Received: 26 July 2021 | Revised: 19 November 2021 | Accepted: 23 December 2021

DOI: 10.1111/mec.16339

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Long-distance migration is a major factor driving local adaptation at continental scale in Coho salmon

Quentin Rougemont[1,2] | Amanda Xuereb[1] | Xavier Dallaire[1] | Jean-Sébastien Moore[1] | Eric Normandeau[1] | Alysse Perreault-Payette[1] | Bérénice Bougas[1] | Eric B. Rondeau[3,4] | Ruth E. Withler[3] | Donald M. Van Doornik[5] | Penelope A. Crane[6] | Kerry A. Naish[7] | John Carlos Garza[8] | Terry D. Beacham[3] | Ben F. Koop[4] | Louis Bernatchez[1]

[1]Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada

[2]CEFE, Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, CNRS, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier, Montpellier, France

[3]Department of Fisheries and Ocean, Pacific Biological Station, Nanaimo, British Columbia, Canada

[4]Department of Biology, University of Victoria, Victoria, British Columbia, Canada

[5]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center, Manchester Research Station, Port Orchard, Washington, USA

[6]Conservation Genetics Laboratory, U.S. Fish and Wildlife Service, Anchorage, Alaska, USA

[7]School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington, USA

[8]Department of Ocean Sciences and Institute of Marine Sciences, University of California Santa Cruz, Santa Cruz, California, USA

Correspondence
Quentin Rougemont, CEFE, Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175, CNRS, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier, Montpellier, France.
Email: quentinrougemont@orange.fr

Funding information
This research was carried out in conjunction with EPIC4 (Enhanced Production in Coho: Culture, Community, Catch), a project supported in part by University Laval and the government of Canada through Genome Canada, Genome British Columbia, and Genome Quebec. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

[Corrections added on 26 February, after first online publication: Author names and ORCiD information have been added for 'Bérénice Bougas' and 'Alysse Perreault-Payette'.]

Abstract

Inferring the genomic basis of local adaptation is a long-standing goal of evolutionary biology. Beyond its fundamental evolutionary implications, such knowledge can guide conservation decisions for populations of conservation and management concern. Here, we investigated the genomic basis of local adaptation in the Coho salmon (*Oncorhynchus kisutch*) across its entire North American range. We hypothesized that extensive spatial variation in environmental conditions and the species' homing behaviour may promote the establishment of local adaptation. We genotyped 7829 individuals representing 217 sampling locations at more than 100,000 high-quality RADseq loci to investigate how recombination might affect the detection of loci putatively under selection and took advantage of the precise description of the demographic history of the species from our previous work to draw accurate population genomic inferences about local adaptation. The results indicated that genetic differentiation scans and genetic–environment association analyses were both significantly affected by variation in recombination rate as low recombination regions displayed an increased number of outliers. By taking these confounding factors into consideration, we revealed that migration distance was the primary selective factor driving local adaptation and partial parallel divergence among distant populations. Moreover, we identified several candidate single nucleotide polymorphisms associated with long-distance migration and altitude including a gene known to be involved in adaptation to altitude in other species. The evolutionary implications of our findings are discussed along with conservation applications.

KEYWORDS
conservation genetics, landscape genomics, local adaptation, polygenic selection, recombination

1 | INTRODUCTION

Both plant and animal biodiversity are currently declining at unprecedented rates due to human activities (Allendorf, 2017), reducing species' capacity to retain genetic diversity, and their evolutionary potential. In this context, accurate understanding of a species' demographic and genetic responses to environmental change is of major importance for their management and conservation. In particular, spatially heterogeneous environments can impose different selective pressures that could lead to local adaptation of populations (Kawecki & Ebert, 2004). In aquatic ecosystems, there is ample geographical variation in environmental conditions among habitats occupied by different populations of a given species. These conditions include salinity, temperature gradients and geological features (Grummer et al., 2019), all of which can favour spatially varying selection and contribute to local adaptation (Gagnaire et al., 2012; Micheletti et al., 2018; Narum et al., 2018). Landscape genomics provides a valuable framework to study genome-wide adaptive variation and its interaction with ecological conditions (Grummer et al., 2019). Despite the growing evidence from landscape genomics for local adaptation associated with variation in ecological parameters in marine and aquatic habitats (e.g., Cayuela et al., 2020; Gagnaire et al., 2012; Hecht et al., 2012; Micheletti et al., 2018), the field remains dominated by research in terrestrial systems (Grummer et al., 2019).

In recent years, landscape genomic methods (Narum & Hess, 2011) and genotype–environment association analyses (GEAs) (e.g., Bernatchez, 2016; Forester et al., 2018; Villemereuil et al., 2014) have been developed to identify signals of adaptation. Many riverscape genomic studies have investigated the effect of temperature, precipitation regime, geology, various distance variables (elevation, migration distance) or barriers to migration (e.g., dams) on local adaptation (e.g., Moore et al., 2017; Micheletti et al., 2018, Dallaire et al., 2021). A major challenge in thoroughly interpreting genome scans and GEAs is to accurately take into account the effect of demographic history. In particular, vicariance events and spatio-temporal variation in population size and migration rate impact the distribution of genetic variation, primarily by distorting the site frequency spectrum (Galtier et al., 2000). These effects are difficult to model in most genome scan methods (Lotterhos & Whitlock, 2015; Narum & Hess, 2011). This is particularly true for populations occupying aquatic habitats and distributed along fragmented landscapes, with variable rates of connectivity and variable effective population sizes (Bierne et al., 2013; Fourcade et al., 2013). In particular, the signal of isolation by distance (IBD), along with a unidirectional expansion from a common source population, has been shown to confound genome scan results, leading to increased rates of false positives (Battey et al., 2020; Meirmans, 2012). To complicate these inferences further, the biologically meaningful signal of selection may not readily be discernible from demographic stochasticity if adaptation is driven by small shifts in allele frequencies at multiple loci (i.e., polygenic selection (Pritchard et al., 2010), as opposed to adaptation driven by hard selective sweeps that leave more discernible patterns along the genome.

A second, often overlooked, challenge is the effect of variation in recombination rate on genome scans and GEAs (Lotterhos, 2019). Nucleotide diversity and genetic differentiation are known to vary with local variation in recombination rate along the genome, with higher differentiation observed in regions of lower recombination (Charlesworth et al., 1993). A consequence for genome scans is an associated high false positive rate in regions of low recombination (Perrier & Charmantier, 2019) and, conversely, a high false negative rate in regions of high recombination (Booker et al., 2020).

Here, we aim to elucidate the effects of these phenomena on inferences of local adaptation with the largest landscape genomic study in a nonmodel species to date by explicitly accounting for the confounding factors of historical demography and recombination rate. We focused on the Coho salmon (*Oncorynchus kisutch*), a species that has suffered from significant demographic declines over decades (Gustafson et al., 2007; Oke et al., 2020) and which is of high conservation concern because of its importance for the recreational and indigenous subsistence fisheries it supports (Oke et al., 2020). Based on a small subset of samples used in the present study, we recently documented the demographic history of Coho salmon throughout its North American range (Rougemont et al., 2020). This study showed that most of the contemporary species' genetic diversity originates from a single major refugium that was located in the southern part of the species range, and previously glaciated regions further north were colonized following postglacial demographic expansion from this refugium. Coho salmon are distributed from 37 to 67°N across a range of environments, from warm to cold (e.g., 16.3°C in California vs. −2.32°C in Alaska, Worldclim database), with varying precipitation regimes (e.g., over 2438 mm annually in the Haida Gwaii archipelago vs. 848 mm in parts of Alaska), and across different geological rock types (Garrity & Soller, 2009). These environmental variables were shown to be involved in the local adaptation of a congeneric species, the steelhead trout (*Oncorhynchus mykiss*), in part of its range (Micheletti et al., 2018). Moreover, the species exhibits a range of spawning migration distances, from a few kilometres in short coastal rivers to over 2000 km into the Yukon River of Alaska (Sandercock, 1991). The combined effects of migratory distance and total elevation gain to the spawning site could constitute an important selective factor (Moore et al., 2017). Indeed, long-distance migration has an elevated bioenergetic cost and can drive local adaptation of morphological, physiological or behavioural traits for improved migration efficiency (Bernatchez &

Dodson, 1987; Crossin et al., 2004; Eliason et al., 2011). Finally, the natal homing behaviour of salmonids and relatively low dispersal between adjacent rivers (Quinn, 1984) should contribute to the development and maintenance of local adaptive differentiation between populations (Waples et al., 2020). Therefore, the spatial variation in environmental conditions and the species' homing behaviour probably promote the establishment of local adaptation in Coho salmon populations.

Here we assessed the respective roles of various environmental factors in shaping putative adaptive variation among Coho salmon populations throughout its North American range. Knowledge of the species' demographic history enabled us to interpret the distribution of outlier regions in the genome more accurately (Rougemont et al., 2020). Given recent population divergence from a single source post-glaciation, we hypothesized that if local adaptation is due to a few single nucleotide polymorphisms (SNPs) of large effect (e.g., a recent hard sweep), these signals should be widespread among all populations. In this case, we expect (i) large changes in allele frequencies at candidate SNPs, and (ii) similar changes at shared candidate SNPs or genomic regions across different geographical parts of the range where populations undergo similar challenges (e.g., similar challenges to migration to spawning grounds, or thermal stress), resulting in parallel changes in allele frequencies. Alternatively, local adaptation may proceed through polygenic selection (Yeaman, 2015). In this case we expect small, but covarying allele frequency shifts at candidate SNPs as well as more limited parallelism as each genetic group will adapt to their local environment through different path (Barghi et al., 2020). We take advantage of a large genomic data set to: (i) document patterns of adaptive genomic variation and evaluate the occurrence of parallel allele frequency shifts at loci of strong effect across the species range, (ii) identify the most important environmental factors involved in shaping patterns of adaptive genomic variation and (iii) investigate the potential confounding effect of recombination rate variation on the detection of putatively adaptive SNPs and explicitly account for this factor within differentiation-based analyses. Finally, we show the power of a recently developed machine learning method to easily investigate levels of population structure over such a large data set.

2 | METHODS

2.1 | Sampling, environmental data and bioinformatics

A total of 7829 Coho salmon individuals were collected from 217 freshwater locations along the Pacific coast of North America, from California to Alaska, and from one site in Russia (Figure 1; Table S1, Figure S1). The median number of individuals per site was 38 (range 14–70). Sampled individuals were distributed across a heterogeneous environment (Table S1), which we predict would result in a range of selective pressures across populations. We previously used data from approximately one quarter of the current samples to show that contemporary Coho salmon populations mostly radiated postglacially from a single major lineage. This system provides ideal conditions to understand how recent divergence favours or constrains parallel adaptation across space (Rougemont et al., 2020).

Given the variability in temperature and precipitation across the study area, we delimited the catchment area upstream of the sampling sites and calculated the mean, minimum, maximum, range and and standard deviation of 19 climatic variables over these areas, resulting in 95 metrics associated with temperature (°C) and precipitation (mm) described in the Supplementary Methods and provided in Table S1 and Figure S2. These were extracted from the WordclimV2.0 database (Fick & Hijmans, 2017) for the period 1970–2000. We reduced these data to a set of uncorrelated variables using two separate principal components analyses (PCAs) and retained the significant axes of variation. The PCAs were performed using the R package ADE4 (Dray & Dufour, 2007). We also extracted geological variables (rock type, geological era) from the United States Geological Survey database (Garrity & Soller, 2009), since geology has been identified as influencing population genetic structure in other salmonids (e.g., Bourret et al., 2013; Quéméré et al., 2016). Moreover, this parameter has been central to the designation of Evolutionary Significant Units in Coho salmon (Weitkamp et al., 1995).

Different Coho salmon populations undertake adult freshwater migrations from the marine environments to breeding sites that span a wide range of distances (river distances from 10 to 2300 km; Table S1), which we predicted should result in differential selective environments across populations (Olsen et al., 2011). In addition to distance to the breeding site, elevation is expected to exert a strong selective pressure on migration and homing phenotypes (e.g., Bernatchez & Dodson, 1987). Such selection may result in differences in allele frequencies at loci linked to migration phenotypes associated with low- or high-elevation sites. Therefore, we computed the product of river length and altitude gain and standardized to a mean of zero and with a standard deviation of one (Moore et al., 2017). This measure will be referred to hereafter as "normalized distance." We evaluated the collinearity among predictor variables prior to our analysis using the variance inflation factor (VIF) and a correlation plot. No predictor displayed a VIF >10, so all were retained for GEAs.

DNA extractions and library preparation for ddRADseq (double digest restriction-site associated DNA sequencing) were performed following protocols described elsewhere (Rougemont et al., 2020). Library preparation, polymerase chain reaction (PCR) amplification and sequencing on the Ion Proton P1v2 chip were performed at the Genomic analysis platform of IBIS at Université Laval (http://www.ibis.ulaval.ca/). The bioinformatic procedure from Rougemont et al. (2020) was used, using STACKS version 2 (Rochette et al., 2019) and with the latest Coho salmon reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_002021735.2/). The resulting vcf file was filtered using VCFTOOLS (Danecek et al., 2011). SNPs were retained if they displayed a read depth >10 and <120. We then removed SNPs that were not present in

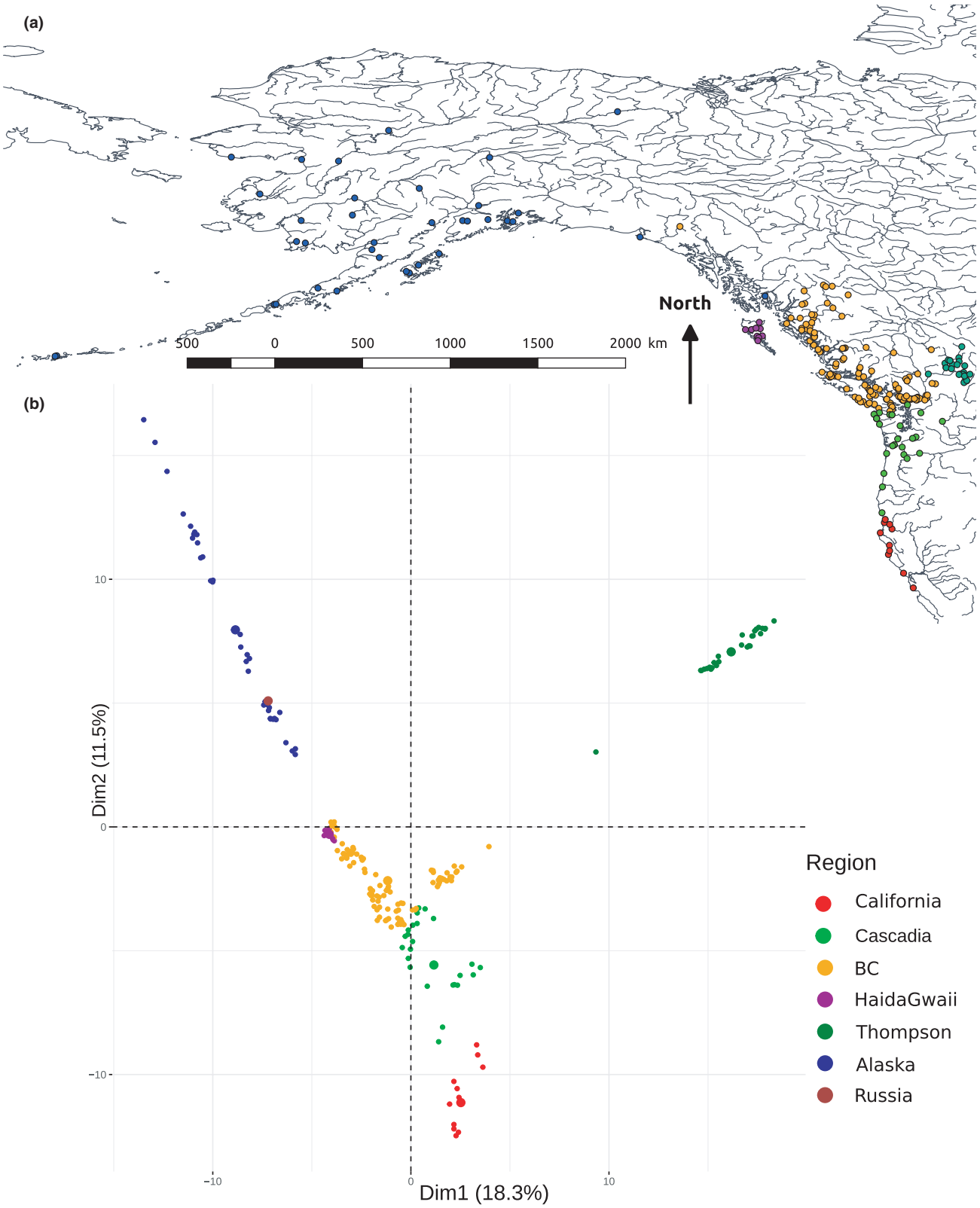


FIGURE 1 PCA genetic structure mirrors geography. (a) Map of the study area; dots represent samples and colours represent seven major regional groups. (b) PCA on allele frequency displaying the genetic structure among 217 Coho salmon populations grouped into the seven different regions for illustrative purposes [Colour figure can be viewed at wileyonlinelibrary.com]

at least 95% of individuals in the data set. For each RAD locus, the single SNP with the highest minor allele frequency was retained. More details about the laboratory protocol and bioinformatic procedure along with a table of the different steps can be found in the Supplementary Methods and Table S2. This filtered data set contained 105,362 SNPs.

2.2 | Population genetics and demography

We previously observed a clinal distribution of decreasing genetic diversity from the south to the north of the Pacific Coast. Using demographic modelling we resolved the species historical demography and showed that it has expanded northward from a major southern refugium, resulting in an increase of the deleterious load due to surfing of deleterious mutations (Rougemont et al., 2020). We seek to verify whether our previous observations that (i) genetic diversity decreased northwards due to population expansion from a major southern refugium and (ii) genetic differentiation increased toward the north could be replicated by increasing the size from our previously published data set from 58 populations to the full set of 217 sampling localities presented here.

As previously, we estimated gene diversity (H_s) and population differentiation using the β_{ST} index (Weir & Goudet, 2017) with the HIERFSTAT R package (Goudet, 1995), and measured the relationship between H_s, β_{ST} and the distance to the southernmost site using linear models. Hierfstat was also used to compute each population's F_{IS} and compute 95% confidence intervals around the parameter using 1000 bootstraps over loci. Oceanographic distances between river mouths were measured with the MARMAP1.0.4 R package (Pante & Simon-Bouhet, 2013) to obtain more accurate distances compared to our previous analyses, where we used Euclidean distances between sample site coordinates.

We investigated population structure by performing a PCA on allele frequencies across loci with the program ADE4 (Dray & Dufour, 2007). A multidimensional scaling plot based on estimates of pairwise relationships (identity-by-state) was constructed using PLINK1.9 (Purcell et al., 2007) and plotted with the GGPLOT2 R package (Wickham, 2016). To better reveal fine-scale population structure that may be spread across many PCA axes, we used variational autoencoders (VAEs) designed for population genetic inference (Battey et al., 2021). VAEs are unsupervised machine learning models that use a neural network (here, feed forward networks are used for the encoder and decoder parts) to regenerate the data and allow visualizing genetic data in a latent space that is a lower dimensional space than PCA while preserving geometry, unlike other recent methods designed for large data sets (Battey et al., 2021).

2.3 | Candidate loci: GEAs

We evaluated the extent to which genetic variation was explained by environmental variables using two approaches. First, we used a redundancy analysis (RDA) implemented in the VEGAN2.5.6 R package (Oksanen et al., 2019). The Russian Reka Saichik River population was excluded from this analysis due to a lack of climatic data and pronounced geographical discontinuity from North American populations. The Mad River population was also excluded due to uncertainty about its original provenance as well as samples from Bonneville Hatchery in Oregon. We applied a minor allele count (MAC) of 15 (discussed in the Supplementary Materials), resulting in 59,115 SNPs with <5% missing data across 7719 individuals. This MAC threshold was used given that outlier detection tests are not designed to accommodate rare variants. The following model was tested: allele frequency ~Normalized_Distance + Temperature + Precipitation + Elevation + Rock + Condition(Latitude). Here, Latitude was used as a conditioning variable in a partial RDA to control for the confounding effect of IBD (Battey et al., 2020; Meirmans, 2012). Latitude was included as a way to partially control for population structure without "over-correcting" for this effect given that the pattern in our data arises due to both population structure and continuous latitudinal IBD following the southern expansions (Rougemont et al., 2020). Making a full correction for population structure may increase Type II error, as demonstrated by Lotterhos & Whitlock (2015) and Forester et al. (2018), with simulations showing that correcting for population structure may result in the exclusion of some true signals of local adaptation and decrease the rate of true positives.

Model significance was tested using an ANOVA with 1000 permutations of the genetic data. A cut-off of three standard deviations from the mean loading was used to identify outlier SNPs on the significant RDA axes. Second, we applied the GEA method based on latent factors mixed models (LFMMs) representing residual population structure. This method is implemented in the R package LFMM1.5 (Caye et al., 2019). We included Latitude as an additional explanatory variable in our model and excluded all outliers associated with this variable. A correction for multiple testing was performed following Benjamini & Hochberg (1995), and a given SNP was considered an outlier if its corrected *p*-value was below .01. Finally, we obtained functional annotations of outlier SNPs shared from the RDA and LFMM analyses using SNPEFF version 3.4 (Cingolani et al., 2012).

2.4 | Allele frequency distribution and parallelism

Given the importance of normalized distance as a candidate selective factor in our GEA analyses (see Section 3), we aimed to dissect the signal explained by adaptation to normalized migration distance. We thus compared populations performing "long-" vs. "short-" distance migrations. Long-distance migration was defined by the distribution of normalized distances observed in our data (i.e., distance >422 km or elevation above 400 m, corresponding to the 15% upper quantile), whereas distance or elevation below this quantile corresponded to short migration. We then performed a PCA on outliers that were shared between RDA and LFMM to test the discriminatory power of these candidate SNPs between long vs. short normalized distances. We worked only on shared outliers as a way to further decrease the rate of false positive detection, without overly stringent corrections. If adaptation to migration distance shows parallel patterns at the genetic level, populations of long vs. short migration distances should form distinct clusters, regardless of geography, and the same alleles at outlier loci should display parallel allele frequency shifts. Conversely, if geographical proximity plays a stronger role and/or if populations are adapting

through small allele frequency changes at different loci (nonparallel adaptation), then they should cluster by geographical groups. To further quantify whether adaptation to long-distance migration was parallel or nonparallel among geographical groups, we tested the following linear model: PC1 ~ Normalized_Distance + Latitude + Normalized_Distance * Latitude. Here PC1 corresponds to the first (and only significant) axis of the previous PCA based on outliers and separating our populations. Under this model, if the Latitude term is significant and explains a large proportion of the variance, it would indicate nonparallel adaptation for migration distance among populations from different geographical areas. In contrast, if only the Normalized Distance term is significant, then this would suggest that populations are adapting through parallel changes at the molecular level, and independently of their geographical proximity. The interaction term represents both parallel and independent adaptation at the molecular level. We performed this analysis using all outliers associated with Normalized Distance detected by both RDA and LFMM and compared the results to a random subset of 400 SNPs, close to the number of SNPs identified by each method. We used the function sample in R after excluding all putative outliers to generate the random subset.

2.5 | Candidate loci: Differentiation-based analysis

To further identify candidate loci involved in parallel (or nonparallel) adaptation to migration distance but displaying large allele frequency changes, we performed a differentiation-based genome scan taking advantage of multiple population pairs with long vs. short normalized distances. Unlike GEA, which can detect subtle allele frequency changes and identify multiple covarying SNPs, differentiation-based genome scans have been developed to identify outlier SNPs that display large allele frequency changes (Lotterhos & Whitlock, 2015). Instead of relying on traditional genome scan methods that are affected by complex demography, we looked at patterns of shared outliers among multiple replicate pairs of populations sharing similar ancestry. To do so, we used pairs of populations from the same watershed with one population exhibiting long-distance migration and one population exhibiting short-distance migration. However, most upstream samples were not independent from each other, often flowing into the same downstream river, so that keeping all pairs would lead to pseudoreplication. This implies a strong reduction in the number of "approximately" independent pairwise comparisons, reducing our sample set to four population pairs (detailed in Table S3). Two pairs were from the Alaska genetic group, one pair from the British Columbia (BC) group and one pair from the Thompson River. Analyses were performed using the Population Branch Statistic values (PBS) (Yi et al., 2010) in ANGSD (Korneliussen et al., 2014). We identified outliers as those above the 0.90th quantile of the PBS distribution (Rougemont et al., 2017). The number of outliers shared among population pairs was then computed to evaluate the extent of parallel adaptation to long-distance migration. If outliers are shared among several pairs of populations across the range, they are unlikely to be confounded by drift, providing stronger evidence for parallel adaptation.

2.6 | Accounting for recombination rate variation

Variation in recombination rate along the genome is known to influence the detection of outliers (Perrier & Charmantier, 2019). To test if the density of RDA, LFMM and differentiation-based outliers increased in areas of low recombination, we estimated population-scaled estimates of recombination rate using LDhat ($\rho = 4^{*}N_e{}^{*}r$ where N_e represents the effective population size and r the recombination rate in cM bp^{-1}, Supplementary Materials 1). These estimates were obtained from 71 individuals sampled from 14 populations distributed from California to Alaska that were sequenced at the whole genome level (Q. Rougemont, EB. Rondeau, B. Koop, L. Bernatchez, unpublished). We then tested for preferential enrichment of "hotspots" or "coldspots" of recombination in outliers (Supplementary Materials 1). We first tested if recombination hotspots (coldspots) contained more (fewer) outliers than "normally" recombining regions using χ^2 tests. Linear mixed-effects models were then used to test the relationship between recombination rate and the distribution of outliers. The response variable was the SNP state, which was considered binomial (0 = nonoutlier, 1 = outlier) and the explanatory variable was the recombination rate. Chromosome identity was included as a random effect. Models were performed using the LME4 R package (Bates et al., 2015). Our results revealed a significant association between detected outliers and recombination. Similarly, differentiation is known to be negatively correlated with recombination, so that many outliers in low recombination regions could be false positives (Booker et al., 2020; Perrier & Charmantier, 2019). The rate of recombination is proportional to the extent of LD across the chromosome (Ohta & Kimura, 1969): when recombination is low, high LD is expected whereas in regions of high recombination, low LD is expected. Therefore, to control for the confounding effect of variation in recombination on differentiation-based outlier detection, we used an approach previously developed to account for large-scale variation in LD (Perrier et al., 2020). To do so, our previous PBS values computed across 1 Mb were used to estimate large-scale differentiation driven by large-scale recombination variation. We subtracted these 1-Mb window values from the PBS values of outliers identified in 50-kb windows to obtain a ΔPBS and corrected for variation in recombination rate. A low ΔPBS is indicative of a low difference between the background (associated with recombination) and the local window, possibly indicating a false positive local window. A large ΔPBS is indicative of a large difference between the background and the local window, so that variation in recombination should not be responsible for driving the pattern of differentiation. We computed the number of outlying windows that remained considering different ΔPBS thresholds. Finally, we used the approach in Rougemont et al. (2020) to test for correlations between PBS values and recombination. We reduced the multiple PBS values using the first

axis of a PCA and tested for a correlation between the first PCA–PBS axis and recombination using mixed linear models including chromosome identity as a random effect.

3 | RESULTS

3.1 | Genetic structure mirrors geography within North American Coho salmon

Global patterns of population genetic structure corroborated our previous results (Rougemont et al., 2020). A PCA performed on 105,362 SNPs among 7829 individuals across the entire North American Coho salmon range revealed that samples that are geographically closer are genetically more similar than more distant samples (Figure 1a,b; Figure S3). The first PC axis was correlated with longitude ($r = .30$, $p < .0001$), whereas the second axis was correlated with latitude ($r = .81$, $p < .0001$), reflecting the decay of genetic similarity as a function of geographical distance ($r = .35$, $p < .0001$, Figure S4). While IBD dominates at broad spatial scales, subtle population structure was also observed, probably due to the species' homing tendency. This fine-scale structure was revealed using the VAE, which separated nearly every river within any given regional group (Figure 2a). As one example, on Haida Gwaii (Figure 2c), all rivers clustered separately except for individuals from four rivers that formed a mixed cluster, consistent with signals of admixture (Figure 2b; Figure S5). These results are probably a product of river-scale homing typical in salmonids (Quinn, 1984).

3.2 | Southern populations are ancestral and the most genetically diverse

Global patterns of variation in intrapopulation genetic diversity also corroborated our previous results (Rougemont et al., 2020). We observed a linear decrease in genetic diversity (H_s) as a function of the distance from the southernmost site ($R^2 = .39$, $p < .0001$, Figure S6a). The Thompson and interior Alaskan populations deviated from the patterns observed for coastal populations (Figure S6a), as expected, due to founding events during upstream postglacial recolonization. We therefore expected that this result would constitute a confounding factor in subsequent analyses (Fourcade et al., 2013).

Negative β_{ST} coefficients revealed that ancestral populations and those with high genetic diversity (Figure S6b) were located in Cascadia and southern BC at the approximate southern boundary of the ice sheet during the last glaciation. A linear decrease in β_{ST} with distance from the south was observed ($R^2 = .39$, $p < .0001$, Figure S06B). The Russian population did not display reduced genetic diversity or increased β_{ST}, as would be expected under a linear expansion from the southeastern Pacific coast. Removing the Russian population increased the R^2 to .42 in our linear models.

3.3 | Genotype environment associations

The PCA on temperature-associated variables revealed four significant axes of variation that were kept for the RDA and LFMM analysis below (Figure S2a,b). Similarly, the PCA on precipitation-associated variables (Figure S2c,d) revealed three significant axes, enabling a decorrelation of the variables among them. The first four axes of the temperature PCA and first three axes of the precipitation PCA respectively captured 87% and 89% of the total variability. We verified that these axes as well as latitude, geology and normalized distance did not display strong correlation using the VIF and a correlation plot (Figure S3). Latitude displayed the highest VIF values (10.35), indicating significant covariation with all other variables considered in our analyses (Figure 3a,b; Figure S7). All other variables displayed a VIF below 10 (Table S4) and were retained, resulting in a total of eight variables.

The RDA permutation and ordination tests revealed that all tested environmental and geographical variables were significant (Table S4). The full model was significant and the environmental variables captured 11.18% of the total genotypic variance ($df = 2$, $p < .0001$). The first seven RDA axes were significant, cumulatively capturing over 94% of the total variance. Outlier SNPs were identified on all significant axes (Table S04) with a total of 1960 outlier SNPs being detected. Fifty-six per cent ($n = 1105$) of those outlier SNPs were associated with normalized distance to the spawning site and mostly associated with the Thompson River (Figure 3a; Figure S8). A further 521 and 292 SNPs were associated with temperature and precipitation, respectively, while 42 SNPs were associated with geology (Figure 3a). The distribution of outlier SNPs across the genome was strongly correlated with chromosome length (adjusted $R^2 = .91$, $p < .0001$, Figure 3d).

Since strong postglacial founder effects may cause high rates of false positives, we also used a mixed latent-factor analysis in LFMM to identify shared outliers between different methods (Capblancq et al., 2018). As for RDA, a high number of outliers were associated with normalized migration distance ($n = 1465$ SNPs or 65% of all detected outliers), and 352, 445, and 0 outliers were associated with precipitation, temperature and geology, respectively. Accordingly, 29% of outliers associated with normalized distance displayed strong selection signals (i.e., −log10(p-values) >5, Figure 3c), whereas other variables displayed a weaker signal (Figure S9). As with the RDA outliers, the outlier SNPs identified by LFMM were distributed throughout the genome with no strong peak. A total of 273 outliers were shared between the RDA and LFMM, among which 181 were associated with normalized distance, representing ~62% of the total number of shared outliers. Of the remaining shared outliers, 12% ($n = 36$) were associated exclusively with precipitation, ~9% ($n = 23$) with temperature, and 18% ($n = 31$) were associated with different environmental variables (often Distance and another variable) by either RDA or LFMM. Only one shared outlier was associated with geology.

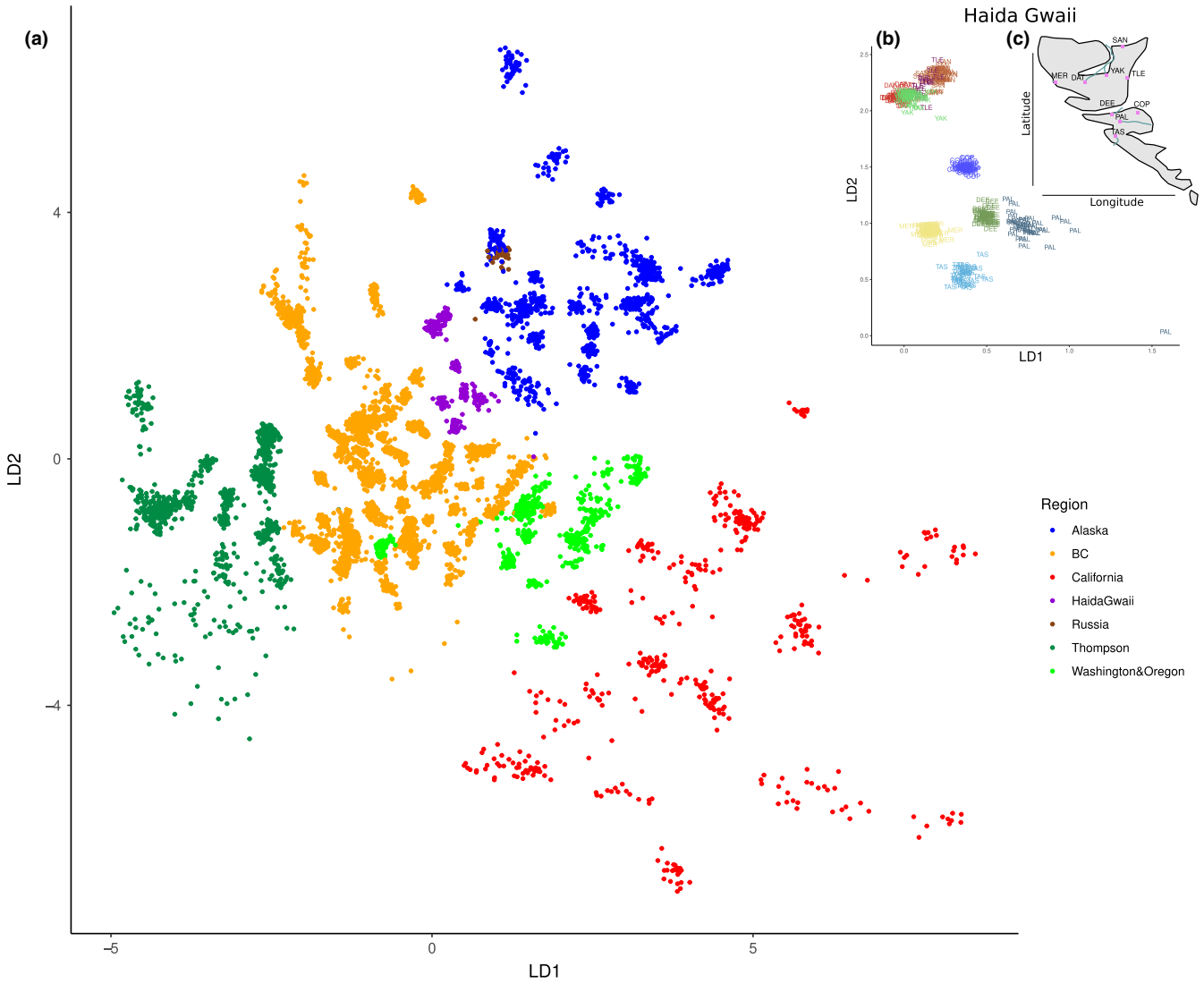
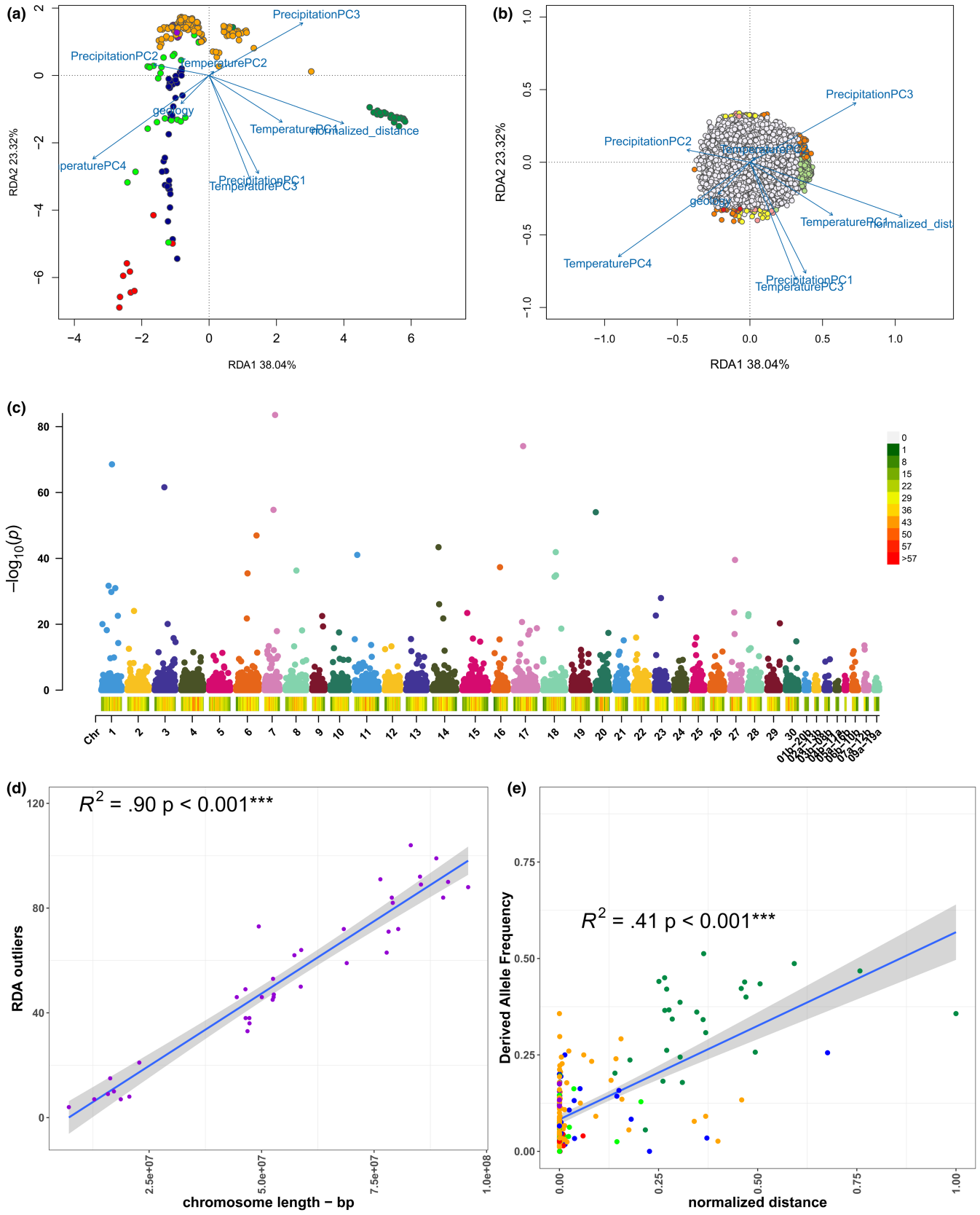


FIGURE 2 Variational autoencoders (VAE) revealing fine scale population structure at the river-scale level. Each point corresponds to an individual. (a) Global structure inferred over the whole data set; each colour defines a major region;. (b) Zoom into one region (Haida Gwaii), highlighting closely related samples within river basins as well as fine-scale structure; each colour defines a river on Haida Gwaii island with the three letters corresponding to the river abbreviation. (c) Map of Haida Gwaii sample locality. Letters match those of river in (b). Closer relationship in the geographical space is reflected in the location of the sample in the VAE. Each additional region is highlighted separately in Figure S3 [Colour figure can be viewed at wileyonlinelibrary.com]

3.4 | Parallelism in outlier distribution

The predominant role of normalized migration distance as a potentially strong selective factor based on the GEA results was investigated further using a PCA to test if populations were effectively separated into groups with long- vs. short-distance migration. We performed a test on all shared outlier SNPs associated with this variable and shared between RDA and LFMM. To capture the strongest signal possible we restricted the RDA outliers SNPs to those showing the strongest correlation with normalized distance ($r > .65$, LFMM p-value <.001) which resulted in a total of $n = 63$ SNPs. As expected, only the first PCA axis was significant and captured ~83% of the total variance. This axis discriminated the Thompson River populations from all others, due to the fact that this river system comprises only "long normalized migration distance" populations (Figure 4a). A linear model revealed a highly significant effect of normalized distance ($p < 2e^{-16}$) and of the interaction between latitude and distance ($p < 2e^{-16}$), but no effect of latitude alone ($p = .366$), indicating that this confounding factor was well accounted for (Table S5). In comparison, a linear model performed on a random set of nonoutlier loci revealed a very strong effect of geography ($p < 2e^{-16}$) as well as normalized migration distance ($p < 1e^{-814}$) and their interaction ($p < 5.21e^{-11}$), highlighting the strong geographical component of our data set that needed to be accounted for (Figure S10). To identify additional outliers associated with geographical regions, we repeated the RDA and LFMM analyses excluding Thompson River populations. The level of outlier sharing was of 237 shared SNPs between RDA and LFMM and 82 outlier SNPs were associated with normalized migration distance (Table S6). Of the 82 SNPs associated with normalized distance that were shared between the two GEA analyses,



20 displayed a strong correlation ($r > .5$) with the significant RDA axes (Figure S11 and Tables S6 and S7). This analysis clearly separated Alaskan populations displaying long normalized migration distance (e.g., Porcupine River, Mile Slough) from the remaining samples (Figure S12) and revealed outliers putatively under weaker selection, or with a more local signal of adaptation, that may have been missed when considering the whole data set. Repeating the linear model approach described above on the set of 48 SNPs displaying the strongest correlations with normalized distance revealed no significant effect of latitude ($p = .675$) but a significant effect of normalized

FIGURE 3 Landscape genomics results. (a) RDA results showing the discrimination of populations along with environmental variables; points are populations coloured according to the population from which they were sampled. (b) RDA results showing outlier SNPs (coloured points) associated with a given environmental variable. Environmental variable (Temperature 1, Temperature 2, Temperature 3, Temperature 4 and Precipitation 1, Precipitation 2, Precipitation 3, Precipitation 4) represents the first four PC axis of the Temperature and Precipitation variable extracted from the worldClim database and decorrelated through the PCA. The grey dots correspond to nonoutlier SNPs. Note that no outlier were associated to the fourth axis of Precipitation and it is therefore not displayed. (c) Manhattan plots of $-\log_{10}(p\text{-value})$ for SNPs associated with migratory distance and identified by LFMM for $K = 20$ groups. *x*-axis, left: Chr1 to Chr 30, diploid chromosomes. *x*-axis, right: Chr01b_20b to chr09a_19a, chromosomes with residual tetraploidy. SNPs falling on unassembled scaffolds were removed from the plot ($n = 10$ SNPs). (d) Relationship between outlier count on each chromosome (*y*-axis) identified by RDA and chromosome length (*x*-axis). The shaded region displays 95% confidence intervals. (e) Relationship between normalized distance (*x*-axis) and derived allele frequency (DAF) (*y*-axis) for the SNP in the vicinity of *epas1*. Each point represents a population and is coloured according to each region of sampling [Colour figure can be viewed at wileyonlinelibrary.com]

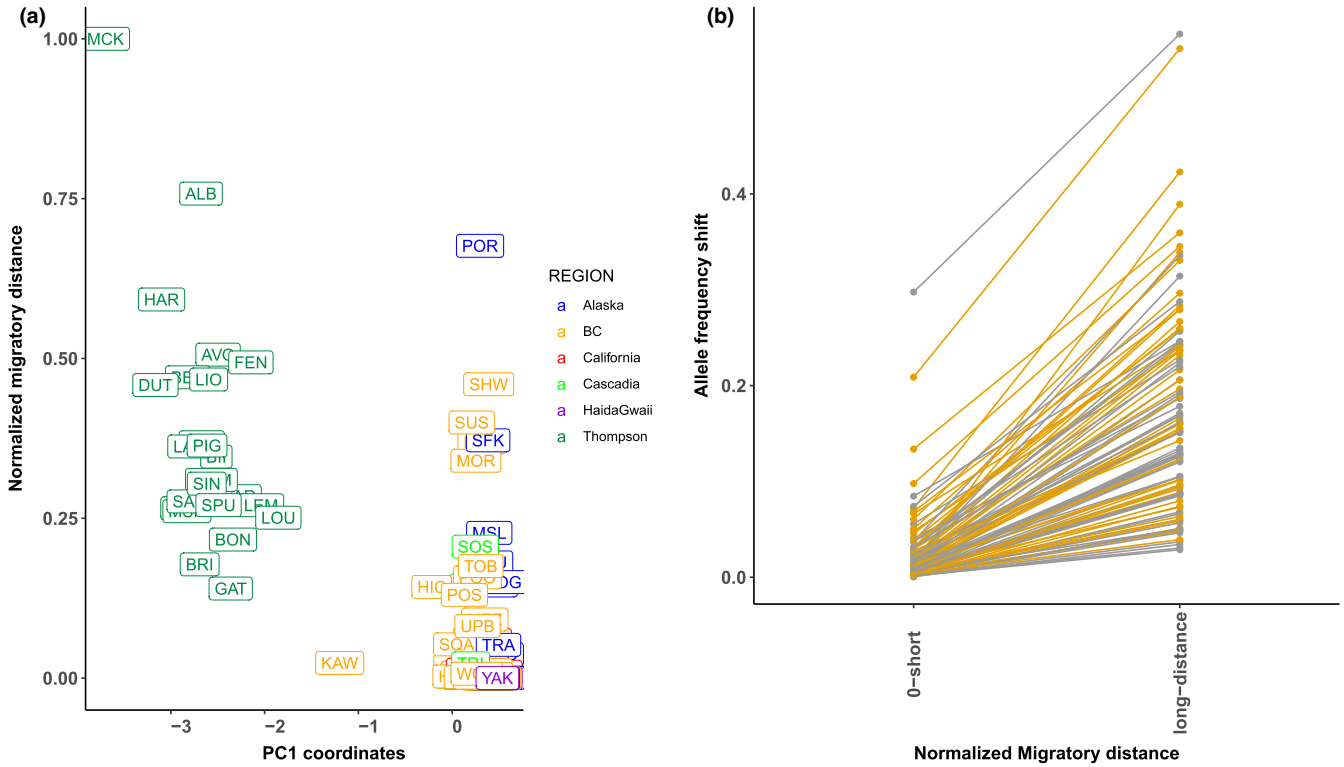


FIGURE 4 Outliers separate samples with short vs. long normalized distance from Thompson and display strong parallel frequency changes across many SNPs. (a) PC1 coordinates based on all sampling locations and the set of the strongest outlier SNPs ($n = 63$) associated with normalized distance that were shared by RDA and LFMM analyses. PC1 separates populations sampled according to the levels of genetic divergence associated with migratory distance. Each box represents a sample location and is coloured according to its regional area of sampling. (b) Parallel allele frequency shift for the 47 SNPs with the strongest signal of selection ($r^2 > .62$). The two colours (grey and yellow) are chosen randomly for ease of visualization. Rivers were divided into two groups representing "short" vs. "long" normalized migratory distance to evaluate allele frequency changes. See Figure S12 for results of the same analyses excluding Thompson River samples [Colour figure can be viewed at wileyonlinelibrary.com]

migration distance ($p < 3e^{-16}$) and of the interaction between latitude and migration distance ($p < 2e^{-16}$, $R^2 = .85$, Table S9).

Averaged allele frequency changes (ΔAF) between the 63 strongest outliers associated with long vs. short normalized migration distance were modest (mean ΔAF full data set = 0.145, mean ΔAF without Thompson = 0.0565), but always in the same direction in the whole data set (Figure 4b), even after removing Thompson populations (Figure S12). This result was in contrast to patterns observed with random SNPs, where allele frequency changes were stochastic (Figures S10 and S12). This suggests that allele frequency changes associated with migration distance responded in a similar way across multiple loci (similar directionality) among Coho salmon populations from different geographical regions.

3.5 | Signature of differentiation along the genome

Next, we looked at population differentiation among six pairs of long vs. short normalized migration distances using PBS values to search for potential 10-kb outlier windows (Figure S13). The top 10% of windows from the empirical PBS distribution identified a total of 3762 windows in the first pair of Alaskan samples, 3756 in the second pair,

3677 windows in BC and 3568 in Thompson. A total of 109 outlier windows were shared among all pairwise comparisons. Shared differentiation among population pairs was much higher at the regional level, with a total of 2206 and 775 shared windows between the two pairs in Alaska and between the pair in Thompson and BC, respectively. Similar results were obtained based on F_{ST} values (Figure S14).

3.6 | Confounding of genome scan and GEA outliers by recombination

A potential pitfall in interpreting outliers as signals of selection is that increased genetic differentiation (F_{ST} or PBS) may also be driven by the confounding effect of background selection (Burri, 2017) or by recombination rate variation (Booker et al., 2020). The ΔPBS value that we used to account for this potential confounder revealed that of the 35 shared windows among all pairs, eight displayed a ΔPBS > 0.2, while only height window displayed a ΔPBS > 0.3, indicating that background differentiation across windows was high and no particular window stands out with a strong signal of differentiation. Accordingly, linear mixed models revealed a low but significant correlation between PC-PBS and recombination rate ($p < 2e^{-16}$, $R^2_c = .0017$, $R^2_m = .0096$), supporting the effect of linked selection on the extent of observed genetic differentiation (Rougemont et al., 2020). The broad background F_{ST} distribution along the genome and among populations was illustrative of the effect of demography and recombination (Figure S14).

Similarly, regions of low recombination may be enriched in GEA outliers (Lotterhos, 2019). Here, shared outliers detected by both RDA and LFMM also fell preferentially in cold spots of recombination ($\chi^2 = 21.11$; $df = 2$, $p = .04e^5$). Accordingly, outlier windows displayed significantly lower recombination rates than "neutral" windows ($W = 7{,}040{,}186$ $p = .00216$, Figure S15) and the GLM confirmed this observation ($p = .005$, $\chi^2 = 7.85$; $R^2_m = .0094$, $R^2_c = .0095$). Therefore, it is possible that linked selection or areas of low recombination along the genome, which can enhance the effect of genetic drift, are not accurately modelled by our GEA and that these regions are enriched in false positives. In salmonids, genomic regions of residual tetraploidy may significantly influence our results. Indeed, we found that outlier SNPs in regions of residual tetraploidy did not differ from nonoutlier SNPs in terms of recombination (see Supplementary Results).

3.7 | Functional annotation of outliers

SNPeff annotations were extracted for all candidate outliers identified by the combined RDA/LFMM analyses with and without the Thompson River populations ($n = 494$ SNPs) and associated with all environmental variables. Among them, ~24% and ~41% fell within intergenic regions and in introns, respectively. Moreover, 7% were composed of missense variants ($n = 34$), 4% of 5′ or 3′ untranslated regions, and the remaining SNPs were synonymous mutations (details in Table S10). A number of putatively important genes were identified (full list provided in Table S10). For instance, 34 SNPs were associated with various zinc finger proteins, 27 were close to transcription factors, and a number of other genes identified were involved in lipoprotein synthesis (e.g., *lmf2* [missense variant]), growth and metabolic processes, spermatogenesis and one missense mutation was involved in cardiomyocite development. Furthermore, two of our four outliers with the lowest *p*-values identified only by LFMM were associated with metabolism and growth (i.e., a fatty acid-synthase like gene, and a tetraspanin-16 gene), the latter of which is known to be involved in cell development, activation and growth, and its role in controlling trygliceride excess (hypertriglyceridemia) has been documented in humans (Weissglas-Volkov et al., 2013).

Lastly, outlier annotation revealed one SNP in the close vicinity (10.485 kb) of *Epas1*, a strong candidate transcription factor gene known to be involved in adaptation to altitude and with a major role in cardiac performance (Yi et al., 2010). This SNP was associated with long normalized migration distance in all methods and displayed an increased derived allele frequency in all Thompson populations (DAF ~ 0.4) and in the most distant Alaskan samples from the Porcupine River (DAF = 0.26) compared to populations with short-distance migration. DAF was correlated with normalized migration distance (linear model $R^2 = .41$, $p < 2e^{-16}$ Figure 3e; Figure S16) and not confounded by either Latitude or Longitude ($p > .05$). Pairwise F_{ST} between populations experiencing long vs. short normalized migration distance at this specific SNP was in the range 0.40–0.60 (Table S11). The SNP is located in an intergenic region between *Epas1* and a *Pcgf5* gene that is known to be important for maintaining the transcriptionally repressive state of many genes throughout development (Yao et al., 2018). On this same chromosome, another downstream outlier gene (*agpt2*) is known to play an important role in oxygen performance (see Section 4).

4 | DISCUSSION

This study is one of the largest landscape genomics efforts performed to date on a nonmodel species. It was made possible by a large multilateral sampling effort combined with an improved and well-annotated genome. The major objective was to document patterns of adaptive genomic variation across the species range and to identify the key environmental factors contributing to such adaptive variation. Below, we discuss these different issues and propose avenues of future research to expand beyond the current limitations of our study.

4.1 | Genetic variation mirrors geography and reveal fine scale homing

We confirmed the geographical component associated with the linear postglacial expansion from a southern refugium toward the north (Rougemont et al., 2020). The PCA axes were correlated with geography, and genetic relatedness between pairs of individuals decayed

with geographical distance. These results indicate that genetic variation mirrors geography for Coho salmon, similar to human populations (Menozzi et al., 1978; Novembre et al., 2008). As previously observed, some Californian samples displayed high genetic diversity while others displayed reduced diversity, strongly suggesting recent population declines in some of these rivers (Rougemont et al., 2020; Gilbert-Horvath et al., 2016; Williams et al., 2011). Accordingly, highest levels of genetic diversity and lower β_{ST} were observed in Cascadia, rather than California, as already noted (Rougemont et al., 2020). We previously hypothesized that recent admixture between populations from several refugia resulted in an increase in genetic diversity at the contact zone, rather than in the refugia themselves (Petit et al., 2001; Rougemont et al., 2020). Interestingly, our sample from a Russian population departs from a strictly linear decrease in diversity as one moves away from the south. This suggests either intercontinental gene flow following the colonization of this area in Russia, or that a second southern refugium existed in the western Pacific, which remains to be tested with the analysis of additional sampling in Russia. A last hypothesis is that this sample site may have been wrongly labelled as originating from Russia. In the absence of further data from the western Pacific, none of these hypotheses can be excluded.

Beyond these broad-scale patterns, the analysis based on VAEs captured fine-scale structure that was not revealed in the PCA. Indeed, the PCA mainly enabled a genealogical interpretation of the results (McVean, 2009), here corresponding broadly to the divergence of the interior Thompson River populations from the rest of the data set and to large-scale IBD from north to south. In contrast, VAEs revealed a biologically meaningful characteristic of Coho salmon. This approach clearly identified sampled populations that were more closely related to each other and may be useful for biomonitoring. In contrast to a PCA, for which the interpretation is well connected with the theory, it is less straightforward to draw theoretical expectations from the VAE, just as with most machine learning-based methods (but see discussion in Battey et al., 2020). Yet, we were able to detect biologically relevant patterns summarized in just two dimensions, highlighting the power of this approach. Similar to our previous work, we observed a decrease in genetic diversity and an increase in population differentiation toward the north, yielding further support for the "out-of-Cascadia" scenario followed by post-glacial founder events moving northward (Rougemont et al., 2020).

4.2 | Multiple environmental variables potentially driving local adaptation

We sought to identify the major environmental variables responsible for driving local adaptation in Coho salmon throughout their North American range. Normalized migratory distance, interpreted as a selective environment that would result in variation in migratory phenotypes, is frequently included in GEA studies of salmonids. Distance is expected to have major selective effects, given the energy expenditure and physiological requirements involved in reproductive migration (Bernatchez & Dodson, 1987; Eliason et al., 2011). Indeed, our GEA suggested a major role of normalized migratory distance (representing 73% of all outliers), a variable commonly identified as a driver of local adaptation in other salmonid species that undergo variable anadromous migratory distances. This is the case in Chinook salmon, *O. tshawytscha* (Brieuc et al., 2015; Hecht et al., 2015), steelhead trout (Hess et al., 2016; Micheletti et al., 2018), and Arctic char, *Salvelinus alpinus* (Moore et al., 2017). A smaller yet non-negligible role of precipitation (10% of all outliers) and temperature (7%) was also observed, while geology appears to play a relatively minor role, in contrast to other studies in salmonids (Quéméré et al., 2016). The roles of precipitation and temperature as selective forces have also been demonstrated in other salmonids (Bourret et al., 2013; Dallaire et al., 2021; Eliason et al., 2011; Micheletti et al., 2018) as well as in other taxa, including in plants (Leroy et al., 2020) and *Drosophila* (Kapun et al., 2016). In salmonids, Micheletti et al. (2018) suggested that extremely high precipitation is likely to impose an additional energetic cost during migration to spawning grounds. Conversely, low flow combined with high temperatures may impede access to spawning grounds and constitute a particularly strong selective force (Gilbert & Tierney, 2018). Temperature, combined with migratory distance, is likely to be a selective force along the distribution range of Coho salmon. For instance, Alaskan populations face high temperatures during the upstream migration (17.9°C), sometimes combined with long migratory distance (e.g., Porcupine River), which we suggest may jointly impose strong selective pressures potentially driving local adaptation (Olsen et al., 2011). Similarly, southern populations in California experience strong selective pressures for local adaptation due to high temperatures, which, if combined with dry periods, may select for resistance to these factors that have important implications for the adaptation of these populations to climate change. Our sampling design may have limited power to detect loci associated with long-distance migration across the whole data set. Indeed, the majority of our sampling localities associated with long-distance migration occurs in the Thompson, five locations were in Alaska whereas only one river with long-distance migration was located in the southern range. Additional sampling focusing on this particular question would therefore be relevant. Another limit of our current study is its reliance on RADseq data, which represent a limited portion of the genome, with approximately one SNP every ~40 kb here. Given the extent of LD, it is plausible that we have missed relevant signals of small to intermediate effect as well (Lowry et al., 2017). Further whole genome sequencing on a set of carefully chosen populations from our study would provide additional clues regarding the genetic architecture of adaptation to long-distance migration.

4.3 | Recombination impacts on genome scans

The distribution of GEA outliers was correlated with chromosome length, which was recently proposed as an indication for a polygenic basis of adaptation (Salmón et al., 2021). However, this result may also be related to the lower rates of recombination in

larger chromosomes and may indicate an increased rate of false positives due to linked selection. Indeed, Rougemont et al. (2020) previously showed that recombination rates were associated with chromosome length in Coho salmon ($R^2 = .48$, $p < .0001$) and that the distribution of deleterious variants was correlated with chromosome length, suggesting linked selection. We have shown that nucleotide diversity is correlated with recombination rate (Rougemont et al., 2020), a pattern that may also be a signal of linked selection (Burri, 2017). F_{ST}-based outlier tests are known to be influenced by variation in recombination rate since the neutral estimates of differentiation will be overly inflated in regions of low genetic diversity (Cruickshank & Hahn, 2014). In contrast, our GEAs consider subtle allele frequency changes related to environmental variation. In comparison to F_{ST}-based outlier tests, recombination rate variation is not expected to significantly bias the observed GEAs (see also Lotterhos, 2019). Similar results were obtained in a study of local adaptation to climate, where group of loci involved in different aspect of the climate response were fond to display low recombination among them (Lotterhos et al., 2018). Yet, our finding that a greater number of GEA outliers are typically observed in regions of lower recombination suggests that false positive detections may occur in these regions. In addition, since RADseq outliers are more likely to be linked SNPs rather than causal variants, this increase in outliers in low recombination regions could also be due to greater retained linkage between RAD SNPs and causal variants. More theoretical work should be undertaken to help interpret these results and derive expectations regarding rates of false positives in GEAs. In addition, accounting for the duplicated nature of salmonid genomes and contrasted rates of recombination in duplicated and nonduplicated regions would require further investigations (Allendorf et al., 2015; Brieuc et al., 2014; Danzmann et al., 2008; Kodama et al., 2014).

Similarly, for the F_{ST}-based analyses, we found a positive correlation between F_{ST}/PBS and recombination rate. This result is indicative of pervasive linked selection (Burri, 2017) or extended regions of low recombination potentially resulting in an excess of false positives (Booker et al., 2020). We attempted to reduce these potential biases using an LD-based approach by correcting PBS values according to values observed across large windows (Perrier et al., 2020). In this way, windows showing a large differentiation relative to the background are more likely to represent true positive detections. This strongly reduced our set of candidate windows. Therefore, it appears that low recombination rate is probably generating false positives that are difficult to account for. A potential way forward would be to perform differentiation-based scans by considering discrete classes of recombination, but the limited number of markers available from RADseq in most studies using this genotyping technique prevents such an approach. We suggest that linked selection in areas of low recombination is likely to shape the landscape of divergence among Coho salmon populations. Similar results have been observed in other salmonids (Lehnert et al., 2020; Rougemont & Bernatchez, 2018) and other species (Burri, 2017; Perrier et al., 2020). A potential way to refine F_{ST}-based analyses would be through neutral envelope estimation obtained by combining coalescent and spatially explicit forward simulations (Haller & Messer, 2019; Kelleher et al., 2016) with heterogeneous recombination. Considering the above, we did not interpret further our F_{ST}-based results and only focused on our GEA-based results, which should be less affected by recombination rate variation (Lotterhos, 2019).

4.4 | Partial parallelism and subtle allele frequency change

As expected under the hypothesis of polygenic selective effects (Barghi et al., 2020), our set of shared SNPs across GEA methods displayed modest allele frequency changes and partial parallelism among all populations (Figure 4b). This suggests an important role of standing genetic variation enabling repeated allele frequency shifts at similar loci (Höllinger et al., 2019). An important caveat, shared by many studies relying on RADseq to investigate parallelism (e.g., Le Moan et al., 2016; Rougemont et al., 2017; Jacobs et al., 2020), is that candidate outliers are more likely to be SNPs linked to causal variants, rather than causal variants themselves (Lowry et al., 2017). In this case, true parallel patterns for causal loci may be masked if outliers represent linked neutral variation that varies among geographically structured populations. With this caveat in mind, we also observed that parallelism was strong at the regional scale (e.g., within Alaska [$n = 2$] or within the Thompson and BC [$n = 2$]) and decreased with distance. This is expected under polygenic selection (Barghi et al., 2020). This is also expected as geographically distant populations showed increased levels of genetic divergence (Bohutínská et al., 2021). For example, a recent study of local adaptation by flowering time in *Arabidopsis* compared global and regional subsamples and reported that genetic effects are distinct in different locally adapted populations (Lopez-Arboleda et al., 2021). When excluding Thompson samples, we detected one outlier, a missense variant (*tmem88* on chr9), involved in cardiomyocite development that was present at appreciable frequency only in Porcupine River (AF = 0.22). This population is the most extreme in terms of long migratory distance, lowest temperature and lowest precipitation, and illustrates the fact that if adaptation is really local then several outliers should only be population-specific, as we observed in several instances. However, the broad sampling we performed here is probably more powerful to detect allelic variants contributing to adaptation globally rather than locally. One caveat is that this population is also at the expansion front and many deleterious alleles may have surfed to high frequencies as previously observed (Rougemont et al., 2020, Rougemont et al. in prep). Therefore, it is also possible that neutral alleles have also surfed to high frequency. The functional role of the numerous missense variants detected here would need further investigation.

In our large data set, we expected adaptation to occur through many genes with small effect (polygenic adaptation) rather than through a single gene of major effect as reported previously in rainbow trout (Hess et al., 2016; Pearse et al., 2019), Atlantic salmon (Ayllon et al., 2015; Barson et al., 2015), and Chinook salmon (Narum et al., 2018; Prince et al., 2017; Thompson et al., 2020). Moreover, in

these species, a binary difference in phenotypes (e.g., late vs. early maturing phenotype in Chinook or Rrainbow trout) was present and the expression of these alternative phenotypes was associated with alternative alleles on single genes. However, such binary phenotypic differences in life history traits has never reported in Coho salmon (Quinn et al., 2016). None of our outlier SNPs mapped exactly onto genes of strong effect previously reported in several of these studies (e.g., *Rock1* and *Greb1L* associated with return timing, and *Vgll3* and *six6*, associated with age at maturity). This is similar to recent findings by Waters et al. (2021), who did not identify any significant associations between the latter two genes and age at maturity in Coho salmon. Moreover, our study did not focus on variation in life history and phenotypic traits among populations, and therefore we did not investigate similar molecular targets of selection.

Among the different genes with a putative role, the transcription factor Endothelial Pas Domain Protein 1 gene (*Epas1*) deserves further investigation. The importance of *Epas1* for altitude adaptation in Tibetan humans (Yi et al., 2010), dogs (Wang et al., 2014), mice (Schweizer et al., 2019) and birds (Xiong et al., 2020) has been well established, with evidence for spatially varying selection. More specifically, this gene is involved in improving cardiac function, muscle activity, oxygen availability and aerobic–anaerobic metabolism under hypoxia (Henderson et al., 2005). Interestingly, physiological differences in cardiac responses between locally adapted populations was demonstrated in a congeneric species, Sockeye salmon (*O. nerka*) (Eliason et al., 2011) as well as in other Pacific salmon including *O. mykiss* and *O. kisutch* (Eliason & Farrell, 2016). In particular, coastal populations with short upriver migration were shown to display a low aerobic and cardiac scope when compared with populations with a long and challenging spawning migration (Eliason et al., 2013). Moreover, salmon facing more challenging spawning migration conditions displayed greater aerobic scope, larger heart and better coronary supply, resulting in adaptation at very local scale (Eliason et al., 2011).

Therefore, in fishes, challenges to migration imposed by the environment are likely to select for individuals displaying higher cardiovascular and muscular abilities, resulting in populations that are locally adapted to reach distant and high-elevation spawning sites (Eliason et al., 2011, 2013). Here, we identify for the first time strong candidate loci potentially underlying the genetic basis of these adaptations.

5 | CONCLUSION

Our results first confirm our previous inference of a south to north, as well as west to east, gradient of genetic diversity and population divergence. This pattern reflects historical contingency associated with a single refugium followed by postglacial expansion. Our findings provide evidence for adaptation of Coho salmon to their local environment, especially with respect to migratory distance and elevation, and to a lesser extent temperature and precipitation PCA axes. We also observed parallel variation in allele frequencies at a small number of outlier loci associated with environmental variation in migratory pathways, but the extent of parallelism decreased with increasing distance among populations, suggesting that local adaptation to similar environmental pressures involves different genomic regions in different populations. Finally, in contrast to previous studies on salmonids that identified a few genes of major effect associated with sexual maturity or migration timing, we failed to detect a significant role for these genes in Coho, but rather found more support for polygenic adaptation, as also reported in other studies (Bernatchez, 2016). Since our study used a partial representation of the genome, it may have missed the signals reported at well-known genes. Therefore, future studies involving whole genome sequencing and functional validation are needed to further unravel the genetic basis of local adaptation in Coho salmon.

ACKNOWLEDGMENTS

We thank C. Hernandez for support in the laboratory. Computations were mainly performed on the Colosse (Calcul Quebec), as well as Graham and Cedar (Compute Canada) servers. We thank three anonymous reviewers as well as the associate editor for their very constructive comments and suggestions. [Correction added on 26 February 2022, after first online publication: ACKNOWLEDGMENTS section has been revised.]

CONFLICT OF INTEREST

The authors have declared that no competing interests exist.

AUTHOR CONTRIBUTIONS

Conceptualization: QR, JSM, LB; Data curation: QR, XD; Formal analysis: QR; Funding acquisition: JSM, LB; Sequencing: BB, APP; Resources: EBR, REW, DMVD, PAC, KAN, JCG, TDB, BFK, BB, APP; Software QR, EN; Writing – original draft: QR, AX; Writing – review & editing: QR, LB, XB, with input from all authors. [Correction added on 26 February 2022, after first online publication: AUTHOR CONTRIBUTIONS section has been revised.]

DATA AVAILABILITY STATEMENT

All scripts to reproduce the analyses at any step can be found on the first author's github page (https://github.com/QuentinRougemont/coho_ldscp_genomics). Raw data have been deposited on NCBI under project PRJNA647050. Vcf files are available on dryad at https://doi.org/10.5061/dryad.bzkh189b2

ORCID

Quentin Rougemont https://orcid.org/0000-0003-2987-3801
Amanda Xuereb https://orcid.org/0000-0002-3975-2299
Xavier Dallaire https://orcid.org/0000-0003-2375-561X
Eric Normandeau https://orcid.org/0000-0003-2841-9391
Alysse Perreault-Payette https://orcid.org/0000-0002-5404-8267
Bérénice Bougas https://orcid.org/0000-0003-4319-9016
Eric B. Rondeau https://orcid.org/0000-0002-6680-7961
Ruth E. Withler https://orcid.org/0000-0003-0393-9657
Kerry A. Naish https://orcid.org/0000-0002-3275-8778
John Carlos Garza https://orcid.org/0000-0002-7325-6803
Terry D. Beacham https://orcid.org/0000-0003-0987-8445

REFERENCES

Allendorf, F. W. (2017). Evolution in a toxic world. *BioScience*, *67*(6), 576–577. https://doi.org/10.1093/biosci/bix029

Allendorf, F. W., Bassham, S., Cresko, W. A., Limborg, M. T., Seeb, L. W., & Seeb, J. E. (2015). Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. *The Journal of Heredity*, *106*(3), 217–227. https://doi.org/10.1093/jhered/esv015

Ayllon, F., Kjærner-Semb, E., Furmanek, T., Wennevik, V., Solberg, M. F., Dahle, G., Taranger, G. L., Glover, K. A., Almén, M. S., Rubin, C. J., Edvardsen, R. B., & Wargelius, A. (2015). The vgll3 locus controls age at maturity in wild and domesticated Atlantic Salmon (*Salmo salar* L.) males. *PLoS Genetics*, *11*(11), e1005628. https://doi.org/10.1371/journal.pgen.1005628

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., … Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*(7582), 405–408. https://doi.org/10.1038/nature16062

Barghi N., Hermisson J., & Schlötterer C. (2020). Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics*, *21*(12), 769–781. http://doi.org/10.1038/s41576-020-0250-z

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Battey, C. J., Coffing, G. C., & Kern, A. D. (2021). Visualizing population structure with variational autoencoders. *G3 Genes|Genomes|Genetics*, *11*(1), 1–11. https://doi.org/10.1093/g3journal/jkaa036

Battey, C. J., Ralph, P. L., & Kern, A. D. (2020). Space is the place: Effects of continuous spatial structure on analysis of population genetic data. *Genetics*, *215*(1), 193–214. https://doi.org/10.1534/genetics.120.303143

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes. *Journal of Fish Biology*, *89*(6), 2519–2556. https://doi.org/10.1111/jfb.13145

Bernatchez, L., & Dodson, J. J. (1987). Relationship between bioenergetics and behavior in anadromous fish migrations. *Canadian Journal of Fisheries and Aquatic Sciences*, *44*(2), 399–407. https://doi.org/10.1139/f87-049

Bierne, N., Roze, D., & Welch, J. J. (2013). Pervasive selection or is it…? Why are FST outliers sometimes so frequent? *Molecular Ecology*, *22*(8), 2061–2064. https://doi.org/10.1111/mec.12241

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., & Kolář, F. (2021). Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21), e2022713118. https://doi.org/10.1073/pnas.2022713118

Booker, T. R., Yeaman, S., & Whitlock, M. C. (2020). Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Molecular Ecology*, *29*(22), 4274–4279. https://doi.org/10.1111/mec.15501

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*(12), 3469–3487. https://doi.org/10.1111/evo.12139

Brieuc, M. S. O., Ono, K., Drinan, D. P., & Naish, K. A. (2015). Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, *24*(11), 2729–2746. https://doi.org/10.1111/mec.13211

Brieuc, M. S. O., Waters, C. D., Seeb, J. E., & Naish, K. A. (2014). A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3 Genes|Genomes|Genetics*, *4*(3), 447–460. https://doi.org/10.1534/g3.113.009316

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*, 118–131. https://doi.org/10.1002/evl3.14

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution*, *36*(4), 852–860. https://doi.org/10.1093/molbev/msz008

Cayuela, H., Rougemont, Q., Laporte, M., Mérot, C., Normandeau, E., Dorant, Y., Tørresen, O. K., Hoff, S. N. K., Jentoft, S., Sirois, P., Castonguay, M., Jansen, T., Praebel, K., Clément, M., & Bernatchez, L. (2020). Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Molecular Ecology*, *29*(13), 2379–2398. https://doi.org/10.1111/mec.15499

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*(4), 1289–1303. https://doi.org/10.1093/genetics/134.4.1289

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Crossin, G. T., Hinch, S. G., Farrell, A. P., Higgs, D. A., Lotto, A. G., Oakes, J. D., & Healey, M. C. (2004). Energetics and morphology of sockeye salmon: Effects of upriver migratory distance and elevation. *Journal of Fish Biology*, *65*(3), 788–810. https://doi.org/10.1111/j.0022-1112.2004.00486.x

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Dallaire, X., Normandeau, É., Mainguy, J., Tremblay, J.-É., Bernatchez, L., & Moore, J.-S. (2021). Genomic data support management of anadromous Arctic Char fisheries in Nunavik by highlighting neutral and putatively adaptive genetic variation. *Evolutionary Applications*, *14*(7), 1880–1897. https://doi.org/10.1111/eva.13248

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danzmann, R. G., Davidson, E. A., Ferguson, M. M., Gharbi, K., Koop, B. F., Hoyheim, B., Lien, S., Lubieniecki, K. P., Moghadam, H. K., Park, J., Phillips, R. B., & Davidson, W. S. (2008). Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow trout and Atlantic salmon). *BMC Genomics*, *9*(1), 557. https://doi.org/10.1186/1471-2164-9-557

de Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models:

When and how much should we trust them? *Molecular Ecology*, *23*(8), 2006–2019. https://doi.org/10.1111/mec.12705

Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*, 1–20.

Eliason, E. J., Clark, T. D., Hague, M. J., Hanson, L. M., Gallagher, Z. S., Jeffries, K. M., Gale, M. K., Patterson, D. A., Hinch, S. G., & Farrell, A. P. (2011). Differences in thermal tolerance among sockeye salmon populations. *Science*, *332*(6025), 109–112. https://doi.org/10.1126/science.1199158

Eliason, E. J., & Farrell, A. P. (2016). Oxygen uptake in Pacific salmon *Oncorhynchus* spp.: When ecology and physiology meet. *Journal of Fish Biology*, *88*(1), 359–388. https://doi.org/10.1111/jfb.12790

Eliason, E. J., Wilson, S. M., Farrell, A. P., Cooke, S. J., & Hinch, S. G. (2013). Low cardiac and aerobic scope in a coastal population of sockeye salmon *Oncorhynchus nerka* with a short upriver migration. *Journal of Fish Biology*, *82*(6), 2104–2112. https://doi.org/10.1111/jfb.12120

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

Fourcade, Y., Chaput-Bardy, A., Secondi, J., Fleurant, C., & Lemaire, C. (2013). Is local selection so widespread in river organisms? Fractal geometrsy of river networks leads to high bias in outlier detection. *Molecular Ecology*, *22*(8), 2065–2073. https://doi.org/10.1111/mec.12158

Gagnaire, P.-A., Normandeau, E., Côté, C., Møller Hansen, M., & Bernatchez, L. (2012). The genetic consequences of spatially varying selection in the panmictic American eel (*Anguilla rostrata*). *Genetics*, *190*(2), 725–736. https://doi.org/10.1534/genetics.111.134825

Galtier, N., Depaulis, F., & Barton, N. H. (2000). Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics*, *155*(2), 981–987. https://doi.org/10.1093/genetics/155.2.981

Garrity, C. P., & Soller, D. (2009). *Database of the Geologic Map of North America; adapted from the map by J.C. Reed, Jr. And others (2005)*. U.S. Geological Survey Data Series 424. Retrieved from https://pubs.usgs.gov/ds/424/

Gilbert, M. J. H., & Tierney, K. B. (2018). Warm northern river temperatures increase post-exercise fatigue in an Arctic migratory salmonid but not in a temperate relative. *Functional Ecology*, *32*(3), 687–700. https://doi.org/10.1111/1365-2435.13006

Gilbert-Horvath, E. A., Pipal, K. A., Spence, B. C., Williams, T. H., & Garza, J. C. (2016). Hierarchical phylogeographic structure of coho salmon in California. *Transactions of the American Fisheries Society*, *145*, 1122–1138. https://doi.org/10.1080/00028487.2016.1201003

Goudet, J. (1995). FSTAT (version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, *86*(6), 485–486. https://doi.org/10.1093/oxfordjournals.jhered.a111627

Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., & Taylor, E. B. (2019). Aquatic landscape genomics and environmental effects on genetic variation. *Trends in Ecology & Evolution*, *34*(7), 641–654. https://doi.org/10.1016/j.tree.2019.02.013

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. https://doi.org/10.1111/j.1523-1739.2007.00693.x

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. https://doi.org/10.1093/molbev/msy228

Hecht, B. C., Hard, J. J., Thrower, F. P., & Nichols, K. M. (2015). Quantitative genetics of migration-related traits in rainbow and steelhead trout. *G3 Genes|Genomes|Genetics*, *5*(5), 873–889. https://doi.org/10.1534/g3.114.016469

Henderson, J., Withford-Cave, J. M., Duffy, D. L., Cole, S. J., Sawyer, N. A., Gulbin, J. P., & Yu, B. (2005). The EPAS1 gene influences the aerobic-anaerobic contribution in elite endurance athletes. *Human Genetics*, *118*(3–4), 416–423. https://doi.org/10.1007/s00439-005-0066-0

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: Biological Sciences*, *283*(1830), https://doi.org/10.1098/rspb.2015.3064

Höllinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genetics*, *15*(3), e1008035. https://doi.org/10.1371/journal.pgen.1008035

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N. V., Alekseyev, S. S., Hooker, O., Leong, J. S., Minkley, D. R., Rondeau, E. B., Koop, B. F., Adams, C. E., & Elmer, K. R. (2020). Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genetics*, *16*(4), e1008658. https://doi.org/10.1371/journal.pgen.100865

Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution*, *33*(5), 1317–1336. https://doi.org/10.1093/molbev/msw016

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Computational Biology*, *12*(5), e1004842. https://doi.org/10.1371/journal.pcbi.1004842

Kodama, M., Brieuc, M. S. O., Devlin, R. H., Hard, J. J., & Naish, K. A. (2014). Comparative mapping between Coho Salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. *G3 Genes|Genomes|Genetics*, *4*(9), 1717–1730. https://doi.org/10.1534/g3.114.012294

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Le Moan, A., Gagnaire, P.-A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, *25*, 3187–3202. https://doi.org/10.1111/mec.13627

Lehnert, S. J., Kess, T., Bentzen, P., Clément, M., & Bradbury, I. R. (2020). Divergent and linked selection shape patterns of genomic differentiation between European and North American Atlantic salmon (*Salmo salar*). *Molecular Ecology*, *29*(12), 2160–2175. https://doi.org/10.1111/mec.15480

Leroy, T., Louvet, J.-M., Lalanne, C., Provost, G. L., Labadie, K., Aury, J.-M., & Kremer, A. (2020). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist*, *226*(4), 1171–1182. https://doi.org/10.1111/nph.16095

Lopez-Arboleda, W. A., Reinert, S., Nordborg, M., & Korte, A. (2021). Global genetic heterogeneity in adaptive traits [Preprint]. *Evolutionary Biology*, *38*(11), 4822–4831. https://doi.org/10.1093/molbev/msab208

Lotterhos K. E., Yeaman S., Degner J., Aitken S., & Hodgins K. A. (2018). Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1545-7

Lotterhos, K. E. (2019). The effect of neutral recombination variation on genome scans for selection. *G3 Genes|Genomes|Genetics*, *9*(6), 1851–1867. https://doi.org/10.1534/g3.119.400088

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*(5), 1031–1046. https://doi.org/10.1111/mec.13100

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Molecular Ecology Resources*, *17*, 366–369. https://doi.org/10.1111/1755-0998.12677

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, *5*(10), e1000686. https://doi.org/10.1371/journal.pgen.1000686

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Menozzi, P., Piazza, A., & Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science (New York, N.Y.)*, *201*(4358), 786–792. https://doi.org/10.1126/science.356262

Micheletti, S. J., Matala, A. R., Matala, A. P., & Narum, S. R. (2018). Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). *Molecular Ecology*, *27*(1), 128–145. https://doi.org/10.1111/mec.14407

Moore, J.-S., Harris, L. N., Le Luyer, J., Sutherland, B. J. G., Rougemont, Q., Tallman, R. F., & Bernatchez, L. (2017). Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic Char. *Molecular Ecology*, *26*(24), 6784–6800. https://doi.org/10.1111/mec.14393

Narum, S. R., Alex, D. G., Micheletti, S. J., & Alejandro, M. (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. https://doi.org/10.1098/rspb.2018.0935

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, *11*(s1), 184–194. https://doi.org/10.1111/j.1755-0998.2011.02987.x

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., & Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98–101. https://doi.org/10.1038/nature07331

Ohta, T., & Kimura, M. (1969). Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics*, *63*(1), 229–238. https://doi.org/10.1093/genetics/63.1.229

Oke, K. B., Cunningham, C. J., Westley, P. A. H., Baskett, M. L., Carlson, S. M., Clark, J. Karatayev, V. A., Kendall, N. W., Kibele, J., Kindsvater, H. K., Kobayashi, K. M., Lewis, B., Munch, S., Reynolds, J. D., Vick, G. K., & Palkovacs, E. P. (2020). Recent declines in salmon body size impact ecosystems and fisheries. *Nature Communications*, *11*(1), 4155. https://doi.org/10.1038/s41467-020-17726-z

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Wagner, H. (2019). vegan: Community Ecology Package (Version 2.5-6). Retrieved from 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. https://doi.org/10.1007/s10592-010-0135-3

Pante, E., & Simon-Bouhet, B. (2013). marmap: A package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS One*, *8*(9), e73051. https://doi.org/10.1371/journal.pone.0073051

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., & Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, *3*(12), 1731–1742. https://doi.org/10.1038/s41559-019-1044-6

Perrier, C., & Charmantier, A. (2019). On the importance of time scales when studying adaptive evolution. *Evolution Letters*, *3*(3), 240–247. https://doi.org/10.1002/evl3.86

Perrier, C., Rougemont, Q., & Charmantier, A. (2020). Demographic history and genomics of local adaptation in blue tit populations. *Evolutionary Applications*, *13*(6), 1145–1165. https://doi.org/10.1111/eva.13035

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., & Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmons the utility of genomics for informing conservation. *Science Advances*, *3*(8), e1603198. https://doi.org/10.1126/sciadv.1603198

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*(4), R208–R215. https://doi.org/10.1016/j.cub.2009.11.055

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575.

Quéméré, E., Baglinière, J.-L., Roussel, J.-M., Evanno, G., McGinnity, P., & Launey, S. (2016). Seascape and its effect on migratory life-history strategy influences gene flow among coastal brown trout (*Salmo trutta*) populations in the English Channel. *Journal of Biogeography*, *43*(3), 498–509. https://doi.org/10.1111/jbi.12632

Quinn, T. P. (1984). Homing and Straying in Pacific Salmon. In JD. McCleave, G. P. Arnold, J. J. Dodson, & W. H. Neill (Eds.), *Mechanisms of migration in fishes* (pp. 357–362). Springer US. doi: https://doi.org/10.1007/978-1-4613-2763-9_21

Quinn, T. P., McGinnity, P., & Reed, T. E. (2016). The paradox of 'premature migration' by adult anadromous salmonid fishes: Patterns and hypotheses. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*, 1015–1030.

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Rougemont, Q., & Bernatchez, L. (2018). The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations*. *Evolution*, *72*(6), 1261–1277. https://doi.org/10.1111/evo.13486

Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S., & Evanno, G. (2017). Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, *26*(1), 142–162. https://doi.org/10.1111/mec.13664

Rougemont, Q., Moore, J.-S., Leroy, T., Normandeau, E., Rondeau, E. B., Withler, R. E., & Bernatchez, L. (2020). Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. *PLoS Genetics*, *16*(8), e1008348. https://doi.org/10.1371/journal.pgen.1008348

Salmón, P., Jacobs, A., Ahrén, D., Biard, C., Dingemanse, N. J., Dominoni, D. M., & Isaksson, C. (2021). Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications*, *12*(1), 2983. https://doi.org/10.1038/s41467-021-23027-w

Sandercock, F. (1991). Life history of coho salmon (*Oncorhynchus kisutch*). In C. Groot, & L. Margolis (Eds.), *Pacific Salmon life histories* (395–445). University of British Columbia Press.

Schweizer, R. M., Velotta, J. P., Ivy, C. M., Jones, M. R., Muir, S. M., Bradburd, G. S., & Cheviron, Z. A. (2019). Physiological and genomic evidence that selection on the transcription factor Epas1 has altered cardiovascular function in high-altitude deer mice. *PLoS Genetics*, *15*(11), e1008420. https://doi.org/10.1371/journal.pgen.1008420

Thompson, N. F., Anderson, E. C., Clemento, A. J., Campbell, M. A., Pearse, D. E., Hearsey, J. W., & Garza, J. C. (2020). A complex phenotype in salmon controlled by a simple change in migratory timing. *Science*, *370*(6516), 609–613. https://doi.org/10.1126/science.aba9059

Wang, G.-D., Fan, R.-X., Zhai, W., Liu, F., Wang, L., Zhong, L., & Zhang, Y.-P. (2014). Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan Plateau. *Genome Biology and Evolution*, *6*(8), 2122–2128. https://doi.org/10.1093/gbe/evu162

Waples, R. S., Naish, K. A., & Primmer, C. R. (2020). Conservation and management of salmon in the age of genomics. *Annual Review of Animal Biosciences*, *8*, 117–143. https://doi.org/10.1146/annurev-animal-021419-083617

Waters, C. D., Clemento, A., Aykanat, T., Garza, J. C., Naish, K. A., Narum, S., & Primmer, C. R. (2021). Heterogeneous genetic basis of age at maturity in salmonid fishes. *Molecular Ecology*, *30*(6), 1435–1456. https://doi.org/10.1111/mec.15822

Weir, B. S., & Goudet, J. (2017). A unified characterization of population structure and relatedness. *Genetics*, *206*(4), 2085–2103. https://doi.org/10.1534/genetics.116.198424

Weissglas-Volkov, D., Aguilar-Salinas, C. A., Nikkola, E., Deere, K. A., Cruz-Bautista, I., Arellano-Campos, O., & Pajukanta, P. (2013). Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. *Journal of Medical Genetics*, *50*(5), 298–308. https://doi.org/10.1136/jmedgenet-2012-101461

Weitkamp, L. A., Wainwright, T. C., Bryant, G. J., Milner, G. B., Teel, D. J., Kope, R. G., & Waples, R. (1995). *Status review of coho salmon from Washington, Oregon, and California*. NOAA. Retrieved from https://repository.library.noaa.gov/view/noaa/6218

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org. ISBN 978-3-319-24277-4.

Williams, T. H., Lindley, S. T., Spence, B. C., & Boughton, D. A. (2011). Status review update for pacific salmon and steelhead listed under the endangered species act: southwest. 106. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.397.9213&rep=rep1&type=pdf

Xiong, Y., Fan, L., Hao, Y., Cheng, Y., Chang, Y., Wang, J., & Lei, F. (2020). Physiological and genetic convergence supports hypoxia resistance in high-altitude songbirds. *PLoS Genetics*, *16*(12), e1009270. https://doi.org/10.1371/journal.pgen.1009270

Yao, M., Zhou, X., Zhou, J., Gong, S., Hu, G., Li, J., & Yao, H. (2018). PCGF5 is required for neural differentiation of embryonic stem cells. *Nature Communications*, *9*(1), 1463. https://doi.org/10.1038/s41467-018-03781-0

Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, *186*(S1), S74–S89. https://doi.org/10.1086/682405

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., & Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, *329*(5987), 75–78. https://doi.org/10.1126/science.1190371

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Rougemont, Q., Xuereb, A., Dallaire, X., Moore, J.-S., Normandeau, E., Perreault-Payette, A., Bougas, B., Rondeau, E. B., Withler, R. E., Van Doornik, D. M., Crane, P. A., Naish, K. A., Garza, J. C., Beacham, T. D., Koop, B. F., & Bernatchez, L. (2023). Long-distance migration is a major factor driving local adaptation at continental scale in Coho salmon. *Molecular Ecology*, 32, 542–559. https://doi.org/10.1111/mec.16339