Revealing the evolutionary history and contemporary population structure of Pacific salmon in the Fraser River through genome resequencing

Kris A. Christensen[1*], Anne-Marie Flores[1], Dionne Sakhrani[2], Carlo A. Biagi[2], Robert H. Devlin[2], Ben J. G. Sutherland[3,4], Ruth E. Withler[5], Eric B. Rondeau[6], Ben F. Koop[1]

1. Department of Biology, University of Victoria, Victoria, BC, Canada, V8W 2Y2
2. Fisheries and Oceans Canada, West Vancouver, BC, Canada, V7V 1H2
3. Sutherland Bioinformatics, Lantzville, BC, Canada V0R 2H0
4. Faculty of Science and Technology, Vancouver Island University, Nanaimo, BC, Canada V9R 5S5
5. Pacific Salmon Foundation, Vancouver, BC, Canada V6H 3V9
6. Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC, Canada V9T 6N7

*Corresponding author
Email: kris.christensen@wsu.edu (KC)

Running title: Genetics of Pacific salmon from the Fraser River

Keywords: Chinook, coho, sockeye, genetics, demography, glacial refugia

Abstract

The Fraser River once supported massive salmon returns, but now years with half of the recorded historical maximum are considered good. There is substantial interest from surrounding communities, governments, and other groups to increase salmon returns for both human use and for functional ecosystems. To help generate resources for this endeavour, we resequenced hundreds of genomes at moderate coverage (~16x) of Chinook (*Oncorhynchus tshawytscha*), coho (*O. kisutch*), and sockeye salmon (*O. nerka*) from the Fraser River. The resequenced genomes are an important resource that can give us new insights. In this study, we found evidence that Chinook salmon have 1.5-2x more polymorphic loci than coho or sockeye salmon. Using principal component analysis (PCA) and admixture analysis, we also identified genetic groups similar to those previously identified with only a few microsatellite markers. As the higher density data supports these previous genetic groups, it suggests that the identity of these groups is not overly sensitive to the number of genetic markers or when the groups were sampled. With the increased resolution from resequenced genomes, we were able to further identify factors influencing these genetic groups, including isolation-by-distance, migration barriers, recolonization from different glacial refugia, and environmental factors like precipitation. We were also able to identify 20 potentially adaptive loci among the genetic groups by analyzing runs of homozygosity. All of the resequenced genomes have been submitted to a public database where they can be used as a reference for the contemporary genomics of Fraser River salmon.

Article Summary

Concerns over Fraser River salmon declines led us to generate resources to better understand the genetics of these salmon. Our findings were similar to previous studies that examined a very small fraction of the genetic markers examined here. We expanded upon previous studies by identifying possible influences on the genetics of Fraser River salmon. These included environmental factors, historical differences, and a lack of connectivity in the river caused by the Fraser Canyon.

Introduction

The Fraser River in Canada may have existed in some form for over 66 million years (Tribe 2005). The modern flow of the river is much more recent as there is evidence of a major change in flow around 760,000 years ago (Andrews *et al.* 2012). During the last glaciation period (the last in the Pleistocene), the entire region where the modern river is located may have been covered by the Cordilleran Ice Sheet, which reached its estimated maximum around 19,000 years ago (Clark *et al.* 2009). The retreat of the Cordilleran Ice Sheet began as early as 18,000 years ago and there were ice-free regions in British Columbia (where the Fraser River is located) as early as 17,700 years ago (Warner *et al.* 1982; Darvill *et al.* 2018). Some regions of British Columbia may have even been ice-free during the glacial maximum (e.g., (Byun *et al.* 1999; McPhail 2007; Stewart *et al.* 2009)). The Cordilleran Ice Sheet fully retreated by around 11,000 years ago, and its only remnants are the glaciers that still exist in modern times (Clague 2017). These massive changes in the Fraser River environment likely caused local extinctions of plants and animals from the river.

Based on fossil evidence, there were salmonids in the Fraser River before it was covered by glaciers (Harington 1996). From paleogeographic evidence mentioned above and this fossil evidence, we can infer that modern salmon in the river must have re-colonized after glaciers receded. There is

similar evidence of colonization of other species to this region, such as different plants, deer, and other ray-finned fishes (Soltis *et al.* 1997; Bernatchez and Wilson 1998; Hewitt 2000; McPhail 2007; Latch *et al.* 2009; Beatty and Provan 2010). Colonization or re-colonization of the 65-66 species of ray-finned fishes native to British Columbia was thought to have started around 13,000 years ago (McPhail 2007). However, fossil evidence indicates that kokanee (*O. nerka*) could have re-colonized the interior of British Columbia, possibly via the Columbia River, as early as 18,000 years ago (Harington 1996). This is consistent with a large isolated and unique kokanee population in the upper Columbia River and Fraser River system that remains today (Beacham and Withler 2017; Christensen *et al.* 2020).

While there is a large body of paleogeographic research and a fossil record describing glaciation and species colonization of the Fraser River after glaciers retreated, we often do not have information on specific locations or populations. We can turn to genetic studies to complement what we understand from these other two fields of study. We expect to observe genetic legacies of re-colonization events in modern populations (Hewitt 1996, 2000). These legacies may influence modern populations, and understanding them will be important for evidence-based conservation and management. Several researchers have suggested that Fraser River salmon have experienced recent genetic bottlenecks that could be the outcome of re-colonization events (Wehrhahn and Powell 1987; Wood *et al.* 1994; Rondeau *et al.* 2023).

Generally, there were two to three Fraser River (e.g., lower, middle, and upper) Chinook (*O. tshawytscha*), sockeye (*O. nerka*), or coho salmon (*O. kisutch*) genetic groups identified from previous studies (Wood *et al.* 1994; Small *et al.* 1998; Teel *et al.* 2000; Beacham *et al.* 2003; Beacham and Withler 2017). This excludes additional groups identified from the Thompson River tributary that were also commonly identified (Small *et al.* 1998; Withler *et al.* 2000; Beacham *et al.* 2003; Beacham and Withler 2017). While these different genetic groups could have originated from different re-colonization events or from different combinations of re-colonization events, they could also have originated as a result of other factors. These factors include: limited gene flow between the groups (e.g., due to barriers like the Fraser Canyon, a phenomenon sometimes referred to as isolation-by-resistance (McRae 2006)), because of stochastic mechanisms like isolation-by-distance, adaptation to environmental conditions (sometimes referred to as isolation-by-environment (Weber *et al.* 2017)), ecotype differentiation in sockeye salmon (Beacham and Withler 2017), or through a combination of these mechanisms. If these genetic groups were a result of different re-colonization events, it is also possible that the re-colonizing populations came from different glacial refugia (Wood *et al.* 1994; Small *et al.* 1998; Teel *et al.* 2000; Withler *et al.* 2000), which would increase their overall distinctiveness.

Wehrhahn and Powell (1987) note that the Fraser Canyon might limit gene flow between lower and upper Fraser River salmon. There are several identified velocity barriers in the Fraser Canyon that could influence salmon migration and particularly limit migration of smaller salmon (Wright 2022). This was dramatically emphasized between 1913-1914 when construction left a massive rockslide in the Fraser Canyon that partially blocked some salmon species, and completely blocked pink salmon (*O. gorbuscha*) (Kew 1992; Grant and Pestal 2009; Pess *et al.* 2012). This single event reduced salmon returns to around a quarter of historic returns to the Fraser River during peak years (Pacific Salmon Commission – psc.org). If the Fraser Canyon acts as a barrier, we might expect unique genetic characteristics above and below the Fraser Canyon for all the species that must traverse it.

Upper and lower Fraser River salmon are also separated by large distances and because they have natal homing with variable straying (reviewed in (Quinn 1993; Keefer and Caudill 2013; Bett *et al.* 2017)), we expect population structure to be influenced by isolation-by-distance (e.g., (Wright 1943; Weber *et al.* 2017; Aguillon *et al.* 2017)). In general, we would expect isolation-by-distance to increase the longer temporally that salmon are isolated from each other and with increased distance. We also expect to observe nearby groups to have higher genetic similarity than those further away for all

species. An exception to this expectation would be salmon with different life history types (e.g., odd and even-year pink salmon (Christensen *et al.* 2021)).

The expectation of increased genetic distance with increased geographic distance will not be the main driver of differentiation if genetic groups were largely influenced by separate re-colonization events. If re-colonization occurred by different genetic groups (e.g., populations from different glacial refugia), we might expect genetic distance to be unrelated to geographic distance in the absence of strong gene-flow or long periods of time. Patterns of re-colonization from different genetic groups would also be expected to be variable among species.

Adaptation to the environment is of particular interest for the conservation and management of a species because salmon from an area could be uniquely suited to that region. The headwaters of the Fraser River flow north down the Rocky Mountain Trench, which leads to the Interior Plateau, followed by the Fraser Canyon, and finally the river outlets through a major delta in the metropolis of Vancouver, British Columbia (Reynoldson *et al.* 2005). These different regions have variable climates, elevations, and river velocities (Reynoldson *et al.* 2005). They are also separated by great distances. Genetic signatures correlated with environmental factors could be from neutral factors such as isolation-by-distance or from non-neutral factors such as local adaptation or other forms of selection. Researchers have not yet found a reliable method for distinguishing among the different mechanisms with genetic data alone (reviewed in (Bierne *et al.* 2013; Lotterhos and Whitlock 2015; Ahrens *et al.* 2018; Saravanan *et al.* 2020)), and other experiments need to be used to firmly establish particular genetic adaptations.

In this study, we analyze 954 resequenced genomes from three species of salmon (317 sockeye, 360 Chinook, and 277 coho salmon), mainly from the Fraser River, to understand differences among the species, identify genetic groups among the collections of samples within each species, and to characterize genetic adaptation to different environments. Data related to genetic structure and local adaptation are particularly important to conservation and management. By using data from multiple species, we compare among species to identify common patterns. This work provides a foundational dataset that will be openly available and of significant use to the salmonid research community for years to come.

Materials and Methods

Sampling and whole genome resequencing

Chinook, coho, and sockeye salmon were sampled from Fraser River tributaries and lakes or were from other bodies of water from previous studies (Figure 1, Figure S1, File S1, (Christensen *et al.* 2020; Rondeau *et al.* 2023)). Some of these sampling locations include hatchery sources (File S1). Most modern hatchery stocks originate from local sources (e.g. (Heard 2012)), and they should reflect the current Fraser River population if not the historical population. The samples were collected during various years (File S1). For coho salmon, there were 125 salmon collected at the Big Bar Landslide between 2019-2021. Spawning locations were unknown for these samples. These were included to identify potentially unknown genetic structure since it is difficult to reach upper Fraser River sites during coho salmon spawning.

Samples were collected by Fisheries and Oceans Canada personnel in compliance with the Canadian Council on Animal Care Guidelines, and under the authority of the Fisheries and Oceans Canada Pacific Region Animal Care Committee (Ex.7.1). Samples were taken either as operculum-clips or as scales. Both were desiccated and stored on Whatman paper.

Genomic DNA was extracted from tissue, after an overnight incubation in 95% ethanol, using a Quick-DNA Kit (Zymo Research). Other genomic DNA samples were previously extracted by Fisheries and Oceans Canada using automated BioSprint extractions, as per manufacturer's instructions (Qiagen). DNA samples were then sent to the Michael Smith Genome Science Centre (Vancouver, BC) for library preparation and whole genome resequencing.

Whole genome sequencing libraries were prepared by shearing the DNA samples individually using a Covaris LE220 (duty cycle: 20%, PIP: 450, cycles per burst: 200, time per run: 90 s; with pulse spin after 45 s). Individual libraries were then constructed using the MGIEasy PCR-Free DNA Library Prep Set (MGI Tech Co.). The indexed libraries were then pooled and sequenced on a MGISeq-G400 sequencer (paired-end 200 bp). Data used in this study were also taken from previous studies (Christensen *et al.* 2020; Rondeau *et al.* 2023) that used Illumina sequencing technology.

SNP calling and filtering

Reads from each individual were aligned to their respective species reference genome assembly (Chinook salmon: GCF_018296145.1 (Christensen *et al.* 2018), coho salmon: GCF_002021735.2 (Rondeau *et al.* 2023), sockeye salmon: unreleased version 2 (Christensen *et al.* 2020)) with BWA (version 0.7.17, parameter -M) (Li and Durbin 2009, 2010; Li 2013). Reads were sorted with SAMtools (version 1.12, default parameters) (Danecek *et al.* 2021). Picard (version 2.26.3, default parameters) (Broad Institute 2019) was used to add read group information and mark reads that were suspected PCR duplicates. GATK (version 3.8) (McKenna *et al.* 2010; Van der Auwera *et al.* 2013) was then used to call nucleotide variants for each individual (parameters: -T HaplotypeCaller, --genotyping_mode DISCOVERY, --emitRefConfidence GVCF) and then combined (parameters: -T GenotypeGVCFs, --max_alternate_alleles 3). Truth and training SNP datasets (File S2) were then used to recalibrate nucleotide variant scores and filter variants using GATK (parameters -T ApplyRecalibration, --mode SNP –ts_filter_level 99.5). The truth SNPs came from multiple studies (Brieuc *et al.* 2014; Meek *et al.* 2016; Nichols *et al.* 2016; Larson *et al.* 2017; Veale and Russello 2017; Rondeau *et al.* 2023), and the training SNPs are described in File S2.

Additional filters were used to remove indels, SNPs with more or fewer than two alleles, or that were missing genotypes in more than 10% of the individuals. SNPs were removed if their mean depths were outside a range of 8-100x, and if they had less than 0.01 minor allele frequency (MAF). All filtering was performed using VCFtools (version 0.1.15) (Danecek *et al.* 2011). The MAF would eliminate alleles found in fewer than 6-8 heterozygous individuals (or 3-4 homozygous individuals), depending on the species and its respective sample size. This threshold was chosen to reduce sequencing errors, but to still keep all but the rarest variants. The MAF filter was not used for the SMC++ analysis. Linkage disequilibrium was evaluated and used to filter SNPs in some analyses, as noted below. The prune add-on for BCFtools (version 1.9) (Danecek *et al.* 2021) was used to filter based on linkage disequilibrium (parameters: +prune, -w 20kb, -l 0.4, -n 2). The filters used for all analyses are shown in Table 1.

Mapping variants among species

To compare analyses among species, variants were mapped from the coho and sockeye salmon genomes to the Chinook salmon genome using a pipeline, *MapVCF2NewGenome* (see *Data Availability*). This pipeline was also used to map sockeye salmon variants, which were on scaffolds, to a chromosome level assembly (this assembly was submitted to the NCBI as version 2 of the reference genome assembly – now available as GCA_034236695.1). The sockeye salmon chromosome mapped version was needed for several analyses.

Coverage, relatedness metrics, and runs of homozygosity (ROH)

Coverage was assessed using a python script, *VCFStats* (see *Data Availability*). This script finds the average depth of all SNPs per individual. It also counts the number and type of genotypes per individual. This information was used to remove individuals, using VCFtools, that had average depth/coverage less than 8x, as a preliminary PCA analysis showed clustering of individuals based on coverage rather than geography if they were below this threshold. These individuals are not discussed and were not analyzed further. Individuals were filtered with an average depth less than 15x depending on the analysis (noted for each analysis). This was because the frequency of heterozygous genotypes dropped when coverage was below 15x and some analyses were sensitive to this issue (Table 1). Genotype information and coverage was visualized in R (R Core Team 2022) using ggplot2 (Wickham 2016).

The related package (version 0.8, parameters of coancestry: ritland = 1) (Pew *et al.* 2015) was used to calculate relatedness (Ritland 1996) in R. Values below zero (e.g., when markers are not shared) and above one (e.g., when rare markers are shared) are expected with this type of estimation (Ritland 1996). There were many more private alleles in coastal populations of Chinook and coho salmon, which may have made them appear much more related. The linkage disequilibrium filtered SNP datasets were converted to the related format using a python script, *VCF2Relate* (see *Data Availability*). The reshape2 (Wickham 2007) and pheatmap (Kolde 2019) packages were used to visualize relatedness. Related individuals were included in most analyses, but we tested what would happen if related individuals were removed from PCA analyses. A relatedness score of 0.15 (an arbitrary value greater than what might be expected for first cousins, but less than for half-sibs – depending on dataset) was used to filter all but one individual from pairs that might be related. Most samples were taken as adults and related individuals were not expected to be common within or among sampling collections.

Runs of homozygosity were identified using PLINK (version 1.9, parameters: --homozyg –double-id, --allow-extra-chr, --homozyg-snp 25, --homozyg-kb, --homozyg-density, --homozyg-gap, --homozyg-window-het 1, --homozyg-window-snp 50, --homozyg-window-threshold 0.05, --homozyg-window-missing 5) (Chang *et al.* 2015; “PLINK 1.9”). This was to compare species levels of runs of homozygosity. All individuals were included in these analyses. No consistent relationship was observed between coverage and runs of homozygosity (e.g., runs of homozygosity might be expected to increase as coverage decreases since there are fewer heterozygous genotypes as noted above). We plotted each individual when comparing runs of homozygosity among species to determine if coverage or relatedness might influence species comparisons.

Population structure and clustering analyses

PCA and admixture analyses were used to identify distinct genetic groups for each species. The admixture software requires independent variants, and PCAs can be sensitive to LD blocks, which can cause clustering that is unrelated to population structure. For this reason, the LD filtered SNPs were used in these analyses. To perform the admixture analyses, the LD filtered data was converted from VCF format to a suitable format using PLINK (version 1.9, parameters: --double-id, --allow-extra-chr). We converted the chromosome names to numbers using Unix commands. The admixture analysis was performed using ADMIXTURE (version 1.3.0, parameters: --cv) (Alexander *et al.* 2009). Cluster values 1-20 were tested and we accepted the value with the lowest cross-validation score. Admixture ancestry values were visualized in R for individuals and in QGIS (QGIS Development Team 2022) as an average score per sampling site.

PCA were used to assess the groupings produced from the admixture analyses. PLINK (parameters: --pca, --double-id, --allow-extra-chr) was used to perform the PCA, and they were visualized in R using ggplot2, reshape2, and ggrepel (Slowikowski 2021). PCA were tried with all the individuals, individuals with ≥15x coverage, and with individuals that were highly related removed. The outputs of these three approaches were compared to determine the influence of coverage and relatedness on clustering and the groups produced from the admixture analyses. We also verified that samples among the new and previous datasets clustered together if they were from the same sampling site.

Environmental variable PCA

To determine whether population structure was associated with environmental factors, we clustered sampling sites by environmental factors using PCA in R (parameters: prcomp, scale=T). Each sampling site was assigned to an admixture group if the average ancestry value was ≥ 0.7 for a particular site. Environmental factors were taken from the WorldClim version 2.1 dataset (Fick and Hijmans 2017). Elevation for each site was either estimated from Google maps, mapcarta, or from data downloaded from the Federal Geospatial Platform (maps.canada.ca) and viewed in QGIS. Distance to the ocean was estimated as the river distance with QGIS or Google maps.

Historical estimates of effective population size

To estimate effective population size through time, we used the program SMC++ (parameters: -c 1000000) (Terhorst *et al.* 2017). The only mutation rate estimate for these species available at the time of writing was for coho salmon (8e-9 from (Rougemont *et al.* 2020)). A correction was applied to this mutation rate for Chinook and sockeye salmon based on the ratio of total SNPs from these species to the total SNPs from coho salmon. For Chinook salmon, this ratio was 1.7274 for the individuals subset from a similar geographic range and for the same number of individuals (mutation rate of 1.382e-8). For sockeye salmon, this was 0.79785342 for the individuals subset, with a mutation rate of 6.38283e-9. Figures were plotted in R using ggplot2, and ocean surface temperature data was taken from previous studies (Zachos *et al.* 2008; Hansen *et al.* 2013).

Genetic diversity

The percent of heterozygous genotypes per individual was calculated by dividing the number of heterozygous genotypes (see *Coverage, relatedness metrics, and runs of homozygosity* section) by all genotypes (including missing data to standardize against all variants) and multiplying by 100. If missing genotypes were excluded from the calculation, values changed by 0.09% on average, and the difference ranged from 0.03%-1.4%. Nucleotide diversity (Pi) and the number of polymorphic loci was calculated using the Stacks populations module (version 2.54, default parameters) (Catchen *et al.* 2011, 2013).

The percent of heterozygous genotypes per individual and nucleotide diversity per sampling site were both calculated from the subset of individuals with ≥15x coverage because these analyses were sensitive to coverage (i.e., as coverage increased, so did the percent of heterozygous genotypes until ≥15x coverage). The number of sampling sites were included in the reported percent of heterozygous genotypes per individual to show if the reported value could be sensitive to relatedness (e.g., fewer locations could be more sensitive to highly related individuals from a single location). Since the nucleotide diversity was plotted per sampling site using the inverse distance weighting interpolation analysis in QGIS, other nearby sites can also be used to determine if relatedness influenced nucleotide diversity regionally.

Admixture group private alleles

A python script was used to identify private alleles among admixture groups of each species (*PrivateAllele* see *Data Availability*). Alleles were identified that were unique to each group if they were present for a specified number of individuals in that group (parameter: -min 3). Individuals were assigned to an admixture group if they had ancestry values ≥ 0.7 for a particular group. Salmon with ancestry values below 0.7 were excluded from these analyses. The private allele counts were visualized in R using ggplot and the gridExtra library (Auguie 2017).

The number of private alleles per individual among admixture groups was identified using another python script, *PrivateAllelePerInd* (see *Data Availability*). This metric identifies if there are individuals or sampling sites that were more responsible for the number of private alleles within a genetic group. The minimum number of individuals with the private allele was set to one. While decreasing this parameter to one could increase the chance that SNP calling errors are included, it also allows us to get an idea for the number of private alleles from a group that are within the genome of each individual. We expect sequencing errors to be evenly distributed among sampling locations and similar among individuals.

Potentially adaptive variants

The rehh library (Gautier and Vitalis 2012; Gautier *et al.* 2017) in R was used to identify extended haplotype homozygosity within and among admixture groups (assignments were made for individuals with ≥ 0.7 ancestry values). To perform the rehh analyses, we used SHAPEIT5 (Hofmeister *et al.* 2023) (version 5.1.1, parameter: phase_common_static) to first phase the genotypes of the MAF 0.01 filtered variants. The output was then converted to VCF format using BCFtools, and the different admixture groups were separated using VCFtools. The rehh command *data2haplohh* was then used to read in each admixture group VCF file, and the *scan_hh* command (parameters, polarized=FALSE, interpolate=FALSE) was used to calculate extended haplotype homozygosity. The within population metric iHS was calculated using the *ihh2ihs* command (default parameters), and the pairwise population metric Rsb was calculated using the *ines2rsb* command (default parameters). The *calc_candidate_regions* function (parameters: threshold=10, window_size=10000, min_n_extr_mrk=8) was used to identify regions with significant differences in extended haplotype homozygosity (the Bonferroni correction p-value threshold was between 8.8-9.13 after -log10 transformation depending on species). A higher threshold was used to identify only the strongest candidates as this is only a preliminary study. We also examined overlapping 10 kbp windows among all three species to identify candidates of potential convergent evolution.

Results

Coverage, relatedness metrics, and runs of homozygosity (ROH)

We identified ~6-13 million SNPs per species (Figure 2a). Chinook salmon had 1.58-2.07x more SNPs than either coho or sockeye salmon, respectively. Between 1-1.75% of all SNP loci were common among species, and only up to 0.1% were found to be polymorphic in all species (Figure 2b). Chinook salmon had a reduction in the total length of runs of homozygosity (ROH), consistent with a higher number of polymorphic loci (Figure 2c). Sockeye and coho salmon had on average ~3.5-3.6x longer total lengths of ROH than Chinook salmon. The length of ROH decreased as the fraction of heterozygous genotypes increased, but generally sockeye and coho salmon had longer ROH when

fractions of heterozygous genotypes were below 0.2 (Figure 2d). No consistent relationship among species was observed between SNP coverage and the total length of ROH (Figure S2).

The average SNP coverage of the combined 954 salmon was 16x (Figure 3a). The average coverage of Chinook salmon was 16.5x (n=360) , coho salmon 18.1x (n=277), and sockeye salmon 13.7x (n=317). The percent of heterozygous genotypes for each individual appeared to depend on the SNP coverage until a depth of 15x (Figure 3a). This relationship was consistent in all species (Figure 3b). This issue did not appear to influence the scale of the number of SNPs among species. The average number of heterozygous genotypes per salmon for those with 15x coverage or greater reflected the trend observed for the number of SNPs per species. These values were 1,667,558 in coho salmon, 1,305,776 in sockeye salmon, and 2,636,167 in Chinook salmon (1.58-2.02x greater in Chinook salmon).

Coastal samples (all sites downstream of the Thompson River or outside the Fraser River – see Figure 1) appeared to be highly related in Chinook and coho salmon (Figure 3c and 3d). These high relatedness scores were independent of the coverage or the technology used. In Chinook and coho salmon, coastal samples had much higher counts of private alleles (see below), which is known to increase relatedness metrics (Ritland 1996). Relatedness values were greatly reduced by sub-setting individuals by coastal or non-coastal (data not shown).

Population structure and clustering analyses

We focused on large-scale resolution of genetic groups in this study, but there was still substantial genetic variation within each of the genetic groups (Figure S3). Each species had three supported Fraser River admixture groups (Figures 4 and S3). All groups were based on geography and were supported by PCA (Figure S4). These included a lower/coastal, mid, and upper Fraser River admixture group.

All of the lower Fraser River (LFR) admixture groups included sample sites in the Fraser Valley south of the Fraser Canyon (approximately where the river turns from a southern flow to a western flow before its outlet to the Pacific Ocean) (Figure 4). In Chinook and coho salmon, the LFR admixture groups also included nearby coastal sites. For coho salmon, this included locations from Alaska to California.

We sampled fewer coho salmon north of the Fraser Valley than for the other species due to accessibility issues during their spawning season, but those north of the valley clustered in a pattern similar to Chinook salmon (Figure 4). All samples, excluding those from McKinley Creek (a tributary of the Horsefly River), and some samples without known natal streams (sampled at the Big Bar Landslide), form one admixture group. In Chinook salmon, a similar admixture group is formed from the collection of sites west of the Horsefly River and from the Nechako River south. This geographic region is known as the Interior Plateau. In sockeye salmon, the mid Fraser River (MFR) group is composed of bodies of water from the Chilcotin and Quesnel River watersheds (Figure 4).

The upper Fraser River (UFR) coho and Chinook salmon admixture groups include most of the Quesnel watershed locations, with the exception of the lower Cariboo River near the confluence with the Fraser River (Figure 4). For coho salmon, this was from a single known sampling location and potentially other locations from salmon that were collected at the Big Bar Landslide. The Chinook salmon UFR group also included sampling sites above the Nechako River (in or near Robson Valley, a part of the Rocky Mountain Trench). The UFR sockeye salmon admixture group is comprised of Nechako River watershed sites.

The LFR admixture groups, in all species, have the greatest genetic differentiation (i.e., F_{st}) from the other admixture groups (Figure 4). In addition, the UFR admixture groups have higher

differentiation from the LFR admixture groups than the MFR groups (Figure 4). F_{st} values were generally much lower between UFR and MFR admixture groups. While there were some significant differences in the percent of heterozygous genotypes per individual among groups, these do not show a consistent pattern (Figure 4). In Chinook salmon, the significant differences in heterozygous genotypes per individual between the MFR admixture group and the other groups were driven by high values in upper Chilcotin River samples (File S1).

Environmental factors influencing genetic structure

Sampling locations from different admixture groups clustered by environmental factors (Figure 5, Figure S5). Sites with predominately LFR ancestry values were in regions with higher annual precipitation and higher mean annual temperatures (Figure S5). While there was some overlap between MFR and UFR admixture groups, there was still distinct environmental factors between them overall. This was the case for all three species.

Historical estimates of effective population size

Historical estimates of effective population size (N_e) are strikingly different among species (Figure 6, Figure S6). Prior to 200,000 years before present, almost all Chinook and multiple coho salmon sampled experienced drops in estimates of effective population size (Figure 6, Figure S6). This coincides with a time when there was an increase in ocean surface temperatures (Figure 6d).

Between 120,000 and 150,000 years before present, coho and sockeye salmon both experienced major declines in estimated effective population size (Figures 6 and 7). In sockeye salmon, there are distinct times when these decreases occurred (Figure 7). For some sockeye salmon locations and for most coho salmon sites, the decrease in effective population size occurred around the penultimate glacial maximum, when ocean temperatures were at their lowest. Neither species recovered to previous estimates of effective population size for over 100,000 years (Figures 6 and S6). This is consistent with the decrease in polymorphic loci and increases of runs of homozygosity compared to Chinook salmon. Interestingly, Chinook salmon effective population size increased during this same time frame. A similar reversal among species was observed between 50,000 and 90,000 years before present, where there were increases in estimates of effective population size for coho and sockeye salmon, but decreases in Chinook salmon (Figure 6, Figure S6).

In all species, there were decreases in effective population size around the last glacial maximum, 19,000 years before present (Figures 6 and 7, Figure S6). The timing of these decreases in effective population size varied in time by species, admixture group, and sampling sites within admixture groups (Figures 6 and 7, Figure S6). Most of the demographic histories have a step pattern near or after the last glacial maximum. Some coastal Fraser River groups have earlier decreases in effective population sizes than the other Fraser River groups. The exception being LFR sockeye salmon sites, which have similar or more recent decreases in population size than MFR and UFR locations (Figure S6).

Understanding the genetic diversity of Fraser River salmon

On average, total runs of homozygosity (ROH) were shorter in LFR locations than MFR and UFR sites of Chinook and coho salmon (Figure 8). Lower ROH is a proxy for greater genetic diversity, suggesting that the Chinook and coho salmon LFR admixture groups have higher diversity than MFR and UFR groups. Sampling sites with the longest average total ROH were the Yakoun River (Haida Gwaii) and Salmon River (tributary of the Thompson River) for Chinook and coho salmon, respectively (Figure S7). Locations with the shortest average total ROH (i.e., highest diversity) for

Chinook and coho salmon, respectively, were the upper Chilcotin River (tributary of the Fraser River) and the Deschutes River (tributary of the Columbia River) (Figure S7). ROH levels were most variable in the Morkill and upper Chilcotin River locations for Chinook salmon. The individuals with the shorter ROH in the upper Chilcotin River drove the average to the lowest for Chinook salmon.

There was a reverse trend for ROH in sockeye salmon compared to Chinook and coho salmon (Figure 8). Most of the sockeye salmon sites with the longest ROH (i.e., the lowest diversity) were from the LFR (i.e., Cultus Lake, Pitt River, and Widgeon Slough). Cultus Lake and Widgeon Slough have small population sizes, often under 1000 salmon in recent generations (DFO 2020; Doutaz *et al.* 2023). The Harrison River, although not a part of this admixture group due to being below the 0.7 admixture threshold, had the lowest average ROH of all sockeye salmon collections (Figure S7). The Harrison River, in the early 1980s, had around 300,000 spawning adults (Doutaz *et al.* 2023). Widgeon Slough and possibly the Harrison River were also the only samples of the sea-type ecotype taken from the Fraser River (Beacham and Withler 2017; Doutaz *et al.* 2023), all the other Fraser River samples were likely lake-type. Also, some Harrison River salmon are sea and some are lake-type. Bodies of water in the Yukon, Alaska, Russia, and the upper Columbia River had slightly lower average ROH (i.e, more diverse) compared to the LFR, but higher than sites from the MFR and UFR (Figure 8 and Figure S7).

Nucleotide diversity (π), which is a population-level metric rather than an individual-level metric like ROH, had a more complex pattern. Nucleotide diversity ranged from 0.147 (Chinook salmon – Yakoun River) to 0.22 (sockeye salmon – Pinchi Creek). All sockeye salmon sampling locations had nucleotide diversity metrics above 0.19 except for Widgeon Slough, which was 0.17. Cultus Lake might have had lower nucleotide diversity than Widgeon Slough based on its longer ROH, but this location was excluded from this analysis since it had lower SNP coverage than 15x. Similarly, most northern coastal and upper Columbia River samples were excluded from the nucleotide diversity analysis because they had SNP coverage below the threshold. Most Fraser River salmon locations had comparable nucleotide diversity to upper Columbia River Kokanee that were available for comparison. Similar to the ROH results, the LFR salmon had lower diversity than the MFR and UFR sockeye salmon.

Most Chinook salmon locations had π values between 0.18 and 0.20. However, the Yakoun River (Haida Gwaii) had the lowest π of all sites for Chinook salmon ($\pi = 0.147$). Many of the sampling locations with the highest nucleotide diversity were from the MFR (Figure 8). The Fraser River Chinook salmon sampling sites had average nucleotide diversity values comparable or lower than other coastal locations along the BC coast (Figure 8).

Coho salmon ranged in nucleotide diversity from 0.17 to 0.21, with the lowest collection site being the Salmon River in the South Thompson River drainage and the highest being the Kawkawa Creek below the Fraser Canyon (Figure 8). Fraser River coho salmon had comparable or higher levels of nucleotide diversity as those from California to Alaska. The most extreme northern (Kwethluk River) and southern sites (Klamath River) had the second and third lowest nucleotide diversity metrics respectively.

The number of polymorphic loci was highest in Chinook salmon sampling sites, followed by coho salmon and then sockeye salmon, which is consistent with the total number of SNPs identified in each species (Figure S8, Figure 2). While the number of polymorphic loci was dependent on the number of samples per location, it remained higher in Chinook salmon at comparable sample counts per location (Figure S8). Even though Chinook salmon had the greatest number of polymorphic loci per location, nucleotide diversity was often lower than for coho and sockeye salmon locations (Figure S8). This is consistent with higher individual percent heterozygous genotypes in coho and sockeye salmon (Figure 3).

Admixture group private alleles

There were more private alleles identified in the LFR admixture groups of Chinook and coho salmon than the other admixture groups, with 5-13% of all the identified SNPs being private to the LFR group (Figure 9a). Sockeye salmon had a more even distribution of private alleles among admixture groups than Chinook or coho salmon (Figure 9a). However, the sample size of the LFR sockeye salmon group was the lowest of all comparisons, and this may have impacted the number of private alleles identified. Also, sockeye salmon from the Harrison River (in the coastal region) had admixture ancestry value below 0.7 and were excluded from this analysis. This site had greater genetic diversity metrics than most other sockeye salmon locations (see previous section), and so the removal of this river would be expected to reduce the observed private alleles from this region.

Analyses of individual private alleles from Chinook and coho salmon provided more evidence that LFR individuals have more unique alleles than individuals from MFR and UFR groups (Figure 9b-d). Each coastal individual had roughly 130,000 private alleles that were not found in the other admixture groups. Other admixture groups, on average, have much fewer individual private alleles compared to LFR salmon (~ 0.04% - 10%). The upper Chilcotin River and Spius River contributed more private alleles to the Chinook salmon MFR admixture group than other rivers in that group (Figure 9b). Likewise, the Kawkawa River contributed fewer private alleles to the coho salmon LFR group than the other rivers (Figure 9d).

Sockeye salmon do not have a comparable admixture group with the large store of private alleles relative to that observed in the LFR coho and Chinook salmon. The northern coast and upper Columbia River kokanee admixture groups have the highest average individual private allele counts (Figure 9c). Most sampling locations in the Fraser River have lower individual private alleles compared to those outside the Fraser River. The main exceptions are samples from the Quesnel River watershed, which have similar values as those outside of the Fraser River (Figure 9c). This is consistent with a higher nucleotide diversity in the Quesnel River watershed than the Chilcotin River watershed (see previous section).

Potentially adaptive variants

Twenty candidate adaptive loci were identified in pairwise comparisons among admixture groups based on expanded haplotype homozygosity analyses (Figure S9, Table 2, File S3). No significant loci were identified from within admixture group analyses. Most candidate loci were identified among Chinook salmon admixture groups (14 out of 20). Twelve of these loci overlapped with the boundary of a gene (Table 2). Two different homologs of the *trace amine-associated receptor 13C* (TAAR13C) gene overlapped with candidate loci identified in comparisons of admixture groups of sockeye salmon (Figure S9, Table 2). One of these genes was identified from the comparison with the LFR group and the MFR group. The other was identified from the comparison of the LFR group and the UFR group.

To identify if there were potential adaptive variants shared among species, we searched for overlapping regions with high -log10 p-values (we were searching for signals of convergent evolution with this analysis). The overlap of 10 kb regions with the highest -log10 p-value was 3.6 among species. The chance of having an overlap with this high of a value was expected by random at least twice by chance given the distribution of p-values. These overlaps were not considered significant and they are not shown or discussed below.

Discussion

The Fraser River drains a very large and diverse section of British Columbia from the Rocky Mountains to its outlet in the city of Vancouver (Reynoldson *et al.* 2005). This river has at times been the largest producer of salmon in North America (Milne 1964; Northcote and Atagi 1997), and remains important commercially (Henderson and Healey 1993) and culturally for its salmon fisheries. The massive declines in salmon, predominately sockeye and pink salmon from a 1913-1914 landslide caused by the construction of a railway, have never fully recovered (Pacific Salmon Commission – psc.org). There are many pressures potentially for why recovery has not occurred and why there has been continued declines in Fraser River salmon (Arbeider *et al.* 2020; DFO 2021; Doutaz *et al.* 2023). Understanding these pressures and their influence on salmon will be valuable for governance and conservation. Our goal is to provide the genetic tools that will facilitate this understanding and help with recovery.

While there have been many studies to examine population structure in the Fraser River (Small *et al.* 1998; Teel *et al.* 2000; Withler *et al.* 2000; Beacham *et al.* 2001, 2003, 2006b, 2006a, 2017; Nelson *et al.* 2001; Beacham and Withler 2017; Xuereb *et al.* 2022), this is the first to use whole genome resequencing data for a large distribution of multiple salmon species. The improved resolution of resequenced genomes allowed us to compare the genetic diversity among salmon species, calculate nucleotide diversity at the genome level, and evaluate historical influences on population structure. These types of analyses would suffer using fewer genetic markers. As an example, genetic and nucleotide diversity estimates are dependent on the nucleotide variants used to estimate them (e.g., polymorphic loci in one region might not be polymorphic in another region) and whole genome analyses allow unbiased estimates among locations. These genome sequences also us to capture a better snapshot in time that can be used in comparisons for future studies.

Species level differences of polymorphic loci

Genetic diversity can be understood at both a species and population level. We first discuss the diversity at a species level and will return to the population level later. From previous research, we expected different levels of polymorphic loci among Pacific salmon, with a generally lower genetic diversity of Pacific salmon compared to other species such as rainbow trout (*O. mykiss*) (Allendorf and Utter 1979). Several studies observed that sockeye salmon, and sometimes coho salmon, have the lowest levels of polymorphic loci or average heterozygosity (Utter *et al.* 1973; Allendorf and Utter 1979; Wood *et al.* 1994).

Our findings indicate that Chinook salmon have ~1.5-2x the number of polymorphic loci and a corresponding drop in the length of ROH (~27-29% of the ROH length) when compared to coho or sockeye salmon. We observed the same trend based on the total number of heterozygous genotypes among the species when only looking at individuals with higher than 15x coverage. While sequencing coverage likely impacts the number of polymorphic loci identified, the scale of the differences we observed, the greater geographic sampling distance covered by the species with fewer polymorphic loci, a similar trend in ROH, and a similar trend with heterozygous genotypes of individuals with at least 15x coverage supports that these findings are robust. If we assume that salmon species had similar levels of genetic diversity during speciation and that mutation rates are similar among species, coho and sockeye salmon likely had a greater reduction in standing genetic variation than Chinook salmon.

From modelling of effective population size through time, there is evidence that sockeye and coho salmon experienced a large drop in effective population size around the penultimate glacial maximum that Chinook salmon did not experience (Chinook salmon did experience an earlier drop in effective population size, but the size rebounded after ~50,000 years). We must consider that we do not have precise estimates of mutation rates for all of these species. Even without precise dates or effective population size estimates, however, we note that the drop in effective population size did not rebound to prior levels in sockeye and coho salmon as it did in most Chinook salmon sites. This could have resulted in a loss of polymorphic loci in both sockeye and coho salmon not observed in Chinook salmon.

While the SNP count, effective population size, and runs of homozygosity data are consistent with this hypothesis, nucleotide diversity and private allele analyses appear to be inconsistent at first glance. For example, why would sockeye and coho salmon have higher nucleotide diversity if Chinook salmon have more polymorphic loci overall (sockeye salmon average π: 0.205, coho salmon: 0.197, Chinook salmon: 0.189)? Also, why would the average private allele count be higher in the LFR coho salmon admixture group than the Chinook salmon LFR group if the Chinook salmon have more polymorphic loci?

First, we note that some coastal samples of sockeye salmon were removed due to low coverage in these analyses, and these samples would have likely lowered the average nucleotide diversity score. Technical artifacts like this and others (e.g., how nucleotide diversity was measured, sequence coverage, or differences among genome assemblies) might help to explain a part of these discrepancies, but they appear to be inadequate for all of the data. For example, at comparable SNP coverages, sockeye salmon had higher proportions of heterozygous genotypes than coho or Chinook salmon even though coho and Chinook salmon have more polymorphic loci overall.

The timing of re-colonization and the distribution of groups with unique genetic characteristics could help to account for some of the discrepancies among genetic diversity metrics. In Chinook and coho salmon, for example, the coastal genetic groups have the largest stores of private alleles and the highest metrics of genetic diversity (comparable between species, even if slightly higher in coho salmon). If there was a difference in the timing or the size of the groups that re-colonized the Fraser River from these coastal genetic groups, we might expect the fraction of private alleles from a group to reflect these differences.

Estimates of effective population size dip in Chinook salmon locations soon after the last glacial maximum when compared to coho salmon. If we assume these dips are founding events and the times are roughly accurate, we infer that Chinook salmon re-colonized the Fraser River earlier. The founding events also appeared to have reduced the effective population size of Chinook salmon sites less than they did for coho salmon, perhaps reflective of larger founding populations in Chinook salmon. With this in mind, even with fewer polymorphic loci overall, the LFR coho salmon might appear to have higher levels of private alleles per individual than Chinook salmon when comparing the LFR group to the MFR and UFR groups because those groups have lower genetic diversity as a result of more recent and smaller founding events. The MFR and UFR coho salmon groups could have retained comparable nucleotide diversity metrics to Chinook salmon groups due to gene flow, or because not as much time has passed for genetic drift to impact this metric as it has for Chinook salmon. Also, heterozygosity excess (e.g., (Wang *et al.* 2016)), where allele frequency varies by chance more often between males and females due to a low breeding population, is another viable explanation for how MFR or UFR coho salmon might have comparable nucleotide diversity to Chinook salmon.

Nucleotide diversity might also be higher in a species with lower overall counts of polymorphic loci and genetic diversity if there was recent admixture between diverse populations. In sockeye salmon, there were at least two distinct demographic patterns identified based on when effective

population size declined around the penultimate glacial maximum (perhaps indicative of groups from different glacial refugia as previously suggested (Wood *et al.* 1994; Beacham *et al.* 2006b)). In the Fraser River, both demographic patterns were observed, with the MFR and UFR admixture groups mostly having the earlier decline in effective population size (earlier than 150,000 years before present), while the LFR locations had mostly the later (after 150,000 years before). Gene flow between these groups could explain the higher nucleotide diversity in the MFR and UFR even though sockeye salmon as a whole have fewer polymorphic loci.

Overall, the data point to a model where Chinook salmon have more polymorphic loci, but less diversity at these loci than coho or sockeye salmon. Demographic modelling points to one explanation for how this might have occurred, but other explanations are possible. For example, heterozygosity excess is expected in populations with smaller effective population size (Robertson 1965; Wang *et al.* 2016).

Influences on the population structure of Fraser River salmon

Turning from the species level to the population level, researchers have known the basic population structure of Fraser River salmon since the mid 1990's to the mid 2000's (Wood *et al.* 1994; Small *et al.* 1998; Teel *et al.* 2000; Withler *et al.* 2000; Nelson *et al.* 2001; Beacham *et al.* 2003, 2006a, 2006b, 2017; Beacham and Withler 2017; Xuereb *et al.* 2022). This knowledge was based on relatively few genetic markers but often many samples from a wide distribution. In the current work, we sampled a moderate number of locations and individuals, but resequenced entire genomes. The basic population structure identified in the Fraser River from resequencing genomes was similar to findings from earlier studies using microsatellite and allozyme genetic markers (e.g., lower, mid, and upper groups), except the Thompson River groups were missing since we did not sample the Thompson River drainage enough to reconstruct them. We interpret the consistency among studies to mean that these genetic clusters are stable through time (as sampling took place at various times, even within studies (Beacham *et al.* 2003, 2006a)), the clusters are robust since they can be identified with only a few genetic markers, and the entire genome is impacted by these groupings.

Understanding why we observe these genetic groups is important (e.g., (Nadeau *et al.* 2016; Rougemont *et al.* 2023)). Are they a technical artifact from trying to cluster genomes influenced by isolation-by-distance? Are they an artifact from the transfers between regions of the Fraser River (Withler 1982)? Were they formed from different re-colonization events (or a combinations of events)? Are they a consequence of environmental hurdles to gene-flow? Were they formed from adaptations to specific environmental conditions?

If the groups are technical artifacts from isolation-by-distance, they may not be useful from a conservation or management perspective because the groups were shaped by stochastic processes rather than adaptations to different environments. If instead, groups were shaped by different re-colonization events, they might reflect those colonization histories rather than the environments they occupy. If the groups are a result of limited gene flow, we might expect the salmon to have specific adaptations to environmental factors that influence gene flow (e.g., the ability to pass regions of the Fraser River such as the Fraser Canyon), but not necessarily to other environmental conditions. These are only some examples of why these groups might exist. The genetic groups could also be influenced by a combination of these processes.

Previous researchers have suggested that isolation-by-distance is an important factor influencing salmon genetics (Withler *et al.* 2000; Rougemont *et al.* 2020, 2023). Others have suggested that different re-colonization events (e.g., from different glacial refugia) influence population

structure in Fraser River salmon (Wood *et al.* 1994; Small *et al.* 1998; Teel *et al.* 2000; Withler *et al.* 2000). Still others have suggested that barriers to gene-flow like the Fraser Canyon (specifically Hell's Canyon) are important for population structure (Wehrhahn and Powell 1987). Finally, researchers have also suggested genetic adaptation to environmental conditions are important (Small *et al.* 1998). Analyzing multiple species can help with distinguishing which hypotheses are more supported (Nadeau *et al.* 2016).

While we observed lower F_{st} between neighbouring admixture groups in all species, which would be consistent with isolation-by-distance, we also observed that private alleles were more common in the UFR group of Chinook salmon than in the MFR group. If populations were delineated strictly by isolation-by-distance, we might expect UFR salmon to have the lowest count of private alleles. Also, in all three species, sampling sites of an admixture group clustered together based on environmental factors in a PCA. If admixture groups were the result of neutral genetic variation or different colonization events alone, we would expect random sorting of groups or for it to be based on distance between locations. This suggests that the environment is a major factor influencing genetic variation in Fraser River salmon, but still does not exclude isolation-by-distance or separate colonization events.

Indeed, from demographic modelling, separate colonization events appear to have been an important source of sockeye salmon in the Fraser River and possibly for Chinook salmon. For sockeye salmon, there are at least two unique demographic histories, with one common to LFR sites and the other to MFR and UFR locations. These groups have unique demographic histories starting as early as the penultimate glacial maximum. In Chinook salmon, LFR sampling locations have an earlier decrease in effective population size that may be indicative of an earlier colonization history. Overall, the variability in demographic histories of the species suggest a complex re-colonization history.

In all three species, there was a delineation between admixture and PCA groups between the LFR group and the MFR group. The LFR groups had higher F_{st} values among groups, and 5-13% of the SNPs we identified in Chinook and coho salmon were private alleles of the LFR group. If we also consider the velocity barriers between these groups (Wright 2022), we have several pieces of evidence to suggest that there is a substantial gene flow and migration barrier due to the Fraser Canyon. In support of this hypothesis, we observed that the Chinook and coho salmon from the LFR groups had higher nucleotide diversity and that longer runs of homozygosity in coho salmon appeared to be demarcated near Hell's Canyon in the Fraser Canyon. If the Fraser Canyon is a migration barrier, we would expect to observe these genetic delineations in all species and generally this is what was observed.

In a study of Fraser River sockeye salmon, researchers observed possible adaptations to environmental differences among populations (Eliason *et al.* 2011). They observed coastal populations had significantly different cardiac morphology and performance from other groups, and that cardiac morphology was correlated with migration difficulty. River temperature was also correlated with cardiac performance. This study highlights the importance of the environment on population structure, migration difficulty, and relates the genetic structure among admixture groups to phenotypic variation. We note that in a comparison of extended haplotype homozygosity between the LFR coho salmon with the MFR group, a potential adaptive locus was identified that overlapped with the *Ankyrin Repeat And SOCS Box Containing 2* (ASB2) gene. This gene is thought to be involved in heart development (Yamak *et al.* 2020; Min *et al.* 2021).

Influence of adaptation on population structure of Fraser River salmon

When comparing genetic groups, we identified twenty candidate adaptive loci. No significant loci overlapped among species, which would have been evidence for convergent evolution. Rather than discuss all twenty regions, we will focus on what these loci reveal in general, and as an example discuss the olfactory receptor gene TAAR13C. Potentially adaptive loci overlapped with two TAAR13C genes and were identified in separate comparisons of the LFR sockeye salmon with MFR and UFR groups.

In general, adaptive loci among the genetic groups reveal that there are different selective pressures along the Fraser River drainage. While the environmental PCA gave us insight into how genetic groups were organized based on environmental components, adaptive loci can reveal how these and other elements shape the genomes of salmon through generations. In the type of analysis we used, we did not need to know what these elements were. This means we can discover adaptation caused by unknown and unexamined factors.

We may be able to formulate a hypothesis regarding a mode of adaptation in the case of the TAAR13C genes. The olfactory receptor gene, TAAR13C, is directly involved in detecting putrescine, cadaverine, and other diamines (typically associated with decomposing flesh) (Hussain *et al.* 2013; Tessarolo *et al.* 2014; Liberles 2015; Gainetdinov *et al.* 2018). The TAAR13C gene has also previously been found to be associated with sea age at maturity in Atlantic salmon (*Salmo salar*) (Sinclair-Waters *et al.* 2022), and to possibly be under selection in another study of Atlantic salmon (not peer reviewed at the time of writing) (Miettinen *et al.* 2023). Since the TAAR13C gene appears to be involved in the timing of maturation, one hypothesis is that diamines could act as a signal for maturation for the different genetic groups of sockeye salmon in the Fraser River. One source of diamines to consider for this hypothesis comes from eggs. A study of Arctic charr (*Salvelinus alpinus*) eggs and alevin revealed that during these developmental stages different amounts of putrescine and cadaverine were produced during the spawning season (Srivastava *et al.* 1992). If diamines influenced maturation of MFR and UFR groups of sockeye salmon, this information would be valuable in conservation and management. Understanding the modes of adaptation of other potential adaptive loci could likewise be useful for these purposes.

Conclusion

From analyzing hundreds of resequenced genomes of Chinook, sockeye, and coho salmon, mostly from the Fraser River, we identified genetic groups that had previously been identified with only a few microsatellites. With data from resequenced genomes, we were able to examine how these groups might have formed. They appear to have been influenced by many factors, including isolation-by-distance, migration barriers, separate glacial refugia in sockeye salmon, and by the diverse environmental factors of the Fraser River drainage. We identified twenty potentially adaptive loci among the different groups of salmon, which is indicative of unique patterns of selection in the various regions of the river. Two of these loci overlapped with homologs of the TAAR13C gene. This gene could be an important target for future studies, to investigate if there is a link between it and timing of maturation in Fraser River sockeye salmon. Finally, by examining three species, we were able to identify commonalities and differences. Patterns of historical effective population size were dramatically different among the three species and could explain the variable genetic diversity currently observed among them. In terms of information useful from a conservation and management perspective, we have generated resequenced genomes that can be reused in future studies, determined metrics for comparison, and identified loci that could impact the timing of maturation of salmon in the Fraser River.

Data availability

Resequenced genomes from this study are available on the NCBI (coho salmon: PRJNA986075, sockeye salmon: PRJNA930425, and Chinook salmon: PRJNA694998 and PRJNA1090956). Truth SNP datasets used for SNP calling are available in File S2. SNP datasets are available in FigShare. Scripts are available at github.com/KrisChristensen (repositories: PrivateAllelePerInd, PrivateAllele, VCF2Relate, VCFstats, and MapVCF2NewGenome).

Acknowledgements

We would like to thank the many people involved with sample collection at Fisheries and Oceans Canada (dfo-mpo.gc.ca). For computer resources, we would like to acknowledge the Digital Research Alliance of Canada (alliancecan.ca) and its regional partner the BC DRI Group at the University of Victoria. For their sequencing services, we would like to thank the Michael Smith Genome Science Centre and the technicians there.

Funding

Funding was provided by the British Columbia Salmon Restoration and Innovation Fund (genetic baseline) through the Pacific Salmon Foundation, and from the Fisheries and Oceans Canada Canadian Regulatory System for Biotechnology.

References

Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark *et al.*, 2017 Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics 13: e1006911.

Ahrens, C. W., P. D. Rymer, A. Stow, J. Bragg, S. Dillon *et al.*, 2018 The search for loci under selection: trends, biases and progress. Molecular Ecology 27: 1342–1356.

Alexander, D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19: 1655–1664.

Allendorf, F. W., and F. M. Utter, 1979 Population genetics, pp. 407–454 in *Fish physiology*, edited by W. S. Hoar, D. J. Randall, and J. R. Brett. Academic Press, New York, NY.

Andrews, G. D. M., J. K. Russell, S. R. Brown, and R. J. Enkin, 2012 Pleistocene reversal of the Fraser River, British Columbia. Geology 40: 111–114.

Arbeider, M., L. M. Ritchie, D. Braun, B. Jenewein, K. Rickards *et al.*, 2020 Interior Fraser Coho Salmon Recovery Potential Assessment: Fisheries and Oceans Canada Canadian Science Advisory Secretariat 2020/025, xi + 211 p. p.

Auguie, B., 2017 gridExtra: Miscellaneous Functions for “Grid” Graphics.

Beacham, T. D., J. R. Candy, K. J. Supernault, T. Ming, B. Deagle *et al.*, 2001 Evaluation and Application of Microsatellite and Major Histocompatibility Complex Variation for Stock Identification of Coho Salmon in British Columbia. Transactions of the American Fisheries Society 130: 1116–1149.

Beacham, T. D., K. L. Jonsen, J. Supernault, M. Wetklo, L. Deng *et al.*, 2006a Pacific Rim Population Structure of Chinook Salmon as Determined from Microsatellite Analysis. Transactions of the American Fisheries Society 135: 1604–1621.

Beacham, T. D., B. McIntosh, C. MacConnachie, K. M. Miller, R. E. Withler *et al.*, 2006b Pacific Rim Population Structure of Sockeye Salmon as Determined from Microsatellite Analysis. Transactions of the American Fisheries Society 135: 174–187.

Beacham, T. D., K. J. Supernault, M. Wetklo, B. Deagle, K. Labaree *et al.*, 2003 The geographic basis for population structure in Fraser River chinook salmon (Oncorhynchus tshawytscha). Fishery Bulletin 101: 229–242.

Beacham, T. D., C. Wallace, C. MacConnachie, K. Jonsen, B. McIntosh *et al.*, 2017 Population and individual identification of coho salmon in British Columbia through parentage-based tagging and genetic stock identification: an alternative to coded-wire tags. Can. J. Fish. Aquat. Sci. 74: 1391–1410.

Beacham, T. D., and R. E. Withler, 2017 Population structure of sea-type and lake-type sockeye salmon and kokanee in the Fraser River and Columbia River drainages. PLOS ONE 12: e0183713.

Beatty, G. E., and J. Provan, 2010 Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant Orthilia secunda. Molecular Ecology 19: 5009–5021.

Bernatchez, L., and C. C. Wilson, 1998 Comparative phylogeography of Nearctic and Palearctic fishes. Molecular Ecology 7: 431–452.

Bett, N. N., S. G. Hinch, N. J. Burnett, M. R. Donaldson, and S. M. Naman, 2017 Causes and Consequences of Straying into Small Populations of Pacific Salmon. Fisheries 42: 220–230.

Bierne, N., D. Roze, and J. J. Welch, 2013 Pervasive selection or is it…? why are FST outliers sometimes so frequent? Molecular Ecology 22: 2061–2064.

Brieuc, M. S. O., C. D. Waters, J. E. Seeb, and K. A. Naish, 2014 A dense linkage map for Chinook salmon (Oncorhynchus tshawytscha) reveals variable chromosomal divergence after an ancestral whole genome duplication event. G3 (Bethesda) 4: 447–460.

Broad Institute, 2019 Picard Toolkit. Broad Institute, GitHub Repository.

Byun, A. S., B. Koop, and T. E. Reimchen, 1999 Coastal Refugia and Postglacial Recolonization Routes: A Reply to Demboski, Stone, and Cook. Evolution 53: 2013–2015.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait, 2011 Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. G3 Genes|Genomes|Genetics 1: 171–182.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013 Stacks: an analysis tool set for population genomics. Mol Ecol 22: 3124–3140.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:.

Christensen, K. A., J. S. Leong, D. Sakhrani, C. A. Biagi, D. R. Minkley *et al.*, 2018 Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. PLOS ONE 13: e0195461.

Christensen, K. A., E. B. Rondeau, D. R. Minkley, D. Sakhrani, C. A. Biagi *et al.*, 2020 The sockeye salmon genome, transcriptome, and analyses identifying population defining regions of the genome. PLOS ONE 15: e0240935.

Christensen, K. A., E. B. Rondeau, D. Sakhrani, C. A. Biagi, H. Johnson *et al.*, 2021 The pink salmon genome: Uncovering the genomic consequences of a two-year life cycle. PLOS ONE 16: e0255752.

Clague, J. J., 2017 Deglaciation of the Cordillera of Western Canada at the end of the Pleistocene. Cuadernos de Investigación Geográfica 43: 449–466.

Clark, P. U., A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark *et al.*, 2009 The Last Glacial Maximum. Science 325: 710–714.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan *et al.*, 2021 Twelve years of SAMtools and BCFtools. GigaScience 10: giab008.

Darvill, C. M., B. Menounos, B. M. Goehring, O. B. Lian, and M. W. Caffee, 2018 Retreat of the Western Cordilleran Ice Sheet Margin During the Last Deglaciation. Geophysical Research Letters 45: 9710–9720.

DFO, 2020 Recovery Potential Assessment – Cultus Lake Sockeye Salmon (Oncorhynchus nerka) (2019): DFO Canadian Science Advisory Secretariat Science Advisory Report 2020/011.

DFO, 2021 Recovery Potential Assessment for 11 Designatable Units of Fraser River Chinook Salmon, Oncorhynchus tshawytscha, Part 2: Elements 12 to 22: Canadian Science Advisory Secretariat Science Advisory Report 2021/030.

Doutaz, D., A.-M. Huang, S. Decker, and T. Vivian, 2023 Recovery Potential Assessment for Fraser River Sockeye Salmon (Oncorhynchus nerka), Nine Designatable Units Part 2: Biology, Habitat, Threats, Mitigations and Allowable Harm - Elements 1-11, 14, 16-18, 22: Fisheries and Oceans Canada Canadian Science Advisory Secretariat 2023/003, xiii + 250 p.

Eliason, E. J., T. D. Clark, M. J. Hague, L. M. Hanson, Z. S. Gallagher *et al.*, 2011 Differences in Thermal Tolerance Among Sockeye Salmon Populations. Science 332: 109–112.

Fick, S. E., and R. J. Hijmans, 2017 WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37: 4302–4315.

Gainetdinov, R. R., M. C. Hoener, and M. D. Berry, 2018 Trace Amines and Their Receptors (J. M. Witkin, Ed.). Pharmacol Rev 70: 549–620.

Gautier, M., A. Klassmann, and R. Vitalis, 2017 rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol Ecol Resour 17: 78–90.

Gautier, M., and R. Vitalis, 2012 rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28: 1176–1177.

Grant, S., and G. Pestal, 2009 CERTIFICATION UNIT PROFILE: FRASER RIVER PINK SALMON: Department of Fisheries and Oceans Canada Canadian Manuscript Report of Fisheries and Aquatic Sciences 2875, vii+36p. p.

Hansen, J., M. Sato, G. Russell, and P. Kharecha, 2013 Climate sensitivity, sea level and atmospheric carbon dioxide. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371: 20120294.

Harington, C. R., 1996 Quaternary Animals: Vertebrates of the Ice Age, pp. 261–273 in *Life in Stone: A Natural History of British Columbia's Fossils*, edited by R. Ludvigsen. University of British Columbia Press, Vancouver.

Heard, W. R., 2012 Overview of salmon stock enhancement in southeast Alaska and compatibility with maintenance of hatchery and wild stocks. Environ Biol Fish 94: 273–283.

Henderson, M. A., and M. C. Healey, 1993 Doubling sockeye salmon production in the Fraser River—Is this sustainable development? Environmental Management 17: 719–728.

Hewitt, G. M., 1996 Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58: 247–276.

Hewitt, G., 2000 The genetic legacy of the Quaternary ice ages. Nature 405: 907–913.

Hofmeister, R. J., D. M. Ribeiro, S. Rubinacci, and O. Delaneau, 2023 Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. Nat Genet 55: 1243–1249.

Hussain, A., L. R. Saraiva, D. M. Ferrero, G. Ahuja, V. S. Krishna *et al.*, 2013 High-affinity olfactory receptor for the death-associated odor cadaverine. Proceedings of the National Academy of Sciences 110: 19579–19584.

Keefer, M. L., and C. C. Caudill, 2013 Homing and straying by anadromous salmonids: a review of mechanisms and rates. Rev Fish Biol Fisheries 24: 333–368.

Kew, M., 1992 Salmon Availability, Technology, and Cultural Adaptation in the Fraser River Watershed, in *A Complex Culture of the British Columbia Plateau*, edited by B. Hayden. UBC Press, Vancouver, BC.

Kolde, R., 2019 pheatmap: Pretty Heatmaps.

Larson, W. A., M. T. Limborg, G. J. McKinney, D. E. Schindler, J. E. Seeb *et al.*, 2017 Genomic islands of divergence linked to ecotypic variation in sockeye salmon. Molecular Ecology 26: 554–570.

Latch, E. K., J. R. Heffelfinger, J. A. Fike, and O. E. RHODES Jr, 2009 Species-wide phylogeography of North American mule deer (Odocoileus hemionus): cryptic glacial refugia and postglacial recolonization. Molecular Ecology 18: 1730–1745.

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio].

Li, H., and R. Durbin, 2010 Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589–595.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

Liberles, S. D., 2015 Trace amine-associated receptors: ligands, neural circuits, and behaviors. Current Opinion in Neurobiology 34: 1–7.

Lotterhos, K. E., and M. C. Whitlock, 2015 The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology 24: 1031–1046.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297–1303.

McPhail, J. D., 2007 *The Freshwater Fishes of British Columbia*. University of Alberta.

McRae, B. H., 2006 ISOLATION BY RESISTANCE. Evolution 60: 1551–1561.

Meek, M. H., M. R. Baerwald, M. R. Stephens, A. Goodbla, M. R. Miller *et al.*, 2016 Sequencing improves our ability to study threatened migratory species: Genetic population assignment in California's Central Valley Chinook salmon. Ecology and Evolution 6: 7706–7716.

Miettinen, A., J. Dannewitz, S. Palm, E. P. Ahi, A. Romakkaniemi *et al.*, 2023 Loci associated with maturation, migration and appetite control are linked with signals of fine-scale local selection in a large Atlantic salmon population. 2023.08.23.553800.

Milne, D. J., 1964 The Chinook and Coho Salmon Fisheries of British Columbia: Fisheries Research Board of Canada 142.

Min, K.-D., M. Asakura, M. Shirai, S. Yamazaki, S. Ito *et al.*, 2021 ASB2 is a novel E3 ligase of SMAD9 required for cardiogenesis. Sci Rep 11: 23056.

Nadeau, S., P. G. Meirmans, S. N. Aitken, K. Ritland, and N. Isabel, 2016 The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecology and Evolution 6: 8649–8664.

Nelson, R., M. Small, T. Beacham, and J. Supernault, 2001 Population structure of Fraser River chinook salmon (Oncorhynchus tshawytscha): An analysis using microsatellite DNA markers. Fishery Bulletin 99: 94–107.

Nichols, K. M., C. C. Kozfkay, and S. R. Narum, 2016 Genomic signatures among Oncorhynchus nerka ecotypes to inform conservation and management of endangered Sockeye Salmon. Evol Appl 9: 1285–1300.

Northcote, T. G., and D. Y. Atagi, 1997 Pacific Salmon Abundance Trends in the Fraser River Watershed Compared with Other British Columbia Systems, pp. 199–219 in *Pacific Salmon & their Ecosystems: Status and Future Options*, edited by D. J. Stouder, P. A. Bisson, and R. J. Naiman. Springer US, Boston, MA.

Pess, G. R., R. Hilborn, K. Kloehn, and T. P. Quinn, 2012 The influence of population dynamics and environmental conditions on pink salmon (Oncorhynchus gorbuscha) recolonization after barrier removal in the Fraser River, British Columbia, Canada. Can. J. Fish. Aquat. Sci. 69: 970–982.

Pew, J., P. H. Muir, J. Wang, and T. R. Frasier, 2015 related: an R package for analysing pairwise relatedness from codominant molecular markers. Molecular Ecology Resources 15: 557–561.

PLINK 1.9 cog-genomics.org.

QGIS Development Team, 2022 QGIS_software.

Quinn, T. P., 1993 A review of homing and straying of wild and hatchery-produced salmon. Fisheries Research 18: 29–44.

R Core Team, 2022 R: A Language and Environment for Statistical Computing.

Reynoldson, T. B., J. Culp, R. Lowell, and J. S. Richardson, 2005 Fraser River Basin, pp. 696–732 in *Rivers of North America*, edited by A. C. Benke and C. E. Cushing. Academic Press, Burlington.

Ritland, K., 1996 Estimators for pairwise relatedness and individual inbreeding coefficients. Genetics Research 67: 175–185.

Robertson, A., 1965 The interpretation of genotypic ratios in domestic animal populations. Animal Science 7: 319–324.

Rondeau, E. B., K. A. Christensen, D. R. Minkley, J. S. Leong, M. T. T. Chan *et al.*, 2023 Population-size history inferences from the coho salmon (Oncorhynchus kisutch) genome. G3 Genes|Genomes|Genetics 13: jkad033.

Rougemont, Q., T. Leroy, E. B. Rondeau, B. Koop, and L. Bernatchez, 2023 Allele surfing causes maladaptation in a Pacific salmon of conservation concern. PLOS Genetics 19: e1010918.

Rougemont, Q., J.-S. Moore, T. Leroy, E. Normandeau, E. B. Rondeau *et al.*, 2020 Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. PLOS Genetics 16: e1008348.

Saravanan, K. A., M. Panigrahi, H. Kumar, B. Bhushan, T. Dutt *et al.*, 2020 Selection signatures in livestock genome: A review of concepts, approaches and applications. Livestock Science 241: 104257.

Sinclair-Waters, M., T. Nome, J. Wang, S. Lien, M. P. Kent *et al.*, 2022 Dissecting the loci underlying maturation timing in Atlantic salmon using haplotype and multi-SNP based association methods. Heredity 129: 356–365.

Slowikowski, K., 2021 ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2."

Small, M. P., T. D. Beacham, R. E. Withler, and R. J. Nelson, 1998 Discriminating coho salmon (Oncorhynchus kisutch) populations within the Fraser River, British Columbia, using microsatellite DNA markers. Molecular Ecology 7: 141–155.

Soltis, D. E., M. A. Gitzendanner, D. D. Strenge, and P. S. Soltis, 1997 Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Syst. Evol. 206: 353–373.

Srivastava, R. K., J. A. Brown, and M. E. Brosnan, 1992 The presence of polyamines during embryonic development of arctic charr, *Salvelinus alpinus*. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 101: 153–157.

Stewart, J. R., A. M. Lister, I. Barnes, and L. Dalén, 2009 Refugia revisited: individualistic responses of species in space and time. Proceedings of the Royal Society B: Biological Sciences 277: 661–671.

Teel, D. J., G. B. Milner, G. A. Winans, and W. S. Grant, 2000 Genetic Population Structure and Origin of Life History Types in Chinook Salmon in British Columbia, Canada. Transactions of the American Fisheries Society 129: 194–209.

Terhorst, J., J. A. Kamm, and Y. S. Song, 2017 Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet 49: 303–309.

Tessarolo, J. A., M. J. Tabesh, M. Nesbitt, and W. S. Davidson, 2014 Genomic Organization and Evolution of the Trace Amine-Associated Receptor (TAAR) Repertoire in Atlantic Salmon (Salmo salar). G3 Genes|Genomes|Genetics 4: 1135–1141.

Tribe, S., 2005 Eocene paleo-physiography and drainage directions, southern Interior Plateau, British Columbia. Can. J. Earth Sci. 42: 215–230.

Utter, F. M., F. W. Allendorf, and H. O. Hodgins, 1973 Genetic Variability and Relationships in Pacific Salmon and Related Trout Based on Protein Variations. Systematic Biology 22: 257–270.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel *et al.*, 2013 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43: 11.10.1-11.10.33.

Veale, A. J., and M. A. Russello, 2017 Genomic Changes Associated with Reproductive and Migratory Ecotypes in Sockeye Salmon (Oncorhynchus nerka). Genome Biol Evol 9: 2921–2939.

Wang, J., E. Santiago, and A. Caballero, 2016 Prediction and estimation of effective population size. Heredity 117: 193–206.

Warner, B. G., R. W. Mathewes, and J. J. Clague, 1982 Ice-Free Conditions on the Queen Charlotte Islands, British Columbia, at the Height of Late Wisconsin Glaciation. Science 218: 675–677.

Weber, J. N., G. S. Bradburd, Y. E. Stuart, W. E. Stutz, and D. I. Bolnick, 2017 Partitioning the effects of isolation by distance, environment, and physical barriers on genomic divergence between parapatric threespine stickleback. Evolution 71: 342–356.

Wehrhahn, C. F., and R. Powell, 1987 Electrophoretic Variation, Regional Differences, and Gene Flow in the Coho Salmon (Oncorhynchus kisutch) of Southern British Colombia. Can. J. Fish. Aquat. Sci. 44: 822–831.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-Statistics for the Analysis of Population Structure. Evolution 38: 1358–1370.

Wickham, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., 2007 Reshaping Data with the reshape Package. Journal of Statistical Software 21: 1–20.

Withler, F. C., 1982 Transplanting Pacific salmon: Canada Department of Fisheries and Oceans Canadian technical report of fisheries and aquatic sciences 1488-5379 1079.

Withler, R. E., K. D. Le, R. J. Nelson, K. M. Miller, and T. D. Beacham, 2000 Intact genetic structure and high levels of genetic diversity in bottlenecked sockeye salmon (Oncorhynchus nerka) populations of the Fraser River, British Columbia, Canada. Can. J. Fish. Aquat. Sci. 57: 1985–1998.

Wood, C. C., B. E. Riddell, D. T. Rutherford, and R. E. Withler, 1994 Biochemical Genetic Survey of Sockeye Salmon (Oncorhynchus nerka) in Canada. Can. J. Fish. Aquat. Sci. 51: 114–131.

Wright, M., 2022 A typology of hydraulic barriers to salmon migration in a bedrock river [Masters]: Simon Fraser University, 57 p.

Wright, S., 1943 Isolation by Distance. Genetics 28: 114–138.

Xuereb, A., Q. Rougemont, X. Dallaire, J.-S. Moore, E. Normandeau *et al.*, 2022 Re-evaluating Coho salmon (Oncorhynchus kisutch) conservation units in Canada using genomic data. Evolutionary Applications 15: 1925–1944.

Yamak, A., D. Hu, N. Mittal, J. W. Buikema, S. Ditta *et al.*, 2020 Loss of Asb2 Impairs Cardiomyocyte Differentiation and Leads to Congenital Double Outlet Right Ventricle. iScience 23: 100959.

Zachos, J. C., G. R. Dickens, and R. E. Zeebe, 2008 An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature 451: 279–283.

Figure Legends

Figure 1. Fraser River salmon sampling locations. a) Each point on the map represents a sockeye salmon sampling site (average of 6.5 samples per location). Each location is shown as the closest point of the specified body of water to the Fraser River. The Fraser River and associated watersheds are highlighted on the map. The Nechako River is highlighted as the top left watershed (darker blue, with the river as a lighter blue), and the Thompson River is highlighted as the bottom right watershed (green). Watershed and geographical data are from ced.org and naturalearthdata.com, respectively. The map was generated using QGIS software. Information on sampling locations outside of the boundaries of this map can be found in File S1. b) Same as a) for Chinook salmon samples (average of 7.8 samples per site). c) Same as a) for coho salmon samples (average of 6.9 samples per site – excluding the Big Bar Landslide sampling site).

Figure 2. Polymorphic loci per species and runs of homozygosity. a) Counts of all SNPs identified for each species with the same pipeline – differing only in score recalibration and reference genome assemblies. The number of samples per species is shown below each column. b) Overlapping SNP loci among species. SNP loci from sockeye and coho salmon were mapped to the Chinook salmon genome assembly. The percent of loci that overlap are shown for each species relative to the number of SNPs identified in that species. c) Boxplots of the total lengths of runs of homozygosity for each species. d) A scatterplot of runs of homozygosity and the fraction of heterozygous genotypes per individual. The sockeye sample on the far left is a doubled-haploid. Lines were fitted using the loess method.

Figure 3. SNP coverage and relatedness metrics. a) A scatterplot of the relationship between the average SNP coverage and the percent of the genotypes that were heterozygous for each salmon (all species). The relationship between the two variables was modelled using the loess method. The vertical line shows 15x coverage, which was used as a threshold in other analyses. b) Same as a), but each salmon is identified by species. c) Heatmap of pairwise relatedness scores between each Chinook salmon. The salmon from a previous study are highlighted under the Illumina banner. A boxplot of SNP depth is displayed in an insert for the previous (Illumina) and current datasets (MGI). Samples from a Thompson River tributary are highlighted from the previous dataset. d) Same as c) for the coho salmon datasets.

Figure 4. Comparison of Fraser River salmon admixture groups. a) (left) Map of sockeye salmon admixture groups plotted in QGIS by overlaying the admixture ancestry raster plots from Figure S3. Each group is labelled (the colour of the group was randomly assigned). The northern coastal admixture group is outside the frame of this map. (right) Average percent of heterozygous genotypes for each of the Fraser River admixture groups. The different Fraser River admixture group names are lower Fraser River – LFR, mid Fraser River – MFR, and upper Fraser River – UFR. The number of samples (≥15x SNP coverage) and sampling sites per admixture group is shown. Weir and Cockerham's (Weir and Cockerham 1984) F_{st} are shown for comparisons between admixture groups. F_{st} comparisons between coastal (LFR) and interior clusters (MFR and UFR) are in bold text. b) Same as a) for Chinook salmon. The LFR group includes coastal sites outside the Fraser River. Significant differences in the percent of heterozygous genotypes among groups are shown with the *p-value* (two-tailed Welch's t-test). c) Same as b) for coho salmon.

Figure 5. PCA of sampling location environmental factors. a) A biplot of a PCA of environmental factors from each sockeye salmon sampling site. These included environmental factors from WorldClim version 2.1, elevation, and distance to the ocean. Environmental factors were simplified as temperature (including: annual mean temperature, minimum temperature of the coldest month, maximum temperature of the warmest month, and temperature range), precipitation (including: driest month precipitation, wettest month precipitation, and annual precipitation), elevation, and distance to the ocean. Figure S5 does not simplify these variables. Locations were coloured by their assignment to admixture genetic groups, but no genetic data was used to produce these PCA. Only sites that had average ancestry values ≥ 0.7 from the LFR, MFR, and UFR genetic groups are shown for simplicity. See Figure S5 for all locations. Each small symbol represents a sampling site, and the larger symbol represents the centre of the ellipse if drawn (ellipse level 0.4). b) Same as a), except for coho salmon. The LFR group includes coastal sites outside the Fraser River. c) Same as b), except for Chinook salmon.

Figure 6. Estimated historical effective population size of Fraser River salmon. a) Effective population size of Chinook salmon estimated using SMC++ and whole genome data. Only one sampling site from each Fraser River admixture group was retained for clarity. Figure S6 shows all sites. The estimated glacial maximum of around 19,000 years ago (Clark *et al.* 2009) is shown with a dashed vertical line. Other regions are highlighted when there were major changes in effective population size for more than one species. b) Same as a) for coho salmon. c) Same as a) for sockeye salmon. d) Estimates of ocean surface temperature (Zachos *et al.* 2008; Hansen *et al.* 2013).

Figure 7. Estimated historical effective population sizes of some admixture groups of sockeye salmon. a) Effective population size of sockeye sampling locations from the UFR admixture group

estimated using SMC++ and whole genome data. b) Same as a), but for sites from the MFR admixture group. c) Same as a), but for locations from the northern coastal admixture group. d) Same as a), but for locations from the upper Columbia River kokanee admixture group.

Figure 8. Genetic diversity of Fraser River salmon. a) (left) Map of sockeye salmon total length (kb) of runs of homozygosity (ROH) averaged for each sampling site. This analysis included all individuals. The colour scale is quantile-based. (middle-left) Barplot of total lengths of runs of homozygosity for samples assigned to the different admixture groups with ancestry values ≥ 0.7. (middle-right) Map of sockeye salmon nucleotide diversity (pi) of each sampling site excluding salmon with less than 15x SNP coverage. The colour scale is quantile-based. (right) Nucleotide diversity barplots of sampling sites with average admixture ancestry values ≥ 0.7 from individuals with at least 15x SNP coverage. b) Same as a) for Chinook salmon. The LFR group contains samples outside the Fraser River. c) Same as b) for coho salmon.

Figure 9. Private allele analyses of Fraser River genetic groups. a) The count of private alleles of Fraser River admixture groups (only individuals with ≥ 0.7 ancestry values were used). Only private alleles identified in at least three individuals from a group were counted. Below the private allele count is the percent that this represents of all the SNPs identified for each species. The circles are scaled by this value. The number of samples per genetic group after these criteria were met are also included. b) Boxplots, with jittered points added, of individual private allele counts of the different Chinook salmon admixture groups. The individual private allele counts are the number of private alleles each individual has from the private alleles common to the admixture group. This is a way to compare how much individuals contribute to the identified private alleles of the group, and to be able to compare these values to understand if sample size might influence group counts. The LFR group has individuals from outside the Fraser River. c) Same as b) for sockeye salmon, except there were five total admixture groups of sockeye salmon and the LFR is strictly Fraser River samples. d) Same as b) for coho salmon.

Tables

Table 1. SNP filtering parameters used for each analysis.

Analyses	Indels, >10% missing, depth	MAF 0.01	LD Filter	15x coverage	Relatedness
SMC++	all				
Rehh	all	all			
Coverage	all	all			
ROH	all	all			
Relatedness	all	all	all		
Admixture	all	all	all		
% Heterozygous	all	all		all	
Nucleotide diversity	all	all		all	
Polymorphic loci	all	all		all	
PCA	all	all	all	some*	some*

*See Figure S4 for all PCAs, including some with the 15x SNP coverage and relatedness filters.

Table 2. Significant extended haplotype homozygosity among admixture groups.

Species	Admixture	Admixture	Admixture	Nearest gene

	comparison LFR vs. MFR	comparison LFR vs. UFR	comparison MFR vs. UFR	
Chinook salmon	LG09:14.36-14.37	LG09:14.36-14.37		CSGALNACT1*
	LG10:5.22-5.23			ENDOD1*
	LG11:36.37-36.38			MACO2*
	LG21:43.68-43.69			CTNNA1
			LG01:54.33-54.34	HCA2
			LG07:46.47-46.48	SRGAP3*
			LG10:2.25-2.26	AIMP1B*
			LG13:37.80-37.81	EPHA7*
			LG13:37.98-37.99	LOC121838999
			LG18:25.30-25.31	METTL24*
			LG18:33.39-33.40	FOSL2
			LG25:28.04-28.05	LRPPRC*
			LG26:1.26-1.27	KLF12B*
			LG30:17.88-17.90	MINK1
Coho salmon	LG18:24.62-24.63			ASB2*
			LG10:24.15-24.16	PTPRD
Sockeye salmon	LG26:45.73-45.74			TAAR13C*
		LG09:23.72-23.73		TAAR13C*
			LG11:31.22-31.53	Two genes overlap
			LG04:28.48-31.57	Many genes

All locations are relative to the Chinook salmon genome assembly for comparison (see File S3 for species specific information). Format of location: Chinook salmon linkage group : start (distance in Mb) – end (distance in Mb). *Location overlaps with gene.

Supporting Information

Figure S1. Expanded maps and sampling locations. Slide 1) Sockeye salmon maps, Slide 2) Chinook salmon map, and Slide 3) Coho salmon maps.

Figure S2. Relationship between average SNP coverage and the total length of runs of homozygosity. a) The average SNP coverage per sockeye salmon sample relative to the total length of runs of homozygosity per sample. The line was determined using the loess function (local regression). b) Same as a) for Chinook salmon. c) Same as a) for coho salmon.

Figure S3. Admixture analyses. Slide 1) Average admixture ancestry values of the different admixture groups were plotted for sampling locations in QGIS using the inverse distance weighted interpolation method. The darker the sampling location, the greater the ancestry assignment to the specified admixture group. Slide 2) Plot of admixture ancestry values for clusters above and below the best supported k=5 for sockeye salmon. Slide 3) Plot of admixture ancestry values for clusters above and below the best supported k=3 for Chinook salmon. The arrows indicate that the name of a location was omitted because of space limitations. Slide 4) Plot of admixture ancestry values for clusters above and below the best supported k=3 for coho salmon. The arrows indicate that the name of a location was omitted because of space limitations.

Figure S4. PCA of each species with different filtering criteria. a) PCA of sockeye salmon. The first column is a PCA of all samples, the second column is a PCA with only one individual from pairs with high relatedness scores (≥ 0.15), and the third column is a PCA of samples with high SNP coverage (>15x coverage). The different admixture groups are noted in the PCA. Admixture numbers were arbitrary, so names were given based on geography for the first column PCA. Individuals with admixture ancestry values < 0.7, were given a different colour from the the main groups. b) Same as a) for Chinook salmon. c) Same as a) for coho salmon.

Figure S5. PCA of sampling locations based on environmental factors. Slide 1) Biplot of sockeye salmon sampling sites based on environmental variables from WorldClim version 2.1 and elevation. Sampling sites with an average ancestry assignment of less than 0.7 were assigned to intermediate genetic groups with the highest fraction named first (e.g., if a site had average ancestry fractions of 0.6 MFR and 0.4 LFR, it was assigned Mid-Coast). The Okanagan was an exception as it had high ancestry values for multiple genetic groups. Each sampling location is represented by a small point (unless there was only one site, in which case it has a larger symbol), larger points show the middle of ellipses, and ellipses were drawn at the 0.4 level. The average annual precipitation and distance to the ocean are shown on the right. Groups were combined if they shared a genetic group as the largest contributor to their ancestry values. Slide 2) Same as Slide 1, but for coho salmon, Slide 3) Same as Slide 1 but for Chinook salmon.

Figure S6. Estimates of historical effective population size for different salmon species and groups. Slide 1) Effective population size for all species and sampling sites. Some Chinook sampling locations are off the graph. These sites may have recent admixture, which can throw off these estimates. These sites were removed in all other figures. Slides 2-3) Fraser River admixture group comparisons of each species. Slides 4-5) Effective population size estimations of different groups of Chinook salmon. Individuals were grouped by admixture and region (Coastal: Northern BC Coast, Vancouver Island, Lower Fraser River; Mid Fraser: Mid-Upper Fraser River, Mid-Coast Fraser River, Mid Fraser River; Upper Fraser: Upper Fraser River, Upper Mid Fraser River). Slides 6-7) Effective population size estimations of different groups of coho salmon (Coastal: Mid Alaska, Northern BC/Southern Alaska, Southern BC, Coastal Fraser River, California-Oregon; Mid Fraser: Mid Fraser-Coastal Fraser, Mid Fraser; Upper Fraser River). Slides 8-9) Effective population size estimations of different groups of sockeye salmon (Coastal: Northern Coast, Coast-Mid Fraser, Coast-Coast Fraser, Coast Fraser; Mid Fraser: Mid-Coast Fraser, Mid-Upper Fraser, Mid Fraser; Upper Fraser: Upper Fraser – Columbia, Upper Fraser River; Columbia: Upper Columbia River Kokanee, Upper Columbia – Upper Fraser; Okanagan Lake).

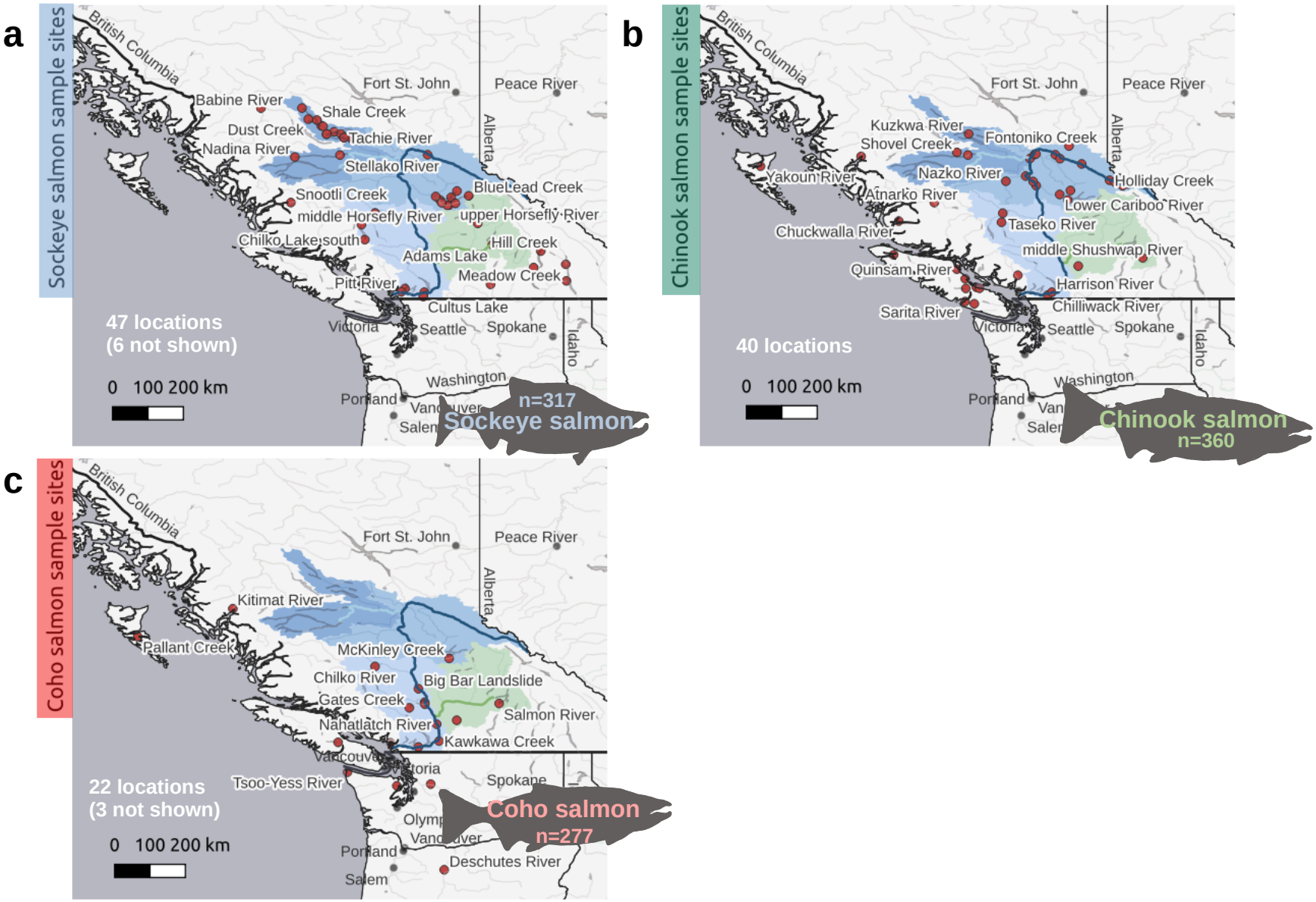
Figure S7. Total runs of homozygosity for each sample location. Slide 1) Boxplot of total runs of homozygosity (kb) for all individuals from each sample location of Chinook salmon. Individuals were also highlighted by Admixture group (individuals with ancestry values < 0.7 were represented by mixes of admixture groups with the largest contributor first). The x-axis represents the sampling site (see File S1 for full name) and the y-axis represents the total length of runs of homozygosity within each individual genome. Slide 2) same as slide 1 except for coho salmon. Slide 3) same as slide 1 except for sockeye salmon.

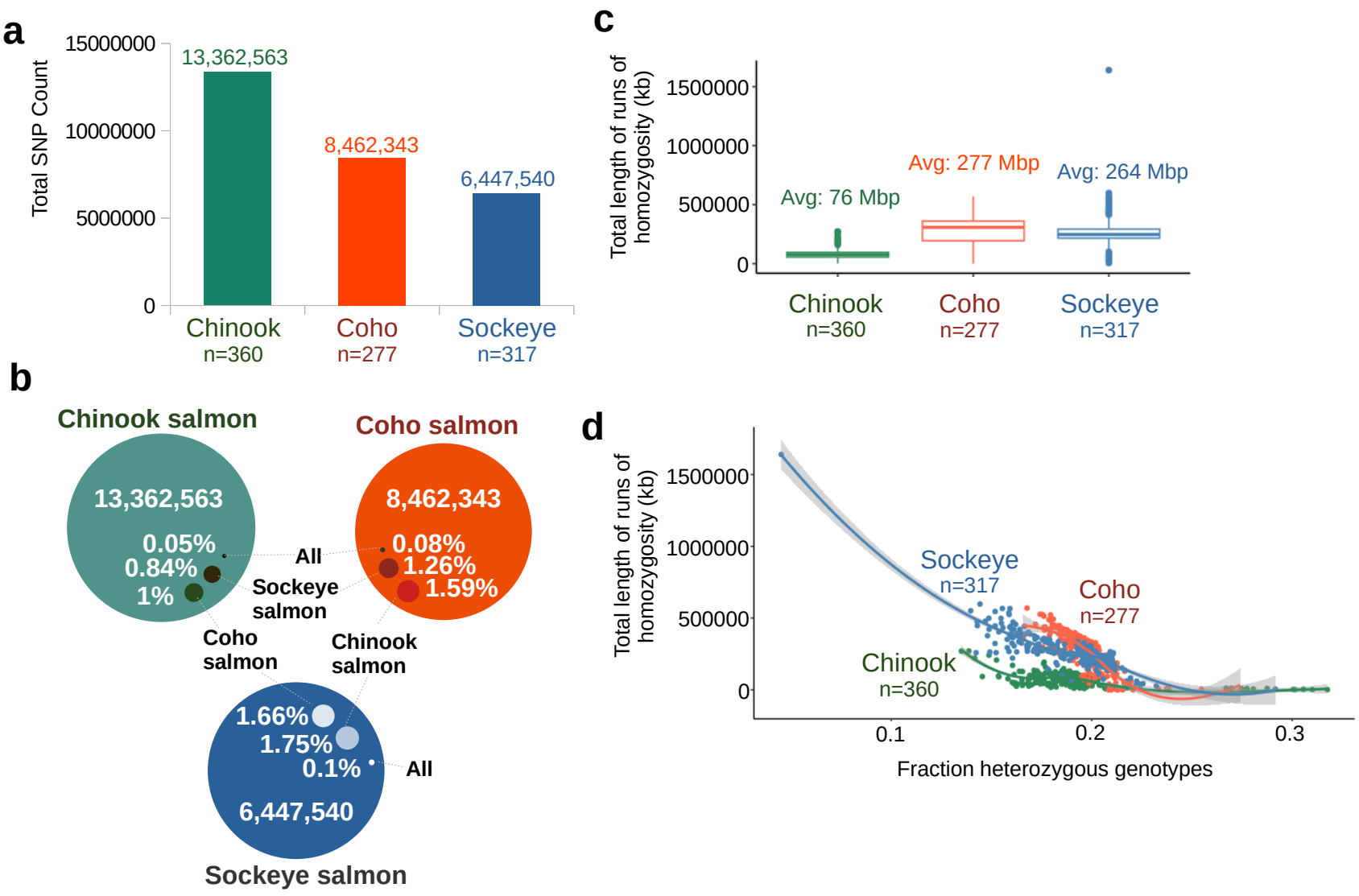
Figure S8. Number of polymorphic loci for each sample location. a) The number of polymorphic loci per location is shown on the top for all species. The number of samples per location is shown on the bottom. b) The relationship between the number of samples per location and the number of

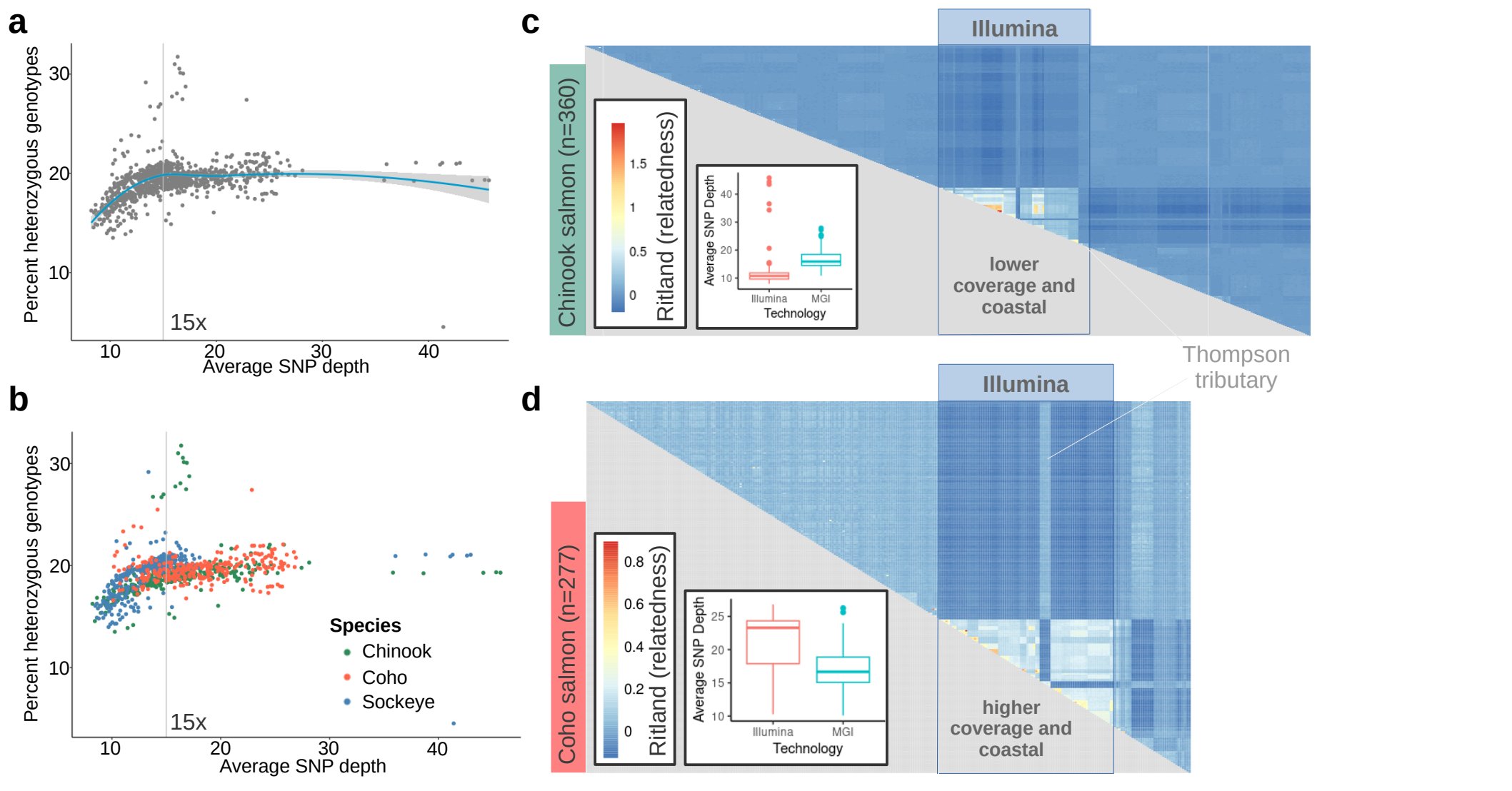
polymorphic loci identified for those locations. c) A boxplot of the number of polymorphic loci based on admixture group. Locations with average admixture group ancestry values ≥ 0.7 only have one label (i.e., lower Fraser River – LFR, mid Fraser River – MFR, and upper Fraser River – UFR). Those with < 0.7 have two labels with the first having the largest ancestry value. Only locations with at least four samples were used for this analysis. d) Relationship between the number of polymorphic loci and nucleotide diversity for locations with at least four individuals per site. Each species has its own linear regression line plotted.

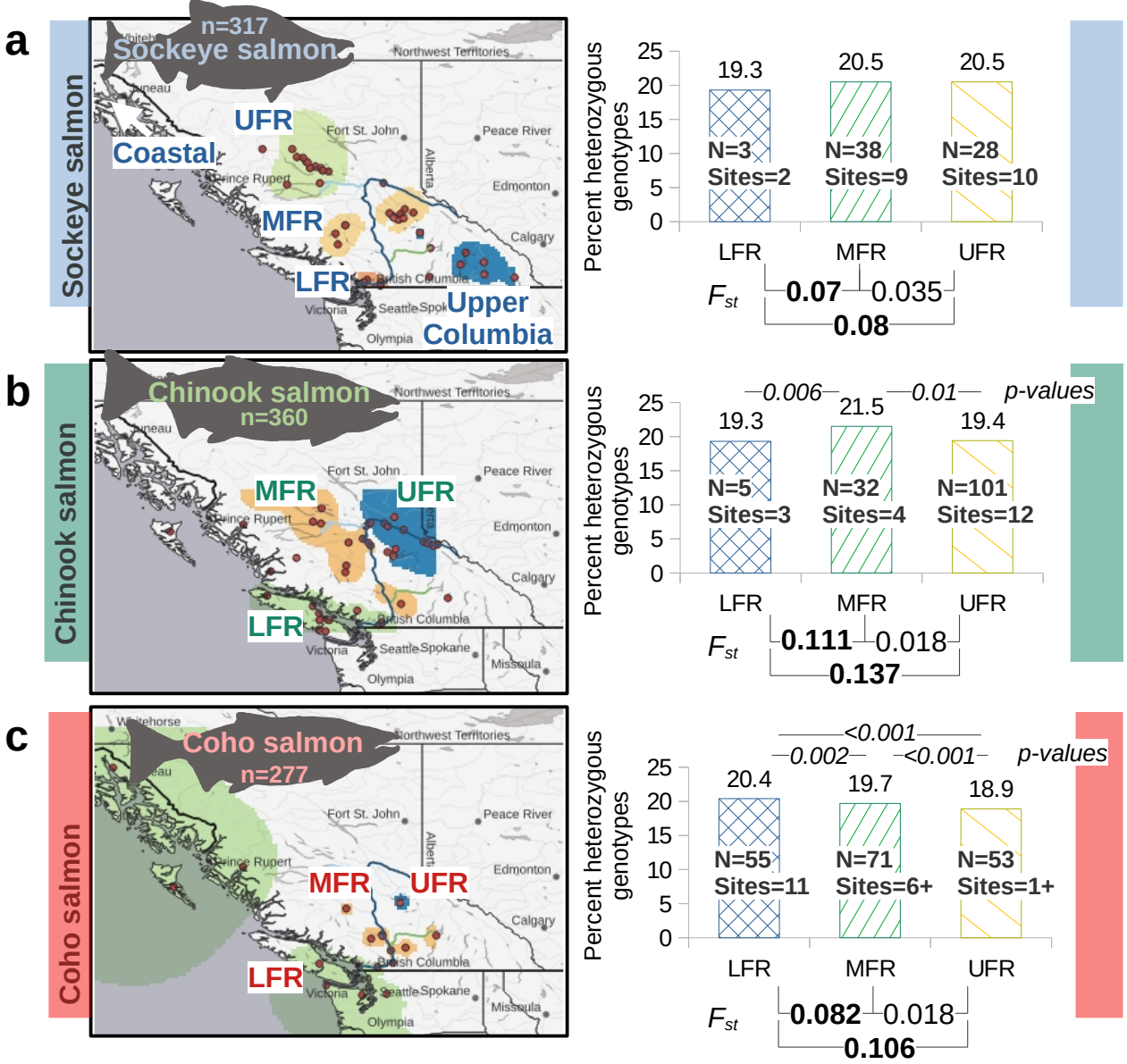
Figure S9. Extended haplotype homozygosity comparisons between LRF and MFR admixture groups. a) A manhattan plot of -log10 p-values from an Rsb analysis between LFR and MFR sockeye salmon admixture groups. Sockeye salmon chromosomes were based on a draft genome assembly submitted to the NCBI (now GCA_034236695.1). Significant peaks were indicated on the graph with a vertical line and a number. b) Same as a) for Chinook salmon. c) Same as a) for coho salmon.

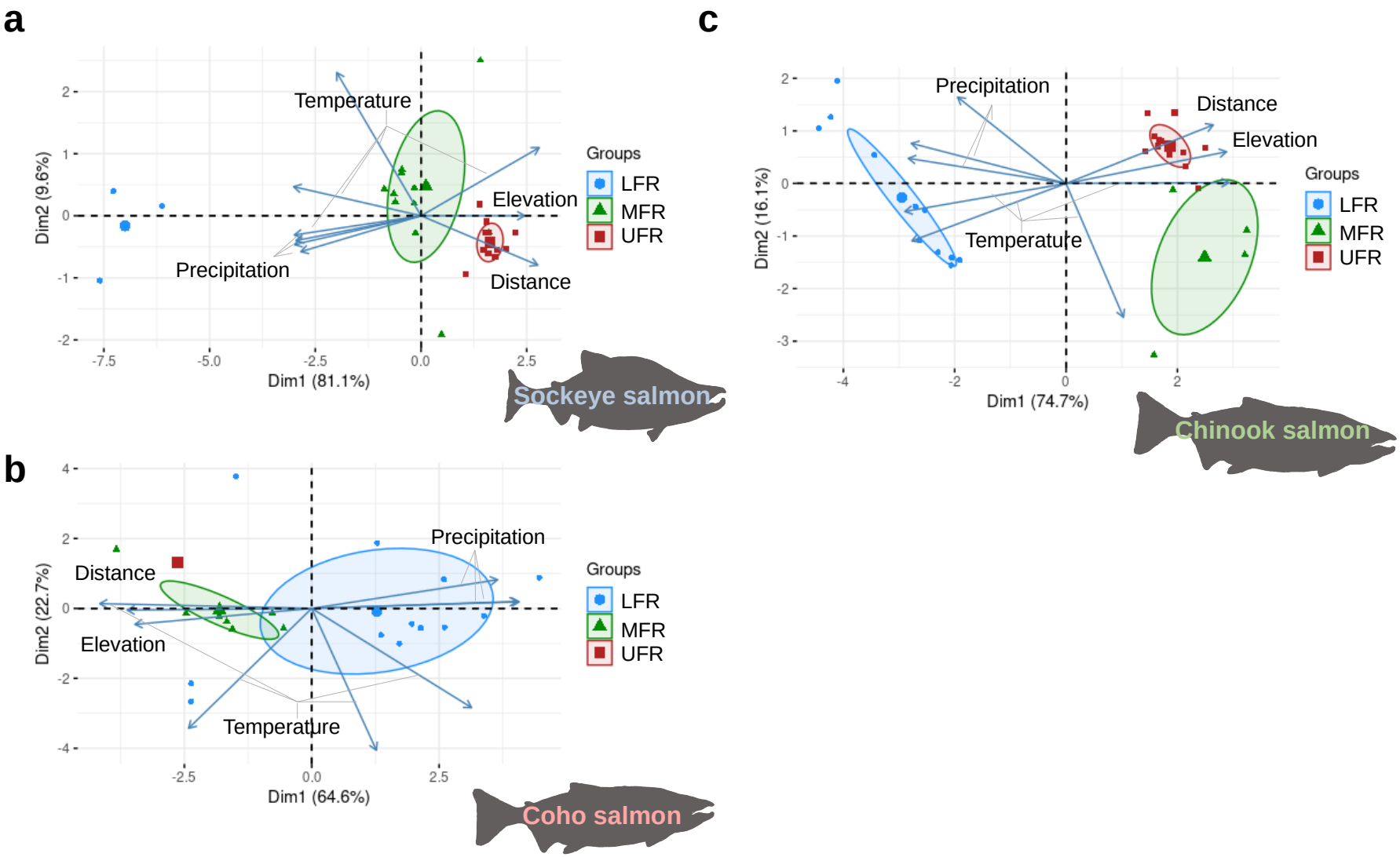
File S1. Sample summary information in a spreadsheet format.
File S2. Truth SNPs used in GATK recalibration for each species (includes a readme file).
File S3. Significant extended haplotype homozygosity from all comparisons among groups.

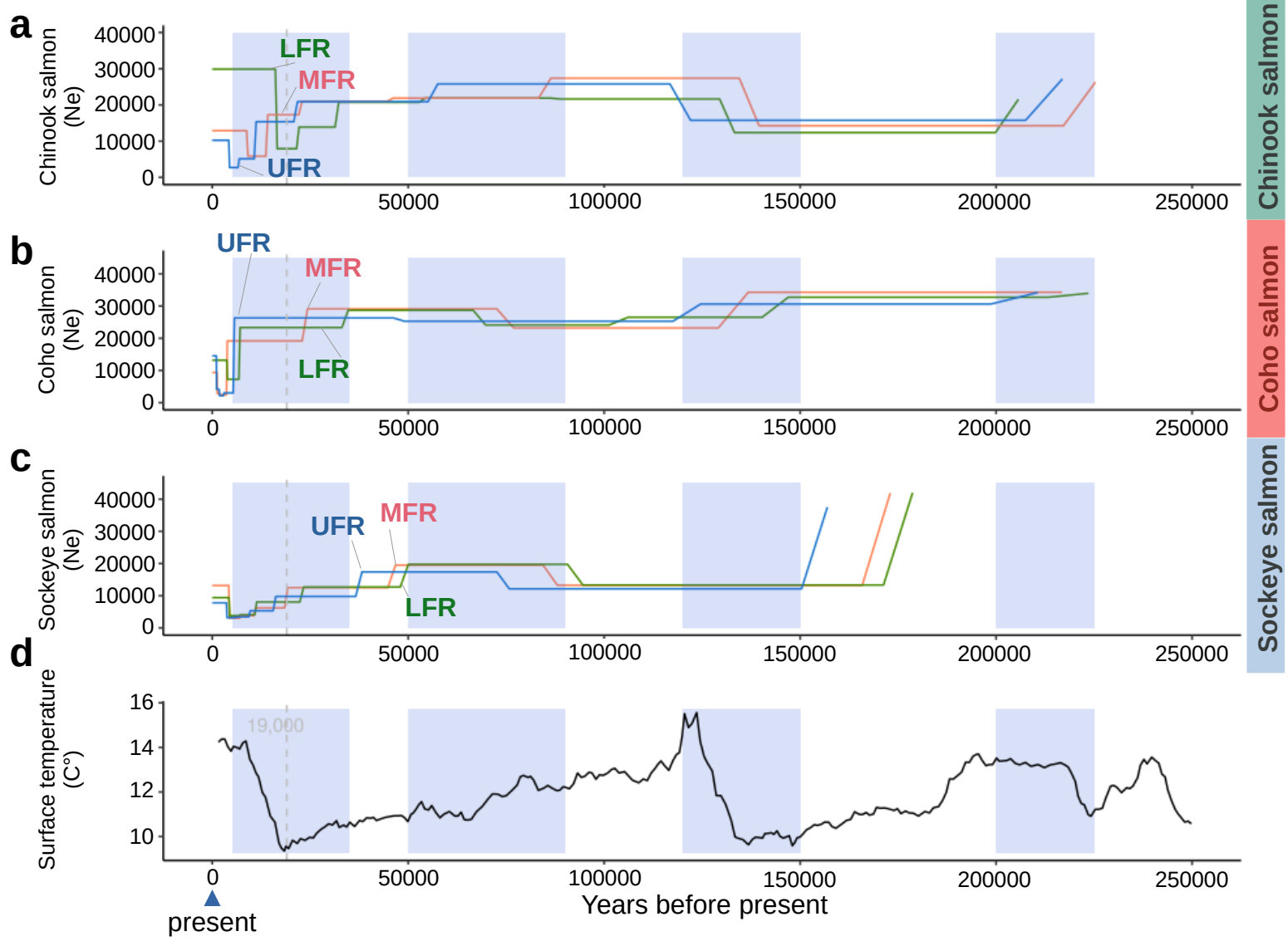

a
Sockeye salmon sample sites
British Columbia
Fort St. John
Peace River
Babine River
Shale Creek
Dust Creek
Tachie River
Nadina River
Stellako River
Alberta
Snootli Creek
Blue Lead Creek
middle Horsefly River
upper Horsefly River
Chilko Lake south
Hill Creek
Adams Lake
Meadow Creek
Pitt River
Cultus Lake
Victoria
Seattle
Spokane
Idaho
Washington
Portland
Salem
47 locations
(6 not shown)
0 100 200 km
n=317
Sockeye salmon
b
Chinook salmon sample sites
British Columbia
Fort St. John
Peace River
Kuzkwa River
Shovel Creek
Fontoniko Creek
Alberta
Yakoun River
Nazko River
Holliday Creek
Atnarko River
Lower Cariboo River
Taseko River
Chuckwalla River
middle Shushwap River
Quinsam River
Harrison River
Sarita River
Chilliwack River
Victoria
Seattle
Spokane
Idaho
Washington
Portland
Salem
40 locations
0 100 200 km
Chinook salmon
n=360
c
Coho salmon sample sites
British Columbia
Fort St. John
Peace River
Kitimat River
Alberta
Pallant Creek
McKinley Creek
Chilko River
Big Bar Landslide
Gates Creek
Salmon River
Nahatlatch River
Kawkawa Creek
Tsoo-Yess River
Spokane
Portland
Deschutes River
Salem
22 locations
(3 not shown)
0 100 200 km
Coho salmon
n=277

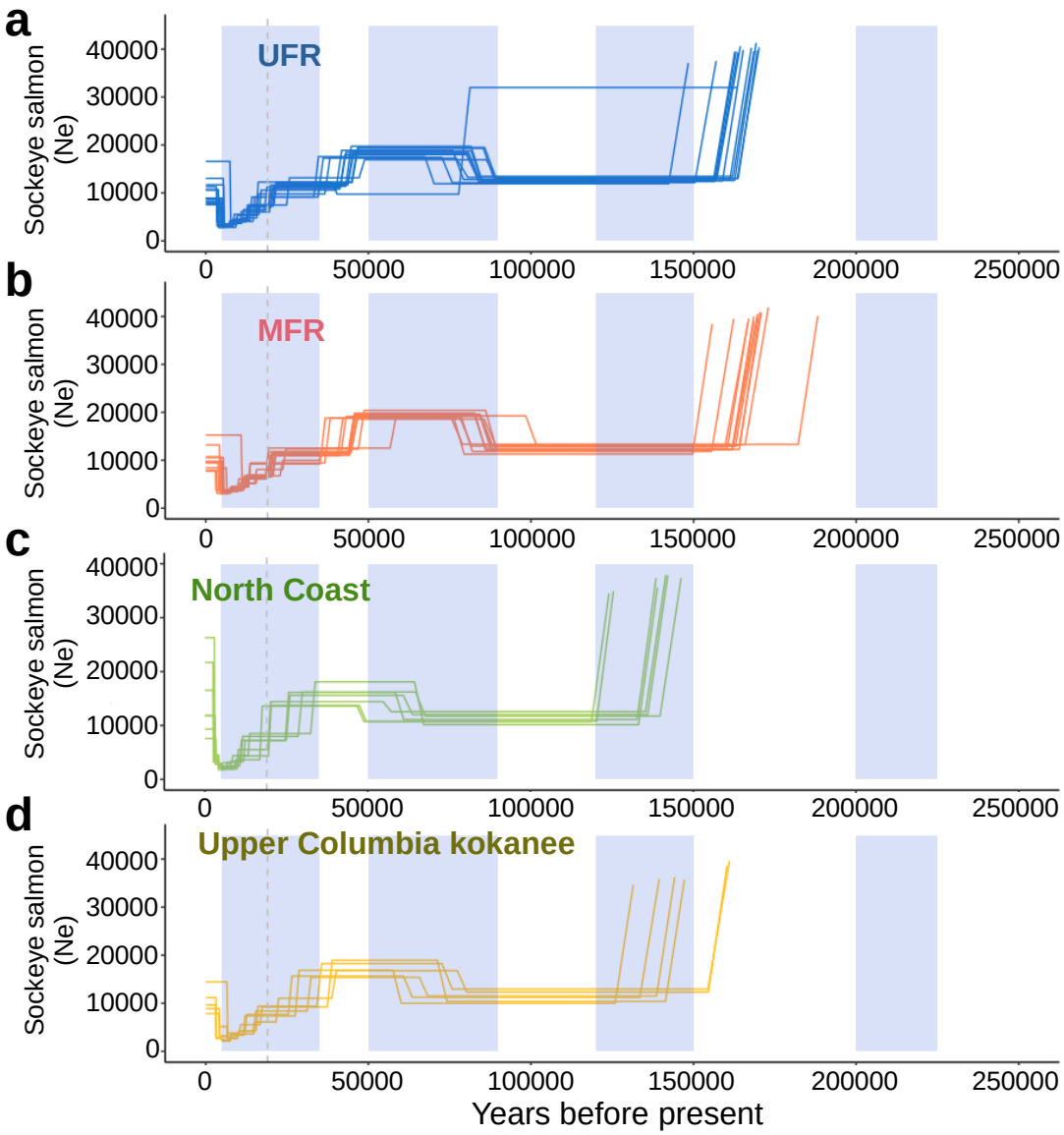

a
Total SNP Count
15000000
10000000
5000000
0
13,362,563
8,462,343
6,447,540
Chinook
n=360
Coho
n=277
Sockeye
n=317
b
Chinook salmon
13,362,563
0.05%
0.84%
1%
All
Sockeye salmon
Coho salmon
Coho salmon
8,462,343
0.08%
1.26%
1.59%
Chinook salmon
1.66%
1.75%
0.1%
All
6,447,540
Sockeye salmon
c
Total length of runs of homozygosity (kb)
1500000
1000000
500000
0
Avg: 76 Mbp
Avg: 277 Mbp
Avg: 264 Mbp
Chinook
n=360
Coho
n=277
Sockeye
n=317
d
Total length of runs of homozygosity (kb)
1500000
1000000
500000
0
Sockeye
n=317
Coho
n=277
Chinook
n=360
0.1
0.2
0.3
Fraction heterozygous genotypes

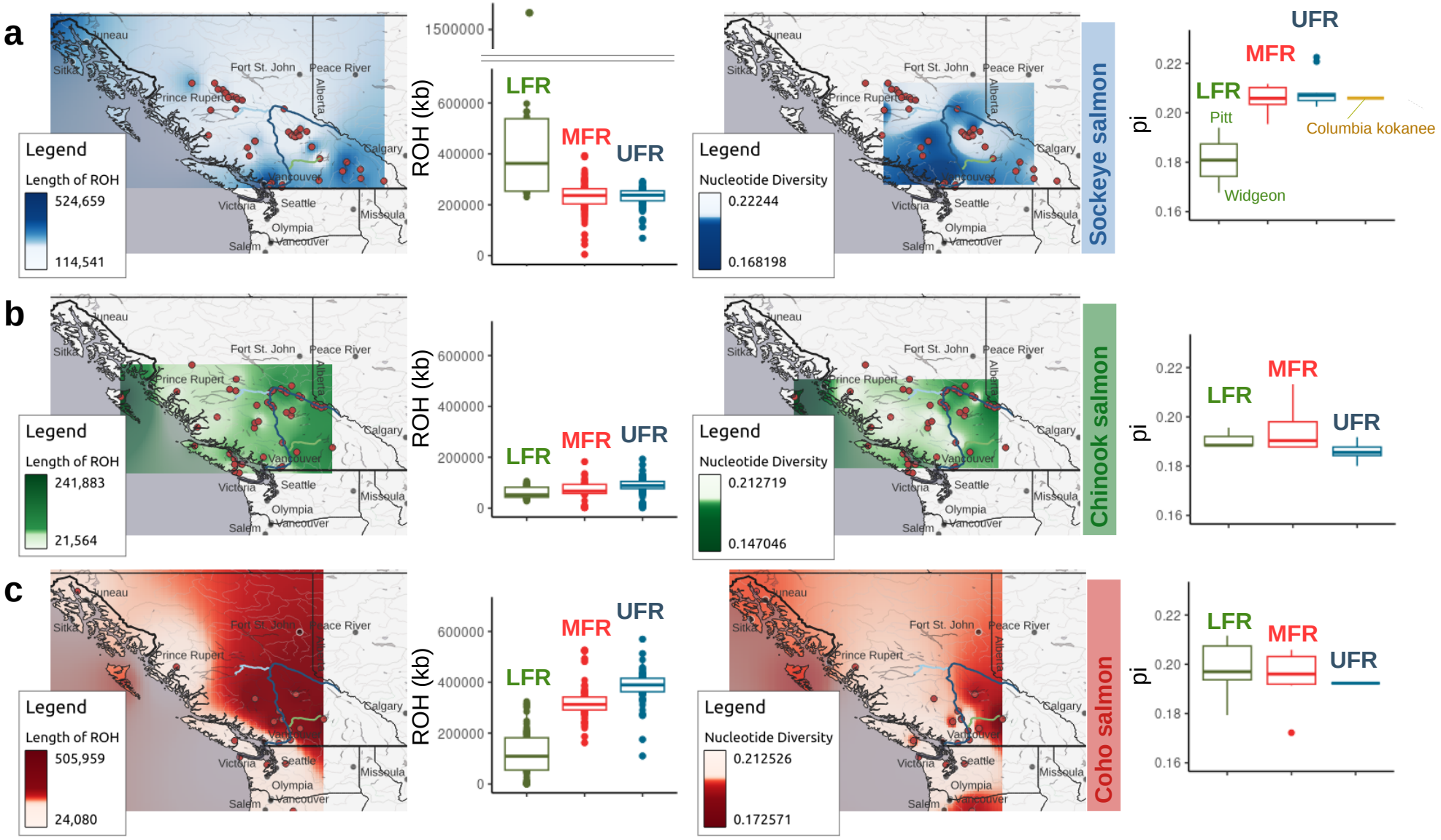

a
Percent heterozygous genotypes
30
20
10
15x
10
20
30
40
Average SNP depth
b
Percent heterozygous genotypes
30
20
10
15x
10
20
30
40
Average SNP depth
Species
Chinook
Coho
Sockeye
c
Illumina
Chinook salmon (n=360)
Ritland (relatedness)
1.5
1
0.5
0
Average SNP Depth
40
30
20
10
Illumina
MGI
Technology
lower coverage and coastal
Thompson tributary
d
Illumina
Coho salmon (n=277)
Ritland (relatedness)
0.8
0.6
0.4
0.2
0
Average SNP Depth
25
20
15
10
Illumina
MGI
Technology
higher coverage and coastal

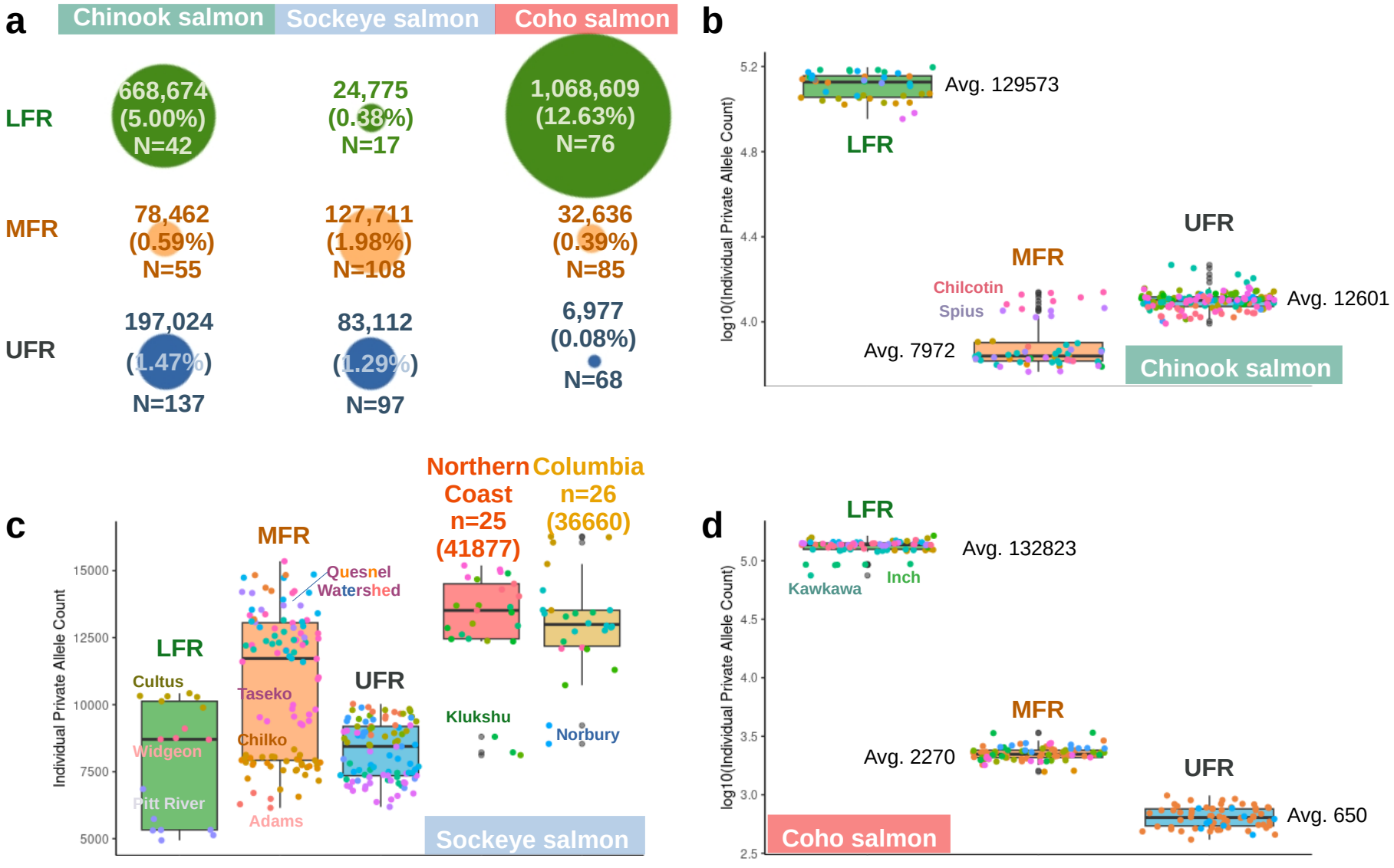

a
Sockeye salmon
n=317
Sockeye salmon
UFR
Coastal
MFR
LFR
Upper Columbia
Northwest Territories
Fort St. John
Peace River
Alberta
Edmonton
Calgary
Prince Rupert
British Columbia
Victoria
Seattle
Spokane
Olympia
Percent heterozygous genotypes
19.3
20.5
20.5
N=3
Sites=2
N=38
Sites=9
N=28
Sites=10
LFR
MFR
UFR
Fst
0.07
0.035
0.08
b
Chinook salmon
n=360
Chinook salmon
MFR
UFR
LFR
Missoula
0.006
0.01
p-values
19.3
21.5
19.4
N=5
Sites=3
N=32
Sites=4
N=101
Sites=12
0.111
0.018
0.137
c
Coho salmon
n=277
Coho salmon
Whitehorse
MFR
UFR
LFR
<0.001
0.002
<0.001
p-values
20.4
19.7
18.9
N=55
Sites=11
N=71
Sites=6+
N=53
Sites=1+
0.082
0.018
0.106


a
Temperature
Elevation
Precipitation
Distance
Groups
LFR
MFR
UFR
Dim2 (9.6%)
Dim1 (81.1%)
Sockeye salmon
b
Precipitation
Distance
Elevation
Temperature
Groups
LFR
MFR
UFR
Dim2 (22.7%)
Dim1 (64.6%)
Coho salmon
c
Precipitation
Distance
Elevation
Temperature
Groups
LFR
MFR
UFR
Dim2 (16.1%)
Dim1 (74.7%)
Chinook salmon


a
Chinook salmon (Ne)
40000
30000
20000
10000
0
LFR
MFR
UFR
0
50000
100000
150000
200000
250000
Chinook salmon
b
Coho salmon (Ne)
UFR
MFR
LFR
Coho salmon
c
Sockeye salmon (Ne)
UFR
MFR
LFR
Sockeye salmon
d
Surface temperature (C°)
16
14
12
10
19,000
present
Years before present


a
UFR
b
MFR
c
North Coast
d
Upper Columbia kokanee
Sockeye salmon (Ne)
40000
30000
20000
10000
0
0
50000
100000
150000
200000
250000
Years before present


a
Legend
Length of ROH
524,659
114,541
ROH (kb)
1500000
600000
400000
200000
0
LFR
MFR
UFR
Legend
Nucleotide Diversity
0.22244
0.168198
Sockeye salmon
pi
0.22
0.20
0.18
0.16
UFR
MFR
LFR
Pitt
Widgeon
Columbia kokanee
b
Legend
Length of ROH
241,883
21,564
ROH (kb)
600000
400000
200000
0
LFR
MFR
UFR
Legend
Nucleotide Diversity
0.212719
0.147046
Chinook salmon
pi
0.22
0.20
0.18
0.16
LFR
MFR
UFR
c
Legend
Length of ROH
505,959
24,080
ROH (kb)
600000
400000
200000
0
LFR
MFR
UFR
Legend
Nucleotide Diversity
0.212526
0.172571
Coho salmon
pi
0.22
0.20
0.18
0.16
LFR
MFR
UFR
Juneau
Sitka
Fort St. John
Peace River
Prince Rupert
Alberta
Calgary
Vancouver
Victoria
Seattle
Missoula
Olympia
Salem


a
Chinook salmon
Sockeye salmon
Coho salmon
LFR
668,674 (5.00%) N=42
24,775 (0.38%) N=17
1,068,609 (12.63%) N=76
MFR
78,462 (0.59%) N=55
127,711 (1.98%) N=108
32,636 (0.39%) N=85
UFR
197,024 (1.47%) N=137
83,112 (1.29%) N=97
6,977 (0.08%) N=68
b
log10(Individual Private Allele Count)
5.2
4.8
4.4
4.0
Avg. 129573
LFR
UFR
MFR
Chilcotin
Spius
Avg. 12601
Avg. 7972
Chinook salmon
c
Northern Coast n=25 (41877)
Columbia n=26 (36660)
Individual Private Allele Count
15000
12500
10000
7500
5000
MFR
Quesnel Watershed
LFR
Cultus
Taseko
Chilko
Widgeon
Pitt River
Adams
UFR
Klukshu
Norbury
Sockeye salmon
d
LFR
Avg. 132823
Kawkawa
Inch
log10(Individual Private Allele Count)
5.0
4.5
4.0
3.5
3.0
2.5
MFR
Avg. 2270
UFR
Avg. 650
Coho salmon