POTENTIAL INVERSIONS AND LOCAL ADAPTATION AMONG SMALL BROOK TROUT (*Salvelinus fontinalis*) POPULATIONS

by

Caitlin Maria Nemeczek

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
April 2023

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

Dedication Page

I dedicate this thesis to my mom, Lynne Purkiss. Your love and support made me the person I am today, and I thank you for making me realise that doing my masters was the best decision I could make despite how scared I was about all aspects of it. My strong work ethic and never give up attitude comes from you and while sometimes that makes me crash, it has also allowed me to achieve all that I have thus far in my life, and I wouldn't change it for anything. You will always be an inspiration to me, and I know this thesis will make you extremely proud.

Table of Contents

List of Tables

List of Figures

Abstract

Chromosomal inversions can play a role in local adaptation as advantageous alleles become linked by supressed recombination in heterokaryotypes. Understanding the role of inversions in adaptation among small, isolated populations is an important addition for robust conservation strategies. As such, we conducted low-coverage whole genome sequencing on N=192 brook trout (*Salvelinus fontinalis*) collected from nine small streams in Nova Scotia. Individuals were sequenced at ~3x depth using paired end sequencing on the Illumina NovaSeq and genotype likelihoods calculated with ANGSD. Four potential inversions were discovered only in individuals from western streams that have lower streamflow and higher maximum daily water temperatures. Population genomics methods of LD, admixture and heterozygosity were used to support the detection of potential inversions. Some genes found within these putative inversion regions play a role in biological processes that are linked to thermotolerance and suggest evidence for potential local adaptation.

List of Abbreviations and Symbols Used

lcWGS low-coverage whole genome sequencing

lcWGR low-coverage whole genome resequencing

LD linkage disequilibrium

H_{obs} observed proportion of heterozygotes

CI chromosomal inversion

SV structural variant

CNV copy number variant

cfs cubic feet per second

cms cubic meters per second

PCA principal components analysis

SAF sample allele frequency

MAF minor allele frequency

2D-SFS two-dimensional site frequency spectrum

D_{xy} nucleotide divergence

π nucleotide diversity

F_{ST} population genetic differentiation

GLM generalized linear model

GO gene ontology

Cerk ceramide kinase

Cerkl ceramide kinase like

Acknowledgements

To my twin sister Becky, it has been extremely hard living so far away from you for almost 3 years with very limited visits due to the pandemic when we were so used to always being around each other. Thank you so much for all your support along this crazy adventure of ups and downs and I promise I will come home more often when this is done.

I remember during my undergrad in 2015 I learned about whole genome sequencing. Having worked and studied in the field of population genomics, I thought it would be amazing to do whole genome sequencing, but I never thought it would happen. I shouldn't have ever said that because here I am, having done low-coverage whole genome sequencing to find potential inversions in one of my favorite fish. I came into this degree with quite a bit of lab experience that helped me succeed from a lab perspective and as an individual to which I would like to thank Caleigh and Woz. You two taught me everything I know about population genetics, lab work, troubleshooting, and how to be an excellent lab mate which I applied during my masters. Woz, thank you especially for always understanding my very visual learning style. You have been an inspiration to me, and I hope that I can continue teaching in that way. Before starting my masters, I really wanted to learn bioinformatics for genomic analyses. I came across the bioinformatics data skills book by Vince Buffalo and as I look back, this book was instrumental for my thesis for three reasons. It taught me to always include the version numbers for packages, to keep detailed notes, and to comment scripts. It's hard to even express how important this was. Jumping into the world of bioinformatics when I was used to using genetic software that had GUIs was a massive learning curve. I am thankful to my self for heavily commenting my code so that I could remember what exactly I did when it came to methods, and I am sure anyone who reads this thesis will appreciate it. In part I was driven to do this because while I was learning I found it extremely frustrating to find scripts on GitHub that had no comments. Thank you also to Claire Mérot for helping me understand the use of percentiles in capturing LD and her wonderfully commented scripts that allowed me to do LD heatmaps and various other analyses involved in inversion discovery.

Brook trout…where do I begin. I did quite a bit of work on brook trout in my previous role as a lab technician which included providing a lot of support at a research hatchery where I spent quite a bit of time weighing, measuring, fin clipping, moving, and spawning brook trout. It was there that I learned phenotypic differences between different strains and understood their true beauty. My masters included a field component to which I was able study brook trout in small streams. Thank you, John MacMillan and Sam Andrews, for all of your help with electrofishing, sampling brook trout, installing temperature loggers and teaching me about the streams. This is an experience that I truly appreciate because I was able to make a connection between the genomic side of things and ecological side. I was so used to only looking at things from a genetic perspective, but having the opportunity to look at environmental impacts on brook trout was a truly unique experience that I will never forget. The brook trout genome is massive and while it made me cry many times due to large data problems, it allowed me to discover something that hasn't been done yet which could be used in conservation of the species.

I knew that completing my masters would be challenging, but the learning curve of bioinformatics and programing is indescribable for someone who had essentially zero experience prior. Thank you to the support staff of ACENET for help troubleshooting programs and script issues on the Digital Research Alliance of Canada clusters. I want to thank everyone in the Ruzzante lab for their support and help through this process, but most importantly Lisette. Lisette thank you so much for all that you taught me in Linux, bash scripting, Compute Canada, and all programs relevant to low-coverage whole genome sequencing. It is particularly challenging to try and troubleshoot errors in bash let alone when you are locked out of the university due to a pandemic. Thanks, Lisette, for your patience and support in helping me either in the office or online when we had to, for being there to vent about how frustrating of a program ANGSD is to deal with, and for being a great friend. Thank you, Mallory, for all your help in R and supporting me while I often got mad at my computer behind you. You taught me so much about manipulating and plotting various kinds of data which allowed me to make the pretty figures seen in this thesis. To James and Ellie, thank you for your emotional support and help with R things, scripting, writing and various others, but most importantly for being awesome friends. Chris, you may not have helped directly with my thesis, but thank you for being

an awesome friend and fishing buddy. The escape from my data analyses to learn more about fishing was very much appreciated.

To my supervisor Daniel, I am so thankful to have had you as a supervisor. You are the true definition of supportive. Throughout this journey I always felt like you believed in what I was doing and that I would figure out whatever difficult analyses or errors I got. Your constant positive attitude and feedback was very much appreciated. I will forever remember the excitement on your face every time I came into your office with a new plot to show you. Thank you for letting me take the thesis in the direction that we went, despite it being quite different than what it was supposed to be. This itself is a learning experience I am very grateful for. Thank you to my committee members, Paul, and Ian for the suggestion to do lcWGS and for fruitful discussions along the way.

Moving to a city far from my hometown where I lived for my entire life was very difficult but every part of the experience of living in Halifax has been amazing and I will never forget it. I drove here in the height of the pandemic with whatever fit in the rental SUV of my belongings and quickly began meeting the most amazing people including my partner, Megan. We have been through quite a bit over the past two years and having you by my side has been so special. Your constant positive attitude and outlook on life as well as your love and support has allowed me to get through the toughest times. I can't imagine having done this without you. To Cat, Joc, Noor, Scott, Tay Tay, and GG thank you all for being the greatest humans and for being so welcoming to me! The times we spent together during my degree were all so fun and I will never forget it. To Dash, while we have drifted apart since I moved here, I thank you for always being there to listen to my issues, for making me move beyond my comfort zone in different aspects of life, and for suggesting that doing a masters would be a great idea. I appreciate all the help and advice you offered from your own life experiences in completing this chapter of my life.

Chapter 1: Introduction

1.1 Local Adaptation And Speciation

When populations are fragmented and small, genetic drift is expected to prevail over natural selection and local adaptation. The relatively recent unraveling of various types of structural variants (SV) in genomes among a wide variety of taxonomic groups (Cohen et al., 2023; Liu et al., 2022; Lundberg et al., 2023; Zhang et al., 2023) however, has led to a better understanding of the roles of natural selection and genetic drift in generating local adaptation. Chromosomal inversions (CIs) are a type of genomic SV which occur when part of the chromosome breaks and the orientation between the breakpoints becomes reversed (Kirkpatrick, 2010).

Inversions can lead to local adaptation if the inverted chromosomal regions contain genes and alleles that play a role in biological processes such that individuals carrying an inversion exhibit a fitness advantage over individuals not carrying it (Blanquart et al., 2013; Kirkpatrick, 2010). The role of inversions in local adaptation and speciation comes primarily from the suppressed recombination that takes place in heterokaryotypes which have one copy of the inverted and one copy of the non-inverted sequence arrangement. Speciation can occur if inversions harbor duplications, deletions, and insertions that make heterokaryotypes inviable, thus creating a barrier between individuals with and without inversions and is known as underdominance (Kirkpatrick, 2010). Speciation via inversions with underdominance is likely less common than speciation via recombination because they will only become established if they occur when populations are small such that genetic drift prevails, and when local adaptation with gene flow allows the inversion to be maintained (Hoffmann & Rieseberg, 2008). Speciation can occur via recombination if there are alleles within an inversion that cause reproductive isolation even when there is gene flow. Dobzhansky-Müller incompatibilities from negative epistatic interactions can prevail and have a greater impact on fitness than for example, locally adapted alleles (Faria, Johannesson, et al., 2019; Hoffmann & Rieseberg, 2008). Alternatively, alleles leading to reproductive isolation could be locally adapted and will become established in populations if they have a positive impact on fitness (Hoffmann & Rieseberg, 2008). Speciation by recombination can also occur if divergent selection plays a major role or if

there is assortative mating. When an inversion develops, it evolves by natural selection and genetic drift where the roles of drift, mutations, and recombination change through time thus shaping the persistence or removal of inversions. Current research is allowing us to better understand how inversions evolve by examining the roles of various types of balancing and divergent selection (Faria, Johannesson, et al., 2019). Inversions can reach fixation by selection or drift in some populations, and they can be maintained by balancing selection. Balancing selection can result from overdominance where the heterokaryotypes are selected over homokaryotypes or associative overdominance in which recessive alleles accumulate causing lower fitness in homokaryotypes.

1.2 Chromosomal Inversions And Other Structural Variants

The recent explosion of CI research has suggested that inversions are often large with an average size of 8.4 megabases (Wellenreuther & Bernatchez, 2018) and frequent as shown in deer mice where 13 large inversions have been described (Harringmeyer & Hoekstra, 2022). CIs can be paracentric when the break occurs along a chromosome arm, or pericentric when they straddle the centromeres, however, these are less common due to the generation of inviable gametes (Hoffmann & Rieseberg, 2008; Wellenreuther & Bernatchez, 2018). The role of inversions in local adaptation has been documented in many species such as the monkey flower where inversions are responsible for annual and perennial ecotypes (Twyford & Friedman, 2015), and common quails where a large inversion controls phenotypic differences in size, throat colouration, wings, migration, and flight efficiency (Sanchez-Donoso et al., 2022). Chromosomal inversions are also implicated in the differences in migration behavior and more broadly in adaptation to environmental differences in willow warbler (Lundberg et al., 2023), and Atlantic cod (Matschiner et al., 2022). Other types of SVs can also be linked to local adaptation and speciation. Transposable elements, gene duplications, chromosomal fusions, translocations, insertions, and deletions are all SVs that can change the nucleotide sequence, sometimes having an adaptive advantage and can therefore contribute to speciation (Mérot et al., 2020). For example, recently a chromosomal translocation was discovered among Newfoundland populations of Atlantic salmon that is involved in population structuring and its frequency is influenced by temperature suggesting evidence of local adaptation (Watson et al., 2022).

1.3 Whole Genome Sequencing

While CIs are the most documented of SVs with adaptive advantages and have been studied for many years (e.g., Sturtevant, 1921), they have recently gained increased attention following the decreased cost in whole-genome sequencing. Low-coverage whole genome sequencing (lcWGS) has facilitated a more robust understanding of the importance of CIs in local adaptation (Faria, Johannesson, et al., 2019). lcWGS is a recently developed sequencing technique that is becoming more predominant in population genomic studies as the cost is similar to other highly used sequencing methods such as RAD-seq, though this is dependent on sample and genome size (Fuentes-Pardo & Ruzzante, 2017; Lou et al., 2021). While there are different forms of whole genome sequencing, in lcWGS each sample is individually barcoded and sequencing coverage per individual is between 1-5x. It thus requires the use of genotype likelihoods rather than direct genotype calls to account for uncertainty in a probabilistic framework (Lou et al., 2021). Inversions have been discovered via other sequencing techniques such as long-read sequencing, but these methods are more expensive and are more labour intensive in terms of bioinformatics (Harringmeyer & Hoekstra, 2022; Stenløkk et al., 2022)

1.4 Brook Trout

Brook trout are economically important and one of the top sportfish in Nova Scotia. Populations in some lakes and rivers of Nova Scotia have been maintained by stocking programs since the early 1900s (Lehnert et al., 2020), but brook trout remain vulnerable to habitat loss, overexploitation, invasive species that compete for food and habitat as well as climate change causing low streamflow which is particularly challenging given their thermal tolerance (Cherry et al., 1977; MacMillan et al., 2008; Nova Scotia Department of Agriculture and Fisheries Inland Fisheries Division, 2005). Some salmonids have been well studied from a local adaptation perspective as seen in a recent study on Coho salmon that covered their entire North American range and found that migration distance and elevation are important to local adaptation (Rougemont et al., 2023). Local adaptation has also been documented in redband trout from the Columbia River where there is evidence that temperature variation has a role in local adaptation and that genes effecting migration and developmental timing are found in adaptive regions (K. R. Andrews et al., 2023; Z. Chen & Narum, 2021). However, studies on local

adaptation in brook trout are very scarce. Brookes et al. (2022) examined introduced brook populations in western Canada and did not find evidence of local adaptation. Inversions have been reported in salmonids and in rainbow trout a known inversion contains genes important for distinguishing migratory ecotypes (Arostegui et al., 2019; Pearse et al., 2019). An inversion polymorphism has also recently been discovered in Arctic charr with genes that have a role in cell adhesion and signal transduction (Hale et al., 2021) further adding to our understanding of the role of inversions in local adaptation. Despite current ongoing research on inversions and local adaptation, no studies to my knowledge have captured the role of inversions among brook trout and their potential role in local adaptation. Such knowledge could improve our understanding of brook trout vulnerability to population decline and help build a basis for new conservation efforts. Brook trout in the current study originate from populations that have not been stocked (John MacMillan, personal communications, July 2022) thus providing a unique scenario to examine the role of inversions in pristine populations exhibiting limited to no gene flow (Ruzzante et al., 2016) which are likely shaped by colonization history. Eastern Canada exhibits evidence of having been colonized by brook trout originating from three glacial refugia: Mississippian, Acadian and Atlantic refugia (Black et al., 1986; Danzmann et al., 1998). Glaciation and postglacial colonization history is known to play a role in shaping population distribution and structure (Bernatchez & Wilson, 1998) and could potentially help explain population differences in inversion patterns.

1.5 Objectives

In this thesis I investigate the presence of potential chromosomal inversions among brook trout populations and assess their contribution to local adaptation. I hypothesize that there are potential chromosomal inversions, and that they contain genes which are locally adapted to environmental variables such as water temperature. Brook trout were sampled from nine streams in the North Mountain, Nova Scotia that show differences in surficial geology, streamflow, and temperature between eastern and western streams. lcWGS was performed on individuals with a target depth of 4x and genotype likelihoods analyzed using methods of LD, population structure, heterozygosity, nucleotide diversity and divergence. We report on four potential chromosomal inversions discovered only in a

subset of brook trout populations examined and explore the relevance of genes in these regions to local adaptation.

Chapter 2: Methods

2.1 Field Work

2.1.1 Brook Trout Sampling

Brook trout (N=360) were sampled non-lethally by backpack electrofishing from nine streams along the North Mountain, Annapolis Valley, Nova Scotia in June 2021. In each stream, fish (N=20) were collected from each of 2 locations for a total of N=40 individuals per stream. Sampling locations were separated by ~ 600 m from each other (Table 1 and Fig. 1). These streams differed in environmental and physical characteristics such as surficial geology, area of drainage basin, stream gradient (Table 1), pH, water temperature, and streamflow (Table A1, A2). Fish were measured (fork length) and fin clipped. Fin clips were stored in ETOH for subsequent DNA analysis.

Table 1 Physical characteristics of the nine streams sampled for brook trout in the North Mountain, Annapolis Valley, Nova Scotia and average and median fork lengths of 40 brook trout sampled from each stream.

Stream	Surficial Geology	Area of drainage basin (km^2)	Stream gradient % (slope)	Average fork length (cm)	Median fork length (cm)
Ross Creek	Scoured layer of glacial till	6.9	3.42	10.33	9.5
Woodworth	Scoured layer of glacial till	10.78	2.93	11.01	10.8
Black Hole	Scoured layer of glacial till	6.42	3.66	11.77	10.6
Church Vault	Thick stony-granite derived till	10.7	2.70	11.32	11
Saunders	Thick stony-granite derived till	8.50	3.33	11.63	11cm
Robinson	Thick stony-granite derived till	12.3	3.08	10.80	10.45
Sheep Shearer	Silt	7.90	3.77	15.11	14.4
Healeys	Silt	4.9	4.5	11.48	10.95
Poole	Silt	6.89	6.03	12.87	12.4

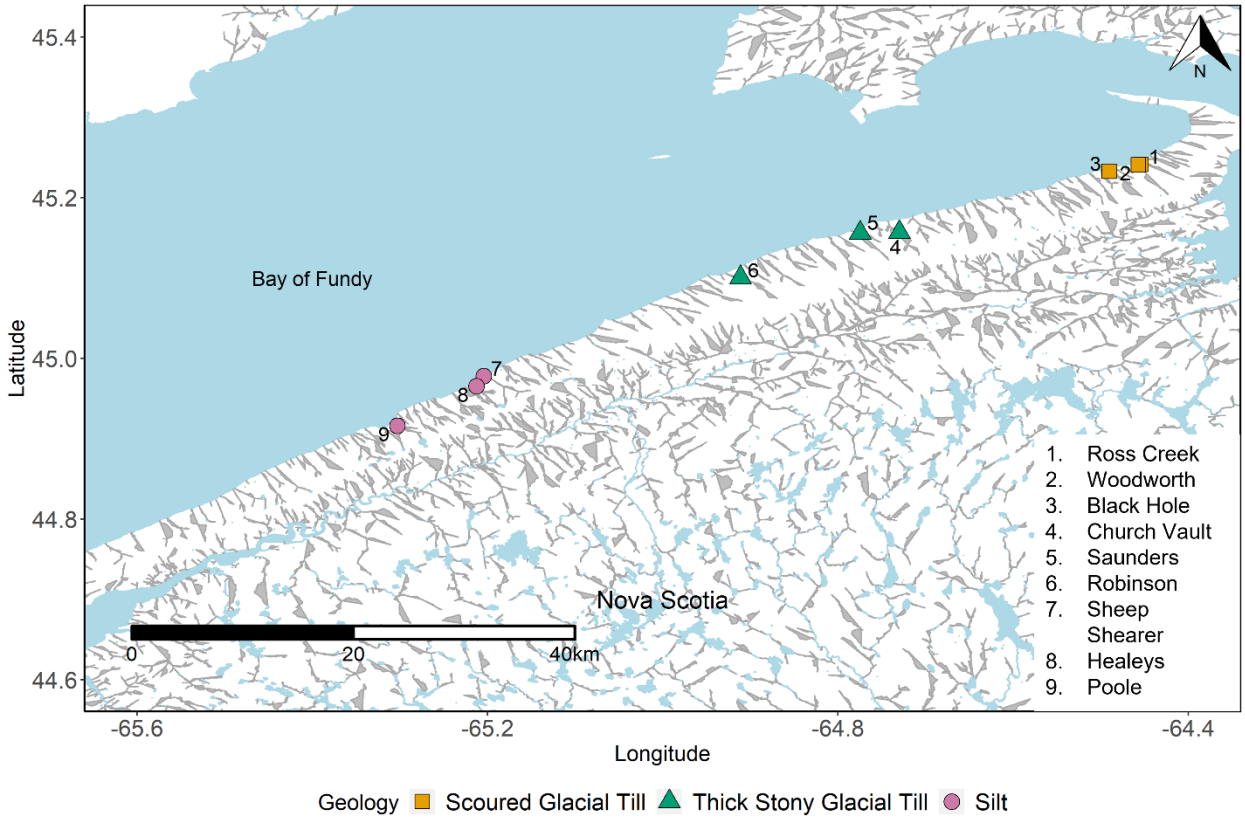

Figure 1 Nine streams in the North Mountain, Nova Scotia sampled for brook trout. Shapes indicate the type of surficial geology that make up the streams.

pH was measured with the pHep^{+} H198108 (Hanna instruments) in the middle of the stream and at both banks. Electrofishing time, water temperature, and air temperature were recorded at each stream.

2.1.2 Temperature Loggers

HOBO pendant temperature loggers were installed in six streams on April 29, 2021. A logger was placed at Church Vault Brook on a tree to record air temperature. The logger at Healeys Brook was moved downstream on June 22, 2021, because where it was initially installed had dried up. A logger was installed at Poole Brook on this day.

2.1.3 Streamflow

Streamflow was measured using the float method as a velocity-area measurement in all nine streams once a month from July to November 2021. An A-Just-A-Bubble fishing float (Rainbow plastics©) filled half-full with water was used as the float. Streamflow

was estimated in straight stream sections with minimal rock obstructions. A reel-measuring tape was used to measure the width of the stream in both an upstream and a downstream site. The length of the stream section was set at approximately 3x the width. In most cases the length was at least 2x the width. We also measured depth at every 30 cm interval in both the upstream and downstream transects. A float was dropped approximately 60 cm upstream of the upstream transect marker such that it would gain speed before entering the area of streamflow measurement. The time for the float to reach the downstream marker was taken with a stopwatch and five measurements were taken per stream. Surface velocity in ft/sec for each of 5 floats from every stream was calculated. The depths of each of the upstream and downstream sections were averaged and the area calculated by multiplying the width of the stream section by the average depth. An average area was calculated from the area upstream and downstream for every stream. Surface velocity was calculated and multiplied by 0.66 to account for depth as the average channel velocity (United States Department of the Interior Bureau of Reclamation, 1997). Streamflow or discharge in cubic feet per second (cfs) was calculated by multiplying the average velocity by the average area and converted to cubic meters per second (cms). pH and surface water temperature were recorded at each streamflow measurement site in the middle of the measured section and at both banks.

2.2 DNA Extractions And Quantification

DNA was extracted from N=22 individuals per stream chosen to represent a single age class based on fork length. For DNA extraction I used the Omega E.Z.N.A tissue DNA kit (Omega Bio-Tek) following the manufacturer's protocol. DNA was quantified (in ng/μL) using the Quanti-iT Picogreen dsDNA assay (Invitrogen) with a microplate reader and each sample was run in triplicate to get an average quantity (in ng/μL).

2.3 Library Preparation And Low-Coverage Whole-Genome (lcWGS) Sequencing

Six libraries were prepared for lcWGS following the Illumina DNA Prep protocol with modifications of the Illumina Nextera-Flex protocol, Hackflex (Gaio et al., 2022), multiplexed library prep from Baym et al. (2015), and the protocol from Therkildsen and Palumbi (2017). For a description of library preparation steps see Figure 2. The pooled

library of 192 individuals was sequenced on 4 lanes of paired-end 2x 150bp reads on a NovaSeq S4 flowcell.

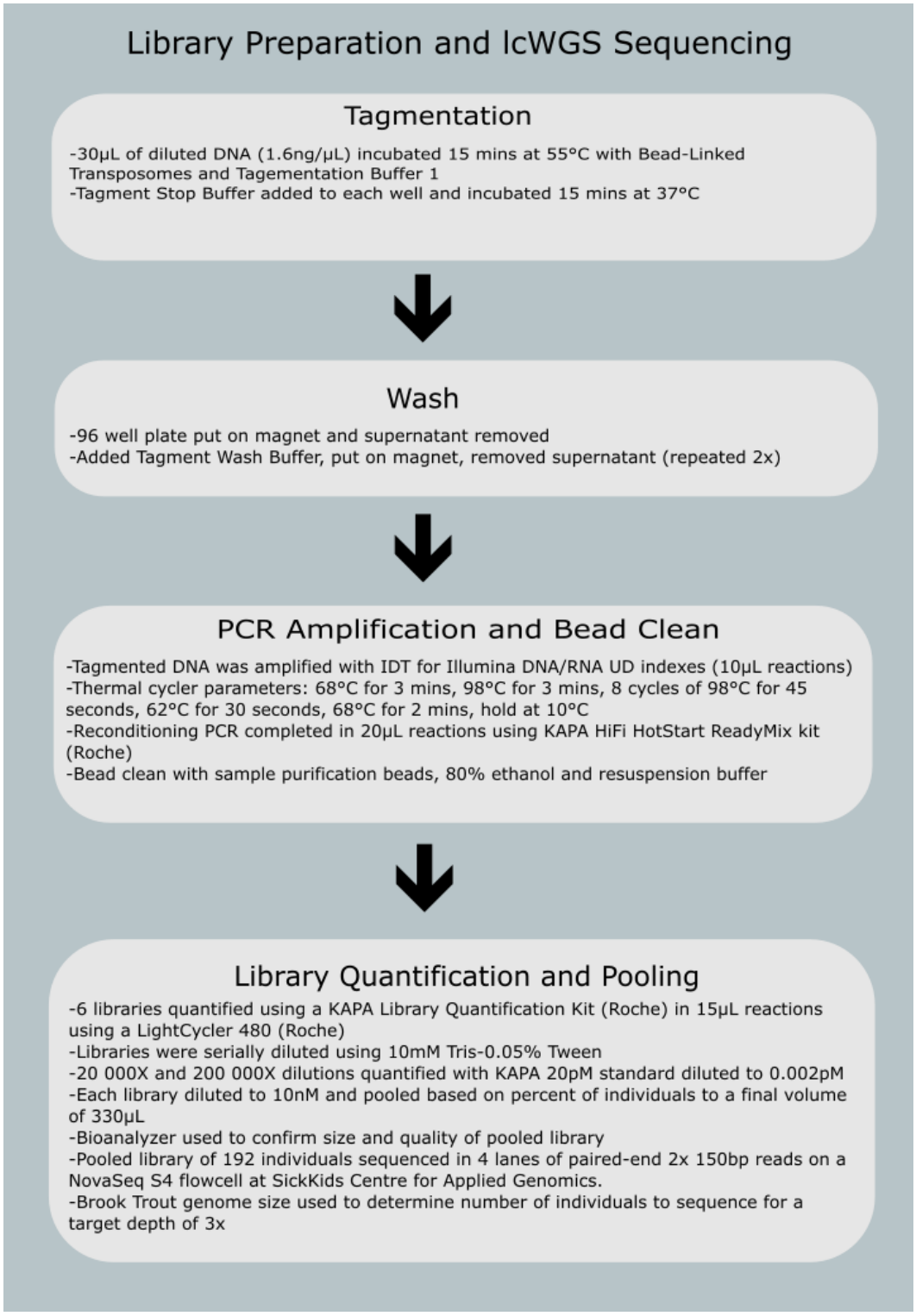

Figure 2 Flowchart representing a summary of the lab procedures used for lcWGS library preparation. The Illumina DNA prep protocol was used with modifications following Illumina Nextera-Flex protocol, Hackflex (Gaio et al., 2022), and the protocol from Therkildsen and Palumbi (Therkildsen & Palumbi, 2017).

2.4 Data Filtering

Raw fastq files were assessed using FastQC v.0.11.9 (S. Andrews, 2010) and MultiQC v.1.11 (Ewels et al., 2016) and adapters removed with Trimmomatic v.0.39 (Bolger et al., 2014) in paired end mode. Fastp v.0.23.1 (S. Chen et al., 2018) was used to remove poly-G tails. FastQC v.0.11.9 (S. Andrews, 2010) and MultiQC v.1.11 (Ewels et al., 2016) were run using the trimmed fastq files to confirm that adaptors and poly-G tails were removed.

2.5 Reference Genome Mapping And Sequence Filtering

Filtered reads were mapped to the lake trout (*Salvelinus namaycush*) genome SaNama_1.0 (GenBank assembly accession: GCF_016432855.1; (Smith et al., 2022) with Bowtie2 v.2.4.4 (Langmead & Salzberg, 2012). The indices of the reference genome and dictionary were generated using SAMtools v.1.13 (Li et al., 2009) and Picard tools v.2.26.3 ("Picard Toolkit," 2019). Aligned bam files were filtered with SAMtools v.1.13 (Li et al., 2009) to remove discordantly mapped pairs and those with mapping quality score below 20. The pooled library was run on each of the 4 lanes of the flowcell resulting in a forward and reverse read for every individual on each lane following recommendations regarding batch effects (Lou & Therkildsen, 2022). SAMtools v.1.13 (Li et al., 2009) was used to merge the mapped bam files to produce one file per individual. Picard tools v.2.26.3 ("Picard Toolkit," 2019) MarkDuplicates command was used to identify and remove PCR and optical duplicate reads and overlapping paired reads were removed using the Bamutils v.1.0.14 (Jun et al., 2015) clipOverlap command. Filtered reads were aligned around indels with GATKs v.3.7 IndelRealigner (McKenna et al., 2010) to account for mapping errors and to increase consistency of read alignments in regions with indels. Following indel realignment, individual read depth was calculated by stream (population) using the SAMtools v.1.13 depth command (Li et al., 2009). Average depth per individual and stream (population) were calculated from the depth files produced by SAMtools v.1.13 depth command (Li et al., 2009). Average depth per stream across 21-22 individuals ranged between 1.8x and 2.6x.

2.6 SNP Calling And Genotype Likelihoods

ANGSD v.0.933 (Korneliussen et al., 2014) was used to estimate minimum and maximum depth thresholds to improve accuracy of genotype likelihoods. After depth filtering, the .mafs.gz files showing the major, minor, and reference alleles, and frequency of minor allele were used to generate a SNP list. The ANGSD v.0.933 (Korneliussen et al., 2014) sites index function was used to generate index files. Genotype likelihoods and minor allele frequencies of 5% were estimated using ANGSD and SAMtools identified 1,745,248 SNPs while GATK identified 1,798,343 SNPs. SAMtools was chosen for further ANGSD analyses because it is more conserved than GATK, which is the more appropriate option given that a reference genome of a different species was used. The covariance matrices produced by ANGSD were used to perform a principal components analysis (PCA) in R v.4.1.2 (R Studio Team, 2021) with ggplot2 v.3.3.5 (Wickham, 2016), dplyr v.1.0.7 (Wickham et al. 2021), and tidyverse v.1.3.1 (Wickham et al. 2019). Eigen vectors and values were calculated from the covariance matrices and used to plot principal components explaining the highest percent of variation in the SNP data. ANGSD was used to calculate genotype likelihoods. The .mafs.gz and beagle.gz files produced from running ANGSD with a supplied SNP list were used for the estimation of linkage disequilibrium (LD). PCAngsd v.1.10 (Meisner & Albrechtsen, 2018) was used to generate a covariance matrix of individual allele frequencies using the LD pruned genotype likelihoods .beagle file produced from ANGSD, filtering out data with a minor allele frequency below 5%. The covariance matrix was used to perform a PCA in R v.4.1.2 (RStudio Team, 2021) as described above.

2.7 Linkage Disequilibrium And Decay

The minor allele frequency file (.mafs.gz) and genotype likelihoods file (beagle.gz) produced from ANGSD with the supplied SNP list were used to calculate LD using ngsLD v.1.1.1 (Fox et al., 2019). LD was tested over three different window sizes (100, 200, 500kb) with –max_kb_dist. The rate of decay in LD began to slow down at approximately 20kb after which SNPs are unlikely to be linked and this was used to calculate LD with parameters discussed in the appendix (Figure A1). SNPs in LD were

pruned using ngsLD's prune_graph.pl resulting in 302,953 SNPs. For pruning, the minimum weight for assuming SNPs were connected based on r^2 values (--min_weight) was set to 0.4 based on the LD decay plot and was done by chromosome.

2.8 LD By Chromosome

To look for evidence of potential inversions the minor allele frequency and genotype likelihood files mentioned above from all 192 Brook Trout were separated into individual files based on chromosome. ngsLD v.1.1.1 (Fox et al., 2019) was used to calculate LD for each chromosome using –max_kb_dist and –max_snp_dist of 0 while randomly sampling 50% of SNPs. The r^2 values from the LD data was divided into percentiles and plotted as heatmaps in R v.4.0.2 (RStudio Team, 2021) using ggplot2 v.3.3.6 (Wickham, 2016) following scripts from Mérot et al. (2021). Heatmaps of each chromosome were used to find areas of high LD (r^2 above ~0.6) and genotype likelihoods were calculated using ANGSD on SNPs only in those regions. PCAs for chromosomes with high LD blocks were produced using ggplot2 v.3.3.6 (Wickham, 2016) in R v.4.0.2 (RStudio Team, 2021). A potential inversion should show three groups representing two homokaryotypes (i.e., one with and one without the inversion) and heterokaryotypes with one copy of the inversion (Huang & Rieseberg, 2020).

A subset of each of chromosomes 9, 14, 19, 24 (NC_052315.1, NC_052320.1, NC_052325.1, NC_052330.1) was chosen for further analyses based on size of blocks of LD. LD was calculated for all nine populations pooled, for each population separately, and for sets of populations pooled by geology (scoured glacial till, thick glacial till, silt) using ngsLD v.1.1.1 (Fox et al. 2019). Heatmaps of LD were produced as described above (see appendix for more detail) to determine if blocks of LD were in all nine populations or a subset of populations, and to assess the effect of sample size on LD.

Subsets of five individuals (unless there were not 5 in a group) from each of the three groups in PCAs of chromosomes were used to estimate F_{ST}. The sample allele frequency (SAF) was calculated using ANGSD v.0.933 (Korneliussen et al., 2014) with the -doSaf option. Two-dimensional site frequency spectrum (2D-SFS) was calculated between SAF groups with realSFS in ANGSD with the -fold 1 parameter. F_{ST} was estimated using realSFS in ANGSD in sliding windows of 10kb. F_{ST} data was used to produce Manhattan

plots in R v.4.1.2 (RStudio Team, 2021) with ggplot2 v.3.3.6 (Wickham, 2016), and dplyr v.1.0.9 (Wickham et al. 2022) to see if there were high F_{ST} SNPs on those chromosomes.

2.9 Heterozygosity

Heterozygosity is used as a method for identifying potential inversions as heterokaryotype individuals represented in the middle of PCAs will have one copy of the inverted arrangement and have higher heterozygosity than homokaryotypes with and without potential inversions. ANGSD was used to calculate allele frequencies and Hardy-Weinberg equilibrium with -doHWE on individuals representing the 3 potential inversion groups for each of chromosome 9, 14, 19, and 24 using SNP lists for entire chromosomes. Observed proportion of heterozygotes (H_{obs}) was calculated and average H_{obs} in 10kb windows with a 10kb slide was obtained using the R package WindowScanR (Tavares, 2022) across entire chromosomes. Average H_{obs} was also calculated within potential inversion regions using a window size that was equal to the potential inversion region and boxplots generated using R v.4.0.2 (RStudio Team, 2021) and ggplot2 v.3.3.6 (Wickham, 2016). H_{obs} represents the per base position across individuals in each of the inverted homokaryotype, heterokaryotype, or non-inverted homokaryotype groups. For each chromosome and group (inverted homokaryotypes, heterokaryotypes, non-inverted homokaryotypes) I estimated the Bonferroni corrected ($p<0.001$) number of SNPs within the corresponding region that deviated from Hardy-Weinberg.

2.10 Nucleotide Diversity And Divergence

Nucleotide diversity (π) and divergence (d_{xy}) can be used to look at age of inversions where younger inversions can have lower diversity and divergence because there has not been as much time for genetic variation to accumulate compared to the ancestral arrangement. Selection is also known to impact patterns of diversity. The same individuals used for heterozygosity representing the 3 potential inversion groups were used for diversity and divergence calculations. ANGSD v.0.933 was used to get global counts of potential inversion groups to select appropriate maximum and minimum depth filters for SAF and run again with -doSaf. SAFs were used to get the folded site

frequency spectrum using realSFS in ANGSD v.0.933 and saf2theta used to output theta values. A 10kb sliding window analysis with a slide of 10kb was done on the thetas using do_stat from thetaStat in ANGSD v.0.933. π was plotted inside and outside potential inversion regions for homokaryotypes with inversions and homokaryotypes without inversions. ANGSD v.0.933 was run using -doMaf and SNP lists for regions inside potential inversions and the minor allele frequencies were used to calculate d_{xy} using modifications to https://github.com/mfumagalli/ngsPopGen/blob/master/scripts/calcDxy.R. The R package WindowScanR (Tavares, 2022) was used to produce average d_{xy} in windows of 10kb with a 10kb slide. D_{xy} was plotted for inverted and non-inverted individuals inside potential inversion regions.

2.11 Population Genomics

Pairwise weighted F_{ST} were estimated between each pair of brook trout populations using ANGSD v.0.936 (Korneliussen et al., 2014) and between populations in the three different geological groupings (scoured glacial till, thick glacial till, silt). The SAF was first estimated for each population (-doSaf where SAF is the population folded minor allele frequency). It is folded because the reference genome is used as the ancestral state rather than an outgroup. Genotype likelihoods and posterior probabilities are used to estimate the SAF for each population before the site frequency spectrum (SFS) is generated. The SAF files per population were used to obtain the folded two-dimensional site frequency spectrum (2D-SFS) between all 36 population combinations using the realSFS program in ANGSD. The -fold 1 parameter was used with the unfolded .saf files to generate the folded 2D-SFS. F_{ST} was estimated between all 36 population combinations and between the three geological groupings (scoured glacial till, thick glacial till, silt). ANGSDs realSFS was used to first estimate the per site F_{ST} index with unfolded SAF and folded SFS using the fst index command with -whichFst 1. F_{ST} was then estimated using realSFS fst stats2 command with a window size of 10kb and a step size of 10kb. The resulting text files of windowed F_{ST} estimates were used to generate Manhattan plots in R v.4.1.2 (RStudio Team, 2021), with ggplot2 v.3.3.6 (Wickham, 2016), and dplyr v.1.0.9 (Wickham et al. 2022). The same method was used to observe

potential inversions on chromosome 9, 14, 19, and 24. For each chromosome, 5 individuals (except for some inversion groups where there are 4 individuals) representing each of homokaryotypes with inversions, homokaryotypes without inversions and heterokaryotypes were analyzed using the above steps and Manhattan plots generated across the genome and on each of the 4 chromosomes.

2.12 Admixture

NGSadmix (Skotte et al., 2013) is part of ANGSD (v.0.933) and was used to infer the likely number of populations through calculation of likelihood ratios from admixture proportions using the LD pruned genotype likelihoods (.beagle.gz). NGSadmix was run setting -K from 1-10 and repeating each K run 5 times with a filter for minor allele frequency of 5%. CLUMPAK (Kopelman et al., 2015) was used to calculate delta K and the corresponding admixture proportion file (.qopt) was plotted in R v.4.1.2 (RStudio Team, 2021).

2.13 Geology Vs Distance As Predictors Of F_{ST}

The type of glacial till, referred to as surficial geology throughout is an environmental characteristic that differs between groups of populations. Scoured glacial till is exposed bedrock with thin layers of leftover glacial material, thick glacial till is a mixture of stone and sandy material from glaciers left on top of bedrock, and silt streams consist of a compact clay and sand like material (Richard & Bradley, 1992). Geology could be related to the distance between streams and a generalized linear model (GLM) was used to find a significant predictor of linearized F_{ST}. Pairwise weighted F_{ST} were estimated between all 36 pairs of population combinations using ANGSD v.0.936 (Korneliussen et al., 2014) as described in the population genomics section. R studio v.4.1.2 was used to calculate mean F_{ST} per population and linearized following the method in Rousset (1997). The stats package 4.1.2 (R Core Team, 2020) in R studio v.4.1.2 (RStudio Team, 2021) was used to fit a GLM using a gaussian function with distance in km and whether populations were within a geology group or between geology groups as predictors of linearized F_{ST}. Predicted linearized F_{ST} was plotted against actual linearized F_{ST} to visualize the model prediction using ggplot2 v.3.3.6 in R studio v.4.1.2 (RStudio Team, 2021).

2.14 Gene Annotations

Regions of potential inversions on chromosomes 9, 14, 19, and 24 were used to determine genes and gene annotations in those regions to search for a potential connection to local adaptation. Size of potential inversions were adjusted for some chromosomes after plotting observed heterozygosity as it was easier to determine start and stop positions from those plots. The UCSC genome browser (Kent et al., 2002) was used to search potential inverted regions on chromosomes 9, 14, 19, and 24 of the lake trout genome and the resulting gene table was sorted in R v.4.0.2 (RStudio Team, 2021) to search for unique gene symbols. The lake trout genome is not fully annotated and as such resulting loci not corresponding to any genes were removed from further analyses. Lists of gene symbols were searched in Metascape v.3.5 (Zhou et al., 2019) with input species as any and analysis as human to find GO terms of biological processes, pathways, and protein interactions. Gene symbols were also searched on google scholar using the gene symbol and salmonid to determine if there were any relevant genes in already published data.

Chapter 3: Results

3.1 Genome Mapping, Filtering, And Sequence Depth

The target sequencing depth was 3x and average depth per stream shows a range between 1.8x and 2.6x after filtering out low quality reads (Figure 3).

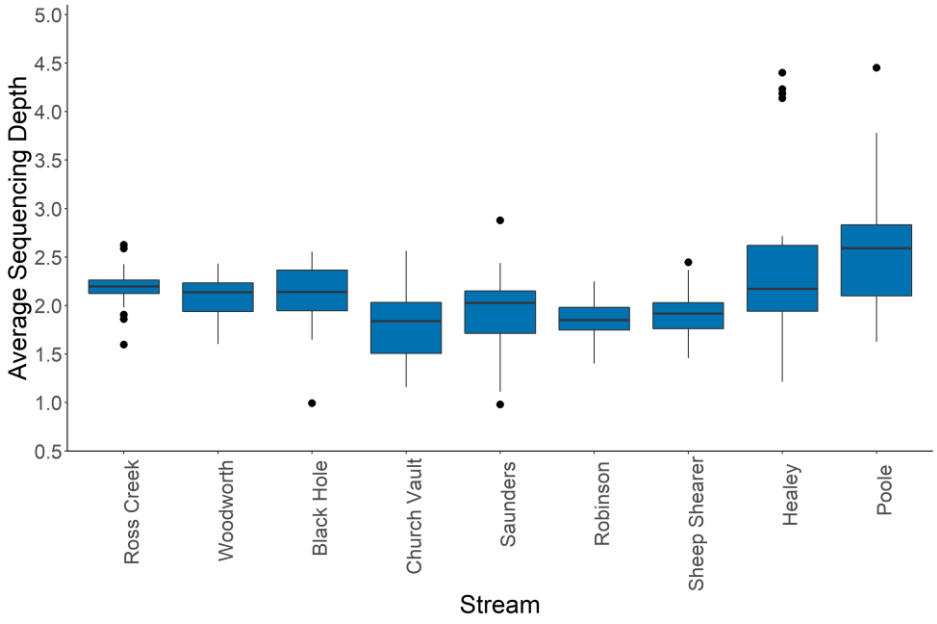

Figure 3 Average sequencing depth per stream (population) calculated at all positions covered by reads using the lake trout (*Salvelinus namaycush*) reference genome. Error bars are 95% confidence intervals. Streams are ordered from East to West among the North Mountain, Nova Scotia.

3.2 ANGSD SNP Calling And Genotype Likelihoods

PCA of the covariance matrix representing 1,745,248 SNPs show grouping of individuals by population and geology except for Poole Brook (Figure 4A). PC1 explains 7.92% of variation and PC2, 5.68% of variation. PCA done using likelihoods and frequencies determined with the SNP list following SNP discovery shows similar groupings (Figure S1). An LD pruned SNP list was used with ANGSD to produce a PCA where PC1 explains 6.41% of variation and PC2, 5.01% of variation in genotype likelihood data (Figure 4B). Individuals from scoured glacial till and thick stony glacial till group closer together on PC1 and PC2 than figure 4A that is based on all SNPs. Results using GATK for genotype likelihoods were similar (Figure S2).

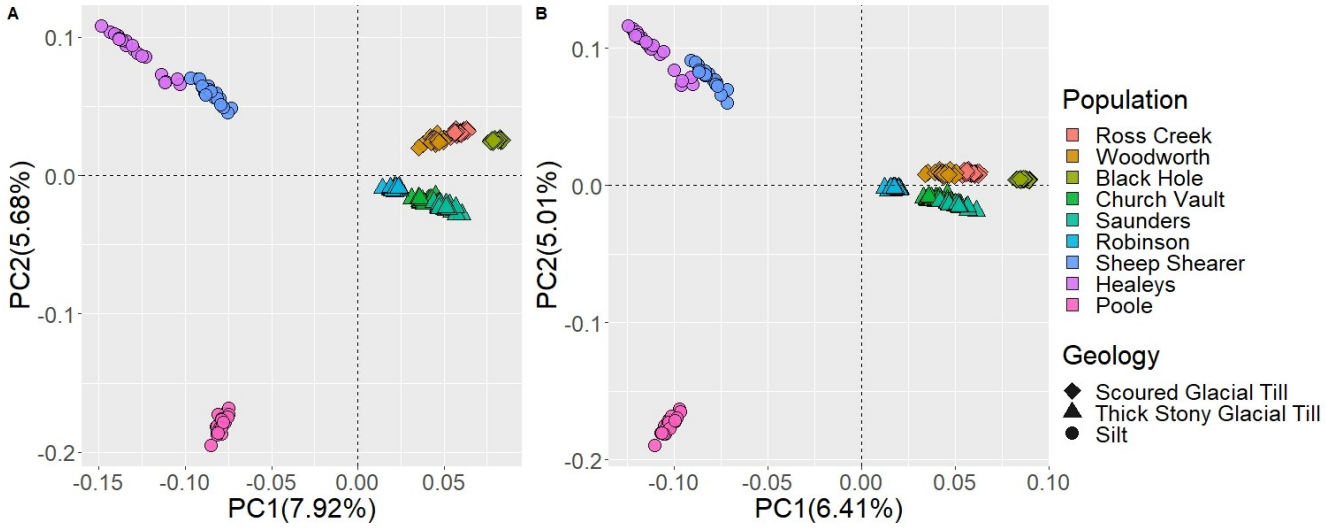

Figure 4 PCA with genotype likelihoods for N=192 brook trout. A) Covariance matrix based on 1,745,248 SNPs. B) Covariance matrix based on 302,953 SNPs (i.e., SNPs left after pruning for SNPs in LD). Colors indicate streams and shapes indicate geology.

3.3 Admixture

The best K as determined from delta K in CLUMPAK is 4. Admixture proportions of K4 indicate that individuals are grouped based on geology except for Poole Brook which is different from the neighboring Healeys Brook and Sheep Shearer both located within the same silt geology type (Figure 5).

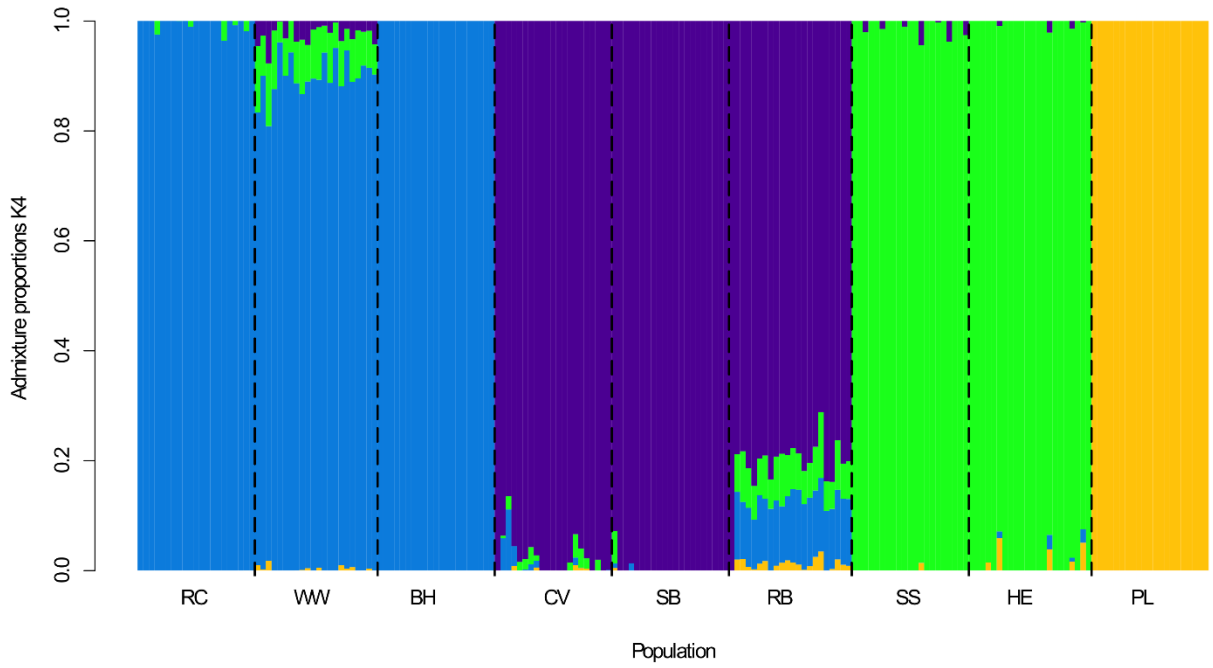

Figure 5 Admixture proportions for K=4. Each bar represents an individual brook trout. RC= Ross Creek, WW= Woodworth, BH= Black Hole, CV=Church Vault, SB= Saunders Brook, RB= Robinson, SS= Sheep Shearer, HE = Healeys, PL= Poole Brook. Admixture analysis was done using the LD pruned SNP list. Populations group by geological background with the exception of Poole.

3.4 Potential Inversion Support

3.4.1 Linkage Disequilibrium

Heatmaps of LD for all populations pooled using the lake trout genome indicate multiple blocks of LD ($r^2 > 0.60$) on 30 chromosomes with chromosome 9, 14, 19, and 24 (NC_052315.1, NC_052320.1, NC_052325.1, and NC_052330.1) containing the largest LD blocks (Figure 6, Table 2). The size of LD blocks seen on these chromosomes differ with the largest on chromosome 14 and 19 while chromosome 24 contains two LD blocks (Figure 6, Table 2). Genotype likelihoods and minor allele frequency were estimated for SNPs in LD blocks for all 30 chromosomes which contain LD blocks (Figure 6) using all individuals and PCAs generated (not included). Chromosome 9 contains 26,792 SNPs, chromosome 14 21,488 SNPs, chromosome 19 25,705 SNPs, and chromosome 24 31,569 SNPs.

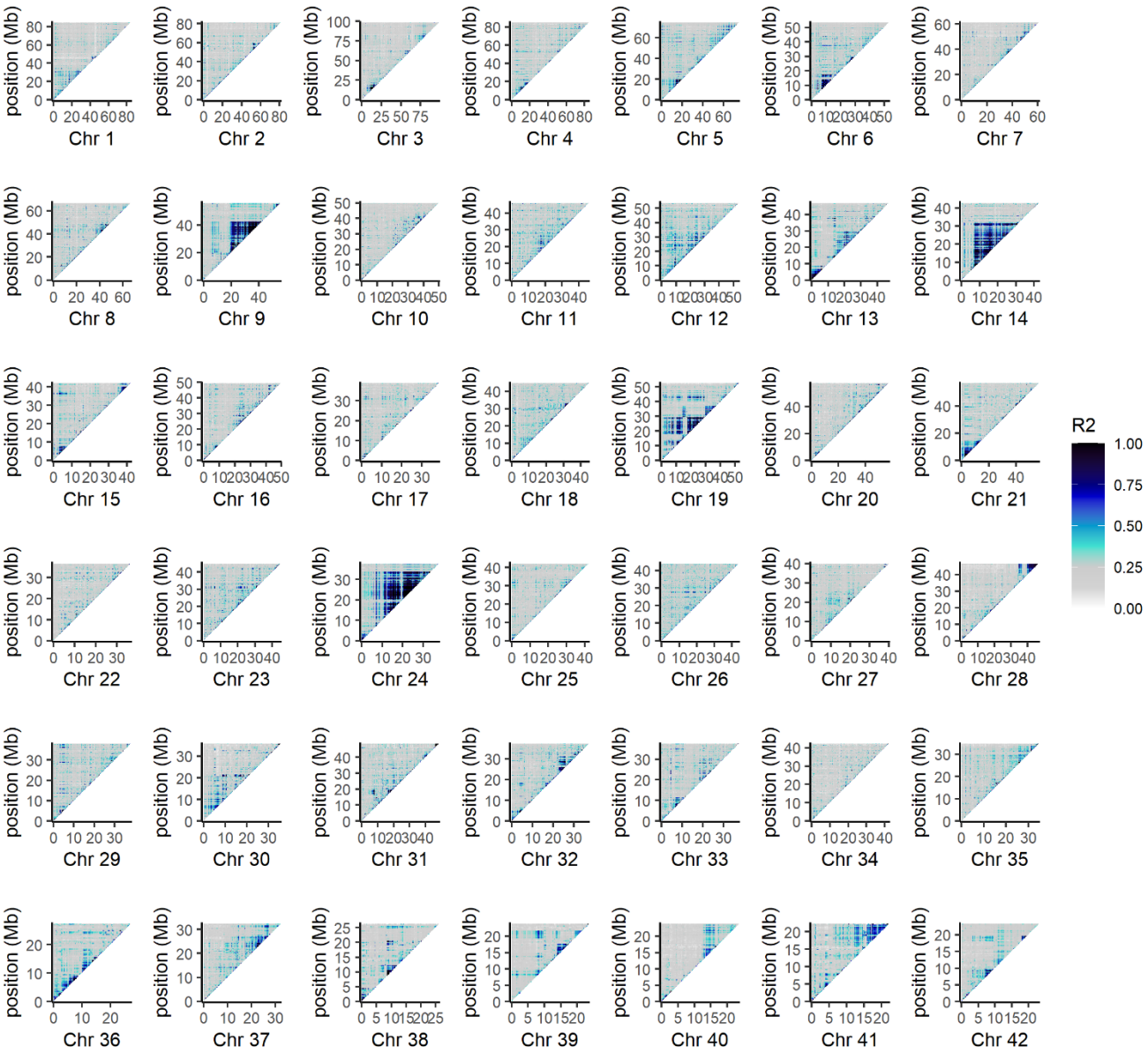

Figure 6 Heatmaps of r^2 from LD calculations for 192 brook trout on all 42 chromosomes in the lake trout genome. The second percentile of r^2 values in 250 kb windows were used for plotting. High r^2 values and dark colours represent areas on a chromosome with high LD.

Table 2 Size of LD blocks on four lake trout chromosomes. The size was determined from the r^2 of LD depicted in heatmaps produced for these chromosomes using all 192 brook trout.

Chromosome	Size of LD block (mb)
9 (NC_052315.1)	20-45
14 (NC_052320.1)	5-30
19 (NC_052325.1)	0-31
24 (NC_052330.1)	11-19 and 21-35

Heatmaps were also produced with populations pooled by geology (scoured glacial till, thick glacial till, and silt), and individually by populations to further look at patterns of LD.

3.4.1.1 Chromosome 9, 19, and 24

Large LD blocks are seen on chromosomes 9, 19, and 24 when all populations are pooled (Figure 6). When individuals from streams in the west (Healeys, Sheep Shearer, Poole) are excluded, the LD pattern disappears in the remaining 6 populations pooled (Figure 7A and 7B) suggesting the LD patterns in chromosomes 9, 19, and 24 appear exclusively in the three western populations (Healeys, Sheep Shearer, Poole). For all three chromosomes (i.e., chromosomes 9, 19, 24) I then plotted the LD patterns individually by populations (panels A, B, and C in Figures 8-10) and pooled within substrate type (Figures 8D, 9D, 10D). Once again, no LD was observed in the populations from either the scoured or the thick stony glacial till, but LD persisted in the three western populations from the silt substrate, whether examined individually (panels A,B,C in Figs 8-10) or pooled (panel D in Figures 8-10). That this LD pattern persists when the samples from the three populations are pooled suggests the LD is not due to chance due to small sample size.

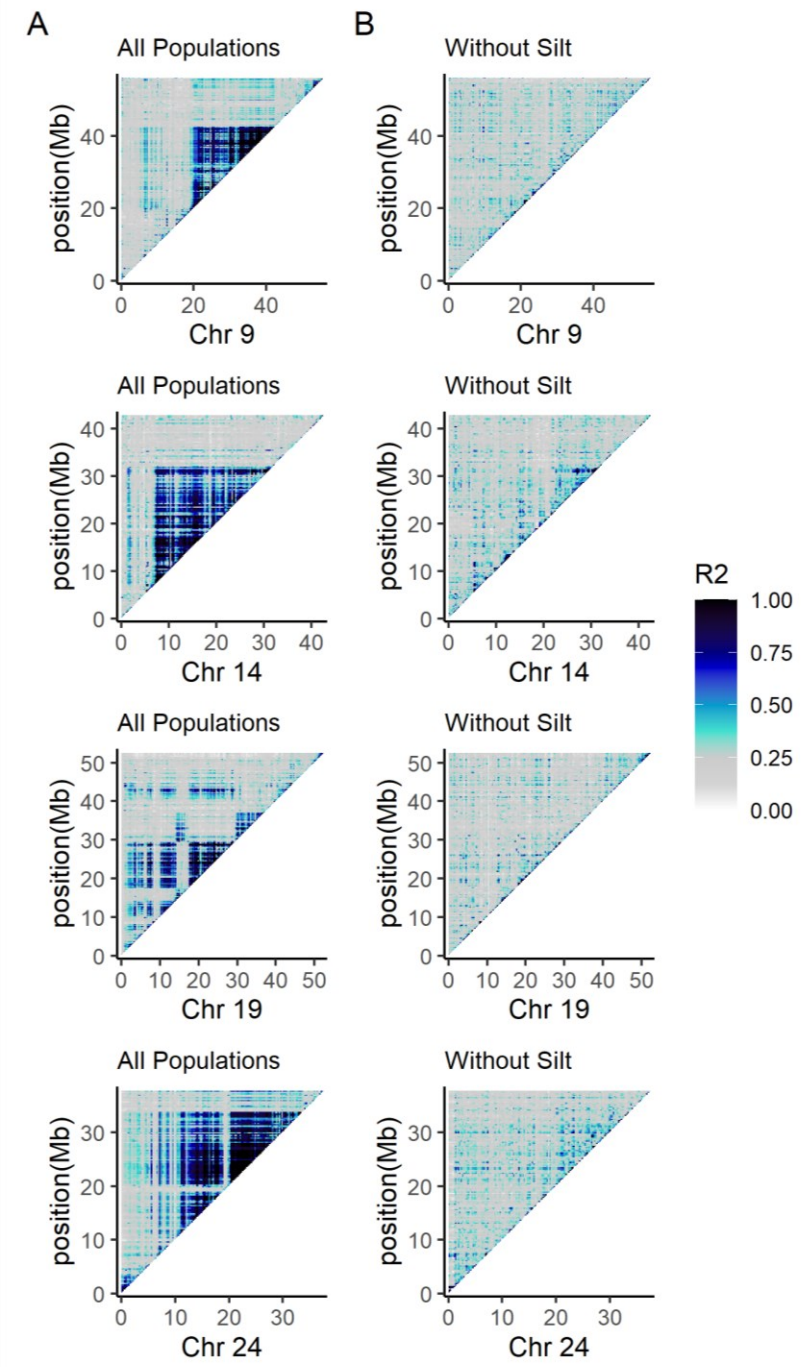

Figure 7 Heatmaps of chromosomes 9, 14, 19, 24 (NC_052315.1, NC_052320.1, NC_052325.1, NC_052330.1) from the lake trout genome. r^2 values in 250kb windows of the second percentile from LD calculations are shown. A) Heatmaps of all 9 populations pooled B) 6 populations excluding individuals from silt streams (Healeys, Sheep Shearer, Poole). High r^2 values are those with dark colours and represent areas of high LD.

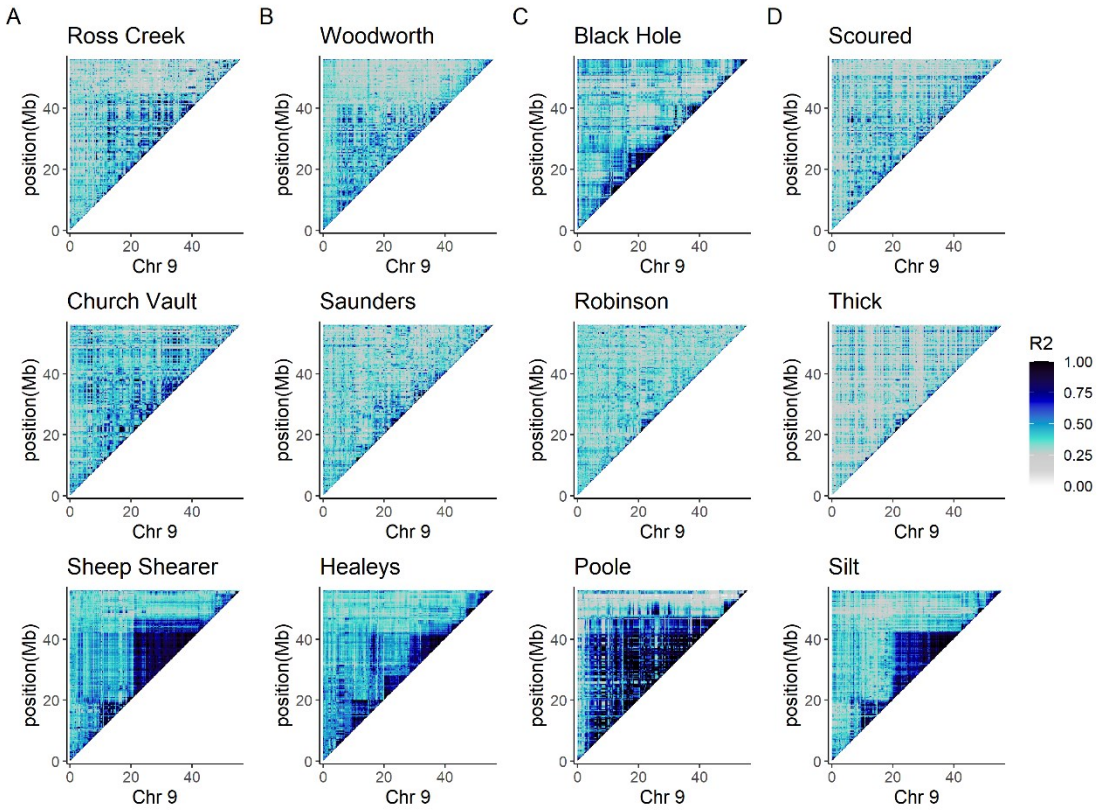

Figure 8 Heatmaps of chromosome 9 (NC_052315.1) from the lake trout genome. r^2 values from LD calculations in percentiles are depicted. High r^2 values are those with dark colours and represent areas of high LD. A-C) Individual populations, D) populations pooled by geology type (Scoured glacial till, thick glacial till, silt).

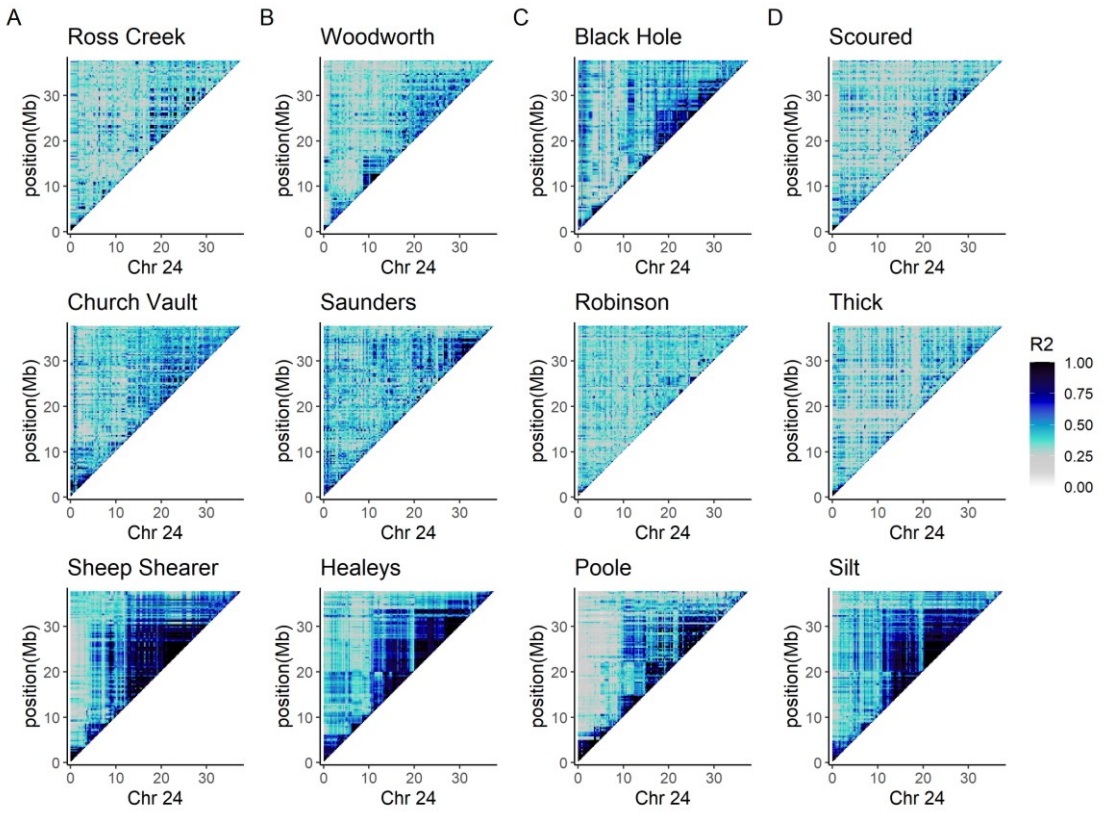

Figure 9 Heatmaps of chromosome 24 (NC_052330.1) from the lake trout genome. r^2 values from LD calculations in percentiles are depicted. High r^2 values are those with dark colours and represent areas of high LD. A-C) Individual populations, D) populations pooled by geology type (Scoured glacial till, thick glacial till, silt).

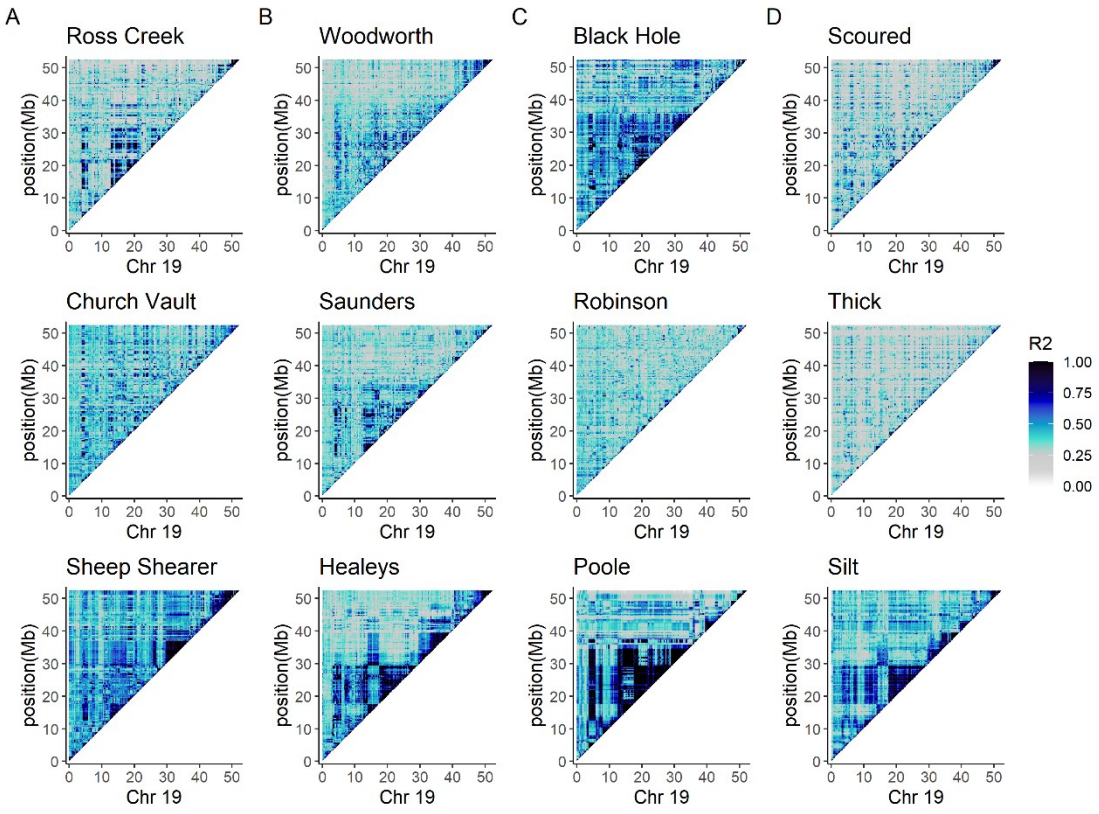

Figure 10 Heatmaps of chromosome 19 (NC_052325.1) from the lake trout genome. r^2 values from LD calculations in percentiles are depicted. High r^2 values are those with dark colours and represent areas of high LD. A-C) Individual populations, D) populations pooled by geology type (Scoured glacial till, thick glacial till, silt).

3.4.1.2 Chromosome 14

A large LD block is also seen on chromosomes 14 when all populations are pooled, and the LD block disappears when streams in the west are excluded (Figure 7). LD blocks are seen in western streams individually and when pooled by geology suggesting that LD in individual populations is not due to sample size (Figure 11). On chromosome 14, a large LD block is also present among individuals from Woodworth and this pattern persists when populations are pooled by geology in scoured glacial till but disappears when western populations are excluded (Figure 7, 11). This unique pattern is also present to a lesser extent in individuals from Ross Creek and Black Hole. This pattern is not seen on chromosomes 9, 19, and 24.

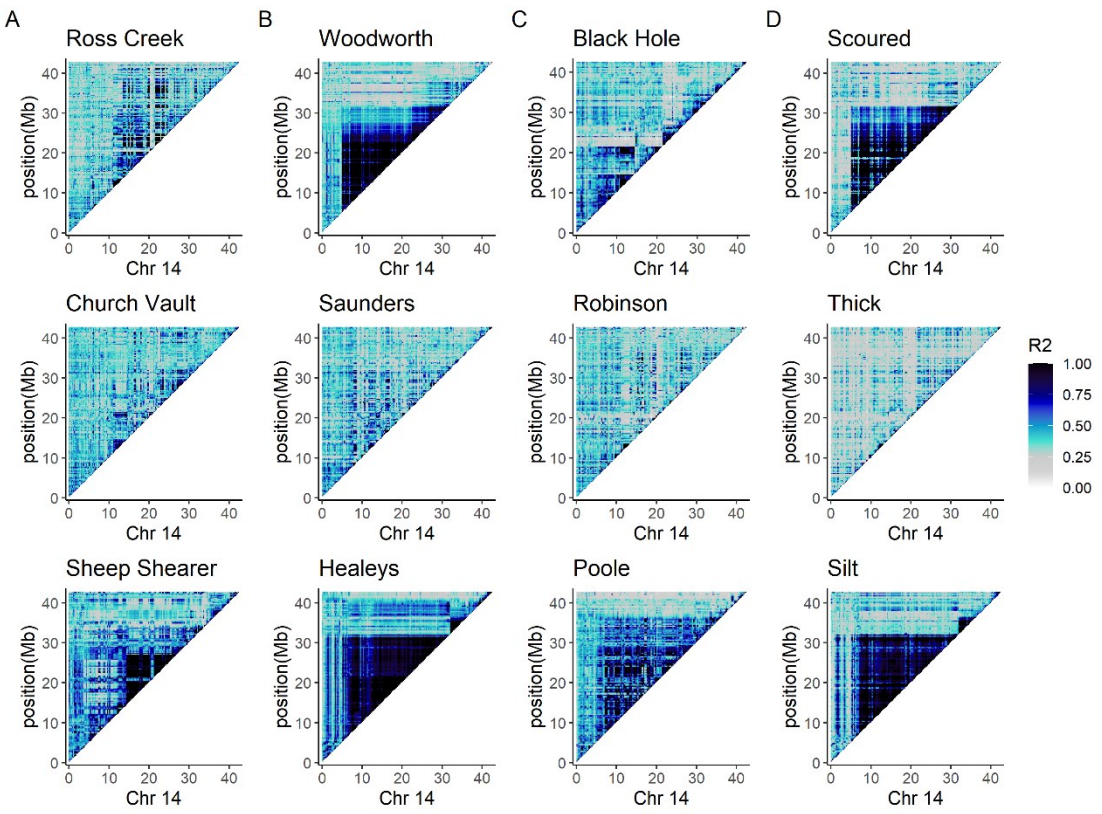

Figure 11 Heatmaps of chromosome 14 (NC_052320.1) from the lake trout genome. r^2 values from LD calculations in percentiles are depicted. High r^2 values are those with dark colours and represent areas of high LD. A-C) Individual populations, D) populations pooled by geology type (Scoured glacial till, thick glacial till, silt).

3.4.2 Genotype Likelihoods

PCA of chromosomes 9, 14, 19, and 24 show 3 groups with individuals clustering as heterokaryotypes with and without inversions on the left and right side and heterokaryotypes in the middle which is expected for potential chromosomal inversions (Figure 12). For all chromosomes, the number of individuals with the potential inversion is 5 or fewer and most individuals fit into the homokaryotypes without inversions group. Chromosome 14 and the first block of LD on chromosome 24 have 4 individuals with the potentially inverted arrangement in PCAs (Figure 12). On chromosome 9 individuals from Sheep Shearer and Healeys are found in each of the 3 groups indicating there are individuals with and without potential inversions. A similar pattern is found in both LD blocks on chromosome 14 and 24 where there are individuals with and without inversions from Healeys. Individuals from Woodworth show high levels of LD on chromosome 14

(Figure 11) and in the PCA some of the individuals from Woodworth group closer to the middle representing heterokaryotypes (Figure 12). On chromosome 19 there are no individuals from Poole that have the non-inverted arrangement (Figure 12).

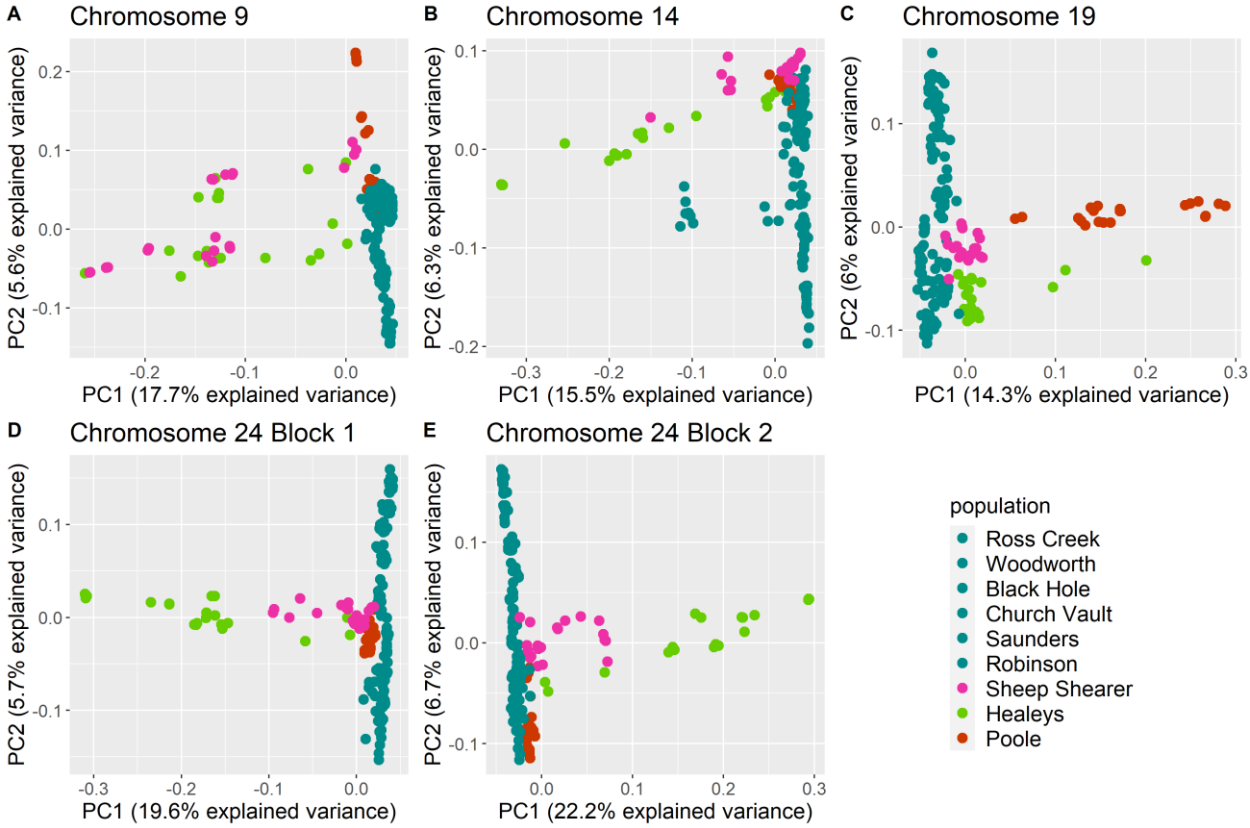

Figure 12 A-E PCA for LD blocks on Chromosomes 9, 14, 19, and 24. There were two distinct LD blocks on chromosome 24 where D) represents the first block and E) represents the second. The 3 clusters in A-E represent the different karyotypes expected when there is an inversion. The middle group on each PCA represents heterokaryotypes that have one copy of the inversion, the left and right groups represent either individuals with two copies of the inversion arrangement or no inversion arrangements.

3.4.3 Genetic Differentiation

Manhattan plots of five individuals from each of the 3 groups except chromosome 9 and the first LD block of chromosome 24 where there were four individuals show a peak of high F_{ST} SNPs on potentially inverted chromosomes despite genome wide high levels of F_{ST} (Figure S8-12). Some peaks such as on chromosome 9 are more distinct than other potentially inverted chromosomes, but these high levels of F_{ST} provide further support for potential inversions.

3.4.4 Observed Proportion Of Heterozygotes (H_{obs})

Average observed heterozygosity (H_{obs}) was calculated in 10kb windows across chromosomes 9, 14, 19, and 24 to further support potential inversions. Lower average H_{obs} is seen among individuals homozygous for the potential inversion and for the alternate arrangement where heterokaryotypes show expected higher levels of average H_{obs} (Figure 13, S14). The start and stop position of potential inversions was more precisely estimated from average H_{obs} compared to the LD blocks and 3 of the 4 inversion sizes were adjusted. The potential inversion on chromosome 9 is from 22-42Mbp, chromosome 14 was unchanged at 5-30Mbp, and chromosome 19 is from 13-25Mbp. Chromosome 24 was thought to have two potentially separate inverted regions, and average H_{obs} suggests the potential inversion is in the second chunk further along the chromosome at 20-32Mbp (Figure S13).

Average H_{obs} was calculated within potentially inverted regions on chromosomes 9, 14, 19, and 24 using windows the size of potential inversion regions. This produced the expected pattern of very low heterozygosity within the LD block in individuals that have two copies of the inverted alleles (left column), intermediate H_{obs} in homokaryotypes with two copies of the non-inverted alleles (right column), and very high H_{obs} particularly within the LD block among heterokaryotypes have the highest H_{obs} on all 4 chromosomes (Figure 13).

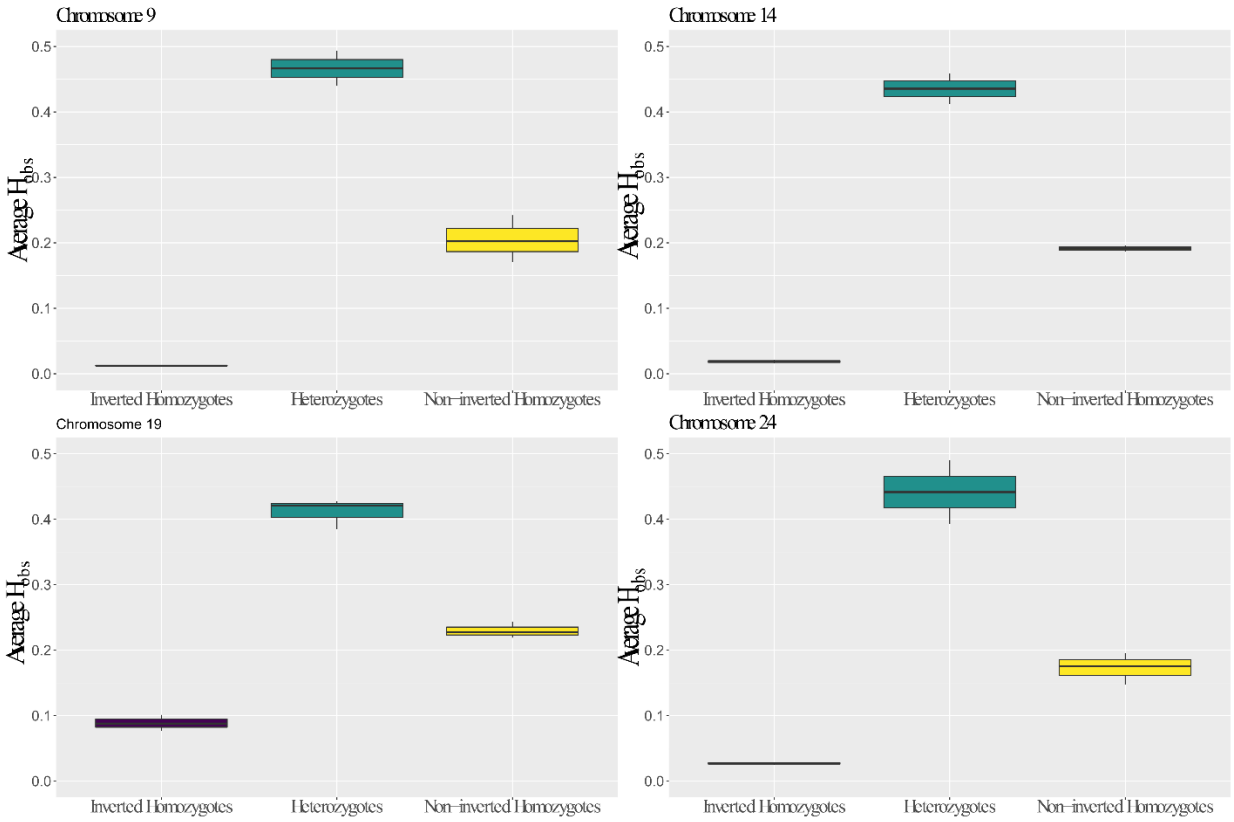

Figure 13 Average observed proportion of heterozygotes (H_{obs}) in windows the size of potential inversion regions. Individuals included in each of inverted homozygotes, heterozygotes, and non-inverted homozygotes are those represented by the 3 different karyotype groups in PCAs of LD blocks (Figure 12). For chromosome 9 there are 5 potential inverted homokaryotypes, 23 potential heterokaryotypes, and 140 potential non-inverted homokaryotypes. For chromosome 14 there are 4 potential inverted homokaryotypes, 9 potential heterokaryotypes, and 163 potential non-inverted homokaryotypes. For chromosome 19 there are 7 potential inverted homokaryotypes, 12 potential heterokaryotypes, and 169 potential non-inverted homokaryotypes. For chromosome 14 there are 4 potential inverted homokaryotypes, 11 potential heterokaryotypes, and 166 potential non-inverted homokaryotypes. All individuals from each of the inverted, non-inverted homokaryotypes and heterokaryotypes were used to calculate H_{obs}. H_{obs} is calculated as a population representing inverted homozygotes, heterozygotes, or non-inverted homozygotes per SNP position then averaged.

3.5 Nucleotide Diversity (π) And Divergence (D_{xy})

For all 4 chromosomes there is a consistent pattern of slightly higher π among potentially inverted homokaryotypes both within and outside (collinear region) the putatively inverted region (Figure 14, groups 1 and 2, respectively) than in the putatively non-inverted homokaryotypes (Figure 14, groups 3 (within) and 4 (outside) the region).

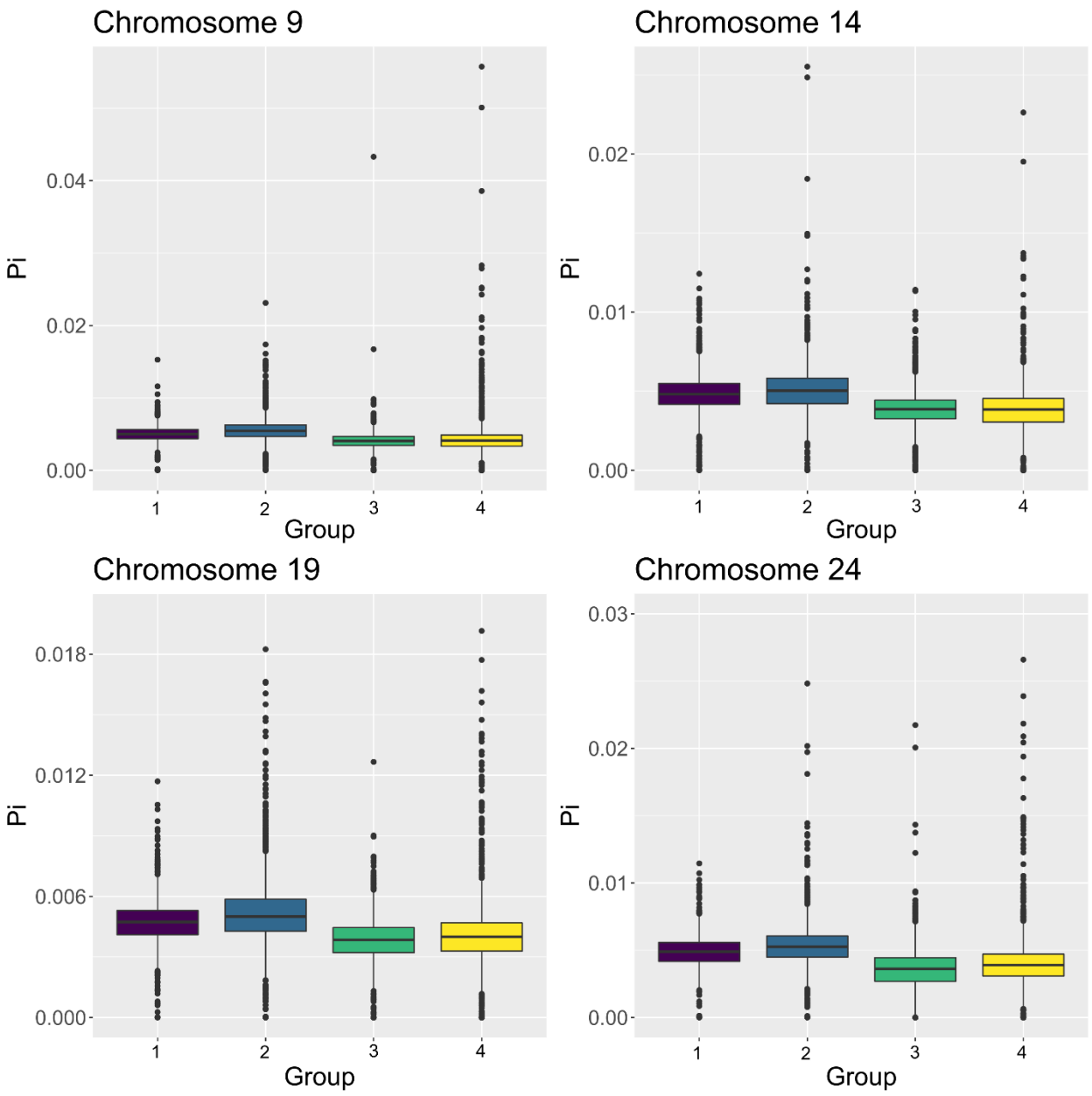

Figure 14 Nucleotide diversity (Pi) within the potential inversion region of individuals representing the homokaryotypes with and without potential inversions compared to collinear regions on chromosomes 9, 14, 19, 24. Individuals representing homokaryotypes with and without potential inversions are those in the left and right clusters of PCAs on LD blocks of chromosomes (Figure 12). Group 1 are homokaryotypes with the potential inversion where Pi is inside potential inversion regions, group 2 are homokaryotypes with the potential inversion where Pi is the collinear region of each chromosome. Group 1 and 2 are the same individuals, but Pi was calculated inside potential inversion regions and outside. Group 3 are homokaryotypes without the potential inversion where Pi is inside potential inversion regions and group 4 homokaryotypes without the potential inversion where Pi is collinear regions on each

chromosome. Group 3 and 4 are the same individuals, but Pi was calculated inside potential inversion region and outside. For chromosome 9 there are 5 potentially inverted homokaryotypes and 140 potential non-inverted homokaryotypes. For chromosome 14 there are 4 potential inverted homokaryotypes and 163 potential non-inverted homokaryotypes. For chromosome 19 there are 7 potential inverted homokaryotypes and 169 potential non-inverted homokaryotypes. For chromosome 14 there are 4 potential inverted homokaryotypes and 166 potential non-inverted homokaryotypes.

Nucleotide divergence is greater among collinear regions of homokaryotypes with inversions than inside inversion regions on chromosomes 9, 14, 19, and 24 (Figure 15). This pattern is expected if diversity was higher in homokaryotypes without inversions although homokaryotypes with and without the inversions show similar diversity (Figure 14).

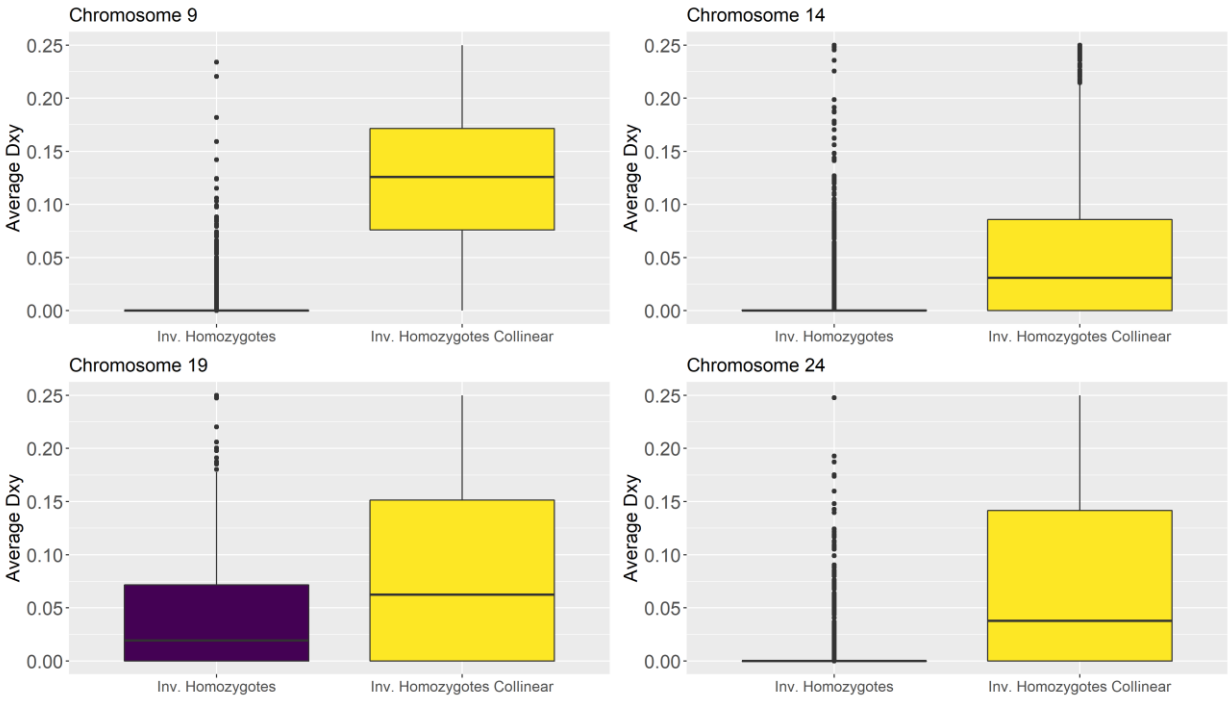

Figure 15 Average nucleotide divergence (d_{xy}) within and outside (i.e., collinear region) the inverted regions of all individuals representing homokaryotypes with potential inversions from PCAs of LD blocks shown in figure 12 which are those in the small cluster on the left for on chromosomes 9 and 14 and those on the right in clusters on chromosomes 19 and 24. For chromosome 9 there are 5 potentially inverted homokaryotypes. For chromosome 14 there are 4 potential inverted homokaryotypes. For chromosome 19 there are 7 potential inverted homokaryotypes. For chromosome 14 there are 4 potential inverted homokaryotypes.

3.6 Gene Annotations

Lists of gene symbols within potential inversions on chromosomes 9, 14, 19, and 24 were searched for gene ontology (GO) terms to look for potential biological processes, pathways, and protein interactions that might be important to local adaptation. Some genes located on all four chromosomes have a potential connection to response to temperature and western streams that these individuals are from have higher maximum daily water temperatures (Table 3,4 and Figure 16).

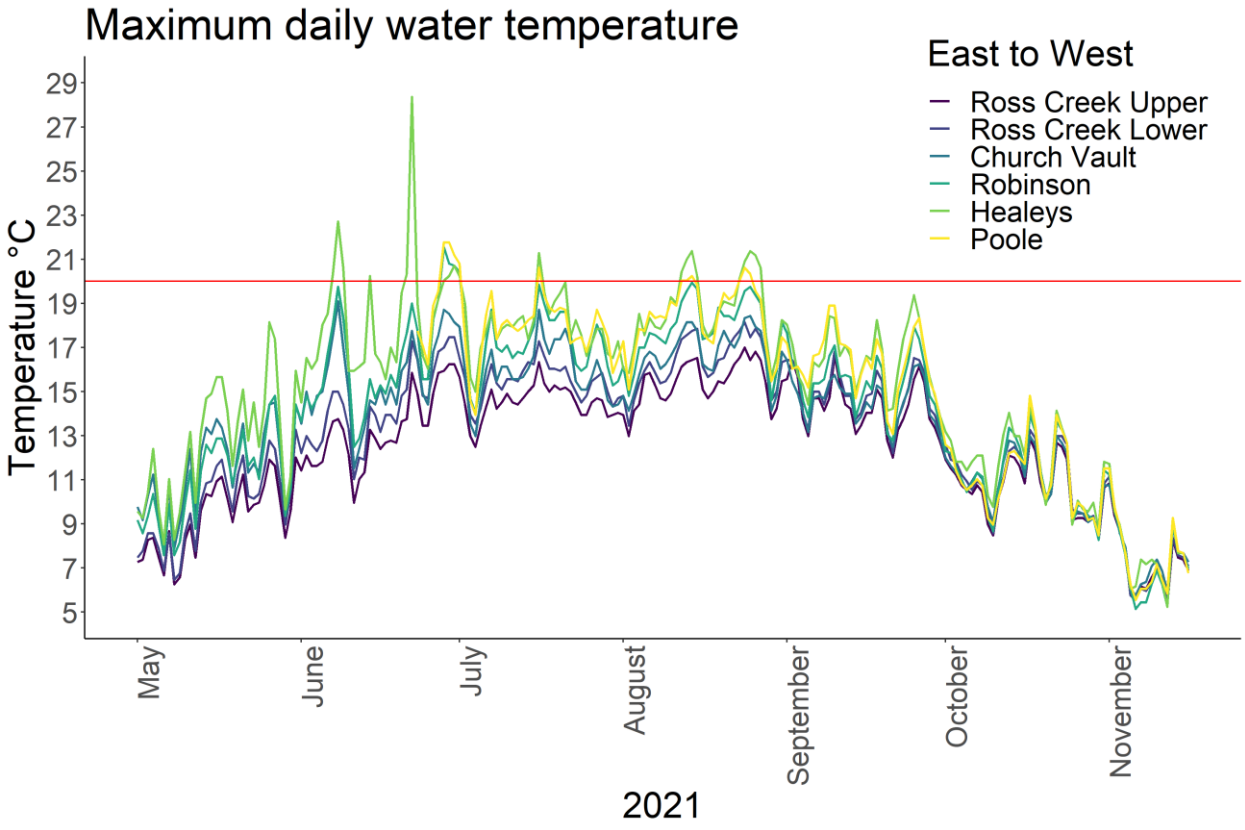

Figure 16 Maximum daily water temperatures calculated from water temperature recorded every two hours from May-November 2021 with HOBO pendant loggers. Temperature was recorded in a subset of the nine streams. The red line represents temperature that brook trout avoid.

Table 3 Biological processes and pathways involved in statistically significant gene symbols from chromosomes 9, 14, 19, and 24. Processes with a -log10(P) greater than 3 were included in the table.

Chromosome	GO biological process, KEGG Pathways, Reactome Gene Sets
9	Platelet activation, axon guidance
14	Glutathione metabolism, positive regulation of intracellular protein transport, regulation of transcription initiation by RNA polymerase II, negative regulation of binding, positive regulation of cell growth, response to temperature stimulus
19	COPII-mediated vesicle transport, regulation of cell-substrate adhesion, positive regulation of apoptotic process, positive regulation of protein localization to membrane
24	Epithelial tube morphogenesis, cellular lipid catabolic process, protein folding, muscle contraction

Table 4 Protein-enrichment for gene symbols from chromosomes 9, 14, 19, and 24.

Chromosome	Protein-protein Interaction Enrichment
9	Inflammatory mediator regulation of TRP channels, C-type lectin receptor signaling pathway, growth hormone synthesis, secretion and action
14	Regulation of TP53 activity through phosphorylation, vesicle-mediated transport in synapse, mRNA metabolic processes
19	COPII-mediated vesicle transport, ER to Golgi anterograde transport, transport to the golgi and subsequent modification
24	Muscle structure development

3.7 Distance And Geology As Predictors Of F_{ST}

GLM of distance and geology as predictors of linearized F_{ST} suggest a small positive and significant relationship between distance and linearized F_{ST} slope= 0.001894, $p<0.05$ (Table 5) suggesting that with every increase in F_{ST} the distance increases. Whether individuals are within a geology group has a negative and non-significant relationship with F_{ST} slope= -0.00643, $p>0.05$ (Table 5) suggesting that individuals within a geology group have lower F_{ST} than between geology groups, but this is not significant. The actual linearized F_{ST} plotted against the models predicted F_{ST} suggests that individuals in streams that are further away tend to have higher F_{ST} which is also seen when average F_{ST} per population comparison is plotted again distance in km (Figure 17 and Figure S7).

Table 5 Summary coefficients for GLM of distance in km and geology group as predictors of linearized F_{ST}. The model was produced using the gaussian family. Geology predictor represents individuals within a geology group or between different geology groups (scoured glacial till, thick glacial till, silt).

	Estimate	p-value
Intercept	0.214114	1.55E-08
Distance (km)	0.001894	0.00272
Within geology group	-0.00643	0.83888

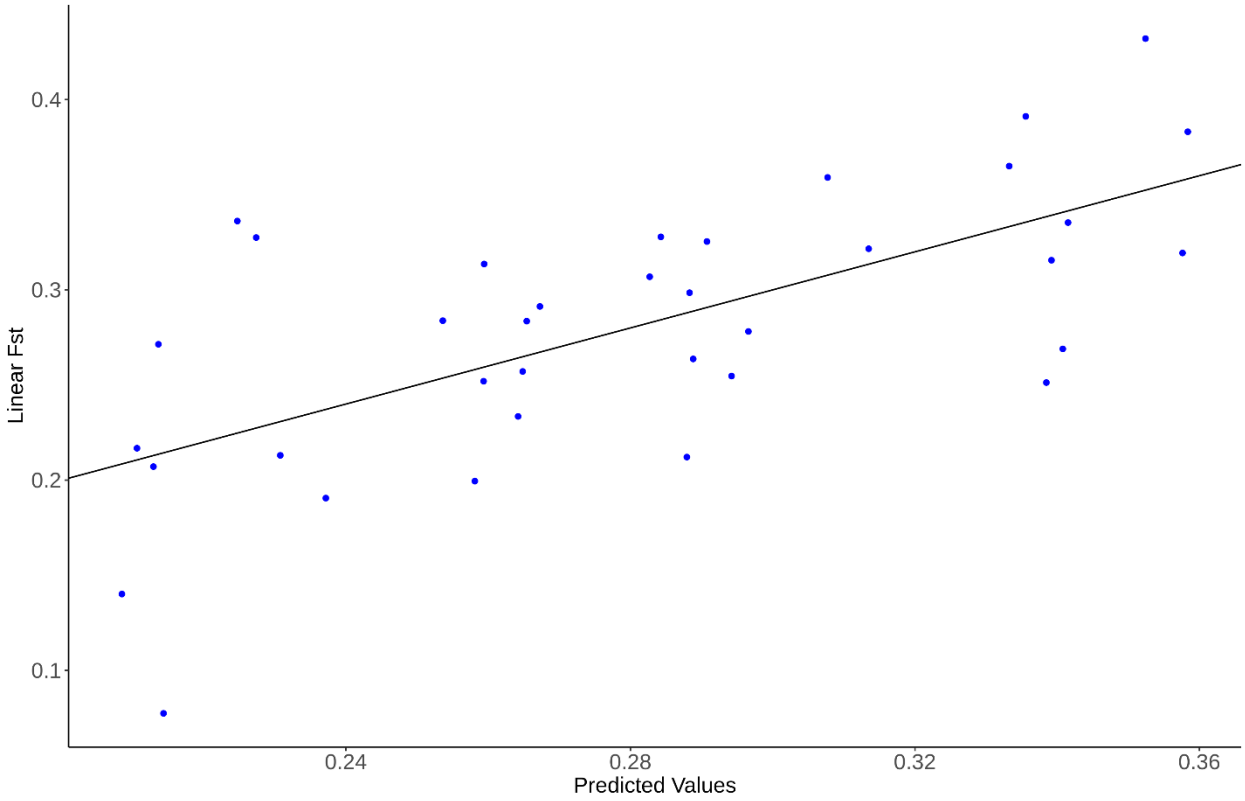

Figure 17 Predicted linearized F_{ST} based on a GLM of distance and geology as predictors of linearized F_{ST}. The y axis is the actual linear F_{ST} values and the x axis are linear F_{ST} predicted from the GLM.

Chapter 4: Discussion

4.1 Overall Findings And Importance

Four large putative chromosomal inversions representing 3% of the lake trout genome were discovered among western populations of brook trout where the surficial geology differs from the other streams. Smaller haploblocks were detected in other chromosomes and they suggest the existence of other potential structural variants, haploblocks under selection or potential smaller inversions. Patterns of LD are similar among chromosomes 9, 19, and 24 showing individuals from western streams exhibiting large LD blocks. On chromosome 14 individuals from western streams and from one stream in the east, Woodworth, show evidence of a large LD block. The fact that LD patterns occur in individual streams from the west and persists when populations are combined, suggest that the LD patterns observed in individual populations are not due to random departures from equilibrium due to small sample size. Patterns of LD and H_{obs} suggest these are potential inversions rather than other types of structural variants although they could also be large haploblocks under selection; they are also rare given that they were generally found in a low number of individuals. Gene ontology of some genes within potential inversion regions suggest biological processes that could be important to local adaptation among the western populations particularly with regards to temperature tolerance. The discovery of new potential inversions in brook trout from lcWGS adds to previous work using reduced representation sequencing methods (Sutherland et al., 2016) and provides a better understanding of inversion differences and similarities with the closely related lake trout (Boeberitz et al., 2020). The occurrence of potential inversions among predominantly western populations indicates a potential role in local adaptation likely due to temperature differences among streams suggesting local adaptation and stream colonization history could be important factors contributing to the observed differences. The cerkl (ceramide kinase like) gene found within the inversion on chromosome 19 is a relative of cerk (ceramide kinase) which is known to effect thermal adaptation in other salmonids. Cerkl could play an important role in local adaptation to temperature differences.

4.2 Caveats And Limitations To lcWGS For Inversion Detection In Species With Small N_e, Genetic Drift, And Limited To No Gene Flow

The use of whole genome sequencing has facilitated the detection of structural variants including chromosomal inversions allowing us to further understand the role of local adaptation among populations. In addition, lcWGS makes this more accessible as the cost of sequencing continues to drop and library preparation is cost effective (Lou et al., 2021; Therkildsen & Palumbi, 2017). By using lcWGS, we were able to calculate LD across entire chromosomes providing initial support for probable inversions, common in inversion discovery (Faria, Chaube, et al., 2019; Harringmeyer & Hoekstra, 2022; Mérot et al., 2021), but this technique also has caveats to consider. Inversion discovery via methods such as LD have been biased to large regions of genomes because the ability to detect reduced recombination in heterokaryotypes increases with inversion size as more SNPs will be in LD (Hale et al., 2021; Wellenreuther & Bernatchez, 2018). Large inversions are also common because they contain more alleles and genes that could be beneficial and therefore become established (Connallon & Olito, 2022). It is also important to note that LD blocks can be the result of other structural variants such as fusions or haploblocks under selection and this is why LD is the first step in inversion detection (Mérot et al., 2020). In addition, lcWGS provides genotype likelihoods (probability of observing a genotype at each site in an individual given that if diploid, there are 10 potential genotypes) rather than direct genotype calls and available software for environmental analysis often require direct genotype calls. Genotype environmental analyses such as RDA require direct genotype calls which can be obtained by converting genotype likelihoods, but this is not ideal because the confidence in calls will be low due to low coverage. There is a need therefore to develop software that can use likelihoods for genotype environmental type analyses. This shortcoming is currently addressed by genotyping a few individuals at high depth therefore increasing confidence in calls via imputation (see e.g., VanRaden et al. 2015). Using genotype likelihoods has its own drawbacks that can be addressed using filters for sequencing quality scores and parameters in commonly used programs such as ANGSD for depth and minor allele frequencies as an example (Lou et al., 2021). See the appendix for all filtering used. Small sample size from each stream in the current study likely added to biases of certain

analyses such as nucleotide diversity and F_{ST}. Both are calculated from the sample allele frequency (SAF) which considers all individuals from each population. In addition, Lou et al. (2021) showed through simulations that diversity is generally either under or overestimated depending on sample size, depth and genotype likelihood model used, whether GATK or SAMtools. While more studies are using lcWGS, the role sample size plays in analyses is not entirely known and remains an area of discussion, especially for study systems where species have small effective sizes, little to no gene flow and strong genetic drift (Fuentes-Pardo & Ruzzante, 2017; Lou et al., 2021) Previous research on some of these brook trout populations suggest that they have small effective sizes (Ruzzante et al., 2016). This allowed for discovering limitations of analyses that can be done with lcWGS. Many studies that use lcWGS have done so on marine species with large population sizes and high gene flow (e.g., Atlantic cod, Atlantic silverside; O'Donnell & Sullivan, 2021; Wilder et al., 2020). A large population size will facilitate the identification of outlier loci that might be linked to local adaptation as genetic drift does not have as large an impact and therefore genome wide F_{ST} estimates are relatively low. While F_{ST} was used in the current study as a secondary method to identify potential inversions this method is not recommended for the finding of outlier loci in small populations because drift can cause issues with false positives as seen in a study on small populations of Alpine ibex (Leigh et al., 2021)**.** In addition, rates of recombination can also impact false positive detection (Booker et al., 2020).

Another limitation to this study was the use of the lake trout reference genome due to the lack of available brook trout genome. Although brook trout is more closely related to lake trout than to other salmonids with available genome sequence (Lecaudey et al., 2018), salmonids in general have been through whole genome duplication events which could affect the number and type of structural variants between two species. While the lake trout genome is fully annotated, there are regions on the potential inversion chromosomes where there is no gene information available. Newer versions of the lake trout genome could lead to increased confidence in the placement of scaffolds on chromosomes and the discovery of more gene locations. In addition, with the discovery of these potential inversions, it remains unknown whether the reference genome of the lake trout has the ancestral or derived arrangements making it difficult to determine which arrangements in

the current study are derived or ancestral. Mapping to the lake trout genome is not ideal because it can introduce bias to analyses as regions that are most conserved will have better mapping success and it can produce lower mapping quality reads (Bohling, 2020; Fuentes-Pardo & Ruzzante, 2017). As such, steps were taken to account for this. Parameters described in the methods and appendix such as indel realignment and removal of duplicate reads were used to be more stringent. It is important to note that potential inversions discovered here could also simply be from using the lake trout genome however if this were the case, I would not expect to see these large haploblocks only in a subset of the streams.

4.3 Environmental Characteristics Vs Genetic Distance

The nine streams sampled along the North Mountain, Nova Scotia have different environmental characteristics, some of which were examined in this thesis such as the surficial geology, pH, water temperature and streamflow. Ruzzante et al. (2016) suggested environmental differences between streams could explain the large variation in effective to census population size ratios and I explored such potential relationships further. PCA of all nine populations show individuals grouping based on the surficial geology with the exception of Poole brook (Figure 4), and admixture analysis show the same pattern (Figure 5) suggesting there are likely 4 population groups. The results based on over one million SNPs add to the findings of Ruzzante et al. (2016) and suggest that pairs of populations closer together are more similar to each other than pairs of populations further apart. While the grouping of populations by surficial geology could be important, it requires further consideration. In this study system, streams on the same surficial geology type are also located closer to each other than to streams on different surficial geology types. It would be ideal if the surficial geology types were dispersed throughout the North Mountain to help further understand this potential relationship. That the streams on the same surficial geology type are close together could confound the patterns I have found in my thesis. Surficial geology and bedrock characteristics are important factors to consider when looking at local adaptation to stream dwelling fish because of its interconnectedness with streamflow and water temperature via surface drainage and conductivity (Newton et al., 1987; Webster et al., 2006). The North Mountain consists of three flow units, upper, middle, and lower which have basalt

formations of variation in thickness underlying surficial geology (Kontak, 2001). Streams in the west generally cut through the lower and middle flow units while all other streams are among the middle flow unit or cut through the middle and upper flow unit. Evidence from the current study suggests that streams in the west are warmer and have lower streamflow which can in part be explained by silty till which due to its thinness, would reduce groundwater availability. While surficial geology seemed important to population structuring, GLM analysis of distance between streams and surficial geology (Table 5, Figure 17) indicates distance is a significant predictor of linearized F_{ST}. This pattern of isolation by distance (IBD) was likely the result of the progressive colonization of streams from west to east currently maintained by low levels of gene flow. Although surficial geology is not a significant predictor, it is important to note that western streams dried up in the summer months and made streamflow nil or at least impossible to measure. That surficial geology was not a significant predictor is not to say it does not play a role in the patterns I see. Linearized F_{ST} was calculated across the genome and not just within potential inversion regions. It is thus possible that if this analysis was done focusing on potential inversion regions that it could be more significant. A complete genotype-environment analyses would be required to more effectively examine the potential role of surficial geology. Future work might focus on the effects of habitat fragmentation due to drying of streams and its role in local adaptation which presumably will intensify as climate change continues (Blackman et al., 2021). Genetic drift is pronounced in small populations and can outcompete the role of natural selection. While we do not know which is the case here, extensive research on brook trout in Cape Race Newfoundland suggests that despite small population sizes, brook trout are able to adapt in fragmented habitats through phenotype-environment associations (Wood et al., 2014, 2016; Wood & Fraser, 2015; Zastavniouk et al., 2017). The putative inversions discovered here suggest they could be important in local adaptation and might have a connection to phenotypic plasticity if genes in potentially inverted regions are related to characteristics such as growth, colouration and metabolism. While more work is needed for gene annotation, results indicate that all four chromosomes contain genes that could be involved thermal adaptation (Table 3).

4.4 Stocking And Hybrids

It is important to note that the presence of potential inversions in only western streams is not likely due to stocking. To our knowledge none of the studied streams have been stocked with brook trout. Rumsey Lake, which drains into Poole Brook, was stocked once with brook trout (~200) in 1997 and multiple times with rainbow trout starting in the 1970s and continuing to the present. Elliot Lake, which drains into Healeys and Sheep Shearer brook, was stocked with brown trout and tiger trout a few times in the 1990s and early 2000s (John MacMillan, personal communications, July 2022). Two brown trout individuals were sampled in each of Healeys and Sheep Shearer but they are phenotypically quite distinct from brook trout, so it is highly unlikely we sampled brown trout that we thought were brook trout. Tiger trout are sterile suggesting its unlikely that any mating occurred between brook trout and tiger trout that could cause genetic differences among populations. No tiger trout were sampled in either steam. Although there is potential for undocumented stocking, an extensive study on brook trout in Nova Scotia based on microsatellite DNA suggests that introgression of hatchery and wild stocks is limited and therefore unlikely to have an impact (Lehnert et al., 2020). It is worth noting that the presence of brown trout in Healeys and Sheep Shearer brook could create competition for resources. In all other sampled streams, American eels are the only competitors (John MacMillan, personal communication, May 2021).

4.5 Deglaciation, History Of Colonization And Inversions

The discovery of potential inversions unique to western streams could be explained by the post glacial colonization history of brook trout in the North Mountain as a result of deglaciation of Atlantic Canada. Different patterns of ice advance and retreat took place in the North Mountain region of Nova Scotia during the late Wisconsinan phase approximately 14,000-10,000 years before present (Stea et al., 1998). During this period, different marine incursion and recession events generated a post-glacial history that led the western streams (Healeys, Sheep Shearer, Poole) to deglaciate sooner than the eastern streams (Shaw et al., 2006; Stea, 2004, John Gosse, personal communications, November 2022). The specific timing of post-glacial history is unknown for the study streams. We do know, however, that the western streams deglaciated first. The timing of deglaciation

is expected to have affected the stream colonization sequence contributing to the IBD pattern we observed in the present populations aided by minimal gene flow among streams. This is also consistent among other populations of brook trout where IBD patterns created from postglacial colonization help genetic structuring with limited gene flow for example see Pilgrim et al., 2012. In a more recent study on brook trout from Québec by Ferchaud et al. (2020), they found that high levels of genetic differentiation and low levels of IBD affecting adaptive variants can be partially explained by colonization history. Colonization history of brook trout involves 3 refugia of which eastern Canada contains individuals with Atlantic and Acadian refugia assemblages (Danzmann et al., 1998). Mitochondrial DNA (mtDNA) work on brook trout from New Brunswick, Nova Scotia, and Cape Breton suggests however, that all populations in this region derive from one lineage (Jones et al., 1996). It is important to note that no original research on brook trout assemblages used samples from any of the streams sampled in the present study, so it is possible that another assemblage exists given the history of deglaciation. The colonization history of brook trout in Labrador is hypothesized to be from Mississippian and Atlantic refugia and/or an Acadian refugium (Black et al., 1986; Danzmann et al., 1998). Results based on microsatellite analyses of brook trout in Labrador suggest a mix of the two hypotheses (Pilgrim et al., 2012) so it seems possible that North Mountain brook trout could be of different lineages. Perhaps when the western streams deglaciated, brook trout founded those streams and underwent a population bottleneck generating the genetic differences that we see currently. Analysis of mitochondrial DNA would help answer the question of whether there are different lineages of brook trout in the North Mountain.

4.6 Inversion Detection And Establishment

The discovery of chromosomal inversions has been facilitated by whole genome sequencing, and it is allowing us to better understand the role of inversions in local adaptation and speciation. Four potential chromosomal inversions were discovered among individuals from three western streams Healeys, Sheep Shearer and Poole as determined by high LD, grouping of karyotypes on PCAs and H_{obs} (Figures 7, 12, 13). Lower H_{obs} among individuals in the homokaryotype group with potential inversions than among individuals in the homokaryotype group without them (Figure 13) also provides

support that the potential inversions on all four chromosomes are likely the derived state as opposed to ancestral because we expect a younger arrangement to have lower diversity due to less time for the accumulation of mutations. That these putative inversions are only found in some populations suggests they could be adaptive despite known high levels of genetic drift and limited to no gene flow in all studied populations. While LD blocks are a first step indication that inversions might be present, genetic hitchhiking during selective sweeps and natural selection could also generate haplotype blocks that are not SVs. This is seen in Atlantic herring where large haplotype blocks are adaptive and maintained without evidence of structural variation (Lamichhaney & Andersson, 2019). In addition, other types of SVs such as chromosomal fusions can generate haploblocks such is the case in Atlantic salmon (Wellband et al., 2019). It is unlikely that the putative inversions are other types of SVs although they could represent large haploblocks under selection. Filtering of data after genome alignment removed the possibility for the presence of indels and duplications, which rules out some types of transposable elements and CNVs. While some CNVs can cause reduced recombination, it is unlikely the case here given the large size of potential inversions discovered (Table 2).

The patterns of diversity discovered here (Figure 14) are unlikely the result of hard selective sweeps as was the case in a study on marine snails (Faria et al. 2018). If selective sweeps, which can last many generations, were driving the large LD blocks, we would not see the three different karyotypes in PCAs (Hale et al., 2021). In such a case, individuals with potentially inverted sequences are expected to exhibit lower diversity (Cruickshank & Hahn, 2014) and this is not what we see (Figure 14). On the other hand, if an inversion is young relative to the alternate arrangement, there will also be lower diversity in the homokaryotype with inversions than in the homokaryotypes without inversions (Faria, Chaube, et al., 2019). Additionally, divergence can be used to determine age of inversions. Mérot et al. (2021) found higher divergence among inverted arrangements and estimated time of divergence using a molecular clock and generation times. We see slightly higher diversity in homokaryotypes with potential inversions than in homokaryotypes without potential inversions (Figure 14) and divergence is much lower in potentially inverted regions among homokaryotypes with inversions than in collinear regions (Figure 15). Over time differences in diversity between arrangements is

expected to decline as mutations accumulate while divergence should increase from supressed recombination in heterokaryotypes (Faria, Chaube, et al., 2019), but this can also be impacted by types of selection. In a study on mimetic butterflies, divergence time was calculated using a molecular clock and within inversion regions, divergence time was lower than in collinear regions suggesting the inversion occurred after lineages split and gene flow is responsible for inversion maintenance (Jay et al., 2018). Further research on mitochondrial DNA would need to be conducted to improve our understanding of whether putative inversions discovered here are young or old given the contrasting patterns of diversity and divergence. It is also important to note that we do not know whether the individuals used to generate the lake trout reference genome have the alternate or derived arrangement. Gene flux which refers to the exchange of alleles from inverted regions to collinear regions in heterokaryotypes could also potentially explain why we see lower divergence among individuals with potential inversions (Guerrero et al., 2012; Navarro et al., 1997). PCAs of LD blocks (Figure 12) show some individuals grouping between homokaryotypes and heterokaryotypes, an indication that there is gene flux, though this could also be from more than one inversion on a chromosome such is the case for rainbow trout on Omy05 (Pearse et al., 2019), or that the size of the inversion is actually smaller or larger than captured here. It is possible that the size of the potential inversions were not optimally estimated with LD. If these are potential inversions, long read sequencing would provide much more accurate locations of breakpoints and could help determine if size can be used to explain these individuals. Gene flux can, in this situation, be a form of gene flow or migration between the different karyotypes and in that sense, partially explain the lower divergence in homokaryotypes for the potential inversion arrangement (Guerrero et al., 2012).

The lack of similar findings where divergence is lower among individuals with inverted arrangements than in those with collinear regions could also be due to the complicated relationship involved in inversion establishment and maintenance that can change between arrangements through time (Faria, Chaube, et al., 2019). While I am not sure what the exact age of the potential inversions might be, the age could be affected by the timing at which environments developed differences that contribute to possible local adaptation. Here we see more heterokaryotypes relative to homokaryotypes with potential

inversions which suggests that the establishment of these inversions are type II in that we see heterokaryotypes and homokaryotypes for the inversions (Faria, Johannesson, et al., 2019). It is possible that two forms of balancing selection in overdominance and associative overdominance are acting among the western populations where potential inversions could be accumulating a deleterious mutational load (Berdan et al., 2021). In this case, homokaryotypes are expected to have reduced fitness and this could help explain the greater number of heterokaryotypes than homokaryotypes. This has been seen in the supergene of butterflies (Jay et al., 2021) and balancing selection in general is responsible for maintaining inversions in other species such as rainbow trout (Pearse et al., 2019) and the stick insect (Lindtke et al., 2017). Additionally, we know that these populations have small effective sizes and inbreeding is prevalent (Ruzzante et al., 2016) which would possibly make it harder for deleterious mutations to be purged, however the heterokaryotype advantage allows the maintenance of the potential inversions, or likely has thus far. In addition, evidence of local adaptation for thermal tolerance suggests that divergent selection is also potentially playing a role in the current study system as these genes may provide a fitness advantage. While we do not know the age of these potential inversions, over time the maintenance of inversions could change with selection, drift, and recombination in combination with accumulations of mutations and epistatic interactions which could have an impact on diversity and divergence.

4.7 Inversions In Salmonids And Adaptive Advantage

These four potential inversions add to the growing number of structural variants discovered among salmonid species. Although the lake trout reference genome was used in this study, we find similarities to previously discovered inversions and add potential diagnostic ones not found in other salmonids such as Arctic charr, rainbow trout, and Atlantic salmon. Chromosome 24 (Sna 24) of the lake trout genome is known to be differentiated between Arctic charr, rainbow trout, and Atlantic salmon and results here further support that this inversion is likely polymorphic between lake trout and brook trout. Sna 10, 11, and 34 are shared chromosomes with inversions between lake trout and brook trout, however I did not further assess them because they did not have large LD blocks. It is possible that we do not see LD blocks with high r^2 because the brook trout sampled do not have this inversion which is probable considering that North Mountain

brook trout have not been included in inversion studies to date or that this is an artefact of using the lake trout reference genome. Sna 12, Sna 23, and Sna 28 have been previously shown to differentiate lake trout and brook trout (Boeberitz et al., 2020) however these did not contain large LD blocks. Sna 28 does have a small LD block near the end of the chromosome (Figure 6) and perhaps should be further studied to examine if it is differentiated between species. To my knowledge, potential inversions discovered on Sna 9, 14, and 19 in brook trout are new and perhaps differentiated from lake trout as these chromosomes are not known to hold inversions among lake trout (Boeberitz et al., 2020). This further adds to the growing research on inversions among salmonids in uncovering adaptive divergence and speciation events.

In addition to inversion discovery, our findings also provide insight into gene function inside potential inversions. Two hundred and fifty seven genes were identified within the four potential inversions and gene annotations revealed some key pathways such as platelet activation on chromosome 9 which is known to have a role in immune response (Yun et al., 2016). This is particularly interesting because previous research on cutthroat trout found SNPs responsible for thermal tolerance to be linked to functions in immune response (Amish et al., 2019). Given that the streams where the potential inversions were found have higher maximum daily water temperatures (Figure 16) it seems probable that this could be an adaptive strategy with regards to thermal tolerance such as was described for cutthroat trout. Recently a study on redband trout which occupy high elevation forest and low-elevation deserts found, using lcWGR, that the cerk gene that has a role in cardiovascular systems could be significant in thermal adaptation (Z. Chen & Narum, 2021). This gene has also been linked to thermal adaptation in rainbow trout because of its role in metabolic and wound-healing rates (Willoughby et al., 2018). Particularly interesting in the current study is the finding of the cerkl gene among the potential inversion on chromosome 19. Cerkl protein is related to cerk and they both have a role in metabolism of signalling lipids, however cerkl does not phosphorylate ceramide (Bornancin et al., 2005; Chalfant & Spiegel, 2005). The role of cerk in stress related signalling pathways suggests it might be involved in cardiovascular function from reduced oxygen at higher water temperatures and heat shock response (Z. Chen & Narum, 2021). Cerkl plays a role in stress-induced apoptosis and is known to be regulated

during oxidative stress such as hypoxia where overexpression protects cells from apoptosis as found in a study looking at its role in retinal degeneration (Tuson et al., 2009). While this has been studied with regards to its specific connection to retinal degeneration (J. Chen et al., 2015; Tuson et al., 2009), we do not know if it has a similar role in thermal adaptation among other cell types but further research should be conducted to understand this potential relationship.

On chromosome 24 I found genes that are connected to cellular lipid catabolic processes effecting the breakdown of lipids. This has potential significance because lipids are used for maturation and migration in Atlantic salmon (Mobley et al., 2021). Perhaps brook trout utilize a similar pathway and are growing faster as a response to the increased temperature. This was seen in a study on rainbow trout where resident trout grew faster in warmer streams and those in cooler streams showed higher lipid content (McMillan et al., 2012). Also on this chromosome are processes linked to muscle contraction which could be connected to temperature tolerance via cardiovascular function, swimming, and foraging. Of particular interest is muscle contraction with regards to swimming and foraging and could be indirectly affected by streamflow. Rainbow trout were studied in two Californian coastal streams that differ in streamflow, temperature and stream cover where those in the warmer, lower flow reached their growth potential earlier due to adjustments in their foraging behaviour (Rossi et al., 2022). There is not a clear pattern in streamflow among eastern streams studied, but western streams show lower flow than the eastern streams (Figure S14) and this could be important for adaptation with regards to muscle contraction if the brook trout need to adapt from temperature stress, and both food and habitat availability. Interestingly, SNPs within the dystrobrevin alpha gene that has a role in muscle contraction were found to be adaptive among brook trout in Québec (Ferchaud et al., 2020). This gene is found on Dolly Varden chromosome 19 which corresponds to lake trout chromosome 25. That other genes involved in muscle contraction among salmonids have been linked to local adaptation provides further support that it could be important in the current study. It is important to remember that although there are genes located within these potential inversions that indicate they might have some kind of adaptive advantage, we cannot rule out the possibility that some are

not adaptive, but simply along for the ride in the potential inversions where recombination is low to none.

Brook trout are the number one sport fish in Nova Scotia and remain vulnerable to habitat loss via climate change. In Nova Scotia brook trout are stocked to maintain populations for harvest however stocking programs do not incorporate genomic structural variants into determining the best stock for a given environment. The discovery of chromosomal inversions and their role in local adaptation is an exploding avenue of research because of the decreased cost in whole genome sequencing. Given that here I discover potential inversions that have potential adaptative significance to temperature tolerance and that other salmonid research has found supergenes responsible for phenotypic differences, it seems that we should incorporate these advantageous structural variants into conservation of species such as brook trout. I did not collect any phenotypic data in the current study such as body size and size at age, and this would be helpful in future research to try and find phenotypic differences that might also link to genes located in these potential inversion regions. While I discovered four potential inversions among these brook trout populations using indirect detection, more research could be done to further confirm that these are inversions and that the potential inversions are themselves adaptive through a more direct detection approach such as long-read sequencing and genetic mapping (Mérot et al., 2020). With long-read sequencing we would be able to find breakpoints to further understand the role of potential inversions. Breakpoints generate mutations that can cause deletions which can be independent of reduced recombination from inversions and be a source of natural selection (Stenløkk et al., 2022; Villoutreix et al., 2021). An environmental analysis should also be conducted to provide more support for the role of potential local adaptation with regards to temperature and surficial geology type.

References

Amish, S. J., Ali, O., Peacock, M., Miller, M., Robinson, M., Smith, S., Luikart, G., & Neville, H. (2019). Assessing thermal adaptation using family-based association and FST outlier tests in a threatened trout species. *Molecular Ecology*, *28*(10), 2573–2593. https://doi.org/10.1111/mec.15100

Andrews, K. R., Seaborn, T., Egan, J. P., Fagnan, M. W., New, D. D., Chen, Z., Hohenlohe, P. A., Waits, L. P., Caudill, C. C., & Narum, S. R. (2023). Whole genome resequencing identifies local adaptation associated with environmental variation for redband trout. *Molecular Ecology*, *32*(4), 800–818. https://doi.org/10.1111/mec.16810

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arostegui, M. C., Quinn, T. P., Seeb, L. W., Seeb, J. E., & McKinney, G. J. (2019). Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. *Molecular Ecology*, *28*(6), 1412–1427. https://doi.org/10.1111/mec.15037

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. (2015). Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. *PLOS ONE*, *10*(5), e0128036. https://doi.org/10.1371/journal.pone.0128036

Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLOS Genetics*, *17*(3), e1009411. https://doi.org/10.1371/journal.pgen.1009411

Bernatchez, L., & Wilson, C. C. (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, *7*(4), 431–452. https://doi.org/10.1046/j.1365-294x.1998.00319.x

Black, G. A., Dempson, J. B., & Bruce, W. J. (1986). Distribution and postglacial dispersal of freshwater fishes of Labrador. *Canadian Journal of Zoology*, *64*(1), 21–31. https://doi.org/10.1139/z86-005

Blackman, R. C., Altermatt, F., Foulquier, A., Lefébure, T., Gauthier, M., Bouchez, A., Stubbington, R., Weigand, A. M., Leese, F., & Datry, T. (2021). Unlocking our understanding of intermittent rivers and ephemeral streams with genomic tools. *Frontiers in Ecology and the Environment*, *19*(10), 574–583. https://doi.org/10.1002/fee.2404

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*, *16*(9), 1195–1205. https://doi.org/10.1111/ele.12150

Boeberitz, O., Luikart, G., & Scribner, K. T. (2020). *Mapping of Adaptive Traits Enabled by a High-Density Linkage Map for Lake Trout*. *10*(June), 1929–1947.

Bohling, J. (2020). Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets. *Ecology and Evolution*, *10*(14), 7585–7601. https://doi.org/10.1002/ece3.6483

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

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

Bornancin, F., Mechtcheriakova, D., Stora, S., Graf, C., Wlachos, A., Dévay, P., Urtz, N., Baumruker, T., & Billich, A. (2005). Characterization of a ceramide kinase-like protein. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1687*(1), 31–43. https://doi.org/10.1016/j.bbalip.2004.11.012

Brookes, B., Jeon, H., Derry, A. M., Post, J. R., Rogers, S. M., Humphries, S., & Fraser, D. J. (2022). Neutral and adaptive drivers of genomic change in introduced brook trout (Salvelinus fontinalis) populations revealed by pooled sequencing. *Ecology and Evolution*, *12*(2), 1–14. https://doi.org/10.1002/ece3.8584

Brown, C. (2018). *na.tools: Comprehensive Library for Working with Missing (NA) Values in Vectors*. https://CRAN.R-project.org/package=na.tools

Chalfant, C. E., & Spiegel, S. (2005). Sphingosine 1-phosphate and ceramide 1-phosphate: Expanding roles in cell signaling. *Journal of Cell Science*, *118*(20), 4605–4612. https://doi.org/10.1242/jcs.02637

Chen, J., Liu, F., Li, H., Archacki, S., Gao, M., Liu, Y., Liao, S., Huang, M., Wang, J., Yu, S., Li, C., Tang, Z., & Liu, M. (2015). PVHL interacts with Ceramide kinase like (CERKL) protein and ubiquitinates it for oxygen dependent proteasomal degradation. *Cellular Signalling*, *27*(11), 2314–2323. https://doi.org/10.1016/j.cellsig.2015.08.011

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Z., & Narum, S. R. (2021). Whole genome resequencing reveals genomic regions associated with thermal adaptation in redband trout. *Molecular Ecology*, *30*(1), 162–174. https://doi.org/10.1111/mec.15717

Cherry, D. S., Dickson, K. L., Cairns Jr., J., & Stauffer, J. R. (1977). Preferred, Avoided, and Lethal Temperatures of Fish During Rising Temperature Conditions. *Journal of the Fisheries Research Board of Canada*, *34*(2), 239–246. https://doi.org/10.1139/f77-035

Cohen, Z. P., Schoville, S. D., & Hawthorne, D. J. (2023). The role of structural variants in pest adaptation and genome evolution of the Colorado potato beetle, Leptinotarsa decemlineata (Say). *Molecular Ecology*, *32*(6), 1425–1440. https://doi.org/10.1111/mec.16838

Connallon, T., & Olito, C. (2022). Natural selection and the distribution of chromosomal inversion lengths. *Molecular Ecology*, *31*(13), 3627–3641. https://doi.org/10.1111/mec.16091

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

Danzmann, R. G., Morgan II, R. P., Jones, M. W., Bernatchez, L., & Ihssen, P. E. (1998). A major sextet of mitochondrial DNA phylogenetic assemblages extant in eastern North American brook trout (Salvelinus fontinalis): Distribution and postglacial dispersal patterns. *Canadian Journal of Zoology*, *76*(7), 1300–1318. https://doi.org/10.1139/z98-056

Dowle, M., & Srinivasan, A. (2021). *data.table: Extension of `data.frame`*. https://CRAN.R-project.org/package=data.table

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., Rafajlović, M., Panova, M., Ravinet, M., Johannesson, K., Westram, A. M., & Butlin, R. K. (2019). Multiple chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. *Molecular Ecology*, *28*(6), 1375–1393. https://doi.org/10.1111/mec.14972

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving Inversions. *Trends in Ecology & Evolution*, *34*(3), 239–248. https://doi.org/10.1016/j.tree.2018.12.005

Ferchaud, A. L., Leitwein, M., Laporte, M., Boivin-Delisle, D., Bougas, B., Hernandez, C., Normandeau, É., Thibault, I., & Bernatchez, L. (2020). Adaptive and maladaptive genetic diversity in small populations: Insights from the Brook Charr (Salvelinus fontinalis) case study. *Molecular Ecology*, *June 2019*, 1–17. https://doi.org/10.1111/mec.15566

Fox, E. A., Wright, A. E., Fumagalli, M., & Vieira, F. G. (2019). ngsLD: Evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, *35*(19), 3855–3856. https://doi.org/10.1093/bioinformatics/btz200

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, *26*(20), 5369–5406. https://doi.org/10.1111/mec.14264

Gaio, D., Anantanawat, K., To, J., Liu, M., Monahan, L., & Darling, A. E. (2022). Hackflex: Low-cost, high-throughput, Illumina Nextera Flex library construction. *Microbial Genomics*, *8*(1), 000744. https://doi.org/10.1099/mgen.0.000744

Guerrero, R. F., Rousset, F., & Kirkpatrick, M. (2012). Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 430–438. https://doi.org/10.1098/rstb.2011.0246

Hale, M. C., Campbell, M. A., & McKinney, G. J. (2021). A candidate chromosome inversion in Arctic charr (Salvelinus alpinus) identified by population genetic analysis techniques. *G3: Genes, Genomes, Genetics*, *11*(10). https://doi.org/10.1093/g3journal/jkab267

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution*, 1–15. https://doi.org/10.1038/s41559-022-01890-0

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Huang, K., & Rieseberg, L. H. (2020). Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants. *Frontiers in Plant Science*, *11*(March), 1–13. https://doi.org/10.3389/fpls.2020.00296

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, *53*(3), Article 3. https://doi.org/10.1038/s41588-020-00771-1

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology*, *28*(11), 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072

Jones, M. W., Clay, D., & Danzmann, R. G. (1996). *Conservation genetics of brook trout (Salvelinus fontinalis): Population structuring in Fundy National Park, New Brunswick, and eastern Canada*. *53*, 16.

Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population scale DNA sequence data. *Genome Research*, gr.176552.114. https://doi.org/10.1101/gr.176552.114

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. *PLOS Biology*, *8*(9), e1000501. https://doi.org/10.1371/journal.pbio.1000501

Kontak, D. J. (2001). *Internal Stratigraphy of the Jurassic North Mountain Basalt , Southern Nova Scotia*. Minerals and Energy Branch Nova Scotia.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

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

Lamichhaney, S., & Andersson, L. (2019). A comparison of the association between large haplotype blocks under selection and the presence/absence of inversions. *Ecology and Evolution*, *9*(8), 4888–4896. https://doi.org/10.1002/ece3.5094

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lecaudey, L. A., Schliewen, U. K., Osinov, A. G., Taylor, E. B., Bernatchez, L., & Weiss, S. J. (2018). Molecular Phylogenetics and Evolution Inferring phylogenetic structure , hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution*, *124*(March), 82–99. https://doi.org/10.1016/j.ympev.2018.02.022

Lehnert, S. J., Baillie, S. M., MacMillan, J., Paterson, I. G., Buhariwalla, C. F., Bradbury, I. R., & Bentzen, P. (2020). Multiple decades of stocking has resulted in limited hatchery introgression in wild brook trout (Salvelinus fontinalis) populations of Nova Scotia. *Evolutionary Applications*, *13*(5), 1069–1089. https://doi.org/10.1111/eva.12923

Leigh, D. M., Lischer, H. E. L., Guillaume, F., Grossen, C., & Günther, T. (2021). Disentangling adaptation from drift in bottlenecked and reintroduced populations of Alpine ibex. *Molecular Ecology Resources*, *21*(7), 2350–2363. https://doi.org/10.1111/1755-0998.13442

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindtke, D., Lucek, K., Soria-Carrasco, V., Villoutreix, R., Farkas, T. E., Riesch, R., Dennis, S. R., Gompert, Z., & Nosil, P. (2017). Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology*, *26*(22), 6189–6205. https://doi.org/10.1111/mec.14280

Liu, Z., Roesti, M., Marques, D., Hiltbrunner, M., Saladin, V., & Peichel, C. L. (2022). Chromosomal Fusions Facilitate Adaptation to Divergent Environments in Threespine Stickleback. *Molecular Biology and Evolution*, *39*(2), msab358. https://doi.org/10.1093/molbev/msab358

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, *July*, 1–28. https://doi.org/10.1111/mec.16077

Lou, R. N., & Therkildsen, N. O. (2022). Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources*, *22*(5), 1678–1692. https://doi.org/10.1111/1755-0998.13559

Lundberg, M., Mackintosh, A., Petri, A., & Bensch, S. (2023). Inversions maintain differences between migratory phenotypes of a songbird. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-36167-y

MacMillan, J. L., Caissie, D., Marshall, T.J, & Hinks, L. (2008). *Population indices of brook trout (Salvelinus fontinalis), Atlantic salmon (Salmo salar), and salmonid competitors in relation to summer water temperature and habitat parameters in 100 streams in Nova Scotia*. Department of Fisheries and Oceans Gulf Region.

Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Brieuc, M. S. O., Pampoulie, C., Bradbury, I., Jakobsen, K. S., & Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology and Evolution*, *6*(4), 469–481. https://doi.org/10.1038/s41559-022-01661-x

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McMillan, J. R., Dunham, J. B., Reeves, G. H., Mills, J. S., & Jordan, C. E. (2012). Individual condition and stream temperature influence early maturation of rainbow and steelhead trout, Oncorhynchus mykiss. *Environmental Biology of Fishes*, *93*(3), 343–355. https://doi.org/10.1007/s10641-011-9921-0

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731. https://doi.org/10.1534/genetics.118.301336

Mérot, C., Berdan, E. L., Cayuela, H., Djambazian, H., Ferchaud, A. L., Laporte, M., Normandeau, E., Ragoussis, J., Wellenreuther, M., & Bernatchez, L. (2021). Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly. *Molecular Biology and Evolution*, *38*(9), 3953–3971. https://doi.org/10.1093/molbev/msab143

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends in Ecology & Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Mobley, K. B., Aykanat, T., Czorlich, Y., House, A., Kurko, J., Miettinen, A., Moustakas-Verho, J., Salgado, A., Sinclair-Waters, M., Verta, J.-P., & Primmer, C. R. (2021). Maturation in Atlantic salmon (Salmo salar, Salmonidae): A synthesis of ecological, genetic, and molecular processes. *Reviews in Fish Biology and Fisheries*, *31*(3), 523–571. https://doi.org/10.1007/s11160-021-09656-w

Navarro, A., Betrán, E., Barbadilla, A., & Ruiz, A. (1997). Recombination and Gene Flux Caused by Gene Conversion and Crossing Over in Inversion Heterokaryotypes. *Genetics*, *146*(2), 695–709. https://doi.org/10.1093/genetics/146.2.695

Neuwirth, E. (2022). *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer

Newton, R. M., Weintraub, J., & April, R. (1987). The relationship between surface water chemistry and geology in the North Branch of the Moose River. *Biogeochemistry*, *3*(1–3), 21–35. https://doi.org/10.1007/BF02185183

Nova Scotia Department of Agriculture and Fisheries Inland Fisheries Division. (2005). *Nova Scotia Trout Management Plan*. Nova Scotia Department of Agriculture and Fisheries.

O'Donnell, T. P., & Sullivan, T. J. (2021). Low-coverage whole-genome sequencing reveals molecular markers for spawning season and sex identification in Gulf of Maine Atlantic cod (Gadus morhua, Linnaeus 1758). *Ecology and Evolution*, *February*, 10659–10671. https://doi.org/10.1002/ece3.7878

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., Anderson, E. C., Rundio, D. E., Williams, T. H., Naish, K. A., Moen, T., Liu, S., Kent, M., Moser, M., Minkley, D. R., Rondeau, E. B., Brieuc, M. S. O., Sandve, S. R., Miller, M. R., … Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, *3*(12), Article 12. https://doi.org/10.1038/s41559-019-1044-6

Picard toolkit. (2019). In *Broad Institute, GitHub repository*. Broad Institute. https://broadinstitute.github.io/picard/

Pilgrim, B. L., Perry, R. C., Barron, J. L., & Marshall, H. D. (2012). Nucleotide variation in the mitochondrial genome provides evidence for dual routes of postglacial recolonization and genetic recombination in the northeastern brook trout (Salvelinus fontinalis). *Genetics and Molecular Research*, *11*(3), 3466–3481. https://doi.org/10.4238/2012.September.26.2

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Studio Team. (2021). *R Studio: Integrated Development Environment for R*. http://www.rstudio.com/

Richard, D. M., & Bradley, T. R. (1992). *Surficial Geology of the Province of Nova Scotia Map 92-3* [Map]. Nova Scotia Department of Natural Resources Mines and Energy Branches.

Rossi, G. J., Power, M. E., Carlson, S. M., & Grantham, T. E. (2022). Seasonal growth potential of Oncorhynchus mykiss in streams with contrasting prey phenology and streamflow. *Ecosphere*, *13*(9), e4211. https://doi.org/10.1002/ecs2.4211

Rougemont, Q., Xuereb, A., Dallaire, X., Moore, J.-S., Normandeau, E., Perreault-Payette, A., Bougas, B., Rondeau, E. B., Withler, R. E., Van Doornik, D. M., Crane, P. A., Naish, K. A., Garza, J. C., Beacham, T. D., Koop, B. F., & Bernatchez, L. (2023). Long-distance migration is a major factor driving local adaptation at continental scale in Coho salmon. *Molecular Ecology*, *32*(3), 542–559. https://doi.org/10.1111/mec.16339

Rousset, F. (1997). Genetic Differentiation and Estimation of Gene Flow from *F* -Statistics Under Isolation by Distance. *Genetics*, *145*(4), 1219–1228. https://doi.org/10.1093/genetics/145.4.1219

Ruzzante, D. E., McCracken, G. R., Parmelee, S., Hill, K., Corrigan, A., MacMillan, J., & Walde, S. J. (2016). Effective number of breeders, effective population size and their relationship with census size in an iteroparous species, Salvelinus fontinalis. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1823), 20152601. https://doi.org/10.1098/rspb.2015.2601

Sanchez-Donoso, I., Ravagni, S., Rodríguez-Teijeiro, J. D., Christmas, M. J., Huang, Y., Maldonado-Linares, A., Puigcerver, M., Jiménez-Blasco, I., Andrade, P., Gonçalves, D., Friis, G., Roig, I., Webster, M. T., Leonard, J. A., & Vilà, C. (2022). Massive genome inversion drives coexistence of divergent morphs in common quails. *Current Biology*, *32*(2), 462-469.e6. https://doi.org/10.1016/j.cub.2021.11.019

Shaw, J., Piper, D. J. W., Fader, G. B. J., King, E. L., Todd, B. J., Bell, T., Batterson, M. J., & Liverman, D. G. E. (2006). *A conceptual model of the deglaciation of Atlantic Canada*. *25*, 2059–2081. https://doi.org/10.1016/j.quascirev.2006.03.002

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics*, *195*(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Smith, S. R., Normandeau, E., Djambazian, H., Nawarathna, P. M., Berube, P., Muir, A. M., Ragoussis, J., Penney, C. M., Scribner, K. T., Luikart, G., Wilson, C. C., & Bernatchez, L. (2022). A chromosome-anchored genome assembly for Lake Trout (Salvelinus namaycush). *Molecular Ecology Resources*, *22*(2), 679–694. https://doi.org/10.1111/1755-0998.13483

Stea, R. R. (2004). The appalachian glacier complex in maritime canada. In J. Ehlers & P. L. Gibbard (Eds.), *Developments in Quaternary Sciences* (Vol. 2, pp. 213–232). Elsevier. https://doi.org/10.1016/S1571-0866(04)80199-4

Stea, R. R., Piper, D. J. W., Fader, G. B. J., & Boyd, R. (1998). Wisconsinan glacial and sea-level history of Maritime Canada and the adjacent continental shelf: A correlation of land and sea events. *Bulletin of the Geological Society of America*, *110*(7), 821–845. https://doi.org/10.1130/0016-7606(1998)110<0821:WGASLH>2.3.CO;2

Stenløkk, K., Saitou, M., Rud-Johansen, L., Nome, T., Moser, M., Árnyasi, M., Kent, M., Barson, N. J., & Lien, S. (2022). The emergence of supergenes from inversions in Atlantic salmon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1856), 20210195. https://doi.org/10.1098/rstb.2021.0195

Sturtevant, A. H. (1921). A Case of Rearrangement of Genes in Drosophila1. *Proceedings of the National Academy of Sciences*, *7*(8), 235–237. https://doi.org/10.1073/pnas.7.8.235

Sutherland, B. J. G., Gosselin, T., Normandeau, E., Lamothe, M., Isabel, N., Audet, C., & Bernatchez, L. (2016). Salmonid chromosome evolution as revealed by a novel method for comparing radseq linkagemaps. *Genome Biology and Evolution*, *8*(12), 3600–3617. https://doi.org/10.1093/gbe/evw262

Tavares, H. (2022). *windowscanr: Apply functions using sliding windows*.

Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Molecular Ecology Resources*, *17*(2), 194–208. https://doi.org/10.1111/1755-0998.12593

Tuson, M., Garanto, A., Gonzàlez-Duarte, R., & Marfany, G. (2009). Overexpression of CERKL, a gene responsible for retinitis pigmentosa in humans, protects cells from apoptosis induced by oxidative stress. *Molecular Vision*, *15*, 168–180.

Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution*, *69*(6), 1476–1486. https://doi.org/10.1111/evo.12663

United States Department of the Interior Bureau of Reclamation. (1997). *Water measurement manual*.

VanRaden, P. M., Sun, C., & O'Connell, J. R. (2015). Fast imputation using medium or low-coverage sequence data. *BMC Genetics*, *16*(1), 82. https://doi.org/10.1186/s12863-015-0243-7

Villoutreix, R., Ayala, D., Joron, M., Gompert, Z., Feder, J. L., & Nosil, P. (2021). Inversion breakpoints and the evolution of supergenes. *Molecular Ecology*, *30*(12), 2738–2755. https://doi.org/10.1111/mec.15907

Watson, K. B., Lehnert, S. J., Bentzen, P., Kess, T., Einfeldt, A., Duffy, S., Perriman, B., Lien, S., Kent, M., & Bradbury, I. R. (2022). Environmentally associated chromosomal structural variation influences fine-scale population structure of Atlantic Salmon (Salmo salar). *Molecular Ecology*, *31*(4), 1057–1075. https://doi.org/10.1111/mec.16307

Webster, T. L., Murphy, J. B., Gosse, J. C., & Spooner, I. (2006). The application of lidar-derived digital elevation model analysis to geological mapping: An example from the Fundy Basin, Nova Scotia, Canada. *Canadian Journal of Remote Sensing*, *32*(2), 173–193. https://doi.org/10.5589/m06-017

Wellband, K., Mérot, C., Linnansaari, T., Elliott, J. a. K., Curry, R. A., & Bernatchez, L. (2019). Chromosomal fusion and life history-associated genomic variation contribute to within-river local adaptation of Atlantic salmon. *Molecular Ecology*, *28*(6), 1439–1459. https://doi.org/10.1111/mec.14965

Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends in Ecology & Evolution*, *33*(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Girlich, M. (2022). *tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr

Wilder, A. P., Palumbi, S. R., Conover, D. O., & Therkildsen, N. O. (2020). Footprints of local adaptation span hundreds of linked genes in the Atlantic silverside genome. *Evolution Letters*, *4*(5), 430–443. https://doi.org/10.1002/evl3.189

Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T., & Christie, M. R. (2018). Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology*, *27*(20), 4041–4051. https://doi.org/10.1111/mec.14726

Wood, J. L. A., Belmar-Lucero, S., Hutchings, J. A., & Fraser, D. J. (2014). Relationship of habitat variability to population size in a stream fish. *Ecological Applications*, *24*(5), 1085–1100. https://doi.org/10.1890/13-1647.1

Wood, J. L. A., & Fraser, D. J. (2015). Similar plastic responses to elevated temperature among different-sized brook trout populations. *Ecology*, *96*(4), 1010–1019. https://doi.org/10.1890/14-1378.1

Wood, J. L. A., Yates, M. C., & Fraser, D. J. (2016). Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications*, *9*(5), 640–657. https://doi.org/10.1111/eva.12375

Xie, Y. (2015). *Dynamic documents with R and Knitr* (Second edition). CRC Press/Taylor & Francis.

Yu, H. (2002). Rmpi: Parallel Statistical Computing in R. *R News*, *2*(2), 10–14.

Yun, S.-H., Sim, E.-H., Goh, R.-Y., Park, J.-I., & Han, J.-Y. (2016). Platelet Activation: The Mechanisms and Potential Biomarkers. *BioMed Research International*, *2016*, 9060143. https://doi.org/10.1155/2016/9060143

Zastavniouk, C., Weir, L. K., & Fraser, D. J. (2017). The evolutionary consequences of habitat fragmentation: Body morphology and coloration differentiation among brook trout populations of varying size. *Ecology and Evolution*, *7*(17), 6850–6862. https://doi.org/10.1002/ece3.3229

Zhang, W., Tan, C., Hu, H., Pan, R., Xiao, Y., Ouyang, K., Zhou, G., Jia, Y., Zhang, X.-Q., Hill, C. B., Wang, P., Chapman, B., Han, Y., Xu, L., Xu, Y., Angessa, T., Luo, H., Westcott, S., Sharma, D., … Li, C. (2023). Genome architecture and diverged selection shaping pattern of genomic differentiation in wild barley. *Plant Biotechnology Journal*, *21*(1), 46–62. https://doi.org/10.1111/pbi.13917

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

Appendix

DNA Extraction And Quantification

For DNA extraction, a small ~10mg piece of tissue was placed in each 1.5mL tube using flame sterilized forceps. Tissue was left to dry for ~1.5 hours to let any remaining ethanol evaporate off before adding proteinase K and TL buffer. Samples were placed in a 55˚C incubator overnight before completing the DNA isolation and wash steps. DNA was eluted in 100µL of elution buffer and run on a 1.5% agarose gel to confirm quality and quantity. 15µL of GelGreen (Biotium) was added to each 1.5% gel to stain DNA and a 1kb DNA ladder (Biotium) was used to determine size of extracted DNA. 5µL of ladder was loaded into one well on each gel and 2µL of each sample was mixed with 2µL loading dye (New England Biolabs) before being pipetted into each well.

For DNA quantification, 50µL of PicoGreen MasterMix was added to each well of the microplate followed by 49µL of 1X TE buffer and 1µL of extracted DNA. Standards used for the assay were 2ng, 5ng, 10ng, and 20ng representing the quantity range of brook trout samples and volumes of 1X TE and standard added to the plate varied depending on the concentration of standard. All samples were diluted to 1.6 ng/µL using ddH_2O for library preparation.

Data Filtering And Reference Genome Mapping

The processing of raw data was done following the Therkildsen lab GitHub https://github.com/therkildsen-lab/genomic-data-analysis as a general guide. The specific parameters that I used are described below. FastQC v.0.11.9 (S. Andrews, 2010) and MultiQC v.1.11 (Ewels et al., 2016) were used on demultiplexed, raw, and barcode removed fastq files to examine quality of reads and test for the presence of adaptor sequences. Trimmomatic v.0.39 (Bolger et al., 2014) was used to remove i5 and i7 IDT Illumina and Illumina Nextera adapters with the following parameters: paired end mode with seed mismatches 2, palindrome clip threshold 30, simple clip threshold 1, and minimum length of 40. The two-colour chemistry used by the Illumina NovaSeq can lead to the presence of poly-G tails. Fastp v.0.23.1 (Chen et al., 2018) was used on the paired, adaptor trimmed fastq.gz files to remove poly-G tails with the –cut_right option using –cut_right_window_size 4, and –cut_right_mean_quality 20. For genome mapping to the

lake trout reference genome, the unplaced scaffold was not used for alignment to reduce the effects of reads mapping in duplicate locations. End to end mode in Bowtie2 with –sensitive were used to improve percentage of reads mapping to account for the use of a closely related species reference genome and the inter-mate distance for mapping concordant reads was set from 0-1500bp. For indel realignment, GATKs RealignerTargetCreator function was used to make an intervals file with locations of indels and potential indels which the IndelRealigner function uses as input. Reads with minimum base quality < 20 and minimum mapping quality < 20 were removed for depth calculations using SAMtools v.1.13 (Li et al., 2009) The resulting depth files were analyzed in R v.4.1.2 (RStudio Team, 2021) using tidyverse v.1.3.1 (Wickham et al., 2019), ggplot2 v.3.3.5 (Wickham, 2016), Rmpi v.0.6-9.2 (Yu, 2002), dplyr v.1.0.7 (Wickham et al., 2022), knitr v.1.36 (Xie, 2015), and RColorBrewer v.1.1.2 (Neuwirth, 2022).

Depth And Genotype Likelihood Parameters For ANGSD

The -doDepth and -doCounts parameters of ANGSD v.0.933 (Korneliussen et al., 2014) were used for depth calculation and -maxDepth was used to set an initial maximum depth of 10x per individual. -minMapQ 20, -minQ 20, -remove_bads 1, -only_proper_pairs 1 and -C 50 were set to filter out reads with minimum mapping and minimum quality below 20 and to keep proper pairs. Briefly, the distribution of depth counts from the .depthGlobal files were plotted as histograms using R v.4.1.2 (RStudio Team, 2021) with tidyr v.1.2.0 (Wickham & Girlich, 2022), data.table v.1.14.2 (Dowle & Srinivasan, 2021), ggplot2 v.3.3.6 (Wickham, 2016), na.tools v.0.3.1 (Brown, 2018), and dyplr v.1.0.9 (Wickham et al., 2022). Data was filtered based on the normal distribution and the mean and standard deviation was calculated. The maximum and minimum depth were then calculated as the mean + and – the standard deviation. This was done for each stream and for all streams combined. ANGSD v.0.933 was then used to calculate genotype likelihoods (-GL), the major and minor alleles were determined (-doMajorMinor 1), and minor allele frequency estimated (-doMaf 1). Genotype likelihoods were calculated with SAMtools (-GL 1) and GATK (-GL 2) for comparison in number of SNPs discovered. -setMinDepth and -setMaxDepth were set according to the .depthGlobal data from the previous step of ANGSD -doDepth and -doCounts. For genotype likelihoods, other

ANGSD parameters included were -uniqueOnly 1, -remove_bads 1, -only_proper_pairs 1, -trim 0, -C 50, -baq 1, -minInd 10, -minMapQ 20, -minQ 20, -doIBS 1, -makematrix 1, -doCov 1, -SNP_pval 1e-6, and -minMaf 0.05. The -SNP_pval was set at 10^{-6} for calling polymorphic loci and -minMaf was used to filter out SNPs with minor allele frequency below 5%. After determining the number of SNPs, genotype likelihoods were calculated again using the SNP list generated from the previous step in discovering SNPs with the following parameters -doMajorMinor 3 as the SNP list, -rf as the chromosome list, -GL 1, -doMaf 1 where the major and minor allele were determined from the genotype likelihoods given the SNP list, and -doIBS 1, -doCounts1, -doCov 1, and -makematrix 1.

Population Genomics

Parameters for ANGSD to generate SAFs were as follows -ref, -anc, -uniqueOnly 1, -remove_bads 1, -only_proper_pairs 1, -trim 0, -C 50, -minMapQ 20, -minQ 20, -minInd 10, -setMinDepth, -setMaxDepthm -doCounts 1. For SAFs based on geology, -minInd 30 was used because there are three populations in each group and 30 was roughly half of the individuals. Minimum and maximum depth were set based on the previously described depth method used in ANGSD to determine these thresholds.

Linkage Disequilibrium And Decay

A position file was generated from ANGSD's minor allele frequency file (.mafs.gz) with chromosome and position in the genome of all the identified SNPs required for LD calculation. The LD decay rate was unknown for brook trout and determined using the fit_LDdecay.R script of ngsLD so that an appropriate maximum distance in kb between SNPs could be used for LD calculation. Prior to determining the decay rate, LD was calculated with a much larger max_kb_dist than likely which was used to determine LD decay. ngsLD was run with the following parameters --probs, --n_ind 192, --n_sites 1745248, --max_kb_dist 100. The .ld file was compressed and supplied as input for fit_LDdecay.R with the following parameters –ld r2, --n_ind 192, --fit_level 100, --fit_boot 100, --max_kb_dist 200, --fit_bin_size 1000, --plot_data, --plot_scale 3 (Figure A1).

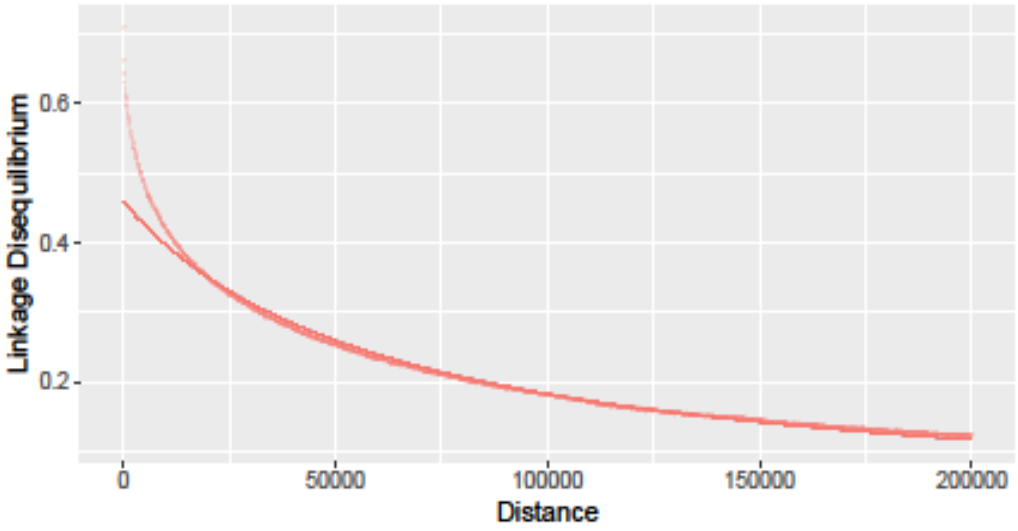

Figure A- 1 LD decay plot produced with fit_LDdecay.R of ngsLD where –max_kb_dist for LD calculation (.ld file supplied to fit_LDdecay.R) was 200kb and –max_kb_dist for fit_LDdecay.R was 200kb. LD represents the r^2 values for LD calculations at each position and distance is displayed in bp.

Linkage Disequilibrium Pruning

The pruned .ld files were used to generate a LD pruned SNP list and used as input for ANGSD to calculate genotype likelihoods and minor allele frequencies using the following parameters -GL 1, -doGlf 2, -doMajorMinor 3, -doMaf 1, -doPost 1, -doIBS 1, -doCounts 1, -doCov 1, -makeMatrix 1.

Linkage Disequilibrium By Chromosome And Testing Of Different Populations

Optimization Of LD Calculation And Percentile Plotting

Individual chromosome minor allele frequency and genotype likelihood files discussed in the methods were used to calculate LD by chromosome with ngsLD v.1.1.1 on all 192 brook trout. The following parameters were used in ngsLD, --posH, --probs, --n_ind, --n_sites (corresponding to the number of SNPs in each .maf.gz file), --max_kb_dist 0, --max_snp_dist 0, --rnd_sample 0.5. The –rnd_sample parameter was used to randomly sample 50% of the frequency and likelihood file comparisons to reduce the time and size of files produced. –max_kb_dist and –max_snp_dist were set to 0 so that LD was calculated across the entire chromosome with no threshold of distance to calculate LD. A maf of 10% and all SNPs were used to test the impact on heatmaps of LD. These

parameters had small or no effect on the results. The original genotype likelihoods generated using maf 5% and randomly sampling 50% of SNPs were used for following analyses. R^2 values were summarized into percentiles using scripts from Mérot et al. (2021) and plotted in R v.4.0.2 (RStudio Team, 2021) using ggplot2 v.3.3.6 (Wickham, 2016). The second percentile of r^2 values was plotted in windows of 250kb for heatmaps using all 192 individuals. Different window sizes and percentiles were tested, and window size had no or very minimal impact on the resulting heatmaps. The second percentile was chosen to be more conservative as the first percentile and higher showed the highest LD, generating larger blocks which could over represent potential inversions.

Exploring Pattern Of LD Blocks

Chromosomes (N=30) with blocks of LD (r^2 above ~0.6) were further analyzed using PCA to look for patterns of inversions in all individuals (3 groups representing homozygotes for the inversion, homozygotes without the inversion, and heterozygotes with one copy of the inversion). SNPs within LD blocks were used to calculate genotype likelihoods using ANGSD v.0.933 with the following parameters -ref, -GL 1, -doGlf 2, -doMaf 1, -doMajorMinor 3, -sites, -rf, -doIBS 1, -doCounts 1, -doCov 1, -makeMatrix 1. PCA was performed in R Studio v.4.0.2 with the eigen function and plotted with ggplot2 v.3.3.6. PCA for a subset of chromosomes NC_052315.1 and NC_052330.1 (largest LD blocks) suggested that streams located most West along the North Mountain were distinct from all other streams (Figures S3 and S6). This resulted in PCA testing using six populations without the Western streams (Sheep Shearer, Healeys, Poole) to determine if three groups would be visible. This was done because we initially thought the distinct Western streams were challenging our ability to see three groups (though still visible) representative of a potential inversion. PCA of LD blocks for some chromosomes did not show the expected three groups and we thought this might be because the LD block was two separate blocks or a smaller block. LD calculated in smaller blocks for some chromosomes made groups clearer in PCA and in general breaking one LD block in two did not change the pattern seen in PCA.

Chromosome 9, 14, 19, 24 (NC_052315.1, NC_052320.1, NC_052325.1, NC_052330.1) were used to test patterns in LD heatmaps based on sample size and grouping of

populations. Heatmaps were made for chromosomes 9, 14, 19, and 24 using all populations, each population individually, and as groups of three populations based on geology (scoured glacial till, thick glacial till, silt) (Figures 8-11). LD was calculated as described above using ngsLD v.1.1.1. The fifth percentile was used for plotting heatmaps of individual streams because LD was higher across the chromosome which could be over representative of the LD block thus the fifth percentile was used. The second percentile was used for all individuals together and grouping by geology.

Heterozygosity And Nucleotide Diversity/Divergence Among Inversion Polymorphisms

Allele frequencies of major and minor alleles across chromosomes 9, 14, 19, and 24 for individuals representing the 3 different groups of a potential inversion were calculated with ANGSD v.0.933 using -doHWE 1, -GL 1, -minMapQ 20, -minQ 20, -remove_bads 1, -doMajorMinor 3, and -sites corresponding to the SNP list for each chromosome. Observed and expected proportions of heterozygotes and homozygotes were calculated using the following script https://github.com/clairemerot/angsd_pipeline/blob/master/01_scripts/Rscripts/Hobs_sliding.r (Mérot, 2021) and the R package WindowScanr v.0.1 (Tavares, 2022) was used to calculate average observed proportion of heterozygotes (H_{obs}) in 10kb windows with a 10kb step. Hardy-Weinberg observed and expected frequencies per SNP position were used to calculate H_{obs}. Average H_{obs} across each chromosome was plotted in R v.4.0.2 (RStudio Team, 2021) using ggplot2 v.3.3.6 (Wickham, 2016). For 3 of the 4 chromosomes, the size of the potential inversion could be more accurately estimated from these plots compared to the LD heatmaps and the SNP lists were adjusted accordingly. H_{obs} was also calculated within potential inversion regions using windows the size of potential inversions and plotted as boxplots in R v.4.0.2 (RStudio Team, 2021) using ggplot2 v.3.3.6 (Wickham, 2016). Results of tests for Hardy-Weinberg equilibrium using -doHWE mentioned above indicate that on all four chromosomes the potentially non-inverted homokaryotypes have SNPs that deviate significantly from Hardy-Weinberg after Bonferroni correction (Table A-2, A-3, A-4, A-5). That there is not a large percentage of SNPs that deviate significantly suggests that there is not a deficit of

homokaryotypes. Potential inverted homokaryotype groups have 5 or fewer individuals and it is expected there would be more departures from Hardy-Weinberg, however no SNPs showed significant deviations (Table A-3, A-4, A-5, A-6). It is important to note that within groups there are individuals from different populations which will affect Hardy-Weinberg frequencies.

Prior to calculating diversity, minimum and maximum depth were determined using the same individuals from H_{obs} representing the 3 inversion groups for chromosomes 9, 14, 19, 24. ANGSD v.0.933 was used to get global depth files for each individual group using -doDepth 1, -maxDepth (based on the number of individuals in each group multiplied by 10), -doCounts 1, -dumpCounts 1, -minMapQ 20, -minQ 20, -remove_bads 1, -only_proper_pairs 1, -C 50. The global depth files were plotted in R v.4.1.2 (RStudio Team, 2021) following the same procedure described above in the depth and genotype likelihood parameters section. ANGSD v.0.933 was used to calculate SAFs for each of the inversion groups using the following parameters -doSaf 1, -GL 1, -doMajorMinor 1, -doCounts 1, -anc as the reference genome, -ref as the reference genome, -minMapQ 20, -minQ 20, -remove_bads 1, -uniqueOnly 1, -only_proper_pairs 1, -trim 0, -C 50, -minInd (set as half the number of individuals in each bamlist), -setMinDepth, and -setMaxDepth. RealSFS in ANGSD was used to get the folded site frequency spectrum from the SAF files. It is folded because the lake trout reference genome was used as the ancestral and reference for -doSaf. RealSFS saf2theta function was used to convert each SAF and folded SFS to a file of thetas. The ThetaStat function in ANGSD was used to get thetas in sliding windows of 10kb with a 10kb step and the pairwise theta (tP) and middle positions in each thetas.windows.gzpestPG was plotted as boxplots in R v.4.0.2 (RStudio Team, 2021) using ggplot2 v.3.3.6 (Wickham, 2016). tP was used because it gives more weight to alleles at intermediate frequencies. To help with the bias between sites with different number of SNPs, windowed tP values were divided by the number of SNPs in the window. For boxplots, data was subset to include only regions inside potential inversions and outside inversions on chromosomes 9, 14, 19, and 24 using sizes adjusted after doing H_{obs}.

ANGSD v.0.933 was used to calculate minor allele frequencies on the inverted and non-inverted homozygote individuals used above for H_{obs} and diversity with the following parameters -GL 1, -doMaf 1, -doGlf 2, -doMajorMinor 3, -sites (SNP list for inside potential inversions), -rf (chromosome name), and -ref (lake trout reference genome).

Nucleotide divergence is calculated from minor allele frequencies and ANGSD v.0.933 was used supplying the same individuals used above for H_{obs} and diversity. Nucleotide divergence is calculated from minor allele frequencies at each site by multiplying minor allele frequency by 1 minus the minor allele frequency. The script used is from https://github.com/mfumagalli/ngsPopGen/blob/master/scripts/calcDxy.R and was modified to only calculate divergence. The R package WindowScanR was used to take divergence values and calculate average divergence in windows of 10kb with a slide of 10kb. Boxplots of divergence for homozygotes with and without inversions and inside potential inversion regions were made with R v.4.0.2 (RStudio Team, 2021) using ggplot2 v.3.3.6 (Wickham, 2016).

Streamflow

Table A- 1 Streamflow measurements in cubic meters per second (cms) taken once a month at the same spot in each stream using the float method as a velocity area measurement of streamflow. See methods for description of how measurements were taken.

Stream	Streamflow (cfs) July 14 2021	Streamflow (cfs) August 9 2021	Streamflow (cfs) September 13 2021	Streamflow (cfs) October 6 2021	Streamflow (cfs) November 17 2021
Ross Creek	0.042	0.138	0.161	0.045	0.083
Woodworth Creek	0.039	0.096	0.164	0.047	0.105
Black Hole Brook	0.020	0.063	0.078	0.022	0.077
Church Vault Brook	0.046	0.128	0.088	0.071	0.074
Saunders Brook	0.087	0.129	0.142	0.076	0.114
Robinson Brook	0.033	0.101	0.291	0.103	0.238
Sheep Shearer Brook	0	0.071	0.027	0.007	0.087
Healeys Brook	0	0.039	0.036	0.008	0.065
Poole	0	0.031	0.050	0.016	0.112

Table A- 2 Average pH calculated from pH measured in the middle and at both banks of each stream where streamflow was measured using a pHep^{+} H198108 (Hanna instruments) once a month.

Stream	Average pH July 14 2021	Average pH August 9 2021	Average pH September 13 2021	Average pH October 6 2021	Average pH November 17 2021
Ross Creek	7.60	7.79	7.73	7.67	7.62
Woodworth Creek	7.78	7.87	7.84	7.65	7.56
Black Hole Brook	7.74	7.83	7.92	7.80	7.64
Church Vault Brook	8.00	8.10	7.98	8.01	7.70
Saunders Brook	8.09	8.16	8.00	8.08	7.82
Robinson Brook	7.88	7.81	7.73	7.86	7.86
Sheep Shearer Brook	7.84	7.51	7.50	7.44	7.70
Healeys Brook	7.42	7.55	7.47	7.36	7.77
Poole	7.59	7.50	7.43	7.17	7.77

Table A- 3 Number of SNPs that differ from Hardy-Weinberg and Bonferroni corrected values for Chromosome 9. Hardy-Weinberg deviations were estimated for each group (inverted homokaryotypes, heterokaryotypes, non-inverted homokaryotypes) within potential inversion regions. Individuals included in each of these groups were those identified in the PCA of the LD block for chromosome 9 (Figure 12) used to identify the 3 different karyotypes groups. A p-value of 0.0001 with a Bonferroni correction for number of tests equal to the number of SNPs (~21,000) within the potential inversion region was used to identify SNPs deviating from HWE. There are 5 potential inverted homokaryotypes, 23 potential heterokaryotypes, and 140 potential non-inverted homokaryotypes.

Chromosome 9	# of SNPs out of HWE (P<0.0001)	# of SNPs out of HWE after Bonferroni correction
Inverted homokaryotypes	0	0
Heterokaryotypes	1701	0
Non-inverted homokaryotypes	1165	213

Table A- 4 Number of SNPs that differ from Hardy-Weinberg and Bonferroni corrected values for Chromosome 14. Hardy-Weinberg deviations were estimated for each group (inverted homokaryotypes, heterokaryotypes, non-inverted homokaryotypes) within potential inversion regions. Individuals included in each of these groups were those identified in the PCA of the LD block for chromosome 14 (Figure 12) representing the 3 different karyotypes groups. A p-value of 0.0001 with a Bonferroni correction for number of tests equal to the number of SNPs (~21,000) within the potential inversion region was used to identify SNPs deviating from HWE. There are 4 potential inverted homokaryotypes, 9 potential heterokaryotypes, and 163 potential non-inverted homokaryotypes.

Chromosome 14	# of SNPs out of HWE (P<0.0001)	# of SNPs out of HWE after Bonferroni correction
Inverted homokaryotypes	0	0
Heterokaryotypes	0	0
Non-inverted homokaryotypes	1849	383

Table A- 5 Number of SNPs that differ from Hardy-Weinberg and Bonferroni corrected values for Chromosome 19. Hardy-Weinberg deviations were estimated for each group (inverted homokaryotypes, heterokaryotypes, non-inverted homokaryotypes) within potential inversion regions. Individuals included in each of these groups were those identified in the PCA of the LD block for chromosome 19 (Figure 12) representing the 3 different karyotypes groups. A p-value of 0.0001 with a Bonferroni correction for number of tests equal to the number of SNPs (~20,000) within the potential inversion region was used to identify SNPs deviating from HWE. There are 7 potential inverted homokaryotypes, 12 potential heterokaryotypes, and 169 potential non-inverted homokaryotypes.

Chromosome 19	# of SNPs out of HWE (P<0.0001)	# of SNPs out of HWE after Bonferroni correction
Inverted homokaryotypes	0	0
Heterokaryotypes	156	0
Non-inverted homokaryotypes	1916	319

Table A- 6 Number of SNPs that differ from Hardy-Weinberg and Bonferroni corrected values for Chromosome 24. Hardy-Weinberg deviations were estimated for each group (inverted homokaryotypes, heterokaryotypes, non-inverted homokaryotypes) within potential inversion regions. Individuals included in each of these groups were those identified in the PCAs of the LD blocks for chromosome 24 (Figure 12) representing the 3 different karyotypes groups. A p-value of 0.0001 with a Bonferroni correction for number of tests equal to the number of SNPs (~10,000) within the potential inversion region was used to identify SNPs deviating from HWE. There are 4 potential inverted homokaryotypes, 11 potential heterokaryotypes, and 166 potential non-inverted homokaryotypes.

Chromosome 24	# of SNPs out of HWE (P<0.0001)	# of SNPs out of HWE after Bonferroni correction
Inverted homokaryotypes	0	0
Heterokaryotypes	22	0
Non-inverted homokaryotypes	1798	401

Supplementary

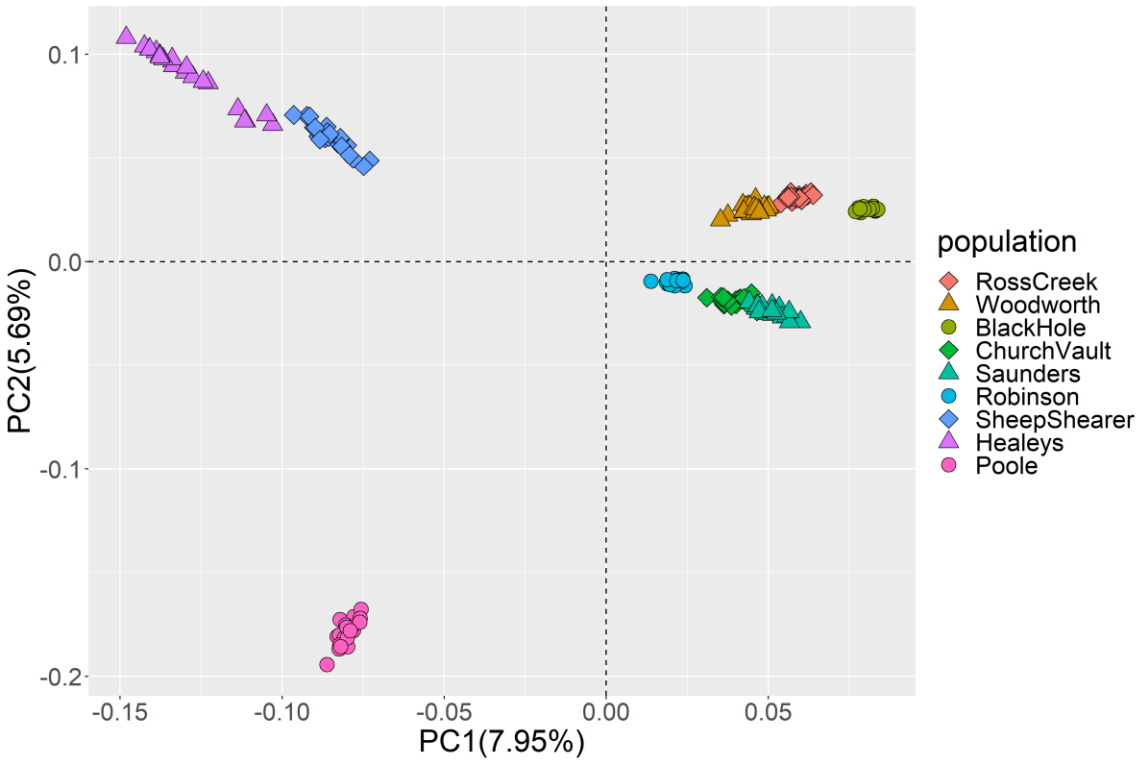

Figure S- 1 PCA of covariance matrix produced from ANGSD with genotype likelihoods calculated with SAMtools. The minor allele frequencies were determined using the SNP list of 1,745,248 SNPs. The PCA represents N=192 brook trout coloured by stream (population).

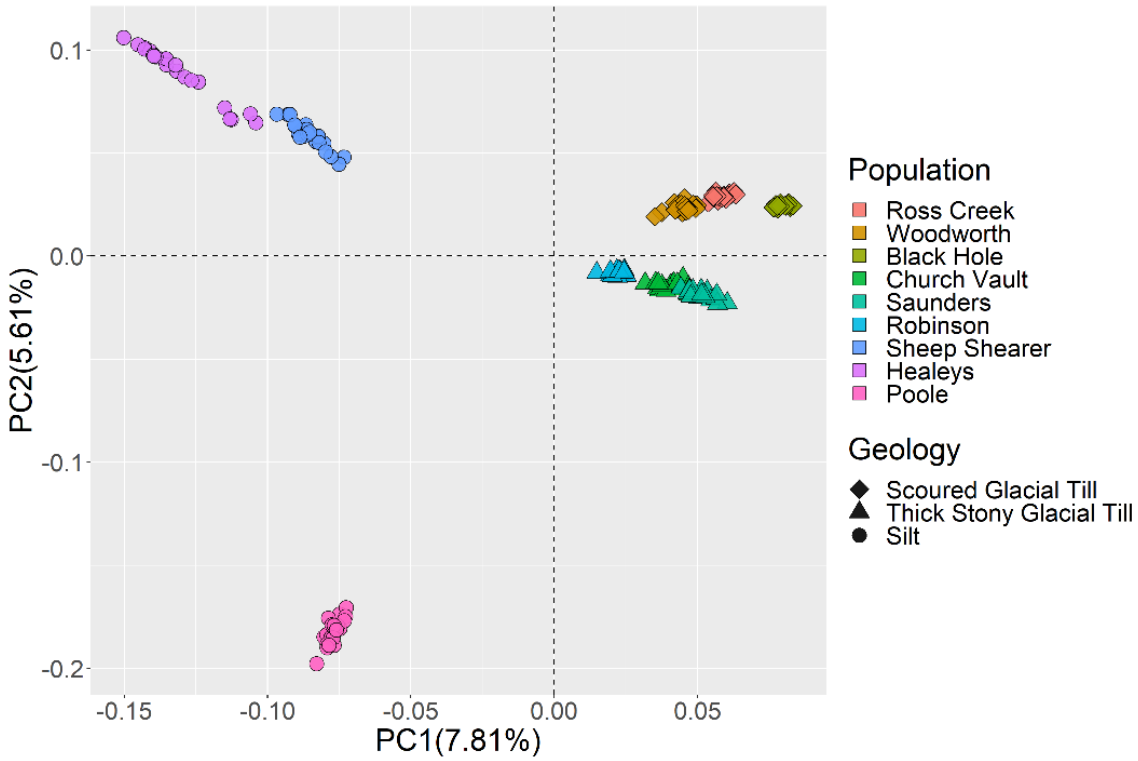

Figure S- 2 PCA of the covariance matrix produced from ANGSD with genotype likelihoods calculated using GATK. The PCA represent N=192 brook trout coloured by stream (population) and shape by geology type. The covariance matrix is based on 1,798,343 SNPs.

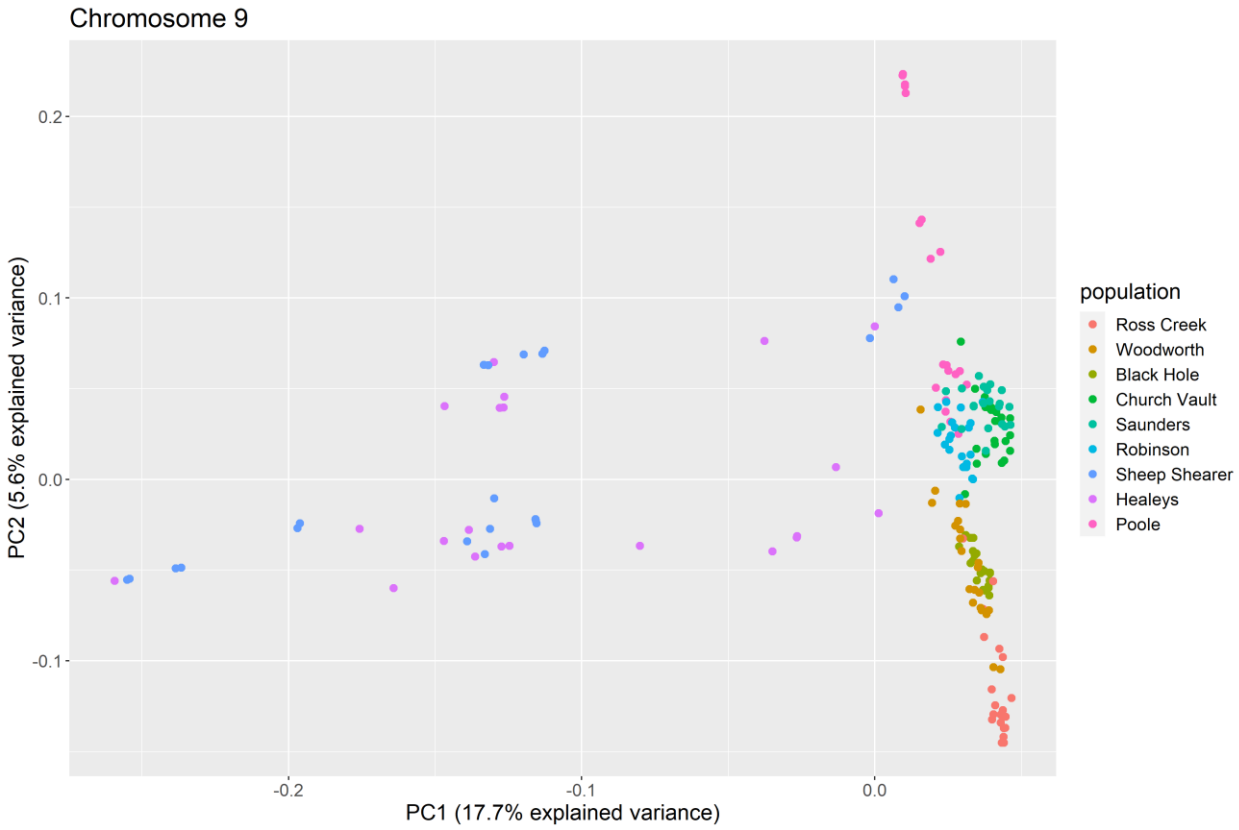

Figure S- 3 PCA of SNPs located in an LD block (20-45mb) on the lake trout chromosome 9 (NC_052315.1) for all 192 brook Trout. Genotype likelihoods and minor allele frequencies were determined using SNPs in the LD block and the covariance matrix was used to produce the PCA.

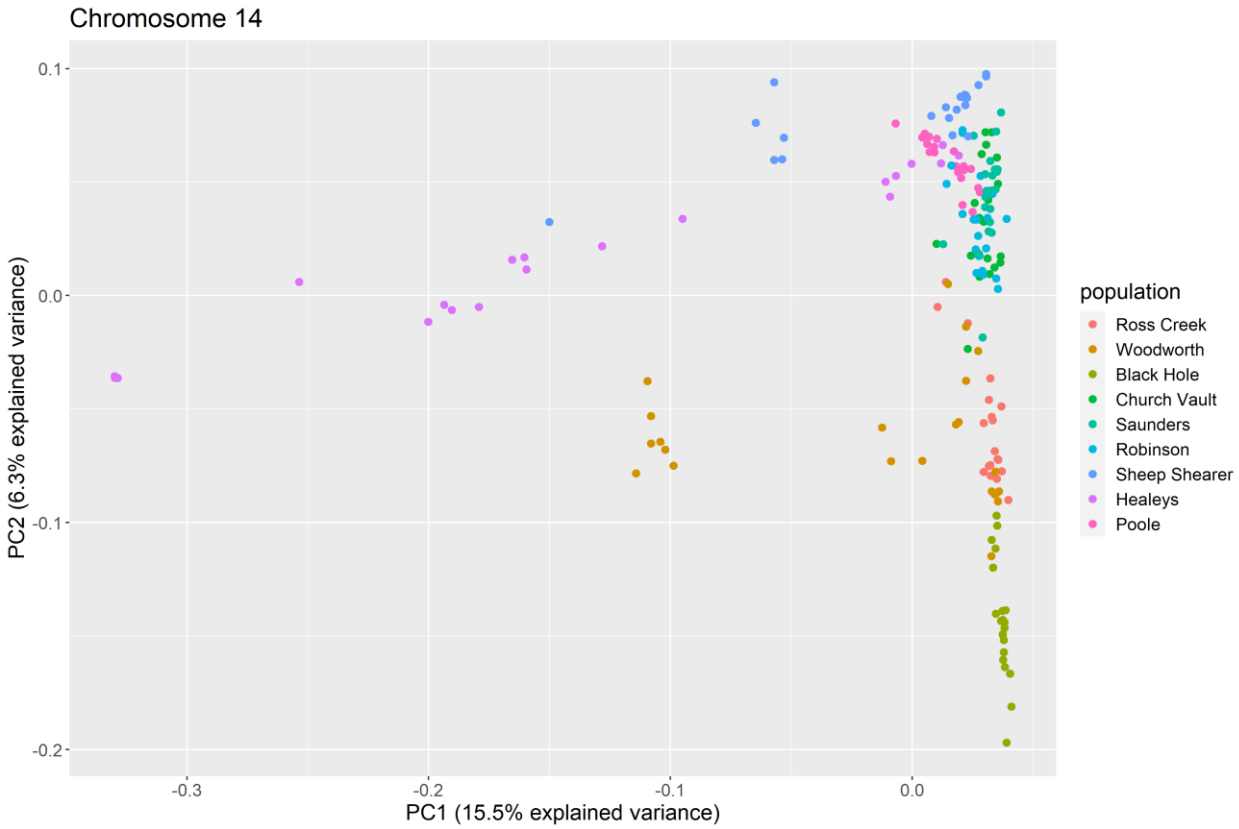

Figure S- 4 PCA of SNPs located in an LD block (5-30mb) on the lake trout chromosome 14 (NC_052320.1) for all 192 brook Trout. Genotype likelihoods and minor allele frequencies were determined using SNPs in the LD block and the covariance matrix was used to produce the PCA.

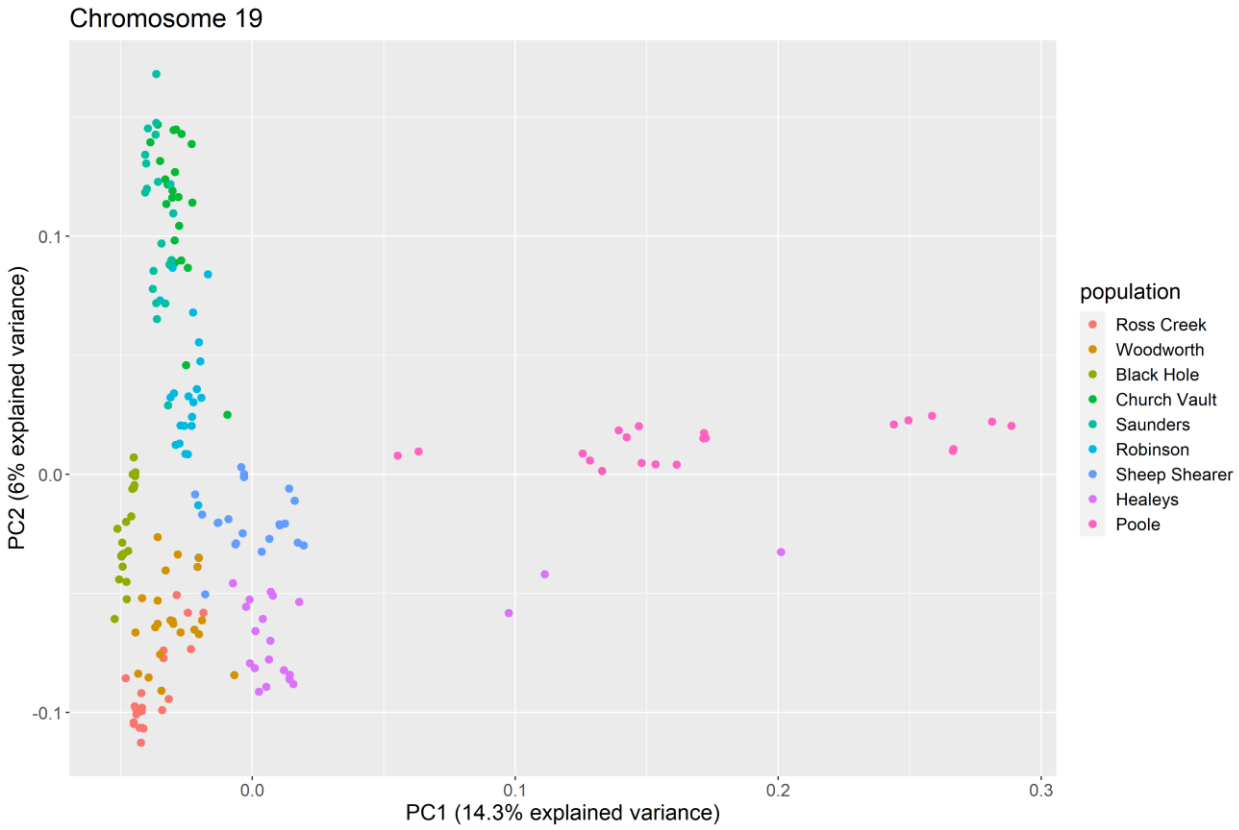

Figure S- 5 PCA of SNPs located in an LD block (0-31mb) on the lake trout chromosome 19 (NC_052325.1) for all 192 brook Trout. Genotype likelihoods and minor allele frequencies were determined using SNPs in the LD block and the covariance matrix was used to produce the PCA.

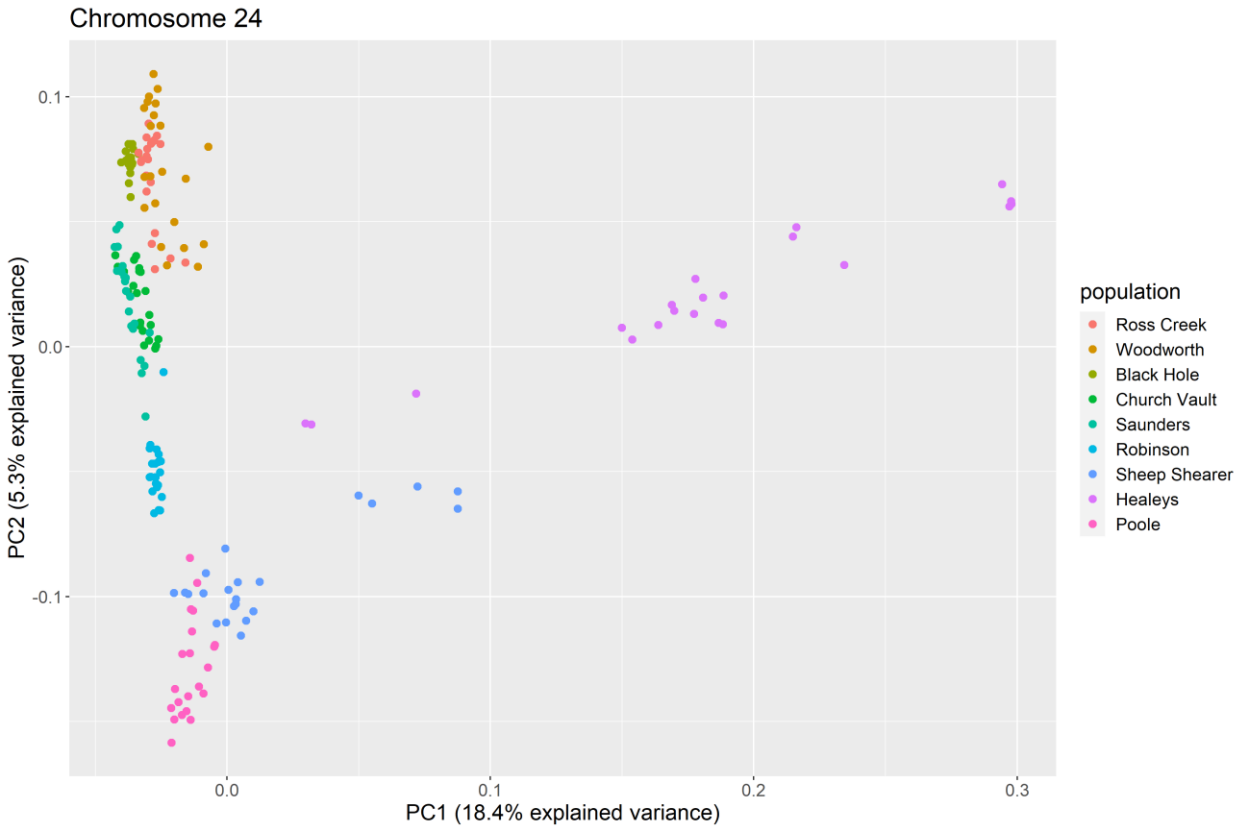

Figure S- 6 PCA of SNPs located in an LD block (7-35mb) on the lake trout chromosome 24 (NC_052330.1) for all 192 brook Trout. Genotype likelihoods and minor allele frequencies were determined using SNPs in the LD block and the covariance matrix was used to produce the PCA.

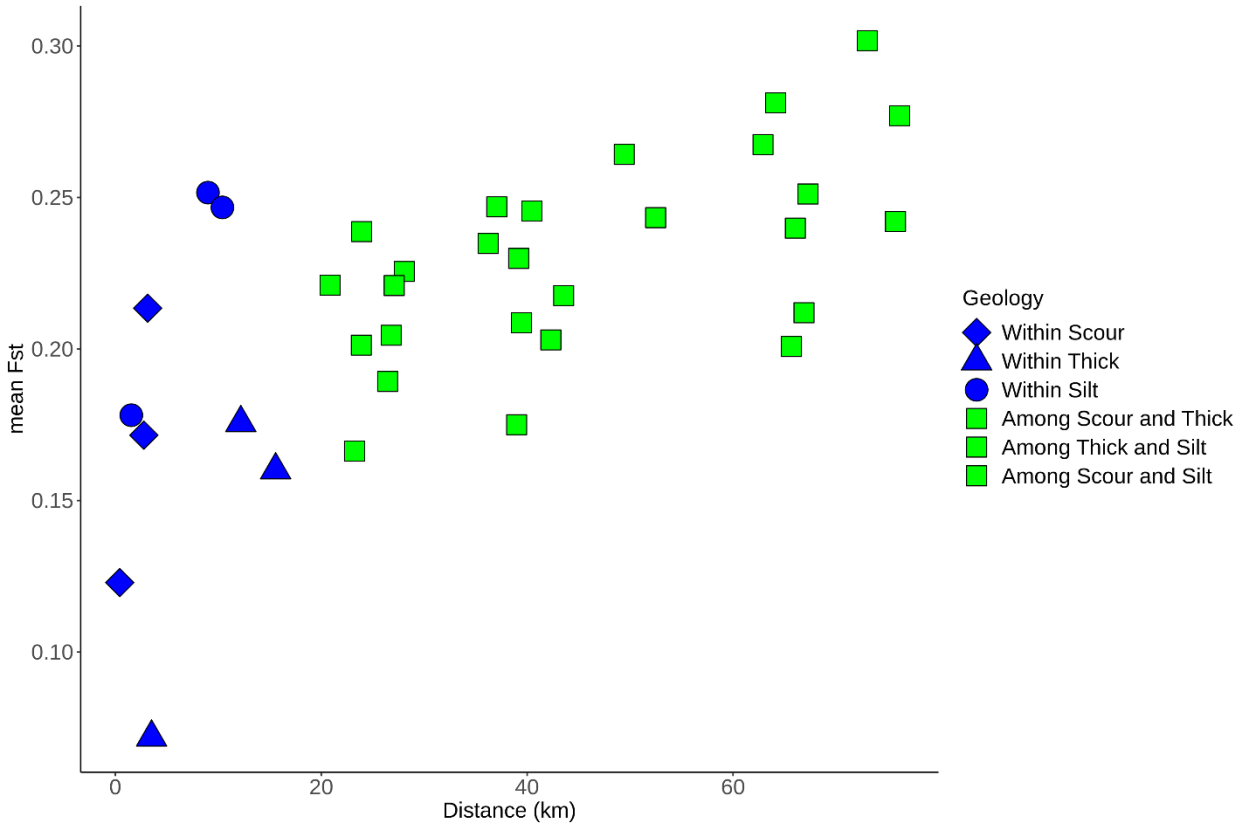

Figure S- 7 Average F_{ST} per population combination where there are 36 populations combinations plotted against the distance in km between each stream. Data was grouped based on geology (scoured glacial till, thick glacial till, silt) where blue shapes represent population comparisons between populations that exist in the same geology and green squares represent comparisons between populations that are in different geology groups.

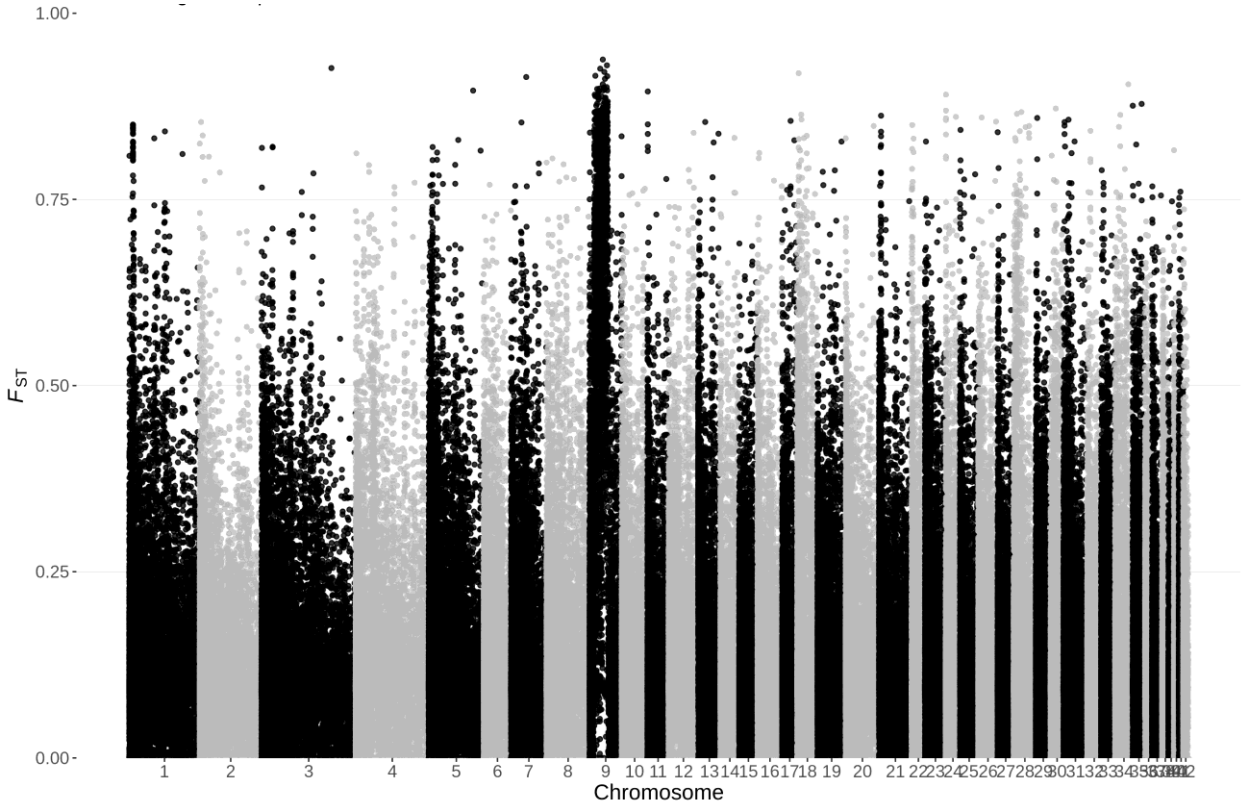

Figure S- 8 Genome wide Manhattan plot of 10 individuals, 5 from each of the inverted homozygote and non-inverted homozygote groups chosen from the PCA of the LD block region for chromosome 9 (NC_052315.1). F_{ST} was calculated in windows of 10kb with a 10kb slide from the 2D-SFS between non-inverted and inverted homozygotes. The lake trout reference genome was used to get SAFs for conversion to the 2D-SFS.

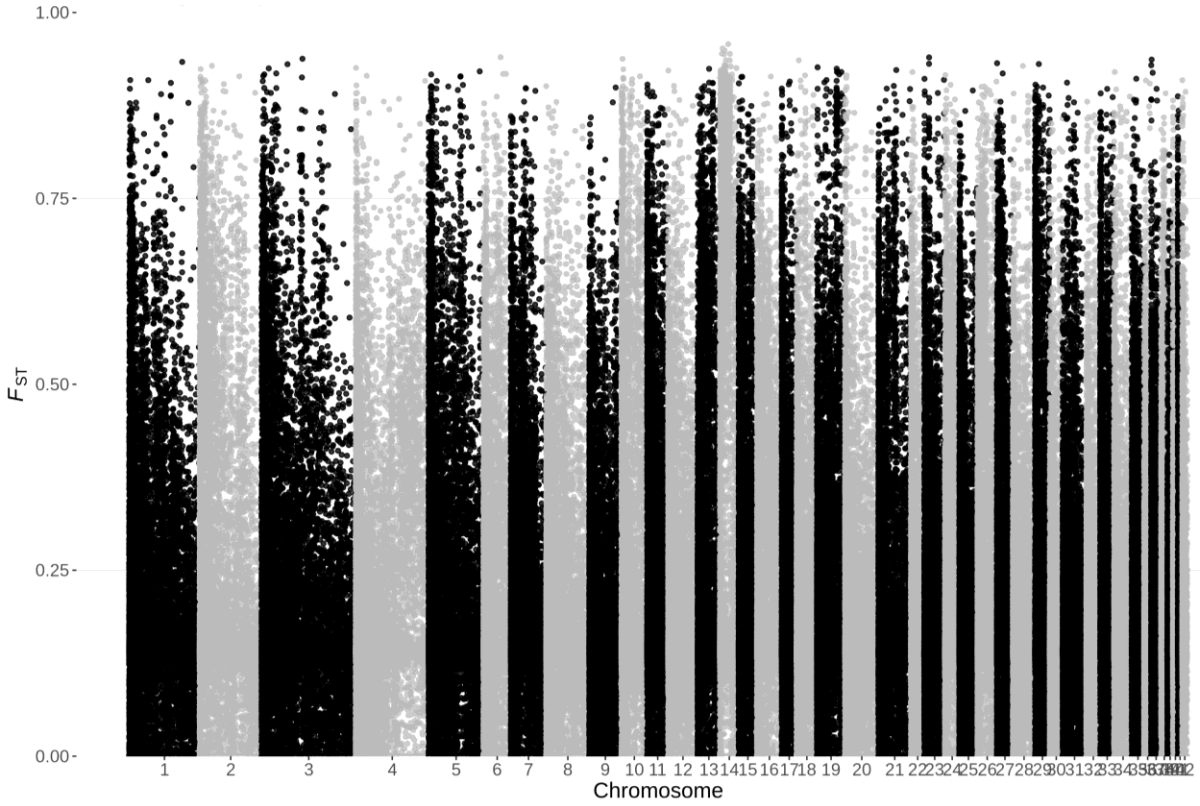

Figure S- 9 Genome wide Manhattan plot of 10 individuals, 5 from each of the inverted homozygote and non-inverted homozygote groups chosen from the PCA of the LD block region for chromosome 14 (NC_052320.1). F_{ST} was calculated in windows of 10kb with a 10kb slide from the 2D-SFS between non-inverted and inverted homozygotes. The lake trout reference genome was used to get SAFs for conversion to the 2D-SFS.

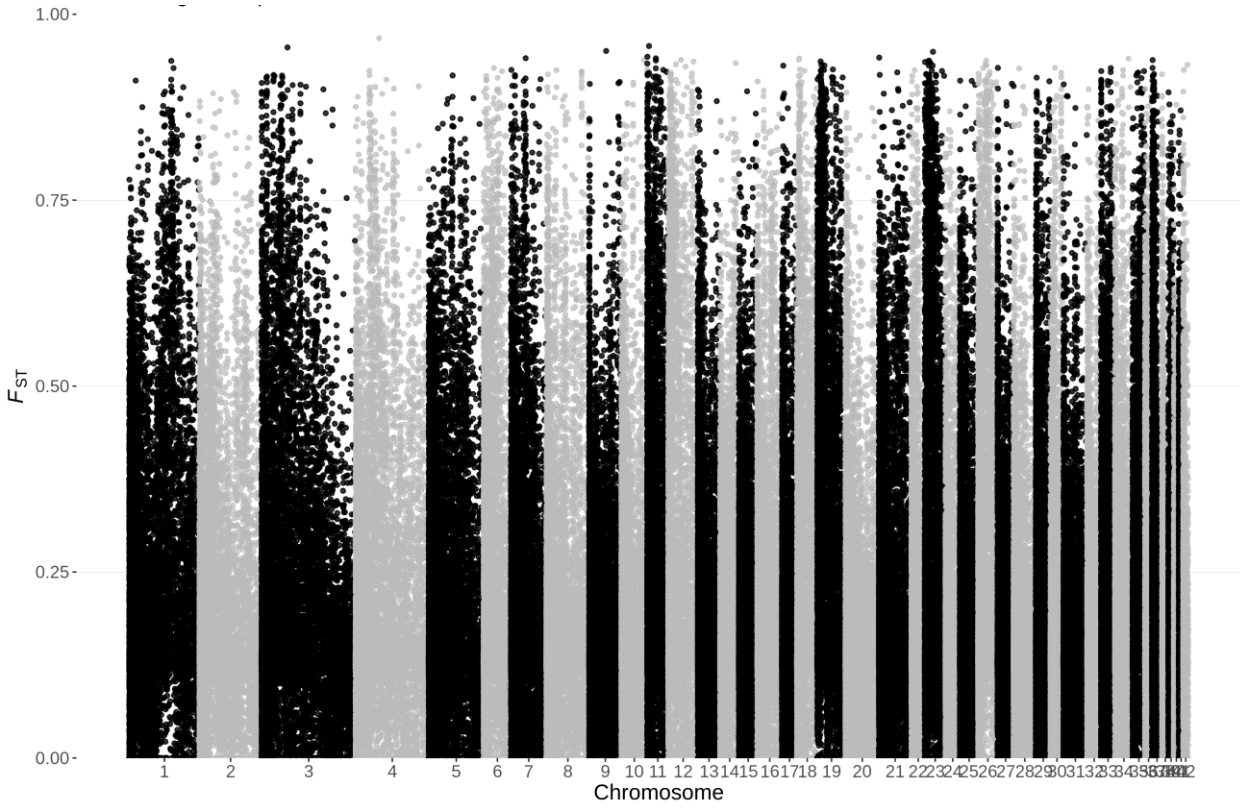

Figure S- 10 Genome wide Manhattan plot of 10 individuals, 5 from each of the inverted homozygote and non-inverted homozygote groups chosen from the PCA of the LD block region for chromosome 19 (NC_052325.1). F_{ST} was calculated in windows of 10kb with a 10kb slide from the 2D-SFS between non-inverted and inverted homozygotes. The lake trout reference genome was used to get SAFs for conversion to the 2D-SFS.

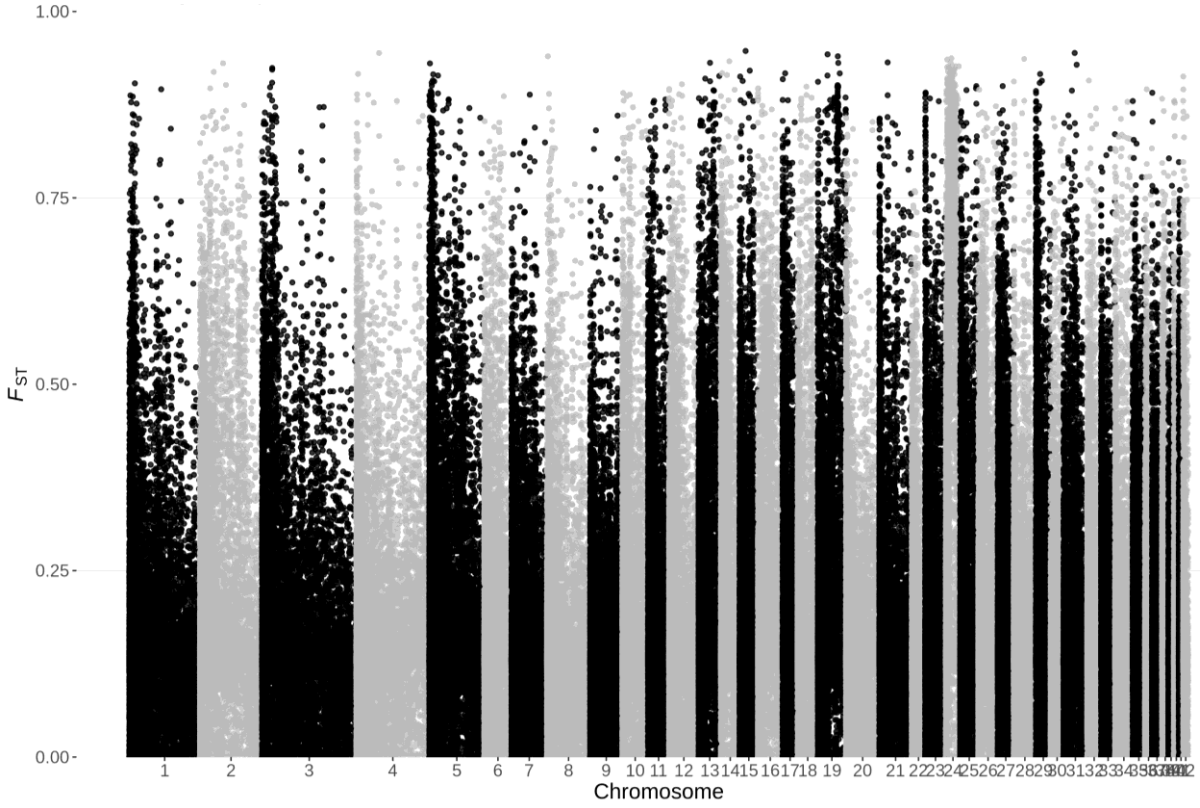

Figure S- 11 Genome wide Manhattan plot of 10 individuals, 5 from each of the inverted homozygote and non-inverted homozygote groups chosen from the PCA of the LD block region for chromosome 24 (NC_052330.1). The LD block on this chromosome was thought to be two separate blocks with this figure representing the first chunk. F_{ST} was calculated in windows of 10kb with a 10kb slide from the 2D-SFS between non-inverted and inverted homozygotes. The lake trout reference genome was used to get SAFs for conversion to the 2D-SFS.

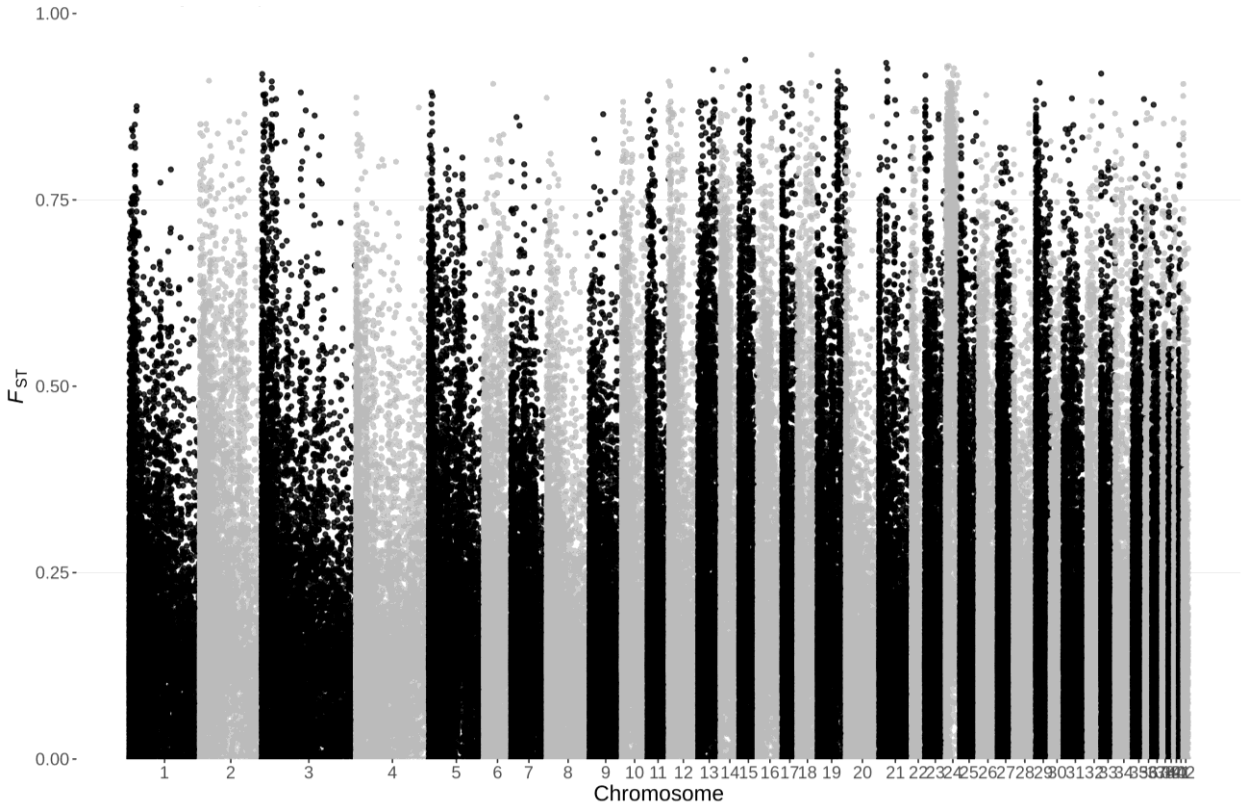

Figure S- 12 Genome wide Manhattan plot of 10 individuals, 5 from each of the inverted homozygote and non-inverted homozygote groups chosen from the PCA of the LD block region for chromosome 24 (NC_052330.1). The LD block on this chromosome was thought to be two separate blocks with this figure representing the second chunk. F_{ST} was calculated in windows of 10kb with a 10kb slide from the 2D-SFS between non-inverted and inverted homozygotes. The lake trout reference genome was used to get SAFs for conversion to the 2D-SFS.

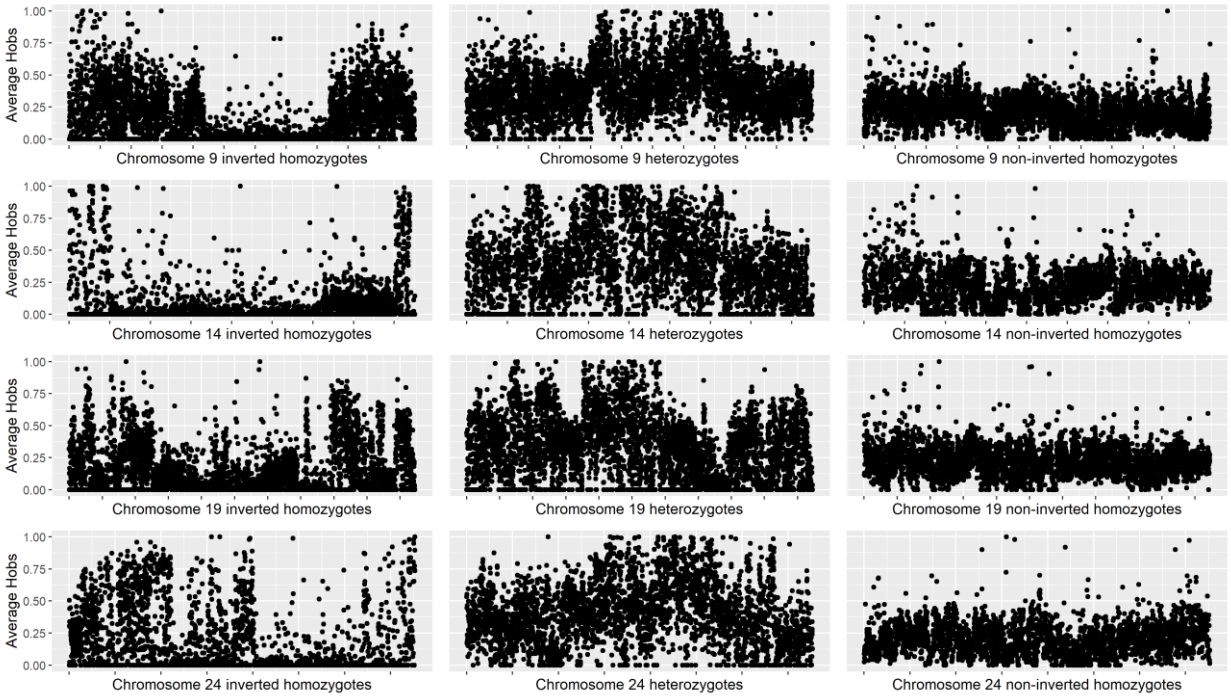

Figure S- 13 Average observed proportion of heterozygotes (H_{obs}) across entire chromosomes of individuals from the 3 different inversion groups from PCAs of LD blocks. All individuals from each of the inverted and non-inverted homozygotes and heterozygotes were used to calculate H_{obs}. Average H_{obs} was calculated in 10kb windows using a 10kb slide and points on the plot represent H_{obs} for SNPs in these windows.

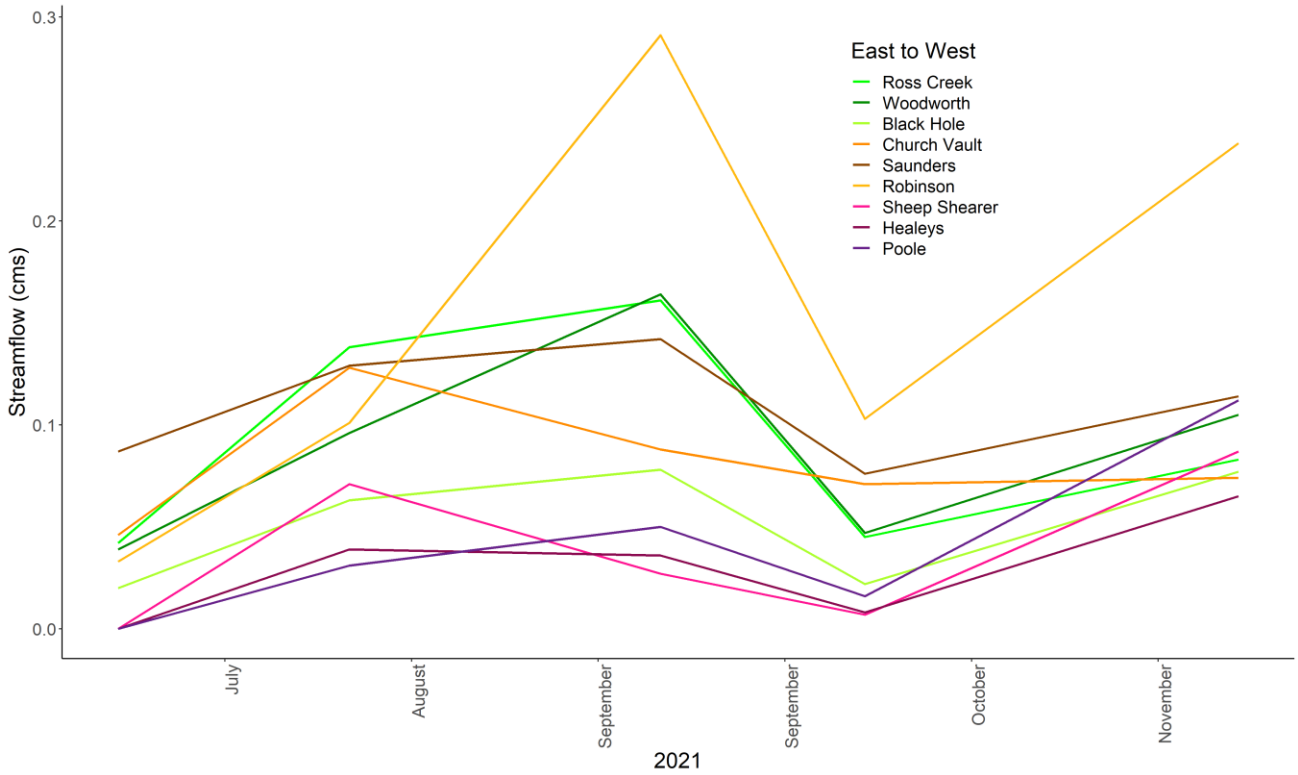

Figure S- 14 Streamflow in cubic meters per second (cms) taken once a month in all nine streams sampled for brook trout using the float method from June to November 2021.

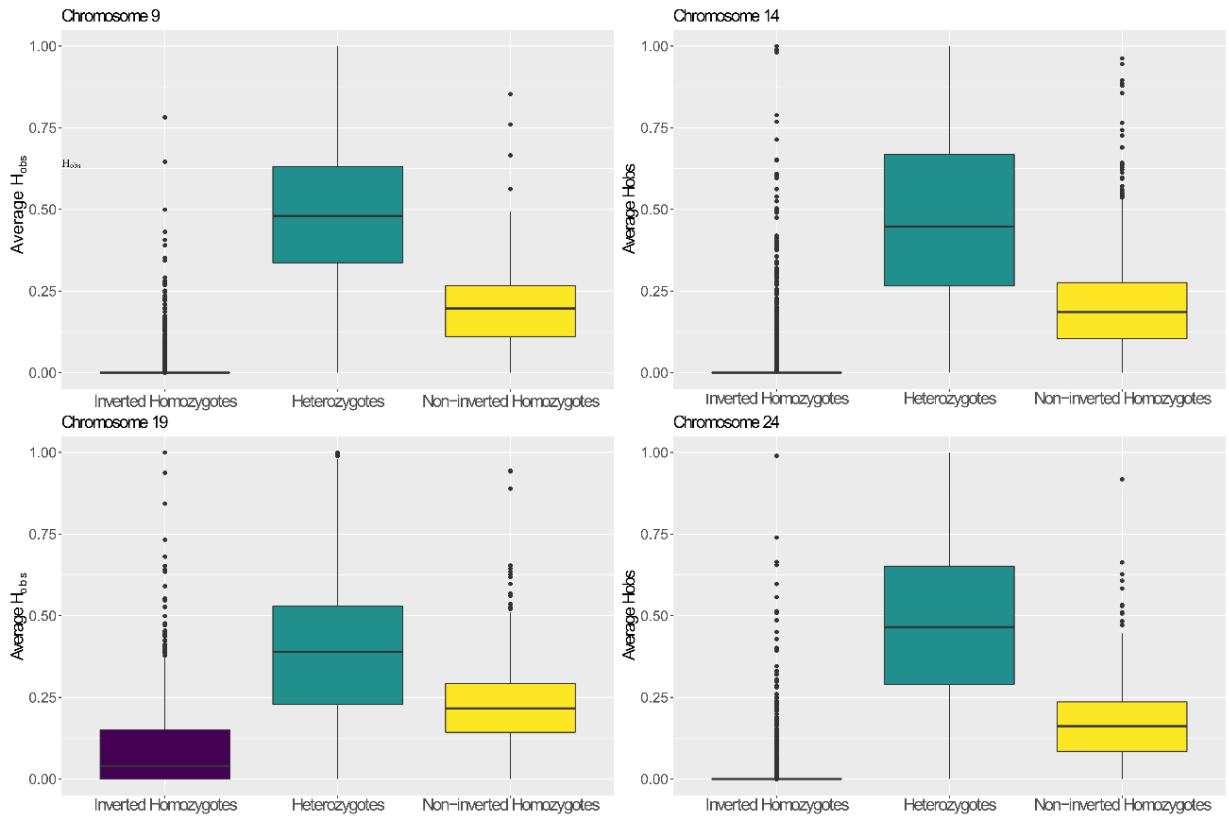

Figure S- 15 Average observed proportion of heterozygotes (H_{obs}) in 10kb windows using a 10kb slide of the 3 different inversion groups identified in the PCAs on LD blocks. For chromosome 9 there are 5 potential inverted homokaryotypes, 23 potential heterokaryotypes, and 140 potential non-inverted homokaryotypes. For chromosome 14 there are 4 potential inverted homokaryotypes, 9 potential heterokaryotypes, and 163 potential non-inverted homokaryotypes. For chromosome 19 there are 7 potential inverted homokaryotypes, 12 potential heterokaryotypes, and 169 potential non-inverted homokaryotypes. For chromosome 24 there are 4 potential inverted homokaryotypes, 11 potential heterokaryotypes, and 166 potential non-inverted homokaryotypes. Average H_{obs} was calculated in potentially inverted regions of chromosomes 9, 14, 19, and 24. H_{obs} is calculated as a population for every SNP and in this case each population is either inverted homozygotes, heterozygotes or non-inverted homozygotes. All individuals from each of the inverted and non-inverted homokaryotypes and heterokaryotypes were used to calculate H_{obs}. Number of points in each boxplot is representative of the number of windows scanned within potentially inverted regions.