

Conserved islands of divergence associated with adaptive variation in sockeye salmon are maintained by multiple mechanisms

Peter T. Euclide¹  | Wesley A. Larson²  | Yue Shi³  | Kristen Gruenthal^{4,5}  |
Kris A. Christensen⁶  | Jim Seeb⁷ | Lisa Seeb⁷ 

¹Department of Forestry and Natural Resources, Illinois-Indiana Sea Grant, Purdue University, West Lafayette, Indiana, USA

²National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, Juneau, Alaska, USA

³College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Juneau, Alaska, USA

⁴Alaska Department of Fish and Game, Juneau, Alaska, USA

⁵Office of Applied Science, Wisconsin Department of Natural Resources, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, Stevens Point, Wisconsin, USA

⁶Department of Biology, University of Victoria, Victoria, British Columbia, Canada

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

Correspondence

Peter T. Euclide, Department of Forestry and Natural Resources, Purdue University, 195 Marsteller St., West Lafayette, IN 47906, USA.

Email: peuclide@purdue.edu

Wesley A. Larson, National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road, Juneau, AK 99801, USA.

Email: wes.larson@noaa.gov

Handling Editor: Marta Farré

Abstract

Local adaptation is facilitated by loci clustered in relatively few regions of the genome, termed genomic islands of divergence. The mechanisms that create and maintain these islands and how they contribute to adaptive divergence is an active research topic. Here, we use sockeye salmon as a model to investigate both the mechanisms responsible for creating islands of divergence and the patterns of differentiation at these islands. Previous research suggested that multiple islands contributed to adaptive radiation of sockeye salmon. However, the low-density genomic methods used by these studies made it difficult to fully elucidate the mechanisms responsible for islands and connect genotypes to adaptive variation. We used whole genome resequencing to genotype millions of loci to investigate patterns of genetic variation at islands and the mechanisms that potentially created them. We discovered 64 islands, including 16 clustered in four genomic regions shared between two isolated populations. Characterisation of these four regions suggested that three were likely created by structural variation, while one was created by processes not involving structural variation. All four regions were small (< 600kb), suggesting low recombination regions do not have to span megabases to be important for adaptive divergence. Differentiation at islands was not consistently associated with established population attributes. In sum, the landscape of adaptive divergence and the mechanisms that

Peter Euclide and Wesley A. Larson contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

create it are complex; this complexity likely helps to facilitate fine-scale local adaptation unique to each population.

KEYWORDS

adaptive divergence, genomic islands of divergence, inversion, sockeye salmon, whole genome re-sequencing

1 | INTRODUCTION

It is increasingly clear that adaptive loci are often clustered in relatively few regions of the genome, termed genomic islands of divergence (Wolf & Ellegren, 2017). This is especially true for populations, ecotypes or species in the early stages of diverging, where recent or ongoing gene flow appears to homogenise most of the genome, while genomic islands display high differentiation (Aeschbacher et al., 2017; Feder et al., 2012; Kirkpatrick & Barton, 2006).

The creation and maintenance of genomic islands of divergence, during adaptation with gene flow, requires that advantageous alleles can be isolated to promote the formation of favourable allelic combinations while protecting them from the disruptive force of recombination (Tigano & Friesen, 2016; Yeaman, 2013). Multiple potential mechanisms that could aid this process were proposed, including divergence hitchhiking (Ma et al., 2018; Via, 2012), clustering of adaptive loci in low recombination regions (Samuk et al., 2017; Wang et al., 2022) and the utilisation of structural polymorphisms such as chromosomal inversions (Faria et al., 2019; Kirkpatrick & Barton, 2006).

Divergence hitchhiking occurs when strong divergent selection reduces gene flow in genomic regions near genes or other targets of selection (Via, 2012). Regions with elevated divergence can span multiple megabases if recombination is substantially reduced due to assortative mating (Via, 2012). Alternatively, selection can exploit existing low recombination regions to preserve co-adapted loci, leading to clustering of adaptive alleles in these regions (Samuk et al., 2017). Finally, selection can utilise structural variation, such as chromosomal inversions, to isolate adaptively important genes. Chromosomal inversions are a common type of structural variant that occur when segments of DNA break off and reattach within the same chromosome but in reverse orientation. Inversions are not necessarily deleterious and may not impact gene function unless the inversion breakpoint occurs within a gene. In these cases, inversions and the alleles they include may remain common in a population. Recombination across inversion types is rare due to mechanisms that include disruption of pairing and crossing over during meiosis and, in some cases, inviability of recombinant gametes (reviewed in Huang & Rieseberg, 2020). Genes found on inversions are therefore generally shielded from recombination, potentially promoting adaptive divergence even in the face of high gene flow (Feder et al., 2012; Rieseberg, 2001; Tigano & Friesen, 2016). Therefore, these regions can be important sources of adaptive variation between populations with limited reproductive isolation.

Examples of adaptively important inversions that enable divergence with gene flow are increasingly recognised (Wellenreuther & Bernatchez, 2018). Prominent case studies include inversions differentiating ecotypes in Atlantic cod (*Gadus morhua*, Kirubakaran et al., 2016), sunflowers (*Helianthus annuus*, Todesco et al., 2020) and deer mice (*Peromyscus maniculatus*, Hager et al., 2022). Adaptively important structural variants have also been identified in several Salmonid species, including a supergene segregating North American Atlantic salmon populations (Stenløkk et al., 2022) and high structural variation between dwarf and normal Lake Whitefish (Mérot et al., 2023). Previously characterised adaptive inversions tended to be large, spanning multiple megabases and old, reflecting genetic variation that arose hundreds of thousands to millions of years ago (Bernatchez et al., 2017). In addition, there are numerous examples of islands of divergence in populations with at least moderately high gene flow that are likely the result of non-structural mechanisms including divergence hitchhiking (Ma et al., 2018), reduced recombination (Samuk et al., 2017) or other processes not explicitly investigated (Roberts Kingman et al., 2021; Thompson et al., 2020). A recent simulation study indicated that, when gene flow is moderate, islands of divergence created by structural and non-structural variation resulted in similar levels of adaptation (Schaal et al., 2022). When gene flow increased, however, simulations including inversions achieved a higher level of local adaptation. Nevertheless, few empirical studies investigated the frequency of genomic islands associated with structural changes relative to non-structural changes or how gene flow influences the frequency of these mechanisms (but see Shi et al., 2021).

Here, we investigate the mechanisms responsible for creating islands of divergence and the patterns of adaptive variation linked to these islands in sockeye salmon (*Oncorhynchus nerka*). Sockeye salmon have colonised a wide range of spawning habitats, leading to the formation of distinct ecotypes that are often found in close spatial proximity (Quinn, 2005). Although sockeye salmon exhibit strong philopatry, the proximity of their spawning habitats also presents numerous opportunities for gene flow (Peterson et al., 2014). In most of the cases, individuals will return to the same beach or stream reach to spawn, but especially in years when sockeye salmon abundance is large, discrete spawning sites may begin to overlap as individuals are forced to nest in sub-optimal habitats between spawning sites (Quinn, 2005). Sockeye salmon runs can be extremely large in certain years, with millions of individuals returning to the same lake drainage or area to spawn. Because spawning takes place over a short period of time (1–3 months in summer), and space

is limited, nest competition, predation and disease can be intense (Quinn, 2005). These factors have all contributed to high natural selection pressure and diversifying selection that influences sockeye salmon across a small spatial scale in response to local spawning site structure and ecology.

Local adaptation to spawning habitat has resulted in a hierarchical diversity of sockeye salmon ecotypes. Anadromous sockeye salmon are grouped into two primary ecotypes: lake-type and sea/river-type (Wood et al., 2008). Sea/river type sockeye salmon have higher stray rates than lake-type and are thought to have colonised lake systems then subsequently evolved into lake-type (recurrent evolution hypothesis; Wood et al., 2008). Lake-type sockeye salmon have further diversified into phenotypically distinct groups that can be differentiated by spawning habitat. Lake-type sockeye salmon spawn in a variety of habitats, including small streams, deep lake beaches and large rivers (Quinn, 2005). The morphology and phenology of sockeye salmon spawning in each of these habitats can vary substantially (Quinn et al., 1995, 2001). Neutral population genetic studies of lake-type sockeye salmon revealed strong hierarchical structure partitioned by drainage, with much lower genetic differentiation among populations and ecotypes within the same drainages (Beacham et al., 2004; Habicht et al., 2007). In contrast, markers under selection have displayed extremely high differentiation among populations within the same drainages (Ackerman et al., 2013; Creelman et al., 2011; Russello et al., 2012). This suggests that lake-type sockeye salmon likely experience high selection pressure in the presence of high gene flow.

The first genetic evidence of adaptive divergence in sockeye salmon was documented in lake-type individuals at the genes of the major histocompatibility complex, which are highly differentiated among spawning sites that are in the same drainage and separated by as little as 10s of meters (Larson et al., 2014; McClelland et al., 2013; McGlaufflin et al., 2011; Miller et al., 2001). Additional studies using restriction site-associated DNA (RAD) sequencing and targeted sequencing revealed a number of islands of divergence, some of which were found in multiple drainages (Larson et al., 2017, 2019; Limborg et al., 2017; Veale & Russello, 2017a, 2017b). However, genetic variation at islands of divergence is not consistently associated with the same spawning habitats (Larson et al., 2019), with the exception of an island on Chr 12, which displays differentiation between beach and tributary (creeks and rivers) spawners across the species range (Larson et al., 2019; Nichols et al., 2016; Tigano & Russello, 2022; Veale & Russello, 2017a). The body of research on adaptive divergence in sockeye salmon suggests that the same genes and genomic islands are involved in adaptive divergence as new systems are colonised, but that variation in these genomic regions is not necessarily partitioned by spawning habitat and may be influenced by a mosaic of selective pressures.

Although previous research provided strong evidence that islands of divergence are involved in adaptive radiation of sockeye salmon, most studies were based on genome scans with fewer than 4000 SNPs, with only 7–15 SNPs found in the islands of divergence (but see Tigano & Russello, 2022). The low genomic marker density

and lack of a reference genome available to these studies made it difficult to map the architecture of the islands of divergence or elucidate the genomic mechanisms underlying their creation (Benjelloun et al., 2019). Here, we leverage whole genome resequencing and published reference genome (Christensen et al., 2020) to provide a more complete characterisation of the landscape of adaptive divergence in sockeye salmon and to improve our understanding of the mechanisms that create and maintain islands of divergence. We sequenced sockeye salmon from multiple ecotypes across three drainages in Southwest Alaska to investigate fine-scale variation and merged these data with resequencing data from Christensen et al. (2020) to anchor our findings in the context of the full species range. We found that: (1) the landscape of adaptive divergence in sockeye salmon is characterised by many small, but highly divergent islands of SNP markers, (2) many of these islands of divergence may be conserved through structural variation and, to a lesser degree, divergence hitchhiking, (3) linkage among loci on islands of divergence is strong across the sockeye range, suggesting that some may have a long evolutionary history and may be repeated targets of selection and (4) while small, each island contained numerous genes that could be targets of selection and have some adaptive function for certain spawning populations.

2 | METHODS

2.1 | Sampling design

We resequenced genomes of sockeye salmon from seven populations in Southwest Alaska, USA (these samples are a subset of those analysed in Larson et al., 2019). Fin-clips from 27 individuals per population (189 individuals total) were obtained from three lake-type spawning populations in the each of the Kvichak River and Wood River drainages as well as one putatively ancestral sea/river population in the Nushagak River drainage (Table 1). Lake-type samples were further subdivided into the following groups based on spawning habitat: mainland beaches, island beaches, creeks and rivers (Figure S1). Mainland and island beaches are similar except island beaches are found in the middle of lakes where they are highly affected by wind and wave action (Stewart et al., 2003). Creeks are narrow (< 5 m wide) and shallow (< 0.5 m deep on average) while rivers are wide (> 30 m wide), deep (> 0.5 m deep) and fast flowing (Quinn et al., 2001). All samples were collected from spawning adults by Alaska Department of Fish and Game between 1999 and 2013 and provided as extracted DNA (extracted with Qiagen DNAeasy Blood and Tissue Kits, Hilden, Germany).

2.2 | Whole genome library preparation and sequencing

Libraries were prepared according to Baym et al. (2015) and Therikildsen and Palumbi (2017) with the following modifications. Input DNA

TABLE 1 Information on populations included in this study. The sample size for each population was 27.

Population	Abbreviation	Drainage	Spawning type	Spawning habitat	Latitude (N)	Longitude (W)	Collection year
Knutson Beach	KNUT	Kvichak River	Lake	Mainland Beach	59.80	-154.16	1999
Woody Island Beach	WOOD	Kvichak River	Lake	Island Beach	59.75	-154.28	2001
Iliamna River	ILIA	Kvichak River	Lake	River	59.75	-153.87	2004
Anvil Beach	ANVB	Wood River	Lake	Mainland Beach	59.55	-158.79	2011
Teal Creek	TEAL	Wood River	Lake	Creek	59.48	-158.73	2013
Agulowak River	AGUL	Wood River	Lake	River	59.41	-158.88	2009
Klutapuk Creek (Upper Nushagak R)	UPNK	Nushagak River	Sea/River	River	60.34	-157.32	2001

Note: See Larson et al. (2019) for more information on each collection. Six of the seven sites sampled represent lake-type sockeye salmon which are believed to have evolved from an ancestral form of sea/river type sockeye salmon.

was normalised to 10ng for each individual. Steps for 96-well AM-Pure XP (Beckman Colter; Brea, CA) purification; product quantification, normalisation and pooling; and size selection were replaced with a SequalPrep (ThermoFisher Scientific, Waltham, MA, USA) normalisation and pooling protocol, similar to that used in GT-seq (Campbell et al., 2015). We used three SequalPrep plates per each of the two 96-well tagmented and adaptor-ligated DNA library plates and pooled the full eluate per individual DNA library to increase total yield. Normalised pooled libraries were subject to a 0.6X size selection, purification and volume concentration with AMPure XP following Therikildsen and Palumbi (2017). In-house QC consisted of visualisation on a precast 2% agarose E-Gel (ThermoFisher Scientific) and quantification with a Qubit HS dsDNA Assay Kit (ThermoFisher Scientific). We constructed two libraries each containing 96 individuals and each of these libraries was sequenced on three Novaseq S4 lanes (six lanes total) at Novogene (Sacramento, CA, USA).

2.3 | Genotype calling and quality control

Variants and genotypes were called using the Genotype Analysis Toolkit (GATK) version 4.1.7 (DePristo et al., 2011; McKenna et al., 2010) and a protocol that closely followed Christensen et al. (2020). Paired-end reads were aligned to the sockeye salmon genome (GCF_006149115.2; Christensen et al., 2020) with BWA MEM v.0.7.17 (Li, 2013) and indexed and sorted with Samtools v.1.10 (Li et al., 2009). Next, readgroups for each alignment file (bam file) were assigned using Picard v2.22.6 (AddOrReplaceReadGroups; <http://broadinstitute.github.io/picard>). Individual bam files produced on separate sequencing lanes were merged, and PCR duplicates were marked using the MarkDuplicates function from Picard with stringency set to 'LENIENT'. Individual genomic VCF files (gvcf) were generated from alignments using HaplotypeCaller from GATK. A single database was created containing all individual gvcf files using GenomeDBImport from GATK. Once the variants from all individuals had been added to the database, joint-genotyping was conducted using the GenotypeGVCFs function. The resulting variant file (vcf) was then hard filtered

using the VariantFiltration function (filter expression= $QD < 2.0 \parallel FS > 60.0 \parallel SOR < 3.0 \parallel MQRankSum < -12.5 \parallel ReadPosRankSum < -8.0$). All variants that passed hard filter were used in conjunction with three datasets used previously as truth datasets by (Christensen et al., 2020) for GATK's VariantRecalibrator function. The tranches file generated by VariantRecalibrator was subsequently used as the input for the ApplyVQSR function and to produce a corrected vcf file and submitted to additional variant filtration in VCFtools v.0.1.16 (parameters: $--maf 0.05, --max-alleles 2, --min-alleles 2, --max-missing 0.9, --remove-filtered-all --remove-indels$; Danecek et al., 2011). Finally, loci with an allele balance of less than 0.2 were marked. The resulting vcf file constituted our baseline file for all other analysis and downstream processing.

2.4 | Characterising the genomic landscape of differentiation

Genetic relationships among and within drainages were evaluated using a neighbour-joining (NJ) tree of Nei's genetic distance (poppr and ape; Kamvar et al., 2014; Paradis & Schliep, 2018), pairwise Weir and Cockerham F_{ST} (snpR; Hemstrom & Jones, 2023). Associations between individuals was evaluated using principal component analysis (PCA) implemented in pcadapt (Privé et al., 2020). To reduce the impact of linked loci in the dataset, PCAs were conducted using linkage disequilibrium (LD) clumping (size = 500, Th = 0.4). These thresholds were chosen following initial pilot runs of pcadapt that varied windows from 100 to 2000 SNPs and squared coefficients from 0.05 to 0.4 that showed that changes in individual scores plateaued at these settings. Bi-plots of individual loading scores first three principal components were plotted in R to identify the relationships among individual samples.

Locus-specific F_{ST} estimates were calculated in PLINK v1.90 (Purcell et al., 2007) separately within each drainage excluding the Nushagak River samples. Individuals from all three spawning sites were used to calculate F_{ST} , therefore locus F_{ST} values represent the overall allele frequency variation within each drainage. Values of F_{ST} were visualised using Manhattan plots and putative islands of

divergence were quantitatively identified using a Hidden Markov Model (HMM; *HiddenMarkov* version 1.8-11; Hofer et al., 2012). This approach assigned each SNP to one of three underlying states (genomic background, regions of high differentiation, or regions of low differentiation) based on their F_{ST} values, following the methods detailed in Marques et al. (2016) and Shi et al. (2021). To reduce false positives, we retained only islands that met the HMM 'regions of high differentiation' threshold and contained at least 10 high F_{ST} SNPs in the top 0.1% of F_{ST} distribution. Regions that passed these thresholds were defined as HMM islands. HMM islands that fell within 100kb of one another were subsequently grouped to reduce redundancy. This analysis separately for the Kivchak River and Wood River drainages. All islands were summarised to identify which spawning sites they differentiated based on locus-specific pairwise F_{ST} estimates. Pairwise F_{ST} was then calculated among populations within each drainage using PLINK v1.90. To identify which spawning habitats grouped HMM islands differentiated between the average F_{ST} of all SNPs within each island was calculated and visualised as a pairwise heatmap constructed using ggplot2.

Because the purpose of our study was to identify to what degree patterns of local adaptation were consistent across the sockeye range, the majority of the analysis was conducted on genomic regions that contained HMM islands in both drainages. Gene flow between drainages is low, therefore genomic regions that are highly differentiated in both drainages are most likely to either have undergone parallel adaptive divergence independently as these drainages were colonised separately (likely from colonisers with similar genetic variation), or be the result of ancestral polymorphism evolved prior to colonisation. Either of these mechanisms would indicate that these regions may represent evolutionarily important genetic variation is being maintained in the species at a broader level.

2.5 | Investigating the genomic mechanisms underlying islands of divergence

We evaluated patterns of LD and conducted PCA to distinguish between islands that likely arose through structural variation versus non-structural variation. We focused on methods to detect common patterns associated with chromosomal inversion-like structural variants. Putative chromosomal inversions can be indirectly distinguished from other types of islands based on (1) strong boundaries where LD is high within the inversion then decays rapidly around inversion breakpoints and (2) genotypes clustering into three distinct groups in multivariate analyses representing the two inversion orientations, with heterozygotes between the arrangements forming an intermediate cluster (Huang et al., 2020). We considered islands to be putative inversions if they met both criteria. However, it is important to note that inversions can show complex patterns and that similar patterns can be created by non-structural variants (Mahmoud et al., 2019). Therefore, the classification of genomic regions that we conduct should be considered preliminary until regions are further validated with methods such as long read sequencing.

LD (measured as R^2 among pairs of loci) was calculated for each chromosome (Chr) containing an island using PLINK v1.9. We then assessed whether LD among island SNPs was higher than SNPs on the surrounding Chr by randomly selecting 1000 regions outside of the island, but of identical size and calculating the LD among SNPs within these regions (see Shi et al., 2021). Significance was tested by calculating a Z-score of LD for each island using the distribution constructed by resampling random regions then calculating a one-tailed *p*-value ($\alpha = .05$) with a Z-test based on this Z-score.

Heatmaps of LD in each island and the surrounding ~200 Kb were constructed and used to visualise island boundaries in ggplot2 (Wickham, 2009). Islands of divergence contained some areas of low differentiation and low LD. We identified the longest continuous block of mutually linked SNPs containing R^2 values in the top 5% for that Chr to refine island boundaries into tightly linked haploblocks. The starting position of each haploblock was defined as the nucleotide position of the first SNP that was strongly linked (top 5% of LD for a given Chr) to the largest number of downstream SNPs in the 5' direction. The end position was defined as the position of the SNP that was linked to the largest number of upstream SNPs (top 5% of LD) in the 3' direction. In all cases this resulted in a LD threshold for 'strongly linked SNPs' of $R^2 > 0.5$ and < 0.75 . All SNPs within the refined haploblocks were then used to conduct a PCA to assess whether the island expressed characteristics of a chromosomal inversion (ade4; Thioulouse et al., 2018). Only islands that exhibited both large and consistent LD blocks and three distinct groups clustering along PC1 in PCAs were considered putative inversions.

2.6 | Assessing patterns of variation and diversity in islands of divergence

Discrete genotypes at islands of divergence were assigned for each individual using *K*-means clustering ($K=3$) implemented through adegenet (Jombart & Ahmed, 2011) and confirmed by assessing heterozygosity of individuals assigned to each genotype cluster (low heterozygosity = homozygote, high heterozygosity = heterozygote). For islands classified as inversions, the entire refined haploblock was used for all analyses. The heterozygosity of individuals within each cluster was calculated with the assumption that individuals homozygous for a particular haplotype would have substantially lower heterozygosity than heterozygous individuals containing an allele from each haplotype. Genotypes for haploblocks were defined as homozygote allele 1 (AA), heterozygous (AB) and homozygous allele 2 (BB).

A modified version of the above approach was used to identify and genotype haploblocks for the island that was not classified as a chromosomal inversion. Haploblocks were defined using only SNPs that contained high loadings (top 25%) for DAPC axis 1 due to more variable patterns of LD. This was done to reduce the dimensionality of the data by focusing the analysis on the SNPs primarily responsible for explaining the largest amount of variation associated with axis 1. We then subjected this subset of SNPs to the same *K*-means

clustering pipeline but used $K=6$ instead of 3 to account for an apparent tri-allelic state associated with incomplete linkage across the island. Heterozygosity was evaluated for each of the clusters to identify homozygous and heterozygous haplotypes. We calculated population level genotype frequencies and observed heterozygosity once haplotypes were identified for each island.

2.7 | Investigating the conservation of islands of divergence using range-wide data

Because the islands of divergence we identified were consistent among spatially isolated drainages in Alaska, we hypothesised that these regions may be conserved in other sockeye populations. To test this, we merged the dataset generated in the present study with whole-genome data from 78 sockeye salmon (kokanee excluded) from Christensen et al. (2020). This dataset was sequenced to a similar depth of coverage and was processed using an almost identical GATK4 pipeline. The dataset included 16 spawning populations that we grouped into five drainage regions: Bristol Bay ($N=12$ individuals), Fraser/Columbia river basins ($N=47$), Gulf of Alaska ($N=8$), Northern British Columbia ($N=9$) and Russia ($N=2$). The variants identified in Christensen et al. (2020) were merged with ours using bcftools v.1.11 (Danecek et al., 2021) by retaining variants that intersected between the two datasets, had a genotyping rate $>80\%$, and were positioned within one of the refined haploblock regions. The resulting vcf file was phased using BEAGLE v. 5.1 with default settings (Browning & Browning, 2007).

We predicted that individuals would cluster by haplogroup and not geography, if haplogroups were conserved across the species range. This prediction was tested by conducting a PCA for SNPs on each island of divergence and qualitatively evaluating groupings. Individuals were also grouped using a NJ tree, and genotype heatmaps were constructed to show allele patterns across each haplogroup. If haplogroups were conserved, we expected individuals from different drainages would cluster together as neighbours and have similar genotypes across the haplogroup region. It is important to note that sample sizes for populations in Christensen et al. (2020) were unevenly distributed. Therefore, the major goal of our range-wide analyses was to identify whether these samples supported haplotype groupings identified in the primary study location, Bristol Bay, rather than an investigation of range-wide patterns of population structure.

2.8 | Annotation of genes in islands of divergence

The functional role of genes located on each of the conserved islands of divergence were investigated by extracting gene annotations (feature= gene) from the assembled sockeye salmon reference genome using the gene feature table (GCF_006149115.1_Oner_1.0_feature_table.txt). The putative functions of genes located in HMM islands were assessed by reviewing gene summaries and gene

ontology information available on NCBI for target genes and their orthologues. Literature citations were included for specific functions or effects referenced outside of the NCBI summary descriptions.

3 | RESULTS

3.1 | Sequencing and SNP calling

Genome resequencing of 189 individuals produced 28.76 million putative variants of which 1.98 million SNPs were retained (no-indels, biallelic SNPs only, minor allele frequency ≥ 0.05 , and max missingness of 10% for each SNP). Across these 1.98 million SNPs, individuals were missing genotypes at an average of 4.4% of sites and contained an average depth of coverage of 6.8X (SD=1.3X), and the average depth per-variant was 6.6X (SD=3.6X). Following recalibration and filtration, variant quality was moderate and within acceptable range for proposed analyses (Mean [standard deviation]: QD=20.78 [7.17]; FS=6.73 [13.75]; SOR=0.87 [0.76]; MQ=57.81 [4.71]; MQRankSum=-0.32 [0.71]; ReadPosRankSum=0.07 [0.39]; Table S1).

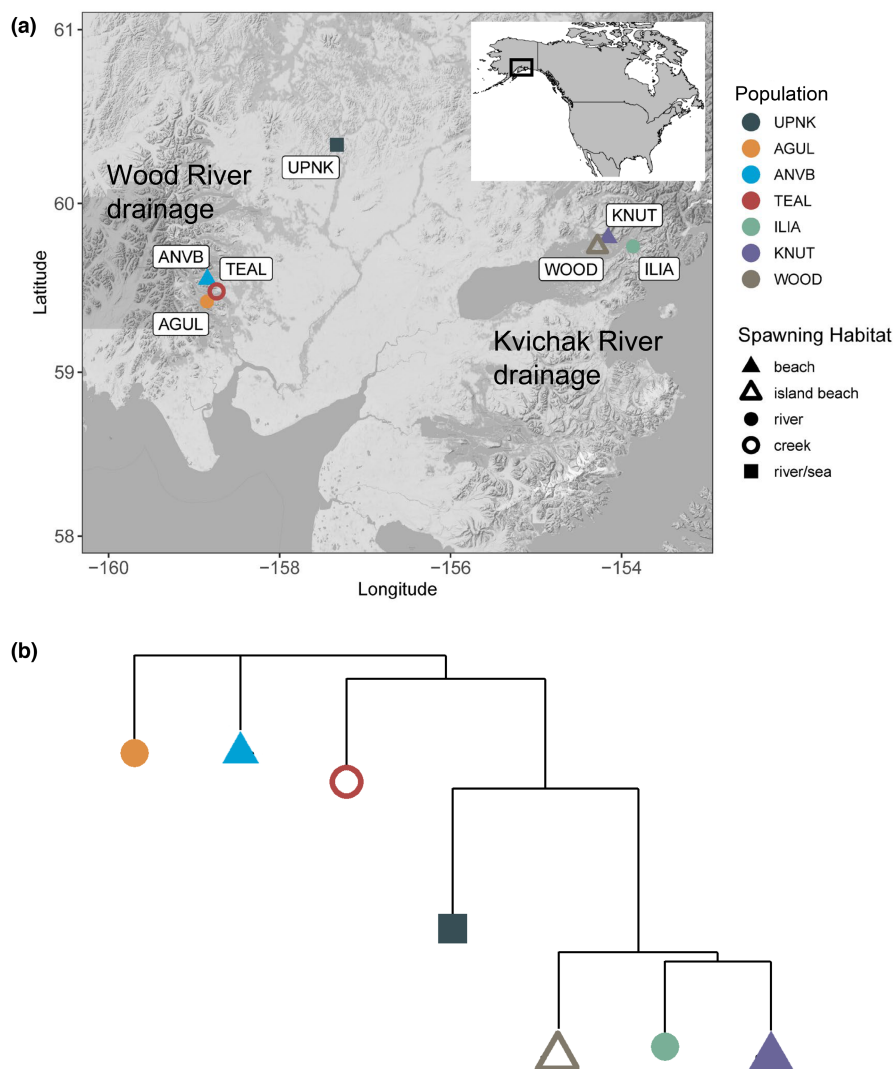
3.2 | Characterising the genomic landscape of differentiation

Spawning populations differed primarily along river drainage boundaries (Figure 1). Pairwise genetic distance among sites was much lower within drainages (Kvichak River average $F_{ST}=0.003$, standard deviation=0.003; Wood River average $F_{ST}=0.005$, standard deviation=0.002; Figure 1) than between drainages (average $F_{ST}=0.035$, standard deviation=0.001). Kvichak River and Wood River populations were equally diverged from the sea/river type sockeye spawning population in Nushagak River drainage (average $F_{ST}=0.027$, standard deviation=0.001 and average $F_{ST}=0.023$, standard deviation=0.001). PCA supported pairwise F_{ST} and neighbour joining tree analysis. Individuals from the same drainage all clustered closely together with PC1 separating samples from the three drainages and PC2 separating the sea/river-type sockeye salmon in the Nushagak River Drainage from the lake-type sockeye salmon in the Kvichak and Wood River drainages (Figure S2a).

Within each of the Kvichak and Wood River drainages, PCA of sockeye salmon clustered individuals by spawning sites in a way that supported the slight sub-branching found in neighbour joining trees (Figure 1b; Figure S2b,c). In the Kvichak River drainage, sockeye salmon from WOOD separated from ILIA and KNUT along the first PC while individuals from ILIA and KNUT overlapped on all the first three axes. In the Wood River drainage, TEAL separated from AGUL and ANVB along the first PC (Figure S2b,c). AGUL and ANVB could also be separated along the first PC but to a lesser degree (Figure S2c).

The markers contributing to intra-drainage differentiation were highly heterogeneous across the genome, with multiple conspicuous

FIGURE 1 (a) Location of seven spawning populations sampled in the Bristol Bay region in Alaska, USA. AGUL, ANVB and TEAL are in the Wood River drainage; UPNK is in the Nushagak River drainage; and ILA, KNUT and WOOD are in the Kvichak River drainage. More information on each population is found in [Table 1](#). (b) NJ tree showing Nei's genetic distance among populations. Populations are grouped by drainage by a random set of 10,000 SNPs.



F_{ST} peaks found among spawning populations within the Kvichak and Wood river drainages (Figure 2). We identified 33 high F_{ST} HMM islands in the Wood River drainage and 31 high F_{ST} HMM islands in the Kvichak River drainage (Figure 2, Table S2). These islands were generally small (minimum = 3.1 kb; maximum = 247 kb) and covered a total of 1.3 Mb of the genome in the Kvichak River drainage, and 1.5 Mb of the genome in the Wood River drainage population. Allele frequency differences among spawning populations at markers within these islands were high (Wood River islands: mean F_{ST} = 0.240; max F_{ST} = 0.834, Kvichak River islands: mean F_{ST} = 0.227; max F_{ST} = 0.725) compared to the genome-wide average. Of 33 HMM islands in the Wood River drainage, and 31 in the Kvichak River drainage 17 and 15 respectively were within 100 kb of an adjacent island. Adjacent islands within 100 kb of each other were combined resulting in a final set of 16 Wood River islands spanning 13 chromosomes and 15 Kvichak River islands spanning 11 chromosomes (Figure 2). In the Wood River drainage, the island with the highest average F_{ST} was on Chr 12 and differentiated ANVB from TEAL (mean F_{ST} = 0.61) and AGUL (mean F_{ST} = 0.37). Islands Chr 3_1, 5_1, 7_1 and 7_2, 13_1, 25_1, and 28_1 all differentiated AGUL from TEAL and ANVB. Islands Chr 18_1, 19_2, 22_2 and 28_2 primarily differentiated TEAL from

AGUL and ANVB. Islands Chr 6_1, 12_1, differentiated ANVB from AGUL and TEAL. The remaining islands had moderate F_{ST} differentiating all spawning sites. In the Kvichak River drainage, the island with the highest average F_{ST} was Chr 3_1 and differentiated WOOD from KNUT (mean F_{ST} = 0.53) and ILIA (mean F_{ST} = 0.34). Comparisons with WOOD generally contained the highest mean F_{ST} , while ILIA and KNUT were only notably differentiated on islands Chr 12_1, 13_2, 13_4 and 22_3. Given that WOOD generally had the highest F_{ST} with the other sites in the drainage, it is notable that ILIA and WOOD had low F_{ST} for the Chr 12_1 (mean F_{ST} \leq 0.01) and Chr 22_3 islands (0.016). In both drainages, pairwise comparisons were either close to zero for one out of three comparisons, or moderately high (generally F_{ST} > 0.1) for all three comparisons (Table S3).

The genomic positions of 13 HMM islands in the Wood River populations directly overlapped with 15 HMM islands in the Kvichak River populations (Table S4). Overlapping HMM islands included the islands Chrs 3_1, 5_1, 12_1, 13_2 and covered 137.4 kb, 90.9 kb, 274.1 kb, and 526.7 kb, respectively (Figure 2). Because the goal of this study was to investigate regions that consistently differentiate intra-drainage sockeye salmon spawning sites, we chose to focus the remaining analyses on these regions.

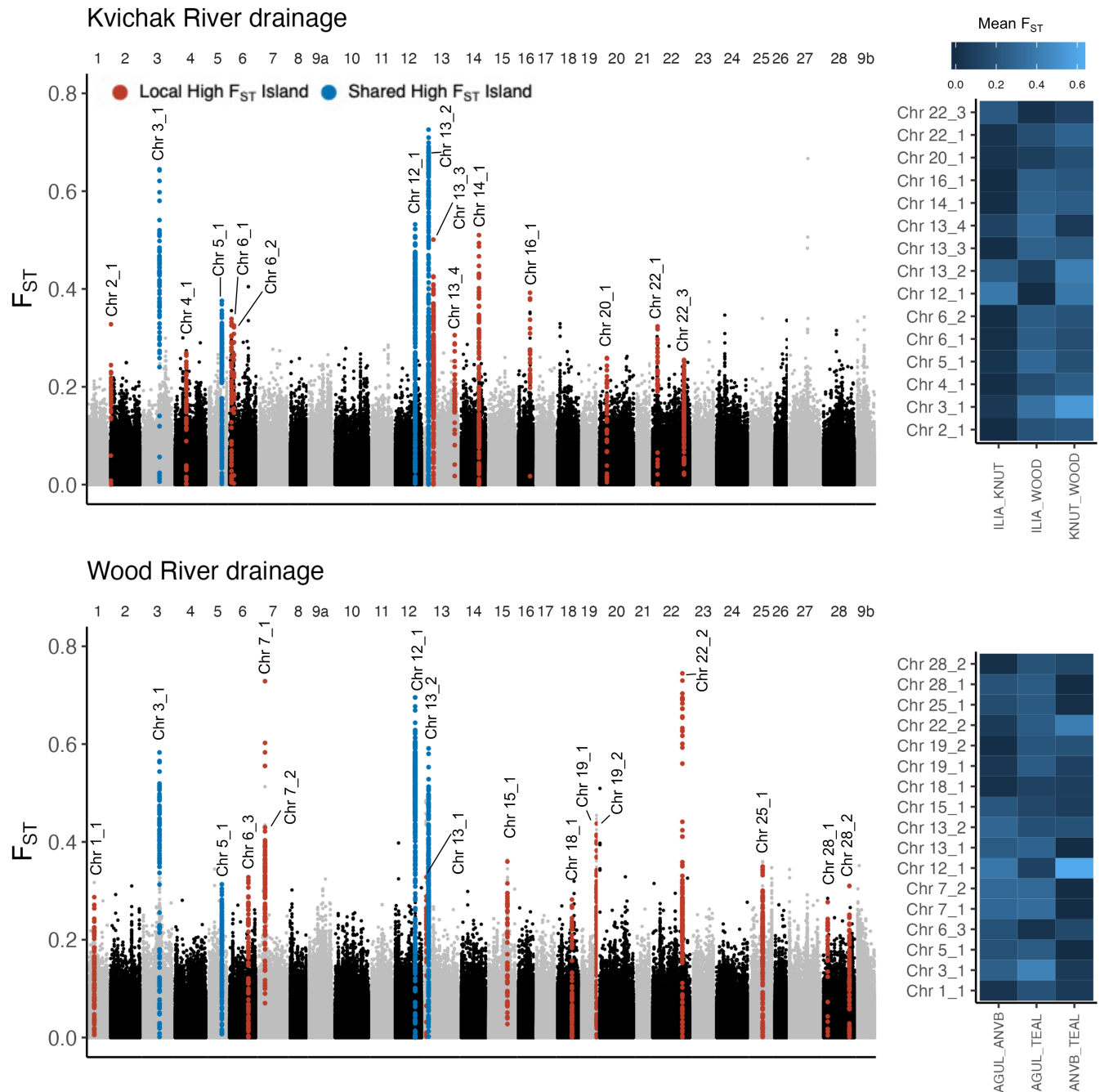


FIGURE 2 Plot of overall Weir and Cockerham F_{ST} (a and c) among calculated among the three spawning sites sampled within each of the Kvichak River (a and b) and Wood River drainages (c and d) and heatmap of pairwise intra-drainage F_{ST} averaged across all loci found within each of the detected islands of divergence (b and d). Each dot in the Manhattan plots represent a single SNP locus, and chromosomes are labelled at the top of each plot. Alternating black and grey colours indicate different chromosomes. Coloured dots indicate SNPs that fell within one of putative islands of divergence identified using HMM analysis. Labels for each coloured putative island of divergence correspond to the y-axis of the heatmap with islands named in sequential order in the 5' direction along the chromosome. The UPNK population was not included in this analysis because it is the only population that we sampled in the Nushagak River drainage. Spawning habitat: KNUT, ANVB=beach; WOOD=island beach; ILIA, AGUL=river; TEAL=creek.

3.3 | Investigating the genomic mechanisms underlying islands of divergence

The islands Chrs 3_1, 5_1, 12_1, 13_2 displayed significantly elevated LD in populations in both drainages compared to the rest of

the genome (Table 2). Patterns of LD among each of the four islands differed substantially (Figure 3). The islands Chrs 3_1, 5_1 and 12_1 contained continuous or almost continuous linkage blocks (i.e. high LD across the entire island) in populations in both drainages. This pattern is suggestive of low recombination commonly observed

TABLE 2 Bootstrapped significance results of LD tests.

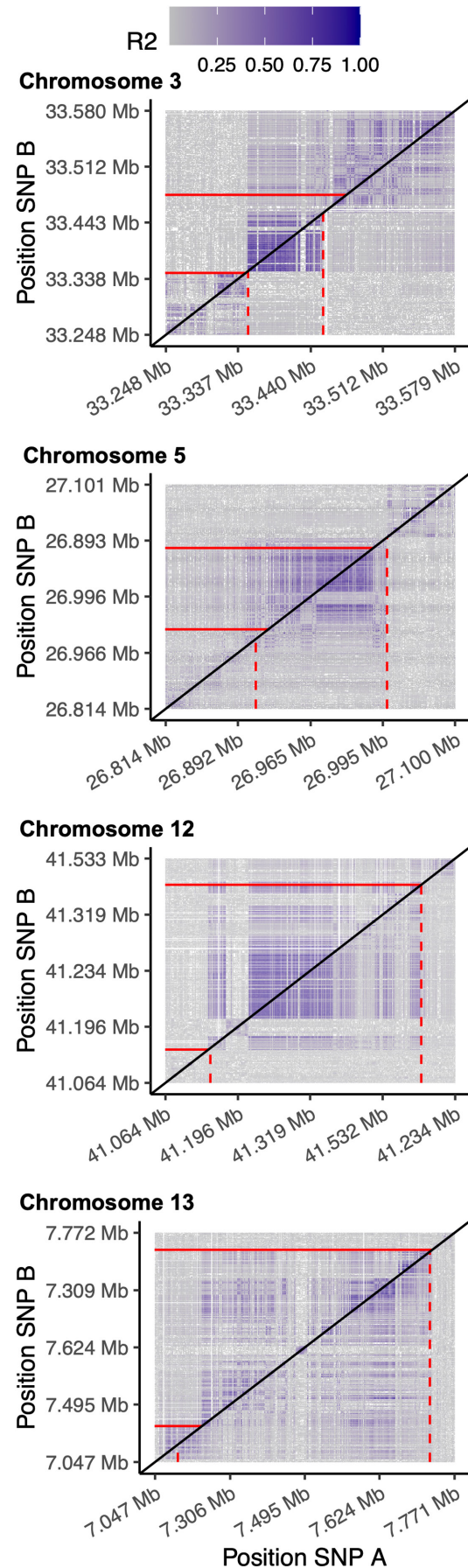
Island	Drainage	Average R^2	Z-score	p-value
Chr 3_1	Kvichak	0.50	6.52	<.0001
Chr 3_1	Wood	0.30	3.97	<.0001
Chr 5_1	Kvichak	0.27	3.50	.0002
Chr 5_1	Wood	0.37	3.58	.0002
Chr 12_1	Kvichak	0.21	3.82	.0001
Chr 12_1	Wood	0.27	5.28	<.0001
Chr 13_2	Kvichak	0.15	3.59	.0002
Chr 13_2	Wood	0.16	3.18	.0007

Note: Bootstrapping was conducted by calculating LD at SNPs from 1000 windows of the same size (in base pairs) and on the same Chr as HMM islands in both the Kvichak and Wood River drainages. The R^2 distribution of the 1000 windows was used to calculate a Z-score to test whether LD among SNPs on the HMM island was higher than expected by chance for the Chr. Average R^2 of SNPs within the HMM windows is reported with the Z-score and the associated one-tailed p-value.

in chromosomal inversions. Furthermore, the edges of the linkage blocks on Chrs 3_1 and 12_1 were abrupt. This was visible due to the clear set of SNPs in consecutive order all in high and equivalent LD with one another in populations in both drainages (Figure 4). The linkage pattern on Chr 5_1 had similarly abrupt edges upstream and downstream of the linkage block in both the Wood River and Kvichak River drainage populations. However, in the Kvichak River drainage populations there was an approximately 20 kb disruption in the centre of the island where SNPs contained lower F_{ST} and linkage when compared to the SNPs on either side of this break. This break in F_{ST} and LD was not present in the Wood River drainage population. The pattern of LD on Chr 13_2 island differed substantially from the other islands. The Chr 13_2 island contained higher-than-average LD compared to the surrounding regions of the Chr, but LD and F_{ST} values were heterogeneous throughout the 526 kb region and there were no abrupt edges of the island. The heterogeneous LD and F_{ST} suggested that this region has likely experienced recombination and is therefore unlikely to be an inversion (Figures 3 and 4).

Individuals clustered into three distinct groups along the first PC when PCA was conducted using SNPs from the islands Chrs 3_1, 5_1 and 12_1 (Figure 5). PC1 explained > 50% of the variance for PCAs of Chrs 3_1, 5_1 and 12_1, providing evidence that the linked SNPs in each island represent most of the genetic variation in each

FIGURE 3 Linkage disequilibrium (LD via R^2) of the islands of divergence shared between the Wood River drainage (above the diagonal) and Kvichak River drainage (below the diagonal) on Chromosome 3 (Chr 3_1), Chromosome 5 (Chr 5_1), Chromosome 12 (Chr 12_1) and Chromosome 13 (Chr 13_2). Darker colours denote higher LD. Red lines denote the boundaries of each island identified using HMM separately for each drainage, with solid red lines indicating boundaries for the Wood River drainage and dotted lines indicating boundaries for the Kvichak River drainage.



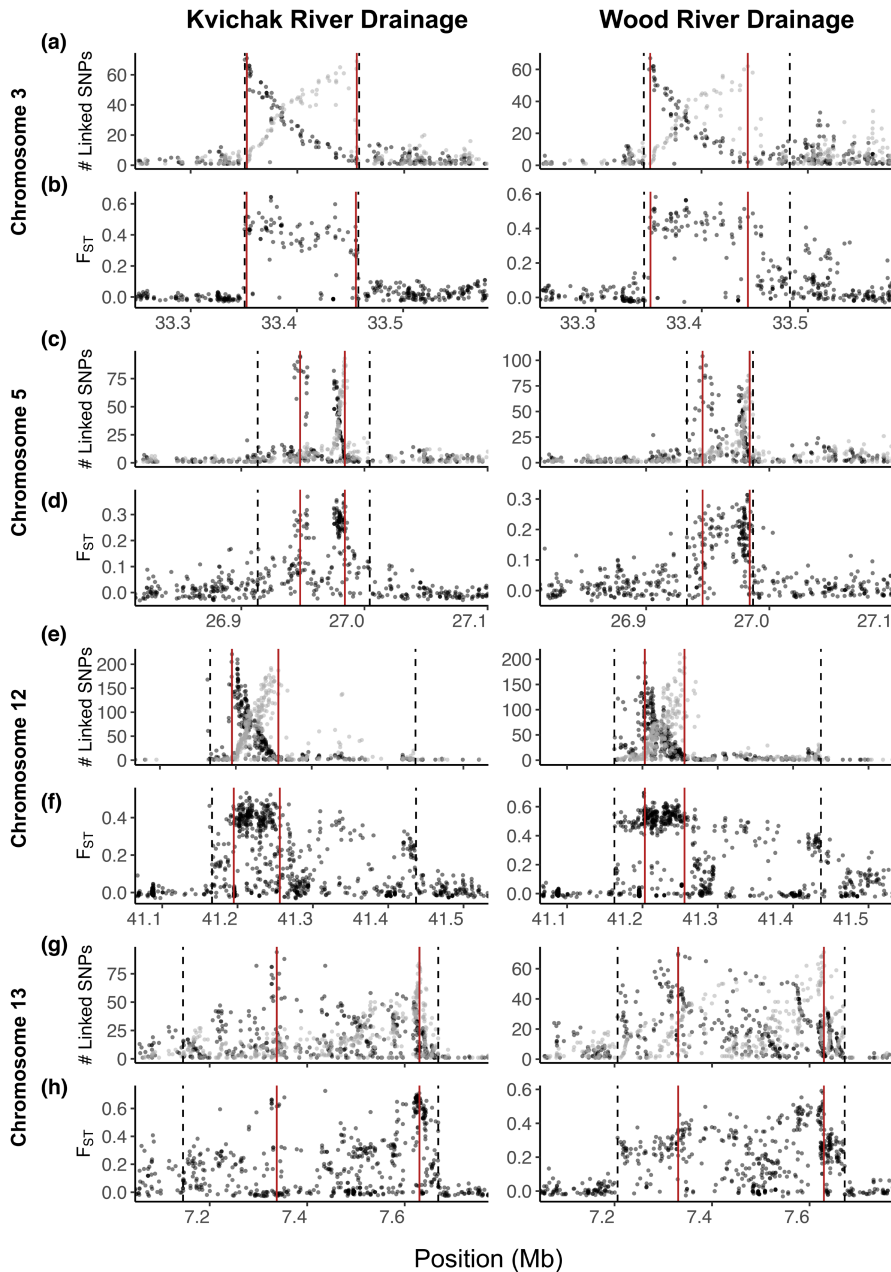


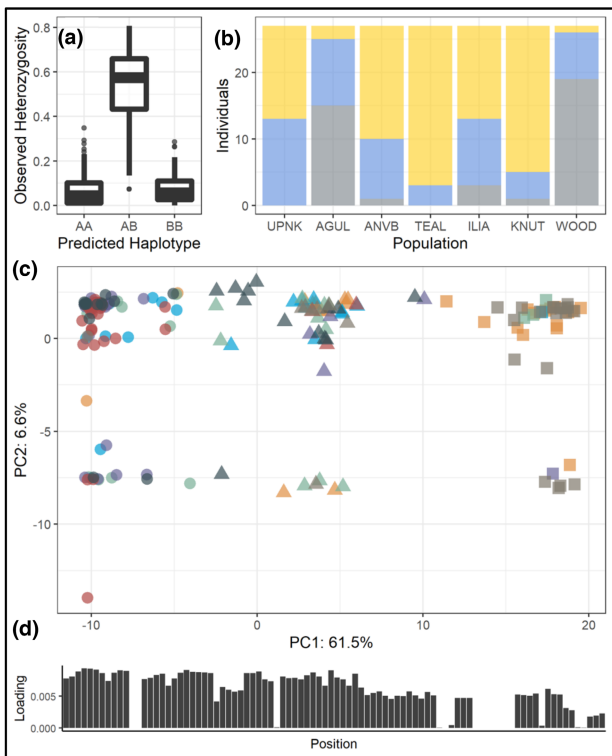
FIGURE 4 Identification of Chr 3_1, 5_1, 12_1 and 13_2 island boundaries (red solid lines) on Chromosome 3 (a and b), Chromosome 5 (c and d), Chromosome 12 (e and f) and Chromosome 13 (g and h) refined from HMM island boundaries (black dotted lines). Plots a, c, e and g, identify the starting point of islands re-defined as the position of the SNP containing strong LD to the largest number of downstream SNPs (black dots), while the ends were defined as the position of the SNP containing strong linkage to the highest number of upstream SNPs (grey dots) for both the Kvichak and Wood River drainages. Plots b, d, f and h show F_{ST} of SNPs on each island of differentiation for both the Kvichak and Wood River drainages.

island, further indicating that they have not been broken up by recombination and may be the result of a chromosomal inversion. All three islands displayed some sub-variation, with between 5 and 10% of variation explained by PC2. Consistent with a two-allele pattern, heterozygosity of the individuals clustered in the middle of the PCA

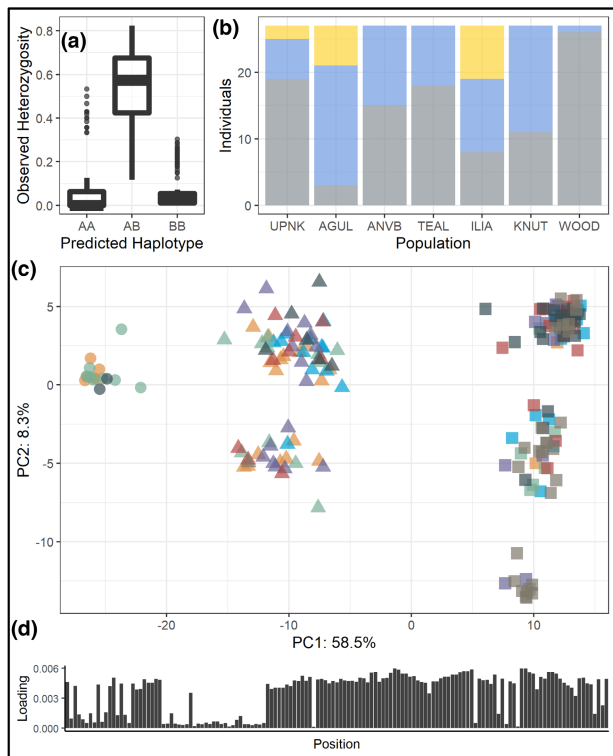
was substantially higher than within the two other groups in islands on Chrs 3_1, 5_1 and 12_1. This pattern is created by individuals in edge groups of the PCA bi-plot being homozygous for alternate arrangements of a chromosomal inversion and the middle group being heterozygous, containing one copy of each arrangement. This

FIGURE 5 Dissection of haplotypes found in the four conserved islands of divergence identified in the Wood and Kvichak River drainages. Panels repeat for islands Chr 3_1, 5_1, 12_1 and 13_2 and share a legend (bottom). (a) Boxplots of observed heterozygosity at loci found within islands grouped by putative genotypes. (b) Genotype frequencies for each predicted genotype in each population. Colours correspond to putative genotypes (see legend). (c) Individual-based PCAs, with colours indicating the sample population and shape indicating the putative genotype as determined by K-means clustering. (d) Plots of the relative DAPC loading scores for each SNP for the first DAPC axis. Higher loadings indicate that a given SNP is contributing more to separation of points along PC1. Chr 3_1, 5_1 and 12_1 display clear patterns of two primary homozygote haplotypes and one heterozygote whereas Chr 13_2 does not. See Figure 6 for more information on the Chromosome 13_2 island. Spawning habitat: KNUT, ANVB=mainland beach; WOOD=island beach; ILIA, AGUL=river; TEAL=creek; UPNK=river.

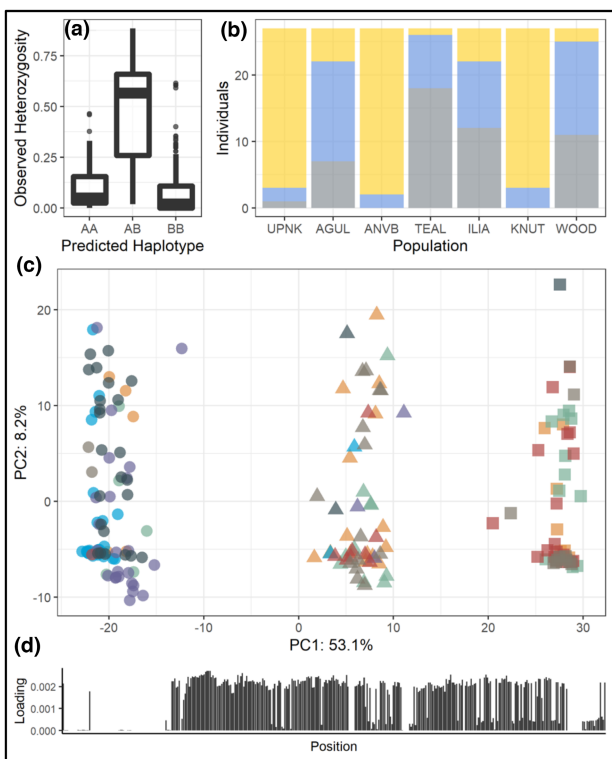
Chromosome 3



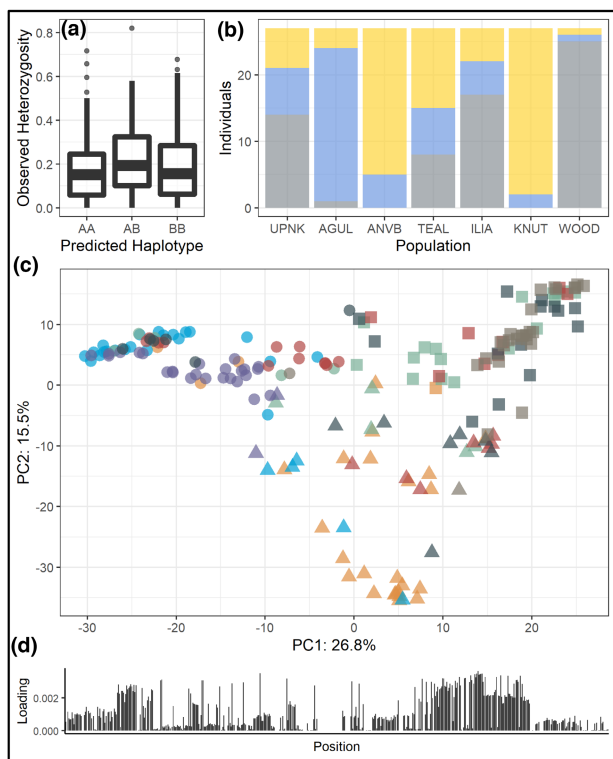
Chromosome 5



Chromosome 12



Chromosome 13



Population ● AGUL ● ILIA ● TEAL ● WOOD ● ANVB ● KNUT ● UPNK

Predicted Haplotype ■ AA ■ AB ■ BB

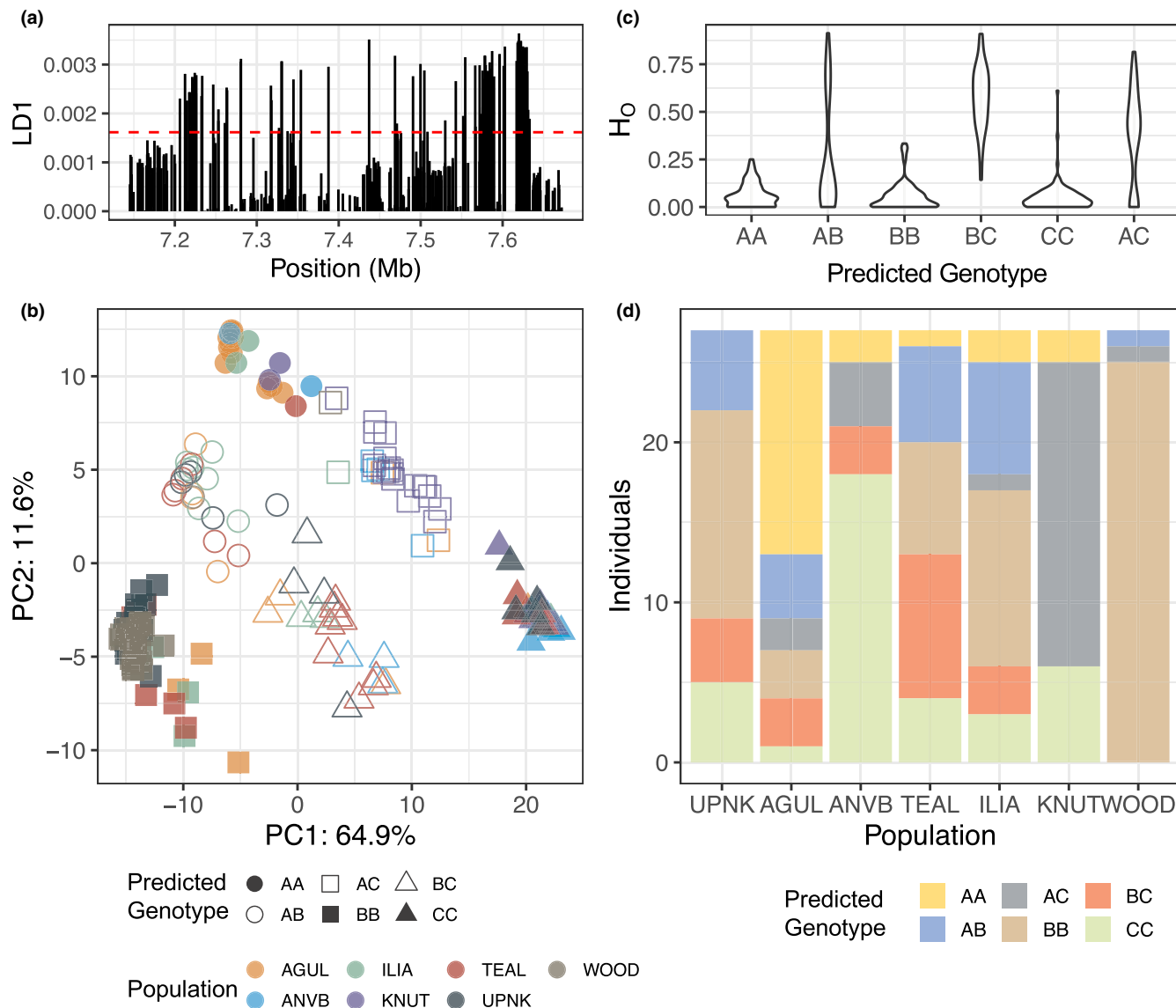


FIGURE 6 In-depth analysis of the Chromosome13 island of divergence Chr 13_2. Chr 13_2 did not fit the typical pattern of two primary haplotypes with a single heterozygotes type found in islands on chromosomes 3, 5 and 12. Instead, it appears to contain three homozygous haplotypes, with all combinations of heterozygotes, for a total of six predicted genotypes. Patterns of linkage are also less clear than for the other islands (see Figure 3). To improve identification of putative genotypes, we focused our analyses on SNPs with loadings on PC1 in the top 25% of the distribution. Panel (a) shows DAPC loading scores for each SNP (x-axis) for the first DAPC axis (LD1). Lines extending above the red line indicate SNPs that contained DAPC loadings in the top 25%. Note the cluster of markers with high loadings between 7.60 and 7.67 Mb. Panel (b) is an individual based PCA, with colours indicating the population and shape indicating the genotype as determined by *K*-means clustering with *K*=6, using the first two PCs and only the high-loading SNPs shown above the red line in Panel (a). Panel (c) visualises observed heterozygosity at loci included in this analysis grouped by putative genotype based on the six clusters that are shown in Panel (b). Panel (d) depicts frequencies of each predicted genotype in each population. Spawning habitat: KNUT, ANVB=mainland beach; WOOD=island beach; ILIA, AGUL=river; TEAL=creek; UPNK=river.

pattern was not observed when SNPs from the island on Chr 13_2 were used to conduct a PCA (Figure 5). Unlike the islands on Chr 3_1, 5_1 and 12_1, PC1 for the Chr 13_2 island explained less total variance (27%), and PC2 had a proportionally larger influence on the dispersion among points (16% of variance).

Because of the strong grouping along PC1, genotypes were assigned for islands on Chrs 3_1, 5_1 and 12_1 using *K*-means clustering (*K*=3). This assigned homozygous and heterozygous genotypes for each individual with alleles named A or B randomly

based on *K*-means designation and the middle genotype in the PCA was assigned to be the heterozygote (i.e. group 1=homozygous genotype AA, group 2=heterozygous genotype AB and group 3=homozygous genotype BB). Genotype frequencies at each island varied substantially among populations (Figure 5b). With the exception of the Chr 12_1 island, genotype frequency did not strongly associate with previously identified selective forces associated with spawning habitat types (i.e., creek/river vs. beach/island beach spawning; Larson et al., 2017). For Chr

3_1, allele A was the most common in all but the Agulowak River (AGUL) and Woody Island Beach (WOOD) populations. For Chr 5_1, allele B was the most common in all but the Agulowak and Iliamna (ILIA) River populations. For Chr 12_1, allele A was most common in populations inhabiting mainland beaches (Anvil Beach [ANVB], Knutson Beach [KNUT]), while allele B was more common in populations inhabiting creeks and rivers. Woody Island Beach population had an allele frequency distribution more similar to creek and river populations than to other beaches. The $K=3$ designation approach could not accurately assign genotypes for the island Chr 13_2, as K -means clustering was unable to detect three clear groupings of individuals (Figure 5).

The inconsistency of Chr 13_2 clustering indicated that a different evolutionary mechanism was responsible for the high linkage and F_{ST} at this island. This difference in structure is illustrated using DAPC loadings, which showed highly heterogeneous levels of explained variance across the genomic region (Figure 5d). To focus analysis on the SNPs responsible for explaining the largest amount of variation at Chr 13_2, variants were filtered to include only the 132 SNPs in the top 25% of loading scores for PC1. When these SNPs were used in a PCA they revealed a clearer six genotype pattern consistent with a triallelic pattern of variation (Figure 6). Variation at the genotypes was clearest in a small region between 7.60 and 7.67 Mb (Figure 6a) that contains highly linked SNPs with the highest F_{ST} values in the island. Heterozygosity at the three putatively heterozygous genotypes was much higher than for putatively homozygous genotypes, indicating that the haplotyping approach (with $K=6$) did identify meaningful variation despite the complicated patterns of LD found in this island (Figure 6). Nevertheless, genotype assignments for the Chr 13_2 region were still somewhat uncertain compared to the other three islands. Using this approach, Alleles B and C were the most common at Chr 13 (Figure 6c). However, the Agulowak River population displayed a high frequency of AA homozygotes. The Upper Nushagak River (UPNK) population, which is part of the putatively ancestral sea/river ecotype, contained at least one copy of all alleles present at the islands we identified, but BB homozygotes were most common.

3.4 | Investigating the evolutionary history of islands of divergence using range-wide data

We successfully merged SNPs called in our dataset with those identified in sockeye salmon collected from five major sockeye populations across their native range from Russia to southern British Columbia (average depth per-variant = 11.1X [SD = 6.79X]; Christensen et al., 2020). These samples included additional samples sequenced by Christensen et al. (2020) from Kvichak and Wood River drainages in Bristol Bay. Following filters for missing data (> 80% genotyping rate), 1,614,900 SNPs remained that were identified in both datasets. This data set included 58 SNPs that mapped to the Chr 3_1 island, 113 SNPs that mapped to the Chr 5_1 island, 146 SNPs that mapped to the Chr 12_1 island and 201 SNPs that mapped

to the Chr 13_2 island. Re-analysis with these SNPs indicated that haploblocks on Chrs 5_1 and 12_1 remained largely intact, with individuals from distant drainages clustering by haplogroup and not geographic location (Figure 7). Interestingly, the area of low LD on Chr 5_1 was also present in some individuals from southern populations, suggesting that variation underlying this small region may be broadly conserved.

The haploblock on Chr 3_1 was less cohesive across the sockeye salmon range than those on Chr 5_1 and 12_1 (Figure 7). More variation was observed in southern populations that was not seen in Bristol Bay populations. This variation may represent divergence of the A allele, while the B allele appears to be well conserved (see discussion for more information). Differentiation among haplotypes at Chr 13_2 was convoluted when examining all SNPs in the island, although the highly diverged region between 7.60 and 7.67 Mb appears conserved across the range, exhibiting the same three haplotypes originally identified in Bristol Bay.

3.5 | Annotation of genes in islands of divergence

Querying of annotations revealed that putative genes are found in all four islands of divergence (Table S5). Two predicted protein coding genes that serve basic functions underlying metabolism and cell function were observed in the Chr 3_1 region: NR4A2 and GPD2. Two genes of broad importance for neurology and cell functions were also identified in the Chr 5_1 region. MGAT5 appears to be important for regulation of the biosynthesis of glycoprotein oligosaccharides that influence cell migration (Marhuenda et al., 2021), and PTH2R is expressed in the brain and pancreas in mammals and involved in pain sensitivity (Bagó et al., 2009). Ten genes and gene-like sequences were observed in the Chr 12_1 region. Genes in this region had a broad range of functions, including liver development (rtn1a, Levi et al., 2009), immune function and environmental stress (Irrc9, Wyżewski et al., 2021; ppm1a), membrane function (pcnx4; dhrr7; ppm1a), protein kinase activity (LOC115138567), DNA binding and eye, limb, and neuronal development (six6; six4; six1), and DNA repair and ion bonding (ppm1a). Most of these genes are well studied in mammalian systems and play important roles in cell function and development. We identified 22 genes in the Chr 13_2 region covering a wide range of functions like cellular transport, transcription and response to DNA damage. Most notably, we documented an oestrogen receptor gene (ESR1) in the highly conserved and highly diverged region of this island (ESR1 at 7.62–7.64 Mb, highly conserved region 7.60–7.67 Mb).

4 | DISCUSSION

Our data confirmed previous findings that sockeye salmon populations in our study region are hierarchically structured, first by river drainage, and then by within-drainage spawning site. Among-drainage population structure reflected genome-wide patterns of

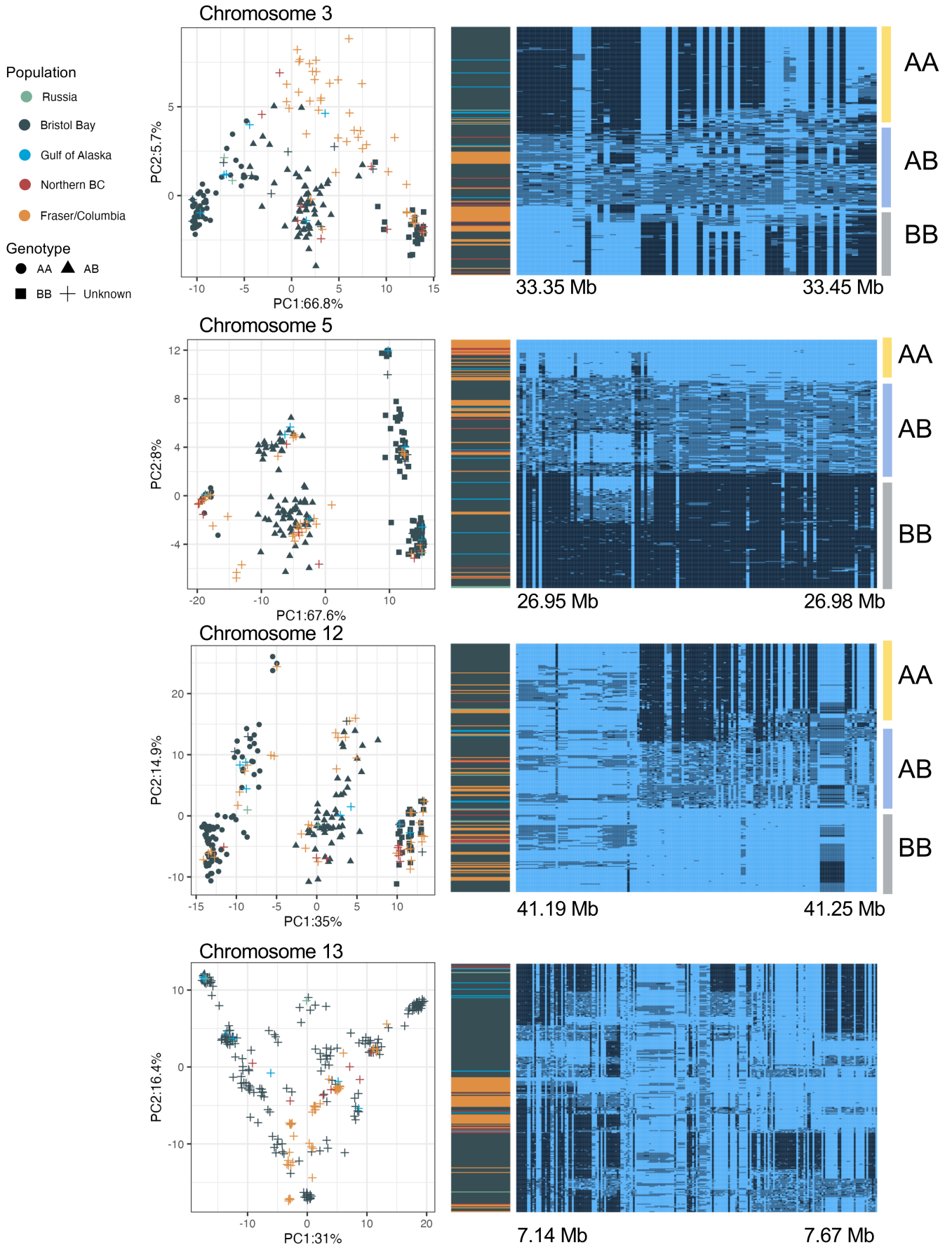


FIGURE 7 PCAs and genotype heatmaps of the four islands constructed using data from individuals sampled across the range of sockeye salmon (Chromosome 3: island Chr 3_1, Chromosome 5: island Chr 5_1, Chromosome 12: island Chr 12_1, Chromosome 13: island Chr 13_2). Data from outside of our main study area (Bristol Bay) are from Christensen et al. (2020). Populations are classified into regional groupings and ordered approximately from north (Russia) to south (Fraser/Columbia). Putative genotypes are denoted for samples sequenced in this study, and range-wide samples are classified as unknowns in PCA biplots (left). Heatmaps (right) display genotype information with individuals represented as rows at SNPs ordered from lowest to highest position in basepairs along the island. The start and ending position of each island is written on the y-axis of the heatmap. Homozygote genotypes AA (i.e. ref allele) are light blue, and alternate homozygote genotypes BB are dark blue. Heterozygote genotype calls (AB) are shown as an intermediate blue colour. Approximate genotype groupings are highlighted along the right y-axis as coloured bars corresponding to the genotype frequency plots colours in Figure 5b. No genotype groupings are shown for Chr 13_2 which has six putative genotypes. Individuals are grouped along the y-axis by putative genotype similarity determined from K-means clustering. The regional group of origin for each sample is colour coded according to the legend.

genetic drift consistent with multiple generations of strong geographical isolation. Meanwhile, within-drainage population structure was dominated by localised islands of divergence that exhibited high allele frequency differences among spawning sites within close physical proximity. Loci on these islands displayed much higher allele frequency differences compared to the rest of the genome, indicating that these loci are likely the result of selection associated with local adaptation to spawning sites. A subset of these islands was conserved in populations across the Wood and Kvichak River drainages. These conserved islands were relatively small (137–527 kb wide) and were found on Chrs 3, 5, 12 and 13.

The mechanisms behind island creation appeared to be a combination of structural variation (likely inversions) and divergence hitchhiking or clustering of loci in low recombination regions. Distinct haplotypes at these four islands could be detected throughout the species range, and all alleles at each island were found in the putatively ancestral sea/river type population from the Nushagak River in Bristol Bay. This pattern suggests that variation at the conserved islands is relatively old, and the islands may contain genes that are important for facilitating adaptive radiation as sockeye salmon colonise new habitats. While it is difficult to pinpoint the direct targets of selection in genomic islands of divergence because they are often composed of large linkage blocks containing many genes (Le Moan et al., 2022; Pampoulie et al., 2022), we identified potentially important genes, including an oestrogen receptor (ESR1) on Chr 13_2 and a gene involved eye development (SIX6) on Chr 12_1 that is associated with life-history variation in other salmon species (Waters et al., 2021; see Tigano & Russello, 2022 for additional discussion of the functional significance of Chr 12).

4.1 | Patterns of adaptive divergence in sockeye salmon

Our study builds on a growing body of research suggesting that adaptive divergence of sockeye salmon is facilitated by conserved islands of divergence (Limborg et al., 2017; Tigano & Russello, 2022; Veale & Russello, 2017a). Previous studies in the Wood and Kvichak river drainages identified similar patterns of differentiation whereby small islands of divergence (including those on Chrs 12_1 and 13_2) had disproportionate influence over intra-drainage genetic structure (Larson et al., 2017,

2019). Additionally, studies focused on more southern populations also found that SNPs in Chr 12_1 were highly differentiated between sockeye salmon spawning in different habitats (Nichols et al., 2016; Tigano & Russello, 2022; Veale & Russello, 2017a). However, nearly all these previous studies used less comprehensive genotyping methods, causing them to fail to detect smaller island such as those like Chrs 3_1 and 5_1, or many of the local islands found only in one of the two drainages and therefore preventing complete characterisation of genomic architecture driving intra-drainage structure.

Our high-resolution data provided evidence that, while some islands are involved in repeated divergence of similar phenotypes (e.g. island Chr 12_1 consistently differentiating beach vs. stream spawners), most of the islands are not associated with repeated differentiation of known phenotypes and are not shared between drainages (at least the two drainages in our system). One particularly interesting example is the Chr 22_2 island, which strongly differentiates the beach and stream populations in the Wood River drainage but does not show a similar pattern in the Kvichak River drainage. Larson et al. (2019) hypothesised that the same loci (in that case loci in islands 12_1, 13_2 and 13_3) are involved in repeated adaptive divergence as sockeye salmon colonise new environments and adapt to a mosaic of selective pressures. Data from the current project provides additional information that can be used to refine this hypothesis. In particular, our data suggest that there are a number (perhaps hundreds) of genomic regions in the sockeye salmon genome that are repeatedly involved in adaptive divergence. These regions often contain putative structural variants and may or may not be under strong selection in a given drainage. Finally, we hypothesise that unique combinations of alleles across these regions have facilitated the formation of thousands of locally adapted populations that have responded to unique selective pressures, creating the massive array of sockeye salmon diversity observed across their range.

4.2 | Identifying potential mechanisms responsible for islands of divergence

Patterns of LD and genetic differentiation in the islands that we identified led us to hypothesise that the islands Chrs 3_1, 5_1 and 12_1 were caused by some sort of structural variation, such as a chromosomal inversion, copy number variant, insertion or deletion, while the island

Chr 13_2 was caused by non-structural mechanisms such as divergence hitchhiking or clustering of loci in low recombination regions. Interestingly, all three putative structural variants that we identified contained sub-variation within structural variant types. This included a region of lower LD within the Chr 5_1 island, potentially indicating that recombination has occurred in this region. The most conspicuous sub-variation occurred on Chr 3_1. When range-wide samples were added, we discovered substantial genetic variation present in the A allele in many individuals from more southern populations. This suggests that a southern subvariant has evolved, potentially during or after the last glacial maxima when southern and northern populations were isolated from each other (Wood et al., 2008). While available evidence still points to the island Chr 3_1 containing an inversion-like structural variant, additional data (e.g., long read data or genome assemblies) would increase support for this hypothesis.

Our observations illustrate that patterns of structural variation can be complex, a result also suggested other empirical (Collins et al., 2017; Matschiner et al., 2022) and simulation (Schaal et al., 2022) studies. In particular, recent results suggest that the barrier to gene flow among inversion types is nuanced, and double recombination and gene conversion events can facilitate exchange of genetic material across types (Korunes & Noor, 2019; Villoutreix et al., 2021). Long read data alone or assembled into genotype-specific reference genomes could help clarify patterns of variation associated with inversions (Kim et al., 2022; Li et al., 2022), although it may not provide conclusive evidence of mechanisms. For example, Tigano and Russello (2022) analysed tributary (creek or river spawning) and beach spawning sockeye salmon from a single lake with short- and long-read data and hypothesised that the Chr 12_1 island was not an inversion. However, the authors also stated that information from additional reference genomes from outside of the sampled range would be useful to clarify this hypothesis. We argue that the conservation of > 200 kb alleles at Chr 12_1 across the species range indicates that structural variation is likely present, but additional long-read data and potentially genome assemblies are necessary for confirmation. The lack of clarity on the mechanism responsible for the Chr 12_1 island (the most conserved in our study) illustrates the importance of obtaining both long- and short-read data from a large portion of a species' range when investigating islands of divergence.

Data from across the species range also helped to identify the small (<10 kb) conserved region and triallelic pattern of variation at the Chr 13_2 island. The lack of fully conserved alleles across the Chr 13_2 island likely represents the legacy of variable selection and recombination across many populations and is consistent with other studies that identify islands not underlain by structural variation (Duranton et al., 2018; Roberts Kingman et al., 2021; Thompson et al., 2020). The triallelic pattern and the variable signals of LD found at the Chr 13_2 island are highly similar to patterns observed at the supergene complex that controls spine and armour plate-related traits found on Chr 7 in stickleback (Roberts Kingman et al., 2021). Roberts Kingman et al. (2021) postulated that high variation and multiple alleles at this supergene may represent a common mechanism that promotes formation of a broad array of

diverse phenotypes. A similar situation could exist in our study system, where adaptive variation at the complex Chr 13_2 island has been utilised to help create the broad array of phenotypic diversity observed in sockeye salmon.

4.3 | What factors influence the size of islands of divergence and the mechanisms underlying them?

Our data indicate that (1) structural variation is not required for the formation of islands in our study system, (2) islands created by different mechanisms (i.e. structural variation vs. other mechanisms) can exist within the same genome and (3) many islands of divergence can be relatively small (<600 kb). Many recent studies investigating the mechanisms that create islands of divergence have focused on large (> 1 Mb) structural polymorphisms, likely because they are easier to identify (reviewed in Wellenreuther & Bernatchez, 2018). However, our data suggest that smaller structural polymorphisms can be adaptively important (also documented in humans; Giner-Delgado et al., 2019), and that structural polymorphisms are not required to create islands.

Simulations suggest that gene flow and selection are the primary forces influencing the size and mechanisms underlying islands of divergence, and that, when gene flow is moderate, structural polymorphisms such as inversions are not required to create islands of divergence (Schaal et al., 2022). Although the interplay between gene flow and selection is complicated, it appears that in general more clustered genetic architectures of adaptation are favoured as either gene flow, strength of selection or both increase (Schaal et al., 2022). Additionally, the genetic basis of a trait under selection appears to influence the genetic architecture of adaptation, with more polygenic traits favouring larger islands often facilitated by inversions and less polygenic traits favouring smaller islands (Schaal et al., 2022). Previous empirical studies also suggest that more clustered architectures of adaptation are favoured as gene flow increases (Shi et al., 2021), with extreme examples of large adaptive inversions in high gene flow marine species (Kirubakaran et al., 2016; Longo et al., 2020). However, it is more difficult to determine how selection influences the size and architecture of islands from empirical studies, where the strength of selection cannot be easily estimated.

Our study system is conceptually similar to two others also exhibiting moderate gene flow (neutral $F_{ST} \approx 0.01$) that have been used to investigate the genetic basis of adaptive divergence: (1) seaweed flies sampled across an environmental gradient (Mérot et al., 2021) and (2) ecotypes of sunflowers (Huang et al., 2020; Todeco et al., 2020). However, these systems are characterised by islands of divergence composed of large inversions and other areas of low recombination spanning megabases, in stark contrast with our system. This contrast highlights the difficulty associated with predicting the genetic architecture of adaptive variation. In fact, previous research has shown that the genetic architecture of adaptive divergence can vary substantially within the same species exposed to the same experimental selection pressures (Atlantic silverside Therkildsen

et al., 2019), or at different areas of their range (steelhead, Pearse et al., 2014, 2019; Weinstein et al., 2019). In silverside, Therikildsen et al. (2019) demonstrated one experimental trial contained a low frequency of an inversion associated with growth and that inversion increased dramatically in frequency in response to experimental selection. In steelhead, Pearse et al. (2014, 2019) documented an extremely large inversion associated with anadromy in the southern portion of the species range whereas Weinstein et al. (2019) found that anadromy was likely controlled by many loci of small effect in the northern range. These studies along with our own illustrate that the interplay between gene flow, selection, and standing genetic variation is extremely complex, making it difficult to predict the genetic architecture of adaptive divergence.

We hypothesise that the substantial differences in the size of islands identified in our study system compared to those mentioned above, despite similar levels of gene flow, may have been caused by differences in the strength of selection and/or the genetic architecture of the traits under selection. Previous studies in salmon have identified small (10s to 100s of kilobases in length) but abundant genomic regions (i.e. 10s of islands all with high F_{ST} across the genome) that control a large proportion of the additive genetic variance for run timing (Prince et al., 2017; Thompson et al., 2020) and age at maturity (Barson et al., 2015). We posit that adaptive variation in our system is encoded by simpler genetic architectures encompassing fewer genes, leading to the smaller islands we observed compared to the islands spanning large regions of chromosomes and containing many genes observed in the studies mentioned above. However, there is still much uncertainty surrounding the interplay between gene flow and selection as well as how these forces influence the characteristics of islands of divergence; this topic should be investigated in future studies.

Our results suggest that islands of divergence in sockeye salmon are the result of more than one mechanism, with three islands created by putative inversions and one likely created through non-structural means. These findings demonstrate that structural variation is not required to create islands of divergence in sockeye salmon. Additionally, the islands that we document were relatively small (< 600kb), indicating that wide islands spanning megabases are not required to facilitate adaptive divergence. Moreover, data collected from across the range of sockeye salmon show that haplotypes at these islands were fully or partially conserved, suggesting this variation is likely old and evolved prior to the colonisation of Bristol Bay. Additionally, we documented putatively important genes within the islands. These results provide strong evidence that the islands of divergence we document were important for facilitating adaptive radiation of sockeye salmon as they colonised new habitats during the Pleistocene epoch.

AUTHOR CONTRIBUTIONS

Wesley A. Larson designed the study, with input from Lisa Seeb and Jim Seeb. Kristen Gruenthal conducted the laboratory work for whole genome sequencing. Peter T. Euclide conducted most of the statistical analyses, with assistance from Yue Shi and Kris A. Christensen, Wesley A. Larson. Peter T. Euclide drafted the manuscript,

with assistance from Yue Shi and Wesley A. Larson. All authors commented on the manuscript and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank the Alaska Salmon Program at the University of Washington, especially Chris Boatright, Jackie Carter, and Daniel Schindler for assisting with sample collection and providing valuable local knowledge. We also thank the Alaska Department of Fish and Game, especially Chris Habicht, Heather Hoyt, and Bill Templin for providing samples. We also thank Sara Schaal for her thoughtful review of the paper and Greg Owens and Peter Sudmant for valuable conversations about this research and the anonymous reviewers who shared many helpful comments.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

FUNDING INFORMATION

This project was partially supported by the Molecular Conservation Genetics Laboratory at the University of Wisconsin Stevens Point, Purdue University, and the University of Washington.

DATA AVAILABILITY STATEMENT

All variant data and associated meta-data are available via Dryad (Access: <https://doi.org/10.5061/dryad.zcrjdfnh5>; Euclide et al., 2023). Sequence alignment files for each individual can be found via the sequence read archive (SRA) on NCBI: (Access: PRJNA1006708).

BENEFIT-SHARING STATEMENT

A broad scale research collaboration was developed with scientists from the Alaska Department of Fish and Game and universities in multiple states, providing genetic samples and bioinformatic and laboratory skills. Many of those collaborators have been included as co-authors. The results of this study are being shared openly with all agencies involved in salmon management and made accessible to the broader scientific community through this publication. Our group is committed to scientific partnerships and to developing a more inclusive and open space for research.

ORCID

Peter T. Euclide  <https://orcid.org/0000-0002-1212-0435>

Wesley A. Larson  <https://orcid.org/0000-0003-4473-3401>

Yue Shi  <https://orcid.org/0000-0002-2860-2666>

Kristen Gruenthal  <https://orcid.org/0000-0002-1985-5071>

Kris A. Christensen  <https://orcid.org/0000-0003-0021-0326>

Lisa Seeb  <https://orcid.org/0000-0002-0626-9743>

REFERENCES

- Ackerman, M. W., Templin, W. D., Seeb, J. E., & Seeb, L. W. (2013). Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. *Conservation Genetics*, 14(2), 483–498. <https://doi.org/10.1007/s10592-012-0401-7>

- Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, 114(27), 7061–7066. <https://doi.org/10.1073/pnas.1616755114>
- Bagó, A. G., Dimitrov, E., Saunders, R., Seress, L., Palkovits, M., Usdin, T. B., & Dobolyi, A. (2009). Parathyroid hormone 2 receptor and its endogenous ligand tuberoinfundibular peptide of 39 residues are concentrated in endocrine, viscerosensory and auditory brain regions in macaque and human. *Neuroscience*, 162(1), 128–147. <https://doi.org/10.1016/j.neuroscience.2009.04.054>
- 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., Særgrov, 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>, <http://www.nature.com/nature/journal/v528/n7582/abs/nature16062.html#supplementary-information>
- Baym, M., Kryazhinskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. (2015). Inexpensive multiplexed library preparation for Megabase-sized genomes. *PLoS One*, 10(5), e0128036. <https://doi.org/10.1371/journal.pone.0128036>
- Beacham, T. D., Lapointe, M., Candy, J. R., McIntosh, B., MacConnachie, C., Tabata, A., Kaukinen, K., Deng, L., Miller, K. M., & Withler, R. E. (2004). Stock identification of Fraser River sockeye salmon using microsatellites and major histocompatibility complex variation. *Transactions of the American Fisheries Society*, 133(5), 1117–1137. <https://doi.org/10.1577/t04-001.1>
- Benjelloun, B., Boyer, F., Streeter, I., Zamani, W., Engelen, S., Alberti, A., Alberto, F. J., BenBati, M., Ibelbachyr, M., Chentouf, M., Bechchari, A., Rezaei, H. R., Naderi, S., Stella, A., Chikhi, A., Clarke, L., Kijas, J., Flicek, P., Taberlet, P., & Pompanon, F. (2019). An evaluation of sequencing coverage and genotyping strategies to assess neutral and adaptive diversity. *Molecular Ecology Resources*, 19(6), 1497–1515. <https://doi.org/10.1111/1755-0998.13070>
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., Maes, G. E., Martinsohn, J. T., Miller, K. M., Naish, K. A., Ovenden, J. R., Primmer, C. R., Young Suk, H., Therkildsen, N. O., & Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology and Evolution*, 32(9), 665–680. <https://doi.org/10.1016/j.tree.2017.06.010>
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, 81(5), 1084–1097. <https://doi.org/10.1086/521987>
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4), 855–867. <https://doi.org/10.1111/1755-0998.12357>
- Christensen, K. A., Rondeau, E. B., Minkley, D. R., Sakhrani, D., Biagi, C. A., Flores, A.-M., Withler, R. E., Pavey, S. A., Beacham, T. D., Godin, T., Taylor, E. B., Russello, M. A., Devlin, R. H., & Koop, B. F. (2020). The sockeye salmon genome, transcriptome, and analyses identifying population defining regions of the genome. *PLoS One*, 15(10), e0240935. <https://doi.org/10.1371/journal.pone.0240935>
- Collins, R. L., Brand, H., Redin, C. E., Hanscom, C., Antolik, C., Stone, M. R., Glessner, J. T., Mason, T., Pregno, G., Dorrani, N., Mandrile, G., Giachino, D., Perrin, D., Walsh, C., Cipicchio, M., Costello, M., Stortchevoi, A., An, J. Y., Currall, B. B., ... Talkowski, M. E. (2017). Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome Biology*, 18(1), 36. <https://doi.org/10.1186/s13059-017-1158-6>
- Creelman, E. K., Hauser, L., Simmons, R. K., Templin, W. D., & Seeb, L. W. (2011). Temporal and geographic genetic divergence: Characterizing sockeye salmon populations in the Chignik watershed, Alaska, using single-nucleotide polymorphisms. *Transactions of the American Fisheries Society*, 140(3), 749–762. <https://doi.org/10.1080/00028487.2011.584494>
- 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. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernysky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498. <https://doi.org/10.1038/ng.806>
- Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nature Communications*, 9(1), 2518. <https://doi.org/10.1038/s41467-018-04963-6>
- Euclide, P. T., Larson, W. A., Shi, Y., Gruenthal, K., Christensen, K., Seeb, J., & Seeb, L. (2023). Data from: Conserved islands of divergence associated with adaptive variation in sockeye salmon are maintained by multiple mechanisms. *Dryad* <https://doi.org/10.5061/dryad.zcrjdfnh5>
- Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology & Evolution*, 34(3), 239–248. <https://doi.org/10.1016/j.tree.2018.12.005>
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7), 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Giner-Delgado, C., Villatoro, S., Lerga-Jaso, J., Gayà-Vidal, M., Oliva, M., Castellano, D., Pantano, L., Bitarello, B. D., Izquierdo, D., Noguera, I., Olalde, I., Delprat, A., Blancher, A., Lalueza-Fox, C., Esko, T., O'Reilly, P. F., Andrés, A. M., Ferretti, L., Puig, M., & Cáceres, M. (2019). Evolutionary and functional impact of common polymorphic inversions in the human genome. *Nature Communications*, 10(1), 4222. <https://doi.org/10.1038/s41467-019-12173-x>
- Habicht, C., Seeb, L. W., & Seeb, J. E. (2007). Genetic and ecological divergence defines population structure of sockeye salmon populations returning to Bristol Bay, Alaska, and provides a tool for admixture analysis. *Transactions of the American Fisheries Society*, 136(1), 82–94. <https://doi.org/10.1577/t06-001.1>
- Hager, E. R., Harringmeyer, O. S., Wooldridge, T. B., Theingi, S., Gable, J. T., McFadden, S., Neugeboren, B., Turner, K. M., Jensen, J. D., & Hoekstra, H. E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*, 377(6604), 399–405. <https://doi.org/10.1126/science.abg0718>
- Hemstrom, W., & Jones, M. (2023). snpR: User friendly population genomics for SNP data sets with categorical metadata. *Molecular Ecology Resources*, 23, 962–973. <https://doi.org/10.1111/1755-0998.13721>
- Hofer, T., Foll, M., & Excoffier, L. (2012). Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics*, 13(1), 107. <https://doi.org/10.1186/1471-2164-13-107>
- Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, 29(14), 2535–2549. <https://doi.org/10.1111/mec.15428>
- Huang, K., & Rieseberg, L. H. (2020). Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science*, 11, 296. <https://doi.org/10.3389/fpls.2020.00296>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>

- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Kim, K.-W., De-Kayne, R., Gordon, I. J., Omufwoko, K. S., Martins, D. J., ffrench-Constant, R., & Martin, S. H. (2022). Stepwise evolution of a butterfly supergene via duplication and inversion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1856), 20210207. <https://doi.org/10.1098/rstb.2021.0207>
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), 419–434. <https://doi.org/10.1534/genetics.105.047985>
- Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., de Rosa, M. C., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S., & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, 25(10), 2130–2143. <https://doi.org/10.1111/mec.13592>
- Korunes, K. L., & Noor, M. A. F. (2019). Pervasive gene conversion in chromosomal inversion heterozygotes. *Molecular Ecology*, 28(6), 1302–1315. <https://doi.org/10.1111/mec.14921>
- Larson, W. A., Dann, T. H., Limborg, M. T., McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2019). Parallel signatures of selection at genomic islands of divergence and the major histocompatibility complex in ecotypes of sockeye salmon across Alaska. *Molecular Ecology*, 28(9), 2254–2271. <https://doi.org/10.1111/mec.15082>
- Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E., & Seeb, L. W. (2017). Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular Ecology*, 26(2), 554–570. <https://doi.org/10.1111/mec.13933>
- Larson, W. A., Seeb, J. E., Dann, T. H., Schindler, D. E., & Seeb, L. W. (2014). Signals of heterogeneous selection at an MHC locus in geographically proximate ecotypes of sockeye salmon. *Molecular Ecology*, 23(22), 5448–5461. <https://doi.org/10.1111/mec.12949>
- Le Moan, A., Panova, M., De Jode, A., Ortega-Martinez, O., Duvetorp, M., Faria, R., Butlin, R., & Johannesson, K. (2022). An allozyme polymorphism is associated with a large chromosomal inversion in the marine snail *Littorina fabalis*. *Evolutionary Applications*, 16, 279–292. <https://doi.org/10.1111/eva.13427>
- Levi, L., Pekarski, I., Gutman, E., Fortina, P., Hyslop, T., Biran, J., Levavi-Sivan, B., & Lubzens, E. (2009). Revealing genes associated with vitellogenesis in the liver of the zebrafish (*Danio rerio*) by transcriptome profiling. *BMC Genomics*, 10(1), 141. <https://doi.org/10.1186/1471-2164-10-141>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Preprint arXiv:1303.3997*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Li, Q., Lindtke, D., Rodríguez-Ramírez, C., Kakioka, R., Takahashi, H., Toyoda, A., Kitano, J., Ehrlich, R. L., Chang Mell, J., & Yeaman, S. (2022). Local adaptation and the evolution of genome architecture in threespine stickleback. *Genome Biology and Evolution*, 14(6), evac075. <https://doi.org/10.1093/gbe/evac075>
- Limborg, M. T., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). Screening of duplicated loci reveals hidden divergence patterns in a complex salmonid genome. *Molecular Ecology*, 26(17), 4509–4522. <https://doi.org/10.1111/mec.14201>
- Longo, G. C., Lam, L., Basnett, B., Samhuri, J., Hamilton, S., Andrews, K., Williams, G., Goetz, G., McClure, M., & Nichols, K. M. (2020). Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small portion of the genome. *Evolutionary Applications*, 13(10), 2536–2554. <https://doi.org/10.1111/eva.13037>
- Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H., Abbott, R. J., Lascoux, M., Nevo, E., & Liu, J. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences*, 115(2), E236–E243. <https://doi.org/10.1073/pnas.1713288114>
- Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biology*, 20(1), 1–14. <https://doi.org/10.1186/S13059-019-1828-7/TABLES/2>
- Marhuenda, E., Fabre, C., Zhang, C., Martin-Fernandez, M., Iskratsch, T., Saleh, A., Bauchet, L., Cambedouze, J., Hugnot, J.-P., Duffau, H., Dennis, J. W., Cornu, D., & Bakalara, N. (2021). Glioma stem cells invasive phenotype at optimal stiffness is driven by MGAT5 dependent mechanosensing. *Journal of Experimental & Clinical Cancer Research*, 40(1), 139. <https://doi.org/10.1186/s13046-021-01925-7>
- Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics*, 12(2), e1005887. <https://doi.org/10.1371/journal.pgen.1005887>
- Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Briec, M. S. O., Pampoulie, C., Bradbury, I., Jakobsen, K. S., & Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology & Evolution*, 6(4), 469–481. <https://doi.org/10.1038/s41559-022-01661-x>
- McClelland, E. K., Ming, T. J., Tabata, A., Kaukinen, K. H., Beacham, T. D., Withler, R. E., & Miller, K. M. (2013). Patterns of selection and allele diversity of class I and class II major histocompatibility loci across the species range of sockeye salmon (*Oncorhynchus nerka*). *Molecular Ecology*, 22(18), 4783–4800. <https://doi.org/10.1111/mec.12424>
- McGlauffin, M. T., Schindler, D. E., Seeb, L. W., Smith, C. T., Habicht, C., & Seeb, J. E. (2011). Spawning habitat and geography influence population structure and juvenile migration timing of sockeye salmon in the Wood River lakes, Alaska. *Transactions of the American Fisheries Society*, 140(3), 763–782. <https://doi.org/10.1080/00028487.2011.584495>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mérot, C., Berdan, E. L., Cayuela, H., Djambazian, H., Ferchaud, A.-L., Laporte, M., Normandeau, E., Ragoussis, J., Wellenreuther, M., & Bernatchez, L. (2021). Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Molecular Biology and Evolution*, 38(9), 3953–3971. <https://doi.org/10.1093/molbev/msab143>
- Mérot, C., Stenløkk, K. S. R., Venney, C., Laporte, M., Moser, M., Normandeau, E., Árnýasi, M., Kent, M., Rougeux, C., Flynn, J. M., Lien, S., & Bernatchez, L. (2023). Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (*Coregonus* sp.) with long and short reads. *Molecular Ecology*, 32(6), 1458–1477. <https://doi.org/10.1111/mec.16468>
- Miller, K. M., Kaukinen, K. H., Beacham, T. D., & Withler, R. E. (2001). Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica*, 111(1–3), 237–257.
- Nichols, K. M., Kozfkay, C. C., & Narum, S. R. (2016). Genomic signatures among *Oncorhynchus nerka* ecotypes to inform conservation and management of endangered sockeye Salmon. *Evolutionary Applications*, 9(10), 1285–1300. <https://doi.org/10.1111/eva.12412>
- Pampoulie, C., Berg, P. R., & Jentoft, S. (2022). Hidden but revealed: After years of genetic studies behavioural monitoring combined with genomics uncover new insight into the population dynamics of Atlantic cod in Icelandic waters. *Evolutionary Applications*, 16, 223–233. <https://doi.org/10.1111/eva.13471>

- Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., Anderson, E. C., Rundio, D. E., Williams, T. H., Naish, K. A., Moen, T., Liu, S., Kent, M., Moser, M., Minkley, D. R., Rondeau, E. B., Briec, M. S. O., Sandve, S. R., Miller, M. R., ... 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>
- Pearse, D. E., Miller, M. R., Abadía-Cardoso, A., & Garza, J. C. (2014). Rapid parallel evolution of standing variation in a single, complex genomic region is associated with life history in steelhead/rainbow trout. *Proceedings of the Royal Society B: Biological Sciences*, 281(1783), 20140012. <https://doi.org/10.1098/RSPB.2014.0012>
- Peterson, D. A., Hilborn, R., & Hauser, L. (2014). Local adaptation limits lifetime reproductive success of dispersers in a wild salmon metapopulation. *Nature Communications*, 5, 3696. <https://doi.org/10.1038/ncomms4696>
- Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., Hotaling, T. J., Spidle, A. P., & Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, 3(8), e1603198. <https://doi.org/10.1126/sciadv.1603198>
- Privé, F., Luu, K., Vilhjálmsson, B. J., Blum, M. G. B., & Rosenberg, M. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution*, 37(7), 2153–2154. <https://doi.org/10.1093/molbev/msaa053>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & 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.
- Quinn, T. P. (2005). *The behavior and ecology of Pacific salmon and trout*. University of Washington Press.
- Quinn, T. P., Hendry, A. P., & Wetzel, L. A. (1995). The influence of life history trade-offs and the size of incubation gravels on egg size variation in sockeye Salmon (*Oncorhynchus nerka*). *Oikos*, 74(3), 425. <https://doi.org/10.2307/3545987>
- Quinn, T. P., Wetzel, L., Bishop, S., Overberg, K., & Rogers, D. E. (2001). Influence of breeding habitat on bear predation and age at maturity and sexual dimorphism of sockeye salmon populations. *Canadian Journal of Zoology*, 79(10), 1782–1793. <https://doi.org/10.1139/cjz-79-10-1782>
- Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16(7), 351–358. [https://doi.org/10.1016/s0169-5347\(01\)02187-5](https://doi.org/10.1016/s0169-5347(01)02187-5)
- Roberts Kingman, G. A., Lee, D., Jones, F. C., Desmet, D., Bell, M. A., & Kingsley, D. M. (2021). Longer or shorter spines: Reciprocal trait evolution in stickleback via triallelic regulatory changes in *Stanniocalcin2a*. *Proceedings of the National Academy of Sciences*, 118(31), e2100694118. <https://doi.org/10.1073/pnas.2100694118>
- Russello, M. A., Kirk, S. L., Frazer, K. K., & Askey, P. J. (2012). Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, 5(1), 39–52. <https://doi.org/10.1111/j.1752-4571.2011.00206.x>
- Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, 26(17), 4378–4390. <https://doi.org/10.1111/mec.14226>
- Schaal, S. M., Haller, B. C., & Lotterhos, K. E. (2022). Inversion invasions: When the genetic basis of local adaptation is concentrated within inversions in the face of gene flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1856), 20210200. <https://doi.org/10.1098/rstb.2021.0200>
- Shi, Y., Bouska, K. L., McKinney, G. J., Dokai, W., Bartels, A., McPhee, M. V., & Larson, W. A. (2021). Gene flow influences the genomic architecture of local adaptation in six riverine fish species. *Molecular Ecology*, 32, 1549–1566. <https://doi.org/10.1111/mec.16317>
- Stenlökk, K., Saitou, M., Rud-Johansen, L., Nome, T., Moser, M., Árnýasi, M., Kent, M., Barson, N. J., & Lien, S. (2022). The emergence of supergenes from inversions in Atlantic salmon. *Philosophical Transactions of the Royal Society B*, 377(1856), 20210195.
- Stewart, I. J., Quinn, T. P., & Bentzen, P. (2003). Evidence for fine-scale natal homing among Island beach spawning sockeye salmon, *Oncorhynchus nerka*. *Environmental Biology of Fishes*, 67(1), 77–85. <https://doi.org/10.1023/a:1024436632183>
- Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Molecular Ecology Resources*, 17(2), 194–208. <https://doi.org/10.1111/1755-0998.12593>
- Therkildsen, N. O., Wilder, A. P., Conover, D. O., Munch, S. B., Baumann, H., & Palumbi, S. R. (2019). Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science*, 365(6452), 487–490. https://doi.org/10.1126/SCIENCE.AAW7271/SUPPL_FILE/AAW7271_THERKILDSEN_SM.PDF
- Thioulouse, J., Dray, S., Dufour, A.-B., Siberchicot, A., Jombart, T., & Pavoine, S. (2018). Multivariate analysis of ecological data with ade4.
- Thompson, N. F., Anderson, E. C., Clemento, A. J., Campbell, M. A., Pearse, D. E., Harsey, J. W., Kinziger, A. P., & 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>
- Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, 25(10), 2144–2164. <https://doi.org/10.1111/mec.13606>
- Tigano, A., & Russello, M. A. (2022). The genomic basis of reproductive and migratory behaviour in a polymorphic salmonid. *Molecular Ecology*, 31(24), 6588–6604. <https://doi.org/10.1111/mec.16724>
- Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., ... Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 584(7822), 602–607. <https://doi.org/10.1038/s41586-020-2467-6>
- Veale, A. J., & Russello, M. A. (2017a). An ancient selective sweep linked to reproductive life history evolution in sockeye salmon. *Scientific Reports*, 7(1), 1747. <https://doi.org/10.1038/s41598-017-01890-2>
- Veale, A. J., & Russello, M. A. (2017b). Genomic changes associated with reproductive and migratory ecotypes in sockeye salmon (*Oncorhynchus nerka*). *Genome Biology and Evolution*, 9(10), 2921–2939. <https://doi.org/10.1093/gbe/evx215>
- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 367(1587), 451–460. <https://doi.org/10.1098/rstb.2011.0260>
- Villoutreix, R., Ayala, D., Joron, M., Gompert, Z., Feder, J. L., & Nosil, P. (2021). Inversion breakpoints and the evolution of supergenes. *Molecular Ecology*, 30(12), 2738–2755. <https://doi.org/10.1111/mec.15907>
- Wang, L., Liu, S., Yang, Y., Meng, Z., & Zhuang, Z. (2022). Linked selection, differential introgression and recombination rate variation promote heterogeneous divergence in a pair of yellow croakers. *Molecular Ecology*, 31(22), 5729–5744. <https://doi.org/10.1111/mec.16693>
- 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>

- Weinstein, S. Y., Thrower, F. P., Nichols, K. M., & Hale, M. C. (2019). A large-scale chromosomal inversion is not associated with life history development in rainbow trout from Southeast Alaska. *PLoS One*, 14(9), e0223018. <https://doi.org/10.1371/JOURNAL.PONE.0223018>
- Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, 33(6), 427–440. <https://doi.org/10.1016/j.tree.2018.04.002>
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Wood, C. C., Bickham, J. W., Nelson, R. J., Foote, C. J., & Patton, J. C. (2008). Recurrent evolution of life history ecotypes in sockeye salmon: Implications for conservation and future evolution. *Evolutionary Applications*, 1(2), 207–221. <https://doi.org/10.1111/j.1752-4571.2008.00028.x>
- Wyżewski, Z., Gradowski, M., Krysińska, M., Dudkiewicz, M., & Pawłowski, K. (2021). A novel predicted ADP-ribosyltransferase-like family conserved in eukaryotic evolution. *PeerJ*, 9, e11051. <https://doi.org/10.7717/peerj.11051>
- Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, 110(19), 1743–1751.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Euclide, P. T., Larson, W. A., Shi, Y., Gruenthal, K., Christensen, K. A., Seeb, J., & Seeb, L. (2023). Conserved islands of divergence associated with adaptive variation in sockeye salmon are maintained by multiple mechanisms. *Molecular Ecology*, 00, 1–21. <https://doi.org/10.1111/mec.17126>