The pink salmon genome: uncovering the genomic consequences of a strict two-year life-cycle

Kris A. Christensen[1,2*], Eric B. Rondeau[1,2,3], Dionne Sakhrani[1], Carlo A. Biagi[1], Hollie Johnson[2], Jay Joshi[2], Anne-Marie Flores[2], Sreeja Leelakumari[4], Richard Moore[4], Pawan K. Pandoh[4], Ruth E. Withler[3], Terry D. Beacham[3], Rosalind A. Leggatt[1], Carolyn M. Tarpey[5], Lisa W. Seeb[5], James E. Seeb[5], Steven J.M. Jones[4], Robert H. Devlin[1], Ben F. Koop[2*]

[1] West Vancouver, Fisheries and Oceans Canada, British Columbia, Canada

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

[3] Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada

[4] Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, British Columbia, Canada

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

[*]Corresponding authors

E-mail: kris.christensen@wsu.edu (KAC), bkoop@uvic.ca (BFK)

Abstract

Pink salmon (*Oncorhynchus gorbuscha*) adults are the smallest of the five Pacific salmon native to the western Pacific Ocean. Pink salmon are also the most abundant of these species and account for a large proportion of the commercial value of the salmon fishery worldwide. A strict two-year life-history of most pink salmon generates temporally isolated populations that spawn either in even-years or odd-years. To uncover the influence of this genetic isolation, reference genome assemblies were generated for each year-class and whole genome re-sequencing data was collected from salmon of both year-classes. The salmon were sampled from six Canadian rivers and one Japanese river. At multiple centromeres we identified peaks of Fst between year-classes that were millions of base-pairs long. The largest Fst peak was also associated with a million base-pair chromosomal polymorphism found in the odd-year genome near a centromere. These Fst peaks may be the result of centromere drive or a combination or reduced recombination and genetic drift, and they could influence speciation. Other regions of the genome influenced by odd-year and even-year temporal isolation and tentatively under selection were mostly associated with genes related to immune function, organ development/maintenance, and behaviour.

Introduction

Pink salmon are an economically important species under heavy exploitation and have been the subject of intense mitigation efforts. Commercial catches of pink salmon comprise roughly half of all Pacific salmon catches by weight and a much greater percentage by count as they are the smallest of the commercially important Pacific salmon (1,2). Since the late 1980s, more than a billion pink salmon are released annually from hatcheries (1) to maintain the abundance of this fishery.

The native range of pink salmon encompasses parts of the southern Arctic Ocean between North America and Asia as well as much of the northern Pacific Ocean (3). Recently, Arctic climate warming has opened previously inaccessible Arctic territory to pink salmon as well (4–6). Pink salmon have been introduced to the Great Lakes in North America (7) and drainage basins of the White Sea (reviewed in (8)), near the border of Russia and Finland.

Pink salmon spend a year and a half at sea before returning to rivers to spawn at two-years of age. This strict two-year life history, unique to this species among salmon, has wide-ranging implications for their evolution, conservation, and possibly for their future as a species. Populations that spawn in an even-numbered year (e.g., 2020) return in an even-year (e.g., 2022) to spawn as adults. Similarly, odd-year spawned pink salmon return in odd-numbered years. Gene flow between year-classes/lineages is consequently limited (9) (this phenomenon is known as allochronic or temporal isolation). Rare exceptions to a strict two-year life-cycle of pink salmon in their native range have been reported (10–12). Outside their native range, three-year-old pink salmon have been observed in the Great Lakes following introduction (7,13). One hypothesis based on experimental rearing in heated sea water is that temperature may play a role in precocious development (i.e., one-year life-cycle) (14).

Within a year-class, population genetic differentiation among rivers tends to be lower than that of other salmon species, which is a possible consequence of increased straying of pink salmon from natal streams during spawning (15,16). Increased straying itself may be a repercussion of the reduced time that pink salmon spend in their natal streams compared to most other salmon species (chum salmon – *Oncorhynchus keta* being an exception) (3). Pink salmon are ready for sea migration as soon as they emerge from gravel and after yolk-sac absorption (17).

In contrast to the regional reduced heterogeneity observed within year-class populations, there is a high level of divergence between year-classes as a result of limited gene flow (9,18–22). Genetic differentiation between odd and even lineages from the same river is greater than within year-class

differentiation, a phenomenon observed across the species natural range (23). There are also phenotypic differences that have been reported between lineages such as gill raker counts (19), length/size (with even-year fish tending to be smaller in Canada) (24–26), and survival/alevin growth in low temperature environments (27).

The divergence of pink salmon from other Pacific salmon species has been estimated to have occurred several million years ago (28–32); this provides a maximal time of odd and even lineage divergence. Based on mitochondrial nucleotide diversity, divergence times between odd and even-year lineages have previously been estimated as 23,600 years (33), 150 – 608 thousand years ago (34), and 0.9 – 1.1 million years ago (22). The relatively recent estimates of divergence are inconsistent with complete temporal isolation between odd and even lineages (potentially for several million years). It has been suggested that low-level gene flow or recolonization of extripated year-classes by alternate year-classes could account for recent estimates of divergence, with recolonization being a favoured explanation (33). Both low-level gene flow and recolonization have been observed in introduced pink salmon in the North American Great Lakes (7,35,36), revealing that it is possible that environment and temperature (suggested in (36)) can alter the strict allochronic isolation observed in modern times. Maturation in pink salmon has been verified to be sensitive to temperature and photoperiod under experimental conditions (37).

While odd-year and even-year pink salmon populations may occupy the same environment (during different years), these lineages can still have different selective pressures (38). For example, the density of pink salmon is known to vary between years (38,39), and density may influence the composition of pink salmon predators, prey, and the number of fish on the spawning grounds (40–42). In years with a high abundance of pink salmon, some studies have reported a decrease in body size of pink salmon at sea (other species of salmon and seabirds have also been adversely influenced during these high abundance years) (41–45). These studies reveal that the intraspecific competition among

other pink salmon and interspecific competition among other species can vary significantly between odd and even-years.

In this study, we present genome assemblies for both odd-year and even-year lineages, develop a transcriptome to annotate these assemblies, and analyze polymorphisms found between groups. We identified regions of the genome that have diverged between odd and even-year lineages with some possibly as a response to selection. We were also able to identify large Fst peaks adjacent to many centromeres and to verify one major fusion or deletion on LG15_El12.1-15.1 by combining polymorphism data with long-read sequencing of both year-classes. These regions of the genome are important aspects of pink salmon biology and provide greater insight into the evolutionary divergence of the lineages.

Materials and Methods

Animal care

Fisheries and Oceans Canada Pacific Region Animal Care Committee (Ex. 7.1) was the authorizing body for animal care carried out in this study. All salmon were reared, collected, or euthanized in compliance with the Canadian Council on Animal Care Guidelines.

Genome assemblies

A mature male pink salmon was sampled from the Big Qualicum River Hatchery (NCBI BioSample: SAMN16688056) on September 19, 2019 (odd-year) by hatchery personnel and euthanized by concussion as specified in section 5.5 of the Canadian Council on Animal Care guidelines. A mature male pink salmon was also sampled from the Quinsam River Hatchery (NCBI BioSample:

SAMN18987060) by hatchery personnel in the same manner on July 28, 2020 (even-year). We dissected liver, spleen, kidney, and heart tissues from the carcasses and flash-froze them on dry ice immediately and stored them at -80°C. We used a Nanobind Tissue Big DNA Kit (Circulomics) to isolate high-molecular DNA following the manufacturer's protocol from multiple tissues. In addition, Short Read Eliminator Kits (Circulomics) were used to reduce the fraction of small DNA fragments in the DNA extractions following the kit protocol for DNA samples to be sequenced on Oxford Nanopore Technologies (ONT) platforms.

We generated sequencing libraries with the prepared DNA using a Ligation Sequencing Kit (SQK-LSK109 ONT) following the manufacturer's protocol. The libraries were sequenced on a Spot On Flow Cell MK1 R9 with a MinION (ONT) or a PromethION (R9.4.1 flow cell). Libraries sequenced on the PromethION were size selected using magnetic beads (0.4:1 ratio). DNase flushes were performed to increase yield according to manufacturer's instructions. We also tried to add 1% DMSO immediately before sequencing to reduce secondary structures that might block pores and reduce sequencing efficiency for one flow-cell (with a minor increase in pore occupancy, more titration will be needed to identify if there are benefits of adding DMSO). FASTQ sequence files were generated either using the Guppy Basecalling Software (version 3.4.3+f4fc735 for sequences from the MinION) with default settings or MinKNOW v3.4.6 (for sequences from the PromethION).

Short-read sequence data were generated for genome polishing for the even-year genome assembly (NCBI SRA accession: SRX10913279 – SRX10913282) and the odd-year genome assembly (NCBI SRA accession: SRX6595859 – SRX6595860). We generated the short-read data for the even-year genome by shearing 1ug of DNA (pink even-year male described above) with a COVARIS LE220 (Covaris) using the following configuration in a 96 microTUBE plate (Covaris): duty 20, pip450, cycles/burst 200, total time 90s, pulse spin in between 45s treatment. The library was then constructed

using the MGIEasy PCR-Free DNA Library Set (MGI) following the manufacturer's protocol. The library was then sequenced on an MGISEQ-200RS Sequencer (150 + 175 PE).

We generated the short-read sequence data for polishing the odd-year genome assembly for a previous assembly that was not published because the contiguity of the assembly was low. The sequences were from an odd-year haploid female produced at Fisheries and Oceans Canada using source material from the Quinsam River Hatchery (NCBI BioSample: SAMN12367892). To produce the haploid salmon, we applied UV irradiation (560 uW/cm^2 for 176 s) to sperm from a Quinsam River male pink salmon (to destroy parental DNA) immediately before fertilizing eggs from a Quinsam River female pink salmon. Prior to sequencing, the individual was confirmed to be haploid using a panel of 11 microsatellites. The details of the library preparations and sequencing technology can be found on the NCBI website (NCBI SRA accession: SRX6595859 – SRX6595860).

We created a Hi-C library for the even-year genome assembly using the Arima-HiC 2.0 kit (Arima Genomics – manufacturer's protocol) with liver tissue from the even-year male (NCBI SRA accession: SRR14496776). The library was then sequenced on an Illumina HiSeq X (PE150). A Hi-C library was only successfully generated for the even-year genome assembly.

After sequencing, we produced initial genome assemblies with the Flye genome assembler (version 2.7-b1587 – odd, 2.8.2-b1695 – even) (46) using ONT sequences (parameters: -g 2.4g, --asm-coverage 30). Racon (version 1.4.16) (47) was then used to find consensus sequences of the Flye assemblies (parameters: -u) after aligning the respective ONT reads to the assemblies using minimap2 (48) (version 2.13, parameters: -x map-ont). We polished the assemblies with Pilon (version 1.22) (49) using the following methods. Paired-end reads were filtered and trimmed using Trimmomatic (50) (version 0.38) (parameters for the odd-year reads: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING:28 TRAILING:28 SLIDINGWINDOW:4:15 MINLEN:200; parameters for the even-year reads: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3

MINLEN:36). The respective reads were aligned to each of the Racon-corrected assemblies using bwa (51,52) (version 0.7.17) with the -M parameter and sorted and indexed using Samtools (53) (version 1.9) prior to polishing with Pilon (default parameters).

After the genome assemblies were polished, we identified the order and orientation of contigs/scaffolds on pseudomolecules/chromosomes for the odd-year genome using a genetic map (54) and synteny to the coho salmon genome (NCBI: GCF_002021735.2). Chromonomer (55) (version 1.10) was used to order the contigs/scaffolds using the genetic map (parameters: --disable_splitting). Ragtag (56) (version 1.0.1) was used to order the contigs/scaffolds using synteny to the coho salmon genome (default parameters). We used a custom script (57) to compare the contig order files output by Chromonomer and Ragtag (.agp files) and manually reviewed the output for discrepancies. The manually curated order and .fasta files were submitted to the NCBI.

To order and orient contigs and scaffolds on pseudomolecules for the even-year genome, we mapped Hi-C reads to the polished assembly using scripts from Arima Genomics (58). The output alignment file was then converted to a .bed file using BEDtool bamtobed (version 2.27.1) (59) with default parameters and sorted using the Unix command 'sort -k 4.' After the Hi-C reads were mapped to the genome assembly, Salsa2 (60,61) was used to further scaffold the contigs and initial scaffolds (parameters: -e GATCGATC,GANTGATC,GANTANTC,GATCANTC). After scaffolding, we mapped the remaining contigs and scaffolds onto pseudomolecules/chromosomes using the same strategy as for the odd-year genome assembly (see above) except a newer genetic map was used (62) (an odd-year genetic map was the only available) and the rainbow trout genome assembly (NCBI: GCF_013265735.2, (63)) was chosen for synteny. The proposed order and orientation was then reviewed manually using Juicebox (version 1.11.08) (64) before submission to the NCBI. The .hic and .assembly files used by Juicebox were produced using the pipeline from Phase Genomics (65). The

nomenclature for the chromosomes was based on the linkage group from the genetic maps and from the Northern pike orthologous chromosomes in an attempt to standardize nomenclature across salmonids (66).

A BUSCO (Benchmarking Universal Single-Copy Orthologs) version 3.0.2 analysis (67) was used to assess assembly quality. We performed these analyses after polishing assemblies, but before mapping contigs/scaffolds onto chromosomes. The lineage dataset used in this analysis was actinopterygii_odb9 (4584 BUSCOs). The parameters used were: -m genome and -sp zebrafish.

A Circos plot was generated from the odd-year genome assembly using Circos software version 0.69-8 (68). We identified homeologous regions of the genome with SyMap version 5.0.6 (69) using a repeat-masked version of the assembly without unplaced scaffolds or contigs (default settings). Repeats had previously been identified by NCBI and were masked by us using Unix commands. The output from SyMap was formatted and summarized using scripts from Christensen et al. (2018) (70). A histogram of repetitive sequence was generated using a python script (71). The Marey map (genetic map markers aligned to a genome) was generated using the methods from Christensen et al. (2018) (70). Centromere positions were taken from the genetic map after it was converted into a Marey map.

Whole-genome re-sequencing

Samples were previously collected by Fisheries and Oceans Canada personnel from the following bodies of water (British Columbia unless otherwise noted): Quinsam River Hatchery, Atnarko River, Kitimat River Salmon Hatchery, Deena River, Yakoun River Hatchery, Snootli Creek Hatchery, Kushiro River (Japan) (S1 File). Samples were chosen to encompass odd-year and even-year samples from the same body of water or from nearby streams (even-year n=30, odd-year n=31).

We extracted DNA from tissues stored either in 100% ethanol or RNAlater (ThermoFisher) using the manufacturer's protocol (72). Whole-genome sequencing libraries were produced at McGill University and Génome Québec Innovation Centre (now the Centre d'expertise et de services Génome Québec). The libraries were generated using the NxSeq AmpFREE Low DNA Library Kit and NxSeq Adaptors (Lucigen). They were then sequenced on an Illumina HiSeq X (PE150).

We identified nucleotide variants using GATK (73–75) (version 3.8). Unfiltered paired-end reads were aligned to the Racon corrected odd-year genome assembly (as other versions were unavailable at the time – available at: https://doi.org/10.6084/m9.figshare.14963721.v1) using bwa mem (parameters: -m) and the sort command from Samtools. Picard's (76) (version 2.18.9) AddOrReplaceReadGroups was used to change read group information (with stringency set to lenient). Samtools was used to index the resulting alignment files, and the MarkDuplicates command from Picard was used to mark possible PCR duplicates (lenient validation stringency). The MarkDuplicates command was also used to merge .bam files if multiple sequencing lanes were used to sequence the sample. Read group information was changed using the Picard command ReplaceSamHeader for these samples so that the library and sample ID were the same, but other information was left the same. This was performed so that GATK would treat the sample appropriately.

HaplotypeCaller (GATK) was then used to generate .gvcf files (parameters: --genotyping_mode DISCOVERY, --emitRefConfidence GVCF) for each sample. The GenotypeGVCF command from GATK was then used to genotype the individuals in 10 Mbp intervals (see (77) for python script used to split into 10 Mbp intervals). The CatVariants command was used to merge the intervals afterwards. Variants were then hard-filtered using vcftools (78) (version 0.1.15) with the following parameters: maf 0.05, max-alleles 2, min-alleles 2, max-missing 0.9, remove-indels, and remove-filtered-all (VCF file available at: https://doi.org/10.6084/m9.figshare.14963739.v1). Additional filtering was done for

some analyses, which are sensitive to linkage disequilibrium. Variants were filtered if heterozygous allele counts were not evenly represented — also known as allele balance (minor allele count < 20% of the major allele count, see (77) for python script). Variants in linkage disequilibrium were thinned using BCFtools (79) (parameters: +prune, -w 20kb, -l 0.4, and -n 2). Custom scripts, bwa mem, and Samtools index were used to map the variants to different genome assemblies (80).

Transcriptome

To better annotate the genome assemblies, we collected a dataset of 19 tissues from a juvenile female pink salmon (NCBI Accessions: SRX6595821-SRX6595839). Euthanasia of this salmon was performed by placing the salmon in a bath of 100 mg/L tricaine methanesulfonate buffered with 200 mg/L sodium bicarbonate. Team dissection was used to quickly remove tissues and each tissue was stored in RNAlater Stabilization Solution (ThermoFisher) as recommended by the manufacturer.

We extracted RNA from the tissue stored in RNAlater Stabilization Solution using the Qiagen RNeasy kit (QIAGEN). Stranded mRNASeq libraries were generated at McGill University and Génome Québec Innovation Centre, with NEBNext dual index adapters. Libraries were then sequenced as a 1/39 fraction of a NovaSeq 6000 S4 PE150 lane at McGill University and Génome Québec Innovation Centre. These datasets were deposited to NCBI for use in gene annotation (BioProject: PRJNA556728).

Population structure

As clustering techniques are sensitive to linkage disequilibrium, we used variants that were hard-filtered (including for allele balance) and filtered for linkage disequilibrium for all population

structure analyses. A DAPC analysis (81) was used to cluster individuals in R (82) using the following packages: adegenet (83), vcfR (84), and ggplot2 (85). The number of DAPC clusters was determined using the find.clusters function and choosing the cluster count with the lowest Bayesian information criterion. Thirty principal components were retained with the dapc function. The variants used for the DAPC analysis were not yet mapped to chromosomes.

To complement the DAPC analysis, we also performed an Admixture (version 1.3.0) analysis (86) to identify clusters of individuals and quantify the admixture between the identified groups. To format the linkage disequilibrium thinned .vcf file, we used a custom Python script to rename the chromosomes to numbers (77) and PLINK (version 1.90b6.15) (87,88) was used to generate .bed files (parameters: --chr-set 26 no-xy, --double-id). PLINK was also used to generate a principal components analysis. The Admixture software was then used to identify the optimal cluster number based on the lowest cross-validation error value. The admixture values from this analysis were plotted in R.

To examine population structure based on the mitochondrion sequence, we generated a phylogenetic tree based on full mitochondria sequences. The genome assembly included a mitochondrion sequence, and this region of the genome was subset from the variant file using vcftools. The resulting file and the SNPRelate (89) package in R was used to generate the phylogenetic tree. The snpgdsVCF2GDS and snpgdsOpen functions were used to import the data, the snpgdsDiss function was used to calculate the individual dissimilarities for pairwise comparisons between samples, the snpgdsHCluster function was used to generate a hierarchical cluster of the dissimilarity matrix, the snpgdsCutTree function was used to determine subgroups, and the snpgdsDrawTree function was used to plot the dendrogram.

From the variants with minimal filtering and the variants after all filters had been applied, the heterozygosity ratio was separately calculated based on the number of heterozygous genotypes divided

by the number of alternative homozygous genotypes (90,91). The number of heterozygous and homozygous genotypes were counted using a python script from Christensen et al. (2020) (77). Heterozygous genotypes per kilobase pair (kbp) was calculated by dividing the heterozygous genotype counts by the genome size (2,528,518,120 bp) and then multiplied by 1000. This calculation was used on the variants with minimal filtering not yet mapped to chromosomes.

The number of shared alleles was calculated as a metric for relatedness using custom scripts for the variants with minimal filtering and which were mapped to chromosomes (92). This value is calculated by counting the number of alleles an individual has in common with another individual and is similar to previous work (93–95). The percent shared alleles was calculated in R (number of shared alleles divided by the total allele count multiplied by 100) and plotted using the reshape2 (96) and pheatmap (97) R packages.

Fst, nucleotide diversity (within populations—pi and between—dxy), and Tajima's D were calculated and plotted using the R packages PopGenome (98), dplyr (99), tidyr (100), stringr, and qqman. In PopGenome, all metrics were calculated using a sliding window of 10 kbp and the data were visualized as a Manhattan plot using qqman. We used the populations module from Stacks version 2.54 (101) to calculate the number of private alleles, percent polymorphic variants, Fis (inbreeding coefficient), and Pi (nucleotide diversity within a population) for odd and even year class samples grouped as populations. These metrics were compared with the sample from Japan as an odd-year sample or as its own population. A comparison was also performed to see how filtering influenced these metrics.

Genomic regions associated with population structure under selection

To identify regions of the genome associated with population structure identified in the DAPC analysis, we performed an eigenGWAS analysis (102). The format of the hard-filtered variants was converted to the appropriate format in PLINK, and the GEAR (103) software was used to run the eigenGWAS analysis (this was performed on a slightly different version of the genome assembly available on the NCBI, but only positions on chromosome 9 were minimally affected). Significance was corrected for using the genomic inflation factor to better identify markers potentially under selection rather than a result of genetic drift between populations. The genomic inflation factor corrected p-values were then plotted in R using the qqman (104) and stringr (105) packages. A Bonferroni correction was applied as a multiple test correction (alpha = 0.05). Only peaks with at least 5 SNPs within 100 kbp of each other were retained to reduce false-positives (nucleotide variants under selection are expected to be in linkage disequilibrium with surrounding variants and significant single variants not in linkage may be a consequence of spurious alignments).

Sex determination and sdY

We utilized a genome-wide association (GWA) of phenotypic sex to identify the region of the genome associated with sex for all pink salmon (individual year-classes were checked as well). This analysis was also used to identify where the contig from the genome assembly with the sdY gene should be placed. This was confirmed with synteny from the rainbow trout Y-chromosome (NC_048593.1) and manual inspection of the Hi-C data (it was placed in the even-year genome assembly). The GWA analyses were performed using PLINK (parameters: --logistic --perm). Synteny was identified from alignments to the rainbow trout genome assembly (GCF_013265735.2, (63)) using CHROMEISTER (106) (default settings).

When manually genotyping the presence/absence of the sdY gene by visualizing alignments in IGV (107), we noticed some males had increased coverage of the sdY gene, and two haplotypes were identified (4 variants in non-coding DNA). The haplotypes were manually genotyped. To estimate the copy number of the sdY gene, we first used a python script to determine the average coverage of all hard-filtered variants (108). The average coverage of the four variants in the sdY gene was then divided by the average coverage of all variants.

Results

Genome assemblies

The odd-year assembly (GCA_017355495.1) had a combined length of ~2.5 Gbp, with 20,664 contigs and a contig N50 of ~1.8 Mbp. The even-year assembly had similar metrics, with a contig N50 of ~1.5 Mbp, 24,235 contigs, and a length of ~2.7 Gbp. We used a BUSCO analysis of known conserved genes to determine the completeness and quality of the genome assembly. Of the 4584 BUSCOs, 95.3% were found to be complete in the odd-year genome assembly (54.9% single-copy and 40.4% duplicated), 1.4% were fragmented, and 3.3% were missing. The even-year assembly also had 95.3% complete BUSCOs (51.5% single-copy and 43.8% duplicated), but more fragmented (1.6%) and fewer missing BUSCOs (3.1%).

The odd-year assembly had 26 linkage groups and extensive homeologous regions between chromosomes (Fig 1). The odd-year genome assembly contained similar levels of repetitive DNA and duplicated regions compared to other salmonids (Fig 1, (70,77,109)). Like other salmon species, increased sequence similarity was also observed at telomeres between duplicated chromosomal arms

(Fig 1). Peaks of increased Fst between odd and even-year lineages were commonly found at putative centromere locations (Fig 1, Table 1).

Fig 1. Circos plot of pink salmon genome assembly. Positions are all based on the odd-year genome assembly. Chromosomes/linkage groups are noted with blue boxes representing the centromere identified in Tarpey et al. (2017) (62). Links between chromosomes are homeologous regions identified using SyMap. A) Fst values between all odd-year and even-year salmon greater than 0.25. Values greater than 0.5 are highlighted red. B) The fraction of repetitive DNA as identified by NCBI (odd-year). Values greater than 0.65 are highlighted red. C) The percent identity between homeologous regions identified by SyMap (scale 75-100%). Values greater than 90% are highlighted red. D) A Marey map with markers from the genetic map (y-axis, 0 – 1, with 1 being the marker with the greatest cM value) placed onto the genome (x-axis, odd-year).

Table 1. Largest Fst peaks between odd and even-year lineages.

Linkage group/ chromosome	Region (Mbp)	Size of peak (Mbp)	Frequency Odd (p*)	Frequency Even (p*)	HWE	Potential cause
LG04_El13.1-02.1	50-53	~3	0.98	0.43	Both	Centromere
LG10_El12.1-15.1	46.5-50	~3.5	0.69	0.22	Both	Centromere
LG14_El18.2-23.2	49-55	~6	0.77	0.12	Both	Centromere
LG15_El08.2-20.1	50-54	~4 Centromere ~1.26 Deletion	0.5**	0**	Both	Centromere/ Deletion-Fusion
LG18_El09.2-17.1	45.5-46.5	~1	1	0	Both	Selection
LG21_El24.2-22.1	31-34.5	~3.5	0.63	0	Both	Centromere
LG25_El23.1-24.1	17.5-19	~1.5	0.65	0.15	Both	Centromere
LG26_El09.1-11.1	11.3-18	~3.5 Centromere ~7.3 Misassembly	0.89	0.67	Both	Centromere/ Misassembly

The odd and even allele frequencies (p) were based on the most clearly defined sub-region rather than the entire region. It is unclear which individuals have the deletion or fusion on LG15_El08.2-20.1.

*reference genome allele frequency

**alternative (to reference) allele frequency

Population structure

A shared allele analysis (Fig 2) and both Admixture and DAPC analyses (Fig 3) revealed a clear delineation between odd and even-year lineages. Parent-progeny and sibling relationships (relationships known during sampling) are highlighted by increased levels of shared alleles, but the majority of clustering appears to be related to geographical distance (Fig 2, S1 File). No apparent admixture was observed in the even-year class (Fig 3B). In the odd-year lineage, estimated ancestry from the even-year group varied from zero to over forty percent (Fig 3B).

Fig 2. Percent of shared alleles among pink salmon. A heatmap of shared alleles between salmon is shown with clustering and a dendrogram. Each square represents the percent shared alleles after minor filtering of variants (bi-allelic SNPs). In addition to the legend displaying the colour representation of percent shared alleles, the sex, year-class, and river system sample information is colour-coded and shown on both rows and columns.

Fig 3. Population structure of pink salmon. A) Sampling locations for odd and even-year pink salmon. B) An admixture analysis based on an optimal group number of two. Sampling site is specified on the left (y-axis) by colour and fraction of alleles inherited from a lineage is shown on the x-axis (orange – even-year, blue – odd-year). On the right, DAPC groups are shown (see S1 File for group and coordinate positions). The DAPC groups matched year-class/lineage designations.

A separate analysis of mitochondrial DNA was performed to further investigate the relationships between the odd and even-year lineages. Odd-year pink salmon had longer branch lengths in mitochondria dendrograms and haplotype networks with more uniform distributions of haplotypes (Figs 4A and B). The even-year salmon had two major haplotypes (Fig 4B). Mitochondrial

sequence analyses revealed 21 unique haplotypes from the 61 mitochondria sequences with 1-19 steps between haplotypes (Fig 4). Based on the length of the sequence analyzed (16,822 bp) this represents a mutation frequency between 0.006% to 0.1%. One haplotype was shared between lineages and the closest haplotype that was not shared had 5 steps between year-classes (Fig 4). The mitochondrial analyses illustrate divergence between the odd and even-year lineages, but also raises questions regarding possible recent admixture based on a shared haplotype and an odd-year haplotype most closely related to an even-year haplotype.

Fig 4. Whole mitochondrial genome comparisons between lineages. A) A dendrogram based on full mitochondrial sequences. The y-axes show dissimilarity scores on the left and coancestry values on the right, which were used to cluster individuals. Year-class/lineage is specified below the dendrogram. B) A full mitochondrial genome haplotype network is shown for the 21 unique haplotypes identified. River names are shown for the haplotype shared between lineages.

Several metrics were calculated to quantify genetic divergence between and within year-classes: heterozygosity ratios, heterozygous genotype per kbp, polymorphic sites, private alleles, and nucleotide diversity. Heterozygosity ratios in odd-year fish ranged from 1.5-4.56, with an average of 2.54 (excluding haploid individuals generated for a previous project) (S1 File). Even-year class individuals ranged from 1.09-1.78, with an average of 1.44 (S1 File). The average heterozygous genotype per kbp (excluding haploids) was 0.71 for odd-year salmon (range: 0.55 – 0.85) and 0.58 for even-year (range: 0.45 – 0.69) pink salmon. The Pearson correlation between heterozygosity ratio and heterozygosity per kb was 0.91 (excluding haploids). Salmon from odd-years had on average higher levels of polymorphic sites, increased private alleles, and increased nucleotide diversity (Table 2). These values varied based on sample inclusion and filtering parameters used for filtering nucleotide variants (Table

2). The average percent of shared alleles among odd-year fish was 76.13%, 74.42% among even-year individuals, and 71.04% between year-classes (S1 File). Most analyses revealed increased genetic diversity among odd-year pink salmon than among even-year pink salmon and fewer shared alleles between odd and even-year populations than within year-class.

Table 2. Population metrics of the two lineages.

	Odd with sample from Japan	**Odd** without sample from Japan	**Even**
Number of variants		3,817,721 (101,594)	
% polymorphic	93.23 (97.91)	92.61 (97.64)	90.66 (87.86)
Private alleles	356,634 (12,333)	302,507 (10,486)	258,335 (2,124*)
Pi	0.276 (0.337)	0.276 (0.337)	0.269 (0.283)
Fis	0.075 (0.143)	0.068 (0.135)	0.122 (0.195)

Variants with minimal filtering are shown first and variants after all filters are shown in parentheses.
*2,126 without sample from Japan

Genomic regions associated with odd and even-year lineages

We identified regions of the genome with divergence between odd and even-year lineages using an eigenGWAS and Fst analysis. Seventeen significant regions of the genome were discovered to contribute to the divergence between odd and even-year lineages (Fig 5, Table 3). These regions are putatively under selection as genetic drift is partially accounted for through the genomic inflation factor. Multiple candidate genes under selection were identified in these regions (Table 3).

Fig 5. Genome-wide divergence between odd and even-year pink salmon lineages. A Manhattan plot of eigenGWAS results, with chromosome positions on the x-axis and p-values (corrected for genetic drift using the genomic inflation factor) on the y-axis identifies region of the genome potentially under selection. The red horizontal line represents a Bonferroni correction at α=0.01 and the blue line at α=0.05. All positions are from the odd-year genome assembly.

Table 3. Top eigenGWAS peaks identified between lineages.

Chromosome	BP range	SNP position with lowest P-value	Candidate Gene	Gene Symbol
LG01_El19.1-16.1	51653225-51738345	51716026	protein tyrosine phosphatase receptor type J	*PTPRJ*
LG02_El19.2-07.2	18075760-18095551	18075760	AT-rich interactive domain-containing protein 3A	*arid3a*
LG02_El19.2-07.2	46961740-47008290	46969254	protein-methionine sulfoxide oxidase mical2b	*mical2b*
LG02_El19.2-07.2	110392052-110493632	110449484	multidrug and toxin extrusion protein 2-like	*SLC47A2*
LG08_El03.2-06.2	60715584-60782108	60767399	polh polymerase (DNA directed), eta	*POLH*
LG14_El18.2-23.2	29365137-29414547	29411435	uncharacterized gene	
LG14_El18.2-23.2	50053735-50217619	50217619	Unknown	
LG15_El08.2-20.1	42753852-42762230	42758791	cystathionine gamma-lyase	*CTH*
LG15_El08.2-20.1	51106314-52224901	51106314	multiple candidates*	
LG18_El09.2-17.1	45516859-45534019	45524530	cell division control protein 42 homolog*	*CDC42*
LG18_El09.2-17.1	46342891-46450909	46347678	H-2 class II histocompatibility antigen, A-U alpha chain*	*H2-Aa*
LG19_El20.2-01.2	22831059-22843129	22836628	B-cell receptor CD22	*CD22*
LG21_El24.2-22.1	9281799-9398462	9391795	histidine N-acetyltransferase	*hisat*
LG22_El03.1	10576168-10595358	10595038	purine nucleoside phosphorylase	*PNP*
LG22_El03.1	15334053-15411821	15405819	multiple candidates	
LG24_El10.1-25.1	47355926-47430898	47398112	microtubule-associated protein 9	*map9*
LG25_El23.1-24.1	37126251-37202999	37137812	uncharacterized gene (ncRNA)	

All positions are relative to the odd-year genome.
*associated with an Fst peak

Seven of the eight largest Fst peaks between year-classes were located in the vicinity of a centromere (Fig 1, Table 1). More detail is presented on one of the largest Fst peaks. This peak is also associated with a large deletion or fusion. The Fst peak on LG15_El12.1-15.1 (Fig 6A) is in Hardy-Weinberg equilibrium in the odd-year lineage (p=0.984 with a chi-square test), but fixed in the even-year lineage (Figs 6B and 6C). When Oxford Nanopore reads from the two year-classes were aligned

to the genome assembly, a heterozygous deletion or fusion from 51,670,144 – 52,926,328 was found in this region of the odd-year male used for genome assembly (S1 Fig.). The ~1.2 Mbp deletion/fusion may explain why the LG15_El12.1-15.1 Fst peak was one of the largest and widest (Fig 1, S1 Fig.).

Fig 6. Chromosome LG15_El12.1-15.1 Fst peak. A) A Manhattan plot of 10 kbp sliding-window Fst values between odd and even-year pink salmon lineages on chromosome LG15_El12.1-15.1. B) Genotypes visualized in IGV. Each row represents an individual pink salmon and each column represents a nucleotide variant (dark blue – homozygous reference, light blue – heterozygous reference, green – homozygous alternative, and white – missing genotype). Individuals were sorted by year-class (shown on the right) and then by assigned genotype (shown on the left). C) Counts of genotypes of the chromosomal polymorphism based on manual genotyping.

It is difficult to distinguish between a deletion and a chromosomal fusion in these analyses and this may represent a fusion instead of a deletion. Previous research support chromosomal variants in pink salmon (110) and a species specific fusion of this chromosome (66), but further research will be needed to confirm this hypothesis. From these analyses, many highly divergent regions of the genome were identified, either from selection or from genetic isolation/population dynamics. The largest reservoirs of divergence between odd and even-year classes appears to be associated with centromeres, but not exclusively and uncommonly for regions putatively under selection (Table 3).

Sex determination and sdY

The sex-determination gene in salmonids, *sdY* (111), was located on a ~110 kbp contig in the pink salmon odd-year genome assembly (NCBI accession: JADWMN010014055.1) and on a contig ~367 kbp that was placed onto a chromosome in the even-year genome assembly. The *sdY* gene can be placed at one of the ends of LG20_El14.2 by using genome-wide association with sex as the trait of interest, Hi-C contact data (even-year genome), and synteny with the rainbow trout Y-chromosome and chromosome 29 (an autosome) of the coho salmon (Fig 7A, S1 and S3 Figs.). LG20_El14.2, has the reverse orientation in the odd-year assembly compared to the genetic map (Fig 1), but was corrected to have the same orientation in the even-year assembly.

Fig 7. The location and counts of the sex-determining gene, *sdY*, in pink salmon. A) A genome-wide association analysis with sex as the phenotype under investigation shown as a Manhattan plot. The putative sex-determining region is indicated with an arrow. B) A scatterplot with the average coverage of the variants across the genome on the x-axis for all the pink salmon, and the estimated *sdY* count on the y-axis (*sdY* has previously been identified as the sex-determining gene in most salmonids). The different colour points represent different year-classes and *sdY* haplotypes.

In both genome assemblies there is only one copy of the *sdY* gene, confirmed with a BLAST alignment of a sdY gene available in the NCBI database (KU556848.1) to the respective assemblies. From a self-alignment of the *sdY*-containing contig, the majority of this contig is highly repetitive, > 90 kbp out of ~110 kbp. From the alignment of the *sdY*-containing region in pink salmon to the coho salmon chromosome 29, only a small portion of the Y-chromosome appears to be unique to the Y-chromosome (S3 Fig.). Genotypes were called for the majority of this region for males and females,

and the main difference related to sex was that all females had large runs of homozygosity while many males had large runs of heterozygosity (S4 Fig, S1 File).

From previous research (112,113), a pseudo growth hormone 2 gene was shown to be tightly linked to sex-determination in pink salmon. Four tandem duplicates of this gene (NCBI: DQ460711.1) were identified on the same contig in the even-year genome assembly, but only two copies were found in the odd-year genome assembly on separate contigs (S1 File). As these contigs were not mapped to a chromosomal position, it is likely that parts of the Y-chromosome specific region remain incomplete in these two assemblies.

There were two *sdY* haplotypes (variants found in non-coding DNA) observed in both odd and even male pink salmon (Fig 7B, S1 File). Additionally, some males possessed multiple copies of the sdY gene (Fig 7B). The CGGA *sdY* haplotype was only identified in a single odd-year male pink salmon, while the TTAC haplotype was evenly distributed between year-classes (S1 File).

Based on manual inspection of the genotypes, long stretches of heterozygosity were observed near the *sdY* gene in some males, but not in others. In males with the TTAC *sdY* haplotype, there were extended or short runs of heterozygosity evenly distributed between year-classes (S1 File). Even-year males with the TTAC *sdY* haplotype and a short run of heterozygosity were more likely to have multiple copies of the *sdY* gene (n=4, average=2.7) than the same group with long runs of heterozygosity (n=4, average=0.9, p=0.017 with one-tailed, unpaired t-test). Any individuals with the CGGA *sdY* haplotype did not have stretches of heterozygosity near the putative location of *sdY*. One hypothesis to explain these results is that individuals with the CGGA *sdY* haplotype have an alternative sex chromosome.

Discussion

Population structure

Similar to previous studies (23,54), pink salmon population structure divergence was found at the whole-genome level to be greater between year-classes rather than based on geography. Shared allele, DAPC, and Admixture analyses point to a clear delineation of odd and even lineages, with the exception of the only sample from Japan. Further sampling will need to be performed to provide an improved picture of the diversity of this species at the whole-genome level within year-class and across their Pacific Rim distribution. In British Columbia, however, the even-year lineage appeared to be more homogeneous than the odd-year lineage based on the admixture analysis and several population metrics such as nucleotide diversity. A similar result was previously observed with microsatellite (15) and SNP markers (23).

Divergence between lineages was also revealed by whole mitochondrial sequences. There were 21 unique mitochondria genotypes among the 61 individuals sampled, and only one of these haplotypes was shared between lineages. While the number of unique haplotypes was the same between lineages, most of the even-year class haplotypes (8 out of 10) were similar in sequence. The two major haplotypes seen in the even-year class were consistent with the Alaskan A and AA haplotypes seen in Churikov and Gharrett (2002) (33), as were the numerous and more distantly related odd-year haplotypes.

The low nucleotide diversity of mitochondrial haplotype networks and the increase of rare haplotypes have led previous studies to conclude that pink salmon (with some local exceptions) have undergone a bottleneck during the Pleistocene interglacial period and rapid expansion since the last glacial maximum or earlier (33,34,114). The interconnected mitochondrial networks in these studies

518 have inner shared haplotypes between year-classes. Churikov and Gharrett (2002) suggested that these
519 observations supported a model where a year-class might go extinct and an alternate year-class would
520 then replace that population rather than continued gene flow between year-classes that would be
521 necessary to otherwise explain the shared haplotypes (incomplete lineage sorting was tested) (33). The
522 mitochondrial network seen in this study is consistent with that hypothesis. An alternative hypothesis
523 is that environmental factors influence maturation timing and the strict two year life-cycle of pink
524 salmon, and gene flow between year-classes only occurs when environmental conditions favour
525 changes to the two year life-cycle, as that seen in the introduction of pink salmon to the Great Lakes
526 (7,35,36).

527 Estimates of divergence based on mitochondrial sequences suggest that odd and even-year
528 lineages (from East Asia and Alaska) are relatively recent for pink salmon as a species (generally less
529 than 1 million years ago) and divergence likely began during the Pleistocene interglacial period or later
530 (22,33,34). If the two-year life-cycle is environmentally influenced, these estimates could be distorted
531 by phases of gene-flow and would suggest that the interglaciation period was the last major period of
532 gene-flow between odd and even-year classes (but does not necessarily mean that was when the strict
533 two year life-cycle evolved).

534 It has previously been reported that the odd-year lineage of pink salmon has higher levels of
535 heterozygosity, private alleles, and allelic richness (23,54). A similar trend was observed in this study
536 with the heterozygosity ratio, heterozygous genotypes per kbp, private alleles, and other metrics
537 assessing nucleotide diversity. Several factors could help explain the reduced levels of nucleotide
538 diversity seen in the sampled even-year populations. Tarpey et al. (2018) suggested three possibilities,
539 1) the odd-year lineage was older and the even lineage was derived from the odd-year lineage, 2) there
540 was a past reduction in even-year lineage(s), and 3) genetic variation was lost during adaptation (23).

Further sampling will be required to understand if this phenomenon is seen in all even-year populations (especially as lower heterozygosity in the even-year lineage is not universally supported, e.g. (20)). This information is important to interpret which hypothesis is better supported or if another model is better suited (e.g., extirpated lineage replaced by alternate year or that recent population demographics are more important).

Fst peaks between odd and even-year lineages

A single major chromosomal polymorphism (either a fusion or deletion) was identified proximal to a centromere on LG15_El12.1-15.1. This region was characterized by ~4 Mbp runs of homozygosity/heterozygosity. This region was identified from an Fst analysis because nearly the entire region was fixed in the even-year lineage, but appeared to segregate as a single locus in Hardy-Weinberg equilibrium in the odd-year lineage.

Interestingly, pink salmon runs of homozygosity/heterozygosity were common at centromeres rather than an effect of chromosomal polymorphisms. Six other major runs of homozygosity/heterozygosity were also located near centromeres and they differed between lineages. All of these Fst peaks extend for at least 1 Mbp. It is expected that regions with reduced recombination, such as centromeres, will have increased runs of homozygosity and reduced genetic diversity (reviewed in (115)). This may help explain why there are long runs of homozygosity at centromeres, but not why there are differences between lineages at these loci. Genetic drift or selection such as centromere drive would also need to be considered.

The centromere drive hypothesis posits that a centromere can be retained in a female gamete more often than an alternative centromere during meiosis due to an advantageous DNA sequence

563 mutation at the centromere or from mutations in centromere associated proteins (reviewed in (116–
564 118)). In populations that become isolated, the competition between centromere sequences can quickly
565 drive differentiation at these regions between the populations and result in hybrid defects should they
566 come into contact again (117). These observations reveal that the pink salmon lineages may be at a
567 point where speciation is a likely outcome as these large centromere differences could cause hybrid
568 defects. In medaka, genomic diversity at non-acrocentric repeats in centromeres were associated with
569 speciation (119).

570 The centromere drive hypothesis may further shed light on the fixation of the Fst peak on
571 LG15_ El12.1-15.1. Robertsonian fusions (assuming that the Fst peak on LG15_El12.1-15.1 is indeed
572 associated with a fusion rather than a deletion) can generate centromeres that are preferentially able to
573 segregate to the egg during female meiosis (118). This could help drive the fusion to fixation in a
574 population. Alternatively, if the telocentric chromosomes instead of the fused metacentric chromosome
575 had more effective centromeres, the telocentric chromosomes would become fixed. Further studies
576 will be needed to confirm if there is indeed a fusion instead of a deletion and that the fusion leads to
577 fixation by centromere drive.

578

579 ## Genomic regions putatively under selection

580 A large component of the genetic and phenotypic diversity between pink salmon year-classes
581 likely originates from genetic drift as there is little evidence for gene flow between lineages. However,
582 in addition to genetic drift, these lineages may experience different selective pressures even if they
583 occupy the same streams. As mentioned in the Introduction, population density between lineages is
584 often different and this can generate different ecological environments. EigenGWAS and Fst analyses
585 were used to identify regions of the genome potentially responsive to these environmental differences

586 between pink salmon year-classes. Candidate genes under selection were organized into three broad
587 categories (immune system, organ development/maintenance, and behavior), and each is discussed
588 below.

589 Immune system

590 Variation in immune related genes is a common phenomenon between salmonid populations
591 (e.g., (77,120,121)). Between odd and even-year pink salmon, five eigenGWAS peaks were identified
592 near or in genes with immune related functions. These include the H-2 class II histocompatibility
593 antigen, A-U alpha chain (*H2-Aa*) (122–126), B-cell receptor CD22 (*CD22*) (127,128), polh
594 polymerase (DNA directed), eta (*POLH*) (129–133), AT-rich interactive domain-containing protein 3A
595 (*arid3a*) (134,135), and purine nucleoside phosphorylase (*PNP*) (136–140).

596 Several factors could influence why these immune related genes might be under selection
597 between odd-year and even-year populations of pink salmon. For example, altered migration patterns
598 (reviewed in (141,142)), increased pathogen loads between year classes due to increased density
599 (reviewed in (142,143)), and increased physiological stress from competition and increased number of
600 predators during years with larger returns (e.g., (142)) could all influence the differences observed in
601 immune related genes. Further investigations into the nature of these genes in pink salmon may
602 uncover the environmental factors and selective pressures relevant to the evolutionary history of these
603 pink salmon lineages.

604 Organ development/maintenance

605 Salmon go through nutritional and behavioural changes that require organ-level alterations and
606 maintenance throughout their life-cycle. This can be observed in developing salmon that transition
607 from plankton to other fish as food sources. In the eye, this transition requires the development of new

608 functionality such as night vision to chase prey. One example of such a transition is the change of UV
609 opsins observed in hatched Pacific salmon to blue later in life by opsin changeover (144,145).

610 Variation in vision related genes have previously been observed between sockeye salmon
611 populations (77). In Atlantic salmon, *six6*, a gene related to eye development, daylight vision
612 (146,147), and fertility (148) was also found to be associated with age at maturity (149,150) and later
613 with stomach fullness during migration (151). These studies suggest that genetic variation influencing
614 organ development, transition, or maintenance are important components influencing salmonid
615 evolution.

616 Similar to *six6* in Atlantic salmon, Protein tyrosine phosphatase receptor type J (*PTPRJ*) (152),
617 histidine N-acetlytransferase (*hisat*) (153–160), and microtubule-associated protein 9 (*MAP9* or *ASAP*)
618 (161) all appear to play roles in proper vision. The variation in these genes may represent differences
619 in selective pressure between odd and even-years and could be driven by the different population
620 dynamics observed between odd and even-year populations.

621 Cystathionine gamma-lyase (*CTH*) may have, among other roles, a function in hearing (162–
622 164), and could have been influenced by similar population dynamics as those suggested for vision-
623 related genes. Multidrug and toxin extrusion protein 2 gene (*SLC47A2*) is not related to a specific
624 organ, though it may have a special role in the blood-brain barrier (165,166). Instead, it may help in
625 removing toxins, which might accumulate in more dense and stressed populations.

626 **Behaviour**

627 Fish display consistent behavioural differences from each other, analogous to human
628 personalities (167). Personality variation in a population may represent adaptive solutions to different
629 environmental pressures (167). In high density populations, such as the odd-year populations, more
630 aggressive behaviours during high-density spawning conditions (40) could result in more offspring, but

might waste energy in lower-density conditions. Associations to genes related to behaviour have previously been identified among sockeye salmon populations (77), and under selection between wild and farmed Atlantic salmon (168). In the present study, protein-methionine sulfoxide oxidase mical2b (*mical2b*) (169,170) and cell division control protein 42 homolog (*CDC42*) (171), both putative genes found in the eigenGWAS analysis between even and odd-year pink salmon, have previously been found to be associated with anxiety/reactiveness and schizophrenia, respectively.

Another gene, we were not able to unambiguously identify as a candidate gene from other nearby genes, on LG15_El08.2-20.1 is worth noting because of its association with behaviour. The gene, 5-hydroxytryptamine receptor 2B (*5-HT2B* and also known as serotonin receptor 2B), is associated with impulsivity and impulsive aggression in humans and mice (172,173). This region of the genome was fixed in the even-year class pink salmon where population density is often lower (in many North American populations, but not necessarily in all of the rivers in this study) and aggression might be maladaptive. These associations suggest that population dynamics might influence the frequency of certain personality traits in pink salmon populations, with density as a possible driving force.

Sex determination and sdY

With the discovery of a novel sex-determining gene in salmonids (111), and previously with closely linked genetic markers (174,175) researchers have been able to identify instances of sex-determination switching between chromosomes in salmonids (176–180). As suggested in Yano et al. (2013), Y-chromosome switching may act in response to (expected) degeneration of the Y-

chromosome due to mutation accumulation from reduced recombination (181). In pink salmon, *sdY* was located on LG20_El14.2, but we suggest there may be an alternative location as well.

Several pieces of information indicate that LG20_El14.2 may not be the only location of the sex-determining gene, *sdY*, in the pink salmon genome. For instance, there were two *sdY* haplotypes and several males had multiple copies of this gene. Also, all males with the CGGA *sdY* haplotype had a run of homozygosity similar to most females on the LG20_El14.2 chromosome near the putative location of *sdY*. It is expected that near the *sdY* gene, recombination is reduced and mutations would accumulate between the X and Y-chromosomes as a result of reduced recombination. Females tend to have long runs of homozygous genotypes where recombination is reduced and males tend to have long stretches of heterozygous genotypes when reads from the X and Y-chromosome align at the same location (77). Since the males with the CGGA *sdY* haplotype have long runs of homozygous genotypes at the LG20_El14.2 region, as most of the females do, we suggest that the CGGA *sdY* is at another location in the genome in these individuals. We were unable to identify a precise putative alternative location because there were too few individuals with the CGGA *sdY* to obtain a signal from a genome wide association analysis, however, the potential discovery of another salmon species with alternative *sdY* locations, further supports the hypothesis of Y-chromosome switching put forth by Yano et al. (2013) for salmonids (181).

Conclusions

We generated reference genome assemblies for both pink salmon lineages, RNA-seq data for genome annotation, and whole genome re-sequencing data to expand the available resources for this commercially important and evolutionarily interesting species. The coupled whole genome re-

674 sequencing study of 61 individuals from several streams in British Columbia (and one from Japan)
675 helped us to characterize regions of the genome that have diverged between the temporally isolated
676 groups. The amount and degree of lineage-specific genomic variation suggests that there is little gene-
677 flow between the year-classes, but the shared variants such as whole mitochondrial and *sdY* haplotypes
678 suggests that there has been enough recent gene-flow or alternative year-class replacement to maintain
679 these similarities. Divergence at centromeres between the two lineages may be a consequence of
680 centromere drive (or genetic drift and reduced recombination) and represent early stages of speciation.
681 Genes related to the immune system, organ development/maintenance, and behaviour were divergent
682 between odd and even-year classes as well. These example lineage defining differences offer us a
683 glimpse into the evolutionary landscape and the selective pressures or demographic histories that may
684 have driven the divergence in these genes and the differences between odd and even lineages of pink
685 salmon.

686

687 # Acknowledgements

688 Extensive sample preparation and sequencing was performed at McGill University and Génome
689 Québec Innovation Centre (now the Centre d'expertise et de services Génome Québec) and we would
690 like to thank the staff and scientists there for their efforts. We would also like to thank the generous
691 computing resources provided by Compute Canada (www.computecanada.ca). Fisheries and Oceans
692 Canada and the University of Victoria facilities and personnel made this work possible. Finally, the
693 authors would like to thank the many Fisheries and Oceans Canada staff who collected samples for
694 analysis in this study.

695

References

1. Statistics – NPAFC [Internet]. [cited 2021 Jan 13]. Available from: https://npafc.org/statistics/

2. Groot G. Pacific Salmon Life Histories. UBC Press; 1991. 602 p.

3. Heard WR. Life history of Pink Salmon (Oncorhynchus gorbuscha). In: Pacific salmon life histories. Vancouver: University of British Columbia Press; 1991. p. 119–230.

4. Farley EV, Murphy JM, Cieciel K, Yasumiishi EM, Dunmall K, Sformo T, et al. Response of Pink salmon to climate warming in the northern Bering Sea. Deep Sea Res Part II Top Stud Oceanogr. 2020 Jul 1;177:104830.

5. Dunmall KM, Reist JD, Carmack EC, Babaluk JA, Heide-Jørgensen MP, Docker MF. Pacific Salmon in the Arctic: Harbingers of Change. In: Responses of Arctic Marine Ecosystems to Climate Change. Alaska Sea Grant, University of Alaska Fairbanks; 2013. p. 141–62.

6. Dunmall KM, McNicholl DG, Reist JD. Community-based Monitoring Demonstrates Increasing Occurrences and Abundances of Pacific Salmon in the Canadian Arctic from 2000 to 2017. North Pacific Anadromous Fish Commission; 2018 p. 87–90. Report No.: 11.

7. Wen-Hwa K, Lawrie AH. Pink Salmon in the Great Lakes. Fisheries. 1981 Mar 1;6(2):2–6.

8. Sandlund OT, Berntsen HH, Fiske P, Kuusela J, Muladal R, Niemelä E, et al. Pink salmon in Norway: the reluctant invader. Biol Invasions. 2019 Apr 1;21(4):1033–54.

9. Aspinwall N. Genetic Analysis of North American Populations of the Pink Salmon, Oncorhynchus gorbuscha, Possible Evidence for the Neutral Mutation-Random Drift Hypothesis. Evolution. 1974;28(2):295–305.

10. Anas RE. Three-year-old Pink Salmon. J Fish Board Can [Internet]. 2011 Apr 13 [cited 2021 Jan 21]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f59-010

11. Foster RW, Bagatell C, Fuss HJ. Return of One-year-old Pink Salmon to a Stream in Puget Sound. Progress Fish-Cult. 1981 Jan 1;43(1):31–31.

12. Turner CE, Bilton HT. Another Pink Salmon (Oncorhynchus gorbuscha) in its Third Year. J Fish Board Can [Internet]. 2011 Apr 10 [cited 2021 Jan 21]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f68-176

13. Wagner WC, Stauffer TM. Three-Year-Old Pink Salmon in Lake Superior Tributaries. Trans Am Fish Soc. 1980 Jul 1;109(4):458–60.

14. MacKinnon CN, Donaldson EM. Environmentally Induced Precocious Sexual Development in the Male Pink Salmon (Oncorhynchus gorbuscha). J Fish Board Can [Internet]. 1976 Nov 1 [cited 2021 Jan 28]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f76-307

15. Beacham TD, McIntosh B, MacConnachie C, Spilsted B, White BA. Population structure of pink salmon (Oncorhynchus gorbuscha) in British Columbia and Washington, determined with microsatellites. Fish Bull. 2012 Apr;110(2):242–56.

16. Thedinga JF, Wertheimer AC, Heintz RA, Maselko JM, Rice SD. Effects of stock, coded-wire tagging, and transplant on straying of pink salmon (Oncorhynchus gorbuscha) in southeastern Alaska. Can J Fish Aquat Sci [Internet]. 2011 Apr 12 [cited 2021 Feb 24]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f00-163

17. Gallagher ZS, Bystriansky JS, Farrell AP, Brauner CJ. A novel pattern of smoltification in the most anadromous salmonid: pink salmon (Oncorhynchus gorbuscha). Can J Fish Aquat Sci [Internet]. 2012 Dec 20 [cited 2021 Jan 13]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/cjfas-2012-0390

18. Beacham TD, Withler RE, Gould AP. Biochemical Genetic Stock Identification of Pink Salmon (Oncorhynchus gorbuscha) in Southern British Columbia and Puget Sound. Can J Fish Aquat Sci [Internet]. 1985 Sep 1 [cited 2021 Jan 28]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f85-185

19. Beacham TD, Withler RE, Murray CB, Barner LW. Variation in Body Size, Morphology, Egg Size, and Biochemical Genetics of Pink Salmon in British Columbia. Trans Am Fish Soc. 1988 Mar 1;117(2):109–26.

20. Hawkins SL, Varnavskaya NV, Matzak EA, Efremov VV, Guthrie CM, Wilmot RL, et al. Population structure of odd-broodline Asian pink salmon and its contrast to the even-broodline structure. J Fish Biol. 2002;60(2):370–88.

21. Phillips RB, Kapuscinski AR. High frequency of translocation heterozygotes in odd year populations of pink salmon (Oncorhynchus gorbuscha). Cytogenet Genome Res. 1988;48(3):178–82.

22. Brykov A vl, Polyakova N, Skurikhina LA, Kukhlevsky AD. Geographical and temporal mitochondrial DNA variability in populations of pink salmon. J Fish Biol. 1996;48(5):899–909.

23. Tarpey CM, Seeb JE, McKinney GJ, Templin WD, Bugaev A, Sato S, et al. Single-nucleotide polymorphism data describe contemporary population structure and diversity in allochronic lineages of pink salmon (Oncorhynchus gorbuscha). Can J Fish Aquat Sci [Internet]. 2018 Jun [cited 2020 Oct 30]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/cjfas-2017-0023

24. Beacham TD, Murray CB. Variation in Length and Body Depth of Pink Salmon (Oncorhynchus gorbuscha) and Chum Salmon (O. keta) in Southern British Columbia. Can J Fish Aquat Sci [Internet]. 2011 Apr 10 [cited 2021 Jan 28]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f85-040

25. Godfrey H. Variations in Annual Average Weights of British Columbia Pink Salmon, 1944–1958. J Fish Board Can [Internet]. 2011 Apr 13 [cited 2021 Jan 28]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f59-026

26. Hoar W. The Chum and Pink Salmon Fisheries of British Columbia 1917-1947. Fisheries Research Board of Canada; 1951 p. 46. Report No.: 90.

27. Beacham TD, Murray CB. Variation in developmental biology of pink salmon (Oncorhynchus gorbuscha) in British Columbia. Can J Zool [Internet]. 2011 Feb 14 [cited 2021 Jan 25]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/z88-388

28. Shedlock AM, Parker JD, Crispin DA, Pietsch TW, Burmer GC. Evolution of the salmonid mitochondrial control region. Mol Phylogenet Evol. 1992 Sep 1;1(3):179–92.

29. Crête-Lafrenière A, Weir LK, Bernatchez L. Framing the Salmonidae Family Phylogenetic Portrait: A More Complete Picture from Increased Taxon Sampling. PLOS ONE. 2012 Oct 5;7(10):e46662.

30. Campbell MA, López JA, Sado T, Miya M. Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes+Salmoniformes based on whole mitochondrial genome sequences. Gene. 2013 Nov 1;530(1):57–65.

31. Smith GR. Introgression in Fishes: Significance for Paleontology, Cladistics, and Evolutionary Rates. Syst Biol. 1992 Mar 1;41(1):41–57.

32. McKay SJ, Devlin RH, Smith MJ. Phylogeny of Pacific salmon and trout based on growth hormone type-2 and mitochondrial NADH dehydrogenase subunit 3 DNA sequences. Can J Fish Aquat Sci. 1996 May 1;53(5):1165–76.

33. Churikov D, Gharrett AJ. Comparative phylogeography of the two pink salmon broodlines: an analysis based on a mitochondrial DNA genealogy. Mol Ecol. 2002;11(6):1077–101.

34. Podlesnykh AV, Kukhlevsky AD, Brykov VA. A comparative analysis of mitochondrial DNA genetic variation and demographic history in populations of even- and odd-year broodline pink salmon, Oncorhynchus gorbuscha (Walbaum, 1792), from Sakhalin Island. Environ Biol Fishes. 2020 Dec 1;103(12):1553–64.

35. Kwain W, Chappel JA. First Evidence for Even-Year Spawning Pink Salmon, Oncorhynchus gorbuscha, in Lake Superior. J Fish Board Can [Internet]. 2011 Apr 13 [cited 2021 Feb 4]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f78-216

36. Bagdovitz MS, Taylor WW, Wagner WC, Nicolette JP, Spangler GR. Pink Salmon Populations in the U.S. Waters of Lake Superior, 1981–1984. J Gt Lakes Res. 1986 Jan 1;12(1):72–81.

37. Beacham TD, Murray CB. Influence of photoperiod and temperature on timing of sexual maturity of pink salmon (Oncorhynchus gorbuscha). Can J Zool [Internet]. 2011 Feb 14 [cited 2021 May 13]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/z88-249

38. Krkošek M, Hilborn R, Peterman RM, Quinn TP. Cycles, stochasticity and density dependence in pink salmon population dynamics. Proc R Soc B Biol Sci. 2011 Jul 7;278(1714):2060–8.

39. Irvine JR, Michielsens CJG, O'Brien M, White BA, Folkes M. Increasing Dominance of Odd-Year Returning Pink Salmon. Trans Am Fish Soc. 2014;143(4):939–56.

40. Quinn TP. Variation in Pacific Salmon Reproductive Behaviour Associated with Species, Sex and Levels of Competition. Behaviour. 1999;136(2):179–204.

41. Springer AM, van Vliet GB. Climate change, pink salmon, and the nexus between bottom-up and top-down forcing in the subarctic Pacific Ocean and Bering Sea. Proc Natl Acad Sci U S A. 2014 May 6;111(18):E1880–8.

42. Ruggerone GT, Nielsen JL. Evidence for competitive dominance of Pink salmon (Oncorhynchus gorbuscha) over other Salmonids in the North Pacific Ocean. Rev Fish Biol Fish. 2004 Sep 1;14(3):371–90.

43. Tadokoro K, Ishida Y, Davis ND, Ueyanagi S, Sugimoto T. Change in chum salmon (Oncorhynchus keta) stomach contents associated with fluctuation of pink salmon (O. gorbuscha) abundance in the central subarctic Pacific and Bering Sea. Fish Oceanogr. 1996;5(2):89–99.

44. Ishida Y, Azumaya T, Fukuwaka M, Davis N. Interannual variability in stock abundance and body size of Pacific salmon in the central Bering Sea. Prog Oceanogr. 2002 Oct 1;55(1):223–34.

45. Kaga T, Sato S, Azumaya T, Davis ND, Fukuwaka M. Lipid content of chum salmon Oncorhynchus keta affected by pink salmon O. gorbuscha abundance in the central Bering Sea. Mar Ecol Prog Ser. 2013 Mar 25;478:211–21.

46. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019 May;37(5):540–6.

47. Vaser R, Sovic I, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017 Jan 18;gr.214270.116.

48. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094–100.

49. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE. 2014 Nov 19;9(11):e112963.

50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

51. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio [Internet]. 2013 Mar 16 [cited 2017 Dec 19]; Available from: http://arxiv.org/abs/1303.3997

52. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009 Aug 15;25(16):2078–9.

54. Limborg MT, Waples RK, Seeb JE, Seeb LW. Temporally Isolated Lineages of Pink Salmon Reveal Unique Signatures of Selection on Distinct Pools of Standing Genetic Variation. J Hered. 2014 Nov 1;105(6):835–45.

55. Catchen J, Amores A, Bassham S. Chromonomer: A Tool Set for Repairing and Enhancing Assembled Genomes Through Integration of Genetic Maps and Conserved Synteny. G3 Genes Genomes Genet. 2020 Nov 1;10(11):4115–28.

56. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019 Oct 28;20(1):224.

57. KrisChristensen. KrisChristensen/CompareAGP [Internet]. 2021 [cited 2021 May 19]. Available from: https://github.com/KrisChristensen/CompareAGP

58. ArimaGenomics/mapping_pipeline [Internet]. Arima Genomics, Inc.; 2021 [cited 2021 May 19]. Available from: https://github.com/ArimaGenomics/mapping_pipeline

59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

60. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLOS Comput Biol. 2019 Aug 21;15(8):e1007273.

61. Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. BMC Genomics. 2017 Jul 12;18(1):527.

62. Tarpey CM, Seeb JE, McKinney GJ, Seeb LW. A dense linkage map for odd-year lineage pink salmon incorporating duplicated loci. School of Aquatic Fishery Sciences: University of Washington; 2017 p. 50. Report No.: COOP-13-085.

63. Gao G, Magadan S, Waldbieser GC, Youngblood RC, Wheeler PA, Scheffler BE, et al. A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. G3 Bethesda Md. 2021 Apr 15;11(4).

64. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. 2016 Jul;3(1):99–101.

65. phasegenomics/juicebox_scripts [Internet]. Phase Genomics; 2021 [cited 2021 May 19]. Available from: https://github.com/phasegenomics/juicebox_scripts

66. Sutherland BJG, Gosselin T, Normandeau E, Lamothe M, Isabel N, Audet C, et al. Salmonid Chromosome Evolution as Revealed by a Novel Method for Comparing RADseq Linkage Maps. Genome Biol Evol. 2016 Dec 1;8(12):3600–17.

67. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct 1;31(19):3210–2.

68. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. Genome Res [Internet]. 2009 Jun 18 [cited 2015 May 21]; Available from: http://genome.cshlp.org/content/early/2009/06/15/gr.092759.109

69. Soderlund C, Bomhoff M, Nelson WM. SyMAP v3.4: a turnkey synteny system with application to plant genomes. Nucleic Acids Res. 2011 May;39(10):e68.

70. Christensen KA, Leong JS, Sakhrani D, Biagi CA, Minkley DR, Withler RE, et al. Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. PLOS ONE. 2018 Apr 5;13(4):e0195461.

71. KrisChristensen. KrisChristensen/NCBIGenomeRepeats [Internet]. 2021 [cited 2021 May 12]. Available from: https://github.com/KrisChristensen/NCBIGenomeRepeats

72. Genomic DNA Preparation from RNAlater™ Preserved Tissues - CA [Internet]. [cited 2019 Dec 19]. Available from: https://www.thermofisher.com/ca/en/home/references/protocols/nucleic-acid-purification-and-analysis/rna-protocol/genomic-dna-preparation-from-rnalater-preserved-tissues.html

73. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–303.

74. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May;43(5):491–8.

75. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43:11.10.1-11.10.33.

76. broadinstitute/picard [Internet]. Broad Institute; 2020 [cited 2020 Dec 9]. Available from: https://github.com/broadinstitute/picard

77. Christensen KA, Rondeau EB, Minkley DR, Sakhrani D, Biagi CA, Flores A-M, et al. The sockeye salmon genome, transcriptome, and analyses identifying population defining regions of the genome. PLOS ONE. 2020 Oct 29;15(10):e0240935.

78. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.

79. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov 1;27(21):2987–93.

80. KrisChristensen. KrisChristensen/MapVCF2NewGenome [Internet]. 2021 [cited 2021 May 19]. Available from: https://github.com/KrisChristensen/MapVCF2NewGenome

81. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010 Oct 15;11(1):94.

82. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

83. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinforma Oxf Engl. 2008 Jun 1;24(11):1403–5.

84. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour. 2017 Jan;17(1):44–53.

85. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. Available from: http://ggplot2.org

86. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep;19(9):1655–64.

87. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience [Internet]. 2015 Dec 1 [cited 2020 Feb 21];4(1). Available from: https://academic.oup.com/gigascience/article/4/1/s13742-015-0047-8/2707533

88. PLINK 1.9 [Internet]. [cited 2018 Jun 1]. Available from: http://www.cog-genomics.org/plink/1.9/

89. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinforma Oxf Engl. 2012 Dec 15;28(24):3326–8.

90. Samuels DC, Wang J, Ye F, He J, Levinson RT, Sheng Q, et al. Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. Genetics. 2016 Nov 1;204(3):893–904.

91. Guo Y, Ye F, Sheng Q, Clark T, Samuels DC. Three-stage quality control strategies for DNA re-sequencing data. Brief Bioinform. 2014 Nov;15(6):879–89.

92. KrisChristensen. KrisChristensen/SharedAllelesVCF [Internet]. 2021 [cited 2021 May 19]. Available from: https://github.com/KrisChristensen/SharedAllelesVCF

93. Chakraborty R, Jin L. A unified approach to study hypervariable polymorphisms: Statistical considerations of determining relatedness and population distances. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ, editors. DNA Fingerprinting: State of the Science [Internet]. Basel:

Birkhäuser; 1993 [cited 2021 May 19]. p. 153–75. (Progress in Systems and Control Theory). Available from: https://doi.org/10.1007/978-3-0348-8583-6_14

94. Mountain JL, Cavalli-Sforza LL. Multilocus genotypes, a tree of individuals, and human evolutionary history. Am J Hum Genet. 1997 Sep;61(3):705–18.

95. Witherspoon DJ, Wooding S, Rogers AR, Marchani EE, Watkins WS, Batzer MA, et al. Genetic Similarities Within and Between Human Populations. Genetics. 2007 May;176(1):351–9.

96. Wickham H. Reshaping Data with the reshape Package. J Stat Softw. 2007 Nov 13;21(1):1–20.

97. pheatmap: Pretty Heatmaps [Internet]. Comprehensive R Archive Network (CRAN); [cited 2021 May 19]. Available from: https://CRAN.R-project.org/package=pheatmap

98. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014 Jul;31(7):1929–36.

99. Wickham H, François R, Henry L, Müller K, RStudio. dplyr: A Grammar of Data Manipulation [Internet]. 2021 [cited 2021 Feb 12]. Available from: https://CRAN.R-project.org/package=dplyr

100. Wickham H, RStudio. tidyr: Tidy Messy Data [Internet]. 2020 [cited 2021 Feb 12]. Available from: https://CRAN.R-project.org/package=tidyr

101. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013 Jun 1;22(11):3124–40.

102. Chen G-B, Lee SH, Zhu Z-X, Benyamin B, Robinson MR. EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. Heredity. 2016 Jul;117(1):51–61.

103. gc5k/GEAR [Internet]. GitHub. [cited 2020 Feb 21]. Available from: https://github.com/gc5k/GEAR

104. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014 Jan 1;005165.

105. Wickham H. stringr: Simple, Consistent Wrappers for Common String Operations [Internet]. 2018. Available from: https://CRAN.R-project.org/package=stringr

106. Pérez-Wohlfeil E, Diaz-del-Pino S, Trelles O. Ultra-fast genome comparison for large-scale genomic experiments. Sci Rep. 2019 Jul 16;9(1):10274.

107. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013 Jan 3;14(2):178–92.

108. KrisChristensen. KrisChristensen/VCFstats [Internet]. 2021 [cited 2021 May 19]. Available from: https://github.com/KrisChristensen/VCFstats

109. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016 May 12;533(7602):200–5.

110. Phillips RB, Matsuoka MP, Smoker WW, Gharrett AJ. Inheritance of a chromosomal polymorphism in odd-year pink salmon from southeastern Alaska. Genome [Internet]. 2011 Feb 15 [cited 2021 Jan 29]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/g99-010

111. Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, et al. An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, Oncorhynchus mykiss. Curr Biol. 2012 Aug 7;22(15):1423–8.

112. Devlin RH, Biagi CA, Smailus DE. Genetic mapping of Y-chromosomal DNA markers in Pacific salmon. Genetica. 2001;111(1–3):43–58.

113. Muttray AF, Sakhrani D, Smith JL, Nakayama I, Davidson WS, Park L, et al. Deletion and Copy Number Variation of Y-Chromosomal Regions in Coho Salmon, Chum Salmon, and Pink Salmon Populations. Trans Am Fish Soc. 2017 Mar 4;146(2):240–51.

114. Sato S, Urawa S. Genetic variation of Japanese pink salmon populations inferred from nucleotide sequence analysis of the mitochondrial DNA control region. Environ Biol Fishes. 2017 Oct 1;100(10):1355–72.

115. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018 Apr;19(4):220–34.

116. Lampson MA, Black BE. Cellular and Molecular Mechanisms of Centromere Drive. Cold Spring Harb Symp Quant Biol. 2017;82:249–57.

117. Henikoff S, Ahmad K, Malik HS. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. Science. 2001 Aug 10;293(5532):1098–102.

118. Chmátal L, Schultz RM, Black BE, Lampson MA. Cell Biology of Cheating—Transmission of Centromeres and Other Selfish Elements Through Asymmetric Meiosis. In: Black BE, editor. Centromeres and Kinetochores: Discovering the Molecular Mechanisms Underlying Chromosome Inheritance [Internet]. Cham: Springer International Publishing; 2017. p. 377–96. Available from: https://doi.org/10.1007/978-3-319-58592-5_16

119. Ichikawa K, Tomioka S, Suzuki Y, Nakamura R, Doi K, Yoshimura J, et al. Centromere evolution and CpG methylation during vertebrate speciation. Nat Commun. 2017 Nov 28;8(1):1833.

120. Kjærner-Semb E, Ayllon F, Furmanek T, Wennevik V, Dahle G, Niemelä E, et al. Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection. BMC Genomics. 2016 Aug 11;17(1):610.

121. Zueva KJ, Lumme J, Veselov AE, Kent MP, Lien S, Primmer CR. Footprints of Directional Selection in Wild Atlantic Salmon Populations: Evidence for Parasite-Driven Evolution? PLOS ONE. 2014 Mar 26;9(3):e91672.

122. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. The major histocompatibility complex and its functions. Immunobiol Immune Syst Health Dis 5th Ed [Internet]. 2001 [cited 2021 Mar 9]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK27156/

123. Grimholt U. MHC and Evolution in Teleosts. Biology [Internet]. 2016 Jan 19 [cited 2021 Mar 9];5(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4810163/

124. Langefors Å, Lohm J, Grahn M, Andersen Ø, Schantz T von. Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon. Proc R Soc Lond B Biol Sci. 2001 Mar 7;268(1466):479–85.

125. Miller KM, Winton JR, Schulze AD, Purcell MK, Ming TJ. Major Histocompatibility Complex Loci are Associated with Susceptibility of Atlantic Salmon to Infectious Hematopoietic Necrosis Virus. Environ Biol Fishes. 2004 Mar 1;69(1):307–16.

126. Dionne M, Miller KM, Dodson JJ, Bernatchez L. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Philos Trans R Soc B Biol Sci. 2009 Jun 12;364(1523):1555–65.

127. Clark EA, Giltiay NV. CD22: A Regulator of Innate and Adaptive B Cell Responses and Autoimmunity. Front Immunol [Internet]. 2018 Sep 28 [cited 2021 Mar 8];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6173129/

128. Fernandes VE, Ercoli G, Bénard A, Brandl C, Fahnenstiel H, Müller-Winkler J, et al. The B-cell inhibitory receptor CD22 is a major factor in host resistance to Streptococcus pneumoniae infection. PLOS Pathog. 2020 Apr 23;16(4):e1008464.

129. Seki M, Gearhart PJ, Wood RD. DNA polymerases and somatic hypermutation of immunoglobulin genes. EMBO Rep. 2005 Dec;6(12):1143–8.

130. Yang F, Waldbieser GC, Lobb CJ. The Nucleotide Targets of Somatic Mutation and the Role of Selection in Immunoglobulin Heavy Chains of a Teleost Fish. J Immunol. 2006 Feb 1;176(3):1655–67.

131. Bilal S, Lie KK, Sæle Ø, Hordvik I. T Cell Receptor Alpha Chain Genes in the Teleost Ballan Wrasse (Labrus bergylta) Are Subjected to Somatic Hypermutation. Front Immunol [Internet]. 2018 [cited 2021 Mar 2];9. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2018.01101/full

132. Flajnik MF. A cold-blooded view of adaptive immunity. Nat Rev Immunol. 2018 Jul;18(7):438–53.

133. Lerner LK, Nguyen TV, Castro LP, Vilar JB, Munford V, Le Guillou M, et al. Large deletions in immunoglobulin genes are associated with a sustained absence of DNA Polymerase η. Sci Rep. 2020 Jan 28;10(1):1311.

134. Ratliff MLPD, Templeton TD, Ward JM, Webb CFP. The Bright Side of Hematopoiesis: Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. Front Immunol

[Internet]. 2014 [cited 2021 Feb 19];5. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2014.00113/full

135. Qiu F, Tang R, Zuo X, Shi X, Wei Y, Zheng X, et al. A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. Nat Commun [Internet]. 2017 Apr 20 [cited 2021 Mar 3];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5429142/

136. Bzowska A, Kulikowska E, Shugar D. Purine nucleoside phosphorylases: properties, functions, and clinical aspects. Pharmacol Ther. 2000 Dec 1;88(3):349–425.

137. Ting L-M, Gissot M, Coppi A, Sinnis P, Kim K. Attenuated Plasmodium yoelii lacking purine nucleoside phosphorylase confer protective immunity. Nat Med. 2008 Sep;14(9):954–8.

138. Kang Y-N, Zhang Y, Allan PW, Parker WB, Ting J-W, Chang C-Y, et al. Structure of grouper iridovirus purine nucleoside phosphorylase. Acta Crystallogr D Biol Crystallogr. 2010 Feb;66(Pt 2):155–62.

139. Wang Y, Wang W, Xu L, Zhou X, Shokrollahi E, Felczak K, et al. Cross Talk between Nucleotide Synthesis Pathways with Cellular Immunity in Constraining Hepatitis E Virus Replication. Antimicrob Agents Chemother. 2016 May 1;60(5):2834–48.

140. Dziekan JM, Yu H, Chen D, Dai L, Wirjanata G, Larsson A, et al. Identifying purine nucleoside phosphorylase as the target of quinine using cellular thermal shift assay. Sci Transl Med [Internet]. 2019 Jan 2 [cited 2021 Mar 1];11(473). Available from: https://stm.sciencemag.org/content/11/473/eaau3174

141. Satterfield DA, Marra PP, Sillett TS, Altizer S. Responses of migratory species and their pathogens to supplemental feeding. Philos Trans R Soc B Biol Sci. 2018 May 5;373(1745):20170094.

142. Kołodziej-Sobocińska M. Factors affecting the spread of parasites in populations of wild European terrestrial mammals. Mammal Res. 2019 Jul 1;64(3):301–18.

143. Krkošek M. Host density thresholds and disease control forfisheries and aquaculture. Aquac Environ Interact. 2010 May 20;1:21–32.

144. Cheng CL, Flamarique IN, Hárosi FI, Rickers-Haunerland J, Haunerland NH. Photoreceptor layer of salmonid fishes: transformation and loss of single cones in juvenile fish. J Comp Neurol. 2006 Mar 10;495(2):213–35.

145. Flamarique IN. Light exposure during embryonic and yolk-sac alevin development of Chinook salmon Oncorhynchus tshawytscha does not alter the spectral phenotype of photoreceptors. J Fish Biol. 2019;95(1):214–21.

146. Ogawa Y, Shiraki T, Asano Y, Muto A, Kawakami K, Suzuki Y, et al. Six6 and Six7 coordinately regulate expression of middle-wavelength opsins in zebrafish. Proc Natl Acad Sci. 2019 Mar 5;116(10):4651–60.

147. López-Ríos J, Tessmar K, Loosli F, Wittbrodt J, Bovolenta P. Six3 and Six6 activity is modulated by members of the groucho family. Development. 2003 Jan 1;130(1):185–95.

148. Xie H, Hoffmann HM, Meadows JD, Mayo SL, Trang C, Leming SS, et al. Homeodomain Proteins SIX3 and SIX6 Regulate Gonadotrope-specific Genes During Pituitary Development. Mol Endocrinol. 2015 Jun 1;29(6):842–55.

149. Ayllon F, Kjærner-Semb E, Furmanek T, Wennevik V, Solberg MF, Dahle G, et al. The vgll3 Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (Salmo salar L.) Males. PLOS Genet. 2015 Nov 9;11(11):e1005628.

150. Barson NJ, Aykanat T, Hindar K, Baranski M, Bolstad GH, Fiske P, et al. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature. 2015 Dec;528(7582):405–8.

151. Aykanat T, Rasmussen M, Ozerov M, Niemelä E, Paulin L, Vähä J-P, et al. Life-history genomic regions explain differences in Atlantic salmon marine diet specialization. J Anim Ecol. 2020;89(11):2677–91.

152. Yu Y, Shintani T, Takeuchi Y, Shirasawa T, Noda M. Protein Tyrosine Phosphatase Receptor Type J (PTPRJ) Regulates Retinal Axonal Projections by Inhibiting Eph and Abl Kinases in Mice. J Neurosci. 2018 Sep 26;38(39):8345–63.

153. Baslow MH. Neurosine, its identification with N-acetyl-L-histidine and distribution in aquatic vertebrates. Zoologica. 1965;50:63–6.

154. Baslow MH. A Review of Phylogenetic and Metabolic Relationships Between the Acylamino Acids, N-Acetyl-l-Aspartic Acid and N-Acetyl-l-Histidine, in the Vertebrate Nervous System. J Neurochem. 1997;68(4):1335–44.

155. Yamada S, Furuichi M. Nα-acetylhistidine metabolism in fish—I. Identification of Nα-acetylhistidine in the heart of rainbow trout Salmo gairdneri. Comp Biochem Physiol Part B Comp Biochem. 1990 Jan 1;97(3):539–41.

156. Yamada S, Tanaka Y, Sameshima M, Furuichi M. Effects of starvation and feeding on tissue Nα -acetylhistidine levels in Nile tilapia Oreochromis niloticus. Comp Biochem Physiol A Physiol. 1994 Oct 1;109(2):277–83.

157. Breck O, Rhodes J, Waagbø R, Bjerkås E, Sanderson J. Role of Histidine in Cataract Formation in Atlantic Salmon (Salmo salar L). Invest Ophthalmol Vis Sci. 2003 May 1;44(13):3494–3494.

158. Rhodes JD, Breck O, Waagbo R, Bjerkas E, Sanderson J. N-acetylhistidine, a novel osmolyte in the lens of Atlantic salmon (Salmo salar L.). Am J Physiol-Regul Integr Comp Physiol. 2010 Jul 21;299(4):R1075–81.

159. Yamada S, Arikawa S. An ectotherm homologue of human predicted gene NAT16 encodes histidine N-acetyltransferase responsible for Nα-acetylhistidine synthesis. Biochim Biophys Acta BBA - Gen Subj. 2014 Jan 1;1840(1):434–42.

160. Baslow MH, Guilfoyle DN. N-acetyl-l-histidine, a Prominent Biomolecule in Brain and Eye of Poikilothermic Vertebrates. Biomolecules. 2015 Apr 24;5(2):635–46.

161. Forman OP, Hitti RJ, Boursnell M, Miyadera K, Sargan D, Mellersh C. Canine genome assembly correction facilitates identification of a MAP9 deletion as a potential age of onset modifier for RPGRIP1-associated canine retinal degeneration. Mamm Genome Off J Int Mamm Genome Soc. 2016 Jun;27(5–6):237–45.

162. Li X, Mao X-B, Hei R-Y, Zhang Z-B, Wen L-T, Zhang P-Z, et al. Protective role of hydrogen sulfide against noise-induced cochlear damage: a chronic intracochlear infusion model. PloS One. 2011;6(10):e26728.

163. Kimura H. Production and Physiological Effects of Hydrogen Sulfide. Antioxid Redox Signal. 2014 Feb 10;20(5):783–93.

164. Nagtegaal AP, Broer L, Zilhao NR, Jakobsdottir J, Bishop CE, Brumat M, et al. Genome-wide association meta-analysis identifies five novel loci for age-related hearing impairment. Sci Rep. 2019 Oct 23;9(1):15192.

165. Yonezawa A, Inui K. Importance of the multidrug and toxin extrusion MATE/SLC47A family to pharmacokinetics, pharmacodynamics/toxicodynamics and pharmacogenomics. Br J Pharmacol. 2011 Dec;164(7):1817–25.

166. Lončar J, Popović M, Krznar P, Zaja R, Smital T. The first characterization of multidrug and toxin extrusion (MATE/SLC47) proteins in zebrafish (Danio rerio). Sci Rep. 2016 Jun 30;6(1):28937.

167. Mittelbach GG, Ballew NG, Kjelvik MK. Fish behavioral types and their ecological consequences. Can J Fish Aquat Sci [Internet]. 2014 Feb 26 [cited 2021 Mar 11]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/cjfas-2013-0558

168. López ME, Linderoth T, Norris A, Lhorente JP, Neira R, Yáñez JM. Multiple Selection Signatures in Farmed Atlantic Salmon Adapted to Different Environments Across Hemispheres. Front Genet [Internet]. 2019 [cited 2021 Mar 11];10. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2019.00901/full

169. Jiang P, Scarpa JR, Fitzpatrick K, Losic B, Gao VD, Hao K, et al. A Systems Approach Identifies Networks and Genes Linking Sleep and Stress: Implications for Neuropsychiatric Disorders. Cell Rep. 2015 May 5;11(5):835–48.

170. Gley K, Murani E, Trakooljul N, Zebunke M, Puppe B, Wimmers K, et al. Transcriptome profiles of hypothalamus and adrenal gland linked to haplotype related to coping behavior in pigs. Sci Rep. 2019 Sep 10;9(1):13038.

171. Gilks WP, Hill M, Gill M, Donohoe G, Corvin AP, Morris DW. Functional investigation of a schizophrenia GWAS signal at the CDC42 gene. World J Biol Psychiatry. 2012 Oct 1;13(7):550–4.

172. Bevilacqua L, Goldman D. Genetics of impulsive behaviour. Philos Trans R Soc B Biol Sci. 2013 Apr 5;368(1615):20120380.

173. Montalvo-Ortiz JL, Zhou H, D'Andrea I, Maroteaux L, Lori A, Smith A, et al. Translational studies support a role for serotonin 2B receptor (HTR2B) gene in aggression-related cannabis response. Mol Psychiatry. 2018 Dec;23(12):2277–86.

174. Du SJ, Devlin RH, Hew CL. Genomic structure of growth hormone genes in chinook salmon (Oncorhynchus tshawytscha): presence of two functional genes, GH-I and GH-II, and a male-specific pseudogene, GH-psi. DNA Cell Biol. 1993 Oct;12(8):739–51.

175. Devlin RH, McNeil BK, Groves TDD, Donaldson EM. Isolation of a Y-Chromosomal DNA Probe Capable of Determining Genetic Sex in Chinook Salmon (Oncorhynchus tshawytscha). Can J Fish Aquat Sci [Internet]. 2011 Apr 11 [cited 2021 May 20]; Available from: https://cdnsciencepub.com/doi/abs/10.1139/f91-190

176. Woram RA, Gharbi K, Sakamoto T, Hoyheim B, Holm L-E, Naish K, et al. Comparative Genome Analysis of the Primary Sex-Determining Locus in Salmonid Fishes. Genome Res. 2003 Jan 2;13(2):272–80.

177. Gabián M, Morán P, Fernández AI, Villanueva B, Chtioui A, Kent MP, et al. Identification of genomic regions regulating sex determination in Atlantic salmon using high density SNP data. BMC Genomics. 2019 Oct 22;20(1):764.

178. Kijas J, McWilliam S, Naval Sanchez M, Kube P, King H, Evans B, et al. Evolution of Sex Determination Loci in Atlantic Salmon. Sci Rep. 2018 Apr 4;8(1):5664.

179. Eisbrenner WD, Botwright N, Cook M, Davidson EA, Dominik S, Elliott NG, et al. Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (Salmo salar). Heredity. 2014 Jul;113(1):86–92.

180. McKinney GJ, Nichols KM, Ford MJ. A mobile sex-determining region, male-specific haplotypes and rearing environment influence age at maturity in Chinook salmon. Mol Ecol. 2021;30(1):131–47.

181. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, et al. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. Evol Appl. 2013 Apr;6(3):486–96.

Supporting information

S1 File. Sample information. The sample tab has metadata about each sample, including information on sex, river, and year-class (latitude and longitude locations are approximate). The StatsAllFilters

shows metrics from the .vcf file after filtering for LD (see methods). Stats1stFilter has the same information, but from the .vcf file after only preliminary filtering (see methods). The eigenGWAS tab contains the DAPC values used in the eigenGWAS analysis (see methods). The Mitochondrion tab shows metadata used to generate the mitochondria figures. The GPS tab shows the coordinates used in the sample map. The Admixture tab has the values output from the admixture analysis. For each tab with LG, these sheets have manually genotyped areas and calculations of HWE. The PrivateAlleles tab has metrics output from Stacks. The SharedAlleles tab has a matrix of shared alleles between individuals in long format and statistics on the right. The Y-Chrom tab has information about the *sdY* haplotypes. The GWAS tab has metadata used in the GWAS analysis. The GHp tab displays the alignments of the growth hormone pseudogene and *sdY* gene to the odd and even genome. The even genome placements will change after processing by NCBI.

S1 Fig. Chromosomal polymorphism at centromere on LG15_El08.2-20.1. Depiction of LG15_El08.2-20.1 and a chromosomal polymorphism, either a deletion or evidence of a chromosomal fusion. At the top, LG15_El08.2-20.1 is depicted with the distance and location of the purposed polymorphism. Scaffolds/contigs that comprise the region surrounding the polymorphism are shown below the chromosomal depiction, with a blue arrow showing where multiple small contigs were placed. Below the scaffolds, synteny with rainbow trout and Northern pike is shown based on CHROMEISTER (106) alignments. Finally, ONT/Nanopore reads that were used to generate the genome assemblies were aligned back to the odd-year genome and visualized with IGV. Reads in the odd-year individual are shown flanking the deletion (the display was split because the region was too large to adequately visualize otherwise).

S2 Fig. Sex determining region of the even-year pink salmon compared to the rainbow trout Y-chromosome. A) A CHROMEISTER (106) dotplot between the Y-specific portion (top) and shared

723 portion (bottom) of LG20_El14.2 of the even-year pink salmon genome assembly and the rainbow
724 trout Y-chromosome (63). The location of the *sdY* gene is shown based on the position in the rainbow
725 trout chromosome and the genes annotated by NCBI are shown on the x-axis at the bottom. B) A plot
726 of the Hi-C contact map of the even-year pink salmon genome assembly produced by Juicebox (64).
727 There are multiple inversions between the pink salmon and rainbow trout genome, but the contact map
728 supports the order and orientation for the pink salmon genome assembly and these could represent
729 actual inversions between species instead of assembly errors.

730 **S3 Fig. Sex determining region of the even-year pink salmon compared to the coho salmon**
731 **chromosome 29 autosome.** A) A CHROMEISTER (106) dotplot between the Y-specific portion (top)
732 and shared portion (bottom) of LG20_El14.2 of the even-year pink salmon genome assembly and coho
733 salmon chromosome 29. B) A plot of the Hi-C contact map of the even-year pink salmon genome
734 assembly produced by Juicebox (64). There are multiple inversions between the pink salmon and coho
735 salmon genome, but the contact map supports the order and orientation for the pink salmon genome
736 assembly and these could represent actual inversions between species instead of assembly errors.

737 **S4 Fig. Sex determining region of the even-year pink salmon with genotype information.**
738 Genotypes are shown from an IGV (107) screenshot for the 61 samples of pink salmon for the region
739 with the *sdY* sex-determining gene. The top portion shows the distance of the Y-specific genome
740 region (~3.2 Mbp) and the contig/scaffold boundaries that make up this region are shown as vertical
741 lines. Below the distances, allele frequencies for each locus are shown, and below that individual
742 genotypes. The x-axis of the genotypes represent loci and each line on the Y-axis represents an
743 individual pink salmon. The dark-blue colour is a homozygous reference genotype, the light-blue
744 colour a heterozygous genotype, and the green genotype is for a homozygous alternative locus. There
745 are large stretches (1-2 Mbp) of heterozygosity and homozygosity based on sex. Please note that there

746 is a possible inversion (from a mis-assembly) in this region as the runs of homozygosity and

747 heterozygosity are broken by a section from ~600 kbp and ~1,300 kbp.

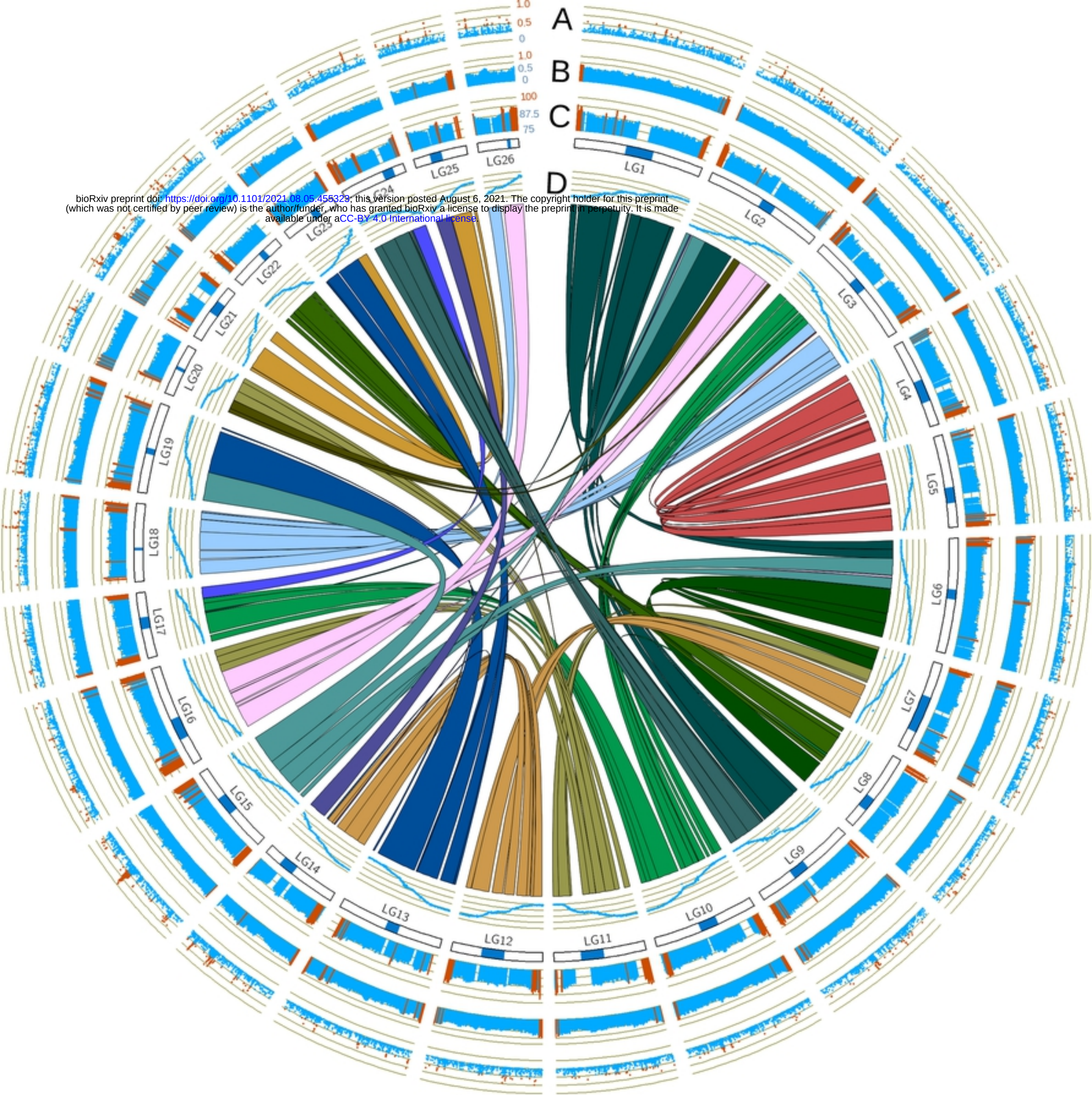


Fig1

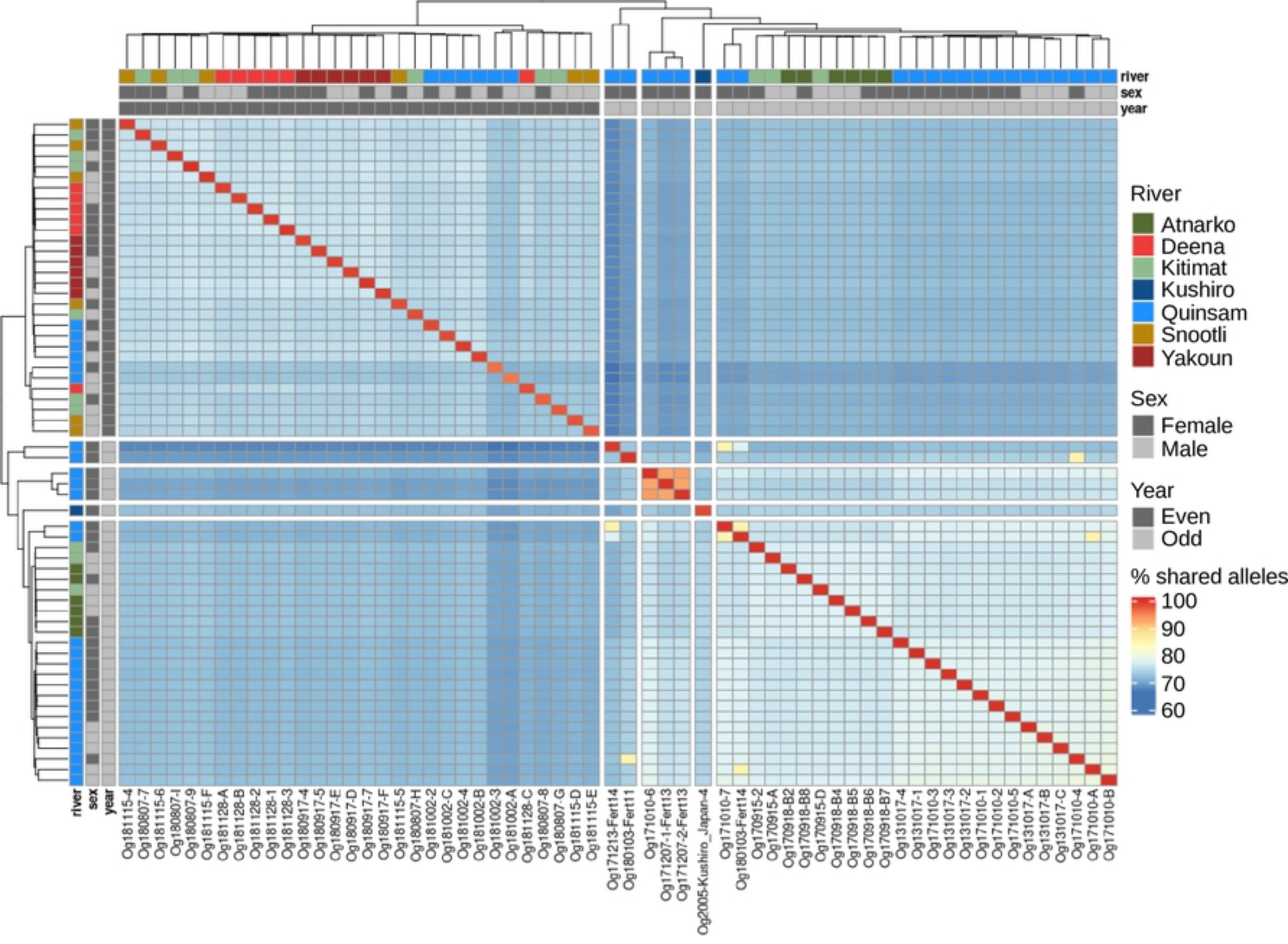
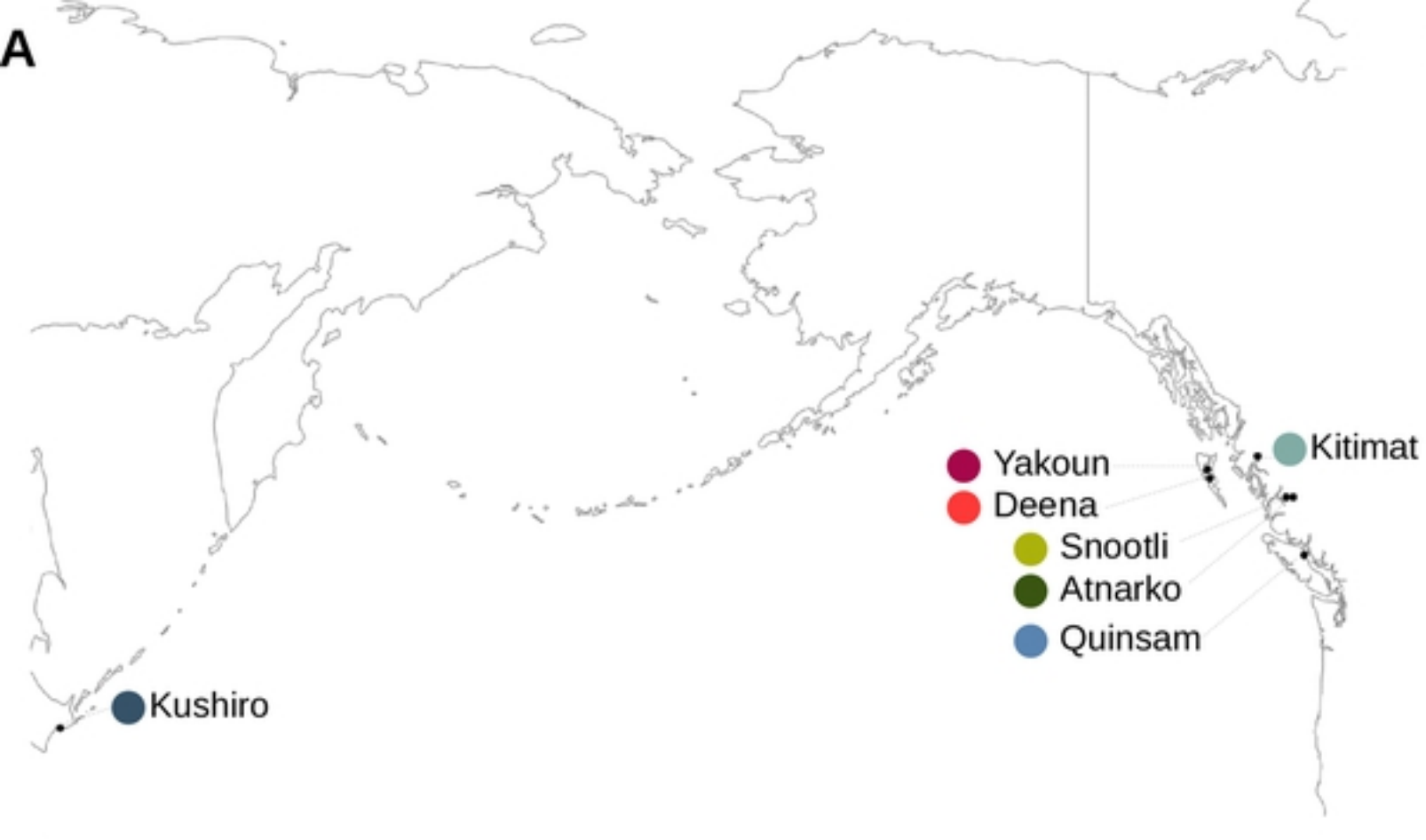

river
sex
year
River
Atnarko
Deena
Kitimat
Kushiro
Quinsam
Snootli
Yakoun
Sex
Female
Male
Year
Even
Odd
% shared alleles
100
90
80
70
60
Og181115-4
Og180807-7
Og181115-6
Og180807-I
Og180807-9
Og181115-F
Og181128-A
Og181128-B
Og181128-2
Og181128-1
Og181128-3
Og180917-4
Og180917-5
Og180917-E
Og180917-D
Og180917-7
Og180917-F
Og181115-5
Og180807-H
Og181002-2
Og181002-C
Og181002-4
Og181002-B
Og181002-3
Og181002-A
Og181128-C
Og180807-8
Og180807-G
Og181115-D
Og181115-E
Og171213-Fert14
Og180103-Fert11
Og171010-6
Og171207-1-Fert13
Og171207-2-Fert13
Og2005-Kushiro_Japan-4
Og171010-7
Og180103-Fert14
Og170915-2
Og170915-A
Og170918-B2
Og170918-B8
Og170915-D
Og170918-B4
Og170918-B5
Og170918-B6
Og170918-B7
Og131017-4
Og131017-1
Og171010-3
Og131017-3
Og131017-2
Og171010-1
Og171010-2
Og171010-5
Og131017-A
Og131017-B
Og131017-C
Og171010-4
Og171010-A
Og171010-B

Fig2


A
Yakoun
Deena
Snootli
Atnarko
Quinsam
Kitimat
Kushiro

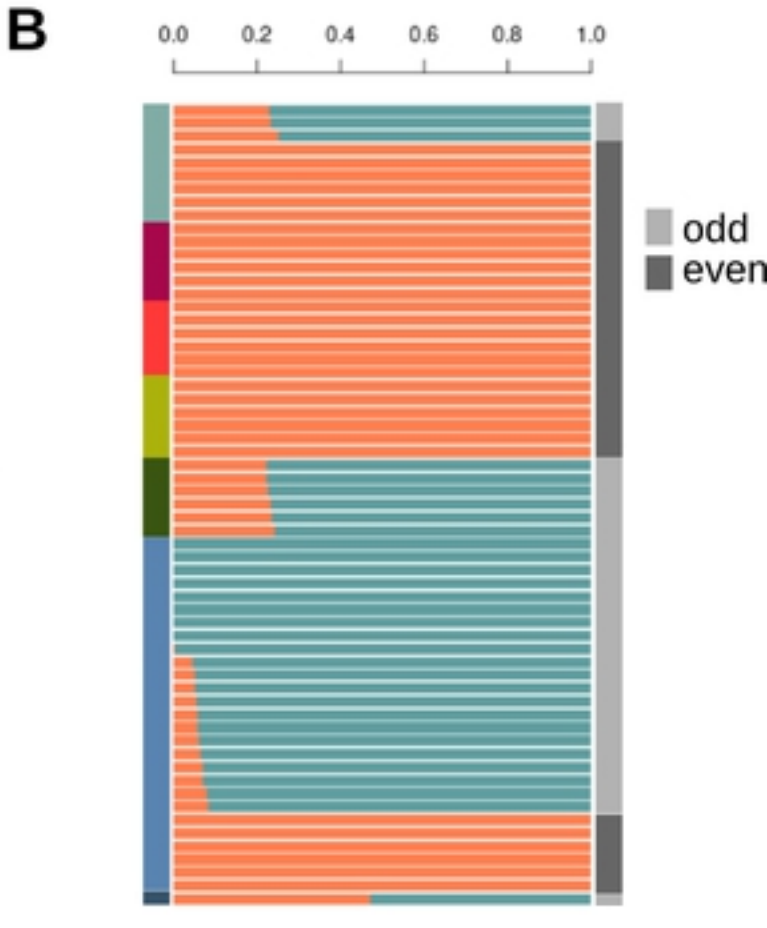
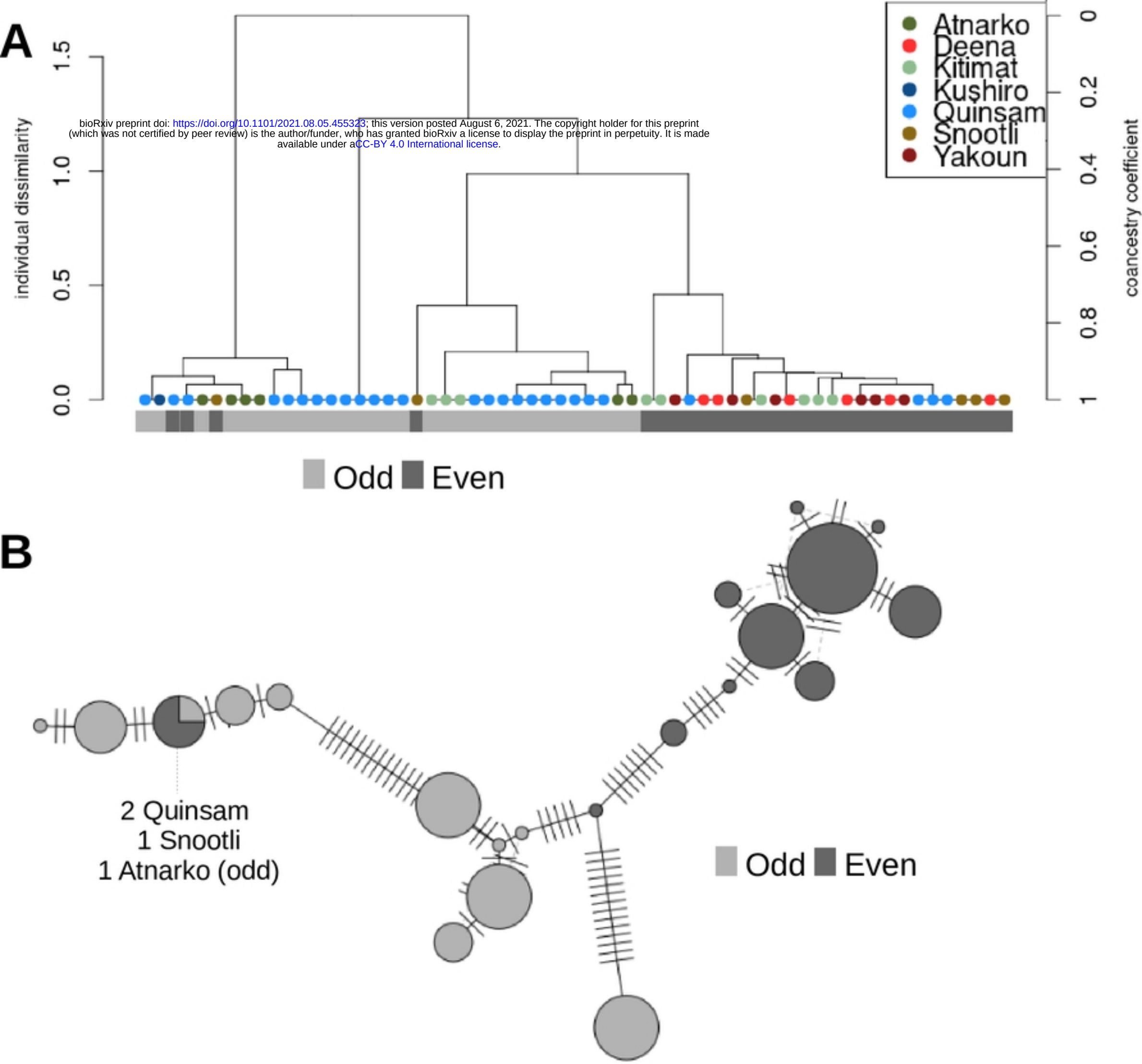

B
0.0
0.2
0.4
0.6
0.8
1.0
odd
even

Fig3

A



B

Fig4

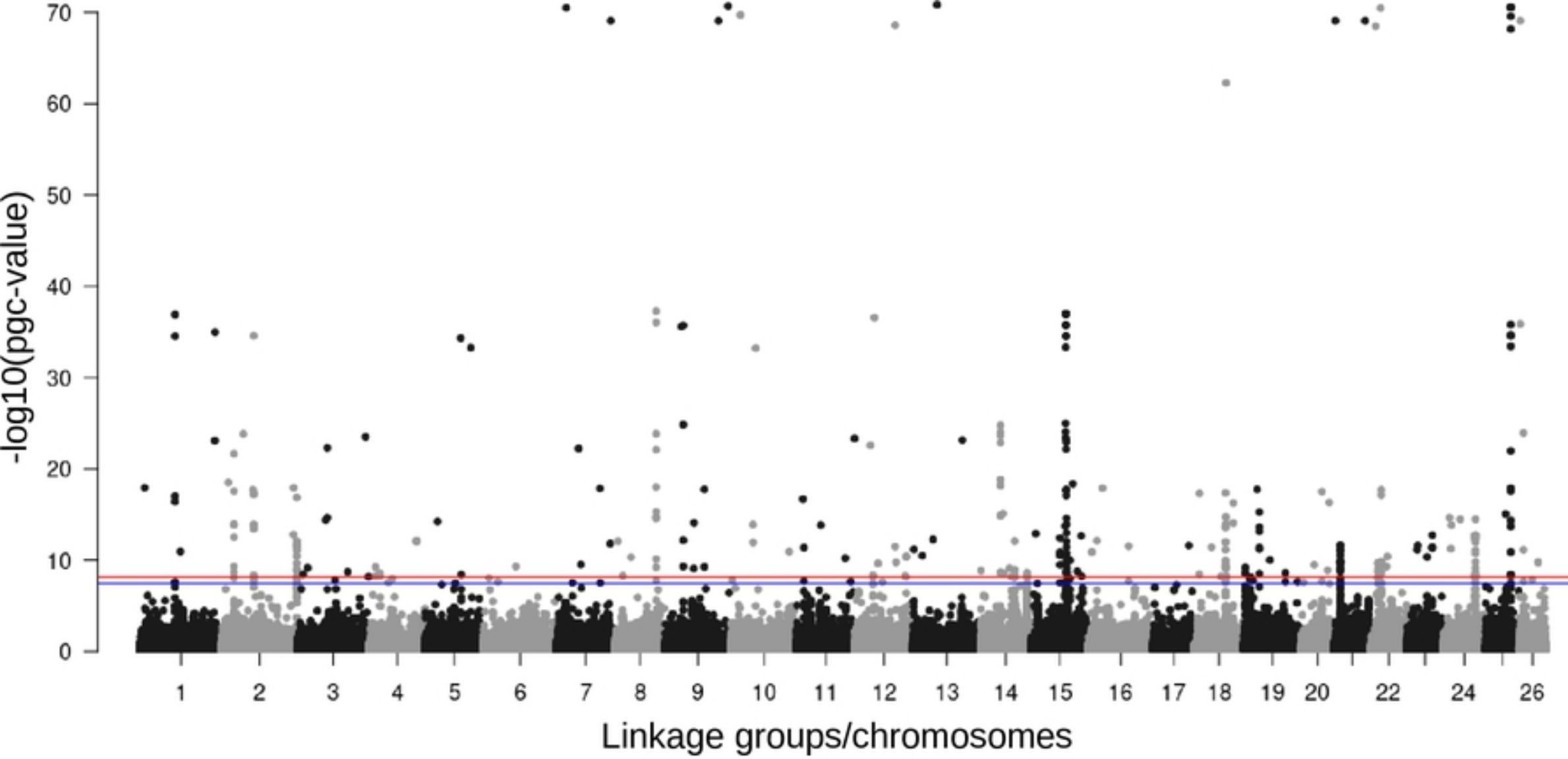

-log10(pgc-value)
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
22
24
26
Linkage groups/chromosomes

Fig5

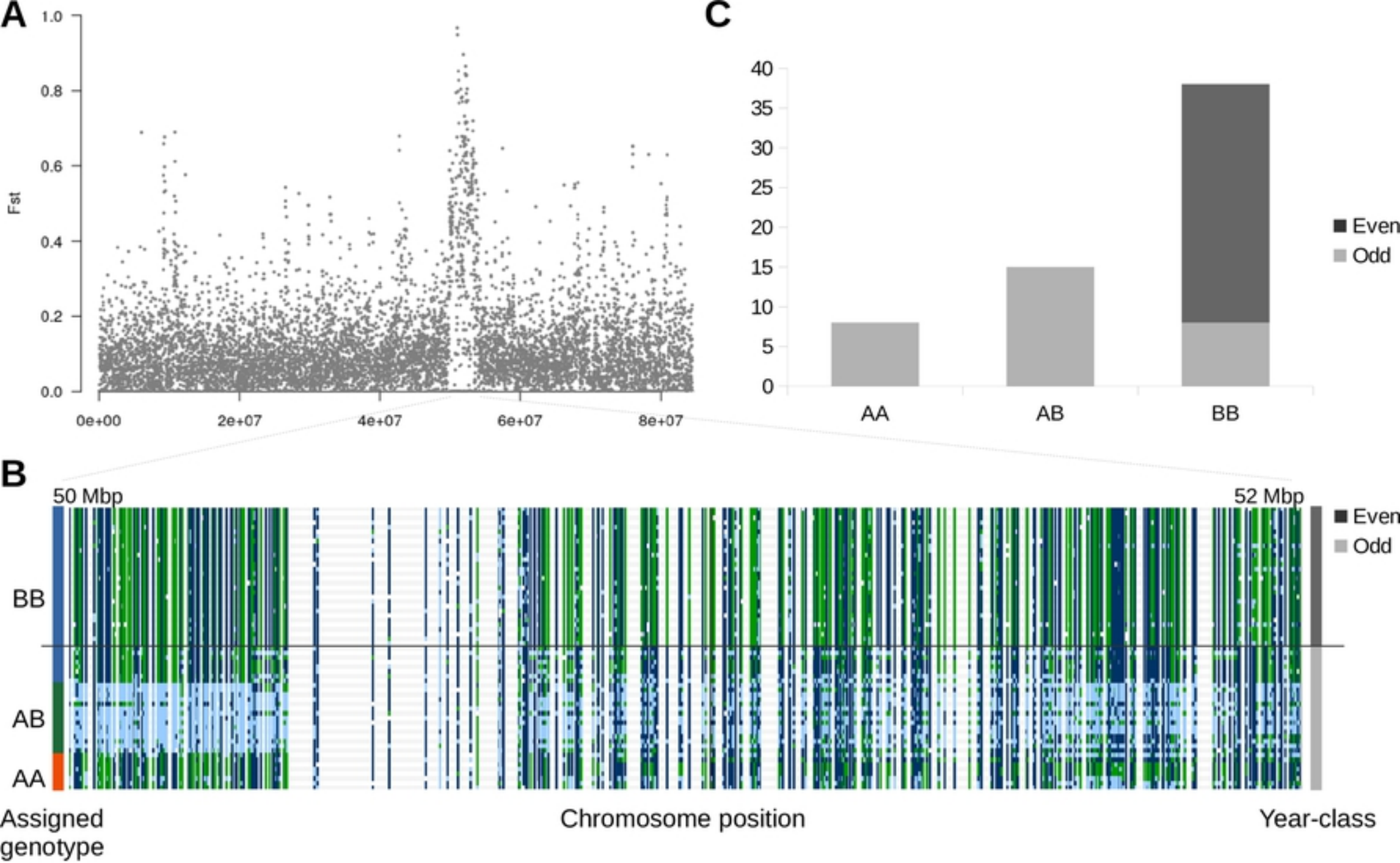
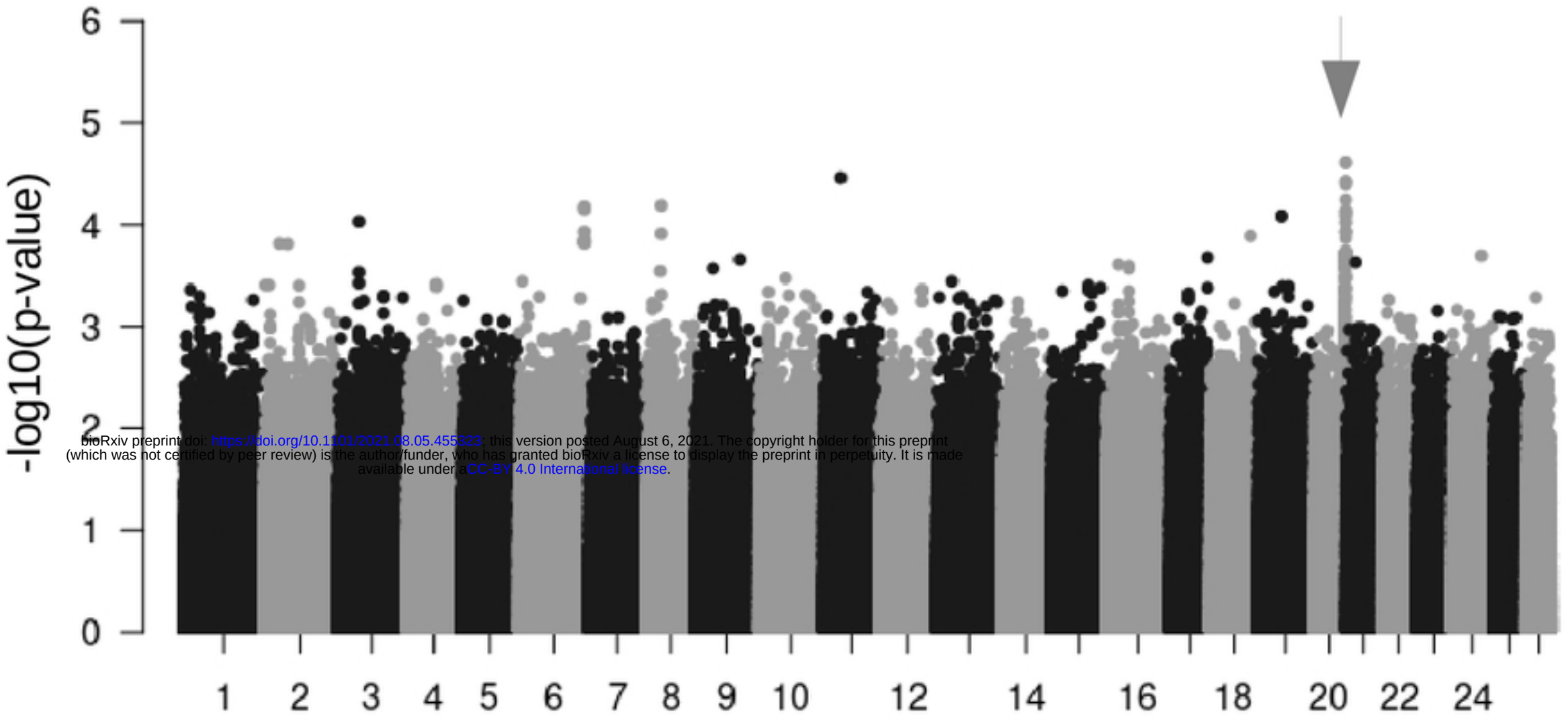

A
Fst
1.0
0.8
0.6
0.4
0.2
0.0
0e+00
2e+07
4e+07
6e+07
8e+07
C
40
35
30
25
20
15
10
5
0
AA
AB
BB
Even
Odd
B
50 Mbp
52 Mbp
BB
AB
AA
Even
Odd
Assigned genotype
Chromosome position
Year-class

Fig6

A



B

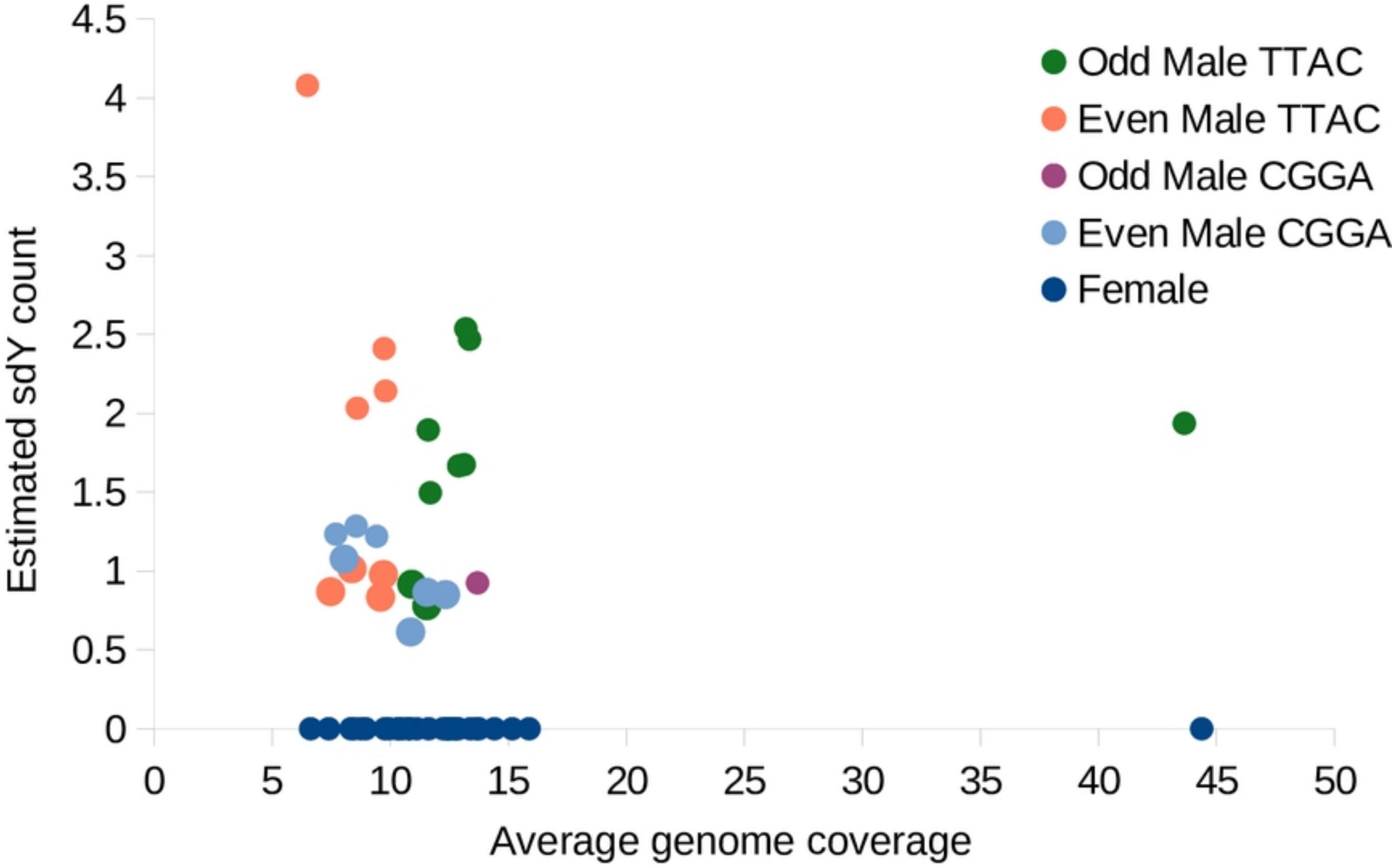


Fig7