Neutral and adaptive drivers of genomic change in introduced brook trout (*Salvelinus fontinalis)* populations revealed by pooled whole-genome re-sequencing

Brent Brookes

A Thesis

In

The Department

Of

Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science (Biology) at

Concordia University

Montreal, Quebec, Canada

August 2020

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Brent Brookes

Entitled: Neutral and adaptive drivers of genomic change in introduced brook trout (*Salvelinus fontinalis)* populations revealed by pooled whole-genome re-sequencing

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
James Grant

_____ Examiner
James Grant

_____ Examiner
Selvadurai Dayanandan

_____ External Examiner
Donald Gray Stirling

_____ Thesis Supervisor
Dylan Fraser

Approved by _____
Robert Weladji

_____2020 _____
André G. Roy

Abstract

Neutral and adaptive drivers of genomic change in introduced brook trout (*Salvelinus fontinalis)* populations revealed by pooled whole-genome re-sequencing

Brent Brookes

Understanding the drivers of successful species invasions is important for conserving native biodiversity and for mitigating the economic impacts of introduced species. However, whole genome resolution investigations of the underlying contributions of neutral and adaptive genetic variation to successful colonization in introduced populations are rare. Increased propagule pressure should result in greater neutral genetic variation, while environmental differences should elicit selective pressures on introduced populations, potentially supporting greater adaptive genetic variation. We investigated neutral and adaptive variation among nine introduced brook trout (*Salvelinus fontinalis*) populations using whole-genome sequencing (28,490,618 SNPs based on pool-seq). The populations inhabit different, isolated lakes in western Canada and descend from a common source, with an average of ~19 (range of 7-41) generations since introduction. We found no evidence of bottleneck events nor strong evidence of purifying selection, and little support that varying propagule pressure or differences in local environments shaped neutral genetic variation. Putative outlier analysis revealed non-convergent patterns of adaptive differentiation among lakes with minimal outlier loci (0.001%-0.15%) which did not correspond with tested environmental variables, despite conditions that should facilitate stronger adaptive differentiation (e.g. abiotic and biotic environmental differences, propagule pressure differences). Our whole genome sequencing analysis provides an example of an introduction success not strongly influenced by genetic variation and suggests that observed differentiation among introduced salmonid populations can be idiosyncratic, population-specific, or stochastic.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Dylan Fraser, for giving me the opportunity of a lifetime researching in the Rocky Mountains for my thesis. Your guidance, patience, edits, and suggestions made this thesis possible and I sincerely appreciate all the growth academic and life experience that your expertise and lab have offered me. Thank you to my advisory committee, Dr. James Grant, Dr. Selvadurai Dayanandan, and Dr. Gray Stirling for their help in structuring the critical components of this work, and their edits and suggestions.

To Shelley Humphries and the staff of Banff, Kootenay and Yoho National Parks, thank you for the amenities and access afforded to our research crew, without which this project would not have been possible.

To Dr. Hyung-Bae Jeon, thank you for your unending aid, advice, leadership, and guidance through every step of this project. You have acted as a big brother to me and your tutelage has taught me much of what I know now about genetics and fish work, and I would likely still be stuck trying to analyse raw data without you. I would like to thank all my lab mates who have aided in the statistics, editing, and suggestions on how to present this work. Jean-Michel Matte, Matthew Yates, Elizabeth Lawrence, Brian Gallagher, Shannon Clarke, Ella Bowles, Thais Bernos, Melia Lagace, Julien Beaulieu, Timothy Law, Alexandra Engler, Ramela Koumrouyan, Gabrielle Nessel, Ashlee Prevost, Nicole Hill, Zach Eisenhauer, Gillian Muir, Clea Montanari, and Richelle Smith your edits have helped get this work to where it is now.

To Dylan Glaser, Camden Moir, Alex Bourassa, Ty Brooks, Kia Marin, Carol Zastavniouk, Adam Baus, Adam Mitton, Kris Jelinski, Dania Goic Mac-Leod, Frank Wu, and every other field volunteer, colleague, and friend that aided in data collection or contributed edits to this work, thank you very much.

To my partner, Danielle, whose support has been unwavering and who has been there for me through thick and thin. You make me a better person at everything I try to accomplish. Your presence and care have kept me going through my hardest times, and your joyful attitude has made the good times all the better. Thank you for your unending help, love, and affection.

To my parents: Mom, Dad, and Kelsey, your support for my drive to work in a field that I am passionate, support through school, and upbringing has been instrumental to my success as a person and who I am today, I will never be able to thank you enough. To Lauren, thank you for always bringing out the best in me and teaching me there is more to life than being competitive. Kieran thank you for beating me in everything imaginable, keeping me grounded while helping me get better where I falter. Finna and Eli, thank you for always being happy to see me and teaching me how to look out for others.

Annick, your company and willingness to engage in shenanigans with me in every way shape and form has kept me from becoming bored, tired, or stressed. Thank you for always being there to lighten the mood, yell with me, and bring laughter to every ridiculous situation we've found ourselves in.

Nia and Ebony, thank you for both setting me straight when I'm clearly wrong and sharing all your opinions, recipies, and headstrong energy with me, it will always be sorely needed. Mila, Vanessa, Haylea, Nick, Alex, Felix, Lloyd, Jeremy, Mattieux, Jordyn, Shannon, Elizabeth, Jessie, Joe, and all new and old friends that helped me get on this path, pursue a masters, and who have been instrumental pillars in my life, thank you for all the friendship and guidance that you've given me.

Sampling was conducted in conjunction with NSERC SPG, 2016 and Parks Canada approved procedures. All procedures followed CCAC guidelines. Funding for this work was provided by NSERC, GRIL, QCBS, FRQNT and Concordia University.

Contribution of Authors

Study concept and design: Brookes, Fraser

Acquisition of data: Brookes, Jeon

Analysis and interpretation of data: Brookes, Jeon, Fraser

Drafting of Manuscript: Brookes

Critical revision: Brookes, Jeon, Fraser

Table of Contents

List of Figures

List of Tables

List of Supplementary Material

Introduction

Human-driven introductions of non-native species have occurred worldwide, both accidental, leading to invasions, and intentional (e.g. for recreation, conservation)(Lee, 2002). As species conservation and extirpation have biodiversity and economic impacts, determining the drivers of successful colonization is important (Lee, 2002). Predicting the roles of environmental and genetic factors on colonization success is particularly challenging, but imperative for mitigating species invasions and for improving reintroduction strategies of endangered species (Lee, 2002; Sakai et al., 2001).

Although it is well understood that successful species introductions often occur into habitats with familiar environmental conditions (Hayes & Barry, 2008; Moyle & Marchetti, 2006), species are also thought to colonize novel environments when adequate propagule pressure ensures that sufficient genetic variation is available for survival and adaptation (Arismendi et al., 2014; Benoit Facon et al., 2006). For example, genetic diversity should be increased by introducing more individuals and/or by carrying out multiple introductions (Allendorf & Lundquist, 2003; Ellstrand & Elam, 1993; Erfmeier et al., 2013, Lavergne & Molofsky, 2007; Walker et al., 2003). However, by itself, genetic diversity associated with propagule pressure may not be a pre-requisite for introduction success. Indeed, in some instances, successful introductions to new environments appear to have occurred despite apparently low levels of genetic diversity from founder effects (Lee, 2002; Prentis et al., 2008). Overall, because many past works lacked the resolution to investigate whole genomic variation in cases with known introduction histories (Halling et al., 2013; Kolbe et al., 2004; Tsutsui et al., 2000), the relative role of propagule pressure in influencing genetic diversity within successful introductions merits further empirical attention.

Until recently, a lack of genomic resolution and tools hindered the understanding of the relative role that adaptation to novel environments plays in the colonization success and genetic diversity of introduced populations. Adaptation is expected to be influenced by the strength of local environmental selective pressures and by demographic factors associated with the size of the introduced habitat (e.g.

larger habitats support larger populations and stronger natural selection; Chevin & Lande, 2011; Endler, 1986; Hamilton et al. 2015; Neville et al., 2009; Via & Lande, 1985; Warren et al., 2012). Recent works have illustrated that introduced species can become locally adapted to both abiotic and biotic environmental factors (Carroll et al., 2001; Filchak et al., 2000), as well as the population- and context-specific nature of such adaptation (Briscoe Runquist et al., 2020; Coulson et al., 2017; Schindler & Parker, 2002). This has led to the supposition that successful introductions are dependent on genetic factors associated with both sufficient propagule pressure and adaptation to the introduced environment (Allendorf & Lundquist, 2003; Prentis et al., 2008). However, studies with the genomic resolution needed to clarify the relative influence of both processes are rare (Dlugosch & Parker, 2008; Narum et al., 2017; Yoshida et al., 2016).

Colonization success of introduced species may also, in part, be linked to adaptive responses by each sex, as local adaptation to environmental conditions can differ between males and females (Fraser et al., 2019; Lasne et al., 2018; Svensson et al., 2019). Males and females can be under varying, concordant, or sometimes antagonistic selective pressures (e.g. intralocus sexual conflict in adult *Drosophila melanogaster* locomotor activity; Cox & Calsbeek, 2009; Connallon & Hall, 2016; Lasne et al., 2018; Long & Rice, 2007). The extent to which features of new environments (e.g. temperature, habitat use) generate phenotypic and genetic changes among the sexes in introduced populations is understudied (Stillwell & Fox, 2009; Svensson et al., 2019) but is thought to provide key information on the fate of populations during colonization and establishment (Iossa, 2019; Lasne et al., 2018; Sprogis et al., 2018),.

New genomic approaches can improve understanding of how propagule pressure, environmental factors, and sex differences affect genetic diversity during introductions into novel environments (Frachon et al., 2019; Micheletti & Narum, 2018; Narum et al., 2017). When genetic diversity is selectively neutral, loci have no effect on fitness, and hence the outcome is commonly linked

to genetic drift, gene flow, bottlenecks, and genetic load (Launey & Hedgecock, 2001; Narum et al., 2017).Conversely, when a genetic variant is putatively adaptive, loci are linked to traits that are under selection and associated with local adaptation to novel environments, commonly inferred by comparing genetic diversity among populations and to a reference genome (Dennenmoser et al., 2017; Hamilton et al., 2015; Hecht et al., 2015). Pool-seq is a method of whole-genome re-sequencing for which groups of individuals (i.e., populations) are sequenced together to generate millions of Single Nucleotide Polymorphisms (SNPs) throughout the genome (Schlötterer et al. 2014). SNP libraries using pool-seq have recently facilitated the analysis of neutral and adaptive genome-wide variation of introduced species in a cost-effective manner (Davey et al., 2011; Kurland et al., 2019; Narum et al., 2013; Stanford, 2019).

Socio-economically important salmonid fishes are amongst the world's most invasive species and provide ideal study species to examine factors influencing colonization success (Krueger & May, 1991; Lecomte et al., 2013; Vigliano et al., 2007). Salmonid invasions are a result of being transplanted worldwide in over 97 countries for sport-fishing or through aquaculture escapees into the wild (Fausch, 2007). Hatchery stock management can strongly influence the level of genetic diversity in introduced salmonid populations, as stocks with high genetic variability may be important for introduction success and for avoiding founder effects or bottlenecks (Bert et al., 2007; Kelly et al., 2006). Because environmental factors such as spawning area availability and habitat (stream/lake) size regulate salmonid abundance (Krueger & May, 1991) and genetic diversity (Bernos & Fraser, 2016; Neville et al., 2011; Rieman & Allendorf, 2001; Wood et al., 2014), they may also create conditions for introduced salmonid genotypes under selection to confer fitness advantages and thereby potentially influence colonization success (Benjamin et al., 2007; Hecht et al., 2015; Kinnison et al., 2008).

In this study, we examine the evolutionary and demographic history of nine introduced, brook trout (*Salvelinus fontinalis*) populations using SNPs generated from a pool-seq approach, with the aim to

quantify the neutral and adaptive genetic variation associated with successful introduction and colonization. To enhance sportfishing opportunities, brook trout were stocked between 1941 to 1973 into mountain lakes in Banff, Kootenay, and Yoho National Parks, Canada (Figure 1; Supplemental Table 1), by hatcheries using the same strain of hatchery trout from the eastern USA (National Parks stocking records). In an effort to restore native aquatic ecosystems, Parks Canada has initiated the manual removal of brook trout in several lakes (Adams et al., 2000; Dunham et al., 2002; Earle, et al., 2007; Fisheries and Oceans Canada, 2014). These lakes represent novel environments for brook trout relative to their eastern North American range: high elevations, covered by ice for 7-9 months of the year, high pH, and drastic seasonal changes in spawning site availability from snowfall runoff (Fassnacht et al., 2018; Power, 1980; Wood et al., 2014; Yates et al., 2019). Parks Canada's mandate provides an opportunity to investigate how brook trout of the same origin colonized and evolved in different environments outside of their native range, over an average span of approximately ~19 generations (range of 7-41, based on 95% of the population at the age of maturity) for the species (Glaser et al., unpublished data).

We hypothesized that standing neutral genetic variation between lakes would be positively associated with propagule pressure, while adaptive genetic variation would be associated with lake-specific environmental variables. Based on the stocking history and environmental data collected from different lakes, we predicted: i) lakes stocked with greater total numbers of fish or through more stocking events would have more neutral genetic variation; ii) neutral genetic variation would be positively correlated with lake (habitat) size; and iii) bottleneck events in the demographic history of the populations and a higher proportion of deleterious mutations would be associated with lower neutral genetic variation. In relation to environmental effects, we further predicted that: iv) an increase in putatively adaptive outlier SNPs would be associated with great environmental differences among lakes; v) signals of adaptive genetic variation would be positively correlated to lake size and to the relative

availability of spawning sites; and vi) signals of adaptive differentiation would vary by sex; specificially that the signals would be stronger in females than males, as there is some evidence that natural selection to abiotic environmental factors can be stronger in female brook trout (Zastavniouk et al., 2017).

Methods

Study Site

Within Banff, Kootenay, and Yoho national parks at the border of Alberta and British Columbia, Canada (Figure 1), nine study lakes were selected for their physical isolation from other lakes, small size, limited inlet/outlet expanse, and brook trout dominance. According to available Parks Canada records, these brook trout originated from a common origin (Paradise Brook Trout company, Pennsylvania USA). The fish were used to establish broodstocks in two Parks Canada hatcheries (Banff, Jasper) for subsequent stocking into park lakes. The specific lakes in this study were: Cobb, Dog, Helen, Margaret, McNair, Mud, Olive, Ross, and Temple (Donald & Alger, 1984; Parks Canada stocking records, Figure 1). The lakes Margaret, Dog, and McNair are potentially open to seasonal gene flow from brook trout populations residing in adjacent bodies of water through otherwise unpassable outlet waterfalls in extreme weather scenarios (Adams et al., 2000; Thompson & Rahel, 1998).

Sampling Methods

Sampling of brook trout was conducted in August 2017, using a standardized, mixed-mesh gill net protocol until 5-10% of fish in each lake were captured; captured fish were euthanized with an overdose of clove oil. Caudal fin tissue was collected from each fish and stored in 95% ethanol for DNA extraction, while sex was determined by abdominal dissection (M=male, F=female). Methodology associated with capture method (i.e. net mesh sizes/lengths, set durations) and census population size estimates obtained in concurrent research and used in analyses below are described in Yates et al. (2020).

To test the hypothesis that habitat size was positively correlated with neutral genetic diversity and adaptive differentiation, a series of limnological, abiotic and biotic variables were quantified in each lake between June and August 2018 (Supplemental Table 2; Appendix C). Seep (groundwater

upwellings), inlet and outlet number, and water discharge, were measured by circumnavigation of each lake as metrics of spawning site availability. Considered a stocking variable, distance to each lake was calculated with Parks Canada hike information and Google Earth 9.2.58.1 along hiking trails or directly from the nearest vehicle access. Connectivity, bathymetry, and lake size were calculated with ArcGIS version 10.3.1 and obtained from Parks Canada records. *YSI* measurements and *HOBO* loggers were used to measure pH and temperature. Detailed descriptions of macroinvertebrate and zooplankton sampling can be found in Appendix C. Briefly, macroinvertebrate sampling was done with a D-frame net, filtered through a 500-µm sieve bucket and stored in 95% EtOH. Zooplankton samples were collected from an average of eight pelagic subsamples of Wisconsin 54- µm whole column net pulls across the lake and stored in 95% EtOH. Jaccard's dissimilarity index for each lake was calculated using presence-absence data collected from the sampling period (R package Adespatial, v 0.3-8; Stéphane Dray et al., 2020); R v 4.0.0 (R Core Team, 2020) and RStudio v 1.2.1335 (RStudio Team, 2020), were used for statistical analyses.

DNA Extraction, Pooling, and Sequencing

DNA extractions from fin tissue were conducted using *Qiagen* blood and tissue kits (Qiagen, Germany) and a modified protocol (Appendix A). DNA quality and quantity were assessed by 1% Agarose gel electrophoresis using HindIII digested Lambda DNA run at 100V to check the degree of degradation. It was further quality tested on a *Qubit* Fluorometer 2.0 (Invitrogen, USA) selecting for quantity >20 ng/µL, and *NanoDrop* spectrophotometer (Thermo Scientific, USA) selecting for quantity >40 ng/µL, and quality 260/230 and 260/280 ratios of >1.8; finally each sample was run through a *Qubit* Fluorometer for a second time to confirm DNA quantity.

Individual DNA was then pooled by sex, and population (18 total pools) with 20 individuals in each pool (Dennenmoser et al., 2017; Schlötterer et al., 2014); exceptions were Cobb females (n=17)

and both McNair sexes (n=8) due to low sample sizes. An amount of 50 µL of each individual sample was selected for each pool at a dilution of 10 ng/µL and confirmed both prior to and post using a *Qubit* Fluorometer. DNA was then pooled together at a concentration of 3 ng/µL and the concentration was confirmed using a *Qubit* Fluorometer.

Pooled DNA was then sent for library prep at Génome Québec Innovation Centre, Montréal, Québec, Canada via a shotgun approach with PCR with *Illuminia TrueSeq* LT adaptors (Illumina, USA). All pools passed quality and quantity requirements and were sequenced via *Illumina HiSeqX* (Illumina), with each pool on two lanes of *NovaSeq 6000 S4* flowcell and paired-end reads of 100 base pairs (bp) for approximately 1X depth of coverage. We estimated the genome size of brook trout as approximately 3Gb based on the Animal Genome Size Database (http://www.genomesize.com/).

Poolseq Pipeline and SNP Discovery

A reference genome was prepared using the arctic charr (*Salvelinus alpinus*) genome available from NCBI (ASM291031v2, Christensen et al., 2018) due to its close phylogenetic relationship and chromosome number and size to brook trout. The reference genome was prepared using Burrows-Wheeler Aligner (BWA) v 0.7.12 (Li & Durbin, 2009), indexed with SAMtools v 1.5 (Li et al., 2009), and a dictionary was created using Picard tools v 2.17.11 (http://broadinstitute.github.io/picard/, accessed 20-11-2019) to permit sequence alignment.

A full version of subsequent genomic analyses through the PoolParty pipeline v 0.8 (Micheletti & Narum, 2018) and additional analyses can be found in Appendix B. Single nucleotide polymorphism (SNP) discovery was facilitated by the PoolParty pipeline by firstly adopting the PPalign module to filter raw reads and map them to the reference genome. Raw reads were trimmed by BBMap v 37.93 (Bushnell, sourceforge.net/projects/bbmap/, accessed 20-11-2019), and filtered using Fastqc v 0.11.7 (Andrews, 2010) for a quality score of 20. BWA-MEM v 0.7.12 was utilized to align sequences to the

reference genome. SAMtools v 1.5 was then used with a mapping quality threshold of 10 to remove any sequences that did not map sufficiently, while SAMblaster v 0.1.24 (Faust & Hall, 2014) marked duplicates in the SAM files and extracted discordant read pairs. SAMtools, BBMap, and Picard tools were used to remove duplicates, sort BAM files, and filter unpaired reads using a minimum fastq trimming length of 25bp before discarding. BCFtools v 1.5 (Li et al., 2009) was then utilized to call SNPs with a quality score of 20, a minimum global allele frequency of 0.05 was required to retain a SNP, and a minimum global coverage of 10 was needed to retain a SNP; while an indel window of 15bp was used for masking SNPs around indels. PPalign created mpileup, variant call format, sync, allele frequency table, and BAM files for use in future analyses, as well as alignment, population, trimming, and quality reports. Pool library specific files for blacklisted minimum global allele frequency and multiallelic SNPs were also produced.

Raw reads were checked for quality using FastQC and MultiQC v1.7 (Ewels et al., 2016). Following alignment, pool reads via previously generated mpileup files were run through the PPstats module of the PoolParty pipeline to provide depth of coverage statistics. Depth of coverage analysis was conducted with a minimum and maximum coverage to retain a genomic position at 1 and 100 respectively.

Neutral Genetic Diversity and Differentiation

To test our hypotheses regarding the relationship between both propagule pressure and environmental variables (habitat size) on neutral genetic diversity, PoPoolation (Kofler et al., 2011) was utilized in analyzing mpileup files derived from PPalign to determine SNPs. PoPoolation settings were: afastq-type of “sanger”, minimum count of 2, minimum and maximum coverage of 4 and 100 respectively, window and step size of 250, and pool size represented by 2x the individuals (diploid). For our nucleotide diversity regressions, a correlation matrix for scaled stocking and environmental

variables was run at a cut-off of 0.7 with the psych package v 1.9.12 (Revelle, 2019). To remove multicollinearity, a linear variance inflation (VIF) analysis with all stocking and environmental variables in car v 3.0-5 (Fox & Weisberg, 2019). The correlation matrix and VIF analysis <10 left only water volume, elevation, pH, zooplankton density, macroinvertebrate density, number of tributaries, and total number of fish stocked as final variables (Bagheri & Midi, 2009; Neter et al., 2004). Beta-regressions were then run in RStudio with betareg v 3.1-2 (Cribari-Neto & Zeileis, 2010) for all variables separately; additive and interaction terms were not calculated as we did not have sufficient power with 18 pools (Supplemental Table 3). Sexes were run separately to avoid pseudo-independence, as they had identical variable data but different dependant variables (nucleotide diversity). Forward model selection was conducted using AIC with the MuMIn package v 1.43.6 (Barton, 2009; Supplemental Table 4). Model visualizations were conducted with ggplot2 v 3.2.1 (Wickham, 2016).

To investigate whether introduced populations experienced inbreeding depression and bottleneck events after introduction, the genetic load of each pool was estimated as loss of function (LOF) variation using SnpEff and SnpSift v 4.3t (Cingolani et al., 2012). LOF mutations are well known for interfering by generating a stop codon in the coding region. Before identifying the LOF mutations, all variants were annotated using SnpEff. The vcf file was retrieved from PPalign, and filtered to remove SNPs on the scaffold, leaving only those on linkage groups. At the beginning of annotation steps, the genome database was built based on both gff3 and fasta files obtained from NCBI (Agarwala et al., 2018). The deleterious LOF variants were sorted by the following four categories, named for their putative impact: “high”, a variant with a significant deleterious impact on the coding region (e.g. start codon lost, stop codon gained, frameshift variant, etc.); “moderate”, a non-disruptive variant that may effect efficacy (e.g. missense variant, splice region variant, etc.); “low”, an innocuous or unlikely deleterious variant; and “modifier”, a variant located in an intergenic region. Finally, the proportion of

heterozygosity from the categories was calculated to compare levels of deleterious mutations among our 18 pools using the "vcfhetcount" argument in vcflib (Garrison, 2016).

To reconstruct the demographic dynamics of introduced brook trout populations and examine the genetic stability of their hatchery source population, we performed a site frequency spectrum (SFS) Stairway plot v 2.0 analysis (Liu & Fu, 2015). The Stairway plot analysis assesses the potential for recent bottleneck events by examining effective population size fluctuation over time and is applicable to pool-seq approaches (Liu & Fu, 2015). The vcf file created from PPalign was converted to an SFS input file using easySFS (easySFS, 2020). A single, grouped population of all nine lakes was used to further examine evidence of bottleneck events post-introduction reflecting the demographic history of the entire population from the source to stocking/introduction. A folded SFS of 5224 singletons and 7623.99 doubletons was used with 67% of the sites retained for training.

To test for genetic structure among pools, pairwise sliding-window fixation indexes (F_{ST}) between all 18 pools were determined with Poolfstat v 1.1.1 (Hivert et al., 2018), an R package designed specifically for computing the F-statistic of pool-seq data. The sync file obtained from PPalign was used to create a popsync object using Poolfstat with a minimum read count of one, a minimum coverage per pool of one, and a maximum coverage per pool of 200, while using the same minor allele frequency as the original file of 0.05 and removing indels, as commonly adopted (Kofler et al., 2011; Micheletti & Narum, 2018). The "Anova" method of Poolfstat was used to calculate pairwise F_{ST} with the same parameters as above aside from the minimum coverage per pool being increased to 20. Visualizing F_{ST} was completed by creating a Principal Coordinate Analysis (PCoA). A distance matrix and eigenvalues were generated and visualized alongside pairwise F_{ST} with ggplot2, using the dist function in R and the pco function of the LabDSV v 1.8-0 (Roberts, 2019) package.

All loci containing at least one multiallelic SNP (considered as blacklisted alleles), were removed from each pool when calculating neutral genetic diversity (Létourneau et al., 2018; Narum et al., 2017; Terekhanova et al., 2019). A second calculation of both nucleotide diversity and pairwise F_{ST} were run without removing the aforementioned loci to determine the difference between procedures. Comparing biallelic and multiallelic filtering methods between these two methods using pairwise t-tests showed a significant difference between the two methods ($p=2.08E^{-7}$ for nucleotide diversity, and $p=6.3E^{-15}$pairwise F_{ST} respectively; Supplemental Figure 1). As such, subsequent analyses were conducted with filtered biallelic datasets. Furthermore, we did not remove outlier loci in calculations of neutral genetic diversity as they only represented 0.15% of all tested SNPs.

Visualization of population structure was conducted using an unweighted pair group method with arithmetic mean (UPGMA) dendrogram of the 18 pools with bootstrap values of 100, from the 28,490,618 binary SNP vcf file of the PPalign module, created through the Rstudio packages vcfR v 1.8.0 (Knaus & Grünwald, 2016 & 2017), ape v 5.3 (Paradis & Schliep, 2018), and poppr v 2.8.3 (Kamvar et al., 2014 & 2015).

Adaptive Genetic Differentiation

To investigate loci under selection, outlier tests were first performed with PCAdapt v 4.3.3 (Privé et al., 2020; Luu et al., 2017). PCAadapt was run under the pool (n=18) designation to represent our pooled dataset on a filtered (biallelic and scaffold removed) allele frequency file of 14,479,563 SNPs with a Bonferroni p-value adjustment and FDR bounded at 0.05. Using a Bayesian framework, PCAdapt determines population structure using *K* z-scores to fit SNPs to *K* principle components based on Cattell's rule (Cattell, 1966), where SNP-specific Mahalanobis distances are used to evaluate outliers from the normal z-score distribution, explained by the *K* factors. We chose PCAdapt because of its ability to run pool-seq data and its strength in examining a divergence model with hierarchical population

structure, while maintaining statistical power and a controlled false discovery rate (FDR) (Luu et al., 2017). Furthermore, PCAdapt has been shown to have consistent strength for detecting outlier loci under weak, moderate, and strong selection (Lotterhos & Whitlock, 2015).

A redundancy analysis (RDA) was also used as a method to examine our hypotheses regarding adaptive differentiation and to determine the driving habitat and environmental factors (n=14; Supplemental Table 2) of putative loci under selection. RDA, a form of multivariate genotype-environment association, was conducted using the package vegan v 2.5-6 (Oksanen et al., 2019), the pooled allele frequency data from PPalign, and based on scaled habitat and environmental predictors. As we were unable to run the RDA on an unfiltered dataset due to computational restrictions, our allele frequency data were filtered to include only the outliers attained from PCAdapt. A correlation matrix for the habitat and environmental variables was then run at a cutoff of 0.7 with package psych package v 1.9.12, where surface area, depth, downstream discharge, and upstream discharge, and temperature variance were removed. Finally, the RDA was conducted against all remaining variables, at which point significance was computed using F-statistics for each constrained axis (Legendre et al., 2010) to examine if specific habitat and environmental predictors explained PCAdapt outlier loci.

The PCAdapt and RDA analyses searched for loci under selection and the habitat and environmental variables responsible across all 18 pools. We also utilized lake-specific outlier analysis based on Cochran-Mantel-Haenszel (CMH) chi-squared tests to examine specific outlier loci between individual lakes, and whether there existed convergent and parallel evolution between annotated gene products. The PPanalyze module of the PoolParty pipeline was used to create sync files of each population comparison (male and female pools combined) with a minimum coverage of 2, maximum coverage of 20 and a minor allele frequency of 0.08. PoPoolation 2 (Kofler et al., 2011) was employed with minimum count of 20, minimum coverage of 20, and maximum coverage of 100. P-values were corrected with an FDR of 0.05 using Benjamini & Hochberg correction. Candidate genes were then run

through SnpEff v 4.3t (Cingolani et al., 2012) to annotate the VCF file from the reference Arctic charr genome, and then run through BLAST (Altschul et al., 1990) and QuickGO (Binns et al., 2009), to determine gene ontology by taking the first available annotated gene product for each BLAST definition. After annotation, SNPs on the scaffold of the vcf file were removed, leaving only biallelic SNPs aligned to linkage groups of the Arctic char reference genome.

Finally, to explore the genomic signature of sexual differentiation, F_{ST} was calculated using 28,490,618 SNPs from the PPanalyze module and a genome wide association study (GWAS) approach (Buniello et al., 2019). As CMH tests examined population-specific differences between lakes (males and females combined), this analysis was conducted with pools separated into nine male and nine female groups, one for each lake population. To examine sex-specific outlier loci, significant SNPs were estimated with a local score analysis based on the Fisher's exact test (FET) for both sexes.

Results

DNA Sequencing

In the 18 pools, raw sequence counts totalled 10,775,432,330 reads, of which unique reads totalled 7,775,705,194, averaging 431,983,622 unique reads in each pool (72.2%) with quality scores of 36 (Supplemental Figure 2). There were 68,836,296 total SNPs called, of which 10,721,236 were removed due to low quality and global coverage parameters. An additional, 25,898,239 SNPs were removed due to global minor allele frequency restrictions and 3,726,203 were removed as indels. Collectively, after all filtering steps, 28,490,618 SNPs remained for the study's analyses. All pools passed base quality scores, and per sequence GC content was normally distributed around 44%. Mean filtered depth of coverage between all 18 pools (9 sex-differentiated populations) was 15.4X with a standard deviation of 1.7 (Supplemental Figure 3). The proportion of the genome anchored to the reference was 69.9%, while the proportion of the genome sufficiently covered with reads was 65.4% (1,418,188,531 bp) of all libraries. After filtering, the proportion of the reference genome sufficiently covered across all pool libraries averaged 70.9% (1,539,082,007 bp) (Supplemental Figure 4). The proportion of each chromosome covered across all libraries except the scaffold averaged 77.2% (Supplemental Figure 5).

Neutral Genetic Diversity and Differentiation

Pools did not differ significantly in nucleotide diversity, with the exception of the Margaret pools which exhibited lower diversity. Average nucleotide diversity was $4.39*E^{-3}$ and ranged from $3.46E^{-3}$ and $3.64E^{-3}$ (Margaret females and males respectively) to $5.05E^{-3}$ and $5.05E^{-3}$ (McNair females and males respectively (Table 1)). Forward model selection indicated that water volume in female populations (AICc = -111.05, p = $3.89E^{-4}$), and water volume and elevation in male populations (AICc = -108.59, p = $1.63E^{-2}$ & AICc = -107.58, p = $2.68E^{-2}$ respectively) were the best fit models (Supplemental Tables 3 & 4). Contrary to our hypotheses, lakes with larger water volume (habitat size) had lower nucleotide diversity,

and metrics of propagule pressure were non-significant and not selected in the final model (Figure 2). All other tested stocking and environmental variables were not important drivers of nucleotide diversity (Figure 2). Log transforming the stocking variables associated with propagule pressure did not change the results.

Examination of LOF mutations showed that heterozygosity of highly deleterious mutations was lower than heterozygosity of moderate and low impact deleterious mutations, suggesting that purifying selection may have been present post introduction (Figure 3). Like nucleotide diversity, the levels of heterozygosity based on deleterious impact categories were similar across populations except for Margaret, which had from 1.6-2.1 times lower heterozygote to homozygote ratio in different mutation categories (Figure 3). The trends observed from the stairway plot (Supplemental Figure 6), suggested a bottleneck event in the more recent generations, implying that founder effects from introduction were present (Supplemental Figure 6). Overall, these results partially supported our prediction that bottleneck events and a greater proportion of deleterious mutations would lead to lower neutral genetic variation, as there was evidence of a bottleneck event, and the lake with the least deleterious mutations had the lowest nucleotide diversity.

Average pairwise F_{ST} was $7.76*E^{-2}$ and ranged from $1.47E^{-4}$ to $1.32E^{-1}$ (Table 2). Pairwise F_{ST} was then visualized with the PCoA plot (Supplemental Figure 7), separating all lakes and sexes along the first axis explained 66.75% of the genetic structure shaped by allelic variation. While most populations were closely clustered, McNair males and females were separated along the second axis (explaining 11.48%) from each other, and in both axes from the other populations, likely due to low sample sizes (Supplemental Figure 7). The UPGMA dendrogram (Figure 3) grouped male and female pools within lakes together, illustrating small genetic distances between the sexes, with bootstrap values of 100% at all nodes. Lake Margaret appeared as an outgroup to the rest of the lakes, showing a greater genetic

distance from the other populations. Overall, these results suggested that variation in neutral diversity was stronger between lakes than between sexes within lakes, and may in part relate to founder effects.

Adaptive Genetic Differentiation

PCAdapt analysis with three principle components (*K* = 3) explained most of the variance (78.23% combined) among our pools (Supplemental Figure 8). After filtering, PCAdapt identified 41,899 outlier loci (0.289%) of 14,479,563 tested SNPs. These outlier loci and principle component groupings were retained for subsequent outlier analysis using RDA. Male and female pools of each lake grouped together in the score plot analysis, while the three distinct groupings (*K* = 3) generated using PCAdapt were Margaret, Cobb, and the remaining lakes clumped together (Figure 3).

Contrary to our hypotheses, RDA determined that the observed loci under selection were not driven by any tested habitat and environmental variables (adjusted R^2 = 2.1%, p = 0.18, with 999 permutations). Re-analysis of the RDA running VIF analysis <10 to avoid overfitting the model remained non-significant (adjusted R^2 = 1.8%, p = 0.198, with 999 permutations), when removing upstream and downstream water discharge, number of tributaries, surface area, depth, temperature variance, and number of discernable spawning sites.

Lake specific outlier analyses with CMH tests estimated from 0-17 outlier loci per pairwise population comparison (n = 9, scaffold removed), suggesting that each population was differentiated by only a few loci (≤17 loci, Table 3). Using CMH test results, we also contrasted the three, PCAdapt-based population groupings by averaging the number of candidate loci based on outliers of each population's individual comparison. The mean number of outliers in these three comparisons were 5.13 ± 2.42 (Cobb), 8 ± 4.78 (Margaret) and 7.76 ± 5.80 (remaining seven lakes; supplemental Figure 9). Of the 286 total candidate loci, 23 either had no exon associated, no results, or an uncharacterized locus. Only four of 286 outlier loci appeared in more than one lake: 1) a peptidyl-tRNA hydrolase 1 homolog, categorized

as a molecular function, associated with aminoacyl-tRNA hydrolase activity was present in Temple-Helen and Olive-Helen comparisons; 2) a spectrin beta chain, erythrocytic, categorized as a molecular function, associated with actin binding was present in Margaret-McNair and Dog- McNair comparisons; 3) a SPT16 homolog, facilitates chromatin remodeling subunit, categorized as a cellular component, associated with the FACT complex was present in Temple-Olive and Ross-Olive; and 4) an oligosaccharyltransferase complex subunit ostc, categorized as a biological process, associated with protein glycosylation was present in Temple-Olive and Mud Lake-Olive comparisons. Lake specific comparisons are found in Supplemental Table 5.

Fisher's exact test to detect sexual differentiation showed 3,403,557 (11.9%) loci changed in allele frequency between sexes. We were unable to determine if selection signals were stronger in females than males as 2,482,830 (72.9%) of these loci were significant. The loci were distributed throughout the genome (Figure 4).

Discussion

Introduction success of a species into novel environments is thought to be directly linked to genomic variation through propagule pressure and adaptation (Dlugosch et al. 2015; Lee, 2002; Sakai et al., 2001) but empirical investigations using whole genome resolution are rare (Dennenmoser et al., 2017; Yoshida et al., 2016). Our study on isolated brook trout populations introduced from a common source into alpine lakes on average ~19 (range of 7-41) generations ago suggests that the amount and nature of genome variation has had little influence on successful colonization of the species. Across 28,490,618 SNPs, we found little support for a role of propagule pressure or habitat or environmental variables in affecting neutral genetic diversity among populations. Furthermore, we did find evidence of bottleneck events associated with introduction, yet no strong evidence of purging genetic load across populations (apart from Margaret). When examining putative outlier loci, there were also only very low levels of adaptive differentiation evidenced between populations; this differentiation was inconsistent across lakes and not associated with tested environmental variables. Inconsistency across lakes could be due to other physiologically acting selection pressures not captured within the tested variables. The lack of major population differences in neutral and adaptive genomic variation is notable given that propagule pressure from stocking, habitat, and environmental conditions varied considerably among the lakes and should have promoted such variation – especially in brook trout which is well known for exhibiting rapid, adaptive differentiation at fine geographic scales in separate lakes and streams (Fraser et al. 2014; Zastavniouk et al. 2017; Ferchaud et al. 2019). While GWAS analysis of sex-differentiation revealed many putative loci under selection across all linkage groups, we also detected stronger inter-population differentiation among lakes than intra-population sex differentiation across all analyses, implying an overall low influence of sex in relation to introduction success.

As founder effects are considered impediments to introduction success, adequate propagule pressure and robust source populations may be sufficient to support colonization through increased

neutral genetic diversity (Allendorf & Lundquist, 2003; Ellstrand & Elam, 1993; Erfmeier et al., 2013, Lavergne & Molofsky, 2007; Walker et al., 2003). Contrary to previous work, we did not find that neutral genetic diversity was positively associated with propagule pressure, habitat size (lake water volume) or biotic factors such as prey availability (Bernos & Fraser, 2016; Bert et al., 2007; Briscoe Runquist et al., 2020; Lachmuth et al., 2010; Narum et al., 2017; Schindler & Parker, 2002); furthermore our populations were closed to gene flow effects and immigration. Instead, nucleotide diversity was negatively associated with water volume in both males and females, and elevation in males, and no relationships were detected for the other five variables tested. The negative trends associated with water volume were driven largely by Margaret, as it is both the largest lake and the lake with the least nucleotide diversity. We suspect that this negative relationship is driven primarily by weak propagule pressure compared to other lakes (Supplemental Table 1); when Margaret was removed, the trends remained weakly negative in females and neutral in males (Supplemental Figure 10). Similarly, the negative association between nucleotide diversity and elevation for male populations may relate to the decreased stocking effort observed with an increase in elevation (Supplemental Tables 1 & 2). Indeed, an increase in stocking effort due to high elevation and large lake size would require a proportional increase in manpower.

We also found, in partial support with our prediction, little indication that high proportions of deleterious mutations were associated with levels of neutral genetic diversity among these introduced brook trout populations. However, stairway plot analysis suggested a bottleneck event in the demographic history of the combined populations, which originated from the same hatchery source. Subsequently, population structure among our pools was weak (average pairwise F_{ST} values = $7.76*E^{-2}$; range of $1.47*E^{-4}$ to 0.133; Table 2) and most populations had similar levels of deleterious mutations. Margaret had fewer deleterious mutations than all other lakes suggesting that purifying selection may be present and purging of deleterious mutations in this population. Conversely, it may suggest a lower

level of neutral genomic heterozygosity (Chen et al., 2017), as low nucleotide diversity (Table 1) and low deleterious mutation ratio values (Figure 3) suggest a more homozygous population comparative to the other lakes. Weak population structure may be driven by the relatively short duration for population differentiation since stocking and lack of major founder effects, or the greater coverage afforded by the pool-seq methodology, as individual genotyping has a greater association with ascertainment bias and selection of highly polymorphic SNPs (Gaughran et al., 2018; Kurland et al., 2019; Malomane et al., 2018).

Despite a near-exhaustive genome-wide approach and large number of SNPs, analyses of 41,899 putative outlier loci suggested only very low levels of adaptive differentiation between populations. Using PCAdapt, more outliers distinguished Cobb and Margaret, perhaps because of a unique food source of amphipods in Cobb and interspecific competition with Westslope Cutthroat trout, lower propagule pressure, and greater habitat/spawning availability in Margaret (Collins, 2011; Martin et al., 1988; Osmond & de Mazancourt, 2013; Schindler & Parker, 2002). The pattern of Cobb and Margaret appearing most differentiated in PCAdapt analyses also mirrored their more divergent relationship from other populations in the UPGMA dendrogram, implying these two lakes may exhibit greater adaptive differentiation. However, putative outliers in the RDA did not relate to the abiotic and biotic variables tested and may not have been relevant as they did not correspond to outliers from the pairwise – based CMH tests. Of the 28,490,618 SNPs in our dataset, there were only a combined total of 286 putative candidate loci across all comparisons with CMH tests, representing 0.001%. Of these loci, there were only four duplicate candidate genes suggesting non-parallel evolution. However, gene ontology searches confirmed that although different at a molecular level, the observed candidate loci had similar functions (Supplemental Table 6). The discontinuity between CMH, PCAdapt, and RDA tests may suggest that (i) candidate genes highlighted by the CMH tests have a polygenic element associated with important biological processes acted on by selection, and/or that (ii) some detected outlier loci are associated with

additional, untested environmental variables or biological processes. We also acknowledge that there is currently no annotated genome for brook trout; our annotation to the Arctic charr genome may have excluded important genetic components despite strong resolution.

Across all our tests for both neutral diversity and putative adaptive differentiation, we did not find that females had more adaptive differentiation than males (Zastavniouk et al., 2017). We found that sexes of our populations grouped together (Figures 3), suggesting that there is more among-population differentiation regardless of sex than there is among sexes within each population. Our GWAS analysis determined 2,482,830 putative candidate loci potentially responsible for sexual differentiation. Unfortunately, the results did not highlight specific chromosomes or regions under selection, therefore we were unable to confirm our hypotheses of outlier loci being associated with specific sexes. We suspect that the lack of specificity in our results is due to the sensitivity of a GWAS approach in a pool-seq study on natural populations. Specifically, hidden population structure, heterogenous sampling within lakes, and insufficient power can create noise within the results (Stram, 2014). Future whole genome approaches examining sex differentiation in species introductions would benefit from an experimental design tailored to examining this aspect in detail, to allow for specific within- and among-population sex comparisons.

Overall, several biological explanations may explain the low and inconsistent level of adaptive differentiation observed among the populations. First, although the mountain lake habitats in this study have distinguishing features from the native range of brook trout that should foster adaptive differentiation (Bernos & Fraser, 2016; Harbicht et al., 2014; Hecht et al., 2015; Krueger & May, 1991; Power, 1980; Rieman & Allendorf, 2001; Schindler & Parker, 2002), perhaps mountain lakes do not ultimately constitute a sufficiently novel environment for the species. Second, ~50 years since introduction (average ~19 generations) may not be enough time to generate stronger adaptive differentiation, though there is evidence that this can happen in related species (e.g. Hendry et al.,

2000). Third, in their native range, brook trout are well-known for exhibiting high levels of phenotypic plasticity and being effective colonizers of small headwater stream habitats (Spens et al., 2007; Wood et al., 2015; Wood & Fraser, 2015; Yates et al., 2019). These attributes may allow brook trout to successfully colonize new environments without requiring a process of strong adaptive differentiation. Fourth, as the study lakes all occur in a similar geological area (alpine environments), the environmental contrasts between them might still be too small to induce major adaptive changes. Finally, nuances in population histories from stocking events and subsequent establishment may generate population-specific idiosyncrasies in neutral and adaptive diversity, concurrent with plasticity to novel environments.

Conclusions

Our analysis of 28,490,618 SNPs facilitated greater resolution for examining the roles of genetic and environmental factors in colonization of introduced species. Understanding the underlying factors that contribute to successful species colonization is crucial for applications in conservation, mitigating effects on endangered species, and population maintenance (Adams et al., 2000; Higgins & Zanden, 2010; Lodge, 1993). Specifically, in our study, environmental factors and propagule pressure related to fish stocking had very little independent effect on colonization success, and wide ranges in both factors did not lead to significant genetic variation among populations. Moreover, population differentiation and signals of local adaptation were not stronger in conditions expected to promote them. Furthermore, we show a large amount of differentiation between sexes among lakes, contrasted by very weak differentiation of sexes within lakes; suggesting that sex specific introduction procedures may not be as important as previously thought. Additionally, our work suggests that stocking and environmental predictors of neutral genetic diversity are not mutually exclusive and should be considered together. To better understand the neutral and adaptive differentiation of introduced species, we encourage future

analyses to use whole genome approaches across a greater range of sample sites that have had a longer time period from introduction to accumulate adaptive differentiation.

Figures

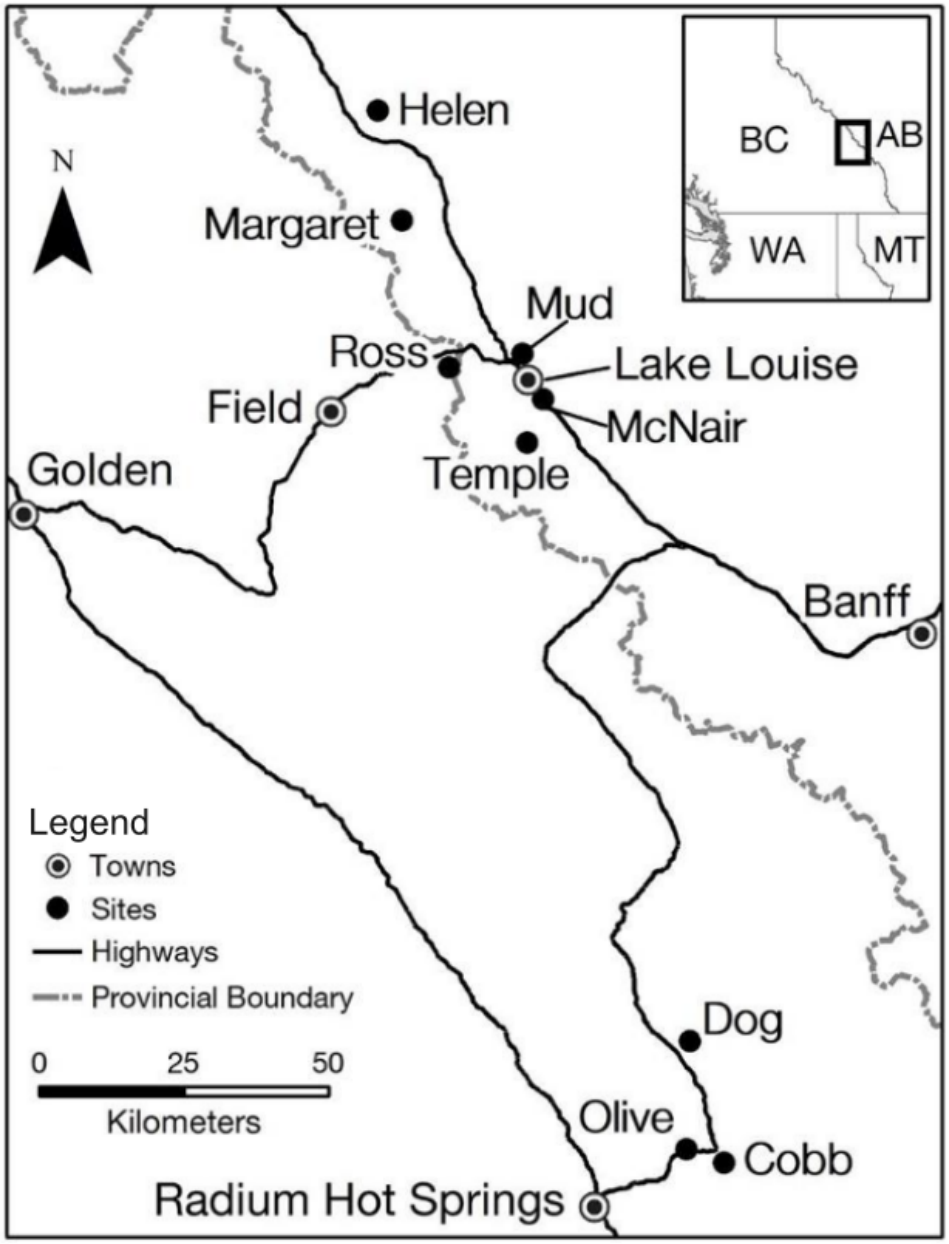


Figure 1. Map of sampled lakes for brook trout in their introduced range of Alberta and British Columbia, Canada.

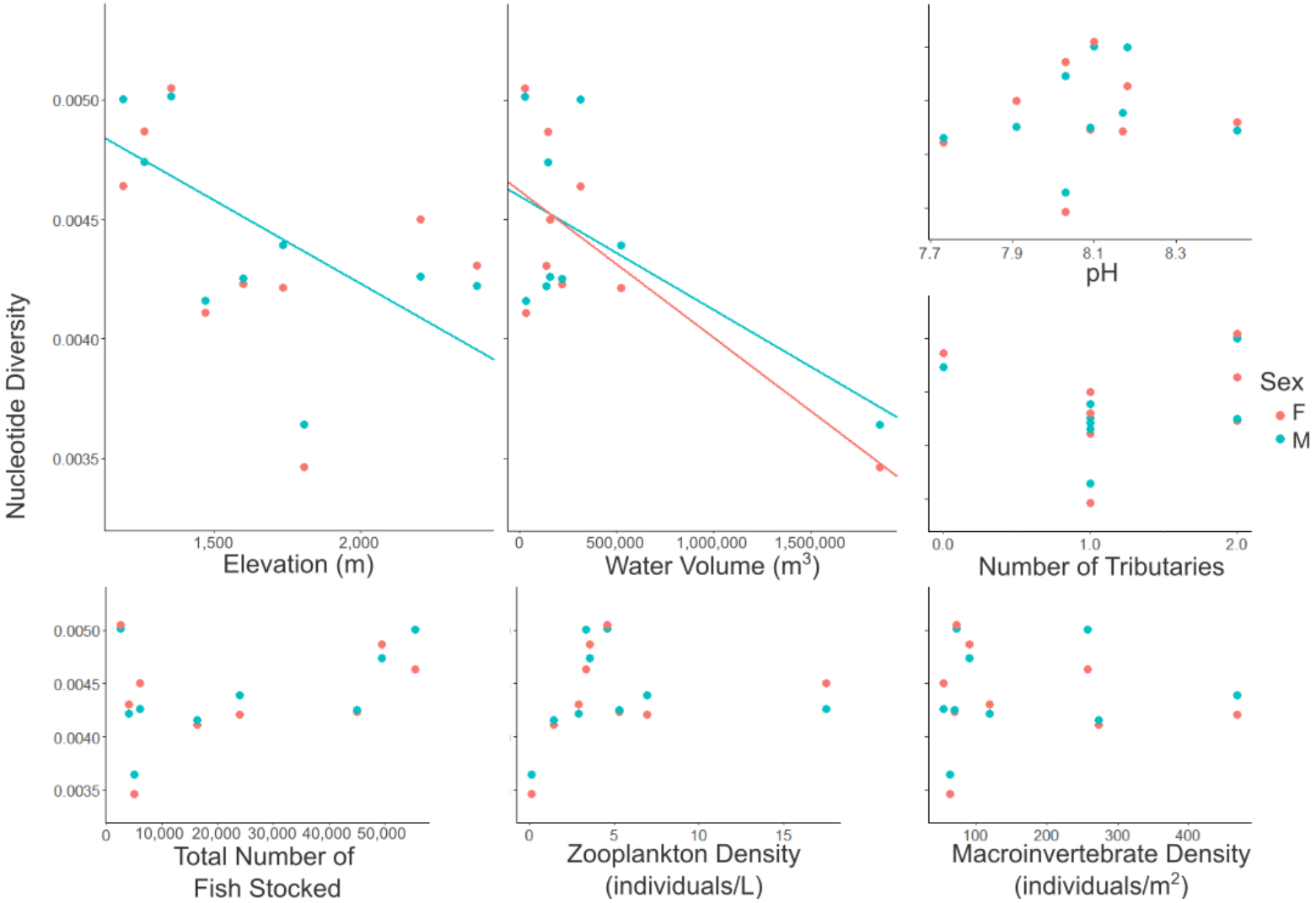


Figure 2. Regressions analyses of nucleotide diversity against tested non collinear variables. Water volume in female populations and water volume and elevation in male populations were significant and selected as the best fit model (trend lines).

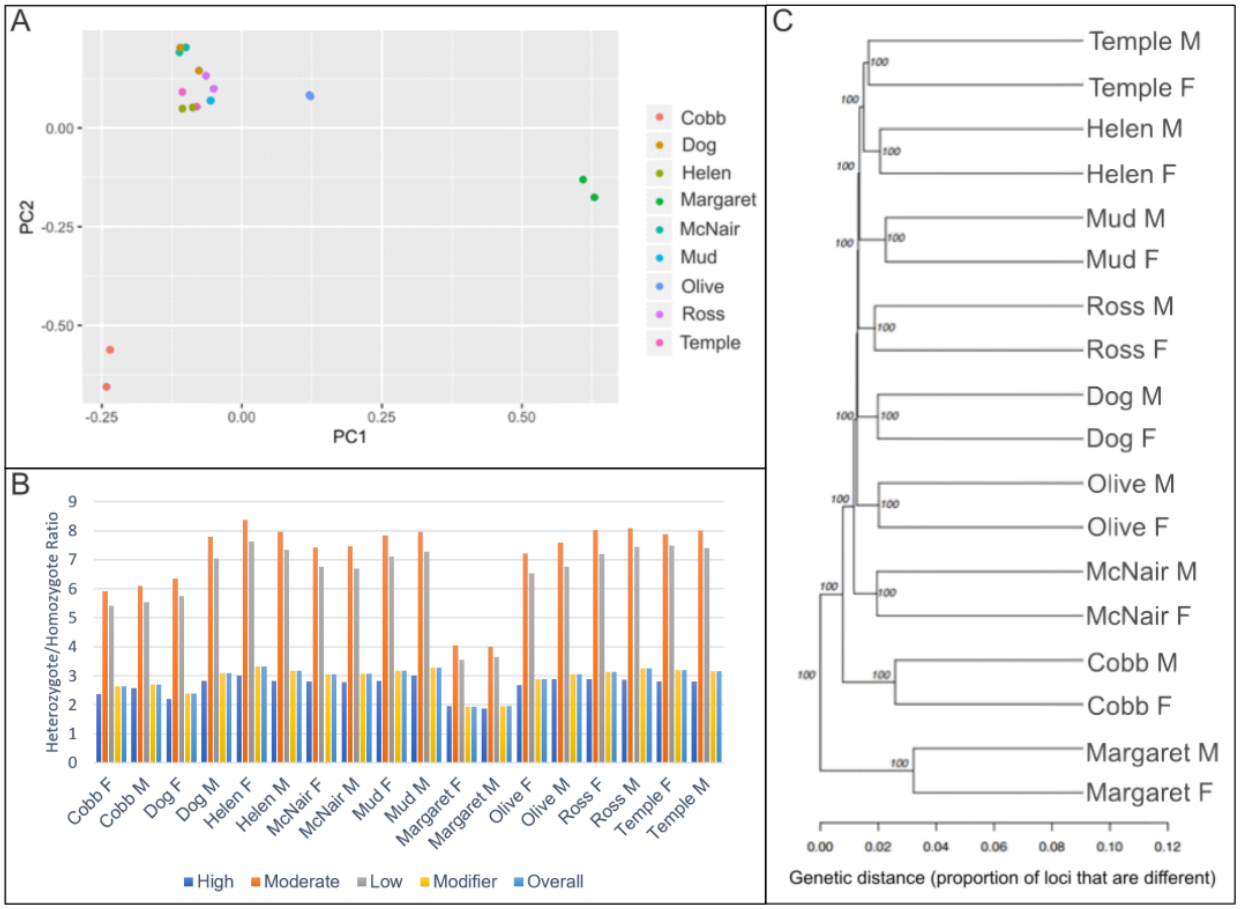


Figure 3. Genetic differentiation of introduced brook trout populations by sex. A) score plot analysis from PCAdapt showing *K*=3; B) Bar graph of the proportion of SNPs among four categories of deleterious effect; C) Unweighted Pair Group Method with Arithmetic Mean (UPGMA) Tree (dendrogram) of all 18 pools (M = male, F= female), with 100% bootstrap values at the nodes; showing Margaret as an outgroup. Male and female groupings can be seen at a population level in all three analyses.

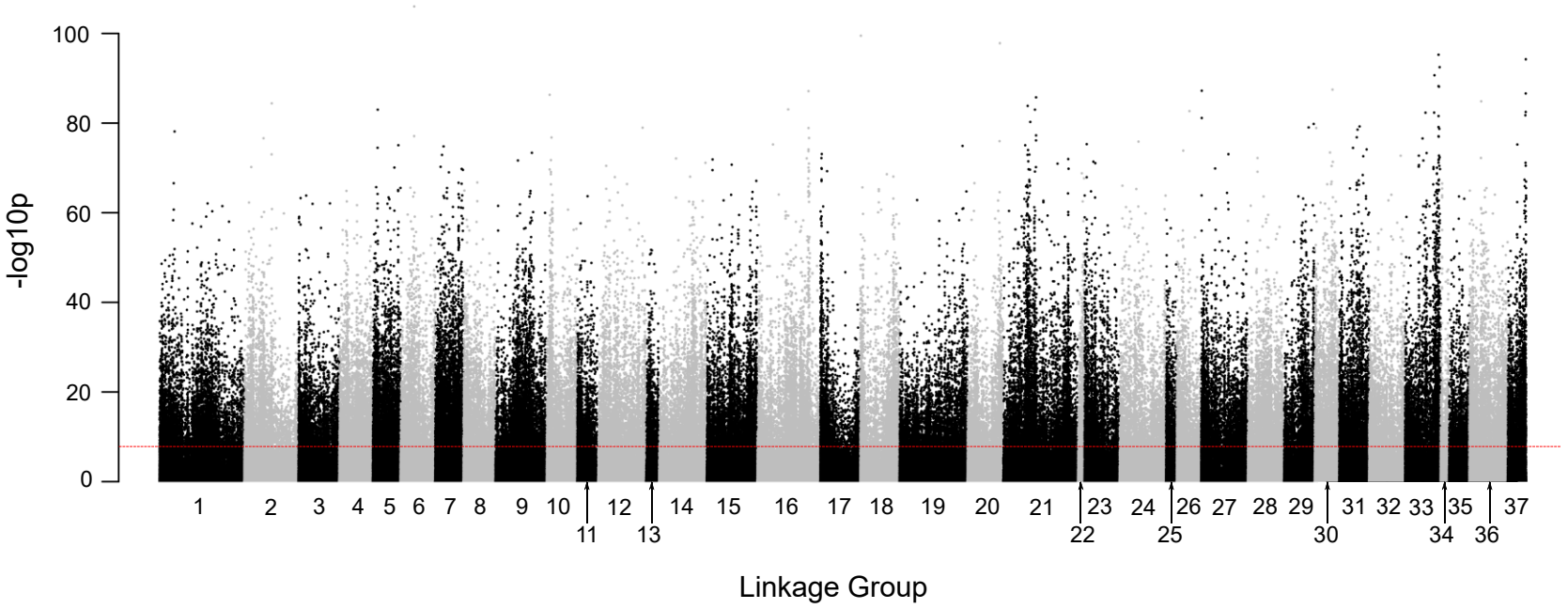


Figure 4. GWAS Manhattan plot comparing male and female introduced brook trout populations with a threshold of 0.05.

Tables

Table 1. Nucleotide diversity (π) and F_{ST} between sex of all nine filtered, biallelic populations. Asterisks denote $p < 0.05$.

Populations	π Male	π Female	F_{ST} between sex
Cobb	0.00474	0.00487	0.00301*
McNair	0.00501	0.00505	0.13243
Dog	0.00500	0.00464	0.04432*
Helen	0.00422	0.00431	0.04536*
Margaret	0.00364	0.00347	0.04340*
Temple	0.00426	0.00450	0.00828*
Mud	0.00425	0.00423	0.03356*
Olive	0.00411	0.00411	0.03011*
Ross	0.00439	0.00421	0.00593*

Table 2. Population differentiation of introduced brook trout populations. Pairwise sliding-window fixation indexes (F_{ST}) between all filtered, biallelic male populations (top left), female populations (bottom left), and male-female comparison within each lake (right). Asterisks denote p < 0.05.

Population (M)	Cobb	McNair	Dog	Helen	Margaret	Temple	Mud	Olive	Ross
Cobb	-	0.04516*	0.00269*	0.01458*	0.00852*	0.00693*	0.01336*	0.00073*	0.00495*
McNair		-	0.07152	0.07084	0.07883	0.05366	0.08314	0.08008	0.08166
Dog			-	0.02019*	0.02702*	0.01022*	0.03591*	0.02446*	0.02837*
Helen				-	0.02814*	0.01341*	0.03663*	0.02429*	0.02743*
Margaret					-	0.02705*	0.03785*	0.03069*	0.03302*
Temple						-	0.03350*	0.01799*	0.02210*
Mud							-	0.03876*	0.03857*
Olive								-	0.04469*
Ross									-
Population (F)	**Cobb**	**McNair**	**Dog**	**Helen**	**Margaret**	**Temple**	**Mud**	**Olive**	**Ross**
Cobb	-	0.06973	0.01176*	0.01603*	0.01413*	0.04330*	0.00052*	0.01870*	0.02068*
McNair		-	0.07293	0.07616	0.07586	0.05694	0.06741	0.07757	0.04137*
Dog			-	0.02894*	0.02728*	0.00029*	0.01074*	0.02606*	0.01318*
Helen				-	0.03257*	0.00527*	0.01990*	0.03122*	0.01413*
Margaret					-	0.00348*	0.02067*	0.03107*	0.00995*
Temple						-	0.01408*	0.00763*	0.04080*
Mud							-	0.02225*	0.03003*
Olive								-	0.00543*
Ross									-

Table 3. Outlier loci amoung introduced brook trout populations determined by independent pairwise CMH analysis (upper), against the number of SNPs in each pairwise analysis (lower).

Lake Name	Cobb	Margaret	Olive	Helen	Dog	Ross	Temple	McNair	Mud
Cobb	-	3	8	4	6	3	7	8	2
Margaret	7,659,675	-	10	13	13	10	10	5	0
Olive	7,271,316	7,025,721	-	15	14	10	14	1	1
Helen	7,318,305	7,068,665	6,838,734	-	11	13	10	6	0
Dog	6,531,154	6,450,080	6,308,517	6,311,949	-	11	14	10	1
Ross	7,636,065	7,385,133	7,041,021	7,132,435	6,469,102	-	17	8	0
Temple	6,951,920	6,840,064	6,619,100	6,653,555	6,163,825	6,892,851	-	4	1
McNair	8,246,277	7,728,579	7,347,432	7,456,360	6,692,374	7,820,175	7,152,265	-	2
Mud	7,439,413	7,052,217	6,727,633	6,833,011	6,204,908	7,165,096	6,541,214	7,483,119	-

References

Adams, S. B., Frissell, C. A., & Rieman, B. E. (2000). Movements of Nonnative Brook Trout in Relation to Stream Channel Slope. *Transactions of the American Fisheries Society*, *129*(3), 623–638.

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., … Zbicz, K. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(1), 8–13.

Allendorf, F. W., & Lundquist, L. L. (2003). Introduction: Population Biology, Evolution, and Control of Invasive Species. *Conservation Biology*, *17*(1), 24–30.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data.

Arismendi, I., Penaluna, B. E., Dunham, J. B., García De Leaniz, C., Soto, D., Fleming, I. A., … Vargas, P. V. (2014). Differential invasion success of salmonids in southern Chile: patterns and hypotheses. *Reviews in Fish Biology and Fisheries*, *24*, 919–941.

Bagheri, A., & Midi, H. (2009). Robust estimations as a remedy for multicollinearity caused by multiple high leverage points. *Journal of Mathematics and Statistics*, *5*(4), 311–321.

Barton, K., & Barton, M. K. (2019). Package 'MuMIn'. R package version, *1*(6).

Benjamin, J. R., Dunham, J. B., & Dare, M. R. (2007). Invasion by nonnative brook trout in Panther Creek, Idaho: Roles of local habitat quality, biotic resistance, and connectivity to source habitats. *Transactions of the American Fisheries Society*, *136*(4), 875–888.

Bernos, T. A., & Fraser, D. J. (2016). Spatiotemporal relationship between adult census size and genetic

population size across a wide population size gradient. *Molecular Ecology*, *25*(18), 4472–4487.

Bert, T. M., Crawford, C. R., Tringali, M. D., Seyoum, S., Galvin, J. L., Higham, M., & Lund, C. (2007). Genetic management of hatchery-based stock enhancement. *Ecological and Genetic Implications of Aquaculture Activities*, 123–174.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. (2009). QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics*, *25*(22), 3045–3046.

Briscoe Runquist, R. D., Gorton, A. J., Yoder, J. B., Deacon, N. J., Grossman, J. J., Kothari, S., … Moeller, D. A. (2020). Context dependence of local adaptation to abiotic and biotic environments: A quantitative and qualitative synthesis. *The American Naturalist*, *195*(3), 412–431.

Buniello, A., Macarthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., … Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(1), 1005–1012.

Bushnell, B. (2014). BBMap: a fast, accurate, splice-aware aligner (No. LBNL-7065E). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Carroll, S. P., Dingle, H., Famula, T. R., & Fox, C. W. (2001). Genetic architecture of adaptive differentiation in evolving host races of the soapberry bug, Jadera haematoloma. *Genetica*, *112–113*, 257–272.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution*, *34*(6), 1417–1428.

Chevin, L.-M., & Lande, R. (2011). Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology*, *24*(7), 1462–1476.

Christensen, K. A., Rondeau, E. B., Minkley, D. R., Leong, J. S., Nugent, C. M., Danzmann, R. G., … Koop, B. F. (2018). The arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLoS ONE*, *13*(9).

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92.

Collins, S. (2011). Competition limits adaptation and productivity in a photosynthetic alga at elevated CO_2. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1703), 247–255.

Connallon, T., & Hall, M. D. (2016). Genetic correlations and sex-specific adaptation in changing environments. *Evolution*, *70*(10), 2186–2198.

Coulson, T., Kendall, B. E., Barthold, J., Plard, F., Schindler, S., Ozgul, A., & Gaillard, J. M. (2017). Modeling adaptive and nonadaptive responses of populations to environmental change. *American Naturalist*, *190*(3), 313–336.

Cox, R. M., & Calsbeek, R. (2009). Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *American Naturalist*, *173*(2), 176–187.

Cribari-Neto, F., Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software, 34(*2), 1–24.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510.

Dennenmoser, S., Vamosi, S. M., Nolte, A. W., & Rogers, S. M. (2017). Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (*Cottus asper*) revealed by Pool-Seq. *Molecular Ecology*, *26*(1), 25–42.

Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, *17,* 431–449.

Dlugosch, Katrina M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, *24*(9), 2095–2111.

Donald, D. B., & Alger, D. J. (1984). *Limnological studies in Kootenay national park part one: the lakes*. Canadian Wildlife Service.

Dunham, J. B., Adams, S. B., Schroeter, R. E., & Novinger, D. C. (2002). Alien invasions in aquatic ecosystems: Toward an understanding of brook trout invasions and potential impacts on inland cutthroat trout in western North America. *Reviews in Fish Biology and Fisheries*, *12*, 373–391.

Earle, J. E., Stelfox, J. D., & Meagher, B. E. (2007). Quirk Creek brook trout suppression project – 2004 to 2006. Unpublished report, Fish and Wildlife Division, Alberta Sustainable Resource Development, Cochrane, Alberta.

EasySFS. (2020). https://github.com/isaacovercast/easySFS#easysfs

Ellstrand, N. C., & Elam, D. R. (1993). Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics*, *24*(1), 217–242.

Endler, J. A. (1986). Natural selection in the wild. Princeton University Press.

Erfmeier, A., Hantsch, L., & Bruelheide, H. (2013). The role of propagule pressure, genetic diversity and

microsite availability for Senecio vernalis invasion. *PloS One*, *8*(2), e57029.

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.

Facon, B., Genton, B. J., Shykoff, J., Jarne, P., Estoup, A., & David, P. (2006). A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology and Evolution, 21*(3), 130–135.

Fassnacht, S., Venable, N., McGrath, D., & Patterson, G. (2018). Sub-seasonal snowpack trends in the Rocky Mountain National Park Area, Colorado, USA. *Water*, *10*(5), 562.

Fausch, K. D. (2007). Introduction, establishment and effects of non-native salmonids: considering the risk of rainbow trout invasion in the United Kingdom. *Journal of Fish Biology*, *71*, 1–32.

Faust, G. G., & Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*, *30*(17), 2503–2505.

Ferchaud, A.-L., Leitwein, M., Laporte, M., Boivin-Delisle, D., Bougas, B., Hernandez, C., … Bernatchez, L. (2019). Adaptive and maladaptive genetic diversity in small populations; insights from the Brook Charr (*Salvelinus fontinalis*) case study. *BioRxiv*, 660621.

Filchak, K. E., Roethele, J. B., & Feder, J. L. (2000). Natural selection and sympatric divergence in the apple maggot Rhagoletis pomonella. *Nature*, *407*(6805), 739–742.

Fisheries and Oceans Canada. (2014). Recovery Strategy for the Alberta populations of Westslope Cutthroat Trout (*Oncorhynchus clarkii lewisi*) in Canada [Final]. Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa. iv + 28 pp + Appendices.

Frachon, L., Mayjonade, B., Bartoli, C., Hautekèete, N.-C., & Roux, F. (2019). Adaptation to plant communities across the genome of *Arabidopsis thaliana*. *Molecular Biology and Evolution*, *36*(7),

1442–1456.

Fraser, D. J., Debes, P. V., Bernatchez, L., & Hutchings, J. A. (2014). Population size, habitat fragmentation, and the nature of adaptive variation in a stream fish. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1790), 20140370.

Fraser, Dylan J., Walker, L., Yates, M. C., Marin, K., Wood, J. L. A., Bernos, T. A., & Zastavniouk, C. (2019). Population correlates of rapid captive-induced maladaptation in a wild fish. *Evolutionary Applications*, *12*(7), 1305–1317.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression, Third edition*. Sage, Thousand Oaks CA.

Garrison, E. (2016). Vcflib, a simple C++ library for parsing and manipulating VCF files.

Gaughran, S. J., Quinzin, M. C., Miller, J. M., Garrick, R. C., Edwards, D. L., Russello, M. A., … Caccone, A. (2018). Theory, practice, and conservation in the age of genomics: The Galápagos giant tortoise as a case study. *Evolutionary Applications*, *11*(7), 1084–1093.

Halling, C., Wikström, S. A., Lilliesköld-Sjöö, G., Mörk, E., Lundsør, E., & Zuccarello, G. C. (2013). Introduction of Asian strains and low genetic variation in farmed seaweeds: Indications for new management practices. *Journal of Applied Phycology, 25*(1), 89–95.

Hamilton, J. A., Okada, M., Korves, T., & Schmitt, J. (2015). The role of climate adaptation in colonization success in *Arabidopsis thaliana*. *Molecular Ecology*, *24*(9), 2253–2263.

Harbicht, A. B., Alshamlih, M., Wilson, C. C., & Fraser, D. J. (2014). Anthropogenic and habitat correlates of hybridization between hatchery and wild brook trout. *Canadian Journal of Fisheries and Aqutic Sciences, 71*(5): 668–697.

Hayes, K. R., & Barry, S. C. (2008). Are there any consistent predictors of invasion success? *Biological Invasions*, *10*(4), 483–506.

Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology*, *24*(22), 5573–5595.

Hendry, A. P., Wenburg, J. K., Bentzen, P., Volk, E. C., & Quinn, T. P. (2000). Rapid evolution of reproductive isolation in the wild: Evidence from introduced salmon. *Science, 290*(5491), 516–518.

Higgins, S. N., & Zanden, M. J. Vander. (2010). What a difference a species makes: a meta–analysis of dreissenid mussel impacts on freshwater ecosystems. *Ecological Monographs*, *80*(2), 179–196.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. *Genetics*, *210*(1), 315–330.

Iossa, G. (2019, June 1). Sex-specific differences in thermal fertility limits. *Trends in Ecology and Evolution*, *34*, 490–492.

Kamvar, Z. N., Tabima, J. F., Grünwald, N. J. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ, 2*, e281.

Kamvar, Z. N., Brooks, J. C., Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet*., 6, 208.

Kelly, D. W., Muirhead, J. R., Heath, D. D., & Macisaac, H. J. (2006). Contrasting patterns in genetic diversity following multiple invasions of fresh and brackish waters. *Molecular Ecology*, *15*(12), 3641–3653.

Kinnison, M. T., Unwin, M. J., & Quinn, T. P. (2008). Eco-evolutionary vs. habitat contributions to

invasion in salmon: experimental evaluation in the wild. *Molecular Ecology*, *17*(1), 405–414.

Knaus, B. J., Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.

Knaus, B. J., Grünwald, N. J. (2016). VcfR: an R package to manipulate and visualize VCF format data. *BioRxiv*.

Kofler, R., Pandey, R. V., & Schlotterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436.

Kofler, Robert, Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., … Schlötterer, C. (2011). PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE*, *6*(1), e15925.

Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature, 431*(7005), 177–181.

Krueger, C. C., & May, B. (1991). Ecological and Genetic Effects of Salmonid Introductions in North America. *Canadian Journal of Fisheries and Aquatic Sciences*, *48*(S1), 66–77.

Kurland, S., Wheat, C. W., Paz Celorio Mancera, M., Kutschera, V. E., Hill, J., Andersson, A., … Laikre, L. (2019). Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. *Ecology and Evolution*, *9*(19), 11448–11463.

Lachmuth, S., Druka, W., & Schurr, F. M. (2010). The making of a rapid plant invader: genetic diversity and differentiation in the native and invaded range of Senecio inaequidens. *Molecular Ecology*, *19*(18), 3952–3967.

Lasne, C., Hangartner, S. B., Connallon, T., & Sgrò, C. M. (2018). Cross-sex genetic correlations and the evolution of sex-specific local adaptation: Insights from classical trait clines in *Drosophila melanogaster*. *Evolution*, *72*(6), 1317–1327.

Launey, S., & Hedgecock, D. (2001). High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics, 159*(1), 255-265.

Lavergne, S., & Molofsky, J. (2007). Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(10), 3883–3888.

Lecomte, F., Beall, E., Chat, J., Davaine, P., & Gaudin, P. (2013). The complete history of salmonid introductions in the Kerguelen Islands, Southern Ocean. *Polar Biology*, *36*(4), 457–475.

Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, *17*(8), 386–391.

Legendre, P., Oksanen, J., & ter Braak, C. J. F. (2010). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, *2*(3), 269–277.

Létourneau, J., Ferchaud, A.-L., Le Luyer, J., Laporte, M., Garant, D., & Bernatchez, L. (2018). Predicting the genetic impact of stocking in Brook Charr (*Salvelinus fontinalis*) by combining RAD sequencing and modeling of explanatory variables. *Evolutionary Applications*, *11*(5), 577–592.

Li, H., & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Liu, X., & Fu, Y. X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, *47*(5), 555–559.

Lodge, D. M. (1993). Biological invasions: Lessons for ecology. *Trends in Ecology & Evolution*, *8*(4), 133–137.

Long, T. A., & Rice, W. R. (2007). Adult locomotory activity mediates intralocus sexual conflict in a laboratory-adapted population of Drosophila melanogaster. *Proceedings of the Royal Society B. Biological sciences, 274*(1629), 3105–3112.

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

Luu, K., Bazin, E., & Blum, M. G. B. (2017). *pcadapt* : an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77.

Malomane, D. K., Reimer, C., Weigend, S., Weigend, A., Sharifi, A. R., & Simianer, H. (2018). Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. *BMC Genomics*, *19*(1), 22.

Martin, M. J., Pérez-Tomé, J. M., & Toro, M. A. (1988). Competition and genotypic variability in drosophila melanogaster. *Heredity*, *60*(1), 119–123.

Micheletti, S. J., & Narum, S. R. (2018). Utility of pooled sequencing for association mapping in nonmodel organisms. *Molecular Ecology Resources*, *18*(4), 825–837.

Moyle, P. B., & Marchetti, M. P. (2006). Predicting invasion success: Freshwater fishes in California as a model. *BioScience*, *56*(6), 515.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-

sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841–2847.

Narum, S. R., Gallardo, P., Correa, C., Matala, A., Hasselman, D., Sutherland, B. J. G., & Bernatchez, L. (2017). Genomic patterns of diversity and divergence of two introduced salmonid species in Patagonia, South America. *Evolutionary Applications*, *10*(4), 402–416.

Neter, J., Kutner, M. H., Wasserman, W., & Nachtsheim, C. J. (2004). *Applied Linear Regression Models.* Third Edition., MacGraw-Hill, New York.

Neville, H., Dunham, J., Rosenberger, A., Umek, J., & Nelson, B. (2009). Influences of wildfire, habitat size, and connectivity on trout in headwater streams revealed by patterns of genetic diversity. *Transactions of the American Fisheries Society*, *138*(6), 1314–1327.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Oksanen, M. J. (2013). Package 'vegan'. Community ecology package, version, *2*(9), 1-295.

Osmond, M. M., & de Mazancourt, C. (2013). How competition affects evolutionary rescue. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1610).

Paradis, E., Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526-528.

Parks Canada (n.d.). National Parks Stocking Records

Power, G. (1980). The Brook Charr, Salvelinus fontinalis. In *E. K. Balon's Charrs, salmonid fishes of the genus Salvelinus*, Kluwer Boston. 141-203.

Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, *13*(6), 288–294.

Privé, F., Luu, K., Vilhjálmsson, B. J., Blum, M. G. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution, 37*(7), 2153-2154.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Revelle, W. (2019). psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois. R package version 1.9.12.

Rieman, B. E., & Allendorf, F. W. (2001). Effective population size and genetic conservation criteria for Bull Trout. *North American Journal of Fisheries Management*, *21*, 756–764.

Roberts, D. W., David, M., & Roberts, W. (2019). Package "labdsv" Ordination and Multivariate Analysis for Ecology version 2.0-1.

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., … Weller, S. G. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics*, *32*(1), 305–332.

Schindler, D. W., & Parker, B. R. (2002). Biological pollutants: alien fishes in mountain lakes. *Water, Air, & Soil Pollution: Focus*, *2*(2), 379–397.

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763.

Spens, J., Alanärä, A., & Eriksson, L. O. (2007). Nonnative brook trout (*Salvelinus fontinalis*) and the demise of native brown trout (*Salmo trutta*) in northern boreal lakes: Stealthy, long-term patterns? *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(4), 654–664.

Sprogis, K. R., Christiansen, F., Raudino, H. C., Kobryn, H. T., Wells, R. S., & Bejder, L. (2018). Sex-specific differences in the seasonal habitat use of a coastal dolphin population. *Biodiversity and Conservation*, *27*(14), 3637–3656.

Stanford, B. (2019). Conservation genomics of the endangered Banff Springs Snail (*Physella johnsoni*) using Pool-seq. (Unpublished master's thesis). University of Calgary, Calgary, AB.

Stéphane Dray, A., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., … Wagner, H. H. (2020). Package "adespatial" Multivariate Multiscale Spatial Analysis.

Stillwell, C. R., & Fox, C. W. (2009). Geographic variation in body size, sexual size dimorphism and fitness components of a seed beetle: local adaptation versus phenotypic plasticity. *Oikos*, *118*(5), 703–712.

Stram, D. O. (2014). *Design, analysis, and interpretation of genome-wide association scans 15*, 334. New York, NY, USA: Springer.

Svensson, E. I., Goedert, D., Gómez-Llano, M. A., Spagopoulou, F., Nava-Bolaños, A., & Booksmythe, I. (2019). Sex differences in local adaptation: What can we learn from reciprocal transplant experiments? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*.

Terekhanova, N. V, Barmintseva, A. E., Kondrashov, A. S., Bazykin, G. A., & Mugue, N. S. (2019). Architecture of parallel adaptation in ten lacustrine Threespine Stickleback populations from the White Sea Area. *Genome Biology and Evolution*, *11*(9), 2605–2618.

Thompson, P. D., & Rahel, F. J. (1998). Evaluation of artificial barriers in small Rocky Mountain streams for preventing the upstream movement of Brook Trout. *North American Journal of Fisheries Management*, *18*(1), 206–210.

Tsutsui, N. D., Suarez, A. V., Holway, D. A., & Case, T. J. (2000). Reduced genetic variation and the

success of an invasive species. *Proceedings of the National Academy of Sciences of the United States of America, 97*(11), 5948–5953.

Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, *39*(3), 505–522.

Vigliano, P. H., Alonso, M. F., & Aquaculture, M. (2007). Salmonid introductions in Patagonia: A mixed blessing. In *Ecological and Genetic Implications of Aquaculture Activities*, 315–331. Springer, Dordrecht.

Walker, N. F., Hulme, P. E., & Hoelzel, A. R. (2003). Population genetics of an invasive species, *Heracleum mantegazzianum*: implications for the role of life history, demographics and independent introductions. *Molecular Ecology*, *12*(7), 1747–1756.

Warren, R. J., Bahn, V., & Bradford, M. A. (2012). The interaction between propagule pressure, habitat suitability and density-dependent reproduction in species invasion. *Oikos*, *121*(6), 874–881.

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer.

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.

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.

Wood, J. L. A., Tezel, D., Joyal, D., & Fraser, D. J. (2015). Population size is weakly related to quantitative genetic variation and trait differentiation in a stream fish. *Evolution*, *69*(9), 2303–2318.

Yates, M. C., Glaser, D., Post, J., Cristescu, M. E., Fraser, D. J., & Derry, A. M. (2020). The relationship between eDNA particle concentration and organism abundance in nature is strengthened by

allometric scaling. *Molecular Ecology*.

Yates, M. C., Bowles, E., & Fraser, D. J. (2019). Small population size and low genomic diversity have no effect on fitness in experimental translocations of a wild fish. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1916), 20191989.

Yoshida, K., Miyagi, R., Mori, S., Takahashi, A., Makino, T., Toyoda, A., … Kitano, J. (2016). Whole-genome sequencing reveals small genomic regions of introgression in an introduced crater lake population of threespine stickleback. *Ecology and Evolution*, *6*(7), 2190–2204.

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.

Supplemental Figures and Tables

***Supplemental Table 1*.** Parks Canada stocking information summarized by lake from 1941 - 1973.

Lake Name	First Year Stocked	Final Year Stocked	Number of Years Stocked	Total Number of Fish Stocked	Mean Number of Fish Stocked per Event (range)	Census Size	Distance from Nearest Road (m)
Cobb	1947	1973	19	49,500	2,475 (200-7,000)	95	2,700
Dog	1941	1972	16	55,500	3,468 (500-9,000)	2,140	2,600
Helen	1964	1965	2	4,000	2,000	1,037	6,000
Margaret	1963	1963	1	5,000	5,000	1,720	5,070
McNair	1963	1968	6	2,500	416 (250-1,000)	269	5
Mud	1954	1968	10	45,050	4,505 (800-11,000)	1,431	1,600
Olive	1947	1972	22	16,250	706 (200-2,500)	2,416	5
Ross	1954	1967	7	24,000	3,428 (1,000-6,000)	2,053	3,200
Temple	1964	1968	3	6,000	2,000	2,650	2,000

Supplemental Table 2. Summarized environmental information by lake 2017.

Lake Name	Surface Area (ha)	Water Volume (m^3)	Depth (m)	Elevation (m)	pH	Mean Seasonal Temperature (°C)	Temperature Variance (°C)	Number of Tributaries
Cobb	2.29	149,000	8	1260	8.03	16.7	4.3	0
Dog	11.5	316,000	3.8	1185	8.18	15.3	9.8	2
Helen	2.48	139,621	15	2400	8.45	11.5	4.8	1
Margaret	18	1,851,206	28.2	1808	8.03	11.7	3.0	1
McNair	1.66	30,135	4	1352	8.1	11.6	9.2	2
Mud	7.2	217,351	7.2	1600	8.09	13.2	9.2	2
Olive	1.66	32,833	3.6	1470	7.73	10.6	6.5	1
Ross	6.61	525,347	21.5	1735	8.17	10.0	4.8	1
Temple	3.25	156,332	14.7	2207	7.91	10.5	6.1	1

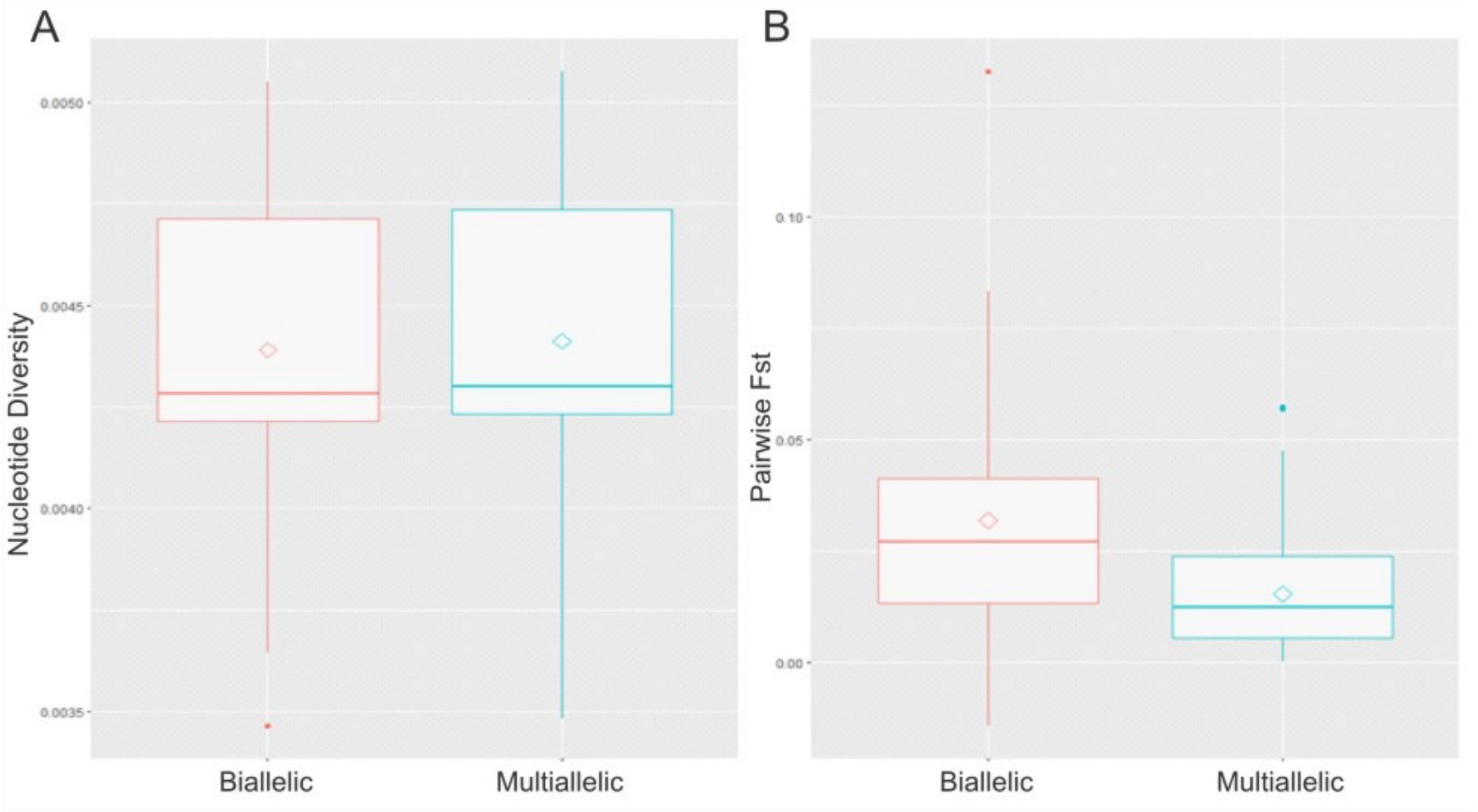
Lake Name	Upstream Discharge (m^3/s)	Downstream Discharge (m^3/s)	Number of Discernable Spawning Sites	Jaccard Species Dissimilarity Index	Zooplankton Density (individuals/L)	Macroinvertebrate Density (individuals/m^2)
Cobb	0	0	0	0.40	3.52	90.53
Dog	254	538	2	0.88	3.31	257.19
Helen	78	186	4	0.40	2.90	118.54
Margaret	4860	5364	2	0.72	0.12	62.42
McNair	1054	974	2	0.79	4.59	72.38
Mud	32	310	3	0.76	5.30	69.41
Olive	72	355	2	0.40	1.40	272.74
Ross	509	6251	4	0.40	6.93	468.75
Temple	0	501	2	0.40	17.50	53.16

***Supplemental Table 3*.** All beta-regression summaries for variable terms.

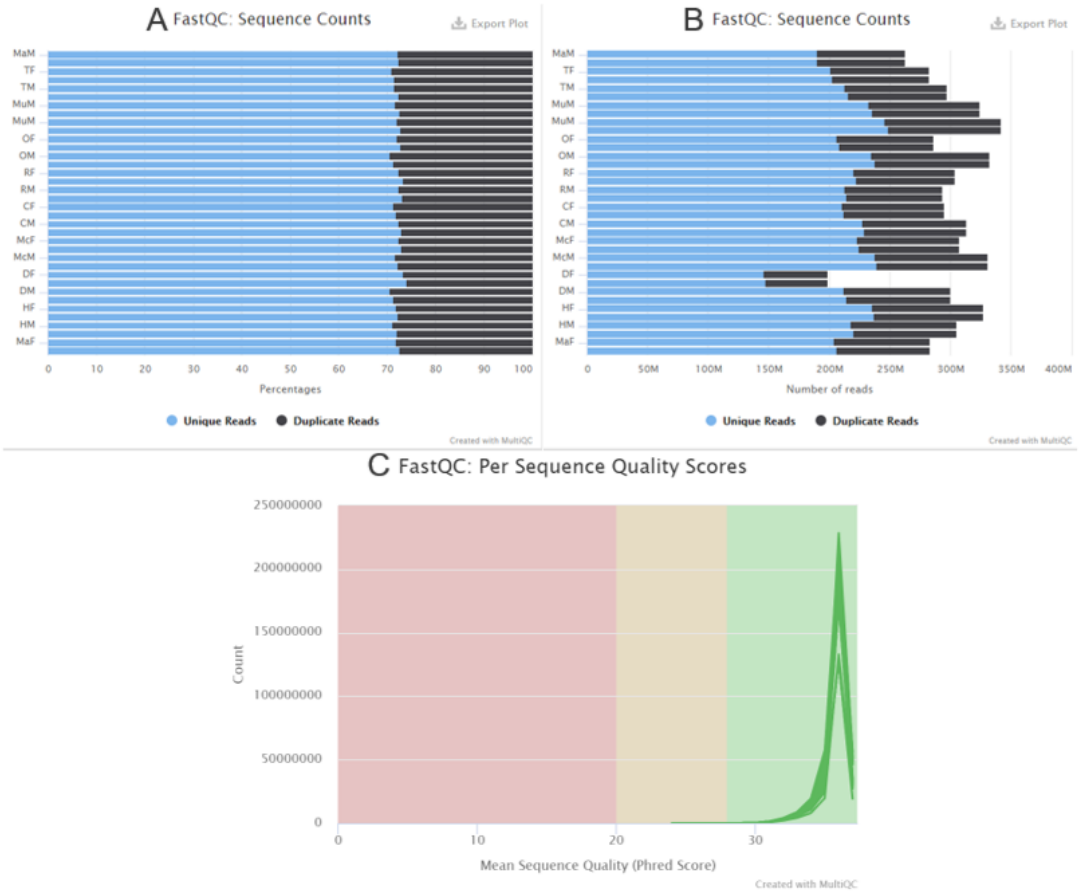
Terms	Estimate	Std. Error	z value	p value	Pseudo R-squared
Nd ~ Elevation F	-0.04124	0.033	-1.25	0.211	0.1332
Nd ~ Elevation M	-0.05945	0.02684	-2.215	0.0268	0.3323
Nd ~ Water volume F	-0.08519	0.02401	-3.547	0.000389	0.637
Nd ~ Water volume M	-0.06408	.02669	-2.401	0.0163	0.4383
Nd ~ pH F	0.0133	0.03488	0.381	0.703	0.01632
Nd ~ pH M	0.01797	0.03198	0.562	0.574	0.0363
Nd ~ Total number of fish stocked F	0.03133	0.03341	0.938	0.348	0.083
Nd ~ Total number of fish stocked M	0.04504	0.02871	1.568	0.117	0.2031
Nd ~ Number of tributaries F	0.01411	0.03495	0.404	0.686	0.01494
Nd ~ Number of tributaries M	0.03102	0.03145	0.986	0.324	0.08463
Nd ~ Zooplankton density F	0.03052	0.03274	0.932	0.351	0.09132
Nd ~ Zooplankton density M	0.009286	0.032186	0.288	0.773	0.01012
Nd ~ Macroinvertebrate density F	-0.00799	0.03526	-0.227	0.821	0.005748
Nd ~ Macroinvertebrate density M	0.01319	0.03204	0.412	0.681	0.0188

Supplemental Table 4. Model selection using the MuMIn package, selecting for water volume in females, and both water volume, and elevation in males. Nd = Nucleotide Diversity.

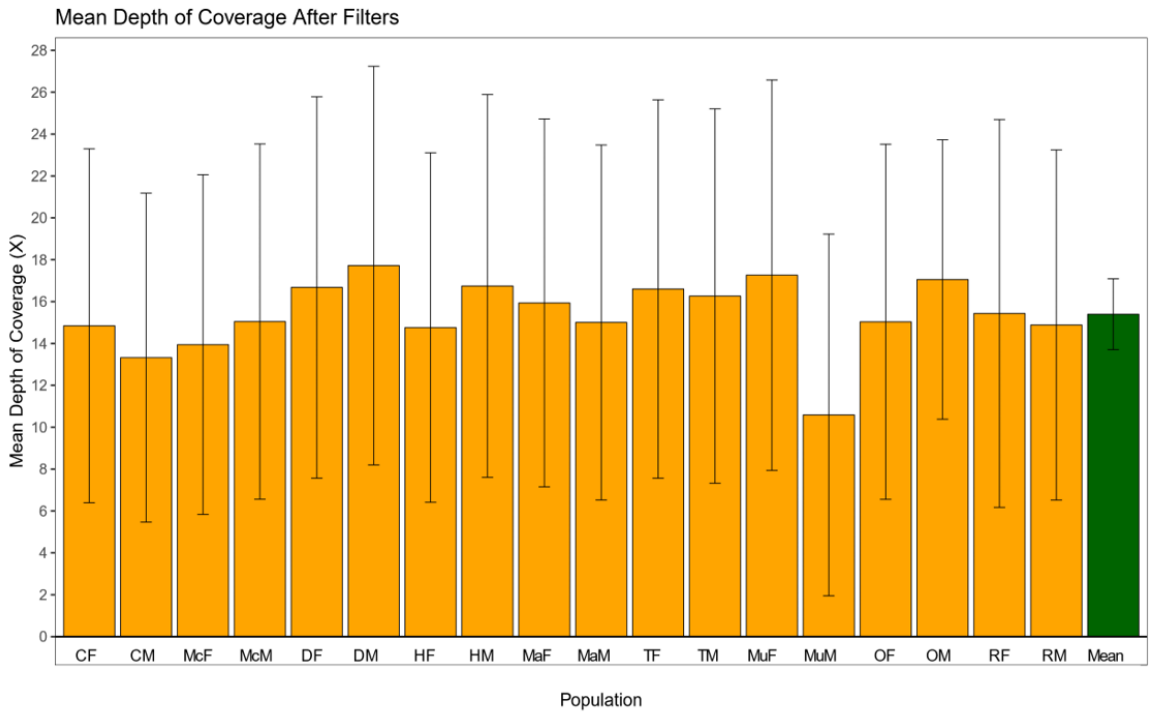
Terms	df	AICc	
Nd ~ Elevation F	3	-103.8945	
Nd ~ Water volume F	3	-111.0454	
Nd ~ pH F	3	-102.6549	
Nd ~ Total number of fish stocked F	3	-103.3243	
Nd ~ Number of tributaries F	3	-102.6570	
Nd ~ Zooplankton density F	3	-103.3535	
Nd ~ Macroinvertebrate density F	3	-102.5594	
Nd ~ Elevation M	3	-107.5823	
Nd ~ Water volume M	3	-108.5852	
Nd ~ pH M	3	-104.0708	
Nd ~ Total number of fish stocked M	3	-105.8632	
Nd ~ Number of tributaries M	3	-104.5973	
Nd ~ Zooplankton density M	3	-103.8361	
Nd ~ Macroinvertebrate density M	3	-103.9183	



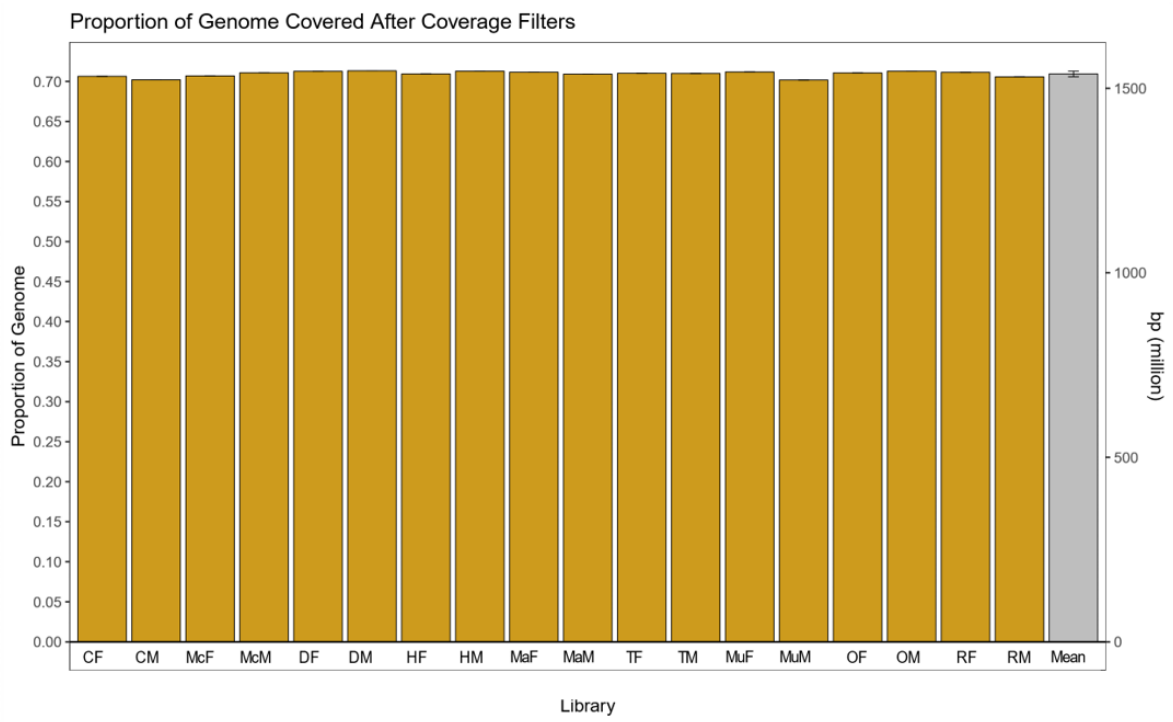
Supplemental Figure 1. Box plot illustrating the difference between biallelic SNPs only, and multiallelic and blacklisted SNPs with A) nucleotide diversity ($p=2.08E^{-7}$) and B) Pairwise F_{ST} ($p=6.3E^{-15}$) with 95% confidence intervals.



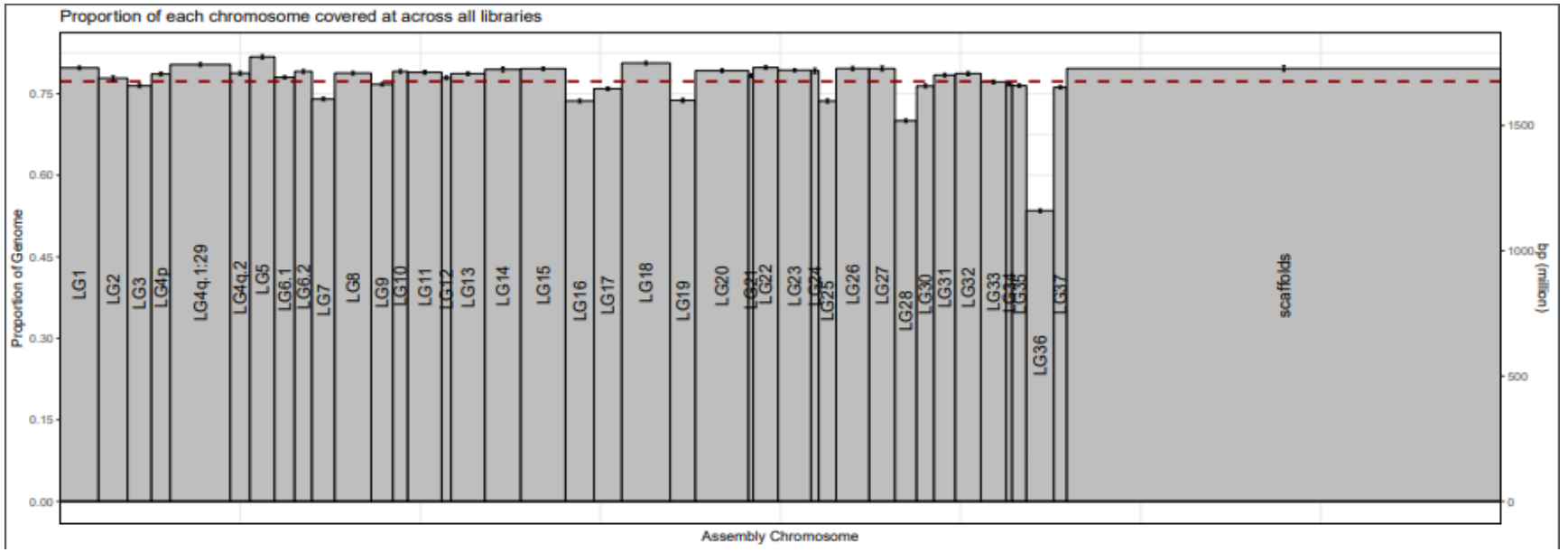
Supplemental Figure 2. FastQC sequence counts visualized as A) percentages and B) number of reads, differentiated between unique and duplicate reads, and C) mean per sequence quality scores. Plots created with MultiQC v 1.7 (Ewels et al., 2016).



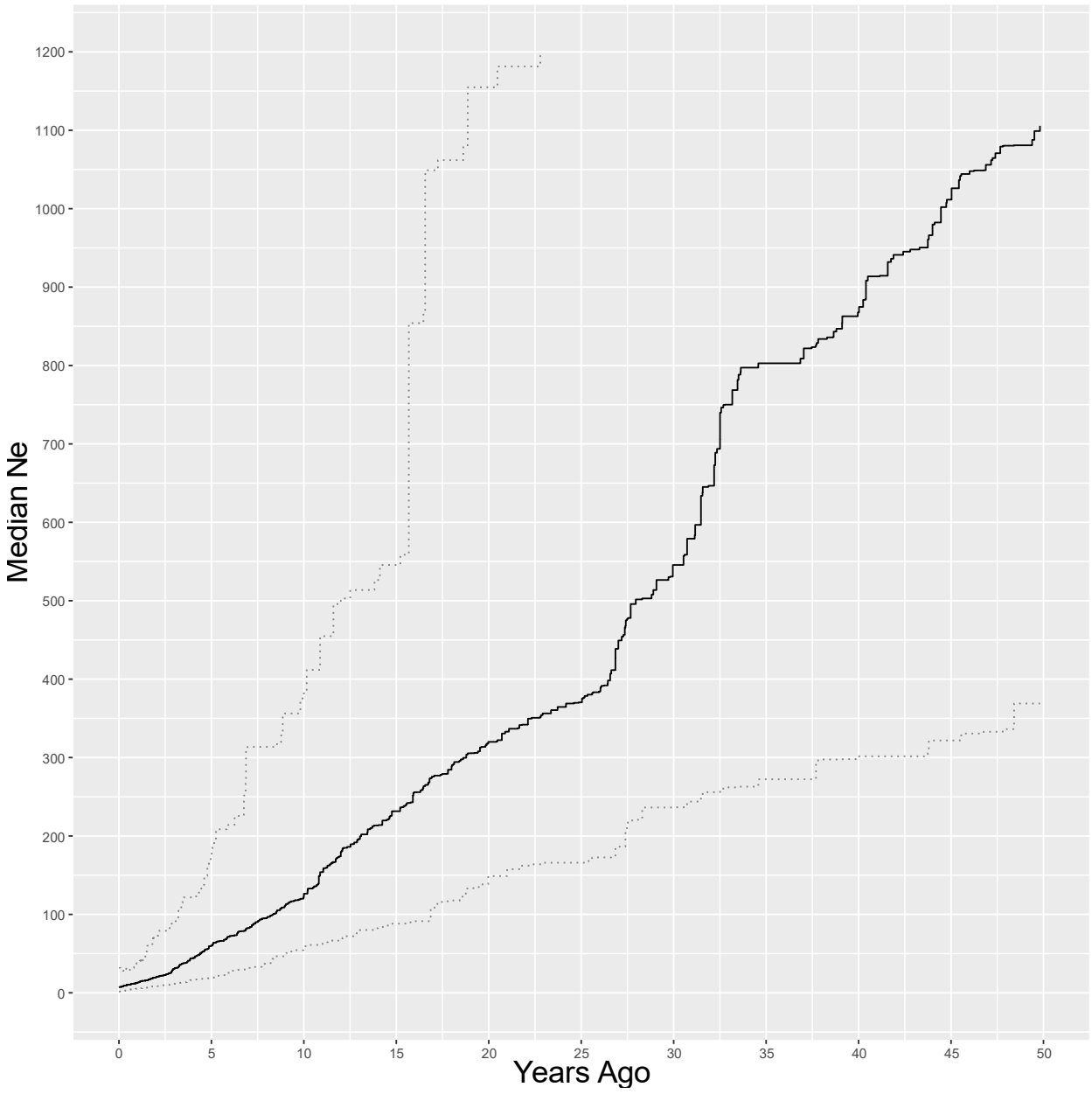
***Supplemental Figure 3*.** Mean depth of coverage across all populations calculated with PPstats in PoolParty v 0.8 (Micheletti & Narum, 2018).



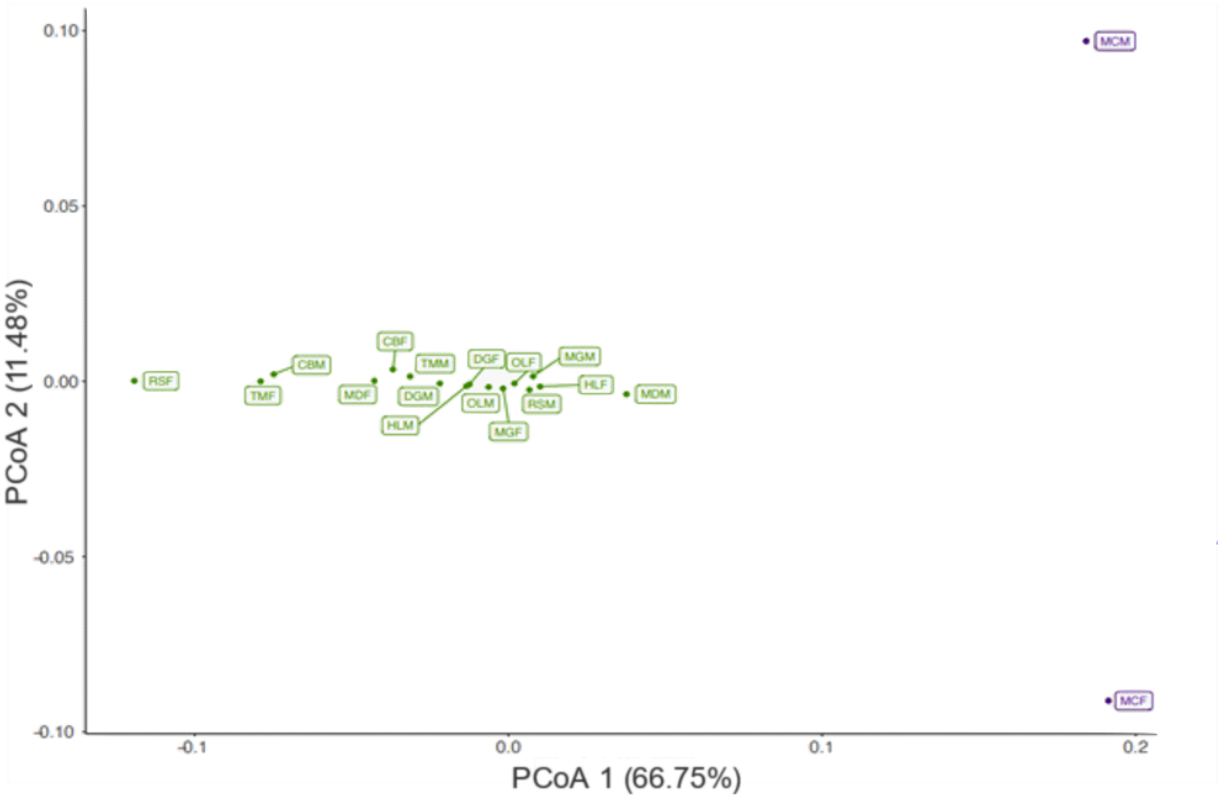
Supplemental Figure 4. Proportion of the genome covered after coverage filters.



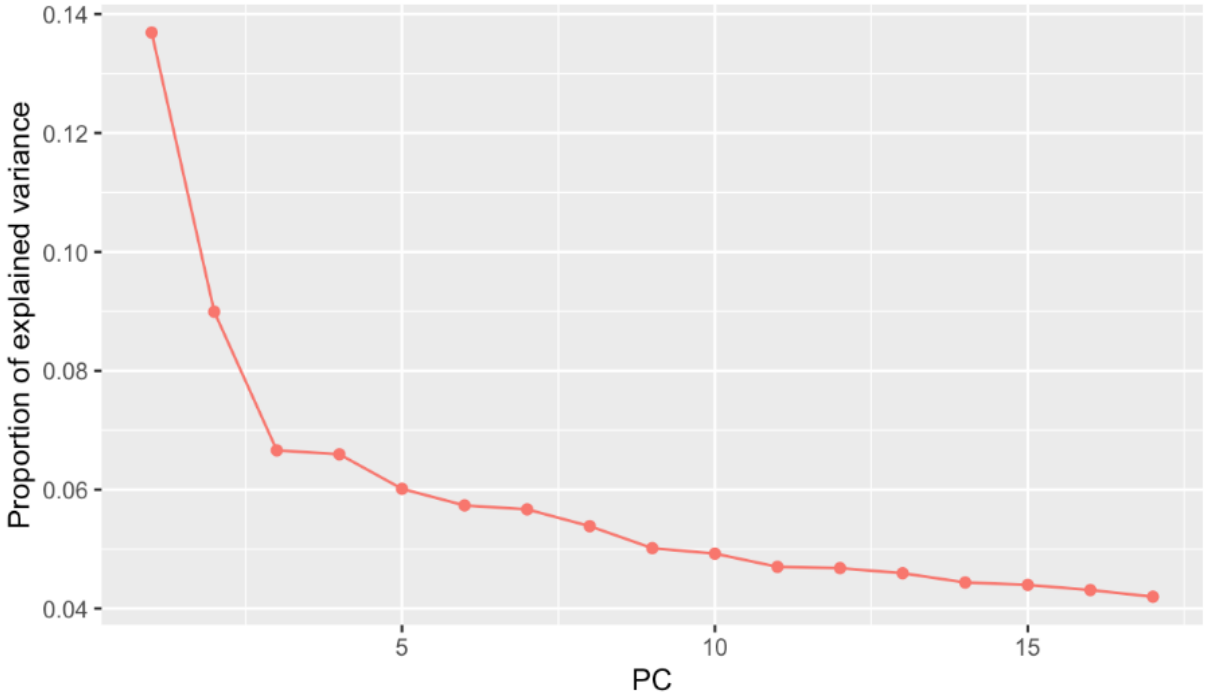
***Supplemental Figure 5*.** Proportion of each chromosome covered by all libraries.



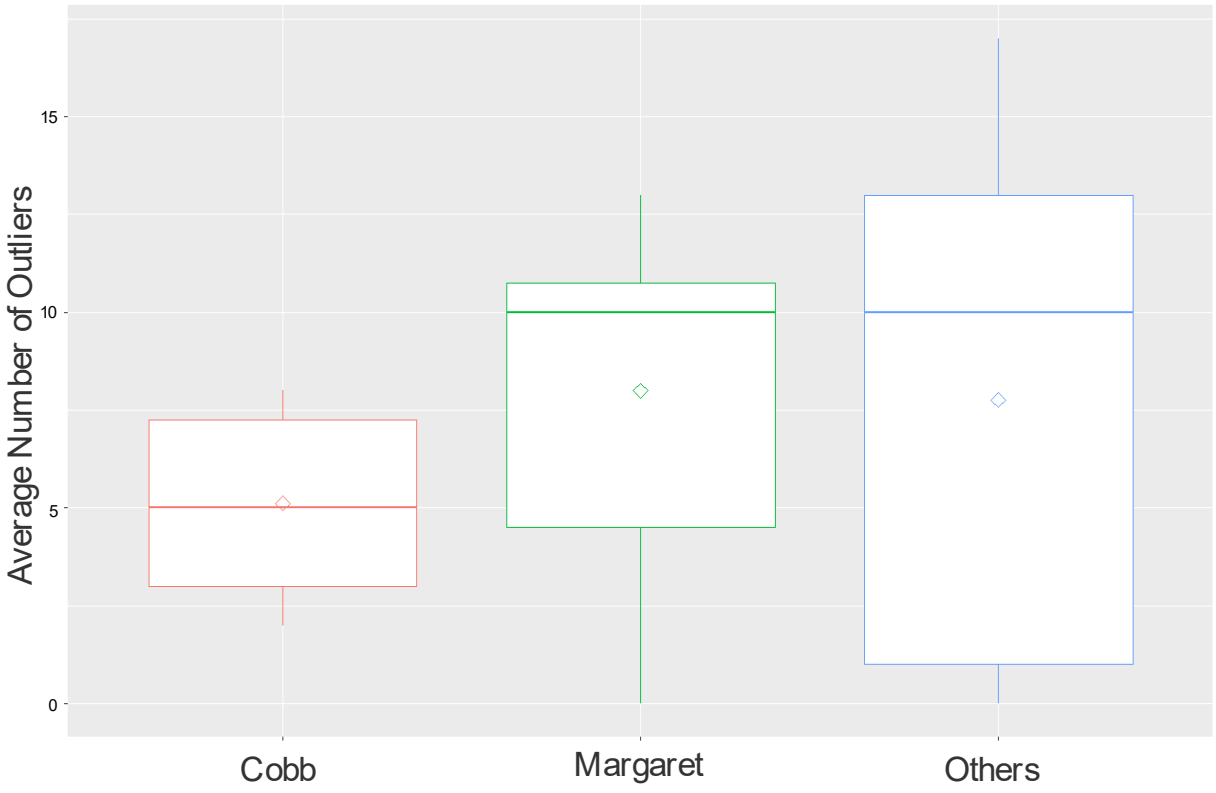
Supplemental Figure 6. Stairway plot of all 18 pools, nine lakes both male and female, showing the population size fluctuation over past generations, grey lines represent 95% confidence intervals.



Supplemental Figure 7. Principle coordinate analysis for pairwise F_{ST} between all populations calculated using Poolfstat with a minimum read count of one, minimum coverage per pool of 20, and maximum coverage per pool of 200, minor allele frequency of 0.05, and the "Anova" method. Lakes are abbreviated C = Cobb, D = Dog, H = Helen, Mc = McNair, Mu = Mud, Ma = Margaret, O = Olive, R = Ross, and T = Temple; while M = Male and F = Female.



Supplemental Figure 8. PCAdapt scree plot confirming the number of parameters that explain the most variance of K=3.



Supplemental Figure 9. Box plot illustrating the number and variance in outliers of the *K*=3 groups. Cobb had a mean of 5.13 ± 2.26 outliers per comparison, Margaret had a mean of 8 ± 4.47, and the remaining group of seven lakes (others) had a mean of 7.76 ± 5.66, showing no significant difference between lakes (p=0.417, type III Anova).

Supplemental Table 5. Individual lake comparisons through CMH analysis, with BLAST function, gene ontology category, gene ontology function and the species related to the gene ontology results. Lakes are abbreviated C = Cobb, D = Dog, H = Helen, Mc = McNair, Mu = Mud, Ma = Margaret, O = Olive, R = Ross, and T = Temple.

Lakes	Chromosome	Base Pair	SNP	P value	BLAST Function	Gene Ontology Category	Gene Ontology Function	Species
TvMc	LG6.1	18227277	snp7	0.02242028	myelin basic protein-like (LOC111965264), transcript variant X1, mRNA	molecular function	structural constituent of myelin sheath	*Austrofundulus limnaeus*
TvMc	LG8	41620402	snp11	0.04147309	putative sodium-coupled neutral amino acid transporter 10 (LOC111967963), mRNA	cellular component	membrane	*Liparis tanakae*
TvMc	LG15	29900965	snp15	0.01294434	transport and golgi organization 2 homolog (tango2), transcript variant X3, mRNA	cellular component	Golgi apparatus	*Equus caballus*
TvMc	LG37	13533282	snp36	0.000146782	long-chain-fatty-acid--CoA ligase ACSBG2 (LOC111972157), transcript variant X2, mRNA	molecular function	ligase activity	*Phytophthora nicotianae*
CvMc	LG4q.1:29	60884645	snp5	0.01817296	SNF-related serine/threonine-protein kinase-like (LOC111962281), mRNA	molecular function	protein kinase activity	*Scleropages formosus*
CvMc	LG15	34473273	snp6	0.02917038	inactive carboxypeptidase-like protein X2 (LOC111974979), mRNA	molecular function	metallocarboxypeptidase activity	*Mus musculus*
CvMc	LG15	46322243	snp7	5.05E-06	homeobox protein orthopedia B-like (LOC111975202), transcript variant X1, mRNA	cellular component	nucleus	*Aurelia aurita*
CvMc	LG15	64538541	snp8	0.04471552	DENN domain containing 5A (dennd5a), transcript variant X4, mRNA	cellular component	trans-Golgi network	*Aotus nancymaae*
CvMc	LG16	10436059	snp9	0.02562949	protein FAM43A-like (LOC111976060), mRNA	cellular component	mitochondrion	*Ectocarpus siliculosus*
CvMc	LG28	19394862	snp17	0.01951748	protein phosphatase 4 regulatory subunit 4 (ppp4r4), transcript variant X1, mRNA	cellular component	cytosol	*Gorilla gorilla gorilla*
CvMc	LG32	20614849	snp21	0.006108301	G-protein coupled receptor 54 (LOC111956315), mRNA	molecular function	G protein-coupled receptor activity	*Ophiophagus hannah*

CvMc	LG36	40398713	snp24	0.006956964	dedicator of cytokinesis protein 9-like (LOC111959519), mRNA	molecular function	guanyl-nucleotide exchange factor activity	*Scleropages formosus*
TvH	LG5	14518627	snp8	9.72E-05	collagen alpha-2(XI) chain-like (LOC111963946), mRNA	cellular component	collagen trimer	*Penaeus vannamei*
TvH	LG6.1	22847223	snp10	0.001711872	transmembrane emp24 domain-containing protein 11 (LOC111965330), transcript variant X1, mRNA	cellular component	Golgi apparatus	*Mus musculus*
TvH	LG6.2	16908776	snp12	0.000167943	multidrug resistance-associated protein 5 (LOC111965891), mRNA	molecular function	organic anion transmembrane transporter activity	*Phaethon lepturus*
TvH	LG8	32338519	snp13	0.02055054	glutamate receptor ionotropic, delta-1-like (LOC111967743), mRNA	molecular function	ionotropic glutamate receptor activity	*Ictalurus punctatus*
TvH	LG15	14478967	snp17	0.001781122	peptidyl-tRNA hydrolase 1 homolog (ptrh1), transcript variant X4, mRNA	molecular function	aminoacyl-tRNA hydrolase activity	*Paramormyrops kingsleyae*
TvH	LG22	5565375	snp22	9.78E-06	B-cell lymphoma/leukemia 11A-like (LOC111949579), transcript variant X1, mRNA	molecular function	nucleic acid binding	*Austrofundulus limnaeus*
TvH	LG23	33592297	snp25	5.48E-06	seizure protein 6-like (LOC111950797), transcript variant X1, mRNA	cellular component	membrane	*Macaca fascicularis*
TvH	LG23	43182717	snp26	0.03394847	CCR4-NOT transcription complex subunit 6 (LOC111950253), mRNA	cellular component	membrane	*Trachymyrmex cornetzi*
TvH	LG28	14531622	snp31	0.000900496	human immunodeficiency virus type I enhancer binding protein 2 (hivep2), mRNA	molecular function	nucleic acid binding	*Cebus capucinus imitator*
TvH	LG28	31400198	snp33	0.003921092	WAS protein family member 1 (wasf1), mRNA	biological process	actin cytoskeleton organization	*Lates calcarifer*
MavMc	LG1	8555253	snp2	0.01917777	mitochondrial import receptor subunit TOM20 homolog (LOC111960523), mRNA	cellular component	mitochondrial outer membrane translocase complex	*Austrofundulus limnaeus*
MavMc	LG7	10773859	snp13	0.000381234	tumor necrosis factor receptor superfamily member 6B-like (LOC111966075), mRNA	biological process	signal transduction	*Callorhinchus milii*
MavMc	LG8	46763424	snp15	7.61E-06	S1 RNA binding domain 1 (srbd1), transcript variant X1, mRNA	cellular component	membrane	*Xenopus tropicalis*

MavMc	LG9	3762720	snp18	0.00767866	spectrin beta chain, erythrocytic (LOC111968456), transcript variant X2, mRNA	molecular function	actin binding	*Tinamus guttatus*
MavMc	LG10	2902731	snp19	0.02638717	protein FMC1 homolog (LOC111980476), mRNA	cellular component	mitochondrion	*Xenopus tropicalis*
CvMu	LG9	15711057	snp1	0.00740331	CDK-activating kinase assembly factor MAT1-like (LOC111968663), transcript variant X1, mRNA	cellular component	transcription factor TFIIH holo complex	*Leptotrombidium deliense*
CvMu	LG11	31603784	snp3	0.006384625	SPRY domain-containing protein 3 (LOC111970186), transcript variant X3, mRNA	molecular function	Dna binding	*Clunio marinus*
MavH	LG1	36458623	snp2	0.00767866	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4 (LOC111966555), transcript variant X2, mRNA	molecular function	6-phosphofructo-2-kinase activity	*Macaca fascicularis*
MavH	LG4p	19447940	snp6	0.02399268	regulation of nuclear pre-mRNA domain-containing protein 2 (LOC111960658), transcript variant X1, misc_RNA	molecular function	aminoacyl-tRNA ligase activity	*Chelonia mydas*
MavH	LG4p	22774374	snp7	0.02175842	LIM/homeobox protein Lhx1-like (LOC111960709), transcript variant X1, mRNA	molecular function	sequence-specific DNA binding	*Austrofundulus limnaeus*
MavH	LG6.2	16763358	snp12	0.000355827	Bruton tyrosine kinase (btk), transcript variant X4, mRNA	biological process	negative regulation of intrinsic apoptotic signaling pathway	*Xenopus tropicalis*
MavH	LG7	14675173	snp14	0.006384625	rho GTPase-activating protein 39 (LOC111966455), mRNA	biological process	signal transduction	*Lygus hesperus*
MavH	LG13	19370915	snp20	0.04147309	protein-glutamine gamma-glutamyltransferase K (LOC111972247), mRNA	molecular function	transferase activity	*Mizuhopecten yessoensis*
MavH	LG17	37935612	snp25	0.03163034	ATP-sensitive inward rectifier potassium channel 12-like (LOC111976506), mRNA	molecular function	voltage-gated ion channel activity	*Scleropages formosus*
MavH	LG20	70063720	snp29	0.007104126	interferon alpha-like (LOC111980627), mRNA	molecular function	cytokine receptor binding	*Enhydra lutris kenyoni*
MavH	LG25	10937222	snp33	0.000712784	collagen alpha-1(XIII) chain (LOC111951857), transcript variant X26, mRNA	cellular component	collagen trimer	*Fukomys damarensis*
MavH	LG27	5827389	snp38	0.01402871	activin receptor type-2B (LOC111953775), mRNA	molecular function	transferase activity	*Rattus norvegicus*

MavH	LG27	13913073	snp40	1.68E-06	dnaJ homolog subfamily C member 13 (LOC111953250), transcript variant X8, mRNA	cellular component	endosome	*Homo sapiens*
MavH	LG34	8953008	snp47	0.03776503	E3 ubiquitin-protein ligase ZFP91 (LOC111958372), mRNA	molecular function	nucleic acid binding	*Phoenicopterus ruber ruber*
MavH	LG36	9940106	snp49	0.04335784	bromodomain adjacent to zinc finger domain 2B (baz2b), transcript variant X16, mRNA	cellular component	host cell nucleus	*Poecilia formosa*
CvH	LG3	19366919	snp3	0.03907941	Nance-Horan syndrome protein-like (LOC111956283), mRNA	biological process	actin cytoskeleton organization	*Scleropages formosus*
CvH	LG21	6029309	snp10	0.01378619	zinc finger protein 180-like (LOC111982340), transcript variant X4, mRNA	molecular function	nucleic acid binding	*Salmo salar*
CvH	LG23	21752291	snp11	0.000104212	protein FAM114A2 (LOC111950452), transcript variant X1, mRNA	biological process	biological process	*Mus musculus*
CvH	LG24	5494728	snp12	0.00259454	metabotropic glutamate receptor 8-like (LOC111951335), mRNA	molecular function	G protein-coupled receptor activity	*Limosa lapponica baueri*
CvT	LG6.1	28928483	snp5	0.001865465	platelet-derived growth factor receptor beta (LOC111965401), transcript variant X3, mRNA	molecular function	platelet-derived growth factor beta-receptor activity	*Chlorocebus sabaeus*
CvT	LG6.2	11612437	snp6	9.23E-05	elongation of very long chain fatty acids protein 6 (LOC111965827), mRNA	cellular component	endoplasmic reticulum membrane	*Larimichthys crocea*
CvT	LG8	22898258	snp8	7.43E-05	short transient receptor potential channel 3-like (LOC111967605), mRNA	molecular function	calcium channel activity	*Brachionus plicatilis*
CvT	LG8	35963603	snp10	0.001686152	carbonic anhydrase-related protein 10 (LOC111967825), transcript variant X1, mRNA	molecular function	zinc ion binding	*Camelus dromedarius*
CvT	LG11	21851796	snp14	0.000635082	heterochromatin protein 1-binding protein 3 (LOC111969980), transcript variant X1, mRNA	cellular component	nucleosome	*Heterocephalus glaber*
CvT	LG13	34965927	snp15	2.12E-05	low-density lipoprotein receptor-related protein 8-like (LOC111971375), transcript variant X2, mRNA	molecular function	calcium ion binding	*Ictalurus punctatus*
CvT	LG20	31257947	snp20	0.000217396	family with sequence similarity 168 member A (fam168a), transcript variant X2, mRNA	biological process	positive regulation of base-excision repair	*Ursus maritimus*

MavT	LG1	20615374	snp3	0.01917777	disks large-associated protein 4-like (LOC111962313), mRNA	biological process	signaling	*Physeter macrocephalus*
MavT	LG2	19816607	snp6	0.02153945	rho-related GTP-binding protein RhoE (LOC111976449), mRNA	molecular function	GTP binding	*Nipponia nippon*
MavT	LG9	9097812	snp17	0.04257932	zinc finger MYM-type protein 4 (LOC111968537), mRNA	molecular function	zinc ion binding	*Eurypyga helias*
MavT	LG11	14766062	snp18	0.00767866	CaM kinase like vesicle associated (camkv), mRNA	molecular function	zinc ion binding	*Xenopus tropicalis*
MavT	LG12	11887432	snp23	0.000614582	neural cell adhesion molecule 2 (ncam2), mRNA	biological process	cell adhesion	*Nothobranchius furzeri*
MavT	LG13	18242153	snp24	0.007956303	neuroligin-1-like (LOC111972231), mRNA	cellular component	membrane	*Tropilaelaps mercedesae*
MavT	LG15	48555482	snp32	0.003123204	serine/threonine-protein kinase BRSK2 (LOC111974396), mRNA	molecular function	protein kinase activity	*Lygus hesperus*
MavT	LG17	28566718	snp35	0.02562949	activin receptor type-1B-like (LOC111976988), transcript variant X2, mRNA	biological process	transmembrane receptor protein serine/threonine kinase signaling pathway	*Xenopus*
MavT	LG23	14747480	snp43	0.003032185	roundabout homolog 3 (LOC111950617), transcript variant X3, mRNA	cellular component	plasma membrane	*Nothobranchius rachovii*
MavT	LG33	14386258	snp52	0.01263083	frizzled-10-B (LOC111957788), mRNA	biological process	Wnt signaling pathway	*Ascaris suum*
CvR	LG4p	25157850	snp3	0.006108301	chloride channel protein 2-like (LOC111960749), mRNA	cellular component	membrane	*Salmo salar*
CvR	LG25	20912172	snp15	0.02175842	neuregulin 3 (nrg3), transcript variant X1, mRNA	molecular function	signaling receptor binding	*Macaca nemestrina*
CvR	LG36	8603537	snp20	0.02803754	double-stranded RNA-specific editase 1 (LOC111959749), transcript variant X4, mRNA	molecular function	adenosine deaminase activity	*Lygus hesperus*
MavR	LG3	26942053	snp5	0.003217101	arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 1-like (LOC111953962), mRNA	molecular function	metal ion binding	*Scleropages formosus*
MavR	LG4q.1:29	81602808	snp9	0.000678933	cat eye syndrome critical region protein 2 (LOC111962616), transcript variant X1, mRNA	biological process	ATP-dependent chromatin remodeling	*Trichinella patagoniensis*
MavR	LG11	37588105	snp16	0.000175444	U3 small nucleolar RNA-interacting protein 2 (LOC111970321), mRNA	biological process	rRNA processing	*Malassezia restricta CBS 7877*
MavR	LG15	23946954	snp17	6.15E-05	arginine/serine-rich coiled-coil protein 2 (LOC111974230), mRNA	biological process	biological process	*Mus musculus*

MavR	LG17	33103754	snp24	2.57E-07	multiple epidermal growth factor-like domains protein 6 (LOC111976271), mRNA	molecular function	calcium ion binding	*Trichinella pseudospiralis*
MavR	LG18	25704590	snp26	0.000821241	peroxisome proliferator-activated receptor gamma coactivator-related protein 1 (LOC111978040), mRNA	molecular function	transcription coregulator activity	*Mus musculus*
MavR	LG31	5864471	snp34	0.02175842	choline transporter-like protein 4 (LOC111955469), mRNA	cellular component	membrane	*Operophtera brumata*
MavR	LG33	1089947	snp37	0.001111816	protein fem-1 homolog C (LOC111958058), mRNA	cellular component	nucleoplasm	*Physeter macrocephalus*
MavR	LG33	29830872	snp39	0.000295767	serine/threonine-protein kinase A-Raf (LOC111957871), transcript variant X4, mRNA	molecular function	protein kinase activity	*Termitomyces sp. J132*
MavR	LG36	19606343	snp40	0.004247934	rho guanine nucleotide exchange factor 7-like	molecular function	guanyl-nucleotide exchange factor activity	*Scleropages formosus*
CvD	LG4q.1:29	81001122	snp8	0.002057949	transmembrane and TPR repeat-containing protein 2-like (LOC111962607), transcript variant X1, mRNA	cellular component	integral component of membrane	*Scleropages formosus*
CvD	LG8	16993680	snp10	0.005706209	Bardet-Biedl syndrome 12 (bbs12), mRNA	biological process	convergent extension involved in gastrulation	*Fundulus heteroclitus*
CvD	LG15	48376592	snp16	9.78E-06	Ras association domain family member 7 (rassf7), transcript variant X1, mRNA	biological process	regulation of microtubule cytoskeleton organization	*Ursus maritimus*
CvD	LG16	145476	snp17	0.000254794	vitellogenin-like (LOC111975854), mRNA	molecular function	lipid transporter activity	*Agrilus planipennis*
CvD	LG19	35153409	snp19	4.91E-06	CMRF35-like molecule 4 (LOC111979088), mRNA	cellular component	membrane	*Labeo rohita*
CvD	LG20	69342944	snp20	1.42E-05	protein FAM171A2 (LOC111980925), mRNA	cellular component	integral component of membrane	*Ursus maritimus*
MavD	LG1	3654619	snp1	0.0175152600	keratin, type I cytoskeletal 13-like (LOC111982613), mRNA	molecular function	structural molecule activity	*Callorhinchus milii*
MavD	LG13	4510701	snp24	0.0039210920	troponin I, fast skeletal muscle (LOC111971996), transcript variant X1, mRNA	cellular component	troponin complex	*Fundulus heteroclitus*
MavD	LG13	1312365	snp23	0.0057062090	transcription factor SOX-6-like (LOC111971958), mRNA	molecular function	DNA binding	*Salmo salar*
MavD	LG14	48718398	snp27	0.0425477400	coagulation factor XIII A chain-like (LOC111973567), mRNA	cellular component	membrane	*Chlorella sorokiniana*

MavD	LG18	44275061	snp32	0.0006789325	pro-neuregulin-3, membrane-bound isoform-like (LOC111978446), mRNA	cellular component	integral component of membrane	*Austrofundulus limnaeus*
MavD	LG18	58843553	snp34	0.0377650300	vinculin (LOC111978256), transcript variant X1, mRNA	molecular function	actin binding	*Scolopendra viridis*
MavD	LG23	44538747	snp38	0.0420096200	junctional adhesion molecule A (LOC111950569), mRNA	biological process	cell adhesion	*Sparus aurata*
MavD	LG26	37452614	snp39	0.0054464470	transmembrane and TPR repeat-containing protein 2-like (LOC111952765), transcript variant X1, mRNA	cellular component	integral component of membrane	*Scleropages formosus*
MavD	LG26	41144937	snp40	0.0398734300	mitochondrial inner membrane protease subunit 2 (LOC111952461), mRNA	biological process	proteolysis	*Kwoniella dejecticola CBS 10117*
MavD	LG27	16829360	snp42	0.0000471113	rab effector MyRIP (LOC111953573), mRNA	molecular function	metal ion binding	*Fukomys damarensis*
MavD	LG28	28727331	snp48	0.0052619750	pre-mRNA-splicing regulator WTAP (LOC111954449), transcript variant X5, mRNA	biological process	regulation of alternative mRNA splicing, via spliceosome	*Opisthocomus hoazin*
MavD	LG3	30073875	snp6	0.0007596940				
MavD	LG3	31827266	snp7	0.0267688100	homeobox protein Hmx-like (LOC111981440), mRNA	cellular component	nucleus	*Clunio marinus*
MavD	LG5	22106851	snp15	0.0012601530	uncharacterized LOC111964057 (LOC111964057), transcript variant X2, mRNA			
MavD	LG5	13076882	snp14	0.0050693100	erbin (LOC111963937), transcript variant X1, mRNA	molecular function	ErbB-2 class receptor binding	*Felis catus*
CvO	LG3	29380169	snp2	0.03698202	Kv channel-interacting protein 4-like (LOC111954260), transcript variant X2, mRNA	molecular function	calcium ion binding	*Scleropages formosus*
CvO	LG4q.2	26534720	snp7	0.000186381	coiled-coil domain containing 170 (ccdc170), transcript variant X1, mRNA	cellular component	integral component of membrane	*Esox lucius*
CvO	LG9	8329517	snp14	0.02188711	A-kinase anchor protein 6-like (LOC111968513), transcript variant X1, mRNA	molecular function	kinase activity	*Austrofundulus limnaeus*
CvO	LG9	30238262	snp17	0.01917777	arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2-like (LOC111968391), mRNA	molecular function	GTPase activator activity	*Hirondellea gigas*
CvO	LG14	48060865	snp18	0.04227848	E3 ubiquitin-protein ligase RNF152 (LOC111973382), mRNA	molecular function	ubiquitin protein ligase activity	*Homo sapiens*

CvO	LG16	1455863	snp20	0.002851459	tubulin tyrosine ligase like 7 (ttll7), transcript variant X1, mRNA	biological process	cellular protein modification process	*Papio anubis*
CvO	LG31	9973521	snp25	1.02E-06	collagen alpha-1(XI) chain-like (LOC111955648), mRNA	cellular component	collagen trimer	*Alligator mississippiensis*
CvO	LG33	8509449	snp26	0.009117259	protein FAM222A (LOC111957935), mRNA	biological process	biological process	*Mus musculus*
MavO	LG4p	19029998	snp2	0.00825757	endothelin-converting enzyme 2-like	biological process	proteolysis	*Scleropages formosus*
MavO	LG4q.1:29	46734565	snp6	0.02676881	leucine rich repeat and Ig domain containing 1 (lingo1), mRNA	molecular function	epidermal growth factor receptor binding	*Macaca mulatta*
MavO	LG5	35326545	snp8	0.009217599	no exon			
MavO	LG6.2	1950395	snp9	0.006108301	pinopsin-like	molecular function	G protein-coupled receptor activity	*Stylophora pistillata*
MavO	LG6.2	3382951	snp10	0.01291537	G protein-coupled receptor activity (mrpl11), transcript variant X1, mRNA	biological process	RNA processing	*Xenopus tropicalis*
MavO	LG6.2	18765852	snp11	0.005261975	extensin	molecular function	structural constituent of cell wall	*Trifolium pratense*
MavO	LG7	33207539	snp12	0.006534903	keratin, type II cytoskeletal 8 (LOC111966854), mRNA	molecular function	protein binding	*Homo sapiens*
MavO	LG9	31657974	snp18	0.04227848	MAM domain containing glycosylphosphatidylinositol anchor 2 (mdga2), mRNA	cellular component	membrane	*Iconisemion striatum*
MavO	LG16	8841445	snp33	0.00722728	paralemmin-1 (LOC111975564), mRNA	biological process	regulation of flower development	*Anthurium amnicola*
MavO	LG16	40816692	snp35	0.001639909	uncharacterized LOC111975783 (LOC111975783), ncRNA			
MavO	LG17	34432365	snp37	0.03389485	mitochondrial carrier homolog 1 (LOC111976240), transcript variant X1, mRNA	cellular component	membrane	*Homo sapiens*
MavO	LG20	74151134	snp40	0.04013898	BAI1 associated protein 3 (baiap3), mRNA	biological process	dense core granule maturation	*Chlorocebus sabaeus*
MavO	LG32	4348632	snp55	0.02487944	uncharacterized LOC111957222 (LOC111957222), transcript variant X1, mRNA			
CvMa	LG6.2	20299408	snp5	0.003217101	PH and SEC7 domain-containing protein 2 (LOC111965965), mRNA	molecular function	ARF guanyl-nucleotide exchange factor activity	*Tinamus guttatus*
CvMa	LG11	6898829	snp9	0.003217101	protein kinase C and casein kinase substrate in neurons protein 1 (LOC111969716), transcript variant X3, mRNA	molecular function	kinase activity	*Trichinella pseudospiralis*

CvMa	LG18	38344797	snp16	0.003921092	protocadherin-15-like (LOC111977924), mRNA	molecular function	calcium ion binding	*Salmo salar*
CvMa	LG31	30807570	snp25	0.00029764	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (LOC111955807), mRNA	no results		
CvMa	LG36	17153082	snp30	0.02796417	uncharacterized LOC111959851 (LOC111959851), ncRNA			
DvH	LG2	39334720	snp5	0.007708939	alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A-like (LOC111979746), mRNA	molecular function	transferase activity	*Scleropages formosus*
DvH	LG4q.2	5198062	snp7	0.04553445	mucolipin-2-like (LOC111962952), mRNA	molecular function	cation channel activity	*Enhydra lutris kenyoni*
DvH	LG14	17507156	snp16	0.02917038	calcipressin-2 (LOC111972542), transcript variant X2, mRNA	molecular function	nucleic acid binding	*Dryobates pubescens*
DvH	LG18	31006510	snp24	0.002621822	echinoderm microtubule associated protein like 4 (eml4), mRNA	cellular component	membrane	*Nothobranchius furzeri*
DvH	LG18	54757101	snp28	8.33E-05	neuromedin-K receptor-like (LOC111977665), mRNA	cellular component	integral component of membrane	*Actinidia chinensis var. chinensis*
DvH	LG20	2274725	snp29	0.015025	histone deacetylase 5-like (LOC111981176), mRNA	molecular function	hydrolase activity	*Nelumbo nucifera*
DvH	LG20	64290763	snp31	0.04394857	protein FAM163B-like (LOC111980058), mRNA	cellular component	membrane	*Scleropages formosus*
DvH	LG22	4205787	snp35	0.02715229	uncharacterized LOC111949697 (LOC111949697), mRNA			
DvH	LG26	20987803	snp39	0.008020971	breast cancer anti-estrogen resistance protein 1 (LOC111952497), transcript variant X3, mRNA	biological process	signal transduction	*Castor canadensis*
DvH	LG26	35585050	snp41	0.04335784	ankyrin repeat and BTB/POZ domain-containing protein BTBD11-B (LOC111952570), transcript variant X1, mRNA	molecular function	metal ion binding	*Human parainfluenza 2 virus*
DvH	LG30	18719397	snp44	0.009419916	zinc finger protein 574 (LOC111954987), transcript variant X3, mRNA	molecular function	nucleic acid binding	*Chlorocebus sabaeus*
DvH	LG32	13801416	snp46	0.00494694	alpha-tocopherol transfer protein (LOC111956960), transcript variant X2, mRNA	cellular component	ribosome	*Camelus dromedarius*

MuvMc	LG4q.1:29	30640310	snp3	0.04227848	RNA exonuclease 4 (LOC111961803), transcript variant X1, mRNA	cellular component	nuclear speck	*Homo sapiens*
MuvMc	LG30	6585886	snp8	0.04054436	ras-related GTP-binding protein C (LOC111955329), transcript variant X2, mRNA	molecular function	GTP binding	*Lygus hesperus*
RvMc	LG3	4229377	snp2	0.01438081	synaptotagmin-1 (LOC111982726), transcript variant X1, mRNA	molecular function	calcium-dependent phospholipid binding	*Trichinella sp. T8*
RvMc	LG4q.1:29	1710168	snp5	0.003217101	U7 small nuclear RNA (LOC111962906), ncRNA	molecular function	protein binding	*Homo sapiens*
RvMc	LG4q.2	9511023	snp7	0.000376754	RAS guanyl-releasing protein 1 (LOC111963252), transcript variant X1, mRNA	molecular function	calcium ion binding	*Alligator mississippiensis*
RvMc	LG16	39253610	snp20	0.003394964	tRNA methyltransferase 1 like (trmt1l), mRNA	biological process	methylation	*Bos indicus x Bos taurus*
RvMc	LG20	67629341	snp24	0.005706209	protein tyrosine phosphatase, non-receptor type 13 (ptpn13), transcript variant X1, mRNA	molecular function	hydrolase activity	*Nothobranchius kuhntae*
RvMc	LG23	44101810	snp27	0.003521334	sentrin-specific protease 3 (LOC111951145), transcript variant X1, mRNA	biological process	proteolysis	*Fukomys damarensis*
RvMc	LG35	10370409	snp33	9.50E-06	thrombospondin type-1 domain-containing protein 7A-like (LOC111958763), mRNA	cellular component	integral component of membrane	*Salmo salar*
RvMc	LG35	11632534	snp34	0.01111573	sialoadhesin-like (LOC111959068), mRNA	cellular component	membrane	*Alligator sinensis*
RvD	LG3	1309919	snp3	0.00016168	plakophilin-2 (LOC111981758), mRNA	biological process	cell-cell adhesion	*Fukomys damarensis*
RvD	LG3	10533240	snp4	9.78E-06	HMG box-containing protein 1-like (LOC111956082), partial mRNA	molecular function	DNA binding	*Scleropages formosus*
RvD	LG3	17008289	snp5	0.001299989	collagen alpha-1(XXV) chain (LOC111951675), transcript variant X1, mRNA	cellular component	collagen trimer	*Amazona aestiva*
RvD	LG8	44874795	snp16	0.000103196	piggyBac transposable element-derived protein 5 (LOC111968011), mRNA	molecular function	transposase activity	*Camelus dromedarius*
RvD	LG18	19639048	snp33	0.04257932	U5 small nuclear ribonucleoprotein 40 kDa protein (LOC111977414), transcript variant X1, mRNA	biological process	cellular protein modification process	*Trichinella pseudospiralis*

RvD	LG22	1712280	snp37	6.14E-06	diacylglycerol kinase delta-like (LOC111949925), mRNA	molecular function	kinase activity	*Hirondellea gigas*
RvD	LG25	12700403	snp41	2.17E-07	potassium voltage-gated channel subfamily KQT member 5-like (LOC111951960), mRNA	molecular function	voltage-gated potassium channel activity	*Vulpes vulpes*
RvD	LG26	36951038	snp42	0.000105928	tetraspanin-9-like (LOC111952939), mRNA	cellular component	integral component of membrane	*Agrilus planipennis*
RvD	LG26	42912672	snp43	2.34E-05	no exon			
RvD	LG30	16331599	snp46	0.000900496	proteasome subunit beta type-4 (LOC111955067), mRNA	cellular component	proteasome complex	*Daphnia magna*
RvD	LG32	12062945	snp49	0.04257932	MAM and LDL-receptor class A domain-containing protein 1-like (LOC111956351), mRNA	cellular component	membrane	*Lingula unguis*
RvD	LG33	37189661	snp52	0.000174384	sin3 histone deacetylase corepressor complex component SDS3 (LOC111957805), transcript variant X1, mRNA	biological process	chromatin organization	*Pongo abelii*
RvD	LG37	4725937	snp57	0.000625001	uncharacterized LOC111960155 (LOC111960155), mRNA			
OvH	LG1	22712048	snp1	0.01951748	rho-related BTB domain-containing protein 3-like (LOC111962653), mRNA	Molecular function	DNA binding	*Clunio marinus*
OvH	LG1	46698092	snp2	0.025139	inter-alpha-trypsin inhibitor heavy chain H3-like (LOC111968607), mRNA	biological process	hyaluronan metabolic process	*Scleropages formosus*
OvH	LG2	1146747	snp3	0.006525255	dachshund homolog 1-like (LOC111973053), mRNA	biological process	microtubule-based process	*Aspergillus flavus*
OvH	LG2	20546200	snp5	0.004247934	tyrosine-protein phosphatase non-receptor type 4 (LOC111976599), transcript variant X1, mRNA	Molecular function	protein tyrosine phosphatase activity	*Cyphomyrmex costatus*
OvH	LG2	24564797	snp6	0.001829857	uncharacterized LOC111972302 (LOC111972302), ncRNA			
OvH	LG3	4021307	snp9	0.00767866	nectin-1-like (LOC111980961), transcript variant X1, mRNA	cellular component	cell-cell junction	*Austrofundulus limnaeus*
OvH	LG5	14562384	snp13	0.02160396	collagen alpha-2(I) chain-like (LOC111964505), mRNA	cellular component	collagen trimer	*Physeter macrocephalus*
OvH	LG8	30570324	snp18	0.03698202	zinc transporter ZIP11-like (LOC111968220), mRNA	biological process	metal ion transport	*Tarsius syrichta*
OvH	LG9	26888301	snp19	0.000158614	arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2-like	Molecular function	GTPase activator activity	*Hirondellea gigas*

					(LOC111968889), transcript variant X1, mRNA			
OvH	LG13	17951333	snp23	0.04335784	neuroligin-1-like (LOC111972231), mRNA	cellular component	membrane	*Tropilaelaps mercedesae*
OvH	LG13	42849796	snp24	5.48E-06	DEAH-box helicase 15 (dhx15), mRNA	Molecular function	helicase activity	*Cebus capucinus imitator*
OvH	LG15	14478967	snp26	0.01515163	peptidyl-tRNA hydrolase 1 homolog (ptrh1), transcript variant X4, mRNA	Molecular function	aminoacyl-tRNA hydrolase activity	*Paramormyrops kingsleyae*
OvH	LG17	36351175	snp28	0.000678933	transcription initiation factor TFIID subunit 10 (LOC111977249), mRNA	biological process	DNA-templated transcription, initiation	*Operophtera brumata*
OvH	LG18	29382824	snp29	0.01489883	cilia and flagella associated protein 46 (cfap46), mRNA	biological process	axoneme assembly	*Chlorocebus sabaeus*
OvH	LG19	38092254	snp31	0.008020971	ATP-binding cassette sub-family F member 2 (LOC111979175), transcript variant X2, mRNA	Molecular function	nucleotide binding	*Dufourea novaeangliae*
OvH	LG36	38841524	snp45	0.009419916	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial (LOC111959764), mRNA	no results		
OvH	LG36	40857574	snp46	0.04394857	homeobox protein Hox-D9a-like (LOC111959924), mRNA	Molecular function	DNA binding	*Saccoglossus kowalevskii*
DvT	LG1	26844658	snp1	0.01951748	receptor-type tyrosine-protein phosphatase gamma-like (LOC111963820), mRNA	molecular function	protein tyrosine phosphatase activity	*Ciona intestinalis*
DvT	LG3	21737474	snp5	0.006619428	lysophospholipid acyltransferase LPCAT4 (LOC111952481), mRNA	molecular function	transferase activity, transferring acyl groups	*Camelus dromedarius*
DvT	LG4q.1:29	621333	snp7	0.02638717	collagen alpha-1(XXVII) chain A-like (LOC111961212), mRNA	cellular component	collagen trimer	*Tropilaelaps mercedesae*
DvT	LG6.1	10095405	snp12	0.04881208	dyslexia-associated protein KIAA0319 homolog (LOC111965041), transcript variant X1, mRNA	cellular component	membrane	*Tursiops truncatus*
DvT	LG8	13864845	snp14	0.002183045	ADP/ATP translocase 2 (LOC111967456), mRNA	cellular component	membrane	*Labeo rohita*
DvT	LG13	29343122	snp20	0.01270873	trafficking protein particle complex subunit 8 (LOC111971168), mRNA	cellular component	integral component of membrane	*Trichinella sp. T9*
DvT	LG17	38050850	snp28	0.04417134	leucine-rich repeat-containing protein 3B-like (LOC111977062), mRNA	cellular component	integral component of membrane	*Salmo salar*

DvT	LG20	4315290	snp33	0.02450734	no exon			
DvT	LG20	38658177	snp35	0.01438081	homeobox protein cut-like 1 (LOC111979971), mRNA	molecular function	DNA binding	*Orchesella cincta*
DvT	LG23	44449131	snp42	8.41E-06	LHFPL tetraspan subfamily member 6 (lhfpl6), transcript variant X3, mRNA	cellular component	membrane	*Poecilia formosa*
DvT	LG25	19083209	snp45	0.033641	ryanodine receptor 2-like (LOC111951725), mRNA	molecular function	calcium channel activity	*Salmo salar*
DvT	LG26	20599424	snp46	5.80E-06	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 9 (LOC111952160), transcript variant X1, mRNA	molecular function	transferase activity, transferring glycosyl groups	*Homo sapiens*
DvT	LG31	15348088	snp51	1.22E-06	potassium two pore domain channel subfamily K member 9 (kcnk9), mRNA	molecular function	potassium channel activity	*Ailuropoda melanoleuca*
DvT	LG36	7883268	snp55	1.56E-07	ATP-dependent RNA helicase DDX18 (LOC111959622), mRNA	molecular function	hydrolase activity	*Acipenser ruthenus*
DvT	LG36	33181251	snp58	0.02303899	semaphorin 5B (sema5b), mRNA	molecular function	semaphorin receptor binding	*Xenopus tropicalis*
McvD	LG4q.1:29	11688130	snp6	0.002693749	WD repeat-containing protein 70-like (LOC111961442), mRNA	molecular function	threonine-type endopeptidase activity	*Clunio marinus*
McvD	LG4q.1:29	12498269	snp7	0.02715229	piggyBac transposable element-derived protein 4-like (LOC111961474), mRNA	molecular function	nucleic acid binding	*Salmo salar*
McvD	LG4q.2	6537561	snp8	0.002151337	chloride intracellular channel protein 4-like (LOC111963185), mRNA	biological process	chloride transport	*Nannochloropsis gaditana*
McvD	LG6.2	6951300	snp11	0.000370257	cholinergic receptor nicotinic beta 3 subunit (chrnb3), mRNA	cellular component	cell junction	*Chlorocebus sabaeus*
McvD	LG7	6940218	snp12	4.92E-05	helicase ARIP4 (LOC111966287), transcript variant X1, mRNA	molecular function	ATP binding	*Clunio marinus*
McvD	LG7	27322574	snp13	0.001463111	prefoldin subunit 4 (LOC111966746), mRNA	biological process	protein folding	*Clunio marinus*
McvD	LG8	35156854	snp15	0.01354193	regulator of G-protein signaling 9 (LOC111968233), mRNA	biological process	G protein-coupled receptor signaling pathway	*Charadrius vociferus*
McvD	LG9	3762720	snp16	0.03163034	spectrin beta chain, erythrocytic (LOC111968456), transcript variant X2, mRNA	molecular function	actin binding	*Tinamus guttatus*
McvD	LG11	16196149	snp21	0.009419916	potassium channel subfamily K member 15-like (LOC111970608), mRNA	molecular function	potassium channel activity	*Salmo salar*

McvD	LG35	3421533	snp43	0.000859749	zinc finger protein, FOG family member 2 (zfpm2), mRNA	molecular function	RNA polymerase II transcription factor binding	*Pan paniscus*
McvH	LG11	900379	snp11	0.04553445	fer-1-like protein 4 (LOC111969638), mRNA	cellular component	integral component of membrane	*Zootermopsis nevadensis*
McvH	LG18	49890639	snp18	0.0258357	transmembrane protein 150A (LOC111978534), transcript variant X1, misc_RNA	cellular component	membrane	*Fundulus heteroclitus*
McvH	LG20	17282628	snp20	0.02450734	sideroflexin-5 (LOC111981706), transcript variant X1, mRNA	biological process	ion transport	*Lygus hesperus*
McvH	LG20	18533025	snp21	0.000320021	peripheral-type benzodiazepine receptor-associated protein 1-like (LOC111981851), mRNA	cellular component	membrane	*Platysternon megacephalum*
McvH	LG32	33238578	snp31	0.04881208	kidney mitochondrial carrier protein 1 (LOC111957071), mRNA	cellular component	membrane	*Glycine soja*
McvH	LG36	1352026	snp34	0.01212841	collagen alpha-1(XVIII) chain (LOC111959381), transcript variant X3, mRNA	cellular component	collagen trimer	*Charadrius vociferus*
TvO	LG2	12700287	snp6	0.000821596	uncharacterized protein C7orf57 homolog (LOC111974535), transcript variant X1, mRNA	biological process	biological process	*Mus musculus*
TvO	LG2	28643751	snp8	0.01981823	neurogenic differentiation factor 1 (LOC111977678), mRNA	biological process	regulation of transcription, DNA-templated	*Camelus dromedarius*
TvO	LG3	159375	snp9	0.000943328	keratocan-like (LOC111955509), partial mRNA	biological process	visual perception	*Poecilia formosa*
TvO	LG3	21775351	snp12	0.02175842	SPT16 homolog, facilitates chromatin remodeling subunit (supt16h), transcript variant X1, mRNA	cellular component	FACT complex	*Papio anubis*
TvO	LG4q.1:29	9564341	snp15	0.04444325	retinoic acid induced 14 (rai14), transcript variant X1, mRNA	cellular component	extracellular matrix	*Xenopus tropicalis*
TvO	LG4q.1:29	21717064	snp16	0.004967303	ankyrin-1 (LOC111961660), mRNA	biological process	signal transduction	*Camelus dromedarius*
TvO	LG6.2	2845079	snp19	0.002312779	uncharacterized LOC111965670 (LOC111965670), transcript variant X2, ncRNA			
TvO	LG6.2	17145963	snp21	7.43E-05	uncharacterized LOC111965565 (LOC111965565), ncRNA			
TvO	LG16	12143852	snp32	0.00056496	uncharacterized LOC111976129 (LOC111976129), transcript variant X2, ncRNA			
TvO	LG18	12394280	snp35	0.007956303	gap junction delta-3 protein (LOC111977403), mRNA	cellular component	plasma membrane	*Takifugu flavidus*
TvO	LG18	52795929	snp36	0.004360239	bone morphogenetic protein 5 (LOC111978064), mRNA	cellular component	integral component of membrane	*Nothobranchius pienaari*

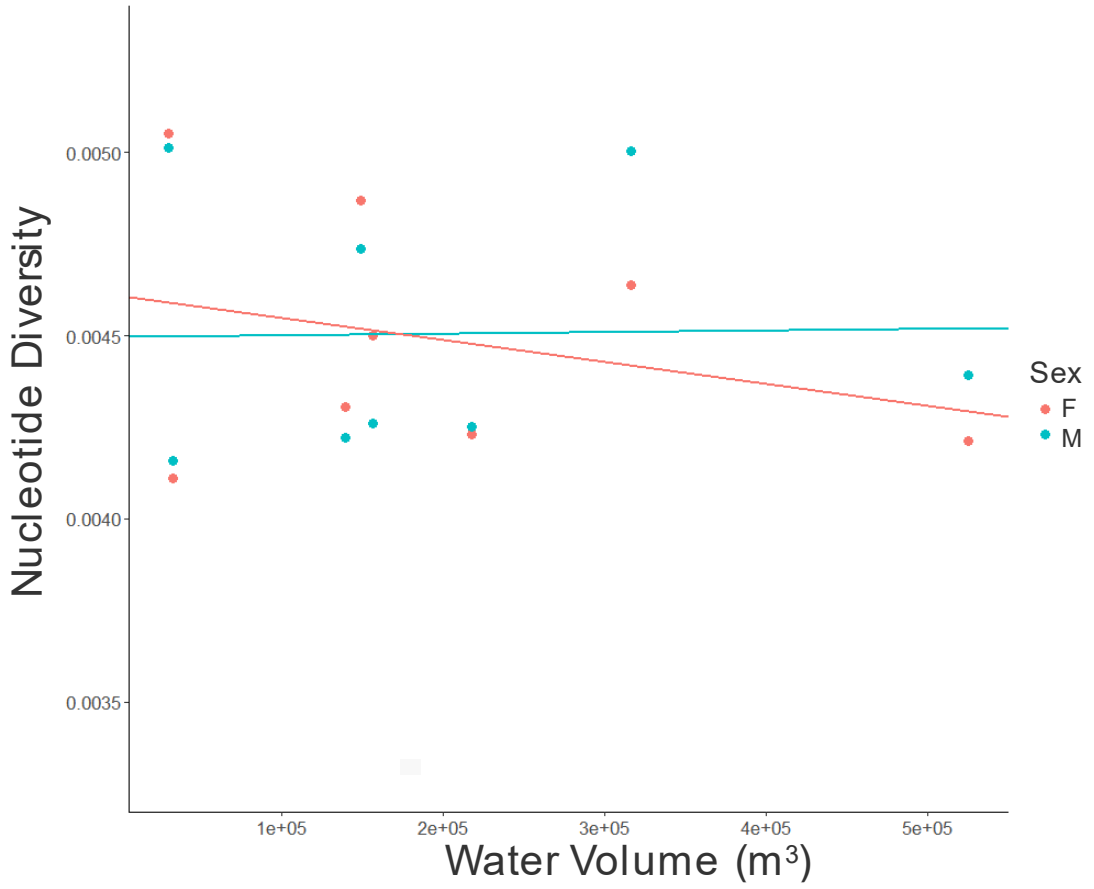
TvO	LG20	41315754	snp41	0.04394857	cytospin-B-like (LOC111981317), transcript variant X3, mRNA	biological process	meiotic cell cycle	*Cucumis melo var. makuwa*
TvO	LG22	9557235	snp43	0.03776503	echinoderm microtubule-associated protein-like 2 (LOC111949414), transcript variant X3, mRNA	molecular function	microtubule binding	*Mus musculus*
TvO	LG26	14313045	snp52	0.009217599	basigin (LOC111952490), transcript variant X2, mRNA	cellular component	plasma membrane	*Camelus dromedarius*
TvO	LG28	18150994	snp58	0.0382894	carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein-like (LOC111954114), mRNA	cellular component	BLOC-1 complex	*Platysternon megacephalum*
TvO	LG30	5859132	snp60	0.02450734	transcriptional repressor GATA binding 1 (trps1), mRNA	molecular function	metal ion binding	*Papio anubis*
TvO	LG37	19080562	snp65	1.70E-05	oligosaccharyltransferase complex subunit ostc (LOC111960315), mRNA	biological process	protein glycosylation	*Zea mays*
RvO	LG3	21775351	snp5	0.000685037	SPT16 homolog, facilitates chromatin remodeling subunit (supt16h), transcript variant X1, mRNA	cellular component	FACT complex	*Papio anubis*
RvO	LG4q.1:29	9560336	snp7	4.88E-06	retinoic acid induced 14 (rai14), transcript variant X1, mRNA	cellular component	extracellular matrix	*Xenopus tropicalis*
RvO	LG6.1	576156	snp9	0.001285517	SET domain bifurcated 1 (setdb1), mRNA	cellular component	nucleus	*Protopterus annectens*
RvO	LG9	18655480	snp11	0.0382894	zinc finger protein DPF3 (LOC111968753), mRNA	molecular function	zinc ion binding	*Trichinella sp. T8*
RvO	LG15	38535385	snp14	0.04463678	proline dehydrogenase 1, mitochondrial (LOC111975162), mRNA	biological process	oxidation-reduction process	*Dictyostelium discoideum*
RvO	LG24	1098108	snp26	0.04147309	protein FAM19A5 (LOC111951564), transcript variant X1, mRNA	cellular component	integral component of membrane	*Callorhinus ursinus*
RvO	LG24	8255688	snp27	0.002151337	vegetative cell wall protein gp1-like (LOC111951357), mRNA	molecular function	structural constituent of cuticle	*Diaphorina citri*
RvO	LG27	784868	snp28	0.04394857	polyhomeotic-like protein 2 (LOC111953970), mRNA	molecular function	zinc ion binding	*Neomonachus schauinslandi*
RvO	LG31	29914955	snp31	0.001500516	extensin-like (LOC111955794), mRNA	cellular component	membrane	*Prunus yedoensis var. nudiflora*
RvO	LG36	2524635	snp32	0.001420932	regulator of nonsense transcripts 3A-like (LOC111959860), transcript variant X2, mRNA	biological process	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	*Plasmodium coatneyi*

McvO	LG8	49368147	snp6	0.006384625	endoplasmic reticulum lectin 1 (LOC111968068), transcript variant X2, mRNA	molecular function	carbohydrate binding	*Lygus hesperus*
McvO	LG32	19188679	snp37	0.03698202	uncharacterized LOC111956372 (LOC111956372), transcript variant X3, ncRNA			
HvR	LG1	7030492	snp1	0.02340302	protein kinase C alpha type (LOC111957424), transcript variant X1, mRNA	molecular function	kinase activity	*Cajanus cajan*
HvR	LG3	11599280	snp2	0.003409886	uncharacterized LOC111981471 (LOC111981471), ncRNA			
HvR	LG4q.1:29	40302530	snp5	0.001112289	serine/threonine-protein kinase BRSK2-like (LOC111960930), mRNA	molecular function	kinase activity	*Tarsius syrichta*
HvR	LG4q.1:29	44597105	snp7	0.03389485	eukaryotic initiation factor 4A-I (LOC111962020), mRNA	molecular function	nucleic acid binding	*Enhydra lutris kenyoni*
HvR	LG13	43861221	snp11	0.000178754	transmembrane protein 131-like (LOC111971549), transcript variant X3, mRNA	cellular component	integral component of membrane	*Mesocricetus auratus*
HvR	LG14	24062137	snp12	0.000759694	T-cell activation Rho GTPase-activating protein (LOC111973443), mRNA	molecular function	GTPase activator activity	*Homo sapiens*
HvR	LG16	38346458	snp16	0.04394857	microtubule-associated serine/threonine-protein kinase 2-like (LOC111975495), mRNA	molecular function	kinase activity	*Mesocricetus auratus*
HvR	LG19	12367471	snp18	0.009217599	myosin light chain 1, cardiac muscle (LOC111978968), mRNA	molecular function	calcium ion binding	*Corvus brachyrhynchos*
HvR	LG19	30136314	snp19	0.02188711	choline-phosphate cytidylyltransferase A (LOC111979607), transcript variant X1, mRNA	molecular function	nucleotidyltransferase activity	*Camelus dromedarius*
HvR	LG20	46736785	snp21	0.01291537	sodium/potassium-transporting ATPase subunit beta-2 (LOC111981831), transcript variant X2, mRNA	biological process	sodium ion transport	*Melipona quadrifasciata*
HvR	LG20	50466081	snp22	0.04394857	transmembrane protein FAM155A-like (LOC111980313), mRNA	cellular component	integral component of membrane	*Lynx pardinus*
HvR	LG26	8099691	snp25	0.001184451	ephrin-A2-like (LOC111953065), mRNA	molecular function	ephrin receptor binding	*Scophthalmus maximus*
HvR	LG33	13525229	snp30	0.009117259	TBC1 domain family member 10A (LOC111957627), mRNA	molecular function	U6 snRNA binding	*Xenopus tropicalis*
HvR	LG33	33323365	snp31	0.01046121	zinc finger protein 131 (LOC111957956), mRNA	cellular component	nucleoplasm	*Papio anubis*

DvO	LG1	24663720	snp2	0.02088757	C->U-editing enzyme APOBEC-2 (LOC111963481), mRNA	molecular function	RNA binding	*Homo sapiens*
DvO	LG4q.1:29	7162706	snp5	0.02160396	soluble lamin-associated protein of 75 kDa-like (LOC111961332), transcript variant X2, mRNA	cellular component	intermediate filament	*Hydra vulgaris*
DvO	LG4q.2	21314780	snp9	0.00173751	bromodomain adjacent to zinc finger domain protein 1A (LOC111963461), transcript variant X3, mRNA	molecular function	metal ion binding	*Dufourea novaeangliae*
DvO	LG7	17418987	snp15	0.00373802	DENN domain-containing protein 2C (LOC111966511), transcript variant X1, mRNA	molecular function	Rab guanyl-nucleotide exchange factor activity	*Homo sapiens*
DvO	LG11	35555994	snp23	0.04335784	sodium- and chloride-dependent GABA transporter 2-like (LOC111970289), mRNA	biological process	transmembrane transport	*Lingula unguis*
DvO	LG16	6530843	snp29	0.00208879	FYVE and coiled-coil domain-containing protein 1 (LOC111975752), transcript variant X2, mRNA	molecular function	metal ion binding	*Phytophthora nicotianae*
DvO	LG16	15161749	snp30	0.000237277	peptidyl-prolyl cis-trans isomerase FKBP8 (LOC111975390), transcript variant X1, mRNA	molecular function	isomerase activity	*Macaca mulatta*
DvO	LG20	27818780	snp36	0.000387177	kin of IRRE-like protein 3 (LOC111981964), mRNA	cellular component	integral component of membrane	*Melipona quadrifasciata*
DvO	LG20	32726302	snp37	1.34E-07	opioid-binding protein/cell adhesion molecule-like (LOC111981966), transcript variant X2, mRNA	molecular function	molecular function	*Mus musculus*
DvO	LG20	43606433	snp38	0.033641	protein EURL homolog (LOC111981575), transcript variant X1, mRNA	biological process	positive regulation of dendritic spine development	*Danio rerio*
DvO	LG20	61024995	snp39	0.001331429	guanine nucleotide-binding protein G(q) subunit alpha (LOC111980819), mRNA	molecular function	G protein-coupled receptor binding	*Melipona quadrifasciata*
DvO	LG24	9530255	snp46	0.00400808	*uncharacterized LOC111951267 (LOC111951267), ncRNA*			
DvO	LG31	17835809	snp53	0.004752522	anillin-like (LOC111955930), transcript variant X1, mRNA	biological process	hemopoiesis	*Euroglyphus maynei*
DvO	LG32	13986667	snp54	0.01046388	chromodomain helicase DNA binding protein 7 (chd7), transcript variant X1, mRNA	biological process	face development	*Papio anubis*

DvO	LG37	3738783	snp61	0.003123204	transmembrane protein 256 (LOC111960416), mRNA	cellular component	integral component of membrane	*Strongyloides ratti*
TvR	LG1	9981469	snp1	0.04951506	son of sevenless homolog 1 (LOC111957031), mRNA	molecular function	guanyl-nucleotide exchange factor activity	*Nothobranchius furzeri*
TvR	LG2	14771845	snp3	0.00573387	serine/threonine-protein kinase tousled-like 1-B (LOC111975038), mRNA	biological process	regulation of chromatin assembly or disassembly	*Danio rerio*
TvR	LG5	30381440	snp10	0.000678933	aristaless-related homeobox protein-like (LOC111964277), mRNA	molecular function	DNA binding	*Tropilaelaps mercedesae*
TvR	LG7	5556872	snp14	9.78E-06	guanine nucleotide-binding protein G(s) subunit alpha (LOC111966254), transcript variant X2, mRNA	biological process	signal transduction	*Trachymyrmex cornetzi*
TvR	LG8	15967402	snp15	0.000915942	no exon			
TvR	LG11	21443443	snp20	0.000556339	membrane-associated guanylate kinase, WW and PDZ domain-containing protein 3 (LOC111969958), transcript variant X3, mRNA	no results		
TvR	LG15	46024167	snp24	0.02088757	U7 small nuclear RNA (LOC111975239), ncRNA	no results		
TvR	LG17	25163238	snp25	0.00056496	basement membrane-specific heparan sulfate proteoglycan core protein-like (LOC111976306), mRNA	cellular component	membrane	*Scleropages formosus*
TvR	LG18	18611124	snp26	0.02088757	G protein-activated inward rectifier potassium channel 1-like (LOC111977554), mRNA	molecular function	inward rectifier potassium channel activity	*Ictalurus punctatus*
TvR	LG20	10018312	snp30	0.01835842	pyruvate dehydrogenase (acetyl-transferring) kinase isozyme 2, mitochondrial (LOC111981212), mRNA	biological process	carbohydrate metabolic process	*Rattus norvegicus*
TvR	LG20	32008925	snp31	8.41E-06	intraflagellar transport 46 (ift46), mRNA	biological process	cilium assembly	*Papio anubis*
TvR	LG23	3579517	snp34	0.02231721	glutaredoxin and cysteine rich domain containing 2 (grxcr2), mRNA	cellular component	microvillus	*Cebus capucinus imitator*
TvR	LG25	1898204	snp37	0.005706209	beta-crystallin S-1-like (LOC111951602), transcript variant X2, mRNA	biological process	glycine betaine biosynthetic process from choline	*Burkholderia plantarii*

TvR	LG25	11075582	snp38	6.85E-06	Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1 (LOC111951704), mRNA	molecular function	ARF guanyl-nucleotide exchange factor activity	*Trichinella sp. T8*
TvR	LG30	16691625	snp41	0.002057949	tumor necrosis factor alpha-induced protein 8-like protein 2 (LOC111955375), mRNA	biological process	regulation of apoptotic process	*Schistosoma haematobium*
TvR	LG33	9156849	snp45	0.02630379	vacuolar protein sorting-associated protein 13A (LOC111957899), transcript variant X1, mRNA	molecular function	catalytic activity	*Talaromyces islandicus*
TvR	LG33	32037434	snp47	0.00016168	growth arrest and DNA damage-inducible protein GADD45 gamma (LOC111957399), mRNA	biological process	regulation of cell cycle	*Ursus maritimus*
TvR	LG36	21320000	snp50	0.000507807	synaptotagmin-like protein 5 (LOC111959478), transcript variant X1, mRNA	molecular function	metal ion binding	*Dufourea novaeangliae*
MuvO	LG37	19080562	snp7	0.02450734	oligosaccharyltransferase complex subunit ostc (LOC111960315), mRNA	biological process	protein glycosylation	*Zea mays*
TvMu	LG11	9577390	snp2	0.001526367	no exon			
TvMu	LG12	9582656	snp3	0.000368374	transmembrane protein 163 (LOC111971006), mRNA	cellular component	integral component of membrane	*Balearica regulorum gibbericeps*
DvMu	LG33	10280610	snp4	0.03163034	cell surface hyaluronidase-like (LOC111958112), mRNA	biological process	viral process	*Duck hepatitis B virus*



Supplemental Figure 10. Nucleotide diversity correlations with water volume for female and male populations corrected without Lake Margaret.

Supplemental Table 6. Additive individual lake comparisons through CMH analysis, with BLAST function, gene ontology category, based on gene ontology function.

BLAST Function	Gene Ontology Category	Gene Ontology Function	CvO	MavO	CvH	MavH	CvD	MavD	CvR	MavR	CvT	MavT	CvMc	MavMc	CvMu	MavMu	CvMa	TvMc	TvH	DvH	MuvMc	RvMc	RvD	OvH	DvT	McvD	McvH	TvO	RvO	McvO	HvR	DvO	TvR	MuvO	TvMu	DvMu	Sum
putative sodium-coupled neutral amino acid transporter 10 (LOC111967963), mRNA	cellular component	membrane	0	2	0	0	1	1	1	1	0	1	0	1	0	0	0	1	2	2	0	1	1	1	3	0	3	0	1	0	0	0	1	0	0	0	24
transmembrane and TPR repeat-containing protein 2-like (LOC111962607), transcript variant X1, mRNA	cellular component	integral component of membrane	1	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	2	0	1	1	1	0	2	2	0	0	1	0	18
B-cell lymphoma/leukemia 11A-like (LOC111949579), transcript variant X1, mRNA	molecular function	nucleic acid binding	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	8
collagen alpha-2(XI) chain-like (LOC111963946), mRNA	cellular component	collagen trimer	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	7
SPRY domain-containing protein 3 (LOC111970186), transcript variant X3, mRNA	molecular function	Dna binding	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	1	0	0	0	7
arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 1-like (LOC111953962), mRNA	molecular function	metal ion binding	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	2	1	0	0	0	7
low-density lipoprotein receptor-related protein 8-like (LOC111971375), transcript variant X2, mRNA	molecular function	calcium ion binding	1	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	6
protein kinase C and casein kinase substrate in neurons protein 1 (LOC111969716), transcript variant X3, mRNA	molecular function	kinase activity	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	6
carbonic anhydrase-related protein 10 (LOC111967825), transcript variant X1, mRNA	molecular function	zinc ion binding	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	5
tumor necrosis factor receptor superfamily member 6B-like (LOC111966075), mRNA	biological process	signal transduction	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	5
protein FAM114A2 (LOC111950452), transcript variant X1, mRNA	biological process	biological process	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4
arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2-like (LOC111968391), mRNA	molecular function	GTPase activator activity	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	3
dedicator of cytokinesis protein 9-like (LOC111959519), mRNA	molecular function	guanyl-nucleotide exchange factor activity	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3
G-protein coupled receptor 54 (LOC111956315), mRNA	molecular function	G protein-coupled receptor activity	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
homeobox protein orthopedia B-like (LOC111975202), transcript variant X1, mRNA	cellular component	nucleus	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3
keratin, type II cytoskeletal 8 (LOC111966854), mRNA	molecular function	protein binding	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3
mitochondrial inner membrane protease subunit 2 (LOC111952461), mRNA	biological process	proteolysis	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
protein-glutamine gamma-glutamyltransferase K (LOC111972247), mRNA	molecular function	transferase activity	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
roundabout homolog 3 (LOC111950617), transcript variant X3, mRNA	cellular component	plasma membrane	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	3
histone deacetylase 5-like (LOC111981176), mRNA	molecular function	hydrolase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3
SNF-related serine/threonine-protein kinase-like (LOC111962281), mRNA	molecular function	protein kinase activity	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
spectrin beta chain, erythrocytic (LOC111968456), transcript variant X2, mRNA	molecular function	actin binding	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3
tubulin tyrosine ligase like 7 (ttll7), transcript variant X1, mRNA	biological process	cellular protein modification process	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
neural cell adhesion molecule 2 (ncam2), mRNA	biological process	cell adhesion	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
oligosaccharyltransferase complex subunit ostc (LOC111960315), mRNA	biological process	protein glycosylation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	2
PH and SEC7 domain-containing protein 2 (LOC111965965), mRNA	molecular function	ARF guanyl-nucleotide exchange factor activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2
potassium two pore domain channel subfamily K member 9 (kcnk9), mRNA	molecular function	potassium channel activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2
protein FAM43A-like (LOC111976060), mRNA	cellular component	mitochondrion	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
protein fem-1 homolog C (LOC111958058), mRNA	cellular component	nucleoplasm	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
retinoic acid induced 14 (rai14), transcript variant X1, mRNA	cellular component	extracellular matrix	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
rho-related GTP-binding protein RhoE (LOC111976449), mRNA	molecular function	GTP binding	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
peptidyl-tRNA hydrolase 1 homolog (ptrh1), transcript variant X4, mRNA	molecular function	aminoacyl-tRNA hydrolase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2
transport and golgi organization 2 homolog (tango2), transcript variant X3, mRNA	cellular component	Golgi apparatus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
tyrosine-protein phosphatase non-receptor type 4 (LOC111976599), transcript variant X1, mRNA	Molecular function	protein tyrosine phosphatase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2
WAS protein family member 1 (wasf1), mRNA	biological process	actin cytoskeleton organization	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
short transient receptor potential channel 3-like (LOC111967605), mRNA	molecular function	calcium channel activity	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2
SPT16 homolog, facilitates chromatin remodeling subunit (supt16h), transcript variant X1, mRNA	cellular component	FACT complex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
TBC1 domain family member 10A (LOC111957627), mRNA	molecular function	U6 snRNA binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	2
piggyBac transposable element-derived protein 5 (LOC111968011), mRNA	molecular function	transposase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
plakophilin-2 (LOC111981758), mRNA	biological process	cell-cell adhesion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
potassium voltage-gated channel subfamily KQT member 5-like (LOC111951960), mRNA	molecular function	voltage-gated potassium channel activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
proteasome subunit beta type-4 (LOC111955067), mRNA	cellular component	proteasome complex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
sin3 histone deacetylase corepressor complex component SDS3 (LOC111957805), transcript variant X1, mRNA	biological process	chromatin organization	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
pyruvate dehydrogenase (acetyl-transferring) kinase isozyme 2, mitochondrial (LOC111981212), mRNA	biological process	carbohydrate metabolic process	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
soluble lamin-associated protein of 75 kDa-like (LOC111961332), transcript variant X2, mRNA	cellular component	intermediate filament	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4 (LOC111966555), transcript variant X2, mRNA	molecular function	6-phosphofructo-2-kinase activity	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
activin receptor type-1B-like (LOC111976988), transcript variant X2, mRNA	biological process	transmembrane receptor protein serine/threonine kinase signaling pathway	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
A-kinase anchor protein 6-like (LOC111968513), transcript variant X1, mRNA	molecular function		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
anillin-like (LOC111955930), transcript variant X1, mRNA	biological process	hemopoiesis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
ATP-sensitive inward rectifier potassium channel 12-like (LOC111976506), mRNA	molecular function	voltage-gated ion channel activity	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
BAI1 associated protein 3 (baiap3), mRNA	biological process	dense core granule maturation	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Bardet-Biedl syndrome 12 (bbs12), mRNA	biological process	convergent extension involved in gastrulation	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
beta-crystallin S-1-like (LOC111951602), transcript variant X2, mRNA	biological process	glycine betaine biosynthetic process from choline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
bromodomain adjacent to zinc finger domain 2B (baz2b), transcript variant X16, mRNA	cellular component	host cell nucleus	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Bruton tyrosine kinase (btk), transcript variant X4, mRNA	biological process	negative regulation of intrinsic apoptotic signaling pathway	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
C->U-editing enzyme APOBEC-2 (LOC111963481), mRNA	molecular function	RNA binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein-like (LOC111954114), mRNA	cellular component	BLOC-1 complex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
cat eye syndrome critical region protein 2 (LOC111962616), transcript variant X1, mRNA	biological process	ATP-dependent chromatin remodeling	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CDK-activating kinase assembly factor MAT1-like (LOC111968663), transcript variant X1, mRNA	cellular component	transcription factor TFIIH holo complex	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
cell surface hyaluronidase-like (LOC111958112), mRNA	biological process	viral process	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
chloride intracellular channel protein 4-like (LOC111963185), mRNA	biological process	chloride transport	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
choline-phosphate cytidylyltransferase A (LOC111979607), transcript variant X1, mRNA	molecular function	nucleotidyltransferase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
cholinergic receptor nicotinic beta 3 subunit (chrnb3), mRNA	cellular component	cell junction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
chromodomain helicase DNA binding protein 7 (chd7), transcript variant X1, mRNA	biological process	face development	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
cytospin-B-like (LOC111981317), transcript variant X3, mRNA	biological process	meiotic cell cycle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
DENN domain containing 5A (dennd5a), transcript variant X4, mRNA	cellular component	trans-Golgi network	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
DENN domain-containing protein 2C (LOC111966511), transcript variant X1, mRNA	molecular function	Rab guanyl-nucleotide exchange factor activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
disks large-associated protein 4-like (LOC111962313), mRNA	biological process	signaling	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
dnaJ homolog subfamily C member 13 (LOC111953250), transcript variant X8, mRNA	cellular component	endosome	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
double-stranded RNA-specific editase 1 (LOC111959749), transcript variant X4, mRNA	molecular function	adenosine deaminase activity	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
E3 ubiquitin-protein ligase RNF152 (LOC111973382), mRNA	molecular function	ubiquitin protein ligase activity	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
echinoderm microtubule-associated protein-like 2 (LOC111949414), transcript variant X3, mRNA	molecular function	microtubule binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
elongation of very long chain fatty acids protein 6 (LOC111965827), mRNA	cellular component	endoplasmic reticulum membrane	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
endoplasmic reticulum lectin 1 (LOC111968068), transcript variant X2, mRNA	molecular function	carbohydrate binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
ephrin-A2-like (LOC111953065), mRNA	molecular function	ephrin receptor binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
erbin (LOC111963937), transcript variant X1, mRNA	molecular function	ErbB-2 class receptor binding	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
extensin	molecular function	structural constituent of cell wall	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
family with sequence similarity 168 member A (fam168a), transcript variant X2, mRNA	biological process	positive regulation of base-excision repair	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
frizzled-10-B (LOC111957788), mRNA	biological process	Wnt signaling pathway	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
G protein-activated inward rectifier potassium channel 1-like (LOC111977554), mRNA	molecular function	inward rectifier potassium channel activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
G protein-coupled receptor activity (mrpl11), transcript variant X1, mRNA	biological process	RNA processing	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
glutaredoxin and cysteine rich domain containing 2 (grxcr2), mRNA	cellular component	microvillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
growth arrest and DNA damage-inducible protein GADD45 gamma (LOC111957399), mRNA	biological process	regulation of cell cycle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
guanine nucleotide-binding protein G(q) subunit alpha (LOC111980819), mRNA	molecular function	G protein-coupled receptor binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
helicase ARIP4 (LOC111966287), transcript variant X1, mRNA	molecular function	ATP binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
heterochromatin protein 1-binding protein 3 (LOC111969980), transcript variant X1, mRNA	cellular component	nucleosome	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
inactive carboxypeptidase-like protein X2 (LOC111974979), mRNA	molecular function	metallocarboxypeptidase activity	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
interferon alpha-like (LOC111980627), mRNA	molecular function	cytokine receptor binding	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
intraflagellar transport 46 (ift46), mRNA	biological process	cilium assembly	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
keratin, type I cytoskeletal 13-like (LOC111982613), mRNA	molecular function	structural molecule activity	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
keratocan-like (LOC111955509), partial mRNA	biological process	visual perception	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
leucine rich repeat and Ig domain containing 1 (lingo1), mRNA	molecular function	epidermal growth factor receptor binding	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
LIM/homeobox protein Lhx1-like (LOC111960709), transcript variant X1, mRNA	molecular function	sequence-specific DNA binding	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
lysophospholipid acyltransferase LPCAT4 (LOC111952481), mRNA	molecular function	transferase activity, transferring acyl groups	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
mitochondrial import receptor subunit TOM20 homolog (LOC111960523), mRNA	cellular component	mitochondrial outer membrane translocase complex	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
neuregulin 3 (nrg3), transcript variant X1, mRNA	molecular function	signaling receptor binding	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
neurogenic differentiation factor 1 (LOC111977678), mRNA	biological process	regulation of transcription, DNA-templated	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
opioid-binding protein/cell adhesion molecule-like (LOC111981966), transcript variant X2, mRNA	molecular function	molecular function	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
paralemmin-1 (LOC111975564), mRNA	biological process	regulation of flower development	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
peptidyl-prolyl cis-trans isomerase FKBP8 (LOC111975390), transcript variant X1, mRNA	molecular function	isomerase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
peroxisome proliferator-activated receptor gamma coactivator-related protein 1 (LOC111978040), mRNA	molecular function	transcription coregulator activity	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
platelet-derived growth factor receptor beta (LOC111965401), transcript variant X3, mRNA	molecular function	platelet-derived growth factor beta-receptor activity	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
prefoldin subunit 4 (LOC111966746), mRNA	biological process	protein folding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
pre-mRNA-splicing regulator WTAP (LOC111954449), transcript variant X5, mRNA	biological process	regulation of alternative mRNA splicing, via spliceosome	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
proline dehydrogenase 1, mitochondrial (LOC111975162), mRNA	biological process	oxidation-reduction process	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
protein EURL homolog (LOC111981575), transcript variant X1, mRNA	biological process	positive regulation of dendritic spine development	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
protein phosphatase 4 regulatory subunit 4 (ppp4r4), transcript variant X1, mRNA	cellular component	cytosol	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ras association domain family member 7 (rassf7), transcript variant X1, mRNA	biological process	regulation of microtubule cytoskeleton organization	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
regulation of nuclear pre-mRNA domain-containing protein 2 (LOC111960658), transcript variant X1, misc_RNA	molecular function	aminoacyl-tRNA ligase activity	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
regulator of G-protein signaling 9 (LOC111968233), mRNA	biological process	G protein-coupled receptor signaling pathway	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
regulator of nonsense transcripts 3A-like (LOC111959860), transcript variant X2, mRNA	biological process	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
alpha-tocopherol transfer protein (LOC111956960), transcript variant X2, mRNA	cellular component	ribosome	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
ATP-binding cassette sub-family F member 2 (LOC111979175), transcript variant X2, mRNA	Molecular function	nucleotide binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
cilia and flagella associated protein 46 (cfap46), mRNA	biological process	axoneme assembly	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
dachshund homolog 1-like (LOC111973053), mRNA	biological process	microtubule-based process	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
DEAH-box helicase 15 (dhx15), mRNA	Molecular function	helicase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
glutamate receptor ionotropic, delta-1-like (LOC111967743), mRNA	molecular function	ionotropic glutamate receptor activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
inter-alpha-trypsin inhibitor heavy chain H3-like (LOC111968607), mRNA	biological process	hyaluronan metabolic process	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
long-chain-fatty-acid--CoA ligase ACSBG2 (LOC111972157), transcript variant X2, mRNA	molecular function	ligase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
mucolipin-2-like (LOC111962952), mRNA	molecular function	cation channel activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
multidrug resistance-associated protein 5 (LOC111965891), mRNA	molecular function	organic anion transmembrane transporter activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
myelin basic protein-like (LOC111965264), transcript variant X1, mRNA	molecular function	structural constituent of myelin sheath	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
nectin-1-like (LOC111980961), transcript variant X1, mRNA	cellular component	cell-cell junction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
RNA exonuclease 4 (LOC111961803), transcript variant X1, mRNA	cellular component	nuclear speck	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
synaptotagmin-1 (LOC111982726), transcript variant X1, mRNA	molecular function	calcium-dependent phospholipid binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
transcription initiation factor TFIID subunit 10 (LOC111977249), mRNA	biological process	DNA-templated transcription, initiation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
tRNA methyltransferase 1 like (trmt1l), mRNA	biological process	methylation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
zinc transporter ZIP11-like (LOC111968220), mRNA	biological process	metal ion transport	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
semaphorin 5B (sema5b), mRNA	molecular function	semaphorin receptor binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
serine/threonine-protein kinase tousled-like 1-B (LOC111975038), mRNA	biological process	regulation of chromatin assembly or disassembly	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
sideroflexin-5 (LOC111981706), transcript variant X1, mRNA	biological process	ion transport	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
sodium- and chloride-dependent GABA transporter 2-like (LOC111970289), mRNA	biological process	transmembrane transport	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
sodium/potassium-transporting ATPase subunit beta-2 (LOC111981831), transcript variant X2, mRNA	biological process	sodium ion transport	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
troponin I, fast skeletal muscle (LOC111971996), transcript variant X1, mRNA	cellular component	troponin complex	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
tumor necrosis factor alpha-induced protein 8-like protein 2 (LOC111955375), mRNA	biological process	regulation of apoptotic process	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
U3 small nucleolar RNA-interacting protein 2 (LOC111970321), mRNA	biological process	rRNA processing	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 9 (LOC111952160), transcript variant X1, mRNA	molecular function	transferase activity, transferring glycosyl groups	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
vacuolar protein sorting-associated protein 13A (LOC111957899), transcript variant X1, mRNA	molecular function	catalytic activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
vegetative cell wall protein gp1-like (LOC111951357), mRNA	molecular function	structural constituent of cuticle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
vitellogenin-like (LOC111975854), mRNA	molecular function	lipid transporter activity	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
WD repeat-containing protein 70-like (LOC111961442), mRNA	molecular function	threonine-type endopeptidase activity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
zinc finger protein, FOG family member 2 (zfpm2), mRNA	molecular function	RNA polymerase II transcription factor binding	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1

Appendix A

Modified Quiagen blood and tissue extraction
Extracting High MW gDNA (Qiagen DNeasy Mini kit) for GBS

Ella and Kia's modified protocol from: Michelle's Optimized Protocol (Dr. Devlin's Lab - DFO)

Legend:
Black/Gray = Qiagen's protocol
Blue = Michelle's optimization
Green = Kia and then Ella's modifications.

Note that Ella has removed some things from Michelle's original protocol that were not necessary, as per testing results

Important points before starting

- If using the DNeasy Blood & Tissue Kit for the first time, read "Important Notes" (page 15).
- All centrifugation steps are carried out at room temperature (15–25°C) in a microcentrifuge.
- vortex DNA only once, very briefly, after removing from the incubator
- RNase A is required for GBS samples. RNase A is not provided in the DNeasy Blood & Tissue Kit (see "Copurification of RNA", page 19).
- Proteinase K must be stored at 4°C upon receiving the kit (Note, Ella is dubious of this, given that it explicitly says to store at room temp, but I do store it in the fridge, and my extractions do not seem to be affected).
- Buffer AW1 and Buffer AW2 are supplied as concentrates. Before using **for the first time**, add the appropriate amount of ethanol (96–100%) as indicated on the bottle to obtain a working solution.

Things to do before starting

- Preheat an incubator to 56°C for use in step 4.
- Ensure you have enough autoclaved tips and 1.5 mL tubes
- Wipe bench with 95% ethanol before starting and at the end of the day. Wipe pipets down with 95% EtOH at the beginning and end of the day. Rinse sample trays at end of the day, first with water and then EtOH.
- Buffer ATL and Buffer AL may form precipitates upon storage. Warm both solutions to 56°C until the precipitates have fully dissolved.
- Label tubes

Procedure

1. Prepare ATL:Proteinase K mixture (made fresh daily). Pulse Vortex 5s. non-filter tips
    - Fin clips (180 µl ATL, 20 µl Proteinase K per sample)
2. Cut up 15 - 20 mg of fish fin (up to 25 mg), dab it with Kimwipe to remove residual ethanol and place it into the 1.5 ml microcentrifuge tube with ATL:Proteinase K(PK):

    A) Two glass containers with 95% ethanol (rinse 1 and rinse 2)
    B) Turn on flame and adjust so blue (and not too high)
    C) Dip tool (scissors and tweezers) in ethanol, wipe, then dip in second wash, shake off extra
    D) Pass it over the flame (it will catch on fire) to burn off the ethanol
    E) Remove tissue from tube, cut with scissors (size of 0 on page)
    F) Dap with Kimwipe (each side) and put in labelled tube

G) Repeat all A-F and dab on different area of Kimwipe or different kimwipe if you aren't certain that you can use a separate area of the tissue

3. Add 200 µl of ATL: Proteinase K mixture into each 1.5 ml microcentrifuge tube with tissue in it. (LABELLED TUBES) non-filter tips

4. Ensure the caps are closed. Place the tubes in an incubator. Incubate at 56°C. (It will become gelatinous)
   - Fin clips 3 hours (Michelle incubates for 1hr, Ella incubates for 3hrs. Length needed depends on how chunky tissue is). Every 45 mins or so, invert tubes to mix. I invert 25 to 30 times

5. During incubation, aliquot some Buffer AE into a 15 mL centrifuge tube and put in incubator.

6. During incubation, prepare AL:EtOH mixture at 1:1 ratio (400 µl of the mixture per sample). Ensure it is at around 56°C. non-filter tips. Also label spin columns

7. After removing tubes from the incubator, pulse vortex very briefly. I usually do two pulses.

8. **Pipet (**Using a 10ul xl filter tip),** add 4 µl RNaseA (100mg/ml) to each sample. Invert the tube 15 times. DO NOT VORTEX or do not pipette up and down. Incubate 10 mins at room temperature.

9. DO THIS STEP QUICKLY. Add 400 µl of Buffer AL:EtOH mix to the sample, and mix thoroughly by inverting (~25-30 times). A white precipitate may form on addition of Buffer AL and ethanol. (Due to the SDS in both solutions are precipitating out from the temperature difference between the 2 solutions mixture). Thus warming the AL:EtOH mixture is very important. Precipitate WILL interfere with the DNeasy procedure in fish tissue especially in 96easy kits. Note that I'm not sure how much I buy this stuff about the precipitate interfering, but heating the buffer does eliminate the precipitate, and I think likely make the DNA sit on the membrane better. Non-filter tips

10. **Pipet (**using a p1000 filter tip)** the mixture from step 7 (including any precipitate) into the DNeasy Mini spin column (**LABELLED**) placed in a 2 ml collection tube (provided). (Be careful not to cross-contaminate). Centrifuge at **6000 x g** (8000 rpm) for 1 min. Discard flow-through and collection tube.*

11. Place the DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW1, and centrifuge for 1 min at **6000** x g (8000 rpm). Discard flow-through and collection tube. Non-filter tips

12. Place the DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW2, and centrifuge for 3 min at **20,000** x g (14,000 rpm) to dry the DNeasy membrane. Discard flow-through and collection tube. Non-filter tips

    It is important to dry the membrane of the DNeasy Mini spin column, since residual ethanol may interfere with subsequent reactions. This centrifugation step ensures that no residual ethanol will be carried over during the following elution.

13. Only for GBS samples, Place the DNeasy Mini spin column in a new 2 ml collection tube (provided), centrifuge for 1 min at 20,000 x g (14,000 rpm) to dry the DNeasy membrane. Discard flow-through and collection tube.

    *** Following the centrifugation step, remove the DNeasy Mini spin column carefully so that the column does not come into contact with the flow-through, since this will result in carryover of ethanol. If carryover of ethanol occurs, empty the collection tube, then reuse it in another centrifugation. **note that ripping off the top of a 1.5ml tube and placing the spin column in it makes this step faster b/c more tubes can fit in the centrifuge.

14. Place the DNeasy Mini spin column in a clean 1.5 ml tube (**FINAL TUBE WITH LABEL**) and pipet 30 µl Buffer AE directly onto the DNeasy membrane. Incubate at 56°C for 5 min, and then centrifuge for 1 min at 6000 x g (8000 rpm) to elute. Non-filter tips
    ** You can elute with TE or AE.

15. On the same DNeasy Mini spin column and 1.5 ml microcentrifuge tube as step 13, pipet another 30 µl Buffer AE directly onto the DNeasy membrane. Incubate at 56°C for 5 min, and then centrifuge for 1 min at 6000 x g (8000 rpm) to elute. Non-filter tips

16. Remove filter (spin column), and the 1.5 ml tube has DNA.

17. Tubes can now be placed in a labeled box and stored in the fridge

For gels:

- Undigested:
    - We get bands 22 – 24,000 BP
    - bands that are ≤ 500 BP = sheered and no good!

Appendix B

Genomic analysis pipeline

The following is the entire pipeline for genomic analysis that we used from PoolParty to graphing CMH testing. The pipeline that follows is not in any specific order and may not be necessary in all genomic pipelines. Futhermore these analyses may not work with different programs and terminals.

Hopefully, this pipeline can assist other researchers that are new to genomics or need help with their pipeline. Again, thank you to Brenna Forrester for her pipeline as it proved instrumental to my understanding and pipeline.

A large part of our analyses was undertaken on Graham and Cedar, both ssh terminals run by Compute Canada. For Mac, terminal is a connection method, for Windows I recommend PuTTY. I suggest familiarizing yourself with Bash script.

Importantly, these scripts are almost absolutely not perfect, but they work (mostly). They are also not exactly the scripts run in my manuscript, some things have been changed to make it easier to digest or because it was an older version

A significant amount of information is available directly from the PoolParty README document found at https://github.com/StevenMicheletti/poolparty/blob/master/README.md

1. Getting Started

Getting access to your program (Graham):

> $ ssh username@graham.computecanada.ca

Moving files:

> $ cd move/to/directory/where/the/file/located

> $ scp <file_to_transfer> <username>@<address>:<directory>

> $ scp Salvelinus_alpinus.zip username@gra-dtn1.computecanada.ca:/home/USER/projects/def-salmo/USER

2. Job Submission

Software installed in the cluster can be loaded using module argument:

> $ module load <tool>/<version>

Switch to a different version of the module:

> $ module switch <tool> <tool>/<version>

Module unload:

> $ module unload <tool>

Check the loaded modules:

> $ module list

Scheduling a job:

~~~ bash

#!/bin/bash

#SBATCH --job-name=example #make this descriptive but short

#SBATCH --account=def-salmo #the account this is under

#SBATCH --cpus-per-task=XX #how many threads you want

#SBATCH --time=0-01:00 #time to run - goes days:hours:minutes

#SBATCH --mem=4000M #amount of memory you want- M = megabytes, G = gigabytes

#SBATCH --output=%x-%j.out #%x=jobname %j=jobid

#SBATCH --mail-user=youremail@email.com

#SBATCH --mail-type=ALL #BEGIN, END, FAIL, REQUEUE

There are plenty of genomics software resources online; check you mother software (eg. <https://docs.computecanada.ca/wiki/Available_software>)

### **3. Packages for PoolParty** ###

**Install all required packages:**

Burrows-Wheeler Aligner (BWA; 07.12) - http://bio-bwa.sourceforge.net/

Fastqc (0.11.7 ) - https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

samblaster (0.1.24) - https://github.com/GregoryFaust/samblaster

samtools (1.5) - http://www.htslib.org/download/

bcftools (1.5) - http://www.htslib.org/download/

Picard Tools (2.17.11) - http://broadinstitute.github.io/picard/

Popoolation2 (1.201) - https://sourceforge.net/p/popoolation2/wiki/Main/

BBMap (37.93) - https://sourceforge.net/projects/bbmap/

**Perl for Popoolation2:**

<https://docs.computecanada.ca/wiki/Perl/en>

<https://sourceforge.net/p/popoolation2/wiki/Manual/>

**R packages:**
~~~

PoolParty will attempts to install required R packages itself; yet, it is highly recommended to manually install packages beforehand:

PPalign: matrixStats, tidyr, stringr, data.table

PPstats: reshape, fBasics, ggplot2, RColorBrewer

PPanalyze: matrixStats, plyr, stringr, data.table, fBasics, ape, metap

4. Preparing a Reference Genome

We used an Arctic char reference genome: <https://www.ncbi.nlm.nih.gov/genome/12179>

Copy chromosomes to a directory:

> $ find ./ -name '*.fa.gz' -exec cp -prv '{}' '/Users/USER/Poolseq.analysis/merge/merge/' ';'

Unzip all zip files within current directory:

> $ find . -name '*.gz' -exec gzip -d {} \;

Rename each fasta file:

> $ brew install rename

> $ rename 's/8036_ref_ASM291031v2_chr/chr/' *.fa

Concatenating multiple chromosome fasta as single fasta file:

> $ cat chr*.fa > genome.fa

Replace fasta header:

directory PATH=/Salvelinus_alpinus/Assembled_chromosomes/seq/

> $ grep ">" genome.fa

~~~

>ref|NC_036860.1| Salvelinus alpinus isolate IW2-2015 linkage group LG20, ASM291031v2, whole genome shotgun sequence

>ref|NC_036861.1| Salvelinus alpinus isolate IW2-2015 linkage group LG21, ASM291031v2, whole genome shotgun sequence

>ref|NC_036862.1| Salvelinus alpinus isolate IW2-2015 linkage group LG22, ASM291031v2, whole genome shotgun sequence

>ref|NW_019942503.1| Salvelinus alpinus isolate IW2-2015 unplaced genomic scaffold, ASM291031v2 Un_scaffold15, whole genome shotgun sequence

>ref|NW_019942504.1| Salvelinus alpinus isolate IW2-2015 unplaced genomic scaffold, ASM291031v2 Un_scaffold16, whole genome shotgun sequence

>ref|NW_019942505.1| Salvelinus alpinus isolate IW2-2015 unplaced genomic scaffold, ASM291031v2 Un_scaffold19, whole genome shotgun sequence
~~~

~~~

Simple liner to get rid of ambiguous header:

~~~ bash

> $ sed 's/ref.*group //' genome.fa > genome2.fa

> $ sed 's/ref.*ASM291031v2 //' genome2.fa > genome3.fa

> $ sed 's/, whole.*sequence//' genome3.fa > genome4.fa

> $ sed 's/, ASM291031v2//' genome4.fa > genome5.fa

> $ sed 's/Un_//' genome5.fa > genome6.fa

~~~

Indexing reference genome using following script:

prep genome.sh

~~~ bash

#!/bin/bash

#SBATCH --job-name=prep_genome

#SBATCH --account=def-salmo

#SBATCH --cpus-per-task=16

#SBATCH --time=0-10:00

#SBATCH --mem=32G #maximum mem for graham

#SBATCH --output=%x-%j.out

#SBATCH --mail-user=youremail@email.com

#SBATCH --mail-type=ALL

export PATH=$PATH:/home/USER/project/dir/

GENOME="saal_lg_genome.fa"

module load nixpkgs/16.09 gcc/7.3.0 bwa/0.7.12

bwa index -a bwtsw ${GENOME}

module load nixpkgs/16.09 gcc/5.4.0 intel/2016.4 samtools/1.5

samtools faidx ${GENOME}

module load picard/2.17.11

java -jar $EBROOTPICARD/picard.jar CreateSequenceDictionary REFERENCE= ${GENOME} OUTPUT=${GENOME}.dict

### **5. Running PPalign** ###

**Name your files:**

The naming convention of the fastq files is essential. The unique ID identifying the library must occur before the first underscore and must match its paired-end mate. The number after the file designates the population or library that the file belongs to - this is particularly useful if individuals are barcoded or populations were sequenced on different lanes.

Your sample name list should be located in your working directory.

**PPalign configuration file:**

> $ nano run_PPalign.sh

> $ <command + v> example

> $ <ctrl + x>

PPalign bash script file: run_PPalign.sh

~~~ bash

#!/bin/bash

#SBATCH --job-name=PPalign

#SBATCH --account=def-salmo

#SBATCH --cpus-per-task=32 #maximum cpu usage of graham

#SBATCH --time=12-00:00 #2 pools require 3 days

#SBATCH --mem=131072M #equivalent of 128G which is maximum memory

#SBATCH --output=%x-%j.out

#SBATCH --mail-user=youremail@email.com

#SBATCH --mail-type=ALL

echo "Starting run at: `date`"

cd /home/USER/projects/def-salmo/01-mis-pool

echo "Current working directory: `pwd`"

echo "Starting load modules"

module load nixpkgs/16.09 gcc/7.3.0 samtools/1.9 bwa/0.7.12 bcftools/1.9 samblaster/0.1.24 r-bundle-bioconductor/3.8 java

module load nixpkgs/16.09 gentoo/2019 fastqc/0.11.7 picard/2.17.11

echo "load completely"
~~~

export PATH=$PATH:/home/USER/projects/def-salmo/01-mis-pool:/home/USER/projects/def-salmo/popoolation2

PPalign.sh align.config

~~~

PPalign configuration file: align.config

~~~ bash

#!/bin/bash

#PPalign configuration file

#Input/Output#

INDIR=/home/USER/projects/def-salmo/USER/poolparty1/00-raw

OUTDIR=/home/USER/projects/def-salmo/USER/poolparty1/01-align

OUTPOP=spg

GENOME=/home/USER/projects/def-salmo/USER/poolparty1/00-gnom/genome.fa

SCAHEAD=scaffold

#Run Parameters#

THREADZ=32

BQUAL=20

MAPQ=10

SNPQ=20

MINLENGTH=25

INWIN=15

MAF=0.05

KMEM=Xmx132g

MINDP=10

#Run-types#

SPLITDISC=on

INDCONT=off

QUALREPORT=on

#Optional Parameters#

ALIGNONLY=

USEVCF=

#Dependency Locations#

BCFTOOLS=bcftools

FASTQC=fastqc

BWA=bwa

SAMBLASTER=samblaster

SAMTOOLS=samtools

PICARDTOOLS=/home/USER/.local/easybuild/software/2017/Core/picard/2.17.11/picard.jar

BBMAPDIR=/cvmfs/soft.computecanada.ca/easybuild/software/2017/avx2/Compiler/intel2016.4/bbmap/37.93/

POOL2=/home/USER/projects/def-salmo/USER/popoolation2/

#Languages needed on system#

#bash/shell

#perl

#java

#R

~~~

Submitting configuration file:

> $ sbatch ./run_PPalign.sh

Checking progress:

> $ squeue -u USER

Checking output files:

> $ grep "ALERT " <log.out> > alert.txt

> $ grep "ERROR" <log.out> > error.txt

> $ grep "duplicates" <log.out> > duplicates.txt

> $ grep "Low quality" <log.out> > low_quality.txt

> $ grep "Contaminant" <log.out> > contaminant.txt

> $ grep "SNPs" <log.out> > SNPs.txt

> $ grep "indel" <log.out> > indel.txt

Quality checks:
~~~

transfer the fastqc output to local

> $ cd move/to/output/directory

> $ scp USER@graham.computecanada.ca:/home/USER/projects/def-salmo/poolparty1/01-align/fastqc/* .

multiqc is versatile tool to check the multiple fastqc outputs.

https://multiqc.info

BAM mapping checks for quality

> $ cd 01-align/reports

> $ more _aln_report.txt

The total number of reads refers to the total number of reads after quality trimming.

Secondary reads are those that have a high chance of aligning to multiple positions in the ref genome.

We should see 100% mapped every time since the report is performed on a bam that has been aligned to ref genome. However, we are generally interested in properly paired as these are the reads that have aligned with their mate in proper orientation and will be retained for SNP calling.

6. Running PPstats

Configuration file:

#!/bin/bash

#PPstats configuration file

#Specify Files#

FAI=/home/USER/projects/def-salmo/SPGBBpoolparty1/00-gnom/genome.fa.fai

MPILEUP=/home/USER/projects/def-salmo/SPGBBpoolparty1/01-SPGBB-align/spg_stats.mpileup

OUTDIR=/home/USER/projects/def-salmo/SPGBBpoolparty1/02-stat

OUTFILE=PP_stats.txt

#Parameters#

SCAFP=scaffold

THREADZ=12

MINCOV=1

MAXCOV=100

#Languages needed on system#

#bash/shell

#R

Submitting configuration file:

> $ sbatch ./run_PPalign.sh

7. Running PPanalyze

SNP calling:

PPmanht

~~~ bash

# loading module

module load nixpkgs/16.09 gcc/7.3.0 r/3.5.2 r-bundle-bioconductor/3.8

cd 01-align

cp mis_CHRbp.txt ../03-analyze/

cd 03-analyze/cheno

../../PPmanhat.sh -i 1-2_analyze.fet -o FET_plot_cheno_00vs11_v1.1 -a -log10p -s scaffold -1 blue -2 red -c ../mis_CHRbp.txt -p TRUE

cd ../papas

../../PPmanhat.sh -i 4-5_analyze.fet -o FET_plot_papas_00vs11_v1 -a -log10p -s scaffold -1 blue -2 red -c ../mis_CHRbp.txt -p TRUE

cd ../papas1116

../../PPmanhat.sh -i 3-5_analyze.fet -o FET_plot_papas_11vs16_v1 -a -log10p -s scaffold -1 blue -2 red -c ../mis_CHRbp.txt -p TRUE

cd ../ruper

../../PPmanhat.sh -i 6-7_analyze.fet -o FET_plot_ruper_00vs11_v1 -a -log10p -s scaffold -1 blue -2 red -c ../mis_CHRbp.txt -p TRUE

~~~

Running Subset Comparisons for CMH output:

#Populations for analysis#

POPS=X:X, X:X

#Input files and names#

PREFIX=lakeX-vs-lakeY

COVFILE=/home/USER/projects/def-salmo/02-spg-pool/spg_coverage.txt

SYNC=/home/USER/projects/def-salmo/02-spg-pool/spg.sync

FZFILE=/home/USER/projects/def-salmo/02-spg-pool/spg.fz

BLACKLIST=/home/USER/projects/def-salmo/02-spg-pool/spg_poly_one.txt

OUTDIR=/home/USER/projects/def-salmo/02-spg-pool/03-analyze/lakeX-lakeY

#Types of Analyses#

FST=on

SLIDINGFST=on

FET=on

NJTREE=on

#Global Parameters#

MINCOV=2

MAXCOV=20

MAF=0.05

#FST Parameters#

FSTTYPE=karlsson

WINDOW=10000

STEP=5000

NIND=20

#NJ Tree Parameters

STRWINDOW=10000

BSTRAP=1000

AFFILT=1

METHOD=mean

Running CMH tests:

When you get a sync file from the ppanalyze, you can run the CMH test using the perl script implemented in Popoolation2.

$ perl cmh-test.pl --input p1_p2_p1_p2.sync --output p1_p2_p1_p2.cmh --min-count 20 --min-coverage 20 --max-coverage 100 --population 1-3,2-4

parameters

Using input dgmg.sync

Using output dgmg.cmh

Using min-count 20

Using min-coverage 20

Using max-coverage 100

Using population 1-3,2-4

Using min-pvalue 1

Remove temporary files 0

Using test 0

Using help 0

Running GWAS perl script:

$ perl <popoolation2-path>/export/cmh2gwas.pl --input p1_p2_p1_p2.cmh --output p1_p2_p1_p2.gwas --min-pvalue 1.0e-20

When you get the .cmh file from the ppanalyze and then you can compare the locus ID with the SnpEff output to identify the candidate genes that changed in the allele frequency.

By the way, SnpEff analysis will return the gene identity from the reference genome (in this case, the Arctic charr). I have run the SnpEff on the Galaxy after making the genome DB.

To generate the genome DB, gff3 and fasta files are required. I have download the files from NCBI.

It has to be matched the chromosome name between genome DB and the vcf file. So I have checked the chromosome name using the SnpEff check-chromosome argument, and then the chromosome name is changed accordingly.

**8. Species Dissimilarity Calculation ###**

```
M<-read.csv("Jaccard Dissimilarity1.csv")
attach(M)
M2 <- M[,-1]
rownames(M2) <- M[,1]
M2
```

```
##     BKTR BLTR LNDC LNSC MNWH RNTR WCTR
## CM     1    0    0    0    0    0    0
## McM    1    1    0    0    0    0    1
## DM     1    1    1    1    1    1    1
## HM     1    0    0    0    0    0    0
```

```
## MaM   1   0   0   0   0   0   1
## TM    1   0   0   0   0   0   0
## MuM   1   0   1   0   0   0   0
## OM    1   0   0   0   0   0   0
## RM    1   0   0   0   0   0   0
```

```
install.packages("adespatial")

library(adespatial)

beta.div(Y = M2, method = "jaccard", sqrt.D = FALSE, samp = F, nperm = 999, adj = T, save.D = T, clock = T)
```

```
## Time for computation = 0.000000  sec

## $beta
##   SStotal   BDtotal
## 1.8187831 0.2273479
##
## $SCBD
## [1] NA
##
## $LCBD
##        CM       McM        DM        HM       MaM        TM       MuM        OM        RM
## 0.04307071 0.19361616 0.27288889 0.04307071 0.14634343 0.04307071 0.17179798 0.04307071
0.04307071
##
## $p.LCBD
##    CM   McM    DM    HM   MaM    TM   MuM    OM    RM
## 1.000 0.015 0.001 1.000 0.152 1.000 0.054 1.000 1.000
##
## $p.adj
##    CM   McM    DM    HM   MaM    TM   MuM    OM    RM
## 1.000 0.120 0.009 1.000 0.912 1.000 0.378 1.000 1.000
##
## $method
## [1] "jaccard"      "sqrt.D=FALSE"
##
## $note
## [1] "Info -- D is Euclidean because beta.div outputs D[jk] = sqrt(1-S[jk])" "For this D functions, use
beta.div with option sqrt.D=FALSE"
##
## $D
##          CM       McM        DM        HM       MaM        TM       MuM        OM
## McM 0.8164966
## DM  0.9258201 0.7559289
## HM  0.0000000 0.8164966 0.9258201
## MaM 0.7071068 0.5773503 0.8451543 0.7071068
```

```
## TM  0.0000000 0.8164966 0.9258201 0.0000000 0.7071068
## MuM 0.7071068 0.8660254 0.8451543 0.7071068 0.8164966 0.7071068
## OM  0.0000000 0.8164966 0.9258201 0.0000000 0.7071068 0.0000000 0.7071068
## RM  0.0000000 0.8164966 0.9258201 0.0000000 0.7071068 0.0000000 0.7071068 0.0000000
##
## attr(,"class")
## [1] "beta.div"
```

9. FST Calculation

```
library(poolfstat)

library(ggplot2)

library(ggrepel)

library(RColorBrewer)
```

```
pool.data = popsync2pooldata((sync.file="M-F_analyze_bl.sync"), poolsizes=c(34, 40, 16, 16, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40), min.rc = 1, min.cov.per.pool = 1, max.cov.per.pool = 200, min.maf = 0.05, noindel = TRUE)
```

```
PW_fst <- computePairwiseFSTmatrix(pool.data, method = "Anova", min.cov.per.pool=20, max.cov.per.pool = 200, min.maf = 0.05, output.snp.values = TRUE)
```

```
pcoa <- read.csv("poolfstat.filtered.fst.csv", header = FALSE) #saved PW_fst above

pcoa.matrix <- data.matrix(pcoa)

euc.matrix <- dist(pcoa.matrix, 'euclidean')


pool.data = popsync2pooldata((sync.file="M-F_analyze_bl.sync"), poolsizes=c(34, 40, 16, 16, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40), min.rc = 1, min.cov.per.pool = 1, max.cov.per.pool = 200, min.maf = 0.05, noindel = TRUE)
```

```
pco <- pco(euc.matrix, k = 2)
```

```
sumeigen=pco$eig[1]+pco$eig[2]+pco$eig[3]+pco$eig[4]+pco$eig[5]+pco$eig[6]+pco$eig[7]+pco$eig[8]
+pco$eig[9]+pco$eig[10]+pco$eig[11]+pco$eig[12]+pco$eig[13]+pco$eig[14]+pco$eig[15]+pco$eig[16]+
pco$eig[17]+pco$eig[18]

eig1=pco$eig[1]

eig2=pco$eig[2]

perc_eig1=eig1/sumeigen

perc_eig2=eig2/sumeigen

plot(pco)
```

```
pco.ggplot <- data.frame(cbind(c("CBF", "CBM", "MCF", "MCM", "DGF", "DGM", "HLF", "HLM", "MGF",
"MGM", "TMF", "TMM", "MDF", "MDM", "OLF", "OLM", "RSM", "RSF"), as.numeric(pco$points[,1]),
as.numeric(pco$points[,2])))

pco.ggplot$X2 <- as.numeric(as.character(pco.ggplot$X2))

pco.ggplot$X3 <- as.numeric(as.character(pco.ggplot$X3))

colnames(pco.ggplot) <- c("site", "PCoA1", "PCoA2")
```

```
filtered <- ggplot(pco.ggplot, aes(x=PCoA1, y=PCoA2)) + geom_point(colour="chartreuse4") +
geom_point(data=pco.ggplot[c(3, 4), ], aes(x=PCoA1, y=PCoA2), colour="purple4") +
geom_label_repel(aes(label = site), size = 3, hjust = 0, nudge_x = 0.003, nudge_y = - 0.00,
colour="purple4") + geom_label_repel(aes(label = site), size = 3, hjust = 0, nudge_x = 0.003, nudge_y = -
0.00, colour="chartreuse4", show.legend = FALSE) + theme(legend.position="none") +
geom_label_repel(data=pco.ggplot[c(3, 4), ], aes(label = site, x=PCoA1, y=PCoA2), colour="purple4", size
= 3, hjust = 0, nudge_x = 0.003, nudge_y = - 0.00, show.legend = FALSE) + theme_bw() +
theme(axis.text=element_text(size=13), axis.title=element_text(size=15,face="bold"), panel.border =
element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank(), axis.line =
element_line(colour = "black")) + labs(x = "PCoA 1 (XX.XX%)", y = "PCoA 2 (XX.XX%)")
```

**10. Stairway Plot ###**

Input settings are formatted as such for the stairway plot program.

popid: pop1 # id of the population (no white space)
nseq: 4 # number of sequences
L: 14817 # total number of observed nucleic sites, including polymorphic and monomorphic
whether_folded: true # whethr the SFS is folded (true or false)
SFS: 5224 7623.999999999998 # snp frequency spectrum: number of singleton, number of doubleton, etc. (separated by white space)

#smallest_size_of_SFS_bin_used_for_estimation: 1 # default is 1; to ignore singletons, change this number to 2
#largest_size_of_SFS_bin_used_for_estimation: 2 # default is nseq/2 for folded SFS
pct_training: 0.67 # percentage of sites for training
nrand: 1 2 3 4 # number of random break points for each try (separated by white space)
project_dir: pop1_fold # project directory
stairway_plot_dir: pop1_2017 # directory to the stairway plot files
ninput: 200 # number of input files to be created for each estimation
#output setting
mu: 1.2e-8 # assumed mutation rate per site per generation
year_per_generation: 3 # assumed generation time (in years)
#plot setting
plot_title: two-epoch_fold # title of the plot
xrange: 0.1,10000 # Time (1k year) range; format: xmin,xmax; "0,0" for default
yrange: 0,0 # Ne (1k individual) range; format: xmin,xmax; "0,0" for default
xspacing: 2 # X axis spacing
yspacing: 2 # Y axis spacing
fontsize: 12 # Font size

**11. UPGMA dendogram ###**

```
library(vcfR)

library(poppr)

library(ape)

library(RColorBrewer)

spgVCF <- read.vcfR("spgtree.VCF") #this is the vcf file from our PPalign step

spggen <- vcfR2genlight(spgVCF)

pop.data <- read.table("popfile.txt", sep = "\t", header = TRUE)

all(colnames(spgVCF@gt)[-1] == pop.data$Pools)

ploidy(spggen) <- 2

pop(spggen) <- pop.data$pops

spggen

tree <- aboot(spggen, tree = "upgma", distance = bitwise.dist, sample = 100, showtree = F, cutoff = 10, quiet = T)

pdf(file = "spg.UPGMAtree.pdf", width = 6, height = 12)

cols <- brewer.pal(n = 9, name = "Paired")

plot.phylo(tree, cex = 0.8, font = 2, adj = 0, tip.color = cols[pop(spggen)])

nodelabels(tree$node.label, adj = c(1.3, -0.5), frame = "n", cex = 0.8, font = 3, xpd = TRUE)
```

```
legend('topleft', legend = c("Cobb", "Mcnair", "Dog", "Helen", "Margaret", "Temple", "Mud", "Olive", "Ross"), fill = cols, border = TRUE, bty = "n", cex = 2)

axis(side = 1)

title(xlab = "Genetic distance (proportion of loci that are different)")

dev.off()
```

12. Nucleotide Diversity

Before Nucleotide diversity could be calculated, mpileup files for each population had to be calculated. These were run in ssh.

Create the configuration file:

```
samtools mpileup -B /home/USER/projects/def-salmo/SPGBBpoolparty1/01-SPGBB-align/BAM/pop_18.bam  -o /home/USER/projects/def-salmo/SPGBBpoolparty1/04-mpileups/pop18.mpileup
```

Run the shell script:

```
#!/bin/bash

# -----------------------------------------------------------------

# Samtools mpileup

# -----------------------------------------------------------------

#SBATCH --job-name=mpileups

#SBATCH --account=def-salmo

#SBATCH --cpus-per-task=16

#SBATCH --time=0-06:00

#SBATCH --mem=1G

# -----------------------------------------------------------------

echo "Current working directory: `pwd`"

echo "Starting run at: `date`"

# -----------------------------------------------------------------

module load nixpkgs/16.09

module load gcc/5.4.0

module load intel/2016.4
```

module load samtools/1.5

/home/USER/path/to/mpileup.config

--

echo "Job finished with exit code $? at: `date`"

--

Convert mpileups to sync files:

#!/bin/bash

#SBATCH --job-name=nucdiv

#SBATCH --account=def-salmo

#SBATCH --cpus-per-task=16

#SBATCH --time=1-00:00

#SBATCH --mem=32000M

#SBATCH --output=%x-%j.out

#SBATCH --mail-user=youremail@email.ca

#SBATCH --mail-type=ALL

echo "Starting run at: `date`"

echo "Current working directory: `pwd`"

echo "Starting load modules"

module load perl/5.22.4

echo "load completely"

perl /home/bbrookes/projects/def-salmo/popoolation_1.2.2/Variance-sliding.pl --measure pi --input /home/bbrookes/projects/def-salmo/SPGBBpoolparty1/04-mpileups/pop10.mpileup --snp-output /home/bbrookes/projects/def-salmo/SPGBBpoolparty1/05-nucdiv/pop10.snps --fastq-type sanger --pool-size 40 --min-count 4 --min-coverage 20 --max-coverage 200 --window-size 250 --step-size 250

Removing blacklisted loci (we skipped this part once as well to confirm the difference between bi and multialleleic loci):

This script runs all populations in parallel to take out a blacklist pattern file from an input text file (removing blacklist loci from population files)

#!/bin/bash

```
#SBATCH --job-name=strictnd
#SBATCH --account=def-salmo
#SBATCH --time=00:30:00
#SBATCH --cpus-per-task=10
#SBATCH --mem-per-cpu=3G
#SBATCH --output=%x-%j.out
#SBATCH --mail-user=youremail@email.ca
#SBATCH --mail-type=ALL
#SBATCH --array=1-18
input_file=pop${SLURM_ARRAY_TASK_ID}tabout.txt
output_file=strictpop${SLURM_ARRAY_TASK_ID}.txt
patterns_file=polyonetaboutfinal.txt
parallel -a $input_file --pipepart --block -1 grep -Fvf $patterns_file > $output_file
```

Running Nucleotide Diversity metrics in R:

```
pop1 <- read.delim("pop1final.txt", header = FALSE, sep = "\t", dec = ".")
colnames(pop1) <- c("chr", "position", "Num.of.SNPs", "frac.of.cov", "pi")
pop1$pi<-as.numeric(as.character(pop1$pi))
pi1<-mean(pop1$pi, na.rm=TRUE) #Do one of these for each pop
pi_matrix <- matrix(c("CbF", "CbM", "McF", "McM", "DgF", "DgM",
"HlF", "HlM", "MgF", "MgM", "TpF", "TpM", "MdF", "MdM",
"OlF", "OlM", "RsF", "RsM", pi1, pi2, pi3, pi4, pi5, pi6, pi7, pi8,
pi9, pi10, pi11, pi12, pi13, pi14, pi15, pi16, pi17, pi18), nrow = 18, ncol = 2)
colnames(pi_matrix) <- c("Site", "NucleotideDiversity")
pi_DF <- as.data.frame(pi_matrix)
write.csv(pi_DF,'pi_DF.csv')
pi_DF<-read.csv("pi_DF.csv")
pi_DF$NucleotideDiversity <- as.numeric(as.character(pi_DF$NucleotideDiversity))
tiff('pi_figure.tiff', units="in", width=5, height=5, res=300)
```

```
ggplot(data=pi_DF, aes(x=Site, y=NucleotideDiversity)) + geom_point(colour="purple4", size = 3) +
labs(y= "Nucleotide Diversity") + geom_point(data=pi_DF[c(1, 3, 5, 7, 9, 11, 13, 15, 17), ], aes(x=Site,
y=NucleotideDiversity),

        colour="chartreuse4", size = 3) + theme_bw() + ylim(0, 0.006) +
theme(axis.text=element_text(size=13), axis.title=element_text(size=15,face="bold"),

    panel.border = element_blank(), panel.grid.major = element_blank(),

    panel.grid.minor = element_blank(), axis.line = element_line(colour = "black"))

dev.off()
```

13. Unfiltered ND vs Filtered ND

As above, Nucleotide diversity was conducted identically, except each population was not filtered with a blacklist file.

14. Nucleotide Diversity graphs and correlations to variables

```
library(ggplot2)

library(dplyr)

library(tidyr)

library(car)

library(readr)

library(betareg)

library(MuMIn)
```

Load in the file variables, and add nucleotide diversity as a column

```
pi_DF<-read.csv("pi_DF.csv")

pi_DF$NucleotideDiversity <- as.numeric(as.character(pi_DF$NucleotideDiversity))

Allvar<-read.csv("Allvariables.csv")

rownames(Allvar)<-c("CF","CM", "McF", "McM", "DF", "DM", "HF", "HM",

            "MaF", "MaM", "TF", "TM", "MuF", "MuM", "OF", "OM",

            "RF", "RM")

attach(Allvar)
```

```
scaled<-scale(Allvar)
write.csv(scaled, "scaledAllVariables.csv")
```

at this point we add nucleotide diversity as a column

```
Allvarscaled<-read.csv("scaledAllVariablesND.csv")
scaleddf<-data.frame(Allvarscaled)
attach(scaleddf)
```

now lets draw a correlation matrix, and remove everything with a 0.7 cuttoff

```
library(psych)
pairs.panels(envframe)
```

now we are going to use vif and remove everything with a cutoff of 10

```
envvif<- lm(scaleddf$ND~., envframe)
vif(envvif)
```

both of the above were done for environmental variables, stocking variables, and then the remaining variables were run together

```
AllvarscaledM<-read.csv("scaledAllVariablesNDM.csv")
scaleddfM<-data.frame(AllvarscaledM)
```

For this example, we are examining remaining variables for male populations and we can see betadiversity and AIC values

```
weleM<-betareg(scaleddfM$ND~scaleddfM$Elevation)
wvolM<-betareg(scaleddfM$ND~scaleddfM$Watervolm3)
wpHM<-betareg(scaleddfM$ND~scaleddfM$pH)
wzoopM<-betareg(scaleddfM$ND~scaleddfM$Zoop)
wtribsM<-betareg(scaleddfM$ND~scaleddfM$Numbertribs)
wmacroM<-betareg(scaleddfM$ND~scaleddfM$Macro)
wstockM<-betareg(scaleddfM$ND~scaleddfM$Numberstocked)


summary(weleM)
summary(wvolM)
```

```
summary(wpHM)

summary(wzoopM)

summary(wtribsM)

summary(wmacroM)

summary(wstockM)


AICc(weleM, wvolM, wpHM, wzoopM, wtribsM, wmacroM, wstockM)
```

Below is an example of how we graphed our data

First, we had to create an abline by predicting x values based on nucleotide diversity values, of which only the

```
predictM<-predict(unscaledvolM, newdata = data.frame((unscaledvarsM$Watervolm3 =
c(0,205689.555556,205689.555556*2,205689.555556*3,205689.555556*4,
                                                    205689.555556*5,205689.555556*6, 205689.555556*7,
205689.555556*9))))

predictF<-predict(unscaledvolF, newdata = data.frame((unscaledvarsF$Watervolm3 =
c(0,205689.555556,205689.555556*2,205689.555556*3,205689.555556*4,
                                                    205689.555556*5,205689.555556*6, 205689.555556*7,
205689.555556*9))))


ggplot(data=unscaledvars, aes(x=Watervolm3, y=ND, colour=sex)) + geom_point(size = 3) +
 labs(x= "Water Volume m3", y= "Nucleotide Diversity")  + theme_bw() +  ylim(0.003, 0.006) +
 theme(axis.text=element_text(size=13), axis.title=element_text(size=15,face="bold"),
      panel.border = element_blank(), panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(), axis.line = element_line(colour = "black")) +
 geom_abline(intercept = 0.004589194, slope= -4.6694727e-10, color = "#00BFC4", size = 1) +
 geom_abline(intercept = 0.004622386, slope= -6.1523137e-10, color = "#F8766D", size = 1)
```

15. Deleterious mutations

Most of the deleterious mutation calculation was done through SnpEff on Galaxy using the vcf file from the PPalign module and our reference genome. We followed the SnpEff manual for completion of this step.

16. PCAdapt

```
library(pcadapt)
pool.data <- read.table("spg.af.txt")
pool.data.transpose <- t(pool.data)
filename <- read.pcadapt(pool.data.transpose, type = "pool")
res <- pcadapt(filename)
plot(res,option="screeplot")
plot(res, option = "manhattan")
plot(res, option = "qqplot")
hist(res$pvalues, xlab = "p-values", main = NULL, breaks = 50, col = "orange")
plot(res, option = "stat.distribution")
plot(res,option="scores",i=1,j=2)
poplist.int <- c(rep("Cobb", 1),rep("Margaret", 1), rep("Temple", 2), rep("Mud", 2), rep("Olive", 2), rep("Ross", 2), rep("Cobb", 1), rep("Mcnair", 2), rep("Dog", 2), rep("Helen", 2), rep("Margaret", 1))
plot(res, option = "scores", pop = poplist.int)
x <- pcadapt(filename,K=3)
summary(x)
plot(x, option="manhattan")
plot(x, option="qqplot", threshold=0.1)
hist(x$pvalues,xlab="p-values", main=NULL, breaks=50)
plot(x, option="stat.distribution")
padj_bonf <- p.adjust(x$pvalues,method="bonferroni")
alpha <- 0.05
outliers <- which(padj_bonf < alpha)
length(outliers)
write.table(outliers, "pcadapt.outliersk3.Bonferroni.0.05.txt")
```

17. RDA

We raised the memory limit so that our machine could handle the size of our allele frequency file. Running all ~28 million SNPs is not possible on a conventional computer due to RAM limitations (to my knowledge)

```
library(psych)
library(vegan)
memory.limit(500000)
RDA.af<-read.table("AF.for.RDA.pos.txt")
```

We load in our environmental variables, then run a pairs.panels and remove all over 0.7 correlation

```
RDAenv0<-read.csv("ScaledEnvRDA0.csv")
pairs.panels(RDAenv0)
RDAenv1<-RDAenv0[,-1]
RDAenv2<-RDAenv1[,-7]
RDAenv3<-RDAenv2[,-7]
RDAenv4<-RDAenv3[,-7]
RDAenv5<-RDAenv4[,-2]
```

Now lets run the RDA

```
RDA<- rda(RDA.af ~ ., data=RDAenv5, scale =T)
RDA
RsquareAdj(RDA)
summary(eigenvals(RDA, model = "constrained"))
```

last lets check the significance (this is where ours showed no significance). If yours is significant continue to follow the pipeline of Brenna Forrester et al.

```
signif.full<- anova.cca(RDA, parallel=getOption("mc.cores"))
signif.full
```

**18. CMH ###**

See step #7 “Running PPanalyze” for how to run CMH tests using the PoolParty Pipeline, and the methods section of our publication on how candidate genes were identified for function.

**19. Sex differences through FST ###**

FST was calculated as above in step #9.

Appendix C

Seep, inlet, and outlet data was measured by circumnavigation of each lake. A GPS point was taken at each location using a *Garmin etrex 20x*, and length measurements for each inlet/outlet/seep were taken up to 2000m from each lake (Gowan & Fausch, 1996), or until a passage barrier was reached, via GPS tracking and confirmed with Google Earth version 9.2.58.1 (Google, available online). Spawning sites were estimated by combining the number of all discernable seeps, inlets, and outlets in each lake; in rare cases where no connectivity was found, data was marked as N/A. Along each length measurement, seven equidistant transects were conducted with five flow-perpendicular sampling points, to measure: depth, substrate, and water velocity along the thalweg (Hydromatch, n.d.). Connectivity to other bodies of water, catchment area, and bathymetry was calculated using ArcGIS version 10.3.1 and obtained from Parks Canada records. Two *YSI* measurements were take at each lake at every 1m of depth, while *HOBO* loggers were set to record every 30 minutes at 0.5m depth; both measurements were conducted at/above the deepest part of each lake and were used to measure pH, and temperature. Macroinvertebrate sampling was done through averaging eight 1m littoral subsamples of kicked substrate collected with a D-frame net and filtered through a 500-µm sieve bucket and stored in 95% EtOH. Zooplankton samples were taken by averaging eight pelagic subsamples of Wisconsin 54-µm whole column net pulls across the lake and stored in 95% EtOH. Lastly, a Jaccard dissimilarity index for each lake was calculated using presence-absence data collected from the sampling period (R package Adespatial, v 0.3-8; Stéphane Dray et al., 2020).