traits, such as divergent migration timing in adults known as winter-run (ocean maturing) and summer-run (freshwater maturing; Hess, Zendt, Matala, & Narum, 2016). Coastal streams often support populations of both summer-run and winter-run steelhead, whereas interior populations east of the Cascades Mountain range are exclusively summer-run (Myers, Aydin, Walker, Fowler, & Dahlberg, 1996). The Cascades also mark a geographic boundary for climatic features such as precipitation and temperature (Zobel, McKee, Hawk, & Dyrness, 1976). Thus, interior populations encounter different climatic conditions than coastal populations which may influence adaptation (e.g., water temperature and flow regimes; Coble, 1961; Richter & Kolmes, 2005). Additionally, due to further distances between the ocean and natal sites, interior populations are expected to face stronger selective pressures from anthropogenic modifications to the landscape (e.g., dams), which may serve as energetic barriers (Raymond, 1988).

In this study, we examined genomic variation in anadromous steelhead trout populations that occur across highly diverse environments to test whether adaptive variation is under selection along migration paths in addition to natal habitat used for spawning and juvenile rearing. Steelhead trout populations in the Columbia River Basin occur in streams throughout a series of water networks representing distinct biomes, ranging from high precipitation streams near the coast to spawning grounds located over 1,400 km upstream in rivers that originate in high mountain tributaries but flow through high desert plains with less hospitable conditions. Consequently, we hypothesized that (i) adaptive genetic variation will be strongly associated with habitat variation experienced along migration paths and (ii) that landscape variables associated with adaptive genetic variation will differ between regional genetic groups.

2 | METHODS

2.1 | Tissue sampling

Between 2000 and 2013, we sampled 3,244 individuals across 56 known steelhead trout natal tributaries throughout the Columbia River Basin (Figure 1; Table 1). Due to their philopatric behaviour, little straying occurs in *O. mykiss* and natal tributaries are expected to support independent populations even when tributaries are proximate (Blankenship et al., 2011; Westley, Quinn, & Dittman, 2013). Fin tissues were sampled nonlethally from either juvenile or adult fish caught by traps, weirs and electrofishing techniques that precede well-characterized natal tributaries and were stored in ethanol or dried on Whatman paper. Collections included localities from diverse coastal and inland environments across the Columbia River Basin which largely differ between the major genetic lineages (Brannon et al., 2004; Quinn, 2011). We specifically targeted populations of anadromous steelhead by excluding sample locations above impassable barriers (i.e., waterfalls and dams) where nonmigratory *O. mykiss* are likely to reside.

2.2 | Landscape measurements

Using natal-site collection coordinates, we determined the migration path between each natal site and the Pacific Ocean with the Landscape Genetics arctoolbox (Etherington, 2011) in ArcGIS 10.3.1 (ESRI, Redlands, CA) along the national hydrography stream network (U.S. Geological Survey, 2013) using a 30-m grid resolution. These tools together determine the most direct route along permanent water sources that lead to the mouth of the Columbia River (Figure 1). We confirmed accuracy of each calculated migration path by manually tracing each path back to the ocean in ArcGIS and comparing those to known migration routes of Columbia River populations.

A variety of commonly tested climatic and topological variables have previously been associated with genetic variation in salmonids

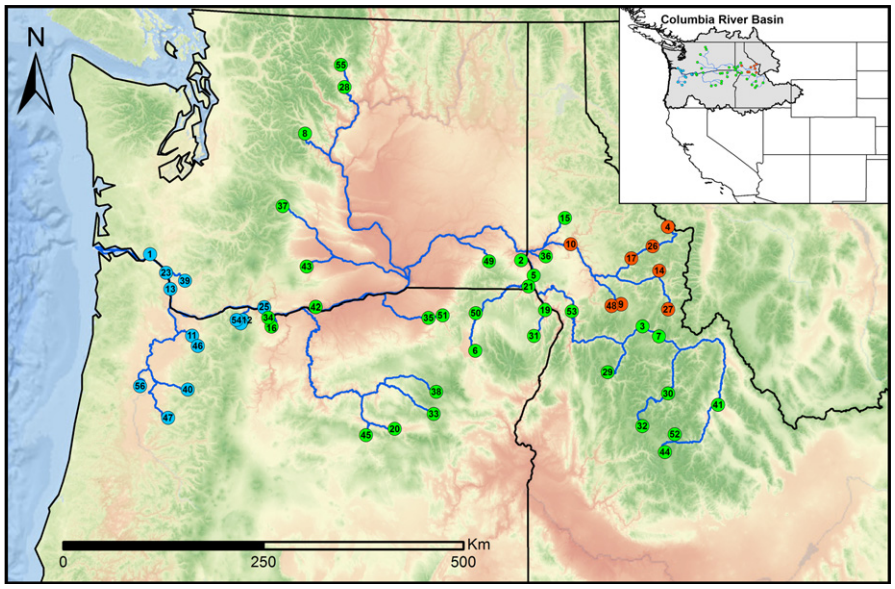

FIGURE 1 Steelhead collection localities coloured by genetic cluster (blue = coastal, green = inland, red = inland-Clearwater). Blue lines indicate each population's migration path to the ocean. Map layer shows mean annual temperature for the warmest quarter with transition from colder (green) to warmer (red) [Colour figure can be viewed at wileyonlinelibrary.com]

(Bourret et al., 2013; Hecht et al., 2015; Matala, Ackerman, Campbell, & Narum, 2014; Meeuwig, Guy, Kalinowski, & Fredenberg, 2010; Narum, Zendt, Graves, & Sharp, 2008). We chose to incorporate a breadth of both frequently and seldom used landscape variables that are categorized into three broad categories that potentially influence adaptation: temperature-related, precipitation-related and topography-related. Specifically, we used the full suite of climate variables from the BIOCLIM database ("bio1"–"bio19"; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), land cover-related variables from the National Land Cover Database (Homer et al., 2015), modelled August stream temperatures from the NORWEST database (Isaak et al., 2011), elevation from the national elevation data set (USGS) and wind velocity and solar radiation from the Vaisala weather database (Vaisala, Vantaa, Finland). We additionally calculated variables directly in ARCGIS 10.3.1 using a digital elevation map (USGS), such as slope, roughness and heat load index. Finally, we extracted stream-related variables from the National Hydrography Dataset, comprising of migration distance to the ocean, stream order and number of dams along migration paths (USGS).

We measured landscape variables both at natal sites and along the migration path for each population using the Spatial Analyst Toolbox in ArcGIS (ESRI, Redlands, CA). For natal-site variables, we measured each landscape value as the mean of a 5-km buffer around the collection site to account for natal habitat variation (Micheletti & Storfer, 2017). Migration path measurements, on the other hand, encompass a long geographic extent, only some of which is shared between populations. Therefore, we extracted the mean, minimum, maximum and range of each landscape variable along each population's migration route. In some instances, we calculated the sum of a variable (e.g., number of dams crossed) or variety (e.g., variety of stream orders crossed). Together, these produced 137 unique measurements across 34 unique landscape variables (available on Dryad). To avoid type I error in downstream analyses, we removed landscape measurements with a pairwise correlation >0.75 (Asuero, Sayago, & Gonzalez, 2006), retaining biologically relevant variables per previous salmonid studies (Hecht et al., 2015; Olsen et al., 2011). This initial filter was intended to remove highly correlated variables and retain putatively informative variables (McGaughran, Morgan, & Sommer, 2014; Micheletti & Storfer, 2017) for further analyses with subsequent identification of autocorrelation. After eliminating highly correlated variables, we ultimately retained 45 measurements across 19 unique landscape variables (Table 2).

These 45 landscape measurements were evaluated to assess habitat heterogeneity across the Columbia River Basin. First, we performed one-way analysis of variance (ANOVA) paired with a Tukey's range test (Tukey, 1949) on each landscape measurement between each identified genetic group to determine significant regional habitat variability. Additionally, we compared environmental characteristics holistically using the "envfit" function in the R package "VEGAN" (Oksanen et al., 2013) to produce a principal component analysis (PCA) plot based solely on landscape measurements. Together, these methods identify landscape features that are prevalent within and among regions inhabited by the identified genetic groups.

2.3 | DNA sequencing and summary statistics

We used restriction site-associated DNA (RAD) sequencing (Baird et al., 2008; Miller, Dunham, Amores, Cresko, & Johnson, 2007) to identify thousands of SNPs from raw Illumina sequence data. We implemented the STACKS v1.40 pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) to discover SNPs de novo and generate genotypes. In this study, the Stacks de novo pipeline was preferred because it was expected to produce fewer missing genotype data and higher genotype accuracy than the similar performing reference-based pipelines (Torkamaneh, Laroche, & Belzile, 2016). The process_radtags module was used for demultiplexing, trimming and quality filtering of raw data, followed by ustacks, cstacks and sstacks modules, which assemble reads, compile a catalog and align individuals to the catalog (Table S1). We implemented stringent quality filters using the "populations" module in Stacks and custom

TABLE 1 Collection information for all anadromous *Oncorhynchus mykiss* populations used in our analyses. Cluster corresponds to the three major genetic groups identified: coastal (C), inland (IN) and inland-Clearwater (ICW), and coordinates correspond to the collection tributary. The life stage of individuals at each collection locality are denoted as either adult (C), juvenile (J) or a mix of both (J/A). Basic descriptive statistics such as number of individuals used per population (*N*), expected average heterozygosity (H_E) and averaged observed heterozygosity (H_O) are included

ID#	Name	Abbr.	Subbasin	Cluster	Stage	Lat	Long	*N*	H_E	H_O
1	Abernathy	ABER	Elochoman	C	A	46.23	−123.15	21	0.16	0.07
2	Asotin	ASOT	Asotin	IN	J/A	46.32	−117.14	45	0.17	0.1
3	Bargamin	BARG	Salmon	IN	J	45.57	−115.19	24	0.16	0.09
4	Boulder	BOUL	Clearwater	ICW	J	46.68	−114.74	47	0.15	0.08
5	Captain John	CAPJ	Snake	IN	J	46.15	−116.93	43	0.16	0.13
6	Catherine	CATH	Grande Ronde	IN	J	45.31	−117.87	36	0.16	0.11
7	Chamberlain	CHAM	Salmon	IN	J	45.45	−114.93	48	0.16	0.11
8	Chiwaukum	CHIW	Wenatchee	IN	J	47.69	−120.74	34	0.17	0.08
9	Crooked	CROO	Clearwater	ICW	A	45.82	−115.53	45	0.15	0.08
10	Dworshak	DWOR	Clearwater	ICW	A	46.5	−116.33	27	0.15	0.13
11	Eagle	EAGL	Willamette	C	J/A	45.35	−122.38	46	0.17	0.09
12	East Fork Hood	EFHJ	Hood	C	J/A	45.56	−121.59	46	0.18	0.08
13	East Fork Lewis	ELEW	Lewis	C	A	45.85	−122.78	38	0.17	0.07
14	East Moose	EMOO	Clearwater	ICW	J	46.19	−114.9	40	0.15	0.09
15	East Potlatch	EPOT	Clearwater	ICW	A	46.8	−116.42	34	0.15	0.09
16	Fifteen	FIFT	Fifteenmile	IN	J	45.51	−121.13	47	0.18	0.15
17	Fish/Lochsa	FISH	Clearwater	ICW	A	46.33	−115.35	47	0.15	0.1
18	Parkdale	PAHH	Hood	C	J	45.52	−121.62	46	0.18	0.11
19	Cow	IMNA	Imnaha	IN	J	45.77	−116.75	14	0.15	0.08
20	John Day Main	JDMA	John Day	IN	J	44.41	−119.12	30	0.16	0.09
21	Joseph	JOSE	Grande Ronde	IN	A	46.03	−117.02	46	0.16	0.13
22	Kalama - Summer	KALS	Kalama	C	A	46.03	−122.87	46	0.17	0.08
23	Kalama - Winter	KALW	Kalama	C	A	46.03	−122.87	36	0.17	0.11
24	Klickitat - Summer	KLIS	Klickitat	C	A	45.72	−121.26	132	0.18	0.11
25	Klickitat - Winter	KLIW	Klickitat	C	A	45.72	−121.26	99	0.18	0.11
26	Lake/Lochsa	LAKE	Clearwater	ICW	J	46.46	−115	42	0.15	0.09
27	Little Clearwater	LCLW	Clearwater	ICW	J	45.75	−114.78	47	0.15	0.09
28	Methow	LIBB	Methow	IN	J	48.23	−120.11	18	0.15	0.1
29	Lick Creek	LICK	Salmon	IN	J	45.06	−115.76	43	0.15	0.06
30	Loon	LOON	Salmon	IN	J	44.81	−114.81	45	0.15	0.07
31	LittleSheep	LSHE	Imnaha	IN	J	45.48	−116.93	37	0.16	0.11
32	Marsh	MARS	Salmon	IN	J	44.45	−115.23	45	0.15	0.1
33	Upper John Day	NFJD	John Day	IN	J	44.59	−118.51	46	0.16	0.08
34	Mill Creek	MILL	Mid. Columbia	IN	J	45.61	−121.19	46	0.18	0.14
35	Minthorn	MINT	Umatilla	IN	J/A	45.67	−118.62	47	0.17	0.13
36	Mission	MISS	Clearwater	IN	J	46.37	−116.74	40	0.16	0.1
37	Naches -Nile	NACH	Yakima	IN	J	46.86	−121.05	46	0.17	0.12
38	Middle John Day	MFJD	John Day	IN	J	44.84	−118.48	47	0.16	0.1
39	North Fork Lewis	NLEW	Lewis	C	A	45.96	−122.56	40	0.17	0.07
40	Little Rock	NSAN	Willamette	C	J	44.75	−122.4	28	0.16	0.07
41	Pahsimeroi	PAHH	Salmon	IN	A	44.66	−114.03	29	0.16	0.09
42	Rock	ROCK	Mid. Columbia	IN	J	45.75	−120.44	40	0.16	0.07
43	Satus	SATU	Yakima	IN	J	46.2	−120.61	47	0.16	0.13

(Continues)

TABLE 1 (Continued)

ID#	Name	Abbr.	Subbasin	Cluster	Stage	Lat	Long	*N*	H_E	H_O
44	Sawtooth	SAWN	Salmon	IN	J/A	44.15	−114.88	41	0.16	0.1
45	South John Day	SFJD	John Day	IN	J	44.33	−119.57	80	0.16	0.1
46	Skamania Stock	SKAM	Willamette	C	A	45.24	−122.28	47	0.17	0.13
47	South Santiam	SSAN	Willamette	C	J/A	44.42	−122.67	31	0.16	0.08
48	Tenmile	TENM	Clearwater	ICW	J	45.81	−115.68	41	0.15	0.08
49	Tucannon	TUCN	Tucannon	IN	A	46.31	−117.66	46	0.17	0.1
50	Grande Ronde	UGRT	Grande Ronde	IN	J	45.73	−117.86	43	0.16	0.12
51	Umatilla	UMAT	Umatilla	IN	J	45.7	−118.4	46	0.17	0.11
52	Yankee Fork	WFYF	Salmon	IN	J	44.35	−114.73	46	0.15	0.1
53	Whitebird	WHIT	Salmon	IN	J	45.75	−116.32	44	0.16	0.11
54	West Fork Hood	WHOO	Hood	C	J	45.56	−121.69	45	0.17	0.15
55	Winthrop	WNFH	Methow	IN	J	48.48	−120.19	47	0.17	0.15
56	West Willamette	WWIL	Willamette	C	J	44.75	−123.15	44	0.16	0.08

TABLE 2 Final list of unique landscape variables used in analyses. All variables were calculated at the collection spawning site (S) and along the migration path to the ocean (M). Migration path is calculated as the mean, maximum, minimum, variety, sum or range of values of each variable. Each landscape variable was categorized as either a temperature-related variable (Temp.), precipitation-related variable (Precip.), or a topography-related variable (Topo). The complete list of unique 45 variables can be found in Table S5

Notation	Description	Unit	Res. (m)	Class.	Source
Atemp.Yr	BIO1 - Annual mean temperature	°C	1,000	Temp.	WorldClim
Trange	BIO2 - Mean diurnal range temperature	°C	1,000	Temp.	WorldClim
Isotherm	BIO3 - Isothermality (BIO2/BIO7)	°C	1,000	Temp.	WorldClim
Atemp.Wm	BIO10 - Mean temperature, warmest ¼	°C	1,000	Temp.	WorldClim
Atemp.Cd	BIO11 - Mean temperature, coldest ¼	°C	1,000	Temp.	WorldClim
Precip.Yr	BIO12 - Annual precipitation	mm	1,000	Precip.	WorldClim
Precip.Wt	BIO16 - Precipitation, wettest ¼	mm	1,000	Precip.	WorldClim
Precip.Dr	BIO17 - Precipitation, driest ¼	mm	1,000	Precip.	WorldClim
Heatload	Head load index	hli	30	Temp.	ESRI
Slope	Stream slope	°	30	Topo.	ESRI
Wtemp	Aug water temperature (20-year mean)	°C	30	Temp.	NorWeST
Elev	Elevation	m	30	Topo.	USGS
Geo	Proportion of primary geological surface	%	500	Topo.	USGS
Strord	Stream order	#	30	Topo.	USGS
Migdist	Migration distance to ocean	km	30	Topo.	USGS
Canopy	Percentage of canopy cover	%	30	Topo.	NLCD
Wind	Wind velocity	m/s	5,000	Topo.	Vaisala
Solar	Solar radiation	w/m^2	3,000	Temp.	Vaisala
Ndam	Number of dams	#	30	Topo.	USGS

scripts to remove poorly sequenced individuals and loci. Quality filters included removing any read with low quality scores and/or an uncalled base. Further, cataloged loci that were not sequenced in at least 70% of individuals, did not have a minor allele frequency >0.01 or had excessive polymorphisms (>4 SNPs per tag) were removed. Because end portions of reads are more prone to errors (Catchen et al., 2013), we also trimmed all reads to 75 bases and only retained the first variable SNP in a tag. Apart from improving read quality, this practice additionally eliminates known physical linkage, which can interfere with population-genetic statistics (Willis, Hollenbeck, Puritz, Gold, & Portnoy, 2017). To remove potential null alleles and homeologs, we tested for deviations from Hardy–Weinberg equilibrium (HWE) by locus and by population using the HWE exact test in GENEPOP 4.2.2 (Rousset, 2008). Additional gene duplications were targeted by aligning RAD sequences from five double-haploid individuals (Hecht, Thrower, Hale, Miller, & Nichols, 2012) which were

expected to be homozygous at all loci. If any of the double-haploid individuals were heterozygous at a locus, it was removed as a potential paralog. We additionally calculated population statistics such as observed heterozygosity, expected heterozygosity and F_{ST} using the R package "ADEGENET" (Jombart & Ahmed, 2011; R Core Team 2015).

2.4 | Population structure

To determine the number of genetic groups represented by our samples in the Columbia River Basin, we performed a discriminant analysis of principal components (DAPC) on individual genotypes in the R package "ADEGENET" (Jombart & Ahmed, 2011). This multivariate method uses sequential *K*-means and model selection to infer genetic groups by partitioning genetic variation into two components, a between-group and a within-group (Jombart, Devillard, & Balloux, 2010). We determined the number of genetic groups by running 25 iterations of the "find.clusters" module of "adegenet" for a range of $K = 1$ to $K = 20$ and averaging the Bayesian information criterion (BIC) across the 25 iterations and standard deviation estimated for each value of *K*. We then chose the most appropriate value of *K* by the rate change in BIC value using the method described in Evanno, Regnaut, and Goudet (2005). To complement this analysis, we also ran FASTSTRUCTURE (Raj, Stephens, & Pritchard, 2014) for $K = 1$ to $K = 20$ and selected the number of groups that maximized marginal likelihood. Finally, to represent structure at the population level, we performed a principal component analysis (PCA) on population major allele frequencies using the "ADEGENET" package in R.

2.5 | Redundancy analyses (RDA)

To estimate the degree to which genomic variation is influenced by environmental variables, we implemented a suite of redundancy analyses (RDA) using the "VEGAN" R package (Oksanen et al., 2013) on a basinwide scale and within each genetic lineage. This approach is a multivariate analog of linear regression, which examines how much of the variation in one set of variables explains the variation in another set of variables (Borcard, Legendre, & Drapeau, 1992). This tool is applicable to the field of landscape genomics because it can determine how a matrix of landscape variables explains variation in adaptive allele frequencies while overcoming linearity assumptions (Kierepka & Latch, 2015). To account for the effect of genetic structure and geographic distance, RDA can be performed on the residuals from covariates bases on genetic group assignment (Price et al., 2006). As collection localities with small stream distances between them tend to be closely related in salmonids (Beacham et al. 2006; Waples, Teel, Myers, & Marshall, 2004), genetic structure indirectly accounts for geographic stream distance among sites while removing the necessity to include additional covariates that may overfit the model (Hecht et al., 2015; Zhang 2014). Constraint scores from RDA provide a measure of the direction and magnitude of the correlation between landscape variables in the model and genomic allele frequency differentiation. Squaring the constraint scores can then be used to assess the magnitude of contribution of the landscape factors on SNP variation (i.e., landscape variables contributing more variation will have larger constraint scores; Wang, 2009). Further, squared SNP RDA scores determine how strongly each SNP is associated with landscape factors, and thus can be used to identify outlier SNPs that are associated with the suite of landscape variables. Using genetic group as a covariate, we first performed permutation tests globally, then within each genetic group using subclustering as a covariate to determine normalized landscape variables with significant effects in the model ($\alpha = 0.05$) based on a comparison to 1000 random permutations of the data. We performed final RDAs using only landscape variables that were significant in the permutation test (Hecht et al., 2015; Wang, 2009), identifying those having the greatest contribution to genomic variation in each genetic group.

2.6 | Outlier tests and environmental associations

We used two different classes of methods to detect non-neutral SNPs: outlier tests and gene–environment association (GEA) analyses (De Mita et al., 2013; Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015; Villemereuil & Gaggiotti, 2015). While RDA can identify outlier SNPs based on multivariate associations with significant landscape variables, it cannot identify univariate associations between single landscape variables and single SNP allele frequencies. Therefore, we performed three univariate GEA analyses to detect SNP loci whose allele frequencies are strongly associated with environmental variation: Bayenv2 (Günther & Coop, 2013), latent factor mixed modelling (LFMM; Frichot, Schoville, Bouchard, & François, 2013) and AutoLM. Bayenv2 uses a covariance matrix to estimate the neutral population structure (Coop, Witonsky, Di Rienzo, & Pritchard, 2010; Günther & Coop, 2013). Bayes factors are then calculated for each locus to compare models in which an environmental variable has a linear effect on allele frequency to a model in which allele frequency is explained by the covariance matrix alone. LFMM uses a Bayesian mixed model that employs a principal component to determine background level of population structure. Environmental variables are used as fixed effects and latent factors as random effects to determine loci–environment associations (Frichot et al., 2013). Finally, we developed a linear model that uses spatial autocorrelation as a covariate that we call AutoLM. AutoLM uses a linear mixed-effects model implemented by "LME4" R package (Bates, Maechler, Bolker, & Walker, 2015) to find significant correlations between allele frequencies and environmental measurements while simultaneously using spatial autocorrelation as a covariate in the model (also see Stucki et al. 2017). In addition to accounting for similarities in environmental characteristics between adjacent populations, another benefit of AutoLM is that it produces both significance (*p*-values) and correlation (R^2) values of variables whereby the fit of a model can be assessed to determine whether a SNP is associated with an environmental variable. Because all three GEA methods use different test statistics, we followed recommended thresholds to determine whether a given SNP was significantly associated with an environmental variable (Table S2).

To determine loci exhibiting high genetic differentiation between populations putatively due to selection, we implemented four different univariate outlier tests: BAYSCAN 2.1 (Foll & Gaggiotti, 2008), OUTFLANK (Whitlock & Lotterhos, 2015), PCAdapt (Luu, Bazin, & Blum, 2017) and BAYENV2 (X_TX;Günther & Coop, 2013). We used a diverse set of outlier tests because they implement different methodology to account for underlying population structure, rather than only determining SNPs that have significantly high or low F_{ST}. For instance, BayeScan compares differences between observed allele frequencies in a population and the expected allele frequencies under neutrality (Foll & Gaggiotti, 2008). OutFLANK determines the neutral distribution of F_{ST} population by trimming off the highest and lowest outliers (Whitlock & Lotterhos, 2015). PCAdapt uses a hierarchical factor model that uses K latent factors to estimate the neutral underlying population structure (Duforet-Frebourg, Bazin, & Blum, 2014) combined with a principal component analysis. Finally, BAYENV2 calculates X_TX, a F_{ST} analog which accounts for the covariance structure of populations (Günther & Coop, 2013). While these outlier techniques have demonstrated low false-positive rates in detecting loci under selection compared to other popular outlier detection methods (e.g., FDist2, Beaumont & Nichols, 1996; LOSITAN, Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008), it is common for significant SNPs to vary between analyses (Lotterhos & Whitlock, 2014; Narum & Hess, 2011). In an effort to standardize outlier tests, we used the method proposed in François, Martins, Caye, and Schoville (2016) to distribute *p*-values using a genomic inflation factor. Briefly, this correction attempts to balance type I error in conservative estimates of outliers and type II error in liberal estimates of outliers, a practice that has gained focus recently (see Lotterhos et al., 2017). Consequently, we used a threshold for significance of $p < .05$ and *q*-value <0.1 among standardized values with FDR corrections for multiple tests. Because the X_TX statistic produces Bayes factors (BF) rather than *p*-values, we did not apply any correction to this metric and used the threshold of $\log_{10}$ BF >2 (Günther & Coop, 2013).

2.7 | Genome assembly and ontology

To determine genomic regions potentially under selection, we aligned our SNP panel to a recent public release of the *O. mykiss reference* genome (Omyk_1.0; *GCA_002163495.1*, USDA/ARS) using Bowtie2 (Langmead & Salzberg, 2012) with a threshold of mapping quality score ≥2. Our SNP panel was displayed against the reference genome using the R package "QQMAN" (Turner, 2014) using squared SNP scores produced by the RDA (significant environmental outliers). We then assigned each SNP to a class of environmental variables (temperature-related, precipitation-related or topography-related; Table 2) based on its consensus association from GEA analyses. We specifically mapped SNPs in this fashion to determine whether any genomic regions were associated with a particular class of landscape variables. Finally, we searched for putative gene function by comparing our SNP sequences against the NCBI nucleotide sequence database in BLAST2GO (Conesa et al., 2005) because matching homologous genes may give insight into the genetic basis for adaptation to our panel of landscape variables.

3 | RESULTS

3.1 | Genetic data and structure

A final sample size of 2451 individuals met quality genotype filters and were thus retained for analysis (Table 1). Additionally, de novo SNP discovery analyses identified 24,526 SNP markers that passed filtering criteria and filtered individuals, on average, had 14% (±5%) of missing genotype data (Table S1). All genetic structure analyses identified three genetic groups based on putatively neutral markers: a coastal, inland and inland-Clearwater group. The individual-based DAPC supported three groups based on changes in BIC values (Figure S1) and illustrated very little introgression among lineages (Figure 2). FASTSTRUCTURE supported these results, identifying that the same three genetic groups maximize marginal likelihood in the model (Figure S2). Finally, an allele frequency PCA showed distinct population clustering of three groups, with two collections intermediate between the coastal and inland groups (KLIS, KLIW), and one population between the inland and inland-Clearwater groups (EPOT; Figure 3). These relationships between genetic groups are further represented by an unrooted neighbour-joining tree (Figure S3). While the aforementioned results were based on putatively neutral markers whereby all potential outliers from outlier tests and GEA analyses were removed, analyses that included all SNPS in the data set had no detectable effect on the outcome of genetic structure (Figure S4).

3.2 | Regional differences in environmental characteristics

Many landscape measurements were heterogeneous between genetic groups, with the largest differences being between the coastal and two inland genetic groups (Table S3). The environmental PCA also illustrated distinct differences between the coastal and

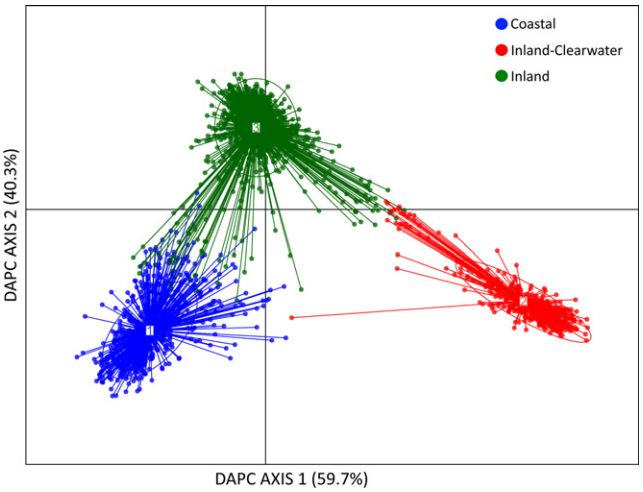

FIGURE 2 Plot of DAPC1 vs DAPC2 using individual genotypes at neutral SNPs showing three distinct clusters. Each circle represents an individual and colours correspond to genetic group (coastal = blue, green = inland, red = inland-Clearwater) [Colour figure can be viewed at wileyonlinelibrary.com]

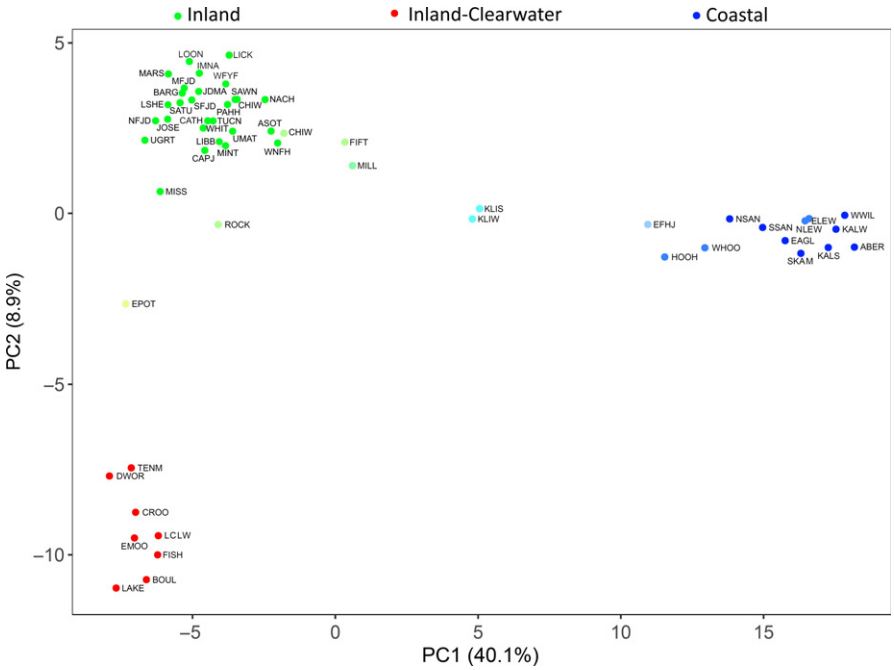

FIGURE 3 Allele frequency principle component analysis displaying PC1 vs. PC2 using neutral SNPs. Populations are coloured according to genetic group assignment, and names correspond to abbreviations in Table 1 [Colour figure can be viewed at wileyonlinelibrary.com]

inland groups (Figure S5). However, the inland-Clearwater group was intermingled with the inland group, suggesting that environment is more similar between these two groups. The ANOVA determined that only three landscape measurements are significantly different between all genetic groups: migration range of water temperature (Wtemp-Range.M; $p < .005$), migration minimum water temperature (Wtemp-Min.M; $p < .003$) and number of dams encountered along migration paths (Ndam-M; $p < .002$; Table S3). There were also unexpected instances where landscape measurements between the coastal and inland-Clearwater groups were not significantly different, but did differ between the inland-Clearwater and inland groups. These include riparian canopy cover (Canopy-Mean.M; Canopy-S) and natal-site precipitation of the driest quarter (Precip.Dr-S).

3.3 | Environmental contribution to divergent selection

Permutation tests with RDA identified that 22 of the 45 landscape measurements were significantly associated with genomic variation across all lineages (Figure 4). Results from RDA, represented as triplots, summarized the three main components of the analyses: RDA values for populations, SNP score values and statistically supported landscape variables. Landscape variables were represented by vectors originating from the centre, where vector length indicates the magnitude of influence of the trait to the overall SNP variance. Although highly correlated variables were removed, triplots indicated relationships between environmental variables and allele frequency variation. Consequently, the angle between vectors represents the correlation between environmental variables with similar relationships, whereby variables <45° were correlated and variables with angles >135° are negatively correlated. Top contributing models for each RDA analysis were determined by squared constraint scores, which also corresponded to the length of the vector in the RDA triplot (Table S4). Among all significant landscape variables, only the mean migration precipitation of the wettest quarter (Precip.Wt-Mean.M) and migration distance to the ocean (Migdist-M) were significant in all four RDA models (Figure 4). Minimum migration water temperature (Wtemp-Min.M) and mean slope of migration (Slope-Mean.M) were group-specific, whereby they were absent in the coastal RDA but present in the two inland RDAs.

Permutation tests performed on a basinwide scale identified 16 landscape variables that were significantly associated with genomic variation (Figure 4a). Of these 16 variables, only three were natal-site landscape measurements. Among all significant variables, eight were temperature-related variables, four were topology-related and four were precipitation-related. The top associated landscape variables, based on vector length and the mean squared constraint scores of RDA axes 1 and 2 (CS^2), were natal-site precipitation of the driest season (Precip.Dr-S; $CS^2 = 0.268$), migration distance (Migdist-M; $CS^2 = 0.197$) and migration mean temperature of the coldest month (Atemp.Cd-Mean.M; $CS^2 = 0.169$; Table S4).

The coastal group had 13 landscape variables that were significantly associated with genomic variation (Figure 4b). Of the 13 variables, three of the variables were natal-site measurements that were correlated with each other based on their position in the RDA triplot. Altogether, six of the variables were related to temperature, four were related to precipitation and three were related to topology. The top associations were with the migration maximum temperature of the warmest month (Atemp.Wm-Max.M; $CS^2 = 0.412$), migration distance (Migdist-M; $CS^2 = 0.375$) and migration maximum water temperature (Wtemp-Max.M; $CS^2 = 0.367$).

The inland group also had 13 significant variables associated with genomic variation (Figure 4c) with only three of the variables relating to natal-site measurements. Together, seven of the variables corresponded to temperature, three were related to precipitation and three were related to topology. The top variables were migration

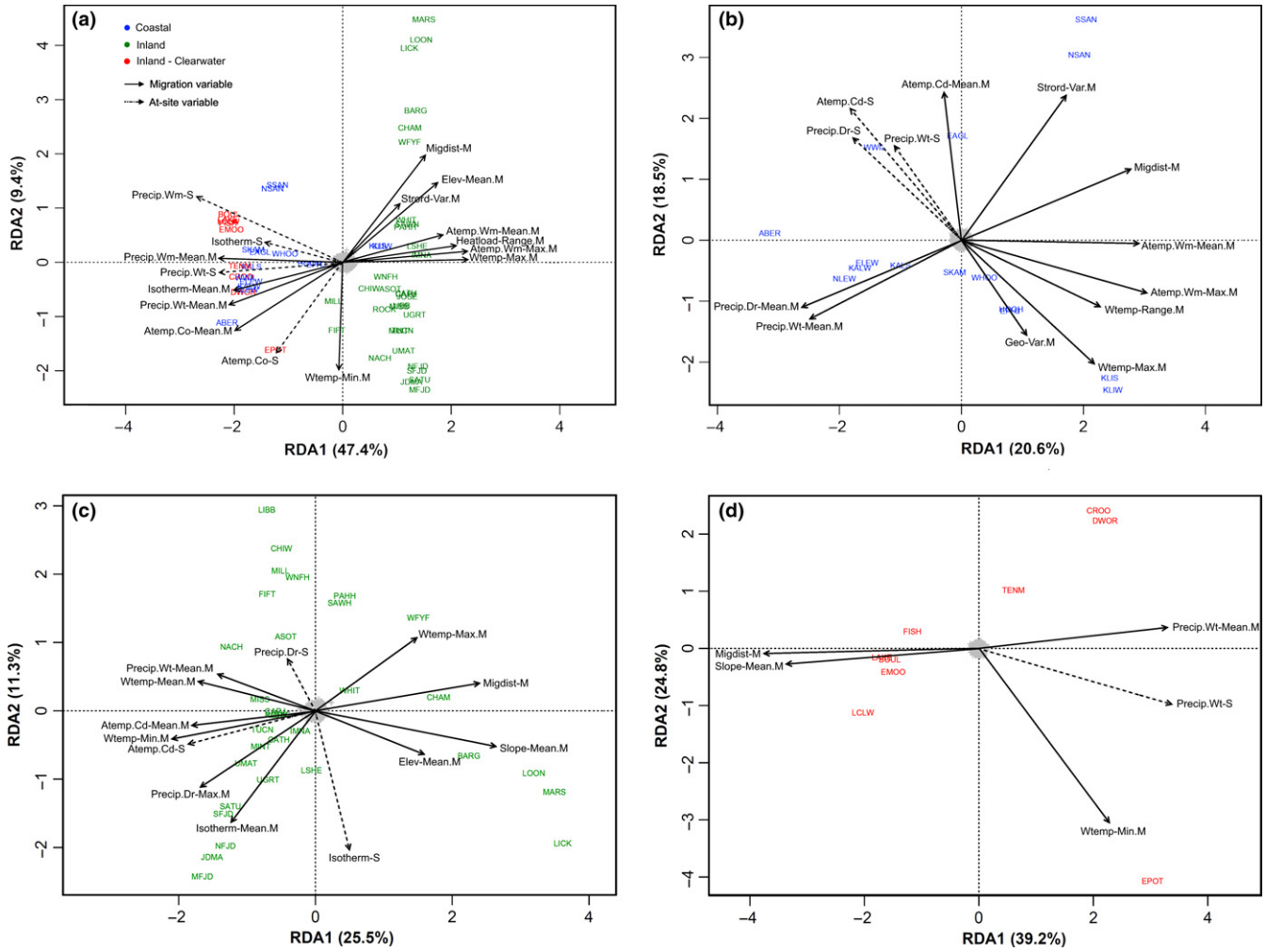

FIGURE 4 Triplots of (a) basinwide collections, (b) coastal group, (c) inland group and (d) inland-Clearwater group. The *x* and *y* axes of a triplot represent scaled RDA values for the sample populations, while individual population values for those axes are illustrated as the four-letter abbreviation for the population name (see Table 1). SNP score values for the RDA axes are represented as "+" points in the centre of the plots. Statistically supported environmental factors in the RDA models are represented as arrows emerging from the centre, whereby longer arrows correspond to a larger magnitude of effect. Natal-site variables are represented by dashed lines whereas migration variables are represented by solid lines [Colour figure can be viewed at wileyonlinelibrary.com]

mean slope (Slope-Mean.M; $CS^2 = 0.322$), migration distance (Migdist-M; $CS^2 = 0.262$) and migration minimum water temperature (Wtemp-Min.M; $CS^2 = 0.202$).

Finally, the inland-Clearwater lineage only had five significant variables (Figure 4d), one of which was a natal-site measurement. Together, one variable was related to temperature, two were related to precipitation and two were related to topology. While all variables had a large contribution to the model, migration minimum water temperature (Wtemp-Min.M; $CS^2 = 0.497$), natal-site mean annual precipitation (Precip.Yr-S; $CS^2 = 0.482$) and migration mean precipitation of the wettest month (Precip.Wt-Mean.M; $CS^2 = 0.462$) were the top contributors.

3.4 | Adaptive loci

Our suite of outlier tests identified 714 unique SNP markers across both basinwide and group-specific scales (Figure S6), and each test identified similar numbers of outliers: X_TX, 266; PCAdapt, 255; BayeScan, 253; OutFlank, 229. Conversely, GEA analyses identified an extensive list of 8691 unique SNP markers significantly associated with landscape variables (Figure S7) and the number of significant SNPs ranged between analyses: LFMM, 1823; AutoLM, 2609; Bayenv, 6713. While many gene–environment associations occurred due to the large number of total landscape variables, only 177 significant SNP markers were identified by all three GEA analyses, and 84 significant outliers were identified by at least three of the four outlier tests (Figures S6 and S7). Of these top outliers, 38 were shared between GEA and outlier tests, indicating that some of the most significant outliers were significantly associated with environmental variation.

Of all the significant GEA SNP markers, 74.2% were associated with a migration-related variables, whereas the other 25.8% were associated with natal-site variables. The primary associations were with temperature-related variables (55.4%), followed by topology (26.7%) and finally precipitation (17.9%). However, when considering

the 177 top GEA SNPs identified by all three analyses, 91.8% of were associated with migration variables and only 8.2% associated with natal-site variables. Following a similar trend, the primary associations between the top GEA SNPs were with temperature-related variables (57.1%), followed by topology (22.5%) and finally precipitation (20.4%).

Among the top 177 GEA SNPs and the 84 outlier SNPs, only 62.8% (164) of the SNPs had significant hits to genes in the NCBI database or were within 10 kb of annotated genes (Table 3). Within this collection of genes, there was no significant gene ontology enrichment based on the Fisher's exact test in BLAST2GO. Of the total 24,526 SNPs, 92% aligned to the *O. mykiss* reference genome with a mapping quality score ≥2 (Figure S8) and only 0.1% of these SNPs aligned to more than one location in the genome. When displayed as a Manhattan plot using squared SNP scores of the first RDA axis, the top 5% of SNPs corresponded to migration-related variables in every lineage (Figure 5).

Other adaptive SNPs had compelling linear, asymptotic and exponential relationships between allele frequencies and landscape measurements, yet did not have any significant BLAST hits (Figure 6). Some of these candidate SNPs showed no relationship on a basin-wide scale, yet were significant when only looking at a single genetic group (Figure S9), illustrating regionally specific selection pressures. SNPs with significant associations, such as these, are candidates for future studies and can be used in genotyping panels for large numbers of individuals (e.g., GT-SEQ; Campbell, Harmon, & Narum, 2015).

4 | DISCUSSION

This study demonstrates that migratory species may experience stronger selection pressure during adult migration than other stages of their life cycle such as breeding and juvenile rearing. Here, we found that adaptive genetic variation was more commonly associated with migratory landscape features as opposed to natal-site landscape features in an anadromous salmonid system with predictable migration routes. We also demonstrated that selective pressures vary among genetic groups, which experience different durations of exposure to habitat conditions due to variation in migration distance. Further, we identified markers exhibiting signatures of divergent selection in association with environmental features that can be used in the future to closely monitor adaptive variation in this species. Our study provides a novel approach for investigating divergent selection related to environmental features of both migratory corridors and natal habitat that should be effective for use in other species with known migratory routes.

TABLE 3 Subset of top outliers and gene–environment associated SNPs that also have putative gene functions. Outlier refers to the outlier tests that identified the SNP as significant: BayENV (X_TX), Outflank (OF), BayeScan (BS) or PcAdapt (PCA). GEA refers to the gene–environment association analyses that identified the SNP as significant: BayENV (X_TX), latent factor mixed models (LFMM) and linear mixed models with spatial autocorrelation (autoLM). Association is the top environmental variables that is associated with the SNP per GEAs. Top BLAST hit refers to the top significant BLAST hit of the 75-bp locus sequence. The full list of candidate outliers can be found on Dryad

SNP ID	Outlier	GEA	Association	Top BLAST hit	Putative function
10198_16	X_TX, OF, PCA, BS	N/A	N/A	GREB1 isoform X2	Maturation timing
15709_53	X_TX, OF, PCA, BS	N/A	N/A	GREB1 isoform X2	Maturation timing
30594_22	X_TX, OF, PCA, BS	N/A	N/A	GREB1 isoform X2	Maturation timing
1751_18	X_TX, OF, PCA, BS	Bay	Isotherm-Min.M	CXCL10-like chemokine	Immune function
12601_68	X_TX, OF, PCA, BS	Bay, autoLM	Atemp.Wm-Max.M	Metal transporter CNNM4-like	Sensory neuron functions
63879_8	X_TX, OF, PCA, BS	Bay, autoLM	Atemp.Wm-Max.M	Junctional associated with coronary artery disease-like	Cell adhesion
24343_29	X_TX, OF, PCA, BS	LFMM, Bay, autoLM	Atemp.Yr-Mean.M	Ras family member 11A-like	Regulator of transcription
25266_23	X_TX, OF, BS	Bay	Canopy-S	Hemicentin-2 isoform X1	Calcium ion binding
29559_69	X_TX, OF, BS	Bay	Geo-Var.M	Fibroblast growth factor 4A-like	Embryonic development
49637_74	X_TX, OF, BS	Bay	Isotherm-S	E3 ubiquitin-ligase RNF43	Protein ubiquitination
28236_38	X_TX, OF, PCA, BS	LFMM, Bay, autoLM	Migdist-M	Guanine nucleotide-binding-like 3	Cell division
35640_63	X_TX, PCA	N/A	N/A	Voltage-dependent R-type calcium channel subunit alpha-1E-like	Muscle contraction
14549_67	OF, BS	N/A	N/A	Homeobox Nkx	Differentiation of myocardium
28403_14	OF, PCA, BS	Bay, autoLM	Precip.Yr-Min.M	A Chain Crystal Structure Of The Sleeping Beauty Transposase Catalytic Domain	DNA binding
54441_29	X_TX, OF, BS	Bay	Slope-S	Hepatocyte growth factor-like	Cell determination
14549_67	X_TX, BS	Bay	Wtemp-Max.M	Sorting nexin-14 isoform X1	Synaptic transmission

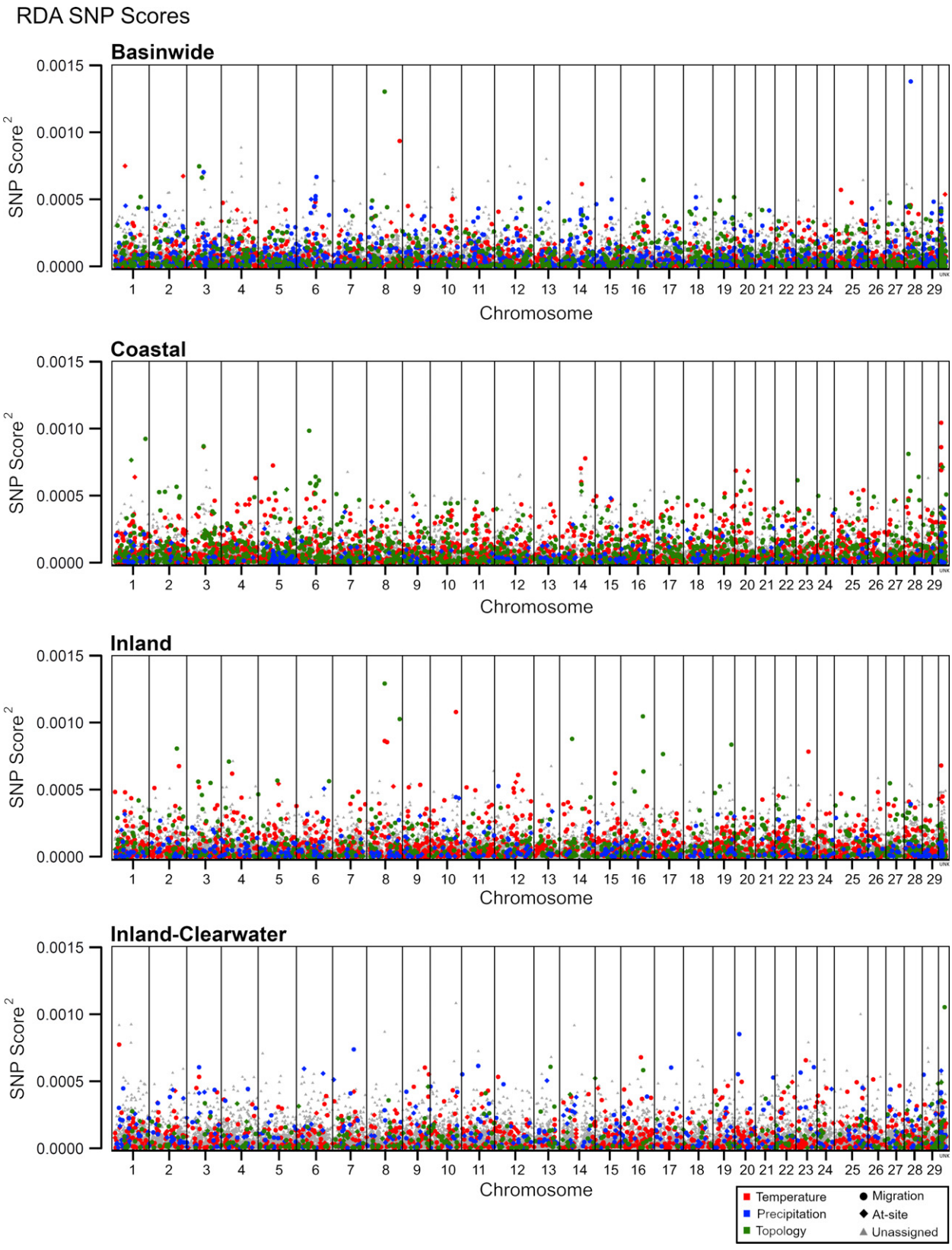

FIGURE 5 Manhattan plots representing the squared SNP scores (correlations) of RDA1 for (a) basinwide, (b) coastal, (c) inland and (d) inland-Clearwater groups across the *Oncorhynchus mykiss* reference genome. Coloured SNPs are those found to have a significant association with an environmental variable and correspond to temperature-related variables (red), precipitation-related variables (blue) and topography-related variables (green). Circular SNPs represent associations with migration landscape variables, whereas diamonds represent associations with natal-site landscape variables. Grey triangles are SNPs that do not have any associations with environmental variables [Colour figure can be viewed at wileyonlinelibrary.com]

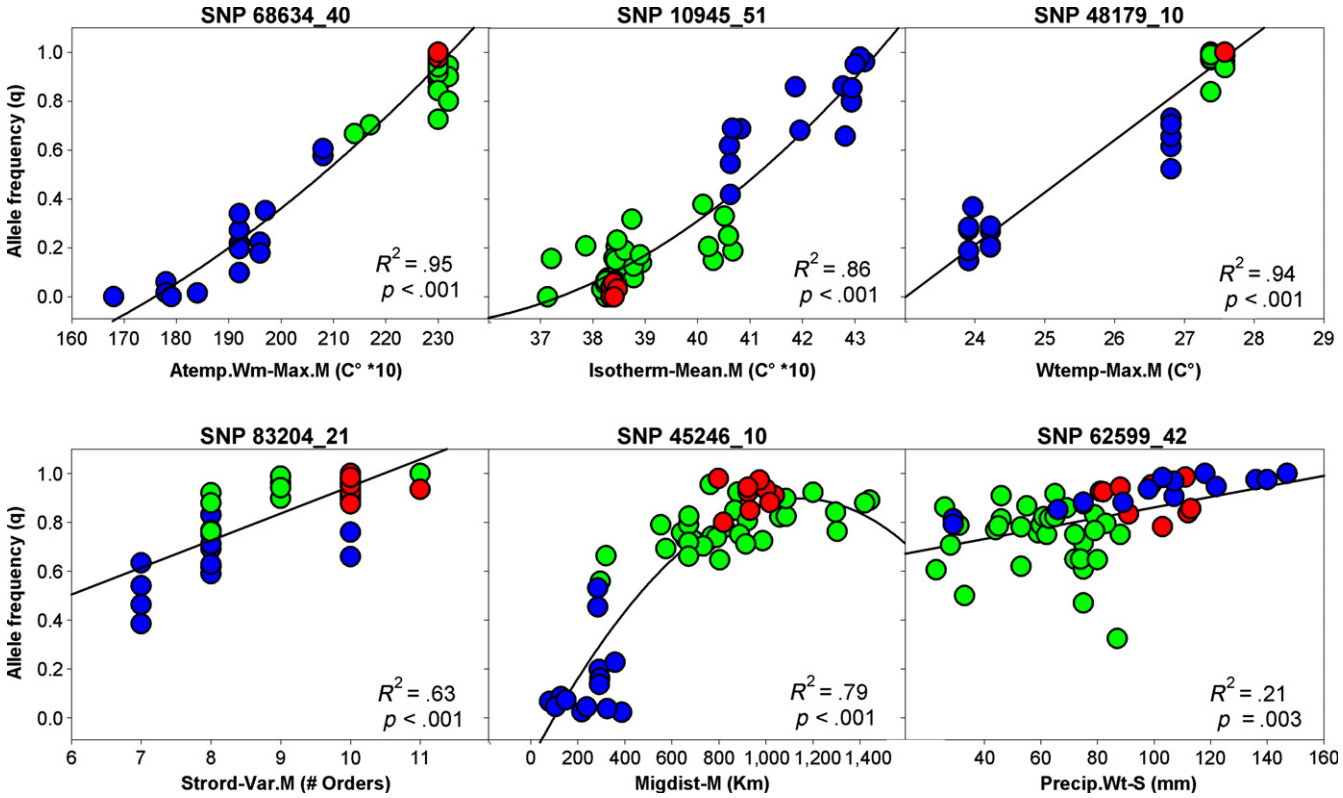

FIGURE 6 Examples of top gene–environment associations SNPs that show compelling relationships between environmental variation and allele frequency across the entire basin. Circles represent populations where colours of the circles correspond to lineage: blue is coastal, green is inland and red is inland-Clearwater. Additional examples can be found in Table S5, and the full list of outliers is available on Dryad [Colour figure can be viewed at wileyonlinelibrary.com]

4.1 | Migration vs. natal-site environmental variables

Our study revealed that significant adaptive variation was more often associated with tested variables in the migratory environment compared to those in the natal-site landscape. This observation likely reflects the erratic and heterogeneous conditions often encountered along migration paths as opposed to spawning and rearing sites where conditions may be extreme but more stable (Richter & Kolmes, 2005). Steelhead generally have specific requirements for spawning sites such as the presence of gravel, specific water temperatures and currents (Quinn, 2011). However, many populations of steelhead must swim through main stem rivers that vary in habitat suitability to reach spawning sites. This is illustrated by variables that did not significantly differ between lineages at natal sites, but were significant along migration paths. For example, natal-site water temperature was not significantly different between any genetic group, yet the migration water temperature range, maximum and minimum were significantly different between lineages. Consequently, multivariate analyses identified migration water temperatures as significant variables, but not natal-site water temperatures. While many populations across the basin spawn in tributaries with similar water temperatures, inland stocks must travel through desert regions with hot climates to reach spawning grounds in mountain streams. Therefore, it is apparent why variables such as water temperature, isothermality, slope and stream order are represented in migration paths, but not at natal sites. While there may be other important natal habitat variables that we did not include, our current data show that selection is occurring most strongly on loci due to conditions encountered during migration.

Selection pressure on migration habitat may be stronger in steelhead than other salmonids due to their iteroparity, especially for populations near the ocean that experience higher rates of repeat spawning. While estimates of repeat spawning are low in Columbia River Basin steelhead (0.5%–9.0%; Keefer, Wertheimer, Evans, Boggs, & Peery, 2008), a proportion of fish will still attempt multiple spawning journeys (Keefer et al., 2017; Matala et al., 2016) and face harsh migration conditions when energetic stores are low after spawning (Penney & Moffitt, 2015). Thus, more research on migration routes in semelparous species is needed to determine whether iteroparous species face stronger migration selection pressure.

4.2 | Significant environmental variables

Significant landscape variables that were identified in every RDA analysis are expected to have a broad impact on adaptation in *O. mykiss* throughout the Columbia River Basin. While these variables may not be the causal mechanism driving local adaptation, they are predicted to be correlated with an abiotic or biotic factor that is directly related to adaptation. Migration distance was one of the few

landscape variables that was consistently associated with adaptive genetic variation across every analysis. Migration distance is a measure of the distance both juvenile and adult steelhead must travel to reach the ocean at the beginning of outmigration or to return to spawning grounds after years in the ocean. In our populations, the distances required to travel to and from spawning sites ranges from 80 to 1440 km, posing obvious differences in selection pressure on energetics and metabolism. Migration distance has been a significant factor in previous landscape studies of anadromous salmonids (Hecht et al., 2015; Olsen et al., 2011). As such, many anadromous species that must make long journeys between the ocean and natal sites likely face similar pressures associated with the energetic costs of migration. While hydroelectric dams pose challenges for migration and contribute to juvenile and adult mortality (Raymond, 1979, 1988), we did not find any direct associations between number of dams and adaptive genetic variation. However, dams greatly influence environmental conditions such as flow and water temperature of the Columbia River, altering natural water conditions (Cushman, 1985). Thus, anthropogenic disturbances, including dams, likely contribute to harsh conditions along migratory corridors that were found to be associated with divergent selection in steelhead populations.

Another variable significant across all lineages was mean precipitation of the wettest quarter along migration paths. Drought conditions of low snowpack or rainfall may be commonly experienced in certain regions that result in low flows with warmer water temperatures than normal, creating extremely poor conditions in certain years with strong selection (Heath, Busch, Kelly, & Atagi, 2002; Petersen & Kitchell, 2001; Richter & Kolmes, 2005). These low flow years may specifically pose selection on genes involved with thermal tolerance (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013) or antipredator defence when refugia becomes limited, especially in juveniles (Brown & Smith, 1997; Valdimarsson & Metcalfe, 1998). In contrast, excessively high precipitation during the wettest months (November–January) increase water volume and stream flow, which may add another energetic barrier to fish attempting to reach spawning sites or the ocean. Therefore, it is possible that high precipitation along migration paths contributes to energetic costs along with migration distance across the entire basin.

Finally, a mix of water and air temperature variables in the migratory environment was associated with adaptation across every genetic group. As fish are ectotherms, warm regions, such as central Washington and Oregon, may pose strong thermal barriers during migration. Temperature has obvious implications on adaptive divergence in relation to cardiac performance and metabolism (Ineno, Tsuchida, Kanda, & Watabe, 2005; Perry, Danzmann, Ferguson, & Gibson, 2001; Taylor, 1991), and studies have found genetic bases for thermal tolerance in salmonids that reflects local adaptation (Eliason et al., 2011; Munoz, Farrell, Heath, & Neff, 2015; Narum et al., 2013). For instance, thermal-tolerant individuals can maintain higher heart rates and stroke volume in warmer temperatures, whereas thermal-intolerant individuals experience arrhythmia which affects metabolism and swimming performance (Ineno et al., 2005; Jain & Farrell, 2003). Therefore, temperature variables are expected to contribute to adaptive genetic variation.

One of the few natal-site variables that had large contributions to multiple RDA analyses based on squared constraint scores was precipitation of the driest quarter. Hecht et al. (2015) also identified this variable as important across North America Chinook salmon populations and attributed this strong association with snowmelt that generally occurs in the driest quarters (March–May), leading to high flows that precede juvenile outmigration. The same is expected in *O. mykiss* (Hand et al., 2016; Matala et al., 2014), and we posit that migration precipitation of the wettest quarter may have a larger impact on adult steelhead migrating to spawning sites, whereas natal-site precipitation of the driest quarter may have a larger impact on juveniles preparing to migrate to the ocean. Nonetheless, the multiple precipitation variables seen across lineages suggest that precipitation is important year-round.

4.3 | Differences in divergent selection among genetic groups

The identification of distinct coastal and inland groups in our study is concordant with genetic studies on steelhead in the Columbia River Basin (Blankenship et al., 2011; Brannon et al., 2004; Matala et al., 2014; Reisenbichler, McIntyre, Solazzi, & Landino, 1992; Utter & Allendorf, 1977). The discrete break between a coastal and inland lineage is due to sharp environmental clines across the Columbia River Basin (Brannon et al., 2004.) For instance, directly west of the Cascades, the coastal group faces significantly lower annual air temperature, higher annual precipitation and exists at lower elevations than the interior lineage. Consequently, many environmental variables that differed between the three genetic groups are reflected by regional-specific associations with adaptive genetic variation. For example, inland populations are generally found in high-elevation, cold-water tributaries and showed strong associations between minimum water temperatures, whereas coastal populations did not. This was also observed with slope, which was significantly higher in inland populations that must swim to higher elevations to spawn. Thus, region-specific associations illustrate that environmental selective pressures differ between genetic groups. These landscape variables that are significantly different between groups and associated with adaptive genetic variation throughout the basin likely contribute to the distinct groups found in the Columbia River Basin. Of the noncorrelated landscape variables, 66% differed between coastal and inland group, whereas only 14% were significantly different between the inland and inland-Clearwater group. While a vast range of variables differentiated the coastal and inland groups, minimum water temperature and water temperature range separated the inland from the inland-Clearwater group. The latter experiences both the coldest and most variable water temperatures during migration and thus may experience unique thermal selective pressures leading to adaptive genetic divergence.

4.4 | Markers associated with environmental characteristics

In addition to landscape variables associated with adaptive genetic variation that were identified by multivariate analyses, there were many independent SNPs with compelling associations and annotation results. Single outlier tests identified suites of significant candidate SNPs; however, we were most interested in adaptive SNPs that were identified by either the majority of outlier tests or gene–environment associations, as these represent consistent candidates for loci under selection. Among these candidate SNPs were three markers that have previously been associated with adult migration run timing in steelhead (Hess et al., 2016), thereby corroborating our methods for detecting true candidate loci under selection. Another candidate SNP (SNP 1751) was identified by all four outlier tests and matched the CXCL10-like chemokine gene, which is associated with immune function (Deng et al., 2008). BayENV also found this gene to be associated with isothermality, indicating a potential immune response to a pathogen whose prevalence may be dependent on isothermality. Aside from these genes, others that involve embryo development (e.g., Hox genes), muscle contraction and myocardium differentiation presumably lay a role in energetic (e.g., the cost of navigating barriers or currents during migration). However, a genomewide association study approach would be more appropriate in determining the direct relationship of these genes with specific phenotypes.

4.5 | Approach to test for divergent selection on migratory conditions

Many methods exist to examine and understand migration networks (Baker, 1978; Egevang et al., 2010; Hobson & Wassenaar, 2008; Shafer et al., 2016), yet few approaches have been developed to test for selection specifically related to migratory conditions. Our method provides a complement to the growing techniques involved with identifying the genetics of migratory patterns (Liedvogel, Akesson, & Bensch, 2011). Specifically, using genetic data in concert with geographic data and information about an organism's movement patterns, our approach can identify landscape variables along migration paths that may be posing selective pressures in any species which can be explicitly compared to variables at residing sites (e.g., overwintering sites, collection sites, natal sites). For instance, many species of North American birds utilize well-characterized flyways to migrate long distances to wintering grounds in Central and South America (Johnsgard, 1975; Martell, Henny, Nye, & Solensky, 2001). In addition to breeding and wintering sites, associating season-specific climate data along entire migratory flyways can elucidate how migratory conditions influence adaptation. Likewise, large mammals, such as pronghorn (*Antilocapra americana*), have been tracked for decades and have predictable and consistent migration paths (Poor, Loucks, Jakes, & Urban, 2012). While conservation effort is already being applied to preserve these narrow paths, little is known about selective pressures beyond human-mediated habitat alteration that may be acting on migratory animals. Finally, as more detailed ocean GIS data become available, such as ocean currents and water temperature, these same methods can be utilized along the well-documented migratory routes of marine mammals such as whales, turtles and seals.

5 | CONCLUSION

Many organisms have complex life histories that involve migration across heterogeneous landscapes. The majority of landscape genomic studies investigate associations between collection locality habitat and adaptive loci (Rellstab et al., 2015). Here, we provide a novel example of associations between migration habitat and adaptive loci in anadromous fish. Our results suggest that, when considering conservation of aquatic species, focusing solely on characteristics of spawning and juvenile rearing sites may not be sufficient in protecting an anadromous species like steelhead. While this method is readily applicable to freshwater aquatic systems that have obvious migration patterns through river networks, it can also be applied to a broader scale in avian, reptile and mammal systems that have obvious migration patterns or corridors in aquatic or terrestrial environments. As next-generation sequencing data and reference genomes become more readily available, landscape genomic methods such as this can be used to determine regions of the genome that may be under selection due to environmental conditions along migration paths.

ACKNOWLEDGEMENTS

We thank Megan Moore, Stephanie Harmon, Jeff Stephenson and Joey Tempest for laboratory support. Samples were contributed by tribal biologists from Nez Perce Tribe, Yakama Nation, Warm Springs and Umatilla, and agencies such as NOAA Fisheries, U.S Fish and Wildlife Service, Idaho Department of Fish and Game, Oregon Department of Fish and Wildlife and Washington Department of Fish and Wildlife.

AUTHOR CONTRIBUTION

S.J.M. performed bioinformatics and data analyses and wrote the manuscript. A.R.M. performed laboratory work and bioinformatics. A.P.M. assisted with study design and organized collection of the samples. S.N.R. designed and directed the project and contributed to analyses. All authors contributed to writing and editing the manuscript.

DATA ACCESSIBILITY

Supplementary data are available in Dryad under Accession ID: https://doi.org/10.5061/dryad.2tm29. Data available in this repository area as follows: (i) RAD sequences, individual locus statistics (Genotype success, HWE *p*-values, MAF, etc.) and other population

statistics; (ii) all landscape variables considered and their measurements at each collection site; (iii) Top BLAST hit results for outlier and GEA loci; (iv) individual read statistics; and (v) pairwise F_{ST} tables. Additionally, raw sequencing reads used in this study are available in the sequence read archive (SRA) with study Accession RP121665.

ORCID

Steven J. Micheletti http://orcid.org/0000-0002-3842-0946

REFERENCES

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a F ST-outlier method. *BMC Bioinformatics*, *9*, 323. https://doi.org/10.1186/1471-2105-9-323

Asuero, A. G., Sayago, A., & Gonzalez, A. G. (2006). The correlation coefficient: An overview. *Critical Reviews in Analytical Chemistry*, *36*, 41–59. https://doi.org/10.1080/10408340500526766

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, e3376. https://doi.org/10.1371/journal.pone.0003376

Baker, R. (1978). *Evolutionary ecology of animal migration*. Teaneck, NJ, USA: Holmes & Meier Publishers.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Beacham, T. D., Candy, J. R., Jonsen, K. L., Supernault, J., Wetklo, M., Deng, L., . . . Varnavskaya, N. (2006). Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim by use of microsatellite variation. *Transactions of the American Fisheries Society*, *135*, 861–888.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, *263*, 1619–1626. https://doi.org/10.1098/rspb.1996.0237

Blankenship, S. M., Campbell, M. R., Hess, J. E., Hess, M. A., Kassler, T. W., Kozfkay, C. C., & Stephenson, J. J. (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. https://doi.org/10.1080/00028487.2011.584487

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*, 1045–1055. https://doi.org/10.2307/1940179

Bourret, V., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape genomics in Atlantic salmon (*Salmo salar*): Searching for gene–environment interactions driving local adaptation. *Evolution*, *67*, 3469–3487.

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*, 99–232. https://doi.org/10.1080/10641260490280313

Brown, G. E., & Smith, R. J. F. (1997). Conspecific skin extracts elicit antipredator responses in juvenile rainbow trout (*Oncorhynchus mykiss*). *Canadian Journal of Zoology*, *75*, 1916–1922. https://doi.org/10.1139/z97-821

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*, 855–867. https://doi.org/10.1111/1755-0998.12357

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. https://doi.org/10.1111/mec.12354

Coble, D. W. (1961). Influence of water exchange and dissolved oxygen in redds on survival of steelhead trout embryos. *Transactions of the American Fisheries Society*, *90*, 469–474. https://doi.org/10.1577/1548-8659(1961)90[469:IOWEAD]2.0.CO;2

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423. https://doi.org/10.1534/genetics.110.114819

Cornelis, M. C., Agrawal, A., Cole, J. W., Hansel, N. N., Barnes, K. C., Beaty, T. H., & Feenstra, B. (2010). The Gene, Environment Association Studies consortium (GENEVA): Maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genetic Epidemiology*, *34*, 364–372. https://doi.org/10.1002/gepi.v34:4

Cushman, R. M. (1985). Review of ecological effects of rapidly varying flows downstream from hydroelectric facilities. *North American Journal of Fisheries Management*, *5*, 330–339. https://doi.org/10.1577/1548-8659(1985)5<330:ROEEOR>2.0.CO;2

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*, 1383–1399. https://doi.org/10.1111/mec.12182

Deng, G., Zhou, G., Zhang, R., Zhai, Y., Zhao, W., Yan, Z., & Zhang, X. (2008). Regulatory polymorphisms in the promoter of CXCL10 gene and disease progression in male hepatitis B virus carriers. *Gastroenterology*, *134*, 716–726. https://doi.org/10.1053/j.gastro.2007.12.044

Dingle, H., & Drake, V. A. (2007). What is migration? *BioScience*, *57*, 113–121. https://doi.org/10.1641/B570206

Duforet-Frebourg, N., Bazin, E., & Blum, M. G. (2014). Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution*, *31*, 2483–2495. https://doi.org/10.1093/molbev/msu182

Egevang, C., Stenhouse, I. J., Phillips, R. A., Petersen, A., Fox, J. W., & Silk, J. R. (2010). Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. *Proceedings of the National Academy of Sciences*, *107*, 2078–2081. https://doi.org/10.1073/pnas.0909493107

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

Etherington, T. R. (2011). Python based GIS tools for landscape genetics: Visualising genetic relatedness and measuring landscape connectivity. *Methods in Ecology and Evolution*, *2*, 52–55. https://doi.org/10.1111/j.2041-210X.2010.00048.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/mec.2005.14.issue-8

Foll, M., & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*, 977–993. https://doi.org/10.1534/genetics.108.092221

François, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*, 454–469. https://doi.org/10.1111/mec.2016.25.issue-2

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*, 1687–1699. https://doi.org/10.1093/molbev/mst063

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*, 205–220. https://doi.org/10.1534/genetics.113.152462

Hand, B. K., Muhlfeld, C. C., Wade, A. A., Kovach, R. P., Whited, D. C., Narum, S. R., … Luikart, G. L. (2016). Climate variables explain neutral and adaptive variation within salmonid metapopulations: The importance of replication in landscape genetics. *Molecular Ecology*, *25*, 689–705. https://doi.org/10.1111/mec.13517

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*, 197–214. https://doi.org/10.1046/j.1365-294X.2002.01434.x

Hecht, B. C., Thrower, F. P., Hale, M. C., Miller, M. R., & Nichols, K. M. (2012). Genetic architecture of migration-related traits in rainbow and steelhead trout, Oncorhynchus mykiss. *G3: Genes, Genomes, Genetics*, *2*, 1113–1127.

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*, 5573–5595. https://doi.org/10.1111/mec.13409

Hess, J. E., Zendt, 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*, *283*, 27170720.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/(ISSN)1097-0088

Hobson, K. A., & Wassenaar, L. I. (2008). *Tracking animal migration with stable isotopes*, Vol. *2*. Cambridge, MA, USA: Academic Press.

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., … Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, *81*, 345–354.

Ineno, T., Tsuchida, S., Kanda, M., & Watabe, S. (2005). Thermal tolerance of a rainbow trout *Oncorhynchus mykiss* strain selected by high-temperature breeding. *Fisheries Science*, *71*, 767–775. https://doi.org/10.1111/fis.2005.71.issue-4

Isaak, D. J., Wenger, S. J., Peterson, E. E., Ver Hoef, J. M., Hostetler, S., Luce, C. H., & Horan, D. (2011). NorWeST: an interagency stream temperature database and model for the Northwest United States. US Fish and Wildlife Service, Great Northern Landscape Conservation Cooperative Grant.

Jain, K. E., & Farrell, A. P. (2003). Influence of seasonal temperature on the repeat swimming performance of rainbow trout *Oncorhynchus mykiss*. *Journal of Experimental Biology*, *206*, 3569–3579. https://doi.org/10.1242/jeb.00588

Johnsgard, P. A. (1975). *Waterfowl of North America*. Bloomington: Indiana University Press.

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

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. https://doi.org/10.1186/1471-2156-11-94

Keefer, M. L., Jepson, M. A., Clabough, T. S., Johnson, E. L., Narum, S. R., Hess, J. E., & Caudill, C. C. (2017). Sea-to-sea survival of late-run adult steelhead (*Oncorhynchus mykiss*) from the Columbia and Snake rivers. *Canadian Journal of Fisheries and Aquatic Sciences*. https://doi.org/10.1139/cjfas-2016-0430

Keefer, M. L., Wertheimer, R. H., Evans, A. F., Boggs, C. T., & Peery, C. A. (2008). Iteroparity in Columbia River summer-run steelhead (*Oncorhynchus mykiss*): Implications for conservation. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*, 2592–2605. https://doi.org/10.1139/F08-160

Kendall, N. W., McMillan, J. R., Sloat, M. R., Buehrens, T. W., Quinn, T. P., Pess, G. R., & Zabel, R. W. (2014). Anadromy and residency in steelhead and rainbow trout (*Oncorhynchus mykiss*): A review of the processes and patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*, 319–342.

Kierepka, E. M., & Latch, E. K. (2015). Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology Resources*, *15*, 512–525. https://doi.org/10.1111/1755-0998.12332

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Liedvogel, M., Akesson, S., & Bensch, S. (2011). The genetics of migration on the move. *Trends in Ecology and Evolution*, *26*, 561–569. https://doi.org/10.1016/j.tree.2011.07.009

Lotterhos, K. E., Card, D. C., Schaal, S. M., Wang, L., Collins, C., & Verity, B. (2017). Composite measures of selection can improve the signal-to-noise ratio in genome scans. *Methods in Ecology and Evolution*, *8*, 717–727.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, *23*, 2178–2192. https://doi.org/10.1111/mec.12725

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*, 67–77. https://doi.org/10.1111/men.2017.17.issue-1

Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., & Fortin, M. J. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, *19*, 3760–3772. https://doi.org/10.1111/j.1365-294X.2010.04717.x

Martell, M. S., Henny, C. J., Nye, P. E., & Solensky, M. J. (2001). Fall migration routes, timing, and wintering sites of North American Ospreys as determined by satellite telemetry. *The Condor*, *103*, 715–724. https://doi.org/10.1650/0010-5422(2001)103[0715:FMRTAW]2.0.CO;2

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*, 682–701. https://doi.org/10.1111/eva.2014.7.issue-6

Matala, A. P., Hatch, D. R., Everett, S., Ackerman, M. W., Bowersox, B., Campbell, M., & Narum, S. (2016). What goes up does not come down: The stock composition and demographic characteristics of upstream migrating steelhead differ from post-spawn emigrating kelts. *ICES Journal of Marine Science: Journal du Conseil*, *73*, 2595–2605. https://doi.org/10.1093/icesjms/fsw109

McGaughran, A., Morgan, K., & Sommer, R. J. (2014). Environmental variables explain genetic structure in a beetle-associated nematode. *PLoS One*, *9*, e87317.

Meeuwig, M. H., Guy, C. S., Kalinowski, S. T., & Fredenberg, W. A. (2010). Landscape influences on genetic differentiation among bull trout populations in a stream-lake network. *Molecular Ecology*, *19*, 3620–3633. https://doi.org/10.1111/j.1365-294X.2010.04655.x

Micheletti, S. J., & Storfer, A. (2017). An approach for identifying cryptic barriers to gene flow that limit species' geographic ranges. *Molecular Ecology*, *26*, 490–504. https://doi.org/10.1111/mec.2017.26.issue-2

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, *17*, 240–248. https://doi.org/10.1101/gr.5681207

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.

Myers, K. W., Aydin, K. Y., Walker, R. V., Fowler, S., & Dahlberg, M. L. (1996). Known ocean ranges of stocks of Pacific salmon and steelhead as shown by tagging experiments. *North Pacific Anadromous Fish Commission Documents*, *192*, 1956–1995.

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*, 2841–2847. https://doi.org/10.1111/mec.12350

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, *11*, 184–194. https://doi.org/10.1111/j.1755-0998.2011.02987.x

Narum, S. R., Zendt, J. S., Graves, D., & Sharp, W. R. (2008). Influence of landscape on resident and anadromous life history types of *Oncorhynchus mykiss*. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*, 1013–1023. https://doi.org/10.1139/F08-025

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin PR, O. R., & Oksanen, M. J. (2013). Package 'vegan'. Community ecology package, version, 2(9).

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*, 223–241. https://doi.org/10.1007/s10592-010-0135-3

Penney, Z. L., & Moffitt, C. M. (2015). Fatty-acid profiles of white muscle and liver in stream-maturing steelhead trout *Oncorhynchus mykiss* from early migration to kelt emigration. *Journal of Fish Biology*, *86*, 105–120. https://doi.org/10.1111/jfb.2015.86.issue-1

Perry, G. M., Danzmann, R. G., Ferguson, M. M., & Gibson, J. P. (2001). Quantitative trait loci for upper thermal tolerance inoutbred strains of rainbow trout (*Oncorhynchus mykiss*). *Heredity*, *86*, 333–341. https://doi.org/10.1046/j.1365-2540.2001.00838.x

Petersen, J. H., & Kitchell, J. F. (2001). Climate regimes and water temperature changes in the Columbia River: Bioenergetic implications for predators of juvenile salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*, 1831–1841. https://doi.org/10.1139/f01-111

Poor, E. E., Loucks, C., Jakes, A., & Urban, D. L. (2012). Comparing habitat suitability and connectivity modeling methods for conserving pronghorn migrations. *PLoS ONE*, *7*, e49390. https://doi.org/10.1371/journal.pone.0049390

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*, 904–909. https://doi.org/10.1038/ng1847

Quinn, T. P. (2011). *The behavior and ecology of Pacific salmon and trout*. Vancouver, BC, Canada: UBC Press.

R Core Team (2015). R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2014.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589. https://doi.org/10.1534/genetics.114.164350

Raymond, H. L. (1979). Effects of dams and impoundments on migrations of juvenile chinook salmon and steelhead from the Snake River, 1966 to 1975. *Transactions of the American Fisheries Society*, *108*, 505–529. https://doi.org/10.1577/1548-8659(1979)108<505:EODAIO>2.0.CO;2

Raymond, H. L. (1988). Effects of hydroelectric development and fisheries enhancement on spring and summer chinook salmon and steelhead in the Columbia River basin. *North American Journal of Fisheries Management*, *8*, 1–24. https://doi.org/10.1577/1548-8675(1988)008<0001:EOHDAF>2.3.CO;2

Reisenbichler, R. R., McIntyre, J. D., Solazzi, M. F., & Landino, S. W. (1992). Genetic variation in steelhead of Oregon and northern California. *Transactions of the American Fisheries Society*, *121*, 158–169. https://doi.org/10.1577/1548-8659(1992)121<0158:GVISOO>2.3.CO;2

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*, 4348–4370. https://doi.org/10.1111/mec.13322

Richter, A., & Kolmes, S. A. (2005). Maximum temperature limits for Chinook, coho, and chum salmon, and steelhead trout in the Pacific Northwest. *Reviews in Fisheries Science*, *13*, 23–49. https://doi.org/10.1080/10641260590885861

Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Shafer, A. B. A., Northrup, J. M., Wikelski, M., Wittemyer, G., & Wolf, J. B. W. (2016). Forecasting ecological genomics: High-tech animal instrumentation meets high-throughput sequencing. *PLoS Biology*, *14*, e1002350. https://doi.org/10.1371/journal.pbio.1002350

Sork, V. L., Aitken, S. N., Dyer, R. J., Eckert, A. J., Legendre, P., & Neale, D. B. (2013). Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics and Genomes*, *9*, 901–911. https://doi.org/10.1007/s11295-013-0596-x

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., & Waits, L. P. (2007). Putting the 'landscape' in landscape genetics. *Heredity*, *98*, 128–142. https://doi.org/10.1038/sj.hdy.6800917

Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., Colli, L., Masembe, C., ... Taberlet, P. (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*, *17*, 1072–1089.

Taylor, E. B. (1991). A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, *98*, 185–207. https://doi.org/10.1016/0044-8486(91)90383-I

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M. C., Schwager, M., & Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: The importance of keystone structures. *Journal of Biogeography*, *31*, 79–92. https://doi.org/10.1046/j.0305-0270.2003.00994.x

Thomas, D. (2010). Gene–environment-wide association studies: Emerging approaches. *Nature Reviews Genetics*, *11*, 259–272. https://doi.org/10.1038/nrg2764

Torkamaneh, D., Laroche, J., & Belzile, F. (2016). Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLoS ONE*, *11*, e0161333. https://doi.org/10.1371/journal.pone.0161333

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

Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *BioRxiv* 005165.

U.S. Geological Survey (2013). *National Hydrography Geodatabase: The National Map viewer available on the World Wide*. Retrieved from Web https://viewer.nationalmap.gov/viewer/nhd.html?p=nhd.

Utter, F. M., & Allendorf, F. W. (1977). Determination of the breeding structure of steelhead populations through gene frequency analysis.

Proceedings of the Genetic Implications of Steelhead Management Symposium, *77*, 44.

Valdimarsson, S. K., & Metcalfe, N. B. (1998). Shelter selection in juvenile Atlantic salmon, or why do salmon seek shelter in winter? *Journal of Fish Biology*, *52*, 42–49. https://doi.org/10.1111/jfb.1998.52.issue-1

Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*, 1248–1258. https://doi.org/10.1111/2041-210X.12418

Wang, H. (2009). Forward regression for ultra-high dimensional variable screening. *Journal of the American Statistical Association*, *104*, 1512–1524. https://doi.org/10.1198/jasa.2008.tm08516

Westley, P. A., Quinn, T. P., & Dittman, A. H. (2013). Rates of straying by hatchery-produced Pacific salmon (*Oncorhynchus* spp.) and steelhead (*Oncorhynchus mykiss*) differ among species, life history types, and populations. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*, 735–746. https://doi.org/10.1139/cjfas-2012-0536

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F ST. *The American Naturalist*, *186*, S24–S36. https://doi.org/10.1086/682949

Willis, S. C., Hollenbeck, C. M., Puritz, J. B., Gold, J. R., & Portnoy, D. S. (2017). Haplotyping RAD loci: An efficient method to filter paralogs and account for physical linkage. *Molecular Ecology Resources*, *17*, 955–965. https://doi.org/10.1111/men.2017.17.issue-5

Zobel, D. B., McKee, A., Hawk, G. M., & Dyrness, C. T. (1976). Relationships of environment to composition, structure, and diversity of forest communities of the central western Cascades of Oregon. *Ecological Monographs*, *46*, 135–156. https://doi.org/10.2307/1942248

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Micheletti SJ, Matala AR, Matala AP, Narum SR. Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). *Mol Ecol*. 2018;27:128–145. https://doi.org/10.1111/mec.14407