AN EXAMINATION OF THE UTILITY OF LARGE GENOMIC DATASETS FOR GENETIC MONITORING: AN ATLANTIC SALMON (*Salmo salar*) CASE STUDY

by

Kristin Bethany Watson

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

at

Dalhousie University
Halifax, Nova Scotia
December 2020

To Charles Watson; the man who taught me to fish

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Effective fisheries management requires an awareness of population demography and the spatial scale of population structuring, yet traditional approaches to quantifying both can be labour intensive and expensive. Here I explore the utility of large genomic datasets to characterize population structure, estimate effective population size, and monitor population status in Atlantic Salmon (*Salmo salar*) in Placentia Bay Newfoundland, Canada using a 220K SNP array. Population structure was linked to a temperature associated chromosomal polymorphism. Populations were small ($\widehat{N}_b < 350$) and currently declining. Simulations suggested that large genomic datasets (≥ 100 microsatellites or ≥ 1000 SNPs) enabled accurate detection of population declines >30%. As such, I demonstrate that large genomic datasets allow the identification of fine-scale spatial structuring, the structuring forces involved, and provide a cost effective and accurate approach to monitor population status in the wild.

LIST OF ABBREVIATIONS AND SYMBOLS

bp	before present
c	Recombination rate
Chr	Chromosome
CI	Confidence interval
CIGENE	Centre of Integrative Genetics
COSEWIC	Committee on the Status of Endangered Wildlife in Canada
cM	Centimorgan
DNA	Deoxyribonucleic acid
F	Fixation index
F_{IS}	Inbreeding coefficent
F_{ST}	Proportion of total genetic variance in a subpopulation
GO	Gene ontology
H_e	Expected heterozygosity
H_o	Observed heterozygosity
IBD	Isolation by distance
ID	Identification
K	Number of populations
Kb	kilobase
km	Kilometre
L	Loci
LD	Linkage disequilibrium
LG	Linkage group
m	metre
MAF	Minimum allele frequency
n	Sample size (empirical)
N	Population size
N_b	Effective number of breeders
$\widehat{N}_b$	Naïve estimate of effective of number of breeders
N_c	Census population size

N_e	Effective population size
$\widehat{N}_e$	Naïve estimate of effective population size
p	p-value: expected false positive rate
PC	Principal component
PCA	Principal components analysis
P_{crit}	Minimum allele frequency
q	q-value: means to control the positive false discovery rate
r^2	Correlation coefficient
RDA	Redundancy analysis
S	Sample size (simulations)
SD	Standard deviation
SNP	Single nucleotide polymorphism
Ssa	Chromosome
t	Number of generations

ACKNOWLEDGEMENTS

There are many people without whom this work would not have been possible. First and foremost, I thank my supervisors Dr. Ian Bradbury and Dr. Paul Bentzen for encouraging me to read widely and think critically, and providing the opportunity to travel, explore, fish, and grow as a young scientist. I also thank Dr. Daniel Ruzzante, Dr. Laura Weir, Dr. Sarah Lehnert, and Dr. Kerry Naish for their valuable contributions as members of my ATC and thesis committees. I thank Dr. Robin Waples for providing guidance and sharing his lifetime of knowledge on effective population size estimation.

I thank Steven Duffy for coordinating sample collection, logistics planning for fieldwork, and sample management, as well as, all members of the Fisheries and Oceans Salmonid Branch in Newfoundland who tirelessly collected samples from around Placentia Bay. I also thank Lorraine Hamilton for assisting with DNA extraction and the Centre for Integrative Genetics (CIGENE) for genotyping.

I thank all members of the Bradbury, Bentzen, and Ruzzante labs for many valuable discussions broadening my scope of knowledge, and analytical assistance. Thanks to Dr. Tony Kess and Dr. Tony Einfeldt for their coding expertise. Dr. Kara Layton, Dr. Melissa Holborn, Ben Perriman, Stanley King, and Robert Fairweather; thanks for editing, thoughtful discussions, and your continued support. Many thanks to Ian Paterson for his patience in training me in the lab. I particularly want to thank Dr. Sarah Lehnert, to whom I am endlessly grateful; her contributions to learning to write code, brainstorming and theoretical discussions, editing, and support have been invaluable.

This work was funded by a Canada Graduate Scholarship from The Natural Sciences and Engineering Research Council of Canada (NSERC), a Nova Scotia Graduate Scholarship (Government of Nova Scotia), the Patrick F. Lett Bursary (Dalhousie University), and the PARR Initiative (Fisheries and Oceans Canada).

Lastly, I want to extend a special thanks to Sarah Salisbury for the many conversations about all things population genetics, Susan Ready for her unwavering confidence, and family, particularly Matthew Conlin, for their unending support and encouragement.

CHAPTER 1 – Introduction

1.1 Introduction

Effective fisheries management requires an awareness of the spatial scale of structuring and population demographic change to ensure sustainable harvest (Hutchings & Reynolds, 2004; Ovenden et al., 2016), forecast future vulnerability (Hare et al., 2011; Fordham, Brook, Moritz, & Nogués-Bravo, 2014; Lehnert et al., 2019a), and inform conservation management actions (Luikart, Ryman, Tallmon, Schwartz, & Allendorf, 2010; Leroy et al., 2018). Traditional monitoring approaches (e.g., angling surveys and counting fences) are often logistically challenging, and there is considerable interest in genetic monitoring that promises to reduce cost and increase the scope and scale of monitoring efforts (Schwartz, Luikart, & Waples, 2007; Palstra & Ruzzante, 2008; Hare et al., 2011; Frankham, Bradshaw, & Brook, 2014). Genome-scale data, now readily available for many non-model organisms, have the potential to improve accuracy of population genetic parameters, such as effective population size (N_e) (Waples, Larson, & Waples, 2016; Allendorf, 2017). Genome-scale data also highlight the role of genomic architecture in adaptive divergence (Campbell, Poelstra, & Yoder, 2018; Wellenreuther, Mérot, Berdan, & Bernatchez, 2019). Thus, parameters based on genome-scale data increase the sensitivity and reliability of genetic monitoring approaches.

Atlantic Salmon (*Salmo salar*), an ecologically, culturally and economically important species, whose range spans the North Atlantic Ocean, has declined in abundance range-wide over the last half century (Chaput, 2012; Lehnert et al., 2019b). Due to management and conservation concern, Atlantic Salmon have been the subject of extensive population genetic (e.g., Nielsen,

1998; Verspoor, 2005) and genomic studies (e.g., Aykanat et al., 2016; Jeffery et al., 2018). As such, the scale of available data and resources (e.g., chromosome level assembly; see Lien et al., 2016) make Atlantic Salmon a prime candidate to examine the utility of large genomic datasets for genetic monitoring.

Atlantic Salmon, an iteroparous species characterized by large-scale ocean migrations (Reddin, 1988) and fine-scale homing behaviour (Stabell, 1984), exhibits hierarchical population structure across its range with genetically, phenotypically, and adaptively differentiated populations (King et al., 2007). Genomic differentiation has been found to be greatest between continents with eastern (European) and western (North American) Atlantic Salmon having diverged more than 600,000 years before present (bp) (Nilsson et al., 2001; King et al., 2007; Rougemont & Bernatchez, 2018). Within Atlantic Canada, regions such as southern Newfoundland have been found to exhibit complex structuring. However, the contributions of neutral (i.e., genetic drift) and adaptive divergence have yet to be resolved (Bradbury et al., 2014).

Southern Newfoundland, in particular the Avalon Peninsula, is unique in that evidence suggests it is an area of trans-Atlantic secondary contact (Bradbury et al., 2015; E. Verspoor and D. Knox, in preparation). Recent research suggests that such trans-Atlantic secondary contact with European salmon near the end of the last glacial maximum approximately 18,000 bp resulted in karyotypic variation and polymorphism of genomic architecture (Brenna-Hansen et al., 2012; Rougemont & Bernatchez, 2018; Lehnert et al., 2019a). However, the extent that chromosomal structural variation associated with trans-Atlantic secondary contact has influenced population structure and adaptive variation in the region remains unknown.

Placentia Bay, a large (145 km wide at the mouth by 125 km long) and deep (240 m) embayment bounded by the Avalon Peninsula to the east and the Burin Peninsula to the west remains a relatively understudied region in south Newfoundland. Demographic estimates obtained using traditional monitoring approaches (i.e., counts of returning mature salmon) suggest declines in this region have ranged from 33 – 96% from the previous three generations (DFO, 2020). Declines have been attributed to a high rate of exploitation (Dempson et al. 2001), and hybridization with aquaculture escapees (Wringe et al., 2018).

The overall objective of this thesis is to generate a multi-year genomic dataset which will serve as a baseline against which future change will be assessed through continued genetic monitoring. The baseline is used here to demonstrate the utility of large genomic datasets to resolve the forces structuring populations regionally and monitor spatial and temporal change in population structure and abundance.

1.2 Thesis structure

Chapter two provides baseline estimates of population size of salmon populations in Placentia Bay, and additionally explores the utility of large genetic marker panels when used in combination with the single-sample linkage disequilibrium (LD) method (Hill, 1981) for genetic monitoring. Chapter three characterizes current spatial structure among Atlantic Salmon populations in Placentia Bay and explores the role of secondary contact in structuring. In each chapter, the implications for conservation and management are addressed both in terms of Atlantic Salmon in the region and more broadly.

Datasets with tens of loci are commonly used to study population demography, describe population structure, and inform conservation and management. Large datasets, with orders of

magnitude more loci, are now readily available for non-model species due to technological advances (Waples et al., 2016). These datasets, with genome-wide coverage, have been shown to possess more power to infer subtle spatial genetic structure and the potential to enable detection of less severe population declines (Wang, 2016; Allendorf, 2017).

Chapter two uses simulations to assess the utility of large panels (L = 100 microsatellites and 1000 or 10,000 SNPs) used in combination with the single-sample linkage disequilibrium (LD) method (Hill, 1981) for genetic monitoring. Simulations, performed with newly developed AGESTRUCN$_b$ (Antao et al., 2020), investigate how number of loci, marker type, initial population size, and sample size impact detection and approximation of declines in population abundance, as well as the precision and bias of estimates of the effective number of breeders ($\widehat{N}_b$). Simulations inform interpretation of accuracy and reliability of empirical $\widehat{N}_b$ generated to provide baseline estimates of population size for 26 rivers, sampled over three discrete years (2016 – 2018), in Placentia Bay, Newfoundland, Canada (Figure 1). The ability to detect trends in wild population abundance with large panels (L = 101 microsatellites and 1000 or 10,000 SNPs) was evaluated both within the previous generation, using the LD method in N$_e$ESTIMATOR v2.1 (Do et al., 2014), and over the recent past, using LINKN$_e$ (Hollenbeck, Portnoy, & Gold, 2016).

Chapter three uses the 220K SNP array to examine both neutral and adaptive population genetic structure of Atlantic Salmon in Placentia Bay, Newfoundland, Canada. Multivariate analysis, across two independent years of samples was used to resolve the scale of structuring. Redundancy analysis was used to identify environmental correlates and a generalized linear model was used to examine the temporal stability in the patterns observed.

Lastly, Chapter four discusses the implications of the results presented in Chapters two and three for the conservation and management of Atlantic Salmon in Placentia Bay, Newfoundland, Canada and highlight potential directions for future research.

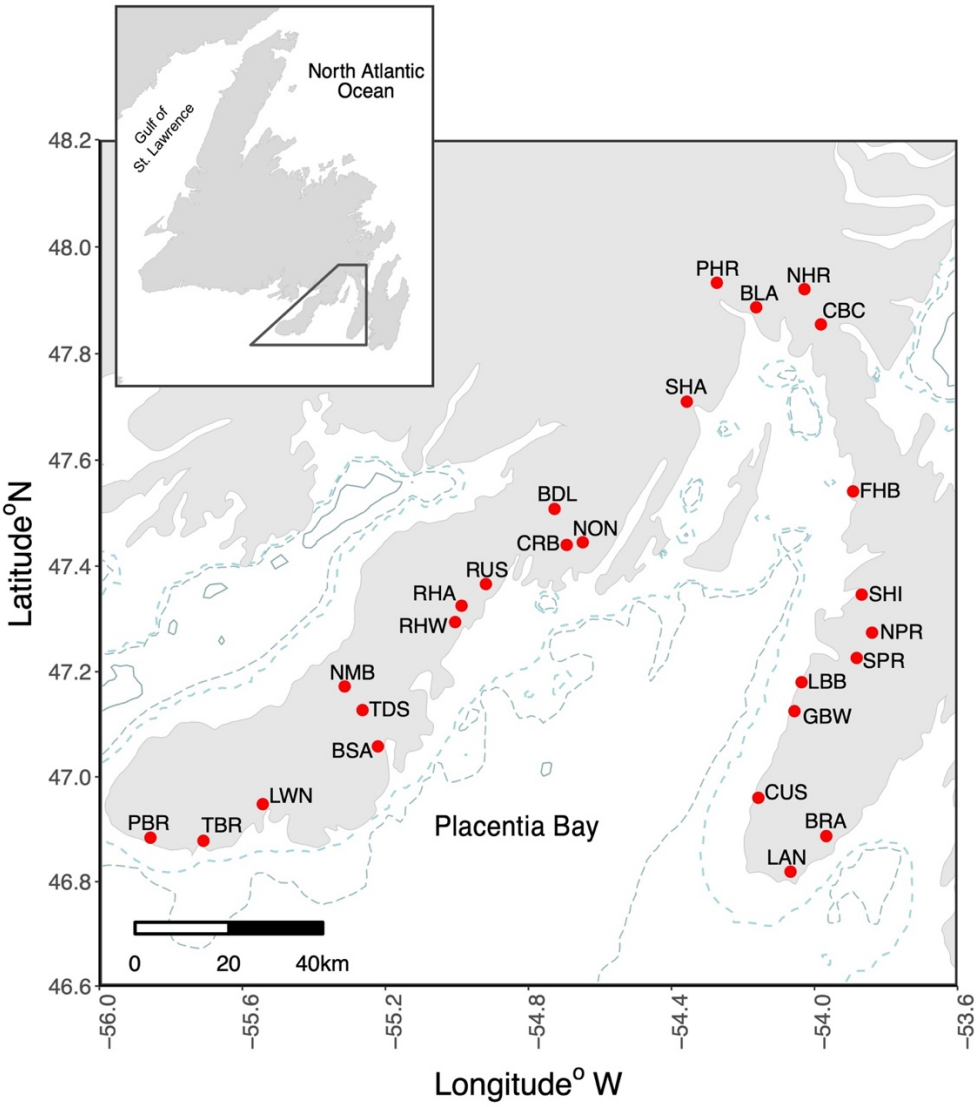

Figure 1 Sampling locations (n = 26, red circles) of juvenile Atlantic Salmon (*Salmo salar*) in Placentia Bay, Newfoundland, Canada. Samples collected in 2016, 2017 and 2018.

CHAPTER 2 – Genetic monitoring in the era of large genomic datasets: early and accurate detection of population decline using effective population size

2.1 Abstract

Early and accurate detection of population declines is essential to inform conservation and management of wild populations. Large genomic panels, containing thousands of genome-wide loci have the potential to improve accuracy and enable detection of population declines. However, gains in power and precision have yet to be quantified. Here I compared the performance of large and small panels of microsatellites and SNPs to detect population decline using repeated sampling and the single-sample linkage disequilibrium (LD) method of effective population size (N_e) estimation. I used simulated data, based on Atlantic Salmon (*Salmo salar*) life history, and varied both population size and per generation rate of decline. Simulation results were compared to empirical N_b estimates for Atlantic Salmon from 26 rivers in Newfoundland, Canada. My results suggest large panels of loci have greater power to both detect declines and resolve the magnitude of decline relative to small panels. Large panels detected minimum declines of 27% when population size was small (N = 100) and sample size was large (S = 50). In addition to panel size, both population size and sample size significantly influenced detection of decline. Power to detect declines decreased as population size increased and large panels were found to outperform large microsatellite panels when sample size was small (S = 30). My results suggest that large panels improve the utility of genetic monitoring as a conservation and management tool by enabling accurate detection of population decline.

2.2 Introduction

Early and accurate detection and quantification of population decline is crucial to fisheries and wildlife conservation and management. Population declines are often associated with a loss of genetic diversity which decreases evolutionary potential and increases risk of extinction (Shaffer, 1981; Frankham, 1995). Over the last few decades, many species have experienced declines in both number and size of populations (Ceballos, Ehrlich, & Dirzo, 2017) with more species anticipated to be affected and greater population declines expected in the coming decades due to continued climatic change and anthropogenic impacts (Hoffman, Sgrò, & Kristensen, 2017). Effective management requires an awareness of population demographic changes to ensure sustainable harvest (Ovenden et al., 2016), forecast future vulnerability (Fordham et al., 2014; Lehnert et al., 2019a), and inform conservation management actions (Luikart et al., 2010; Leroy et al., 2018). However, obtaining demographic data using traditional monitoring approaches is challenging, particularly for populations that are often not directly observable or accessible such as those in marine or freshwater environments, or that are cryptic, elusive, or remote (Luikart et al., 2010). As such, there is considerable interest in genetic monitoring for the detection and identification of population declines on a timescale relevant to population management and conservation (Schwartz et al., 2007; Palstra & Ruzzante, 2008; Hare et al., 2011; Frankham et al., 2014).

Population genetic parameters are commonly used to make inferences regarding population demography (e.g., Bernos & Fraser, 2016; Perrier, April, Cote, Bernatchez, & Dionne, 2016) and to inform conservation and management actions (Palstra & Fraser, 2012; Ferchaud et al., 2016) yet our understanding of how these metrics perform in scenarios of population decline

is limited. Simulations suggest population genetic parameters such as effective population size (N_e) could be used to reflect trends in abundance (Tallmon et al., 2010; Antao, Pérez-Figueroa, & Luikart, 2011; Wang, 2016). Empirically, a positive correlation between N_e and adult census size (N_c) has been detected in some studies (Bernos & Fraser, 2016; Ferchaud et al., 2016; Kuparinen, Hutchings, & Waples, 2016), although it has been lacking in others (Palstra, O'Connell, & Ruzzante, 2009; Serebzov, Jorde, Bernatchez, Olsen, & Vøllestad, 2012; Whiteley et al., 2015; Mueller, Chakarov, Krüger, & Hoffman, 2016). While the reason for disagreement among studies remains uncertain, previous studies have shown that the ability to detect a demographic decline with genetic data is likely strongly affected by population genetic parameter of interest and method of analysis, population size, sample size, and number of loci (Tallmon et al., 2010; Antao et al., 2011; Wang, 2016).

Large panels, with orders of magnitude more loci than commonly used panels, have the potential to improve accuracy of population genetic parameters, potentially enabling detection of less severe population declines (Waples et al., 2016; Allendorf, 2017). Although large panels with genome-wide coverage are becoming increasingly available for non-model species and are commonly used in the genetic monitoring of natural populations, their utility for inferring population size and trend has not been evaluated (Ellegren, 2014; Garner et al., 2015; Pierson, Luikart, & Schwartz, 2015). As such, there is a need to assess the power, precision, and bias of large panels in the context of estimation of contemporary effective population size.

Here, I evaluated the power and accuracy of large panels when used to estimate the effective number of breeders (N_b), the number of parents contributing to a single cohort, for detection and quantification of population declines. I focus on the most commonly used estimator (Jones et al., 2016; Wang, 2016), the single-sample linkage disequilibrium (LD) method (Hill,

1981; Waples, & Do, 2010). I used AGESTRUCN$_b$ (Antao et al., 2020), parameterized for Atlantic Salmon, a species that has experienced range wide population declines over the last half century (Chaput, 2012; Lehnert et al., 2019b), to determine how number of loci, marker type, initial population size, and sample size affect ability to: (i) detect population decline, and (ii) approximate the magnitude of decline. I compare simulation results using empirical data for three discrete cohorts of Atlantic Salmon (*Salmo salar*) for 26 rivers in Placentia Bay, Newfoundland, Canada. Estimates of N_b and N_e will serve as a baseline for genetic monitoring. Atlantic Salmon have experienced range-wide population declines over recent decades (Chaput, 2012). Traditional monitoring approaches such as angling surveys and counting fences are often used to inform conservation and management actions. In Newfoundland, these monitoring approaches are implemented in a small number of salmon rivers. Genetic monitoring provides an opportunity to track population abundance on a broad scale. I discuss both the simulation and empirical results in the context of practical applications of large genomic datasets in combination with the LD method of N_e estimation as a tool for conservation and management.

2.3 Methods

2.3.1 Simulations

I used simulated genetic data to evaluate the utility of a large genomic panel (≥ 1000 SNPs or 100 microsatellites) to detect and approximate population decline with the single-sample linkage disequilibrium (LD) method, the most commonly used estimator of contemporary effective population size (N_e). I created simulation scenarios, based on Atlantic Salmon life history, that varied in population size (number of mature anadromous individuals; N = 100, 500,

1000), number of loci (L = 25 or 100 microsatellites and 100, 1000 or 10,000 SNPs), sample size (S = 30, 50), and rate of decline (0 – 60% per generation) in a fully factorial design (Table S1; see Appendix E for total decline for each generation). Each scenario was replicated 10 times.

Generation of simulated data

I simulated the evolution of populations with controlled per generation rates of decline using the two-step procedure implemented in AGESTRUCN$_b$ (Antao et al., 2020) which integrates the individual-based forward-time simulation environment of SIMUPOP (Peng & Kimmel, 2005) with the estimation of effective number of breeders (N_b) using LDN$_e$ v2 (Do et al., 2014). First, simulations were initialized using population demographic parameters of wild Atlantic Salmon from southern Newfoundland, Canada (Table S2). Simulated datasets were generated for bi-allelic SNPs, initialized with a heterozygosity of 0.5, and microsatellites initialized with 10 alleles, frequencies drawn from a Dirichlet distribution, and a heterozygosity of 0.8. Following eight cycles of burn-in, populations were run for up to 20 generations (4 years per generation) forward in time, or until extirpated, without mutation for SNPs and with a mutation rate of 0.01 for microsatellite markers.

Estimation of effective population size

I then used LDN$_e$ v2 (Do et al., 2014) to estimate N_b and calculate upper and lower bounds of confidence intervals (CI), using the jackknife method (Jones et al., 2016), from a subsample (S = 30 or 50) of age 1 individuals for each year simulated. Due to computational limitations of the LDN$_e$ v2 (Do et al., 2014) as implemented in AGESTRUCN$_b$ (Antao et al., 2020), age 1 individuals

were extracted from the simulated GENEPOP (Rousset, 2008) files for the largest panels (L = 10,000 SNPs and N ≥ 500) using a custom R script prior to N_b estimation. A minimum allele frequency (MAF) threshold of 0.05 was used. A true N_b variance allowed, which determines the allowable threshold of N_b, of 5.0 was chosen to mimic a real-world scenario in which no limiting constraint can be placed on the estimate of N_b generated. Data collection began at generation t_0 (pre-decline) and continued for x generations up to a maximum of 20 generations or until extirpation as determined by the per generation rate of decline (Figure 2 and Figure S1). Populations became functionally extinct prior to 20 generations (80 years) when per generation rate of decline was high. For example, a population with a 5% per generation rate of decline was monitored for 20 generations, at which point total decline was approximately 65%, while a population with a 30% per generation rate of decline was monitored for 12 generations, at which point total decline was approximately 98% and the population was considered functionally extinct.

Rate of false positives and false negatives

I investigated the rate of false positives (type I error) and false negatives (type II error; failure to detect a decline) across panels of varying size using stable (0% decline) populations. Infinite $\widehat{N}_b$ and CI estimates were set to 10,000. Frequency of false positive and negatives was assessed using the distribution of total change in $\widehat{N}_b$; positive values indicated an increase, representative of false negatives, and negative values indicated a decrease, indicating a false positive.

Detection of declines

I evaluated the performance of each panel, varying by marker type and number of loci, from two perspectives: power, and accuracy. I first assessed the likelihood of detecting a decline for each simulated scenario. A decline was detected if $\widehat{N}_b$ (naïve point estimate) at t_x, sampled the last year of each generation, was below $\widehat{N}_b$ at t_0. Power, the probability of detection was regressed against the modeled decline for each scenario. Mean detectable decline per scenario was calculated as the average decline detected with 80% probability (Antao et al., 2011).

Mean $\widehat{N}_b$ was calculated across replicates for each scenario to assess precision. Infinite $\widehat{N}_b$ and CI estimates were set to 10,000. Visual inspection of the range of $\widehat{N}_b$ was used to assess bias. Panels were considered unbiased if the range of $\widehat{N}_b$ was small, and precise if mean $\widehat{N}_b$ fell within the expected range of true N_b; due to the parameterization of Atlantic Salmon life history with precocial male maturation, true N_b ranges from the number of anadromous adults to the total number of mature individuals.

Approximation of declines

I next examined the ability of panels of varying size to accurately recover the modeled decline. Offset, the difference in the modeled and observed decline in $\widehat{N}_b$ between t_0 and t_x, was calculated once per generation for each scenario. For example, a simulated population with an N_0 = 100 at t_0 and a 20% decline per generation would have an N_b = 21 at t_7 (79% decline); if $\widehat{N}_b$ at t_7 was 79% lower than $\widehat{N}_b$ at t_0 then offset = 0 indicating the magnitude of decline was accurately estimated. Infinite $\widehat{N}_b$ were excluded from analysis. Declines were rounded to the nearest 10 (i.e., the 79% decline at t_7, 20% decline per generation, was rounded up to 80% for data visualization).

2.3.2 Empirical data

Genotyping and selection of neutral loci

Juvenile Atlantic Salmon, aged 0 to 2+, were sampled by electrofishing during the period of July to September in 2016 – 2018 from 26 rivers in Placentia Bay, Newfoundland, Canada (Figure 1 and Table 1). Age of individuals was inferred based on length-age relationships (Sylvester et al., 2019). Fin clips were collected and stored in 95% ethanol. Samples collected in 2016 (n = 1470) were genotyped using a panel of 101 microsatellite loci as previously described in Bradbury et al. (2018) (see Appendix F for microsatellite primers). Briefly, microsatellite loci were identified using the ICSASG v2 genome for Atlantic Salmon (Lien et al., 2016). Locus amplification and scoring of genotypes using MEGASAT followed the protocol previously described in Zhan et al. (2017). After filtering individuals for missing genotypes (missingness < 0.05), 1433 remained; of these, a maximum of 30 individuals per river were retained for downstream analysis. Summary statistics, including number of alleles, observed (H_o) and expected heterozygosity (H_e), and global and pairwise population F_{ST} (Weir & Cockerham, 1984) were calculated using GENALEX v 6.5 (Peakall & Smouse, 2006, 2012) (see Appendix G). Samples collected in 2017 (n = 745) and 2018 (n = 684) were genotyped using a 220K bi-allelic SNP Affymetrix Axiom array developed for Atlantic Salmon by the Centre for Integrative Genetics (CIGENE, Ås, Norway) as previously described in Sylvester et al. (2018a). Briefly, these SNPs were a subset of those in the 930K XHD Ssal array (dbSNP accession numbers ss1867919552–ss1868858426) (Barson et al., 2015). Sample preparation and genotyping followed the protocol outlined in Sylvester et al. (2018a). Genotyped data were filtered for a

minor allele frequency (MAF) cut-off of 0.01 across all rivers and years and a missingness threshold of < 0.05 using PLINK v 1.9 (Chang et al., 2015). This resulted in a total of 138,853 SNPs and 677 individuals in 2017 and 684 individuals in 2018.

The LD method of effective population size estimation assumes loci are neutral and unlinked. Following convention, microsatellite loci were assumed to be selectively neutral (Holleley et al., 2014) and unlinked as only one locus per genomic scaffold was selected (Bradbury et al., 2018). A panel of neutral and unlinked SNPs was generated by first removing chromosomes with known structural variants (Ssa01, Ssa08, Ssa23, and Ssa29) (Brenna-Hansen et al., 2012; Lehnert et al., 2019a) and then removing global F_{ST} outlier loci (> 95th percentile) using PLINK v 1.9 (Chang et al., 2015). This putatively neutral dataset was further thinned by count using PLINK v 1.9 (Chang et al., 2015); variants were removed at random, to generate panels of 100, 1000 and 10,000 SNPs.

Estimation of effective population size and detecting population declines

To investigate the applicability of the simulated results to real world populations, N_b was estimated for the 26 rivers in Placentia Bay, Newfoundland, Canada for which juvenile Atlantic Salmon were sampled over three discrete years. To quantify trend in population size within the current generation, I estimated N_b for each year sampled (2016 – 2018) using the LD method in N_eESTIMATOR v2.1 (Do et al., 2014). Rivers with a sample size < 20 were excluded from analysis. Precision was assessed using independent replicates (n = 5) of a sub-set of populations (North Harbour River; NHR, Nonsuch River; NON, and Big Salmonier Brook; BSA) selected based on $\widehat{N}_b$ to represent the range of simulated population sizes. Replicate estimates of N_b were generated

by subsampling the putatively neutral dataset for 2017 using PLINK v 1.9 (Chang et al., 2015) to generate independent replicate panels of varying size (L = 100, 1000 and 10,000 SNPs).

LINKNe (Hollenbeck et al., 2016), which bins pairs of loci with similar recombination rates to reconstruct N_e over time was used to detect trend in $\widehat{N}_e$ over the recent past to characterize expected trend in population size using the panel of 1238 SNPs and following the protocol outlined in Lehnert et al. (2019b). Estimates of N_e were generated for a range of time points over a 100-year period in the past using samples collected in 2017. Briefly, default parameters for bin size (5 cM), MAF (< 0.05), and sample-size bias correction were used. Since the time period to which an LD-based estimate of N_e applies is a function of recombination rate (c), mean recombination rate (c) per bin was used to estimate the number of generations (t) in the past (t = 1/2c) to which each $\widehat{N}_e$ applied. Recombination rates for pairs of loci were calculated using the average length of male (2153 cM) and female (968 cM) linkage maps for North America Atlantic Salmon (Brenna-Hansen et al., 2012). A generation time of 4 years was used, consistent with Atlantic Salmon populations in south Newfoundland (COSEWIC, 2010). A significant decline was detected if CIs for the period 1989 – 2013 (~ 6 generations) were non-overlapping.

2.4 Results

2.4.1 Simulations

Genetic diversity (mean expected heterozygosity; H_e) and gametic disequilibrium (r^2) were estimated across time in simulations by AGESTRUCNb (Cosart & Hand, 2017). Initial population size affected the initial level of genetic diversity; mean H_e was greatest at large initial

population sizes whereas, r^2 was greatest when initial population size was small and increased as population size declined. Following theoretical expectation, small populations (N = 100) lost genetic variation more rapidly than large populations (N ≥ 500) regardless of whether they were stable or in decline, consistent with theoretical expectation. Mean H_e was unchanged by sample or panel size across time, whereas, r^2 was unchanged by panel size or marker type, but decreased with increased sample size. In simulations with large stable (0% decline) populations, mean H_e increased over time because alleles were not lost via genetic drift as in small populations.

Rate of false positives and false negatives

For the stable population scenario (0% decline), the distribution of false positives and false negatives varied considerably with panel size and initial population size (Figure S2). Large panels more frequently inferred a decline in $\widehat{N}_b$ (false positives) than small panels whereas, small panels were more likely than large panels to infer a large increase in $\widehat{N}_b$ (false negatives). Notably, the distribution of total change was narrower (closer to expected 0% change) when N_b was estimated with large panels relative to small panels. With larger initial population size, the rate of false positives increased when using large panels (L = 100 microsatellites or ≥ 1000 SNPs). In contrast, a relatively uniform distribution was observed at moderate to large population sizes when using small panels (L = 25 microsatellites or 100 SNPs).

Detection of declines

Large panels had greater power to detect smaller declines in $\widehat{N}_b$ than small panels (Figure 3 and Figure S3). Large SNP panels had greater power to detect declines relative to the large

microsatellite panel when sample size was small (S = 30) (Figure 3). With a large SNP panel and a small initial population size, declines ≥ 30% were detected more than 80% of the time with a small sample size. The difference in mean detectable decline between marker types with large panels was not apparent when sample size was large (S = 50), as a decline of similar magnitude (≥ 30%) was detected with the large microsatellite panel only when sample size was large (Figure S3). Mean detectable decline decreased with increased initial population size (Table S3). Large panels detected mean declines of approximately 27%, 35%, and 46% in small (N = 100), moderate (N = 500), and large (N = 1000) populations with a large sample size, respectively. While per generation rate of decline appeared to have a slight effect on the severity of decline detected, whereby smaller total declines were detected with 80% probability when per generation rate of decline was low (< 20%) relative to high (>30%), this finding was not explicitly tested. Overall, power to detect a decline was largely dependent on initial population size, and panel size, with effect of sample size on power to detect a decline appearing to be correlated with both marker type and panel size.

Bias and precision of $\widehat{N}_b$

Large panels accurately (+/- 10% of modeled N_b) estimated N_b more frequently than small panels (Table 2). When sample size was large, and initial population size was small, $\widehat{N}_b$ was accurate approximately 23% of the time when using a large panel compared to 15% of the time when using a small panel. The interquartile range of $\widehat{N}_b$ was consistently narrower with large panels relative to small panels indicating replicate estimates were more similar and large panels more precise than small panels regardless of sample size or initial population size (Figure 4).

Notably, few outliers fell outside the expected range of true N_b when population size was small and panel size was large suggesting that when population size was small, large panels exhibited little bias. Interestingly, no negative or infinite $\widehat{N}_b$ were observed when initial population size was moderate (N = 500) and the interquartile range remained within the expected range of true N_b when population size was large. This finding suggests the 10,000 SNP panel remained relatively unbiased as population size increased.

Accuracy of estimated declines

Large panels approximated the magnitude of decline with greater accuracy than small panels (Figure 5 and Figure S4). Accuracy of estimated decline increased with increased sample size (Figure 5), decreased as population size increased (Figure S4), and increased with severity of decline. Distribution of offset, the difference between the modeled and observed decline in $\widehat{N}_b$, was greatest in the no decline scenarios and decreased as decline increased indicating the likelihood of accurately inferring the decline increased as magnitude of decline increased.

Median offset was generally found to be positive but small when using large panels indicating the decline estimated from genetic data was greater than the modeled decline (Table S4). Median offset inferred using small panels was found to be negative (i.e., declines were underestimated) regardless of sample or initial population size. Large SNP panels approximated declines with greater accuracy than the large microsatellite panel when sample size was small; however, when sample size was large, large panels were comparable regardless of marker type (Table S4). Notably, range of offset was consistently narrower with the 10,000 SNP panel relative to both the 100 microsatellite and 1000 SNP panels indicating the probability of

accurately inferring the decline was highest with the 10,000 SNP panel (Table S4). These results suggest that large panels provide the highest accuracy. Small panels are likely to underestimate the magnitude of decline.

2.4.2 Empirical data

Estimation of N_b and detection of decline within a generation

Using an empirical Atlantic Salmon dataset spanning three discrete years I found large panels had greater precision than small panels. Within year, $\widehat{N}_b$ was found to be relatively consistent across panel size; however, both $\widehat{N}_b$ and the upper bound of CIs were less often infinite when using large panels (Figure 6). The upper bound of CIs was found to be infinite three times more often when using a small rather than large panel. When comparing independent replicates within river, I found large panels were surprisingly precise; $\widehat{N}_b$ was consistent across all replicates using the 10,000 SNP panel despite populations ranging from small ($\widehat{N}_b$ ~ 75) to moderately small in size ($\widehat{N}_b$ ~ 230) (Figure S5). Estimated $\widehat{N}_b$ was generally greater and CIs narrower when sample size was large (S = 50 – 60) relative to small (S = 20 – 30) when using a large microsatellite panel (Figure S6). No significant change in $\widehat{N}_b$ was observed within generation when using the criteria of non-overlapping CIs (Figure S7). However, I note that the wide CIs observed with small panels are more likely to impede ability to detect decline.

Estimation of N_e and detection of decline over the recent past

I also used LINKN$_e$ (Hollenbeck et al., 2016), which reconstructs N_e over the recent past, to set expectations and contextualize observed trends in $\widehat{N}_b$within the current generation. Using a

large panel (L = 1238 SNPs; see Lehnert et al., 2019b), I observed a similar pattern of decline across all rivers, with starting $\widehat{N}_e$ dependent on river (Figure S8). Over approximately 6 generations (1989 – 2013; 4 years/generation), $\widehat{N}_e$ significantly declined in all but one of the rivers sampled. While the exception (Piper's Hole River; PHR) was found to decline, the decline was not significant using the criteria of non-overlapping CIs (Figure S9).

2.5 Discussion

Populations can change rapidly in size, particularly in response to climatic or anthropogenic threats (Waples, 2002). Obtaining demographic data with traditional monitoring approaches may be challenging for some species, and detection of decline may be difficult due to limitations imposed by survey methods (Luikart et al., 1998, 2010). As such, there is considerable interest in genetic monitoring, using population genetic parameters such as effective population size (N_e) for the conservation and management of wild populations (Frankham, 2010; Hare et al., 2011; Ovenden et al., 2016). Previous work has shown the linkage disequilibrium (LD) method of N_e estimation can detect population bottlenecks using datasets of a few dozen loci (England et al., 2010; Antao et al., 2011; Wang, 2016). Large genomic datasets, now readily available, have the potential to enable detection of less severe decline and to accurately quantify declines in population size. Here, I examined the utility of large genomic panels for genetic monitoring and addressed implications for precision and bias within a management framework. Using the single-sample LD method to estimate effective number of breeders (N_b), I found that large panels have greater power and more accuracy than panels commonly used today. Here, I

show that large genomic panels improve sensitivity and reliability of genetic monitoring approaches.

2.5.1 Detection and accuracy of population decline

I found r^2 (mean gametic disequilibrium) was unchanged by panel size, and my results support the finding of Waples et al. (2016) that while large panels, with orders of magnitude more loci, increase the number of pairwise comparisons between loci, they do not increase the magnitude of linkage. As the number of loci is increased, the fraction of pairs of loci that are linked is expected to converge on the true fraction, which increases statistical power (Waples et al., 2016). However, not all pairwise comparisons among loci are independent. Confidence intervals estimated with the LD method are expected to be unrealistically precise; a problem that has been shown to be exacerbated as the number of loci is increased (Waples et al., 2016). Jones et al. (2016) found CIs estimated with the LD method and as few as 200 SNPs may only contain true N_e 80% of the time. Here, I found that CIs narrowed with increased panel size. Additionally, the range of $\widehat{N}_b$ across replicates narrowed, with empirical $\widehat{N}_b$ surprisingly consistent across replicates using large SNP panels.

My results suggested that large panels consistently outperformed small panels by being able to detect less severe population declines. Declines detectable by large panels were approximately half as severe as those detectable with small panels comparable to those used in most current empirical studies. Although severity of decline detected was dependent on panel size, large declines were inferred with less bias and greater precision than small declines in all scenarios. Here, I found small panels (L = 25 microsatellites or 100 SNPs) detected minimum

declines of approximately 33 - 40% in small populations (N = 100) and 58 - 65% in moderately large populations (N = 500) with a sample size of 50. Using a small microsatellite panel (L = 20) and the LD method, Wang (2016) similarly detected declines of 40% in small populations (N = 50) with a sample size of 100, and Antao et al. (2011) declines of roughly 67% in moderately large populations (N = 600) with a sample size of 50. In contrast, large panels (L = 100 microsatellites or ≥ 1000 SNPs) simulated here, detected minimum declines of 25 - 29% in small populations (N = 100) and 33 - 40% in moderately large populations (N = 500) with a sample size of 50. Interestingly, when rate of decline was slow (low per generation rate of decline) I observed greater power to detect less severe declines (i.e., the minimum decline detected with 80% probability is lower when per generation rate of decline was 5% relative to 40%).

Large panels also outperformed small panels by inferring the magnitude of decline with greater accuracy. Large panels enabled reasonably accurate quantification of decline likely due to improved accuracy of N_b point estimates. Although false positives (type I error) were observed in all no-decline scenarios regardless of panel size, inferred magnitude of decline was often smaller with large panels relative to small panels. Inference of large population increase in no-decline scenarios were more common with small panels than with large panels. Accurate identification of stable and declining populations has been suggested to be dependent on the number of generations between samples (Tallmon et al., 2010). Here, I find likelihood of detecting a decline using the LD method dependent on panel, sample, and population size only. My results support previous work suggesting declines can be detected within a single generation using the LD method (e.g., England et al., 2010; Wang, 2016). Overall, large panels enabled detection of less severe declines with greater accuracy than small panels. I therefore expect less noise when

monitoring a population over time with a large panel, making trends in abundance easier to detect.

2.5.2 Population and sample size

Population and sample size affected detection and quantification of decline; power and precision increased with larger sample size and decreased with larger initial population sizes, a finding previously noted by Waples & Do (2010). Gains in power with larger sample size were largely dependent on the number of independent pairwise comparisons between loci; larger gains in power with increased sample size were observed when the number of independent comparisons was small and decreased as the number of independent comparisons increased. Due to the smaller number of independent pairwise comparisons, larger sample size improved power of the large microsatellite panel more than the large SNP panels. A similar increase in power was reported by Antao et al. (2011) who noted an increase in sample size was more beneficial than an increase in panel size when using small microsatellite panels. An increase in sample size from 30 to 50 did little to improve power of large SNP panels; mean detectable decline was relatively unchanged. Interestingly, range of offset decreased when sample size was increased from 30 to 50 while using a large SNP panel indicating that although power was not affected, precision increased with increased sample size. While an increase in sample size may improve power and precision more than a proportional increase in the number of loci, I note that technological advances have made increasing panel size using SNPs ten-fold easier than increasing sample size (England et al., 2010; Luikart et al., 2010; Antao et al., 2011). This result has implications for

genetic monitoring, particularly for species of conservation concern for which sample size is often small.

I found $\widehat{N}_b$ was more frequently within the expected range of true N_b, when using large panels relative to small panels indicating large panels estimate N_b with less bias (i.e., more closely approximate true N_b) than small panels. Although I expected bias to be lower with SNPs relative to microsatellites due to rare alleles effects as found by Antao et al. (2011), I found little difference between marker type when panel size was large.

Using genetic monitoring to infer trend in abundance in large populations is challenging because the signal of drift is expected to diminish as N_e increases (Waples & Do, 2010). As expected, I found relatively low accuracy in in populations with large initial size (N = 1000) using a panel of 25 microsatellites and a sample size of 50. While large panels increased power to detect decline in moderately large populations, frequency of false positives increased as population size increased, a finding also reported by Antao et al. (2011). I note that although it has been suggested that the LD method estimates N_e with reasonable accuracy in very large populations (N_e ~ 30,000), using simulations with a panel of 20 microsatellites and a sample size of 100 (Wang, 2016) in practice, the LD method has been found to lack power and precision unless a sufficient number of individuals (~1%) are sampled (Marandel et al., 2019).

2.5.3 Limitations and biases

It is worth noting here that I estimate N_b rather than N_e. It is often easier to estimate N_b than N_e, and while the two quantities are related, they differ in that N_b is important for understanding eco-evolutionary dynamics and mating systems, whereas N_e influences long-term

evolutionary process (Perrier et al. 2014; Waples et al., 2014; Ferchaud et al., 2016). While much remains to be understood about how N_b relates to N_e and N, N_b has been shown to approximate N_e when sampling a single cohort with overlapping generations and declines in N_e have been shown to track declines in N (Waples et al., 2013; Ferchaud et al., 2016; Kuparinen et al., 2016; Ovenden et al., 2016). Genetic monitoring should therefore target sampling designs of consecutive cohorts to estimate N_b using single-sample approaches, with consecutive cohorts being analyzed jointly to estimate N_e to provide a more comprehensive overview of the system over time.

I also average over replicates and acknowledge that accuracy may be lower in practice. However, I provide distribution of $\widehat{N}_b$ to better inform range of estimates possible under each scenario. Detection of decline can be defined many ways, here a decline was detected if a decrease in $\widehat{N}_b$ was observed between t_0 and t_x and a significant decline was detected when confidence intervals between $\widehat{N}_b$ at t_0 and t_x were non-overlapping as in Lehnert et al. (2019b). Lastly, I acknowledge that the conversion of negative and infinite $\widehat{N}_b$ to 10,000 may reduce accuracy in scenarios with small panel or sample size and large population size for which infinite $\widehat{N}_b$ were common. However, conversion of negative and infinite estimates to a finite value of 10,000 is commonly used and as such was the approach taken here unless otherwise stated.

2.5.4 Empirical observations

My results support the finding that large panels are more precise and that larger sample size increases accuracy of large microsatellite panels. Demographic estimates obtained using traditional monitoring approaches (i.e., counts of returning mature salmon) suggest regional

declines in abundance of mature salmon have ranged from 33 – 96% from the previous three generations (DFO, 2020). However, demographic estimates fail to account for precocial (mature) male parr, which have been found to significantly increase $\widehat{N}_e$ and $\widehat{N}_b$ (Johnstone, O'Connell, Palstra & Ruzzante, 2013; Perrier, Normandeau, Dionne, Richard & Bernatchez, 2014). Southern Newfoundland has a high incidence of precocial male maturation (Dalley, Andrews & Green, 1983). As such, trend in demographic and genetic estimates of population abundance may be de-coupled. While no significant change in $\widehat{N}_b$ was observed within a generation, using LINKNe, which detects trend in $\widehat{N}_e$ over the recent past, I found that all rivers sampled had declined in population size over the last six generations. This result provides a baseline expectation for regional trend in $\widehat{N}_e$ and further emphasizes the importance of quantifying decline on a management timeline. Further research is needed to determine the sampling frequency and temporal period of sampling required to detect trend in $\widehat{N}_b$.

2.5.5 Practical considerations

Some useful guidelines on the use of large panels to infer population abundance and trend emerge from my simulations. Under certain sampling and demographic conditions $\widehat{N}_b$ can provide useful insights into population size and trends in abundance. In general, N_b estimated using large panels outperformed small panels by enabling earlier and more accurate detection of less severe population declines. However, limitations to the use of genetic monitoring with N_b remain despite the increase in power and precision with large panels. My results indicate that detection and quantification of decline with reasonable accuracy can be achieved with a large SNP panel and a sample size of 30 when population size is small, suggesting large SNP panels

may be better than microsatellites for species of conservation concern or for which it's difficult to obtain a large sample of a single cohort. Large panels have greater power at moderate to large population sizes than small panels; however, accuracy decreases as population size increases and frequency of false positives increases as population size increases. This finding has implications for the feasibility of genetic monitoring studies; although false positives are generally a greater concern than false negatives, conservation resources are limited and as such future work should explore the relationship between power and precision working to maintain power while improving precision.

Large panels also increased the likelihood of successfully identifying trend in population size and were better able to quantify magnitude of decline with reasonable accuracy than small, commonly used panels. The COSEWIC assessment process in Canada is based on decline in total number of mature adults, as such, accurate quantification of decline is essential. My results suggest large panels accurately recover population declines as small as 30% within a single generation using the LD method to estimate N_b; the larger the decline the greater the accuracy. Further research is needed to characterize how natural fluctuations in population size influence the ability to detect decline using $\widehat{N}_b$ and large panels and to characterize how demographic history may influence precision and bias, since heterozygosity, mutation rate, and population size are all strongly influenced by past demographic events such as bottlenecks, and the LD method has been shown to be slow to track recovery (Hare et al., 2011).

Several important considerations should be addressed when designing a real-world study informed by these simulations. First, I used sample sizes typical of population genetic studies (S = 30 – 50); however, Marandel et al. (2019) found the most common sample size for fisheries

stock assessment was 100, as such, I note that some conclusions, particularly those regarding large population sizes, might not hold if sample size were increased to 100 for both marker types. Second, I used a *p*crit (minor allele frequency cut-off) of 0.05 for all scenarios regardless of sample or population size. The importance of selecting this parameter and its effect on precision has been noted elsewhere (Waples & Do, 2010; Jones et al., 2016; Marandel et al., 2019). Future work should examine how this parameter is affected by the use of large panels and potential differences in $\widehat{N}_b$ when using polymorphic microsatellite or bi-allelic SNP panels. Additionally, since microsatellites are multi-allelic whereas SNPs are bi-allelic, they provide different types of information. I expect microsatellites would have greater power if used to calculate variance N_e (temporal method), which utilizes changes in allele frequency over time. Third, although a wide range of per generation rates of decline were modelled, in real populations abundance is likely to fluctuate stochastically. Future work should examine the frequency of sampling needed to detect trend in population size in a stochastic population.

2.5.6 Conclusions

Over the last few decades, many species have experienced declines in both number and size of populations (Ceballos, Ehrlich, & Dirzo, 2017), with more species anticipated to be affected and greater population declines expected in the coming decades due to continued climatic change and anthropogenic impacts (Hoffman, Sgrò, & Kristensen, 2017). My results suggest using large genomic panels enables more accurate detection of less severe population declines than panels of tens of microsatellites commonly used. Traditionally, genetic monitoring has aimed to discern small from large populations (Waples & Do, 2010), and detect decline

(Tallmon et al., 2010; Antao et al., 2011; Wang, 2016). Large panels with thousands of SNPs, now commonly available, have greater power to detect less severe declines and more accurately quantify magnitude of decline making genetic monitoring using N_b an increasingly informative tool for management and conservation. Detection of less severe declines enables rapid management action and as such avoids irreversible loss of genetic variation, reducing the risk of extirpation or extinction. If sample size is sufficiently high, large panels may enable genetic monitoring of moderate to large populations with reasonable accuracy; early warning would be useful in planning conservation actions and assessing genetic risk by allowing less costly management actions (i.e., quotas and fisheries closures) to be taken and minimizing the need for costly conservation action (i.e., transplantation). The increasing ease with which large panels can be genotyped makes genetic monitoring using N_b more feasible enabling a better understanding of the causes, consequences and severity of population declines.

Table 1 Sampling location, in decimal degrees, of Atlantic Salmon (*Salmo salar*) in Placentia Bay, Newfoundland. Sample sites organized east to west. Axial river length (km) and number of samples (n) genotyped per year (2016 – 2018).

River Name	River ID	Longitude	Latitude	Axial River Length (km)	n		
					2016	2017	2018
Branch River	BRA	-53.97	46.89	22.5	60	30	32
Lance River	LAN	-54.07	46.82	12.9	31	9	-
Cuslett Brook	CUS	-54.16	46.96	8.0	61	30	30
Great Barasway Brook	GBW	-54.06	47.12	14.3	60	30	41
Little Barasway Brook	LBB	-54.04	47.18	11.4	10	16	-
Southeast Placentia River	SPR	-53.88	47.23	21.7	60	30	49
Northeast Placentia River	NPR	-53.84	47.27	23.5	60	30	31
Ship Harbour Brook	SHI	-53.87	47.35	13.2	60	30	32
Fair Haven Brook	FHB	-53.89	47.54	1.6	60	30	30
Come by Chance River	CBC	-53.98	47.86	17.2	61	30	31
North Harbour River	NHR	-54.03	47.92	14.5	60	30	32
Black River	BLA	-54.16	47.89	22.4	60	30	30
Piper's Hole	PHR	-54.27	47.93	40.2	61	30	33
Sandy Harbour River	SHA	-54.36	47.71	43.4	61	36	9
Nonsuch River	NON	-54.65	47.44	11.8	60	30	36
Cape Roger Brook	CRB	-54.69	47.44	20.3	60	30	32
Bay de l'Eau	BDL	-54.73	47.51	27.4	61	30	32
Rushoon River	RUS	-54.92	47.37	14.8	60	30	37
Red Harbour River East	RHA	-54.99	47.32	8.7	60	30	34
Red Harbour River West	RHW	-55.01	47.29	6.9	61	30	30
Northwest Brook	NWB	-55.31	47.17	27.7	60	30	32
Tides Brook	TDS	-55.26	47.13	20.3	60	30	28
Big Salmonier Brook	BSA	-55.22	47.06	12.2	61	30	31
Lawn River	LWN	-55.54	46.95	11.7	60	30	30
Taylor Bay Brook	TBR	-55.71	46.88	15.4	43	30	32
Piercey's Brook	PBR	-55.86	46.88	15.8	60	30	28

Table 2 Accuracy of effective number of breeders ($\widehat{N}_b$) across marker type (microsatellites and SNPs), panel size (L = 25 or 100 microsatellites, or 100, 1000 or 10,000 SNPs), population size (N = 100, 500 and 1000), and sample size (S = 30 or 50). Estimate of N_b was accurate if +/- 10% of simulated N_b; overestimated if >10% lower and underestimated if >10% higher. Proportion of accurate, overestimated, and underestimated $\widehat{N}_b$ shown per scenario.

Marker Type	Panel Size	Initial Population Size	Sample Size	> 10% lower	< 10%	> 10% higher
Microsatellites	25	100	30	43.4	12.0	49.7
			50	41.5	13.1	45.4
		500	30	24.8	6.0	69.8
			50	36.6	10.2	55.0
		1000	30	24.8	5.6	70.2
			50	24.3	5.9	72.5
	100	100	30	38.6	15.8	50.6
			50	45.9	21.0	33.1
		500	30	48.9	11.5	39.5
			50	48.9	12.0	39.1
		1000	30	30.3	7.7	61.9
			50	53.4	9.1	38.3
SNPs	100	100	30	48.2	8.7	47.5
			50	45.9	17.5	36.6
		500	30	27.8	8.4	64.6
			50	37.0	7.4	56.2
		1000	30	24.4	4.5	72.3
			50	31.4	7.1	61.5
	1000	100	30	36.8	20.8	47.7
			50	38.5	24.6	36.9
		500	30	49.7	11.5	38.9
			50	48.0	13.4	39.8
		1000	30	40.3	11.3	48.4
			50	45.9	13.6	40.5
	10,000	100	30	44.5	15.5	45.5
			50	36.8	22.1	41.1
		500	30	44.1	11.9	44.0
			50	53.5	12.4	34.1
		1000	30	53.0	7.5	39.6
			50	50.4	10.1	39.5

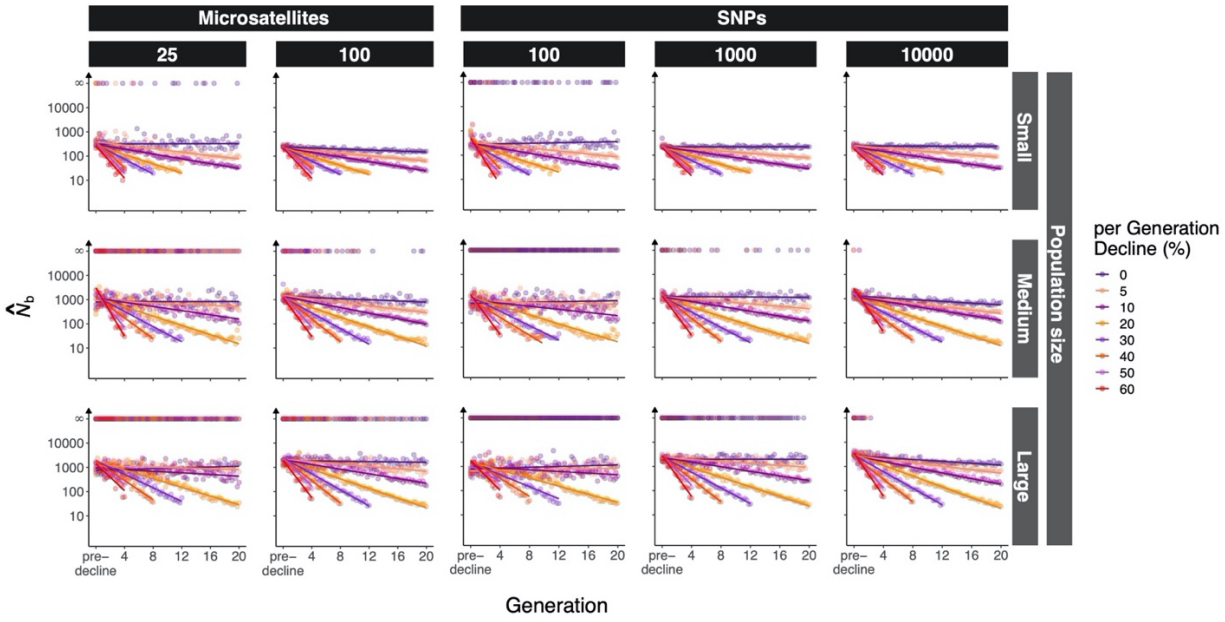

Figure 2 Linear regression of mean $\widehat{N}_b$ per scenario shown with infinite estimates, coloured by per generation rate of decline (%). $\widehat{N}_b$ generated using LDNe v2 as implemented in AGESTRUCNb (Antao et al., 2020), S = 30. Mean $\widehat{N}_b$ calculated across ten replicates for each rate of decline and scenario. Two marker types (microsatellites and SNPs) and five panel sizes (columns: L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) were modelled using three initial population sizes (N = small (100), medium (500), or large (1000)). Eight per generation rates of decline (%) were simulated per scenario, indicated by colour. Monitoring began in the year pre-decline and continued up to 20 generations. Raw data shows both variance of N_b point estimates and the number of infinite estimates decreased as panel size increased.

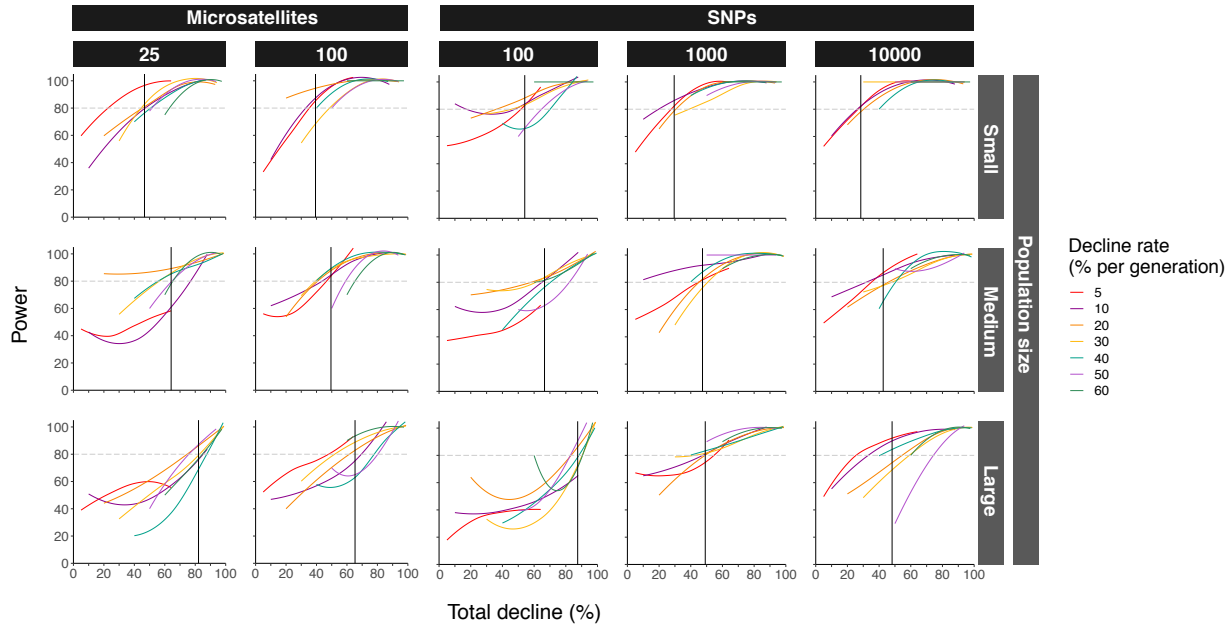

Figure 3 Power to detect a decline in $\widehat{N}_b$ between t_0 and t_x with S = 30 for all simulated scenarios. Point estimates generated using LDNe v2 as implemented in AGESTRUCNb (Antao et al., 2020). Two marker types (microsatellites and SNPs) and 4 panel sizes (columns: L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) compared across three population sizes (rows: N = small, medium and large). Coloured by per generation rates of decline (%). Loess curves are used for visualization of trends in the data. The vertical solid black line is the mean total decline (%) detectable for each scenario. The horizontal grey dashed line is the 80% power threshold. Mean detectable decline decreased as panel size increased.

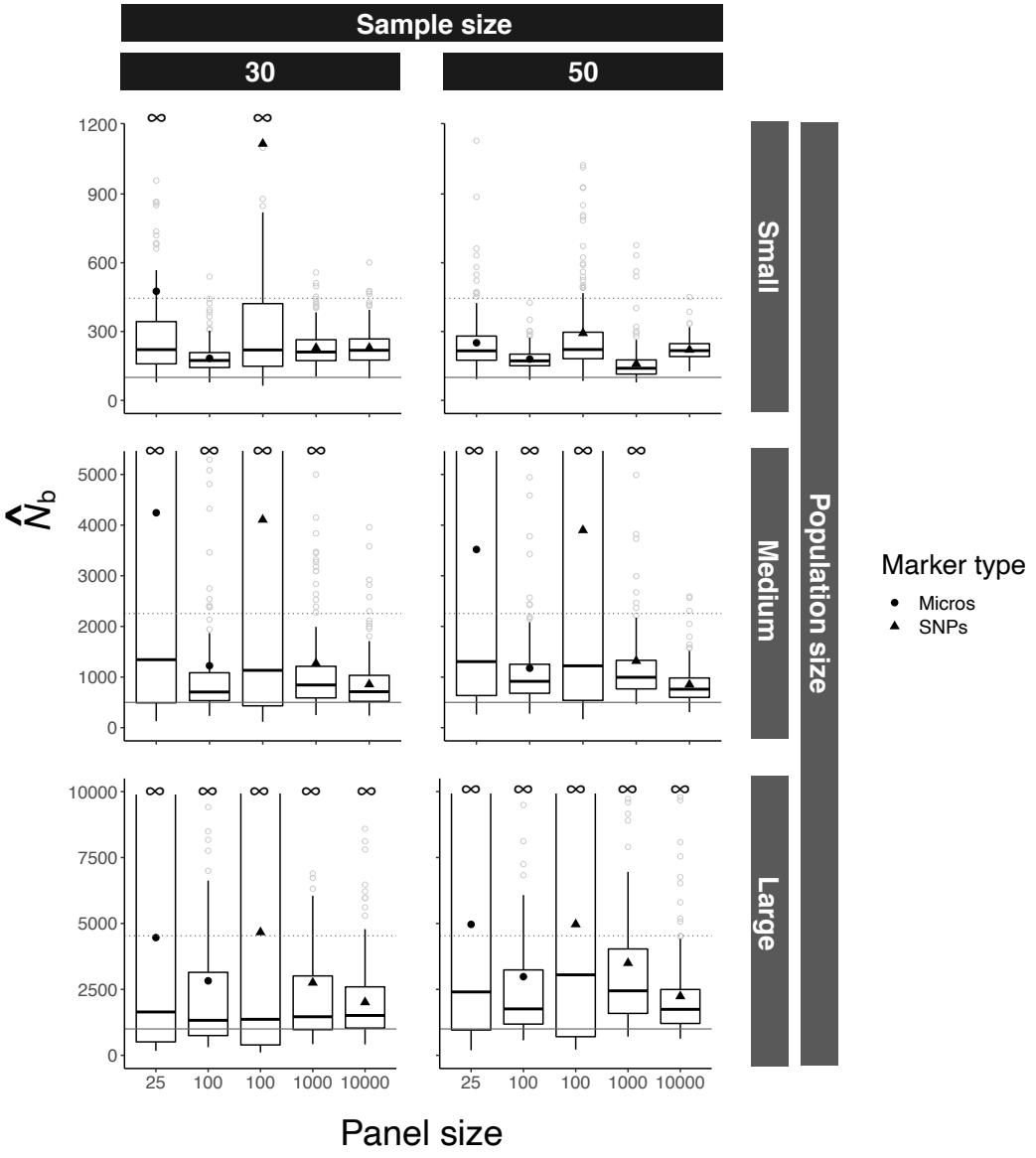

Figure 4 Precision and bias across stable (0% decline) scenarios. Effective number of breeders ($\widehat{N}_b$) and confidence intervals (CI), estimated using the jackknife method, calculated with LDNe v2 as implemented in AGESTRUCNb (Antao et al., 2020). Five panels (L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) were compared across three initial population sizes (rows: N = small, medium, or large), two sample sizes (columns: S = 30 or 50). Boxplots show range of $\widehat{N}_b$ (point estimates) for each scenario. True N_b of Atlantic Salmon is expected to fall between the horizontal dashed line which represents the number of mature individuals (including precocious male parr) and the solid line which represents the number of anadromous individuals (N). Mean $\widehat{N}_b$ for each scenario shown; microsatellites (circles) and SNPs (triangles). Infinite $\widehat{N}_b$ (10,000) represented as ∞. Bias decreased as panel size increased.

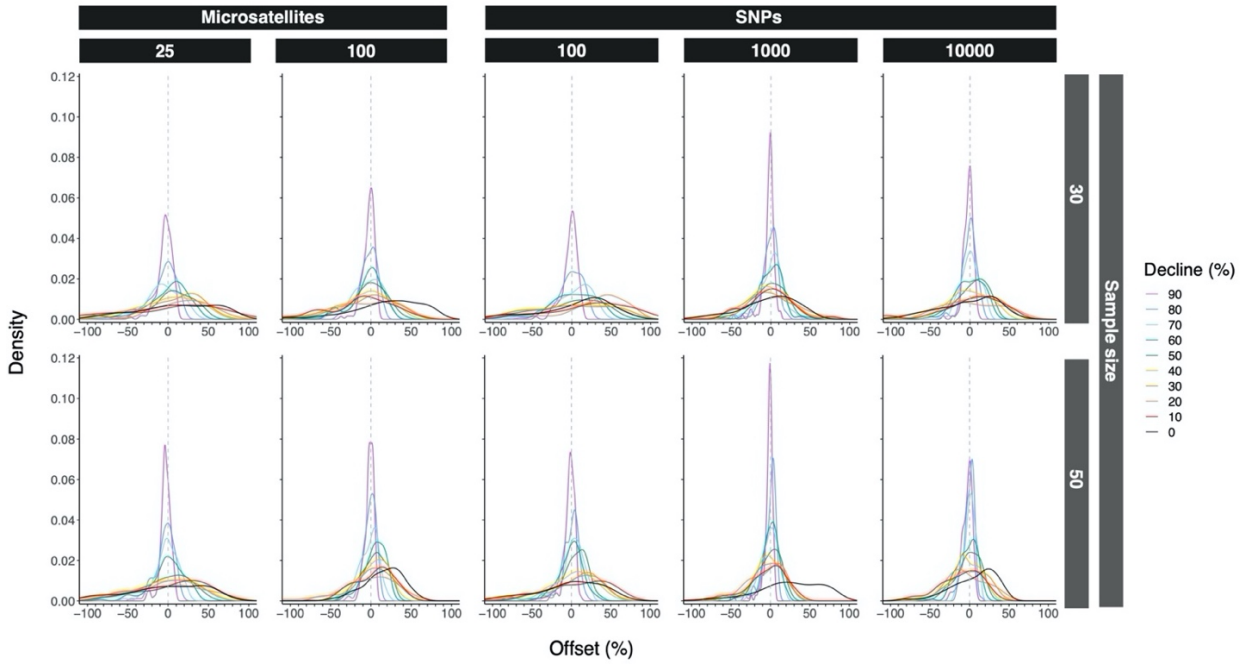

Figure 5 Accuracy of decline in $\widehat{N}_b$ between t_0 and t_x. Offset, the difference between modeled and observed decline in $\widehat{N}_b$, estimated using LDNe v2 as implemented in AGESTRUCNb (Antao et al., 2020) shown across marker type and panel size (columns: L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) for small (S = 30) and large (S = 50) sample sizes when initial population size was small (N = 100). Modeled declines in $\widehat{N}_b$ span a broad range of generations (up to 20) and per generation rates of decline (0 – 60%). The vertical dashed line at 0% offset indicates the observed decline accurately approximates the simulated decline. Negative offset indicates the modeled decline was underestimated and positive offset that the modeled decline was overestimated. Variability in offset indicated by width of the scaled density curve. Plots have been constrained to offset within the range -100 to 100%. See Table S4 for median, minimum and maximum offset per scenario. Precision increased as panel size increased.

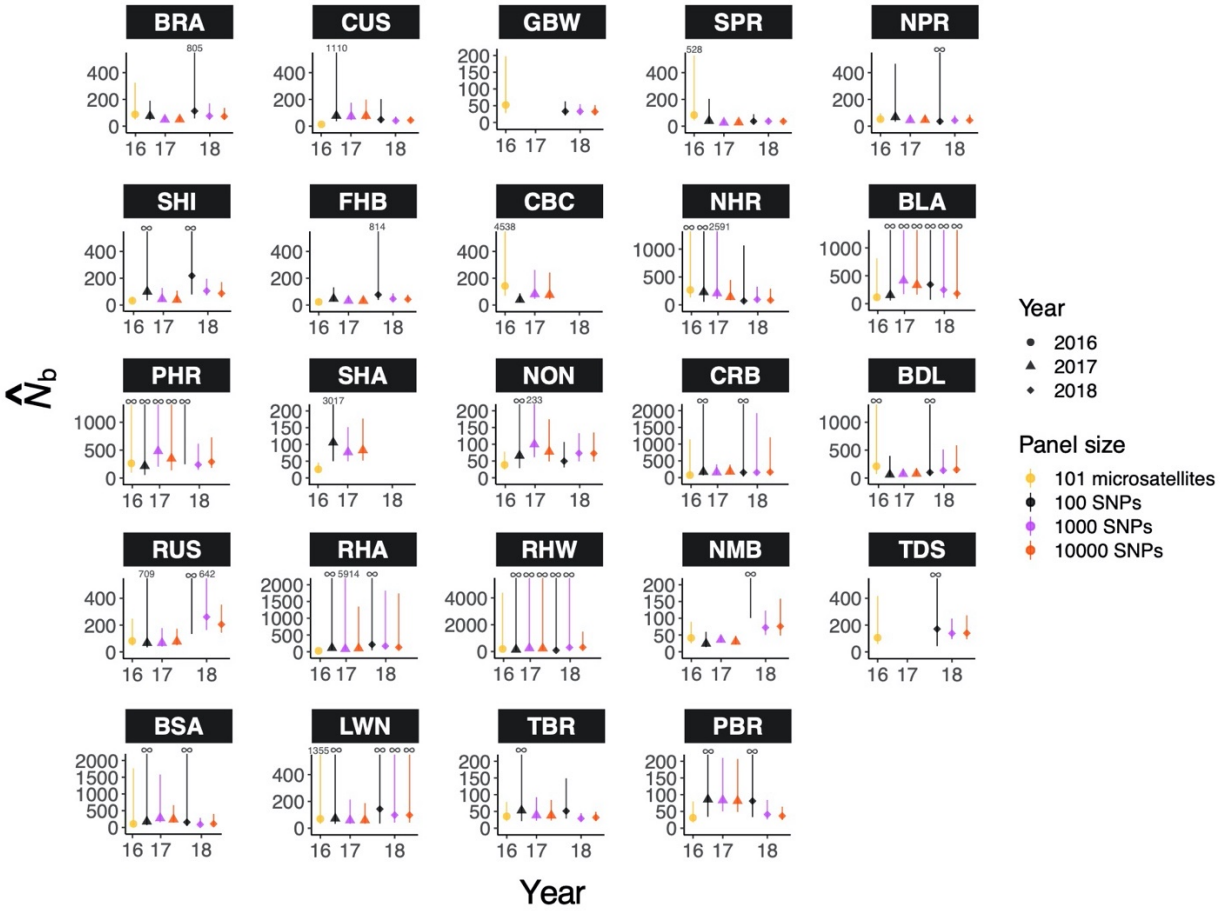

Figure 6 Empirical $\widehat{N}_b$ (point) and confidence intervals (line), jackknifed, for 24 rivers in Placentia Bay, Newfoundland sampled over three discrete years (2016 – 2018). Estimates of N_b generated with small (L = 100; black) and large (L = 1000; purple and 10,000; red) single nucleotide polymorphism (SNPs) panels, subset from a 220K SNP array, and a panel of 101 microsatellites (yellow) using the linkage disequilibrium (LD) method in NeESTIMATOR v2.1 (Do et al., 2014) with a small sample size (S = 20 – 30 individuals). Individuals sampled ranged in age from young of the year (YoY) to 2+; if < 90% individuals were from a single age-class, estimates were considered N_e (*) rather than N_b. Large genomic panels are more accurate than small commonly used panels of ten to a hundred loci.

CHAPTER 3 – Environmentally associated chromosomal structural variation influences fine-scale population structure of Atlantic Salmon (*Salmo salar*)

3.1 Abstract

Chromosomal rearrangements (e.g., inversions, fusions, and translocations) have long been associated with environmental variation in wild populations. New genomic tools provide the opportunity to examine the role of these structural variants in shaping adaptive differences within and among wild populations of non-model organisms. In Atlantic Salmon (*Salmo salar*), variation in chromosomal rearrangements exist across the species natural range, yet the role and importance of these structural variants in maintaining adaptive differences among wild populations remains poorly understood. I genotyped Atlantic Salmon (n = 1429) from 26 populations within a highly genetically structured region of southern Newfoundland, Canada with a 220K SNP array. Multivariate analysis, across two independent years, consistently identified variation in a structural variant (translocation between chromosomes Ssa01 and Ssa23), previously associated with evidence of trans-Atlantic secondary contact, as the dominant factor influencing population structure in the region. Redundancy analysis suggested that variation in the Ssa01/Ssa23 chromosomal translocation is strongly correlated with temperature. My analyses suggest environmentally mediated selection acting on standing genetic variation in genomic architecture introduced through secondary contact may underpin fine-scale local adaptation in Placentia Bay, Newfoundland, Canada, a large and deep embayment, highlighting the importance of chromosomal structural variation as a driver of contemporary adaptive divergence.

3.2 Introduction

Elucidating the genetic basis of adaptation is central to the understanding of evolutionary biology (Schluter, 2009). Adaptive differences, accumulated through selection imposed by spatially and temporally heterogenous environments, enable persistence (Felsenstein, 1976; Savolainen, Lascoux, & Merilä, 2013) and drive the diversification of life (Dobzhansky, 1951). Genome-scale data, now available for many non-model organisms, are highlighting the potential of genomic architecture to facilitate adaptive divergence (Campbell et al., 2018; Wellenreuther et al., 2019) with chromosomal structural variants (e.g., inversions, fusions, and translocations) increasingly being identified and associated with environmental and life history variation (see Wellenreuther & Bernatchez, 2018). However, understanding of (i) the origins of chromosomal structural variants (Rougemont & Bernatchez, 2018; Fuller, Koury, Phadnis, & Schaeffer, 2019; Marques, Meier, & SeeHausen, 2019), (ii) associations between structural variants and complex phenotypes (Lee et al., 2017; Fuller et al., 2019; Jay et al., 2019), and (iii) the effectiveness of different types of structural variants as drivers of adaptive divergence (Rieseberg, 2001; Guerrero & Kirkpatrick, 2014) remains limited.

Chromosomal structural variants can be caused by changes in copy number (insertion, deletion and duplication), orientation (inversion) or position of chromosomes within the genome (translocation and fusion). To date, most work has focused on inversion polymorphisms (Dobigny, Britton-Davidian, & Robinson, 2017) due to the potential for strong suppression of recombination and reduced fertility in heterozygous individuals (Sturtevant, 1917; White, 1978; King, 1993). Inversions are taxonomically widespread (Kirkpatrick & Barton, 2006; Wellenreuther & Bernatchez, 2018), frequently polymorphic within and between species and

populations (Sturtevant, 1938; Dobzhansky, 1951; White, 1978), and commonly align with environmental gradients (Balanyà, Huey, Gilchrist, & Serra, 2009; Kennington & Hoffmann, 2013; Kapun, Fabian, Goudet, & Flatt, 2016) or complex morphological and behavioural phenotypes (Huang, Andrew, Owens, Ostevik, & Rieseberg, 2019; Sinclair-Waters et al., 2018). Unlike inversions that change gene order (Fuller et al., 2019), translocations and fusions physically unlink and/or link genes and reduce recombination within each newly linked chromosome arm (Dumas & Britton-Davidian, 2002) reducing recombination in both heterozygotes and rearranged homozygotes (Bidau, Giménez, Palmer, & Searle, 2001; Castiglia & Capanna, 2002). The potential of translocations and fusions to facilitate adaptation (Charlesworth, 1985) is supported by recent theoretical modeling work (Guerrero & Kirkpatrick, 2014), and empirical evidence of environmental correlations with fusions and translocations (Drosophilia americana, McAllister, 2003; Dichroplus sp., Bidau, Miño, Castillo, & Martí, 2012; Atlantic Salmon, Wellband et al., 2019). However, translocations and fusions remain understudied with little reporting of the distribution or frequency of polymorphisms, and associated phenotypic and environmental variation (Dobigny et al., 2017).

Atlantic Salmon (*Salmo salar*) span the North Atlantic Ocean and exhibit hierarchical spatial structure across their range (King et al., 2007). Genomic differentiation is greatest between continents with eastern (European) and western (North American) populations having diverged more than 600,000 years before present (bp) (Nilsson et al., 2001; King et al., 2007; Rougemont & Bernatchez, 2018). Over the intervening period numerous differences have accumulated including large chromosomal rearrangements, two fusions (Ssa08/Ssa29 and Ssa26/Ssa28) and a translocation (Ssa01/Ssa23) (Brenna-Hansen et al., 2012; Wellband et al., 2019; Lehnert et al., 2019a), which have reduced the number of chromosome pairs from 29 in

Europe to 27 in North America (Hartley, 1988; Phillips & Ráb, 2001). Recent findings suggest that variation in the Ssa01/Ssa23 chromosomal translocation exists in North America, and that this variation was likely introduced through trans-Atlantic secondary contact from European Atlantic Salmon near the end of the last glacial maximum (LGM) approximately 18,000 bp (Bradbury et al., 2015; Rougemont & Bernatchez, 2018). Lehnert et al. (2019a) found that the frequency of the Ssa01/Ssa23 translocation changed with latitude across Atlantic Canada with more northern locations, such as sites in Newfoundland and Labrador, exhibiting the highest frequency of the European karyotype (no Ssa01/Ssa23 translocation) relative to southern populations which were primarily fixed for the North American karyotype (Ssa01/Ssa23 translocation). The importance of this structural variation resulting from secondary contact, and the mechanisms acting to maintain high levels of polymorphism within and across Atlantic Salmon populations in Atlantic Canada remain unknown.

Here I explore fine-scale spatial variation in southern Newfoundland, Canada, an area with pronounced regional spatial structure (Bradbury et al., 2014), and evidence of trans-Atlantic secondary contact (King et al., 2007; Bradbury et al., 2015). I first examine genomic variation within Atlantic Salmon populations among 26 rivers using two discrete years of sampling and identify variation in the Ssa01/Ssa23 translocation as a major driver of population structure. I examine temporal stability of population structure as well as the frequency of this structural variation within and among rivers. Next, I identify environmental associations with population structure and consequently the Ssa01/Ssa23 translocation. I build directly on previous work that identified range-wide polymorphism of a translocation between chromosomes Ssa01 and Ssa23 across Atlantic Canada (Lehnert et al., 2019a) and mitochondrial DNA evidence of trans-Atlantic secondary contact along southeastern Newfoundland (King et al., 2007; Bradbury et al., 2015)

and highlight the potential role of chromosomal structural variation in fine-scale local adaptation and the importance of secondary contact in generating standing genetic variation and driving contemporary adaptive divergence.

3.3 Methods

3.3.1 Sampling and genotyping

Juvenile Atlantic Salmon, young-of-the-year (YOY) and parr (ages 0 to 2+), were collected by electrofishing during the period July to September of 2017 and 2018 from 26 rivers around Placentia Bay, a large (145 km wide at the mouth by 125 km long), deep bay (240 m) separating the Avalon and Burin Peninsulas on the south coast of Newfoundland, Canada (Figure 1 and Table 3). Two rivers, Lance (LAN) and Little Barasway (LBB), were excluded from sampling in 2018 due to small sample size in 2017. Cohorts were assigned based on age-length relationships validated by scale ageing (Sylvester et al., 2019). Fin clips were collected and preserved in 95% ethanol. DNA was extracted using DNeasy Blood and Tissue or DNeasy 96 Blood and Tissue kits (Qiagen, Toronto, ON, Canada) following manufacturer's protocols. Concentration of extracted DNA was assessed using a Nanodrop spectrophotometer and by agarose gel visualization. DNA was standardized to a concentration of 15 ng/µl. A total of 1429 (2017: 745 and 2018: 684) individuals were genotyped by Centre for Integrative Genetics (CIGENE, Ås, Norway) using a 220K bi-allelic single nucleotide polymorphism (SNP) Affymetrix Axiom array developed for Atlantic Salmon as described in Barson et al. (2015). These SNPs were a subset of those in the 930K XHD Ssal array (dbSNP accession numbers ss1867919552–ss1868858426) designed using Norwegian aquaculture salmon. Genotype data were filtered for high quality SNPs based on their clustering patterns and subsequent filtering was

performed using PLINK v 1.9 (Purcell et al., 2007; Chang et al., 2015). SNPs were filtered for a minor allele frequency (MAF) cut-off of 0.01, a missingness threshold of 0.05 across samples, and non-biallelic loci within each year sampled. SNPs were retained only if they passed filtering in both years resulting in a total of 139,038 SNPs. In addition, samples were filtered to retain only individuals with <5% missing genotypes, resulting in a total of 662 and 611 individuals in 2017 and 2018, respectively.

3.3.2 Detection of population structure

For population genetic analyses that require a panel of neutral and unlinked loci I first used PLINK version 1.9 (Chang et al., 2015) to identify outlier loci (F_{ST} > 95th percentile) among sample sites. Loci identified as outliers in both 2017 and 2018 were removed. I then removed SNPs with high physical linkage using a sliding-window approach in PLINK version 1.9. SNPs with a variance inflation factor (VIF) greater than 2 were removed from 50 SNP windows shifted by five SNPs each iteration as in Wellband et al. (2019). This neutral, unlinked dataset was then thinned, keeping one SNP per 200,000 bases, using PLINK version 1.9 (Chang et al., 2015).

Indices of genetic diversity including observed and expected heterozygosity (H_o and H_e) and F_{IS} (Nei, 1987), were calculated per river and year to assess deviation from Hardy-Weinberg equilibrium. Indices were calculated using the basic.stats function in the R package hierfstat (Goudet, 2005) with the neutral, unlinked dataset (n = 6,302 SNPs). Confidence intervals for river-specific F_{IS} were calculated using the boot.ppfis function in hierfstat (Goudet, 2005) with 1,000 bootstrap replicates. Genetic differentiation was assessed by calculating pairwise F_{ST} (Weir and Cockerham, 1984) between rivers using the stamp.fst function, with 100 bootstrap replicates,

in the R package StAMPP (Pembleton, Cogan, & Forster, 2013) and visualized using gplots (Warnes et al., 2016).

To estimate the number of distinct genetic clusters in each year, ADMIXTURE version 1.3 (Alexander, Novembre & Lange, 2009), which calculates individual ancestry proportions using maximum likelihood estimates in a parametric model, was run for K (genetic clusters) 1 to 27 with three different random number seeds with the neutral, unlinked dataset (n = 6,302 SNPs). From ADMIXTURE runs, standard deviation of cross-validation (CV) error was used to select a reasonable range of K as in McCartney-Melstad, Vu & Shaffer (2018). Bar plots of estimated individual ancestry proportions given by the Q-values were generated using R version 3.6.1 (R Core Team, 2019).

The R package pcadapt (Luu et al., 2017) was then used to detect genomic regions associated with population-based differences in genomic architecture across all 26 rivers in each year. Multiple values of K (number of principal components; PCs), ranging from 1 to 100, were explored. The R package qvalue (Storey et al., 2015) was used to correct for false-discovery rate by transforming p-values for all SNPs into q-values which were plotted using the Manhattan plot function in the R package qqman (Turner, 2014). The final number of PC axes retained was K = 2, as I was primarily interested in large scale patterns of population differentiation. The inclusion of additional values of K highlighted inter-individual differences rather than population level differences. Further, upon visual inspection of the Manhattan plots, I found that two strongly divergent genomic regions localized on chromosomes Ssa01 and Ssa23, a known chromosomal translocation (Brenna-Hansen et al., 2012; Lehnert et al., 2019a), were the dominant source of variation across multiple values of K (see Results). These regions associated with the

chromosomal translocation were important drivers of differentiation along the first PC axis and were thus a primary focus of my study.

3.3.3 Environmental association analysis

Spatial patterns of association between genomic variation and climatic (i.e., temperature and precipitation) and habitat (i.e., axial river length, basin relief, number of obstructions, and human population density) variables were investigated using partial redundancy analysis (RDA) as implemented in the R package vegan (Oksanen et al., 2017). Environmental variables were collected from publicly available sources (see Table S5). GPS coordinates of sample sites were used to extract 19 BIOCLIM variables, interpolated monthly climate data at a spatial resolution of 30 arc-seconds averaged for the years 1970 – 2000, from the WorldClim 2.0 database (Fick & Hijmans, 2017). Human population density, a proxy for habitat disturbance, was calculated as in Lehnert et al. (2019b); briefly, human population density in 2000, accessed from NASA NEO, was averaged across a grid (0.5° latitude x 0.5° longitude) around the site. Axial river length, number of obstructions (e.g., dams and waterfalls), and basin relief were obtained from Porter et al. (1974). Axial river length and number of obstructions were approximated using GOOGLE EARTH and maximum elevation from HydroSHEDS digital elevation model (Lehner, Verdin, & Jarvis, 2008) was substituted for basin relief for missing sites (Fair Haven Brook, Red Harbour East, and Piercey's Brook). All climatic and habitat data were standardized using the scale function in R (R Core Team, 2019).

Given that many of the climatic variables were highly correlated ($r > 0.8$), I first performed a variable reduction step using PCA to summarize climatic variation for temperature

and precipitation as in Sylvester et al. (2019). The first two PCs of each PCA were retained to reduce dimensionality and covariance of loadings. Next, latitude and longitude of each river mouth were determined using GOOGLE EARTH. Geographic distance between each river mouth was calculated using the least-cost distance function, constrained to a maximum depth of 100 m below sea-level, in the R package marmap (Pante & Simon-Bouhet, 2013). Multivariate associations between genomic and environmental data were tested using redundancy analysis (RDA), conditioned on geographic distance from the most easterly site (Branch River), with four climatic summary variables (i.e., PC1 and PC2 temperature, and PC1 and PC2 precipitation) and four habitat variables (i.e., basin relief, number of obstructions, axial river length, and human population density) as predictors and individual genotypes as dependent variables. Variance inflation factors were below 2 for all variables indicating no multicollinearity between predictors. The RDA was visualized using the plot function in R version 3.6.1 (R Core Team, 2019). The final number of RDA axes retained (n = 3) was determined by visual inspection of the scree plot. To identify SNPs influenced by climatic and habitat variation, I scaled and centred the raw scores on each constrained RDA axis. I then identified outlier SNPs based on the loadings on each RDA axes, defined here as SNPs more than three standard deviations from the mean. The Manhattan plot function in the R package ggman (Rajagopal, 2020) was used to visualize each of the three retained RDA axes. Visual inspection of the first RDA axis identified outlier SNPs localized on chromosomes Ssa01 and Ssa23, a known chromosomal translocation (Brenna-Hansen et al., 2012; Lehnert et al., 2019a). The correlation between frequency of the Ssa01/Ssa23 chromosomal translocation and environmental variation was further explored in subsequent analyses.

3.3.4 Assignment of translocation karyotype

Variation in the Ssa01/Ssa23 chromosomal translocation was a major source of population level differentiation and associated with environmental variation. Therefore, I next examined variation in translocation frequency. Using the results of pcadapt, outlier SNPs (q-value < 0.05 in both 2017 and 2018, n = 887) in the outlier block regions, Ssa01 (44,000,000 - 53,000,000 bp) and Ssa23 (0 - 9,500,000 bp), were combined as in Lehnert et al. (2019a). Spatial genetic structure of the Ssa01/Ssa23 chromosomal translocation was explored with principal component analyses (PCA) performed using the R package pcadapt (Luu et al., 2017) with K = 3. Based on clustering patterns on the first PC axis, on which individuals were separated into three clusters consistent with Lehnert et al. (2019a), individuals were assigned a karyotype using the kmeans function in R version 3.6.1 (R Core Team, 2019). The three clusters corresponded to three karyotypes: 1) standard North American (homozygous translocated; Ssa01p/Ssa23 and Ssa01q); 2) standard European (homozygous non-translocated; Ssa01p/q and Ssa23); and 3) heterozygous (carrying a translocated and non-translocated copy of the chromosomes). Karyotype assignment followed Lehnert et al. (2019a), which incorporated European samples; greater genetic variation was observed on PC1 and PC2 for the standard European karyotype relative to the standard North American karyotype. This pattern of variation in genetic diversity was consistent in my analysis (see Results).

Neighbor-joining (NJ) trees based on Nei's *D* (Nei, 1972) were generated using outlier loci (q-value < 0.05) within the outlier block regions on Ssa01 and Ssa23 and the R package StAMPP (Pembleton et al., 2013). Trees were visualized using FIGTREE v1.4 (Rambaut, 2012). Genetic differentiation between karyotypes was assessed by calculating pairwise F_{ST} (Weir and Cockerham, 1984) using the R package StAMPP (Pembleton et al., 2013). Linkage

disequilibrium (LD) was calculated among outlier SNPs ($q < 0.05$) on chromosomes Ssa01 and Ssa23 between karyotypes. Pairwise LD (R^2) values were calculated using PLINK v 1.9 (Chang et al., 2015) and visualized using the R package gplots (Warnes et al., 2016).

3.3.5 Frequency of translocation karyotype

Population variation and temporal stability

Heterogeneity in translocation and karyotype frequencies between rivers was tested for each year sampled (2017 and 2018) using an analysis of deviance in a generalized linear model (GLM) with a binomial logistic transformation, followed by a comparison of contrasts as in Mérot et al. (2018), and a pairwise Fisher's exact test adjusted for multiple comparisons. Temporal stability of translocation and karyotype frequency within each river between years was then tested using a pairwise Fisher's exact test. Two rivers, Lance (LAN), and Little Barasway (LBB), were excluded from analysis due to limited sample size. To calculate translocation frequency, I used the equation:

((# homozygous translocated x 2) + (# heterozygotes)) / (# total individuals x 2)

which provides the frequency of the standard North American allele per river.

Environmental associations

Given the association between the Ssa01/Ssa23 chromosomal translocation and environmental variables (see RDA above), I next tested for significant relationships between the identified climatic variables (temperature and precipitation) and translocation frequency using linear regression. In the model, the response variable was the frequency of the North American

allele (Ssa01p/Ssa23) or the standard North American karyotype (homozygous translocated) and the explanatory variable was the climatic summary variables (see Results).

3.3.6 Gene ontology

I examined functional enrichment of genes associated with the Ssa01/Ssa23 chromosomal translocation and identified as significant environmental outliers ($q < 0.05$) on RDA1. I conducted gene ontology (GO) enrichment analysis using GO annotations in the Atlantic Salmon genome from SalmoBase (Samy et al., 2020). A reference set of genes, genes within 10 kb of all 138,038 SNPs from the array, was identified and extracted using BEDTOOLS (Quinlan & Hall, 2010). Outlier sets of genes, genes within 10 Kb of outlier SNPs on RDA1 and located within the translocation were then extracted for 2017 (n = 914) and 2018 (n = 700). The R package topGO version 2.38.1 (Alexa, Rahnenführer, & Lengauer, 2006) was then used to test for over-representation of GO biological processes using a node size of 5 and the 'weight01' algorithm to account for structural relationships among GO terms. Using the same outlier sets of genes I then tested for enrichment of gene profiles using the R package CLUSTERPROFILER (Yu, Wang, Han, & He, 2012). NCBI gene ID numbers were used as search criteria in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, Goto, Sato, Furumichi, & Tanabe, 2012). An alpha level of 0.01 was used to determine significance in both topGO and CLUSTERPROFILER analyses.

3.4 Results

3.4.1 Sampling and genotyping

In total, 662 individuals sampled in 2017 and 611 individuals sampled in 2018 were genotyped and passed quality control thresholds. Exploratory analysis using principal component analysis (PCA) found genetic variation separated Red Harbour East (RHA) from all other rivers along the first two principal component (PC) axes in both 2017 and 2018 (Figure S10). Due to the prevalence of non-anadromous Atlantic Salmon along the south coast of Newfoundland (Verspoor, McGinnity, Bradbury, & Glebe, 2015) and the verification of a significant waterfall at the mouth of RHA, this river was excluded following analyses of neutral genetic structure as these individuals most likely represent a highly divergent landlocked population. A total of 138,451 SNPs, with a high overall genotyping rate (> 99%), and 632 individuals in 2017 and 585 individuals in 2018 were used in downstream analyses.

3.4.2 Detection of population structure

All populations exhibited significant genetic differentiation ($p < 0$; see Figure S11) from each other. Patterns of pairwise F_{ST} clearly indicated strong regional structure within Placentia Bay consistent across both discrete years sampled (Figure S11). Pairwise F_{ST}, which ranged from 0.0062 to 0.11 in 2017 and 0.0077 to 0.083 in 2018, was greatest between rivers along the Burin (Bay de l'Eau (BDL) – Piercey's Brook (PBR)) and Avalon (Branch (BRA) – Ship Harbour (SHI)) Peninsulas indicative of an east-west divide. Rivers along the Avalon Peninsula exhibited a higher degree of genetic differentiation relative to each other as compared to rivers along the Burin Peninsula or head of the bay where neighboring rivers exhibited little genetic

differentiation. Interestingly, Cuslett River (CUS) was found to be genetically similar to more geographically distant rivers along the Burin Peninsula than neighboring rivers on the Avalon Peninsula. Observed (H_o) and expected (H_e) heterozygosity ranged from 0.16 to 0.20 (mean = 0.19) in both 2017 and 2018. Inbreeding coefficient (F_{IS}) ranged from -0.067 to -0.018 (mean = -0.026) in 2017 and -0.036 to 0.00 (mean = -0.023) in 2018 (Table 4).

Fine-scale population structure was observed using a panel of putatively neutral loci (6,302 SNPs with $F_{ST} < 0.05$ globally in both 2017 and 2018; known chromosomal rearrangements excluded). The value of K with the lowest mean CV error in ADMIXTURE was K = 11 in 2017, and K = 9 in 2018. The reasonable range of K, values that had standard deviations that overlapped with the lowest mean CV error were K = 10 and 12 – 14 in 2017, and K = 10 – 11 in 2018 (Figure S12). Red Harbour East (RHA) appeared distinct across all values of K in both years sampled. Although the majority of rivers formed river-specific clusters in both years sampled, the head of the bay (CBC – SHA) appeared to form an admixed cluster, most similar to rivers along southern Burin Peninsula, across all values of K in both 2017 and 2018 (Figure S13).

I also explored spatial structure within Placentia Bay by performing a principal component analysis (PCA) on the full dataset (n = 138,038 SNPs) to detect genomic regions associated with population-based differences in genomic architecture. This analysis similarly found genetic variation separated populations by geographic region along the first two PC axes, with PC1 (variance explained in 2017: 1.8% and 2018: 2.4%) highlighting an east-west divide within Placentia Bay and PC2 (variance explained in 2017: 1.2% and 2018: 1.6%) a north-south divide along the Avalon Peninsula (Figure 7a, b). This pattern of spatial structure was found to be temporally stable with the exception of Northwest Mortier Bay (NMB), which in 2017 clustered more closely with the Avalon Peninsula than the Burin Peninsula on which it is located.

Divergence among clusters identified by PC1 was found to be driven by large outlier block regions (> 8 Mbp; Table S6) on both chromosomes Ssa01 and Ssa23 (Figure 7c, d). Each peak of genetic differentiation (Ssa01 and Ssa23) contained > 400 SNPs that were statistical outliers ($q < 0.05$) in both years. An increasing number of PC axes tested (K = 2 – 100; see Supplementary Information Figure S14) supported Ssa01 and Ssa23 as the dominant factor driving genomic divergence among Atlantic Salmon within Placentia Bay. In a range-wide study, Lehnert et al. (2019a) found genomic divergence driven by Ssa01 and Ssa23 indicative of polymorphism in a known chromosomal translocation that differentiates European and North American salmon (Brenna-Hansen et al., 2012). As in Lehnert et al., (2019a), outlier block regions on Ssa01 and Ssa23 were combined and analyzed together in downstream analyses (see below). I note that physical genomic positions presented in both my study, and that of Lehnert et al., (2019a), are based on the European Atlantic Salmon genome, where Ssa01 and Ssa23 are separate chromosomes (Lien et al., 2016), which differs from the standard North American karyotype where the p arm of Ssa01 has fused to Ssa23 (Brenna-Hansen et al., 2012).

3.4.3 Environmental association analysis

PCA-based reduction of climatic variables was used to construct summary variables for temperature and precipitation. For the PCA of temperature variables, the first PC axis explained 47.1% of site environmental variation and was positively associated with temperature seasonality (BIO4) and temperature annual range (BIO7) but negatively associated with minimum temperature of the coldest month (BIO 6), while the second PC axis explained 2.0% of site environmental variation was most strongly negatively associated with annual mean temperature

(BIO1) (Table S7). For the PCA of precipitation variables, the first PC axis of the summary variable for precipitation explained 65.1% of site environmental variation and was most strongly negatively associated with annual precipitation (BIO12), while the second PC axis explained 2.3% of site environmental variation and was strongly positively associated with precipitation seasonality (BIO15).

In the RDA, total proportion of genetic variance explained by the constraining (environmental) variables was 4.3% and 4.4% with 0.85% and 1.3% of the total proportion of variance explained by the conditioning variable (distance from the most easterly site; BRA) in 2017 and 2018 respectively. I found temperature and precipitation explained the greatest proportion of genetic variance on RDA1 based on vector length and number of significant SNPs. A total of 422 and 222 SNPs were significantly associated with temperature PC1 and 114 and 440 SNPs were significantly associated with precipitation PC1 on RDA1 in 2017 and 2018 respectively. In both discrete years sampled, SNP loadings on RDA1 indicated a strong association with the Ssa01/Ssa23 outlier block regions previously identified (Figure 8c, d), suggesting an association between the translocation and environmental variation. Genetic variance associated with temperature was most strongly driven by rivers at the head of Placentia Bay and along the Burin Peninsula (Figure 8a, b), a pattern that was consistent across years. Whereas, genetic variance associated with precipitation was most strongly driven by rivers along southern Burin and Avalon Peninsulas (Figure 8a, b).

3.4.4 Assignment of translocation karyotype

Given that variation in the Ssa01/Ssa23 chromosomal translocation was identified as a major source of genetic structure, and found to be associated with environmental variation, I next examined translocation and karyotype frequency. Analysis of outlier SNPs (n = 887) within the outlier block regions on Ssa01/Ssa23 found three distinct clusters on PC1 (variance explained in 2017: 54.3% and 2018: 53.6%) (Figure 9a, b), consistent with a chromosomal rearrangement. Similar clustering patterns were found using neighbor-joining (NJ) trees (Figure 9c, d). Karyotype was assigned to each of the three clusters (based on PCA) with individuals assigned as either: 1) standard North American (homozygous translocated; Ssa01p/Ssa23 and Ssa01q), 2) standard European (homozygous non-translocated; Ssa01p/q and Ssa23), and 3) heterozygous (carrying a translocated and non-translocated copy of the chromosomes). The heterokaryotype was found to be intermediate to the homokaryotypes along PC1 (Figure 9). Individuals from throughout Placentia Bay were found in each of the three clusters suggesting karyotype clusters were not completely driven by the geography of the bay. Interestingly, the cluster of individuals found to have the standard European karyotype (homozygous non-translocated; Ssa01p/q and Ssa23) exhibited greater genetic variation along the first two PC axes than the clusters of heterozygous or standard North American karyotype (homozygous translocated; (Ssa01p/Ssa23 and Ssa01q) individuals which exhibited the least amount of genetic variation along the first two PC axes (Figure 9a, b). This was consistent with patterns found by Lehnert et al. (2019a).

Genetic differentiation (F_{ST}) between homokaryotypes was significantly greater ($p < 0.001$) within the Ssa01/Ssa23 outlier block regions (Ssa01: $F_{ST} = 0.55$ and $F_{ST} = 0.52$, and Ssa23: $F_{ST} = 0.71$ and $F_{ST} = 0.70$ in 2017 and 2018 respectively) relative to that observed genome wide ($F_{ST} = 0.0094$ and $F_{ST} = 0.0096$ in 2017 and 2018 respectively). Heatmaps of linkage

disequilibrium (LD) between outlier SNPs ($q < 0.05$) on chromosomes Ssa01 and Ssa23 revealed regions of high LD, in all pairwise comparisons of translocation karyotype (Figure S16). As expected, linkage disequilibrium (R^2) was highest between homokaryotypes. Interestingly, LD between the heterokaryotype and standard European karyotype was found to be lower than that observed between the heterokaryotype and standard North American karyotype. Heterozygosity was found to be four times higher for the standard European karyotype relative to the standard North American karyotype within the outlier block regions (Figure S15).

3.4.5 Frequency of translocation karyotype

Population variation and temporal stability

All rivers were polymorphic for the Ssa01 and Ssa23 chromosomal translocation (Figure S17). Frequency of the translocation did not significantly differ between years ($p = 0.263$) but did differ significantly between rivers within years ($p < 0.01$) (Figure S18). Average translocation frequency (standard North American 'allele') was 61.3% (range: 5.5 – 86.0%) in 2017 and 59.6% (range: 6.7 – 94.5%) in 2018. Frequency of the standard North American karyotype (homozygous translocated; Ssa01p/Ssa23 and Ssa01q) was 41.9% (range: 0 – 72.0%) in 2017 and 41.0% (range: 0 – 89.9%) in 2018 (Table S8). Frequency of the translocation was temporally stable within rivers across the two discrete years sampled with the exception of Northwest Mortier Bay (NMB), Cuslett (CUS), and Big Salmonier (BSA) which differed significantly in both karyotype and translocation frequency between 2017 and 2018 (Table S9). Translocation frequency significantly increased in NMB and significantly decreased in CUS and BSA from 2017 to 2018 (Figure S19).

Environmental associations

The translocation appeared spatially distributed along a longitudinal gradient with the highest frequency of the standard North American karyotype found along central Burin Peninsula and the head of Placentia Bay, and absent or occurring at low frequency along the Avalon Peninsula (Figure 10). A pronounced transition in frequency of the translocation homokaryotype, consistent across both years sampled, was observed between Ship Harbour Brook (SHI) and Fair Haven Brook (FHB). SHI and FHB had an average of 2.3% and 42.7% standard North American karyotype (homozygous translocated) individuals across years, respectively, suggesting an increase in European ancestry (in this genomic region) between sites. Interestingly, variation in translocation frequency between rivers was significantly correlated with temperature PC1 (2017: $p = 0.067$ and 2018: $p = 0.0087$) (Figure 11) but not precipitation PC1 (Figure S20). The translocation (standard North American allele) was found to occur more frequently in rivers that were found to have a lower minimum temperature in the coldest month and exhibited greater variability in temperature both seasonally and annually.

3.4.6 Gene Ontology

I searched the *Salmo salar* genome for annotated genes within 10kb on either side of each SNP within the outlier block regions on Ssa01 and Ssa23 identified as an environmentally associated outlier on RDA1. I identified 260 and 241 unique genes in proximity to the 914 and 700 SNPs identified as RDA1 outliers within the Ssa01/Ssa23 chromosomal translocation in 2017 and 2018 respectively. These genes represent putative targets of selection. I found

functional enrichment of 26 and 35 gene ontology (GO) biological processes ($p < 0.01$, Table S10; see Appendix H for gene annotations) in 2017 and 2018 respectively. Of these, 21 GO biological processes were significantly enriched ($p < 0.01$) in both 2017 and 2018. Of particular interest were processes related to immunity, growth, and oxidative stress. Further, using the outlier SNPs identified above, enrichment tests in KEGG indicated over-representation of the insulin signaling pathway in both 2017 ($p = 0.00763$) and 2018 ($p = 0.00439$).

3.5 Discussion

Chromosomal structural variation is a significant, yet poorly understood source of genetic variation (Wellenreuther et al., 2019) which may underpin complex phenotype and life history variation across a wide range of taxa (Dobigny et al., 2017). In contrast with chromosomal inversions, which have been increasingly associated with adaptive variation (Lamichhaney et al., 2016; Jay et al., 2018; Huang et al., 2019), few well-documented occurrences of fusion or translocation polymorphisms in wild populations have been reported (see Bidau & Martí, 2002; Dobigny et al., 2017; Wellband et al., 2019; Cayuela et al., 2020). My study is among the first to report evidence of an adaptive chromosomal translocation influencing spatial structure in the wild. I find the Ssa01/Ssa23 chromosomal translocation, previously found to reflect trans-Atlantic differences and secondary contact (Brenna-Hansen et al., 2012; Lehnert et al., 2019a), to be polymorphic and associated with fine-scale spatial structure of Atlantic Salmon in Placentia Bay, Newfoundland, Canada. Moreover, I find translocation frequency is significantly correlated with environmental variation in the region. This work extends previous analyses (Bradbury et al., 2015; Lehnert et al., 2019a) providing a high-resolution examination of trans-Atlantic secondary contact in Atlantic Salmon in southern Newfoundland, and highlights the importance of

secondary contact, introgression, and chromosomal structural variation as drivers of adaptive divergence.

3.5.1 Chromosomal translocation drives fine-scale spatial structure

In my study, Atlantic Salmon populations exhibited hierarchical spatial genetic structure within Placentia Bay, with the greatest genetic differentiation occurring between the Avalon and Burin Peninsulas. This is consistent with previous work by Bradbury et al. (2015) which identified two discrete genetic clusters along southern Newfoundland, east and west, with a boundary near the Burin Peninsula. Trans-Atlantic secondary contact has been suggested to be a significant factor structuring this region. Gene flow following secondary contact has been documented for many temperate species that experienced periods of range expansion and contraction throughout the Quaternary (Hewitt, 2000; Tigano & Friesen, 2016). In Atlantic Salmon, secondary contact between European and North American salmon occurred, most recently, during the Pleistocene at the end of the last glacial maximum (King et al., 2007; Rougemont & Bernatchez, 2018) and is supported by evidence of European mitochondrial DNA (King et al., 2007; Bradbury et al., 2015), and recently the identification of a chromosomal polymorphism associated with European ancestry in northern Canada (Lehnert et al., 2019a).

The Ssa01/Ssa23 chromosomal translocation, previously associated with introgression from European Atlantic Salmon into northern Canada (Lehnert et al. 2019a), was found to be polymorphic within Placentia Bay and appeared to be a significant factor contributing to genetic structuring in this region. While whole genome sequencing is required to confirm the presence of a translocation in these populations, this finding strongly supports a hypothesis of trans-Atlantic secondary contact in Placentia Bay. Genetic variation of outlier loci within the outlier block

regions identified on chromosomes Ssa01 and Ssa23 showed three distinct clusters, a pattern observed by Lehnert et al. (2019a) and consistent with a chromosomal rearrangement. Karyotype frequencies showed a longitudinal clinal pattern. Populations on the Avalon Peninsula were predominately composed of individuals with the standard European karyotype (homozygous non-translocated; Ssa01p/q and Ssa23) whereas populations at the head of the bay and on the Burin Peninsula were predominately composed of individuals with the standard North American karyotype (homozygous translocated; Ssa01q and Ssa01p/Ssa23). Interestingly, Lehnert et al. (2019a) found no individuals with the standard European karyotype south of Labrador, although sampling was limited to only four populations in Newfoundland located west of the Burin Peninsula.

Observed patterns of genetic diversity indicate the standard North American karyotype has reduced genetic diversity relative to the standard European karyotype. Furthermore, pairwise comparisons of linkage disequilibrium between karyotypes suggests greater suppression of recombination between the standard North American karyotype and heterokaryotype than between the standard European karyotype and heterokaryotype. This finding is consistent with reports of reduced frequency of recombination and a shift in the distribution of recombination towards the distal ends of chromosomes in fused homokaryotypes and, to a lesser extent, heterokaryotypes (Bidau et al., 2001; Castiglia & Capanna, 2002; Dumas & Britton-Davidian, 2002; Guerrero, & Kirkpatrick, 2014). While North American salmon generally have low levels of European ancestry (~ 3% genome-wide), individuals with the standard European karyotype have been reported to have high levels of European ancestry (> 50%), particularly near the centromeres, within the outlier block regions on Ssa01 and Ssa23 (Lehnert et al., 2019a). Given that the SNP array was developed using European Atlantic Salmon (Barson et al., 2015), it is

possible inferences about diversity may be influenced by ascertainment bias. However, the dataset used here was based on a subset of polymorphic loci and using a similar array and methodology Bradbury et al. (2015) concluded the observed pattern in genetic diversity along southern Newfoundland was the result of historical processes with minimal influence of ascertainment bias.

Interestingly, the outlier block region identified on chromosome Ssa23 (0 – 9.5 Mbp) was larger in size than that reported by Lehnert et al. (2019a) suggesting multiple secondary contact events may have occurred during the colonization of North America following the last glacial maximum, a hypothesis for which Rougemont & Bernatchez (2018) found some support. Moreover, the larger outlier blocks found here may suggest a more recent secondary contact event in Newfoundland compared to northern regions (Labrador). This highlights the need for future studies on the demographic and evolutionary history of Atlantic Salmon in Canada. While other studies have identified evidence that secondary contact influences differentiation range-wide (Lehnert et al., 2019a), I find variation in the Ssa01/Ssa23 chromosomal translocation is significantly associated with genome-wide population structure demonstrating the importance of the Ssa01/Ssa23 chromosomal translocation to fine-scale structuring within Placentia Bay and suggesting the translocation may influence gene flow through incompatibilities or adaptive differences between karyotypes.

3.5.2 Environment correlated with chromosomal translocation

Geographic regions of post-glacial secondary contact can provide opportunities to investigate the evolution and maintenance of chromosomal structural variation and its role in adaptive divergence (Tigano & Friesen, 2016; Lee et al., 2017). Secondary contact events can

generate polymorphism in standing variation, which can then be selected upon in a heterogenous environment (Alcala & Vuilleumier, 2014; Marques et al., 2019). Genomic architecture is an underappreciated source of variation on which selection can act (Wellenreuther et al., 2019). Chromosomal structural variants may be adaptive due to spatially and/or temporally varying selection on i) breakpoints or position effects which cause gene disruption or alter expression (Corbett-Detig, 2016; Puerma, Orengo, & Aguadé, 2016); ii) recombination rate (McDonald, Rice, & Desai, 2016); or iii) alleles captured or accumulated (Wright, 1931; Coyne & Orr, 2004; Fuller et al., 2019). As such, chromosomal structural variants are expected to clearly delineate the genetic boundaries between parapatric populations that straddle an ecotone (Slatkin, 1975; Kirkpatrick & Barton, 2006). Using a fine-spatial scale approach I identified a cline in translocation karyotype that aligned with an environmental gradient consistent with a hypothesis of adaptive significance.

Placentia Bay, a long (125 km) and deep (240 m) embayment in southeastern Newfoundland, spans 145 km at the mouth and narrows towards the head of the bay where summer temperatures are warmer relative to the mouth of the bay (Fisheries and Oceans Canada, 2007). Genotype-environment analysis (redundancy analysis; RDA) indicated temperature range and seasonality best explained the observed spatial genetic structure and highlighted a strong association with the Ssa01/Ssa23 chromosomal translocation. Although the use of air temperature as a proxy for freshwater temperature may not be accurate, particularly across small spatial scales (Hansen, Read, Hansen, & Winslow, 2016), previous work in the region has reported a strong correspondence between air and water temperatures (Bradbury et al., 2014). Furthermore, a significant correlation between translocation frequency and temperature was found, indicating that like inversions, translocations have the potential to be adaptive, a finding supported by

evidence of signatures of positive selection acting on both the standard North American and standard European karyotypes (Lehnert et al., 2019a). Taken together, these findings suggest the standard North American karyotype is adaptive in North America, however, within secondary contact zone(s) where introgression has occurred, the standard European karyotype also confers a locally adaptive advantage, presumably by adding to standing variation on which selection is acts within a heterogenous landscape over a fine-spatial scale. Alternatively, temperature-aligned population structure could be the result of demographic history. Chromosomal structural variants may have been pre-adapted to differing temperature regimes and aligned themselves along the currently observed cline in temperature during post-glacial colonization; consistent with a hypothesis of niche coupling (Knowles, Carstens, & Keat, 2007). Regardless of the influence of colonization history, a significant correlation between contemporary fine-scale population structure and temperature was observed.

Local adaptation has been recognized as an important evolutionary process in salmonids (reviewed in Taylor, 1991; Garcia de Leaniz et al., 2007) with temperature often identified as the dominant factor structuring populations (Larson, Lisi, Seeb, Seeb, & Schindler, 2016). Temperature has been shown to directly influence metabolic and growth rate (Burgerhout et al., 2017; Stehfest et al., 2017; Vikeså, Nankervis, & Hevrøy, 2017), age at smoltification and maturation (Mangel, 1994; Minns et al., 1995; Friedland, 1998), proportion of precocial parr in Atlantic Salmon populations (Valiente et al., 2005; Yates et al. 2015), and migration timing (Jonsson & Jonsson, 2018). While some evidence suggests a difference in proportion of precocial males (Dalley, Andrew, & Green, 1983) between the mouth and head of Placentia Bay, little is known about the life history and ecology of salmon at this spatial scale in this region.

While my results suggest an association between translocation frequency and temperature, I acknowledge that other unmeasured variables that covary with temperature may contribute to the genetic structure observed (Storfer et al., 2006). Inclusion of additional environmental parameters, such as pH and/or geological characteristics (Bourret et al., 2013; Bradbury et al., 2014), and pathogen or parasite diversity (Dionne, Miller, Dodson, & Bernatchez, 2009), may provide further insight into the mechanisms influencing spatial structure in the region. Notably, in regions such as Labrador, previous work has found a higher frequency of the standard European karyotype within populations within a large marine embayment (Lake Meville) where temperatures are warmer compared to coastal populations with a lower frequency of the standard European karyotype (Sylvester et al., 2018a; Lehnert et al., 2019a). While this association has not been formally tested, it highlights another region in North America where polymorphism of the Ssa01/Ssa23 chromosomal translocation may align with clinal variation in temperature. Although these relationships appear to operate in different directions, this may further support the role of multiple secondary contact events in parts of Canada from different regions of Europe.

3.5.3 Chromosomal translocation exhibits temporal stability across ecotone

Translocation frequency was found to be temporally stable across the two discrete years sampled. A pronounced transition in translocation frequency was observed between Ship Harbour (SHI) and Fair Haven Brook (FHB) on the Avalon Peninsula, however, this transition was less apparent on the Burin Peninsula, where Northwest Mortier (NMB) and Big Salmonier (BSA), rivers located in an intermediary zone of clinal variation, were found to significantly differ in translocation frequency between 2017 and 2018. While maintenance of polymorphism and temporal stability suggest the Ssa01/Ssa23 chromosomal translocation is adaptive, it is plausible

that the Ssa01/Ssa23 chromosomal translocation is neutral or near-neutral and polymorphism persists due to demographic factors such as low gene flow and low effective population size in region (Palstra et al., 2007; Bradbury et al., 2015) and drift. Alternatively, polymorphism of the Ssa01/Ssa23 chromosomal translocation may be maintained, not because the variants represent adaptations to divergent habitats, but because the homozygous translocated karyotype carries harmful recessive mutations as suggested by Jay et al. (2019) in *Heliconious numata*. Future work should examine transposable element (TE) dynamics and the rate of non-synonymous to synonymous substitution (dN/dS) within the translocated region.

3.5.4 Functional significance and gene annotations

I identified biological processes which were over-represented using genes within the outlier block regions on Ssa01 and Ssa23 that were located near environmentally associated SNPs. KEGG pathway analysis found enrichment of the insulin signalling pathway in both 2017 and 2018. In addition, analysis of gene ontology found over-representation of biological processes primarily related to regulation of metabolic processes, and immune response. The insulin signalling pathway may relate to both metabolic processes (Babbitt, Warner, Fedrigo, Wall, & Wray, 2010; Zhang et al. 2018) and immunity (Yada & Tort, 2016; Cheng et al., 2017; Wang et al., 2019).

In salmon, genes and/or biological processes related to metabolic processes and immunity can be influenced by environmental factors, such as temperature (Dionne, Miller, Dodson, Caron, & Bernatchez, 2007; Beauregard et al., 2013). Metabolic differences in salmon populations under different temperature regimes have been reported; salmon in warmer conditions grow faster and migrate to sea at a younger age (Power, 1981; Metcalfe & Thorpe, 1990) while salmon in cooler

environments exhibit higher growth rate and more efficient metabolic processes (Nicieza et al., 1994). In Atlantic Salmon, precocial male parr are common in Newfoundland (Dalley et al. 1983), and their occurrence can be influenced by temperature (Valiente et al., 2005; Yates et al. 2015).

Temperature regime has also been correlated with bacterial diversity and as such genetic diversity for immune-related genes (Dionne et al., 2007, 2008) and is thought to be involved in local adaptation to different pathogen communities (Bourret et al., 2013). Parasites have played a major role in mortality of wild fish in Newfoundland (Khan, 2009) with outbreaks and mortality of proliferative kidney disease (PKD) being seasonal and temperature dependent in salmonids (Sterud et al., 2007). Overall these processes point to potential adaptive associations with temperature but remain speculative, and experimental work is needed to better understand these relationships.

3.5.5 Conservation and management implications

Atlantic Salmon in southern Newfoundland have undergone significant declines in abundance over the last few decades (Chaput et al., 2012; Lehnert et al., 2019b) and are currently managed as a single designatable or evolutionarily significant unit (COSEWIC, 2010). My results clearly demonstrate two genetic clusters and provide strong evidence that an adaptive chromosomal translocation associated with trans-Atlantic secondary contact drives fine-scale population structure in the region. Placentia Bay is of particular interest in that it is a geographic region of post-glacial secondary contact and as such provides an opportunity to investigate the evolution and maintenance of chromosomal structural variation and its role in adaptive divergence because evolutionary dynamics of chromosomal structural variants differ from other

parts of the genome, (Tigano & Friesen, 2016; Lee et al., 2017; Wellenreuther & Bernatchez, 2018; Mérot, Oomen, Tigano, & Wellenreuther, 2020). While polymorphism in chromosomal structural variation, such as the Ssa01/Ssa23 chromosomal translocation characterized here, may complicate the process of delineating evolutionarily significant units, particularly when the relative fitness consequences are unclear, I must consider genomic architecture underlying adaptive phenotype or life history variation when predicting the consequences of environmental disturbance and climatic change (Oomen, Kuparinen, & Hutchings, 2020).

3.5.6 Conclusions

Species and populations adapt through selection imposed by spatially and temporally heterogenous environments on new mutation or standing genetic variation (Wright, 1931; Coyne & Orr, 2004). Gene flow is an important source of standing genetic variation (Tigano & Friesen, 2016) promoting adaptation through the re-introduction of previously lost variation (Rieseberg, 2009) and the introgression of novel genetic variants and allelic combinations among meta-populations (Poelstra, Richards, & Martin, 2018). The Ssa01/Ssa23 chromosomal translocation (Brenna-Hansen et al., 2012) has recently been found to be polymorphic within secondary contact zones in North America (Lehnert et al., 2019a). This study further supports secondary contact with European introgression into Atlantic Salmon populations along southeastern Newfoundland and highlights the importance of secondary contact in shaping population genetic structure. Effects of chromosomal structural variants are expected to be most pronounced when fixed in populations prior to secondary contact, with subsequent reproductive isolation maintained by adaptive change involving many genes with small fitness effects (Feder, Nosil, & Flaxman, 2014). Here, I find evidence of an adaptive chromosomal rearrangement and a cline in

translocation frequency aligned with a cline in temperature. These findings suggest the standard North American karyotype is broadly adaptive in North America, however, within secondary contact zone(s) where introgression of the standard European karyotype has occurred, the standard European karyotype also confers an adaptive advantage in local populations. Future work should explore the roles of demography and drift, monitor clinal stability of translocation frequency over an extended period of time to investigate the evolution and maintenance of putatively adaptive translocation in the wild, and use direct temperature measurements and common-garden experiments to investigate differential gene expression between homokaryotypes. This study highlights the importance of chromosomal structural variation as a source of standing variation on which selection can act.

Table 3 Sampling locations of Atlantic Salmon (*Salmo salar*) in Placentia Bay, Newfoundland, Canada. Rivers ordered geographically, east to west around the bay. Samples were collected in 2017 and 2018. Number of samples genotyped (N) per site per year.

River	**RiverID**	**Longitude** (∘W)	**Latitude** (∘N)	$\mathbf{N_{2017}}$	$\mathbf{N_{2018}}$
Branch River	BRA	-53.97	46.89	30	32
Lance River	LAN	-54.07	46.82	9	-
Cuslett Brook	CUS	-54.16	46.96	30	30
Great Barasway Brook	GBW	-54.06	47.13	30	41
Little Barasway Brook	LBB	-54.04	47.18	16	-
Southeast Placentia River	SPR	-53.88	47.23	30	31
Northeast Placentia River	NPR	-53.84	47.27	30	24
Ship Harbour Brook	SHI	-53.87	47.35	30	32
Fair Haven Brook	FHB	-53.89	47.54	30	30
Come By Chance River	CBC	-53.98	47.86	30	31
North Harbour River	NHR	-54.03	47.92	30	32
Black River	BLA	-54.16	47.89	30	30
Pipers Hole Brook	PHR	-54.27	47.93	30	33
Sandy Harbour River	SHA	-54.36	47.71	30	9
Nonsuch River	NON	-54.65	47.45	30	20
Cape Roger Brook	CRB	-54.69	47.44	30	32
Bay de l'Eau River	BDL	-54.73	47.51	30	32
Rushoon River	RUS	-54.92	47.37	30	37
Red Harbour East River	RHA	-54.99	47.33	30	26
Red Harbour West River	RHW	-55.01	47.29	30	30
Northwest Mortier Bay Brook	NMB	-55.31	47.17	30	32
Tides Brook	TDS	-55.26	47.13	30	28
Big Salmonier Brook	BSA	-55.22	47.06	30	31
Lawn River	LWN	-55.54	46.95	30	30
Taylor Bay Brook	TBR	-55.71	46.88	30	10
Piercey's Brook	PBR	-55.86	46.88	30	21

Table 4 Summary of genetic diversity for Atlantic Salmon from 26 rivers in Placentia Bay, Newfoundland, Canada. Rivers are ordered geographically, east to west around the bay. Number of samples (N), mean observed (H_o) and expected (H_e) heterozygosity and median F_{IS} calculated across putatively neutral single nucleotide polymorphism (SNP) loci (n = 6,302) using hierfstat.

RiverID	2017				2018			
	N	H_o	H_e	F_{IS}	N	H_o	H_e	F_{IS}
BRA	30	0.184	0.182	-0.025	32	0.187	0.183	-0.025
LAN	9	0.181	0.170	-0.067	-	-	-	-
CUS	30	0.184	0.185	-0.018	30	0.189	0.185	-0.036
GBW	18	0.188	0.186	-0.030	41	0.189	0.186	-0.026
LBB	15	0.191	0.184	-0.037	-	-	-	-
SPR	27	0.192	0.185	-0.040	30	0.194	0.188	-0.036
NPR	29	0.193	0.192	-0.018	17	0.192	0.191	-0.032
SHI	22	0.202	0.200	-0.024	32	0.204	0.201	-0.018
FHB	30	0.188	0.185	-0.036	25	0.193	0.188	-0.029
CBC	30	0.192	0.194	-0.018	19	0.194	0.192	-0.029
NHR	24	0.188	0.188	-0.022	29	0.192	0.189	-0.018
BLA	24	0.196	0.195	-0.022	29	0.198	0.198	-0.018
PHR	29	0.193	0.193	-0.018	30	0.196	0.194	-0.018
SHA	30	0.198	0.196	-0.018	9	0.197	0.198	0.000
NON	27	0.181	0.177	-0.020	20	0.181	0.180	-0.027
CRB	29	0.190	0.190	-0.018	27	0.183	0.182	-0.020
BDL	30	0.181	0.181	-0.018	32	0.188	0.187	-0.016
RUS	25	0.182	0.183	-0.021	31	0.182	0.182	-0.017
RHA	30	0.158	0.158	-0.018	26	0.159	0.158	-0.020
RHW	27	0.196	0.196	-0.020	24	0.196	0.194	-0.022
NMB	27	0.185	0.181	-0.040	31	0.193	0.189	-0.035
TDS	15	0.179	0.177	-0.037	24	0.186	0.185	-0.022
BSA	30	0.188	0.187	-0.018	24	0.187	0.184	-0.022
LWN	27	0.193	0.190	-0.020	22	0.195	0.194	-0.024
TBR	18	0.193	0.194	-0.030	10	0.192	0.199	0.000
PBR	30	0.194	0.194	-0.018	17	0.194	0.192	-0.032

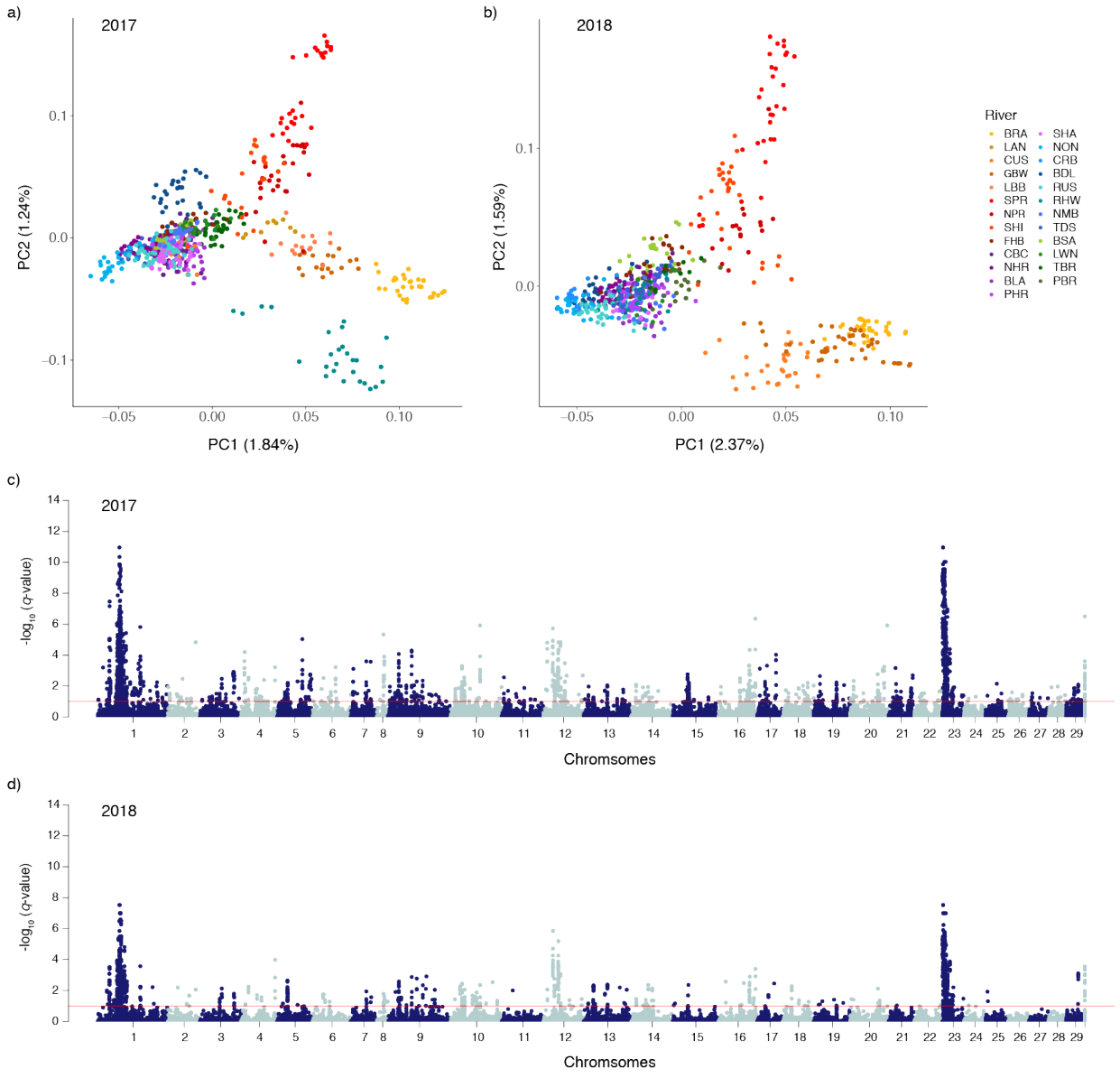

Figure 7 Genomic outlier blocks drive spatial structure of Atlantic Salmon. (a, b) Genetic structure across Placentia Bay, Newfoundland, Canada based on the first two principal component (PC) axes from pcadapt (Luu et al., 2017) using 138,451 SNPs. (c, d) Manhattan plots showing genomic regions of variation based on PC1. Samples collected in (a, c) 2017 (b, d) 2018. Rivers coloured east (yellow-red) to west (green-blue) with head of the bay (purple). Red line represents a genome-wide significance threshold of 5.0e-8.

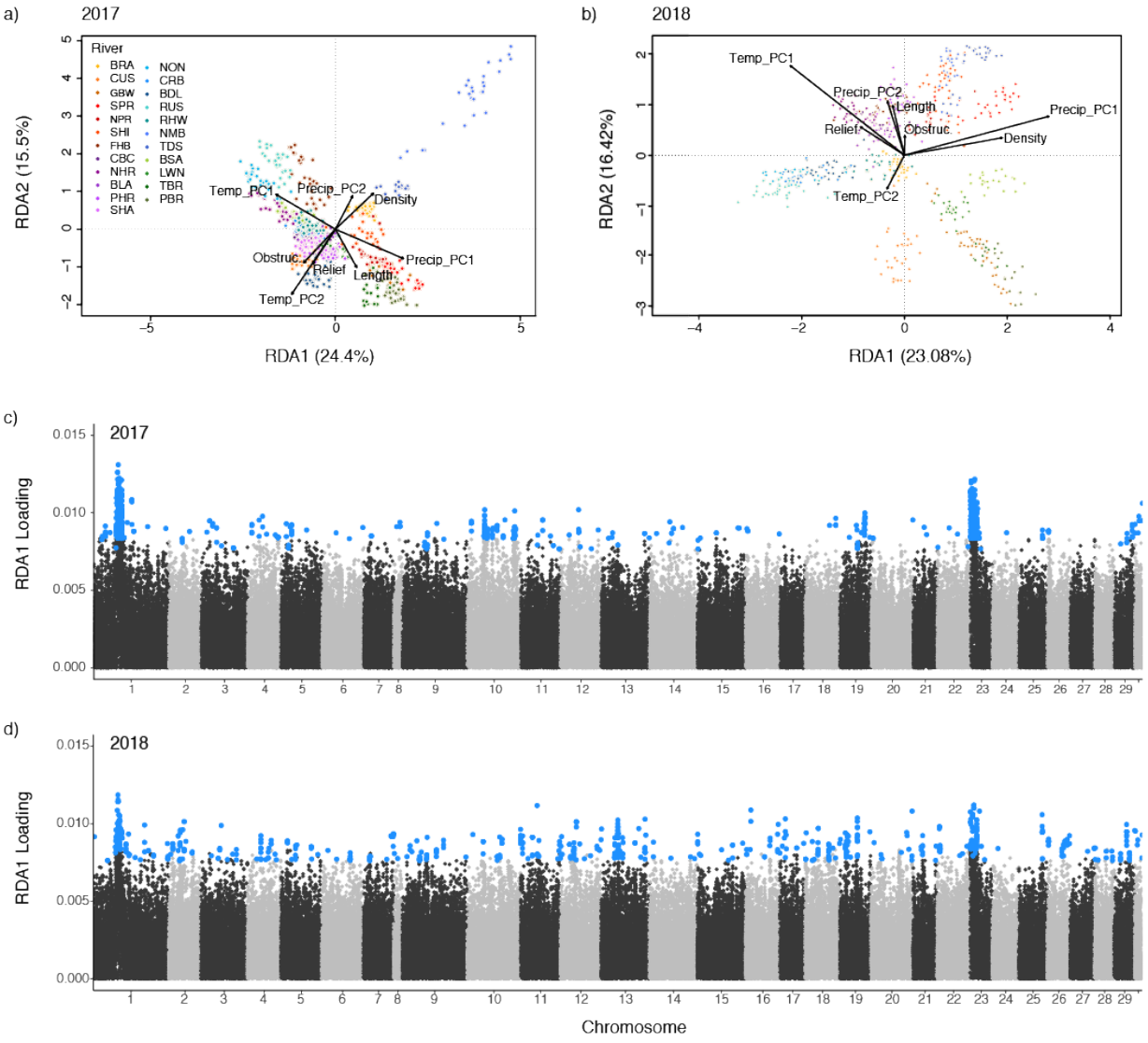

Figure 8 Redundancy analysis (RDA) of (a) 2017 and (b) 2018 based on PCs 1 and 2 of BIOCLIM (WorldClim) temperature and precipitation variables, and habitat variables. Manhattan plots, showing absolute loadings, of the distribution of outlier SNPs (blue) associated with the first RDA axis of (c) 2017 and (d) 2018.

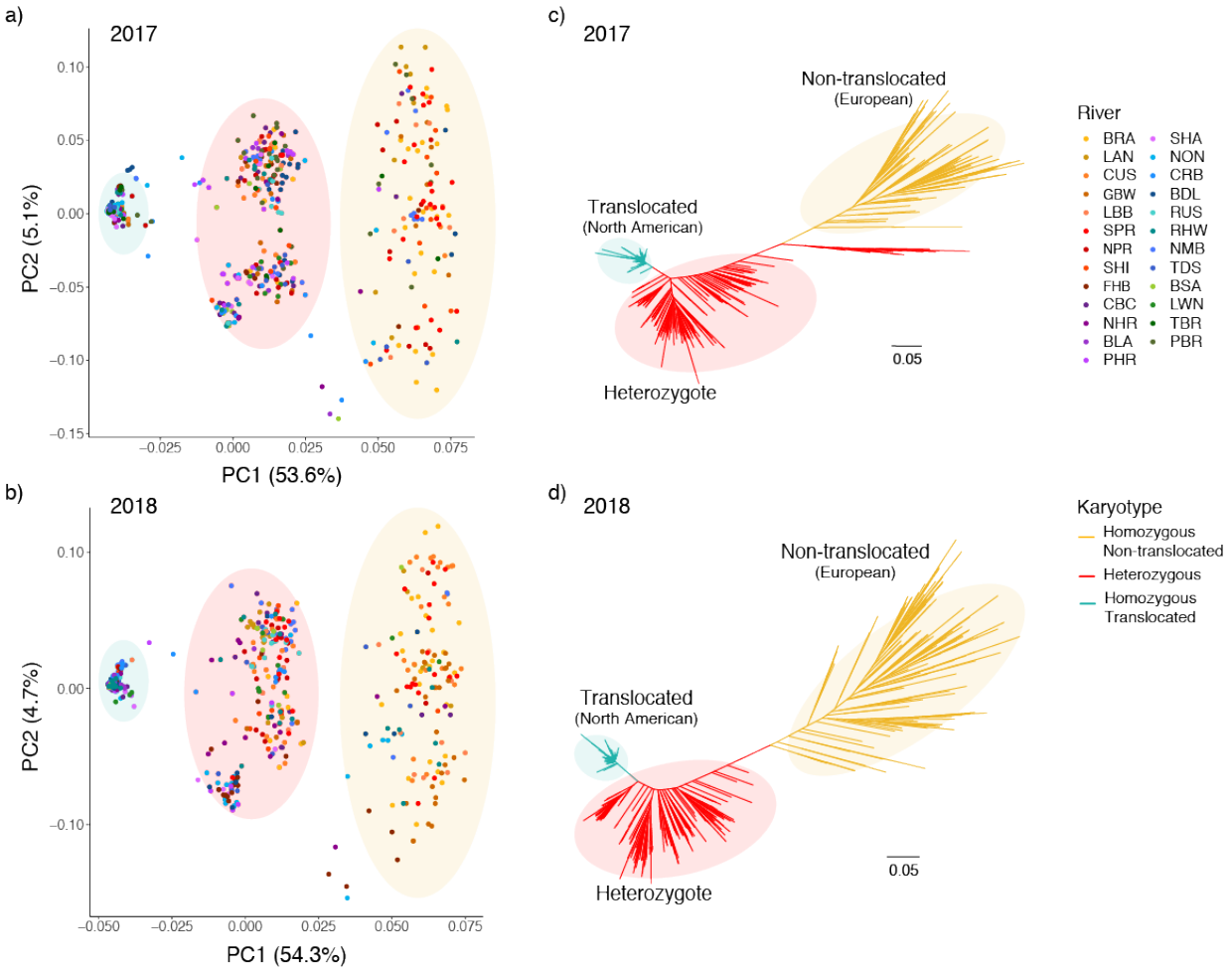

Figure 9 Genetic relationships between individual Atlantic Salmon based on outlier SNPs (q < 0.05) within the Ssa01p/Ssa23 chromosomal translocation. Samples collected in (a) 2017 and (b) 2018. Neighbor-joining (NJ) tree for (c) 2017 and (d) 2018. Homozygous European non-translocated (Ssa01p/q and Ssa23) karyotype (yellow), heterozygous (red), and homozygous North American translocated (Ssa01q and Ssa01p/23) karyotype (blue-green).

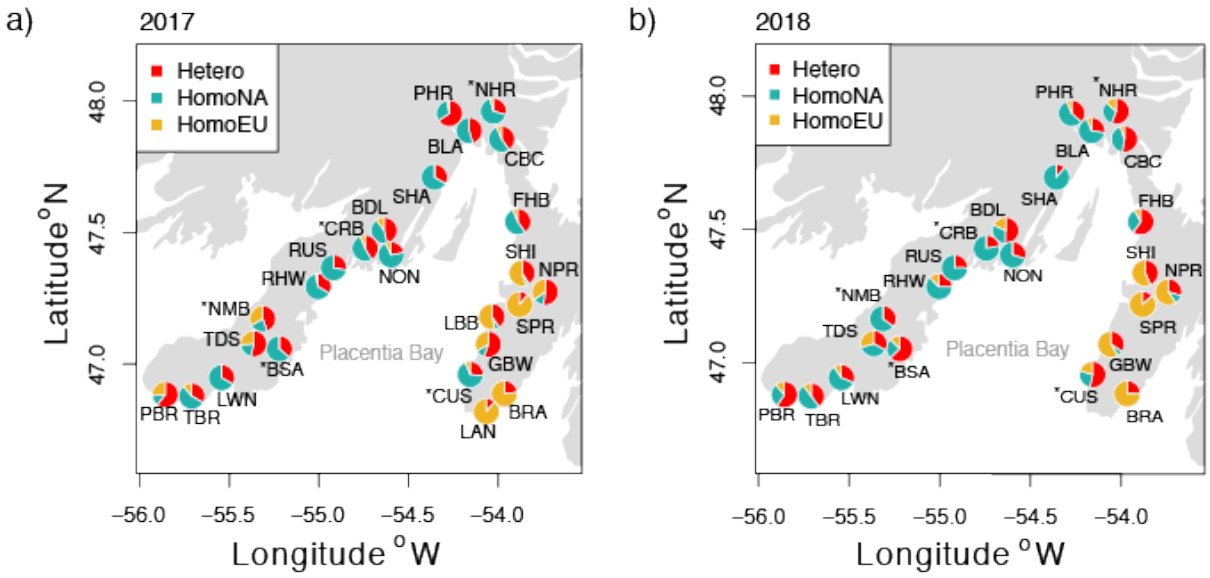

Figure 10 Frequency of Ssa01 and Ssa23 chromosomal translocation exhibits fine-scale spatial variability in Placentia Bay, Newfoundland, Canada. (a, b) Karyotype frequency within river in (a) 2017 and (b) 2018. Asterisk (*) indicates karyotype frequency differed significantly between years. (c) Translocation frequency between rivers across years. Homozygous European non-translocated (Ssa01p/q and Ssa23) karyotype (yellow), heterozygous (red), and homozygous North American translocated (Ssa01q and Ssa01p/23) karyotype (blue-green).

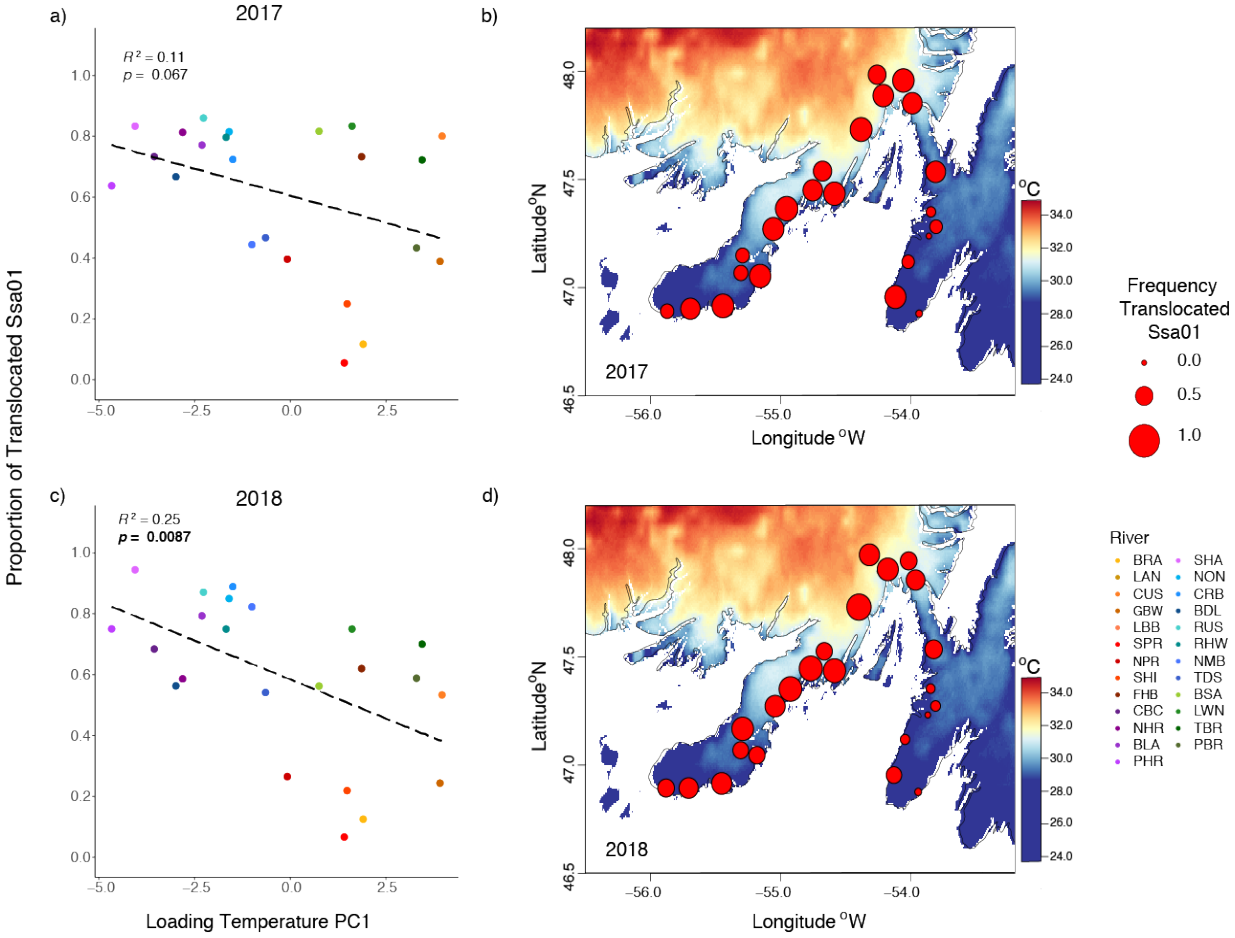

Figure 11 Correlation between temperature and the Ssa01/Ssa23 chromosomal translocation in Atlantic Salmon within Placentia Bay, Newfoundland, Canada. (a, c) Linear regression of the first principal component (PC) of a PCA based on 11 temperature variables (BIOCLIM) and proportion of non-translocated Ssa01p/q sampled in (a) 2017 and (c) 2018. (b, d) Temperature annual range (BIO7), the highest loading variable on temperature PC1. Size of point indicates frequency of non-translocated Ssa01p/q in (b) 2017 and (d) 2018.

CHAPTER 4 – Conclusion

4. 1 Summary

Atlantic Salmon have been assessed as threatened in southern Newfoundland (COSEWIC, 2010). Threats include poor marine survival (Friedland et al., 1998) and direct genetic interaction (i.e., hybridization), genetic change due to interbreeding, with escaped farm salmon (Wringe et al., 2018). In Newfoundland, traditional monitoring methods (i.e., angling surveys and counting fences) are implemented in a small number of all salmon rivers; there is currently a seasonal counting fence in one river (Northeast Placentia; monitored 1984 – 2002 and 2015 – present) in Placentia Bay. There is therefore considerable interest in genetic monitoring to complement traditional demographic monitoring and provide unique insights into population spatial structure, species life history and genetic diversity, and to detect future change in distribution and abundance of populations. Here, I generate a pre-impact genomic baseline against which future change will be assessed. Furthermore, I demonstrate the utility of such a large genomic dataset to resolve both the nature of adaptive diversity and demographic changes in wild populations.

The findings presented in this thesis are the result of analyzing two genomic datasets consisting of 101 polymorphic microsatellite loci and 220K bi-allelic SNPs distributed across the North American Atlantic Salmon genome. These large genomic datasets are used to describe both neutral and adaptive population genetic structure and provide estimates of effective population size for 26 rivers in Placentia Bay, Newfoundland, Canada sampled across three discrete years (2016 – 2018). I additionally examined the utility of genome-scale data for the timely detection

and quantification of contemporary population declines using the single-sample linkage disequilibrium (LD) method.

In Chapter 2, I examined the power of large genomic datasets in combination with the most commonly used method of N_e estimation, the LD method, as a tool for genetic monitoring. I find large panels enable detection of less severe population declines with greater accuracy than small, commonly used panels of tens of microsatellite loci or hundreds of SNPs. In the past, genetic monitoring has aimed to detect population bottlenecks, large declines in population size. Here, large panels not only detected smaller declines but recovered the magnitude of decline with reasonable accuracy; large panels (L = 100 microsatellites or ≥ 1000 SNPs) detected declines in $\widehat{N}_b$ as small as 30% within as little as a single generation when initial population size was small (N = 100). Diminishing returns in power and accuracy were observed when panel size was increased from 1000 to 10,000 SNPs and when sample size was increased from 30 to 50 individuals dependent on marker type and initial population size. Although false positives (type I error) were more frequently observed with large relative to small panels, large panels were less prone to false negatives (type II error; failure to detect decline) than small panels. As such, large panels are highly recommended for genetic monitoring of species of conservation concern (i.e., small population size) for which extinction or extirpation is a risk.

In Chapter 3, I built on recent research which found range-wide polymorphism of the Ssa01/Ssa23 chromosomal translocation in North America was likely introduced through trans-Atlantic secondary contact with European Atlantic Salmon (Lehnert et al., 2019a). I find the Ssa01/Ssa23 chromosomal translocation to be the dominant factor influencing population structure of Atlantic Salmon in Placentia Bay, Newfoundland, Canada. Frequency of the

translocation was temporally stable across the two discrete years sampled (2017 – 2018) and was found to be strongly positively correlated with temperature seasonality and temperature annual range and negatively correlated with minimum temperature of the coldest month.

4. 2 Applications to conservation and management

Despite a full moratorium on the commercial fishery since 2000 and reduced daily and season retention bag limits in the recreational fishery over recent years (Fisheries and Oceans Canada, 2009), populations in Placentia Bay have remained small. Current population productivity, measured as N_b, was low over the three discrete years sampled (mean $\widehat{N}_b < 100$). No trend in $\widehat{N}_b$ was observed suggesting future research is needed to quantify natural fluctuations in the effective number of breeders annually to determine the sampling frequency and temporal period of sampling required to detect trend in $\widehat{N}_b$ with reasonable accuracy for Atlantic Salmon within this region. However, evidence of continued declines in N_e over the recent past (LINKN$_e$) may be cause for immediate conservation concern. Low N_e increases the potential for inbreeding (Palstra & Ruzzante, 2008) and may constrain a populations ability to adapt over evolutionary timescales elevating the risk of extirpation or extinction (Shaffer, 1981; Frankham, 1995). Interestingly, precocial male parr, common in Newfoundland (Dalley et al. 1983), may reduce inbreeding and significantly increase N_e and N_b (Johnstone et al. 2013; Perrier et al., 2014). Future work should estimate river specific abundance of precocial male parr and calculate the inbreeding coefficient of each population Here, preliminary analysis found at least one full- or half-sib pair in the majority of rivers sampled with small clusters of full-sibs more likely in rivers on the Avalon Peninsula (see Appendix I).

Strong genetic differentiation at neutral loci was observed between the Avalon (east) and Burin (west) Peninsulas corroborating previous evidence of an east-west divide within southern Newfoundland (Bradbury et al., 2015) and suggesting multiple discrete units of Atlantic Salmon are present in Placentia Bay, Newfoundland. Here, I find evidence of an adaptive chromosomal polymorphism which aligned with a cline in temperature seasonality. The standard North American (translocated; Ssa01p/Ssa23 and Ssa01q) karyotype appears adaptive as does the standard European (non-translocated; Ssa01p/q and Ssa23) karyotype within this secondary contact zone.

Interestingly, the outlier block region on chromosome Ssa23 was found to be larger than reported in a range-wide study of the Ssa01/Ssa23 chromosomal polymorphism in North America (Lehnert et al., 2019a). Furthermore, the relationship between karyotype and temperature appears to operate in different directions across secondary contact zones (i.e. Lake Melville, Labrador and Placentia Bay, Newfoundland) within the North American range (Sylvester et al., 2018a; Lehnert et al., 2019a). This finding may support a hypothesis of multiple secondary contact events in Atlantic Canada from different regions of Europe as suggested by Rougemont & Bernatchez (2018). Future work should explore the demographic and evolutionary history of these regions with particular focus on timing of colonization and potential source populations. Despite the relationship between translocation frequency and temperature seasonality appearing to operate in opposing directions, rivers in both locations (Lake Melville, Labrador and Placentia Bay, Newfoundland) with a high frequency of the standard European (non-translocated) karyotype generally had lower $\widehat{N}_e$ than rivers in the same embayment with a high frequency of the standard

North American (translocated) karyotype. Overall, these results support a finding of multiple evolutionarily significant units within Placentia Bay, Newfoundland.

Asymmetry in the degree of introgression with evidence of gene flow towards the Avalon Peninsula (east) was reported by Bradbury et al. (2015). As such, it is highly recommended that future contributions to the pre-impact genomic baseline screen wild populations in Placentia Bay for introgression with farm fish. Here, both spatial genetic structure and karyotype frequency were found to be temporally stable over the two discrete years sampled. Future work should continue to monitor clinal stability; a pronounced transition in karyotype frequency was observed between Ship Harbour Brook (SHI) and Fair Haven Brook (FHB) on the Avalon Peninsula and a transition zone was observed south of Northwest Mortier Bay Brook (NMB) on the Burin Peninsula. Long-term maintenance of chromosomal structural polymorphisms is unusual (Charlesworth, 2006), yet polymorphism of the Ssa01/Ssa23 chromosomal translocation is likely to have persisted in south Newfoundland since the end of the last glacial maximum (~ 14,000 bp) (Shaw & Potter, 2016). To investigate the evolution and maintenance of this putatively adaptive chromosomal polymorphism in the wild, future research should utilize common-garden experiments to investigate differential gene expression between homokaryotypes under varying thermal regimes.

Future research should additionally build on the work of Dalley et al. (1983) and Hutchings & Jones (1998) and describe and monitor river specific rates of precocial maturation, growth rate, smolt age, and sea-age at maturity. Temperature regime (seasonal and annual) is expected to influence growth, smolt age, and the frequency of precocial male maturation (Minns et al., 1995; Friedland, 1998; Valiente et al., 2005; Yates et al. 2015). While some evidence suggests a higher rate of precocial male maturation on the Avalon Peninsula relative to the head

of the bay and Burin Peninsula and that smolts from coastal rivers in Labrador are significantly older than those from Lake Melville, little is known about the life history and ecology of Atlantic Salmon at this spatial scale in these regions (Dalley et al., 1983; ICES, 2018). While my results suggest a strong correlation between the Ssa01/Ssa23 chromosomal polymorphism and temperature, future work should characterize other unmeasured regional differences that may covary with temperature such as water chemistry (i.e. pH) (Bourret et al., 2013; Bradbury et al., 2014), and pathogen or parasite diversity (Dionne et al., 2009). South Newfoundland has been found to have unusual phenotypic (i.e., precocial male parr) and genetic diversity (i.e. European mtDNA) with some of the first evidence of an adaptive chromosomal translocation in wild populations. As such, the impact of indirect genetic interactions or changing climate may be greater than elsewhere within the species range making preservation of these discrete and evolutionarily significant populations a key conservation concern (Verspoor et al., 2015).

4. 3 Implications for genetic monitoring

Genetic monitoring is a time-efficient, cost-effective, and often non-lethal method of inferring change in adaptive spatial structure and population size (Blower, 2020). The LD method of N_e estimation has been shown to reliably detect population bottlenecks and track changes in demography (Tallmon et al., 2010; Antao et al., 2011). However, the utility of genetic monitoring is reliant on power achieved through sample size and number of loci. Inadequate power can bias estimates of N_e resulting in incalculable (infinite) upper CIs (Waples & Do, 2010; Blower, 2020). Samples from natural populations are limited by accessibility and rarity, particularly for species of conservation concern. To achieve accurate N_e, sample sizes of >1% of the total population size

have been suggested (Marandel et al., 2019). While an increase in sample size may improve power and accuracy more than a proportional increase in the number of loci, technological advances have made increasing panel size using SNPs ten-fold easier than increasing sample size (England et al., 2010; Luikart et al., 2010; Antao et al., 2011). As such, large genomic panels, with orders of magnitude more loci than commonly used, have the potential to significantly improve power and accuracy of population genetic parameters, particularly when used to monitor species of conservation concern for which population sizes are small and rarity of samples can be problematic (Waples et al., 2016; Allendorf, 2017).

Large panels increased precision and decreased bias of $\widehat{N}_b$ which is expected to reduce noise and improve ability to detect trend in population size relative to small commonly used panels. Diminishing returns in power and accuracy between panels of 1000 and 10,000 SNPs when initial population size is small ($N_b < 100$) were observed; Marandel et al. (2019) reported improvements in accuracy plateaued at panel sizes of 200 – 500 bi-allelic loci. However, the distribution of false negatives (type II error; failure to detect declines) and false positives (type I error) was narrower with the 10,000 SNP panel which was also found to be surprisingly precise in both simulated and empirical datasets. It is therefore recommended that genetic monitoring of Atlantic Salmon rivers in Placentia Bay, Newfoundland, with mean $\widehat{N}_b < 100$ (range: 18 – 336), continue using a large (10,000 SNP) panel with a sample size of 30 (i.e. the number of 1+ individuals that could be consistently sampled across most rivers and years); the use of a large microsatellite panel is not recommended as a large sample size (S = 50) was required to achieve the same power. It is additionally recommended that sampling design continue using consecutive

cohorts of 1+ individuals enabling assessment of both N_b and, when analyzed jointly N_e, to provide a comprehensive overview of the region over time.

4. 5 Final thoughts

The on-going expansion of Atlantic Salmon aquaculture into Placentia Bay, Newfoundland will more than double current production in the region with 30,000 – 45,000 tonnes expected to be harvested annually from 11 sites located near the head of Placentia Bay (Grieg Newfoundland Salmon Ltd., 2016). The genomic baseline, both estimates of population size and adaptive spatial structure, presented in this thesis will serve as the pre-impact baseline against which future change will be assessed. Continued genetic monitoring should aim to i) quantify the degree of introgression between farm and wild individuals, ii) track position of the cline; and iii) quantify differences in magnitude of population decline across karyotypes assessing the contribution of precocial male parr which may buffer against inbreeding depression (Johnstone et al. 2013; Perrier et al., 2014). Use of a large genomic dataset for genetic monitoring is expected to enable detection of change in fine-scale adaptive structure and population declines, as small as 30% with reasonable accuracy, on a timescale relevant to population management and conservation.

REFERENCES

Alcala, N., & Vuilleumier, S. (2014). Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. Proceedings of the Royal Society B, **281**(1704), https://doi.org/10.1098/rspb.2014.1369.

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics, **22**(13), 1600– 1607.

Alexander, D. H., Novmbre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, **19**, 1655 – 1664.

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. Molecular Ecology, **26**(2), 420 – 430. https://doi.org/10.1111/mec.13948

Antao, T., Pérez-Figueroa, A., & Luikart, G. (2011). Early detection of population declines: high power of genetic monitoring using effective population size estimators. Evolutionary Applications, **4**(1), 144 – 154. https://doi.org/10.1111/j.1752-4571.2010.00150.x

Antao, T. (2014). AgeStructureNe, GitHub, Retrieved from: https://github.com/AASJournals/Tutorials/blob/master/Repositories/CitingRepositories.md

ANtao, T., Cosart, T., Trethewey, B., Waples, R.S., Ackerman, M.W., Luikart, G., & Hand, B.K. (2020). AgeStrucNb: Software for simulating and detecting changes in the effective number of breeders (N_b). *Journal of Heredity*, **111**(5), 491 – 497.

Aykanat, T., Lindqvist, M., Pritchard, V.L., & Primmer, C.R. (2016). From population genomics to conservation and management: a workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon Salmo salar. Journal of Fish Biology, **89**(6), 2658 – 2679.

Babbitt, C. C., Warner, L. R., Fedrigo, O., Wall, C. E., & Wray, G. A. (2010). Genomic signatures of diet-related shifts during human origins. Proceedings of the Royal Society B, **278**(1708), https://doi.org/10.1098.rspb/2010.2433.

Balanyà, J., Huey, R. B., Gilchrist, G. W., & Serra, L. (2009). The chromosomal polymorphism of Drosophila subobscura: a microevolutionary weapon to monitor global change. Heredity. **103**, 364 – 367.

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., … & Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature, **528**, 405– 408.

Beauregard, D., Enders, E., & Boisclair, D. (2013). Consequences of circadian fluctuations in water temperature on the standard metabolic rate of Atlantic salmon parr (Salmo salar). Canadian Journal of Fisheries and Aquatic Sciences, **70**(7), 1072 – 1081.

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**(8), 4472 – 4487. https://doi.org/10.1111/mec.13790

Bidau, C. J., Giménez, M. D., Palmer, C. L., & Searle, J. B. (2001). The effects of Robertsonian fusion on chiasma frequency and distribution in the house mouse (Mus musculus domesticus) from a hybrid zone in northern Scotland. Heredity, **87**, 305 – 313.

Bidau, C. J., & Martí, D. A. (2002). Geographic distribution of Robertsonian fusions in Dichroplus pratensis, (Melanoplinae, Acrididae): the central-marginal hypothesis reanalysed. Cytogenetic and Genome Research, **96**, 66 – 74.

Bidau, C. J., Miño, C. I., Castillo, E. R., & Martí, D. A. (2012). Effects of abiotic factors on the geographic distribution of body size variation and chromosomal polymorphisms in two neotropical grasshopper species (Dichroplus: Melanoplinae: Acrididae). Psyche, https://doi.org/10.1155/2012/863947.

Blower., D. (2020). Estimating contemporary abundance, demography, and vulnerability to change for long-lived species with effective population size and population simulation. PhD Thesis, School of Biological Sciences, The University of Queensland. https://doi.org/10.14264/uql.2020.760.

Bourret, V., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape genomics in Atlantic salmon (Salmo salar) searching for gene-environment interactions driving local adaptation. Evolution, **67**(12), 3469 – 3487.

Bradbury, I. R., Hamilton, L. C., Robertson, M. J., Bourgeois, E., Mansour, A., & Dempson, B., (2014). Landscape structure and climatic variation determine Atlantic salmon genetic connectivity in the Northwest Atlantic. Canadian Journal of Fisheries and Aquatic Sciences, **71**(2), 246 – 258.

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. (2015). Transatlantic secondary contact in Atlantic salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. Molecular Ecology, **24**(20), 5130 – 5144.

Bradbury, I.R., Wringe, B.F., Watson, B., Paterson, I., Horne, J., & … Bentzen, P. (2018). Genotyping-by-sequencing of genome-wide microsatellite loci reveals fine-scale harvest composition in a coastal Atlantic salmon fishery. Evolutionary Applications, **11**(6), 918 – 930. https://doi.org/10.1111/eva.12606.

Brenna-Hansen, S., Li, J., Kent, M. P., Boulding, E. G., Dominik, S., Davidson, W. S. & Lien, S. (2012). Chromosomal differences between European and North American Atlantic salmon discovered by linkage mapping and supported by fluorescence in situ hybridization analysis. BMC Genomics, **13**, 432.

Burgerhout, E. (2017). Genetic background and embryonic temperature affect DNA methylation and expression of myogenin and muscle development in Atlantic salmon (Salmo salar). PLoS One, **12**(6): e0179918.

Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biological Journal of the Linnean Society, **124**(4), 1 – 23.

Castiglia, R., & Capanna, E. (2002). Chiasma repatterning across a chromosomal hybrid zone between chromosomal races of Mus musculus domesticus. Genetica, **114**, 35 – 40.

Cayuela, H., Dorant, Y., Merot, C., Laporte, M., Normandeau, E., Gagnon-Harvey, S., Sirois, P., & Bernatchez, L. (2020). Thermal adaptation rather than demographic history drives genetic structure inferred by copy number variants in a marine fish. BioRkiv, https://doi.org/10.1101/2020.04.05.026443.

Ceballos, G., Ehrlich, P.R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences, **114**(3), 6089 – 6096. https://doi.org/10.1073/pnas.1704949114

Chang, C. C., Carson, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, **4**(7), doi:10.1186/s13742-015-0047-8.

Chaput, G. (2012). Overview of the status of Atlantic salmon (Salmo salar) in the North Atlantic and trends in marine mortality. ICES Journal of Marine Science, **69**, 1538 – 1548.

Charlesworth, B. (1985). Recombination, genome size and chromosome number. In T. Cavalier-Smith (Ed.), The evolution of genome size (pp 489 – 513). John Wiley & Sons, Chichester.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. PLoS Genetics, **2**, 379 – 384.

Cheng, C-H., Ye, C-X., Guo, Z-X., & Wang, A-L. (2017). Immune and physiological responses of pufferfish (Takifugu obscurus) under cold stress. Fish & Shellfish Immunology, **64**, 137 – 145.

Corbett-Detig, R. B. (2016). Selection on inversion breakpoints favors proximity to pairing sensitive sites in Drosophila melanogaster, Genetics, **204**(1), 259 – 265.

COSEWIC. (2010). COSEWIC assessment and status report on the Atlantic Salmon Salmo salar (Nunavik population, Labrador population, Northeast Newfoundland population, South Newfoundland population, Southwest Newfoundland population, Northwest Newfoundland population, Quebec Eastern North Shore population, Quebec Western North Shore population, Anticosti Island population, Inner St. Lawrence population, Lake Ontario population, Gaspé-Southern Gulf of St. Lawrence population, Eastern Cape Breton population, Nova Scotia Southern Upland population, Inner Bay of Fundy population, Outer Bay of Fundy population) in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. xlvii + 136 pp.

Coyne, J. A., & Orr, H. A. (2004). Speciation. Sinauer, Sunderland, MA.

Dalley, E. L., Andrews, C. W., & Green, J. M. (1983). Precocious male Atlantic salmon parr (Salmo salar) in insular Newfoundland. Canadian Journal of Fisheries and Aquatic Science, **40**(5), 647 – 652.

Dempson, J.B., Schwarz, C.J., Reddin, D.G., O'Connell, M.F., Mullins, C.C., & Bourgeois, C.E. (2001). Estimation of marine exploitation rates on Atlantic salmon (Salmo salar L.) stocks in Newfoundland, Canada. ICES Journal of Marine Sciences, **58**(1), 331 – 341. https://doi.org/10.1006/jmsc.2000.1014.

DFO. 2020. Stock assessment of Newfoundland and Labrador Atlantic Salmon in 2018. DFO Can. Sci. Advis. Rep. 2020/038.

Dionne, M., Miller, K. M., Dodson, J. J., Caron, F., & Bernatchez, L. (2007). Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution, **61**(9), 2154 – 2164.

Dionne, M., Caron, F., Dodson, J. J., & Bernatchez, L. (2008). Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. Molecular Ecology, **17**(10), 2382 – 2396.

Dionne, M., Miller, K. M., Dodson, J. J., & Bernatchez, L. (2009). MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Philosophical Transactions of the Royal Society B, **364**(1523), https://doi.org/10.1098/rstb.2009.0011.

Do, C., Waples, R.S., Peel, G.M., Tillett, B.J., & Ovenden, J.R. (2014). NeESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. Molecular Ecology Resources, **14**(1), 209 – 214. https://doi.org/10.1111/1755-0988.12157.

Dobigny, G., Britton-Davidian, J., & Robinson, T. J. (2017). Chromosomal polymorphism in mammals: an evolutionary perspective. Biological Reviews, **92**(1), 1 – 21.

Dobzhansky, T. (1951). Genetics and the origin of species (3rd ed.) Columbia University Press.

Dumas, D., & Britton-Davidian, J. (2002). Chromosomal rearrangements and evolution of recombination: comparison of chiasma distribution patterns in standard and Robertsonian populations of the house mouse. Genetics, **162**(3), 1355 – 1366.

Ellegren, H., (2013). Genome sequencing and population genomics in non-model organisms. Trends in Ecology and Evolution, **29**(1), 51 – 63. https://doi.org/10.1016/j.tree.2013.09.008

England, P.R., Luikart, G., & Waples, R.S. (2010). Early detection of population fragmentation using linkage disequilibrium estimation of effective population size. Conservation Genetics, **11**, 2425 – 2430.

Feder, J. L., Nosil, P., & Flaxman, S. M. (2014). Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. Frontiers in Genetics, https://doi.org/10.3389/fgene.2014.00295.

Felsentein, J. (1976). The theoretical population genetics of variable selection and migration. Annual Review of Genetics, **10,** 253 – 280.

Ferchaud, A-L., Perrier, C., April, J., Hernandez, C., Dionne, M., & Bernatchez, L. (2016). Making sense of the relationships between N_e, N_b and N_c towards defining conservation thresholds in Atlantic salmon (Salmo salar). Heredity, **117**, 268 – 278. https://doi.org/10.1038/hdy.2016.62

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, **37**(12), 4302– 4315.

Fisheries and Ocean Canada (2007). Placentia Bay integrated management plan. Fisheries and Oceans Canada. Retrieved from the Fisheries and Marine Institute of Memorial University of Newfoundland website: https://www.mi.mun.ca/media/marineinstitutewwwmimunca/mi/programsandcourses/marinespatialplanning/files/PBIMCIntegratedManagementPlan.pdf

Fisheries and Ocean Canada (2009). Canada's policy for conservation of wild Atlantic salmon. Fisheries and Oceans Canada. Retrieved from the Department of Fisheries and Oceans Canada website: https://www.dfo-mpo.gc.ca/reports-rapports/regs/wasp-pss/wasp-psas-2009-eng.htm

Fordham, D.A., Brook, B.W., Moritz, C., Nogués-Bravo, D. (2014). Better forecasts of range dynamics using genetic data. Trends in Ecology and Evolution, **29**, 436–443. https://doi.org/10.1016/j.tree.2014.05.007

Forseth, T., Barlaup, B. T., Finstad, B., Fiske, P., Gjøsæter, H., Falkegård, M., & … Wennevik, V. (2017). The major threats to Atlantic salmon in Norway. ICES Journal of Marine Science, **74**(6), 1496 – 1513. https://doi.org/10.1093/icesjms/fsx020.

Frankham, R. (1995). Conservation genetics. Annual Review of Genetics, **29**, 305 – 327. https://doi.org/10.1146/annurev.ge.29.120195.001513

Frankham, R. (2010). Where are we in conservation genetics and where do we need to go? Conservation Genetics, **11**, 661 – 663.

Frankham, R., Bradshaw, C.J.A., & Brook, B.W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation, **170**, 56 – 63. https://doi.org/10.1016/j.biocon.2013.12.036

Friedland, K. D. (1998). Ocean climate influences on critical Atlantic salmon (Salmo salar) life history events. Canadian Journal of Fisheries and Aquatic Science, **55**(Supplement 1), 119 – 130.

Fuller, Z. L., Koury, S. A., Phadnis, N., & Schaeffer, S. W. (2019). How chromosomal rearrangements shape adaptation and speciation: case studies in Drosophila pseudoobscura and its sibling species Drosophila persimilis. Molecular Ecology, **28**(6), 1238 – 1301.

Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., & … Quinn, T. P. (2007). A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. Biological Reviews, **82**(2), 173 – 211.

Garner, B.A., Hand, B.K., Amish, S.J., Bernatchez, L., Foster, J.T., Miller, K.M., … & Luikart, G. (2015). Genomics in conservation: case studies and bridging the gap between data and application. Trends in Ecology and Evolution, **31**(2), 81 – 83. https://doi.org/10.1016/j.tree.2015.10.009

Glover, K.A., Solberg, M.F., McGinnity, P., Hindar, K., Verspoor, E., & … Svåsand, T. (2017). Half a century of genetic interaction between farmed and wild Atlantic salmon: status of knowledge and unanswered questions. Fish and Fisheries, **18**(5), 890 – 927. https://doi.org/10.1111/faf.12214.

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes, **5**(1), 184 – 186.

Grieg Newfoundland Salmon Ltd. (2016). Placentia Bay Atlantic salmon aquaculture project: proposed workforce and timeline. Retrieved from the Government of Newfoundland Municipal Affairs and Environment website: https://www.gov.nl.ca/mae/files/env-assessment-projects-y2016-1834-1834-oct25-workforce-plan.pdf

Guerrero, R. F., & Kirkpatrick, M. (2014). Local adaptation and the evolution of chromosomes fusions. Evolution, **68**(10), 2747 – 2756.

Hansen, G. J. A., Read, J. S., Hansen, J. F., & Winslow, L. A. (2016). Projected shifts in fish species dominance in Wisconsin lakes under climate change. Global Change Biology, **23**(4), 1463 – 1476.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R.S., … & Palstra, F. (2011). Understanding and estimating effective population size for practical application in marine species management. Conservation Biology, **25**(3), 438 – 449. https://doi-org.ezproxy.library.dal.ca/10.1111/j.1523-1739.2010.01637.x

Hartley, S. (1988). Cytogenetic studies of Atlantic salmon, Salmo salar L., in Scotland. Journal of Fish Biology, **33**, 735– 740.

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

Hill, W.G. (1981). Estimation of effective population size from data on linkage disequilibrium. Genetic Research, **38**(3), 209 – 216. https://doi.org/10.1017/S0016672300020553

Hoffmann, A.A., Sgrò, C.M., & Kristensen, T.N. (2017). Revisiting adaptive potential, population size and conservation. Trends in Ecology and Evolution, **32**(7), 506 – 517. https://doi.org/10.1016/j.tree.2017.03.012

Holleley, C.E., Nichols, R.A., Whitehead, M.R., Adamack, A.T., Gunn, M.R., & Sherwin, W.B. (2014). Testing single-sample estimators of effective population size in genetically structured populations. Conservation Genetics, **15**, 23 – 35. https://doi.org/10.1007/s10592-013-0518-3

Hollenbeck, C.M., Portnoy, D.S., & Gold, J.R. (2016). A method for detecting recent changes in contemporary effective population size from linkage disequilibrium at linked and unlinked loci. Heredity, **117**(4), 207-216. https://doi.org/10.1038/hdy.2016.30

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K., L., Rieseberg, L. H., (2019). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. BioRkiv, https://doi.org/10.1101/829622.

Hutchings, J.A., & Jones, M.E.B. (1998). Life history variation and growth rate thresholds for maturity in Atlantic salmon, Salmo salar. Canadian Journal of Fisheries and Aquatic Sciences, **55**(S1), 22 – 47. https://doi.org/10.1139.

Hutchings, J.A., & Reynolds, J.D. (2004). Marine fish population collapses: consequences for recovery and extinction risk. BioScience, **54**(4), 297 – 309. https://doi.org/10.1641/0006-3658(2004)054[0297:MFPCCF]2.0.C0;2.

ICES (Ed.) (2018). Report of the working group on North Atlantic salmon (WGNAS). ICES WGNAS Report 2018, ICES Advisory Committee, Woods, Hole, MA: ICES CM.

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara M. A., Nowell, R. W., Mallet, J., … & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. Current Biology, https://doi.org/10.1016/j.cub.2018.04.072.

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Llaurens, V., Parrinello, H., & Joron, M. (2019). Mutation accumulation in chromosomal inversions maintains wing pattern polymorphism in a butterfly. BioRxiv, https://doi.org/10.1101/736504.

Jeffery, N.W., Wringe, B.F., McBride, M.C., Hamilton, L.C., Stanley, R.R.E., & … Bradbury, I.R. (2018). Range-wide regional assignment of Atlantic salmon (Salmo salar) using genome wide single-nucleotide polymoprhisms. Fisheries Research, **206**, 163 – 175.

Johnstone, D.L., O'Connell, M.F., Palstra, F.P., & Ruzzante, D.E. (2013). Mature male parr contribution to the effective size of an anadromous Atlantic salmon (Salmo salar) population over 30 years. Molecular Ecology, **22**(9), 2394 – 2407. https://doi.org/10.1111/mec.12186.

Jones, A.T., Ovenden, J.R., & Wang, Y-G. (2016). Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. Heredity, **117**, 217-223. https://doi.org/10.1038/hdy.2016.19

Jonsson, B., & Jonsson, N. (2018). Egg incubation temperature affects the timing of the Atlantic salmon Salmo salar homing migration. Journal of Fish Biology, **93**(5), 1016 – 1020.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Research, **40**(1), 109– 114.

Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic evidence for adaptive inversion clines in Drosophila melanogaster. Molecular Biology and Evolution, **33**(5), 1317 – 1336.

Kennington, W. J., & Hoffmann, A. A. (2013). Patterns of genetic variation across inversions: geographic variation in the In(2L)t inversion in populations of Drosophilia melanogaster from eastern Australia. BMC Evolutionary Biology, **13**, 100.

Khan, R. A. (2009). Parasites causing disease in wild and cultured fish in Newfoundland. Icelandic Agricultural Sciences, **22**, 29 – 35.

King, M. (1993). Species evolution: the role of chromosomal change. Cambridge University Press, Cambridge, U. K.

King, T. L., Verspoor, E., Spidle, A. P., Gross, R., Phillips, R. B., Koljonen, M. L., & … Morrison, C. L. (2007). Biodiversity and population structure. In E. Verspoor, L. Stradmeyer, & J. Nielsen (Eds.), The Atlantic Salmon: Genetics, conservation and management (pp. 117– 166). Oxford, UK: Blackwell Publishing Ltd.

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. Genetics, **173**, 419 – 434.

Knowles, L. L., Carstens, B. C., & Keat, M. L. (2007). Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. Current Biology, **17**(11), 940–946.

Kuparinen, A., Hutchings, J.A., & Waples, R.S. (2016). Harvest-induced evolution and effective population size. Evolutionary Applications, **9**(5), 658 – 672. https://doi.org/10.1111/eva.12373

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Schwochow Thalmann, D., Hoeppner, M. P., & … Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nature Genetics, **48**, 84 – 88.

Larson, W. A., Lisi, P. J., Seeb, J. E., Seeb, L. W., & Schindler, D. E. (2016). Major histocompatibility complex diversity is positively associated with stream water temperatures in proximate populations of sockeye salmon. Journal of Evolutionary Biology, **29**(9), 1846 – 1859.

Lee, C-R., Wang, B., Mojica, J. P., Mandáková, T., Prasad, K. V. S. K., Goicoechea, J. L., … & Mitchell-Olds, T. (2017). Young inversion with multiple linked QTLs under selection in a hybrid zone. Nature Ecology and Evolution, **1**:0119. doi:10.1038/s41559-017-0019

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. EOS, **89**(10), 93 – 94.

Lehnert, S. J., Bentzen, P., Kess, T., Lien, S., Horne, J. B., Clement, M., & Bradbury, I. R. (2019a). Chromosome polymorphisms track trans-Atlantic divergence and secondary contact in Atlantic salmon. Molecular Ecology, **28**(8), 2074 – 2087.

Lehnert, S. J., Kess, T., Bentzen, P., Kent, M. P., Lien, S., Gilbey, J., …. Bradbury, I. R. (2019b). Genomic signatures and correlates of widespread population declines in salmon. Nature Communications. https://doi.org/10.1038/s41467-019-10972-w

Leroy, G., Carroll, E.L., Bruford, M.W., DeWoody, J.A., Strand, A., Waits, L., & Wang J. (2018). Next-generation metrics for monitoring genetic erosion within population of conservation concern. Evolutionary Applications, **11**(7), 1066 – 1083. https://doi.org/10.1111/eva.12564

Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., … & Davidson, W.S. (2016). The Atlantic salmon genome provides insights into rediploidization. Nature, **533**, 200 – 205. https://doi.org/10.1038/nature17164

Luikart, G., & Cornuet, J-M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology, **12**(1), 228 – 237.

Luikart, G., Ryman, N., Tallmon, D.A., Schwartz, M.K., & Allendorf, F.W. (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. Conservation Genetics, **11**(2), 355 – 372. https://doi.org/10.1007/s10592-010-0050-7

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

McAllister, B. F. (2003). Sequence differentiation associated with an inversion on the neo-X chromosome of Drosophila americana. Genetics, **165**(3), 1317 – 1328.

McCartney-Melstad, E., Vu, J.K., & Shaffer, H.B. (2018). Genomic data recover previously undetectable fragmentation effects in an endangered amphibian. Molecular Ecology, **27**(22), 4430 – 4443.

McDonald, M. J., Rice, D. P., & Desai, M. M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. Nature, **531**(7593), 233–236.

Mangel, M. (1994). Climate change and salmonid life history variation. Deep-Sea Research, **41**, 75 – 106.

Marandel, F., Lorance, P., Berthelé, O., Trenkel, V.M., Waples, R.S., & Lamy, J-B. (2019). Estimating effective population size of large marine populations, is it feasible? Fish and Fisheries, **20**(1), 189 – 198. https://doi.org/10.1111/faf.12338.

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. Trends in Ecology and Evolution, **34**(6), 531 – 544.

Mérot, C., Berdan, E. L., Babin, C., Normandeau, E., Wellenreuther, M., & Bernatchez, L. (2018). Intercontinental karyotype-environment parallelism supports a role for a chromosomal inversion in local adaptation in a seaweed fly. Proceedings of the Royal Society B, **285**(1881), https://doi.org/10.1098/rspb.2018.0519

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. Trends in Ecology & Evolution, (in press). https://doi.org/10.1016/j.tree.2020.03.002.

Metcalfe, N. B., & Thorpe, J. E. (1990). Determinants of geographical variation in the age of seaward-migrating salmon, Salmo salar. Journal of Animal Ecology, **59**(1), 135 – 145.

Minns, C. K., Randall, R. G., Chadwick, E. M. P., Moore, J. E., & Green, R. (1995). In Climate change and northern fish populations. Edited by R. J. Beamish, Canadian Special Publication of Fisheries and Aquatic Sciences, **121**, 699 – 708.

Moore, J-S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., & … Bernatchez, L. (2014). Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. Molecular Ecology, **23**(23), 5860 – 5697. https://doi.org/10.1111/mec.12972

Mueller, A-K., Chakarov, N., Krüger, O., & Hoffman, J.I. (2016). Long-term effective population size dynamics of an intensively monitored vertebrate population. Heredity, **117**, 290 – 299. https://doi.org/10.1038/hdy.2016.67

Nei, M. (1972). Genetic distance between populations. The American Naturalist, **106**, 283 – 292.

Nei, M. (1987). Molecular evolutionary genetics. Columbia University Press, New York.

Nicieza, A G., Reyes-Gavilán, F. G., & Braña, F. (1994). Differentiation in juvenile growth and bimodality patterns between northern and southern populations of Atlantic salmon (Salmo salar L.). Canadian Journal of Fisheries and Aquatic Sciences, **72**(9), 1603 – 1610.

Nielsen, J.L. (1998). Population genetics and the conservation and management of Atlantic salmon (Salmo salar). Canadian Journal of Fisheries and Aquatic Sciences, **55**(S1), 145 – 152.

Nilsson, J., Gross, R., Asplund, T., Dove, O., Jansson, H., Kellonieni, J. & … . Lumme, J. (2001). Matrilineal phylogeography of Atlantic salmon (Salmo salar L.) in Europe and postglacial colonization of the Baltic Sea area. Molecular Ecology, **10**(1): 89 – 92.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … & Wagner, H. (2017). vegan: Community ecology package. R package version 2.4-5. Retrieved from https://github.com/vegandevs/vegan

Oomen, R. A., Kuparinen, A., & Hutchings, J. A. (2020). Consequences of single-locus and tightly linked genomic architectures for evolutionary responses to environmental change. BioRxiv, https://doi.org/10.1101/2020.01.31.928770.

Ovenden, J.R., Leigh, G.M., Blower, D.C., Jones, A.T., Moore, A., Bustamante, C., … & Dudgeon, C.L. (2016). Can estimates of genetic effective population size contribute to fisheries stock assessments? Fish Biology, **89**(6), 2505 – 2518. https://doi.org/10.1111/jfb.13129

Palstra, F. P., O'Connell, M. F., & Ruzzante, D. E. (2007). Population structure and gene flow reversals in Atlantic salmon (Salmo salar) over contemporary and long-term temporal scales: effects of population size and life history. Molecular Ecology, **16**(21), 4504 – 4522.

Palstra, F.P. & Ruzzante, D.E. (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Molecular Ecology, **17**(15), 3428 – 3447. https://doi.org/10.1111/j.1365-294X.2008.03842.x

Palstra, F.P., O'Connell, M.F., & Ruzzante, D.E. (2009). Age structure, changing demography and effective population size in Atlantic salmon (Salmo salar). Genetics, **182**(4), 1233 – 1249. https://doi.org/10.1534/genetics.109.101972

Palstra, F.P., & Fraser, D.J. (2012). Effective/census population size ratio estimation: a compendium and appraisal. Ecology and Evolution, **2**(9), 2357 – 2365. https://doi.org/10.1002/ece3.329

Pante, E., & Simon-Bouhet, B. (2013). marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. PLoS One, **8**(9): e73051.

Peakall, R.O.D., & Smouse, P.E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, **6**(1), 288 – 295. https://doi.org/10.1111/j.1471-8286.2005.01155.x.

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, **13**(5), 946 – 952.

Peng, B., & Kimmel, M. (2005). simuPOP: a forward-time population genetics simulation environment. Bioinformatics, **21**(18), 3683 – 3687. https://doi.org/10.1093.bioinformatics/bti584.

Perrier, C., Normandeau, É., Dionne, M., Richard, A., & Bernatchez, L. (2014). Alternative reproductive tactics increase effective population size and decrease inbreeding in wild Atlantic salmon. Evolutionary Applications, **7**(9), 1094 – 1106. https://doi.org/10.1111/eva.12172.

Perrier, C., April, J., Cote, G., Bernatchez, L. & Dionne, M. (2016). Effective number of breeders in relation to census size as management tools for Atlantic salmon conservation in context of stocked populations. Conservation Genetics, **17**(1), 31 – 44. https://doi.org/10.1007/s10592-015-0758-5

Phillips, R., & Ráb, P. (2001). Chromosome evolution in the Salmonidae (Pisces): an update. Biological Reviews. **76**, 1 – 25.
Pierson, J.C., Luikart, G., & Schwartz, M.K. (2015). The application of genetic indicators in wild populations: Potentials and pitfalls for genetic monitoring. In: D. Lindenmayer, P. S. Barton & J. C. Pierson (Eds.), Indicators and surrogates of biodiversity and environmental change. Clayton, Vic.: CSIRO Publishing.

Poelstra, J. W., Richards, E. J., & Martin, C. H. (2018). Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. Molecular Ecology, **27**(21), 4270 – 4288.

Porter, T. R., Riche, L. G., & Traverse, G. R. (1974). Catalogue of rivers in insular Newfoundland, vol. A., Department of the environment, St. John's, NL.

Power, G. (1981). Stock characteristics and catches of Atlantic salmon (Salmo salar) in Quebec, and Newfoundland and Labrador in relation to environmental variables. Canadian Journal of Fisheries and Aquatic Sciences, **38**(12), 1601 – 1611.

Puerma, E., Orengo, D. J., & Aguadé, M. (2016). Multiple and diverse structural changes affect the breakpoint regions of polymorphic inversions across the Drosophila genus. Scientific Reports, **6**, 36248. https://doi.org/10.1038/srep36248

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, **81**, 559– 575.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics, **26**(6), 841–842

R Core Team. (2019). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rajagopal, V. M. (2020). ggman: R package for Manhattan plots, GitHub repository https://github.com/drveera/ggman/

Rambaut, A. (2012). FigTree (Version 1.4). Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Reddin, D.G. (1988). Ocean life of Atlantic salmon (Salmo salar, L.) in the Northwest Atlantic. In: Atlantic Salmon: Planning for the Future (ed. Milles, D.), pp 438 – 511. Croom Helm, Kent, UK.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. Trends in Ecology and Evolution, **16**(7), 351 – 358.

Rieseberg, L. H. (2009). Evolution: replacing genes and traits through hybridization. Current Biology, **19**(3), R119-R122.

Rougemont, Q., & Bernatchez, L. (2018). The demographic history of Atlantic salmon (Salmo salar) across its distribution range reconstructed from approximate Bayesian computations. Evolution, **72**(6), 1261 – 1277.

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources, **8**(1), 103 – 106. https://doi.org/10.1111/j.1471-8286.2007.01931.x.

Samy, J. K. A., Mulugeta, T. D., Nome, T., Sandve, S. R., Grammes, F., Kent, M. P., … & Våge, D. I. (2020). SalmoBase: an integrated molecular data resource for Salmonid species. BMC Genomics, **18**, 482. Doi:10.1186/s12864-017-3877-1.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics, **14**, 807 – 820.

Schulter, D. (2009). Evidence for ecological speciation and its alternative. Science, **323**(5915), 737 – 741.

Schwartz, M.K., Luikart, G., & Waples, R.S. (2007). Genetic monitoring as a promising tool for conservation and management. Trends in Ecology and Evolution, **22**(1), 25 – 33. https://doi.org/10.1016/j.tree.2006.08.009

Serebzov, D., Jorde, P.E., Bernatchez, L., Olsen, E.M., & Vøllestad, L.A. (2012). Short-term genetic changes: evaluating effective population size estimates in a comprehensively described brown trout (Salmo trutta) population. Genetics, **191**(2), 579 – 592. https://doi.org/10.1534/genetics.111.136580

Shaffer, M.K. (1981). Minimum viable population size for species conservation. Bioscience, **31**, 131 – 134. https://doi.org/10.2307/1308256

Shaw, J., & Potter, D.P. (2016). A glacial landform assemblage in Placentia Bay, Newfoundland, eastern Canada. Geological Society, London, Memoirs, **46**, 139 – 142. https://doi.org/10.1144/M46.141.

Sinclair-Waters, M., Bradbury, I. R., Morris, C. J., Lien, S., Kent, M. P. & Bentzen, P. (2018). Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic cod in the northwest Atlantic. Molecular Ecology, **27**(2), 339 – 351.

Slatkin, M. (1975). Gene flow and selection in a two-locus system. Genetics, **81**(4), 787 – 802.

Stabell, O.B. (1984). Homing and olfaction in salmonids: a critical review with special reference to Atlantic salmon. Biological Reviews, **59**, 333 – 388.

Stehfest, K. M., Carter, C. G., McAllister, J. D., Ross, J. D., & Semmens, J. M. (2017). Response of Atlantic salmon Salmo salar to temperature and dissolved oxygen extremes established using animal-borne environmental sensors. Scientific Reports, **7**, 4545, https://doi.org/10.1038/s41598-017-04806-2

Sterud, E., Forseth, T., Ugedal, O., Poppe, T. T., Jorgensen, A., Bruheim, T., … & Mo, T. A. (2007). Severe mortality in wild Atlantic salmon Salmo salar due to proliferative kidney disease (PKD) caused by Tetracapsuloides bryosalmonae (Myxozoa). Diseases of Aquatic Organisms, **77**, 191 – 198.

Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. (R package 2.10.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://github.com/jdstorey/qvalue

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., &… Waits, L. P. (2006). Putting the 'landscape' in landscape genetics. Heredity, **98**(3), 128 – 142.

Sturtevant, A. H. (1917). Genetic factors affecting the strength of linkage in Drosophilia. Proceedings of the National Academy of Science USA, **3**:555.

Sturtevant, A. H. (1938). The interrelations of inversions, heterosis and recombination. The American Naturalist, **72**:742, 447 - 452.

Sylvester, E.V.A., Beiko, R. G., Bentzen, P., Paterson, I., Horne, J. B., Watson, B., … & Bradbury, I.R. (2018a). Environmental extremes drive population structure at the northern range limit of Atlantic salmon in North America. Molecular Ecology, **27**(20), 4026 – 4040.

Sylvester, E.V.A., Wringe, B.F., Duffy, S.J., Hamilton, L.C., Fleming, I.A., & Bradbury, I.R. (2018b). Migration effort and wild population size influence the prevalence of hybridization between escaped farmed and wild Atlantic salmon. Aquaculture Environment Interactions, **10**, 401 - 411.

Sylvester, E.V.A., Wringe, B.F., Duffy, S.J., Hamilton, L.C., Fleming, I.A., Castellani, M., … & Bradbury, I.R. (2019). Estimating the relative fitness of escaped farmed salmon offspring in the wild and modelling the consequences of invasion for wild populations. Evolutionary Applications, **12**(4), 705 – 717.

Tallmon, D.A., Gregovich, D., Waples, R.S., Baker, C.S., Jackson, J., Taylor, B.L., … & Schwartz, M.K. (2010). When are genetic methods useful for estimating contemporary abundance and detecting population trends? Molecular Ecology Resources, **10**(4), 684 – 692. https://doi.org/10.1111/j.1755-0998.2010.02831.x

Taylor, E. B. (1991). A review of local adaptation in Salmonidae, with particular references to Pacific and Atlantic salmon. Aquaculture, **98**(1 – 3), 185 – 207.

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. Molecular Ecology, **25**(10), 2144 – 2164.

Turner, S. D. (2014). Qqman: An R package for visualizing GWAS results using QQ and manhattan plots. BioRxiv, 005165.

Valiente, A. G., Juanes, F., & Garcia-Vazquez, E. (2005). Reproductive strategies explain genetic diversity in Atlantic salmon, Salmo salar. Environmental Biology of Fishes, **74**, 323 – 334.

Verspoor, E. (2005). Genetic diversity among Atlantic salmon (Salmo salar L.) populations. ICES Journal of Marine Science, **54**, 965 – 973.

Verspoor, E., McGinnity, P., Bradbury, I., & Glebe, B. (2015). The potential direct and indirect genetic consequences for native Newfoundland Atlantic salmon from interbreeding with European-origin farm escapes. DFO Can. Sci. Advis. Sec. Res. Doc. 2015/030. vii + 36 p.

Vikeså, V., Nankervis, L., & Hevrøy, E. M. (2017). Appetite, metabolism and growth regulation in Atlantic salmon (Salmo salar L.) exposed to hypoxia at elevated seawater temperature. Aquaculture Research, **48**(8), 4086 – 4101.

Wang, J. (2016). A comparison of single-sample estimators of effective population sizes from genetic marker data. Molecular Ecology, **25**(19), 4692 – 4711. https://doi.org/10.1111/mec.13725

Wang, Y., Li, C., Pan, C., Liu, E., Zhao, X., & Ling, Q. (2019). Alterations to transcriptomic profile, histopathology, and oxidative stress in liver of pikeperch (Sander lucioperca) under heat stress. Fish & Shellfish Immunology, **95**, 659 – 669.

Waples, R.S. (2002). Effective size of fluctuating salmon populations. Genetics, **161**(2), 783 – 791.

Waples, R.S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evolutionary Applications, **3**(3), 244 - 262. https://doi.org.ezproxy.library.dal.ca/10.1111/j.1752-4571.2009.00104.x

Waples, R.S., Luikart, G., Faulkner, J.R., & Tallmon, D.A. (2013). Simple life-history traits explain key effective population size ratios across diverse taxa. Proceedings of the Royal Society B, **280**(1768). https://doi.org/10.1098/rspb.2013.1339.

Waples, R.S., Antao, T., & Luikart, G. (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics, **197**(2), 769 – 780. https://doi.org/10.1534/genetics.114.164822.

Waples, R.K., Larson, W.A., & Waples, R.S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity, **117**, 233 – 240. https://doi.org/10.1038/hdy.2016.60

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., … Venables, B. (2016)**.** gplots: Various R programming tools for plotting data (R package 3.0.1). Vienna, Austria: R Foundation for Statistical Computing.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38(6), 1358 – 1370.

Wellband, K., Mérot, C., Linnansaari, T., Elliott, J. A, K., Curry, A., Bernatchez, L. (2019). Chromosomal fusion and life history-associated genomic variation contribute to within-river local adaptation of Atlantic salmon. Molecular Ecology, **28**(6), 1439 – 1459.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. Trends in Ecology & Evolution, **33**(6), 427 – 440.

Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. Molecular Ecology, **28**(6), 1203 – 1209.

White, M. J. D. (1978). Modes of speciation. W. H. Freeman, San Francisco, CA.

Whiteley, A.R., Coombs, J.A., Cembrola, M., O'Donnell, M.J., Hudy, M., Nislow, K.H. & Letcher, B.H. (2015). Effective number of breeders provides a link between interannual variation in stream flow and individual reproductive contribution in a stream salmonid. Molecular Ecology, **24**(14), 3585 – 3602. https://doi.org/10.1111/mec.13273

Wright, S. (1931). Evolution in Mendelian populations. University of Chicago, Chicago, Illinois.

Wringe, B.F., Jeffery, N.W., Stanley, R.R.E., Hamilton, L.C., Anderson, E.C., & … Bradbury, I.R. (2018). Extensive hybridization following a large escape of domesticated Atlantic salmon in the Northwest Atlantic. Communication Biology, **1**, 108. https://doi.org/10.1038/s42003-018-0112-9.

Yada, T., & Tort, L. (2016). Stress and disease resistance: Immune system and immunoendocrine interactions. Fish Physiology, **35**, 365 – 403.

Yates, M. C., Debes, P. V., Fraser, D. J., Hutchings, J. A. (2015). The influence of hybridization with domesticated conspecifics on alternate reproductive phenotypes in male Atlantic salmon in multiple temperature regimes. Canadian Journal of Fisheries and Aquatic Sciences, **72**(8), 1138 – 1145.

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology, **16**(5), 284–287.

Zhan, L., Paterson, I.G., Fraser, B.A., Watson, B., Bradbury, I.R., Nadukkalam Ravindran, P., & Bentzen, P. (2017). Megasat: Automated inference of microsatellite genotypes from sequence data. Molecular Ecology Resources, **17**(2), 247 – 256. https://doi.org/10.1111/1755-0998.12561

Zhang, Y., Qin, C., Yang., L., Lu, R., Zhao, X., & Nie, G. (2018). A comparative genomics study of carbohydrate/glucose metabolic genes: from fish to mammals. BMC Genomics, **246**. https://doi.org/10.1186/s12864-018-4647-4.

Appendix A – Supplementary Tables: Chapter 2

Table S1 Simulation parameters for datasets generated with AGESTRUCN$_b$ (Antao et al., 2020). Ten independent replicates were performed for each scenario.

Marker type	**Loci** (L)	**Population size** (N)	**Sample size** (S)
Microsatellites	25, 100	100, 500, 1000	30, 50
SNPs	100, 1000, 10000	100, 500, 1000	30, 50

Table S2 Life history parameters of Atlantic Salmon (*Salmo salar*) of southern Newfoundland used in AGESTRUCN$_b$ (Antao et al., 2020) to simulate populations of varying size.

		Age						
		YoY	**1**	**2**	**3**	**4**	**5**	**6**
Survival		0.10	0.25	0.40	0.098	0.46	0.11	0.00
Proportion of Fecund Individuals	**Male**	0.00	0.59	0.68	0.80	0.50	0.75	1.00
	Female	0.00	0.00	0.00	0.50	0.75	1.00	1.00

Table S3 Mean detectable decline in $\widehat{N}_b$ between t_0 and t_x all simulated scenarios; estimates of N_b generated using LDN$_e$ v2 as implemented in AGESTRUCN$_b$ (Antao et al., 2020). Mean detectable decline, averaged across per generation rates of decline, at which power to detect decline surpasses 80% shown for scenarios varying in marker type (microsatellites and SNPs), panel size (L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs), population size (N = 100, 500 and 1000), and sample size (S = 30 or 50).

Sample size	**Population size**	**Microsatellites**		**SNPs**		
		25	**100**	**100**	**1000**	**10000**
30	100	46.7	39.3	54.1	29.5	28.4
	500	64.2	49.5	66.6	47.4	42.6
	1000	82.1	65.3	87.5	49.1	48.2
50	100	40.3	24.6	32.5	26.2	28.6
	500	57.8	39.6	65.3	33.9	32.9
	1000	76.8	44.5	73.9	48.8	44.7

Table S4 Median and range (minimum and maximum) offset, the difference between modeled and observed decline in $\widehat{N}_b$ between t_0 and t_x, for all simulated scenarios. Point estimates generated using LDNe v2 as implemented in AGESTRUCNb (Antao et al., 2020). Two marker types (microsatellites and SNPs) and 4 panel sizes (L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) compared across three population sizes (N = 100, 500 and 1000) and two sample sizes (S = 30 or 50).

Marker type	Panel size	Pop. size	S = 30			S = 50		
			Min.	Median	Max	Min.	Median	Max.
Micros	25	100	-756.1	-0.7	46.0	-164.8	-1.6	46.3
		500	-1314.6	-16.9	36.9	-422.1	-3.5	44.2
		1000	-906.9	-17.5	41.3	-895.0	-14.7	39.4
	100	100	-109.0	0.4	37.0	-48.9	3.1	34.1
		500	-403.4	0.8	44.6	-232.1	0.9	42.3
		1000	-594.2	-3.8	38.1	-330.1	0.9	43.7
SNPs	100	100	-684.2	0.1	49.0	-211.5	1.0	41.3
		500	-1460.5	-13.0	36.6	-744.2	-3.8	44.9
		1000	-919.9	-24.9	34.6	-733.0	-8.8	43.3
	1000	100	-100.9	-0.4	44.0	-84.2	1.1	28.6
		500	-422.1	1.0	44.9	-181.7	0.4	35.5
		1000	-412.9	-1.2	41.8	-206.5	0.6	41.8
	10000	100	-76.3	0.2	37.5	-56.9	0.5	27.1
		500	-320.1	0.2	47.9	-118.3	2.2	36.7
		1000	-419.3	1.6	45.6	-207.6	1.9	41.9

Appendix B – Supplementary Figures: Chapter 2

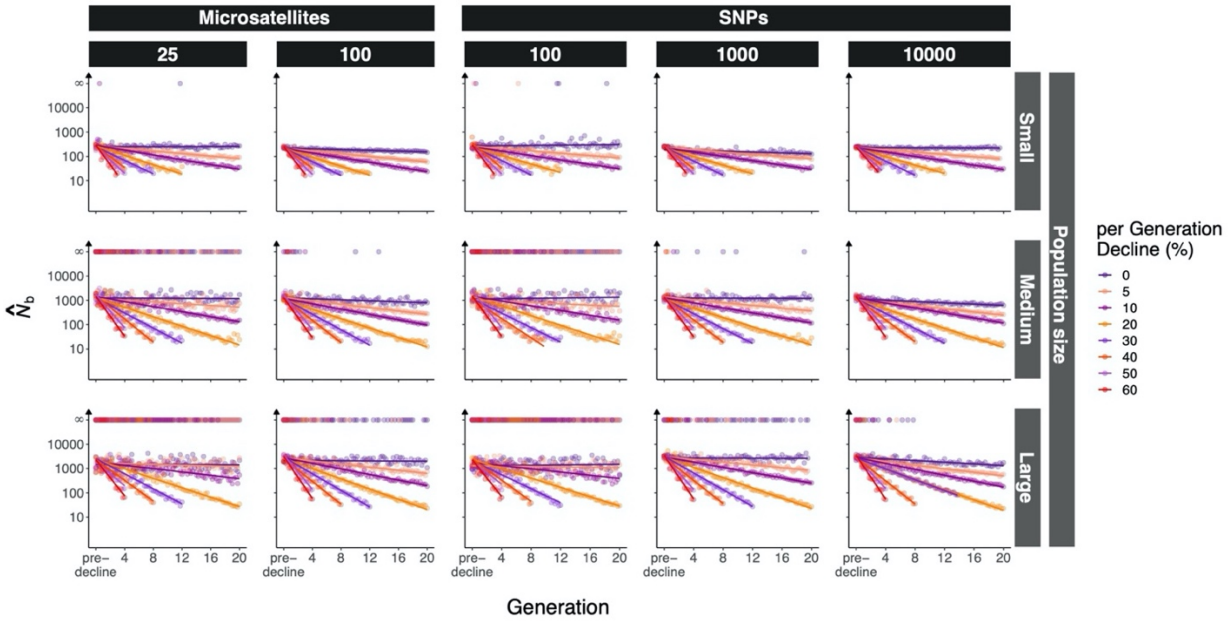

Figure S1 Linear regression of mean $\widehat{N}_b$ per scenario shown with infinite estimates, coloured by per generation rate of decline (%). $\widehat{N}_b$ generated using LDN_e v2 as implemented in $AGESTRUCN_b$ (Antao et al., 2020), S = 50. Mean $\widehat{N}_b$ calculated across ten replicates for each rate of decline and scenario. Two marker types (microsatellites and SNPs) and five panel sizes (columns: L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) were modelled using three initial population sizes (N = small (100), medium (500), or large (1000)). Eight per generation rates of decline (%) were simulated per scenario, indicated by colour. Monitoring began in the year pre-decline and continued up to 20 generations.

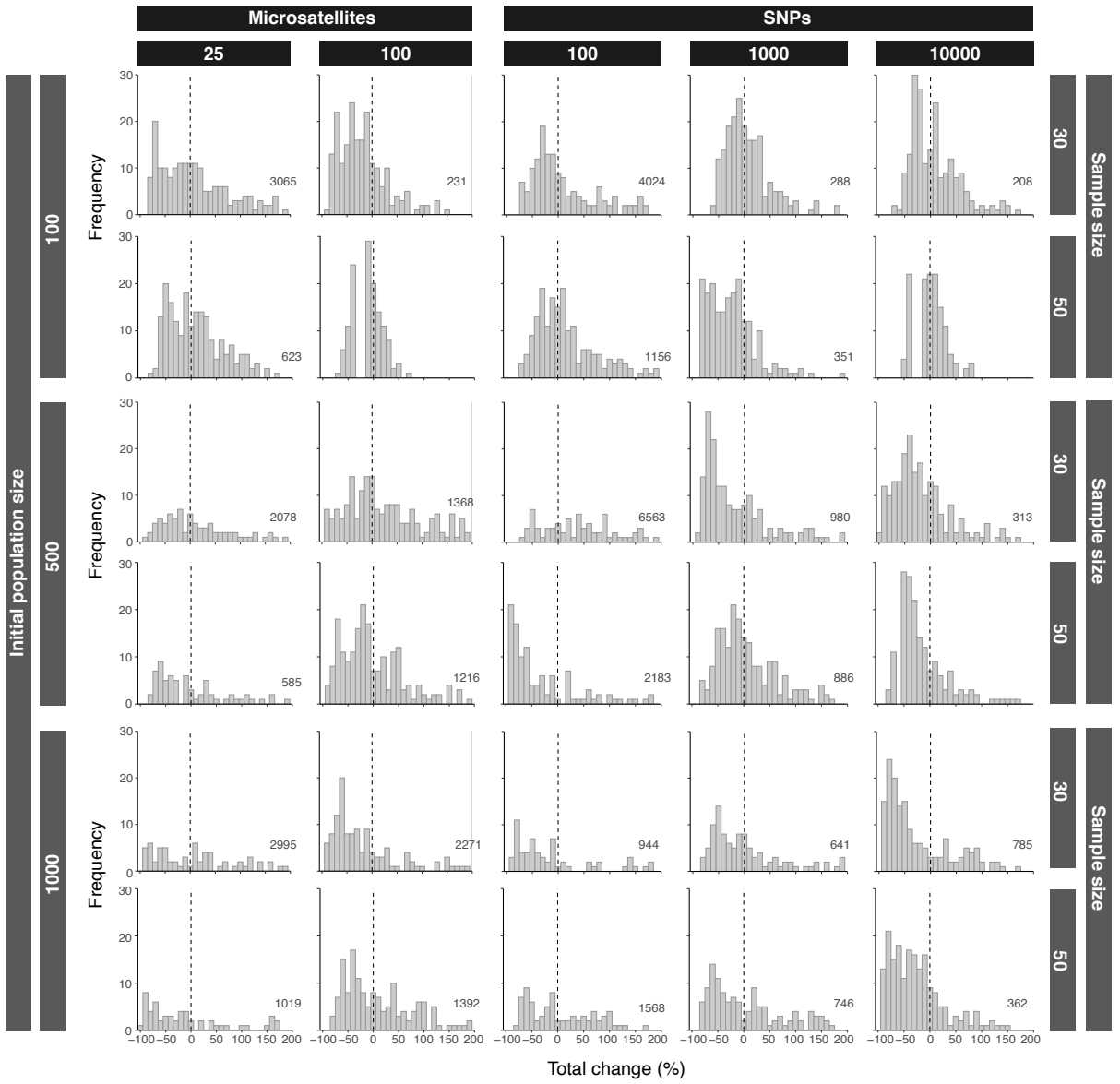

Figure S2 Distribution of total change (%) in $\widehat{N}_{\mathrm{b}}$ (point estimates) between t_0 and t_x for stable populations when sample size was small (S = 30) or large (S = 50). Positive values indicate an increase in population size ($\widehat{N}_{\mathrm{b1}} < \widehat{N}_{\mathrm{bx}}$) and negative values indicate a decrease in population size ($\widehat{N}_{\mathrm{b1}} > \widehat{N}_{\mathrm{bx}}$) was inferred. Dashed vertical lines represent the simulated total change (i.e. 0%). Values in the bottom right of each scenario are the greatest total change (%) observed; shown if beyond the limit of the x-axis. Columns: panel size (L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) and marker type. Rows: initial population size (N = 100, 500 or 1000). Estimates of N_{b} generated using LDNe v2 as implemented in AGESTRUCNb (Antao et al., 2020).

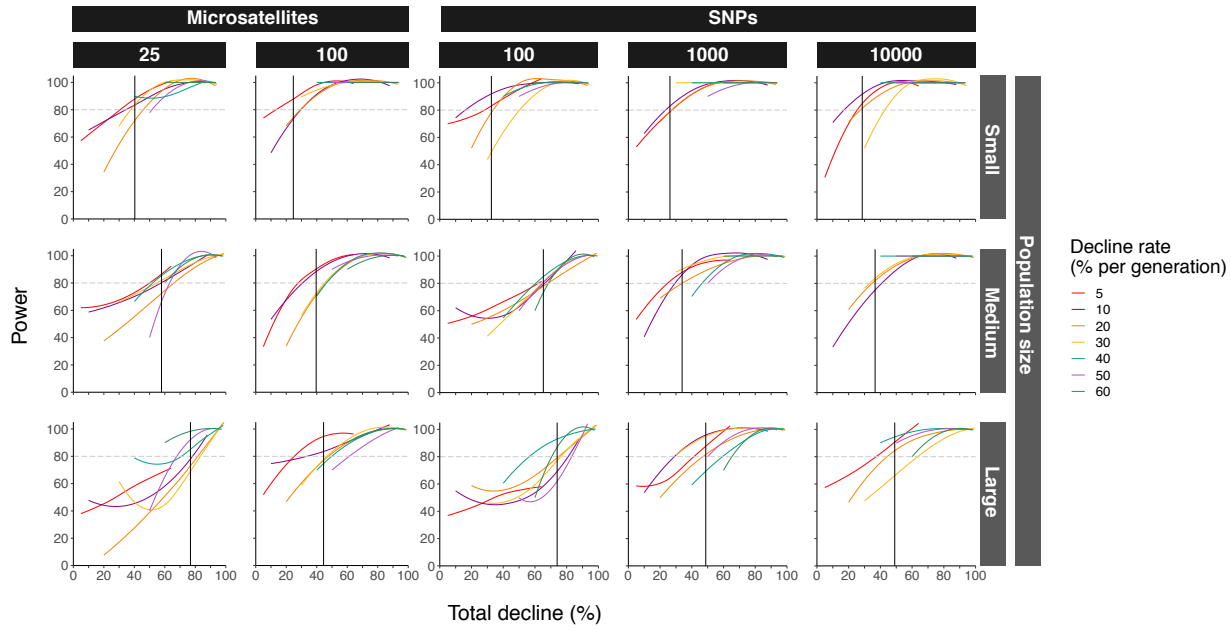

Figure S3 Power to detect a decline in $\widehat{N}_b$ between t_0 and t_x with S = 50 for all simulated scenarios. Point estimates generated using LDNe v2 as implemented in AGESTRUCNb (Antao et al., 2020). Two marker types (microsatellites and SNPs) and 4 panel sizes (columns: L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) compared across three population sizes (rows: N = small, medium and large). Coloured by per generation rates of decline (%). Loess curves are used for visualization of trends in the data. The vertical solid black line is the mean total decline (%) detectable for each scenario. The horizontal grey dashed line is the 80% power threshold.

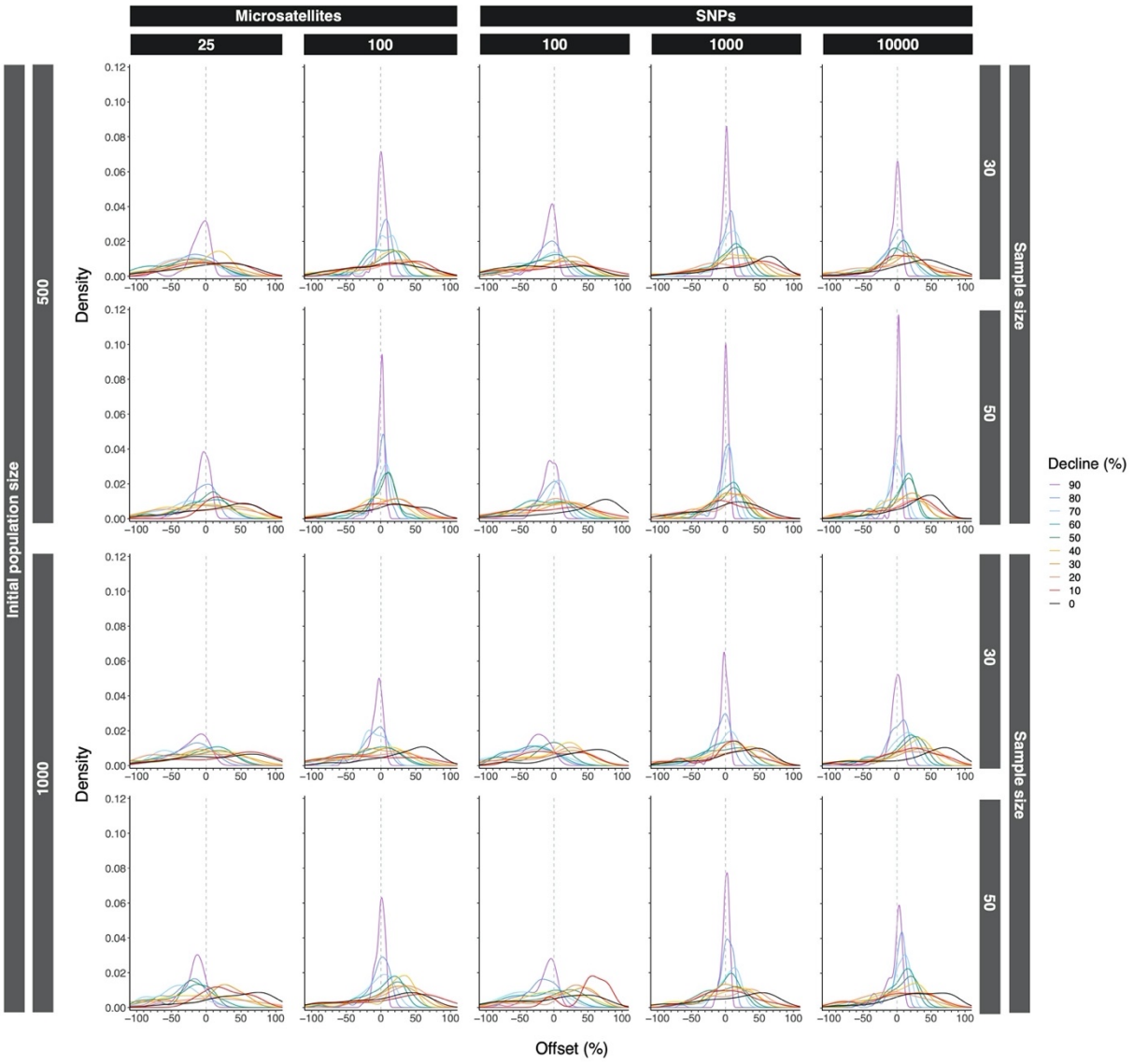

Figure S4 Accuracy of decline in $\widehat{N}_b$ between t_0 and t_x. Offset, the difference between modeled and observed decline in $\widehat{N}_b$, estimated using LDN_e v2 as implemented in $AGESTRUCN_b$ (Antao et al., 2020) shown across marker type and panel size (columns: L = 25 or 100 microsatellites, and 100, 1000 or 10,000 SNPs) for small (S = 30) and large (S = 50) sample sizes and moderate (N = 500) and large (N = 1000) populations. The vertical dashed line at 0% offset indicates the observed decline accurately approximates the simulated decline. Negative offset indicates the simulated decline was underestimated and positive offset that the simulated decline was overestimated. Variability in offset indicated by width of the scaled density curve. Plots have been constrained to offset within the range -500 to 150%. See Table 4 for mean minimum and maximum offset per scenario.

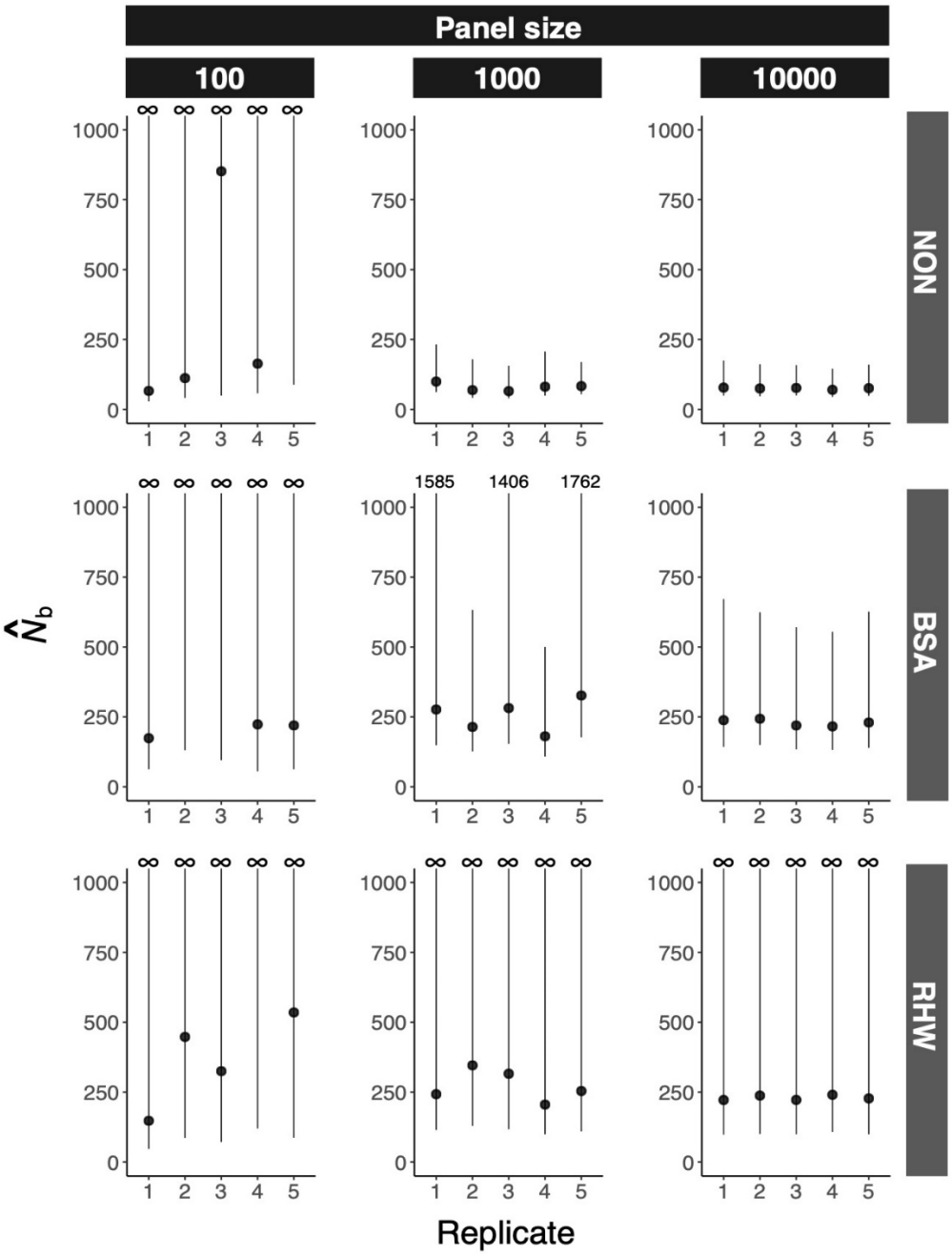

Figure S5 Bias and precision of empirical $\widehat{N}_b$ (point) and upper and lower bounds of CI (line) for three rivers (Rows: Big Salmonier River (BSA), Nonsuch River (NON), and Red Harbour West (RHW)) sampled in Placentia Bay, Newfoundland in 2017 (S = 25 – 30). Five replicate estimates of N_e generated using 100, 1000 or 10,000 SNP panels (columns) in NeESTIMATOR v2.1 (Do et al., 2014) with locus pairing across chromosomes. Rivers arranged according to inferred population size; NON (small, $\widehat{N}_b$ < 100), BSA (small-moderate, $\widehat{N}_b$ = 100 – 200), and RHW (moderate, $\widehat{N}_b$ > 200).

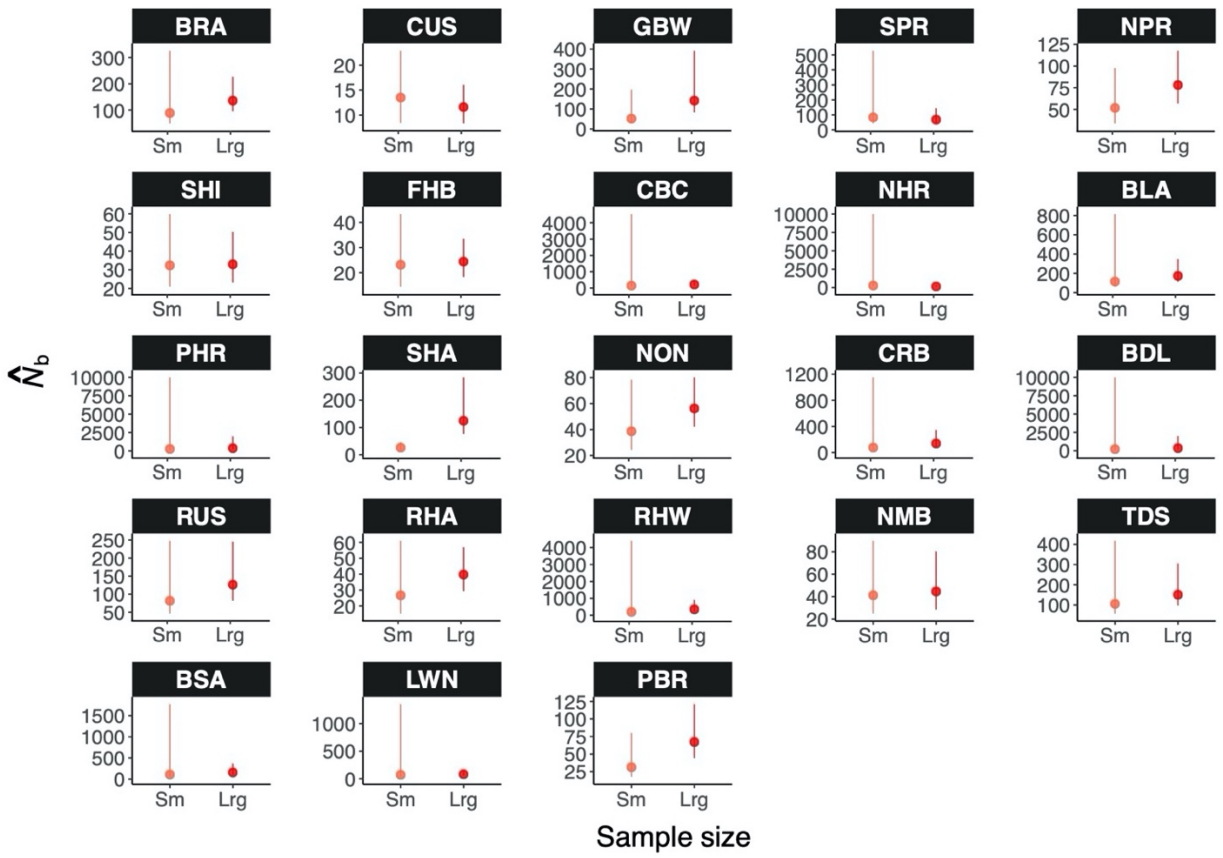

Figure S6 Comparison of empirical $\widehat{N}_b$ generated with a small sample size (orange: S = 20 – 30) or large sample size (red: S = 50 – 60). Estimates of $\widehat{N}_b$ generated using a large genomic panel (L = 101 microsatellites) and the linkage disequilibrium method in NeESTIMATOR v2.1 (Do et al., 2014). Individuals sampled from 23 rivers in Placentia Bay, Newfoundland, Canada in 2016.

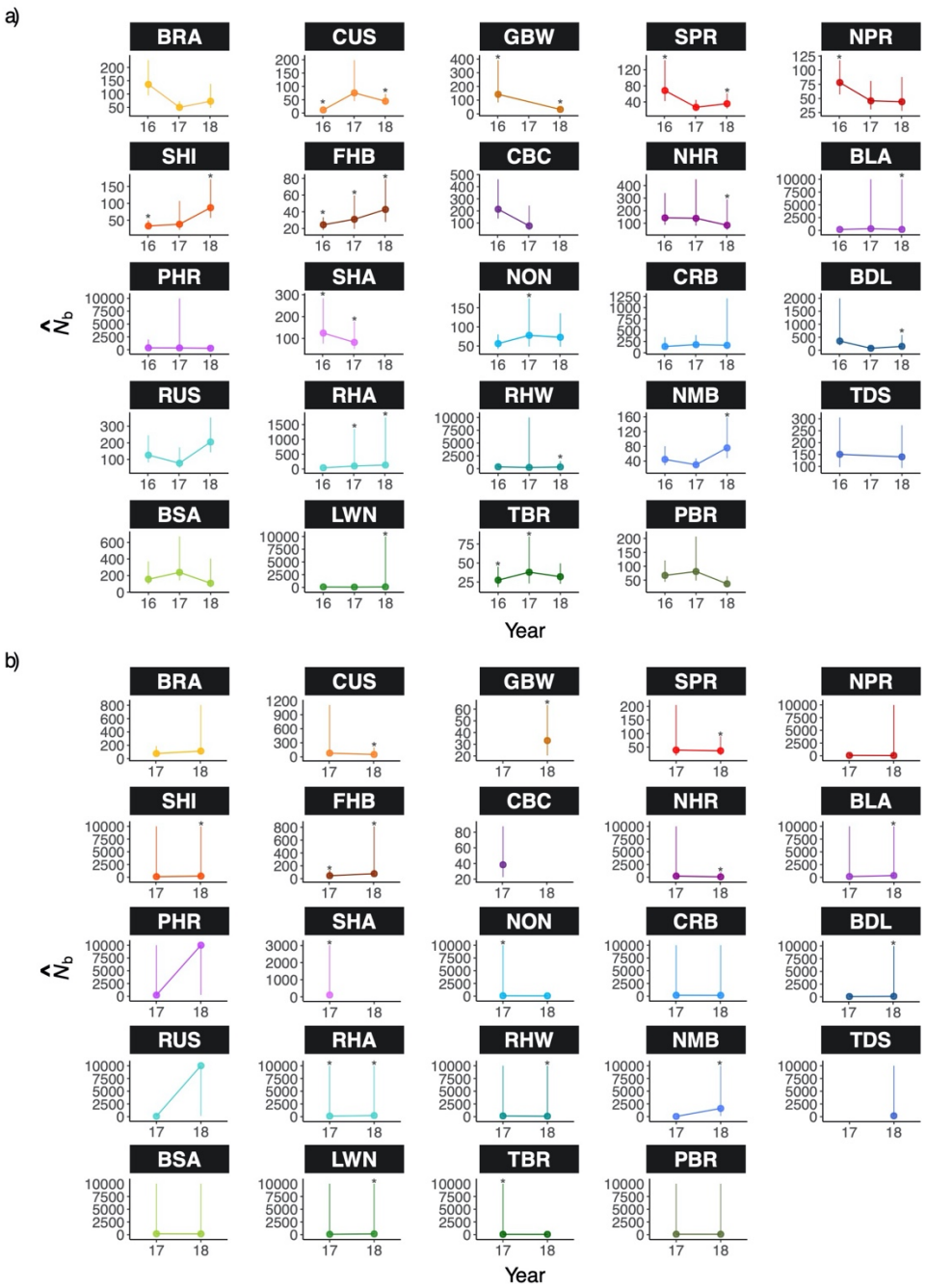

Figure S7 Trend in effective number of breeders over three discrete years in 24 rivers within Placentia Bay, Newfoundland, Canada. Empirical $\widehat{N}_b$ calculated using the linkage disequilibrium (LD) method in NeESTIMATOR v2.1 (Do et al., 2014) and large panels of 101 microsatellites with a sample size of 50 – 60 in 2016, and 10,000 single nucleotide polymorphisms (SNPs) with a sample size of 20 – 30 in 2017 and 2018. When ≤ 90% individuals sampled were from the same age class, estimates were considered N_e (*) rather than N_b.

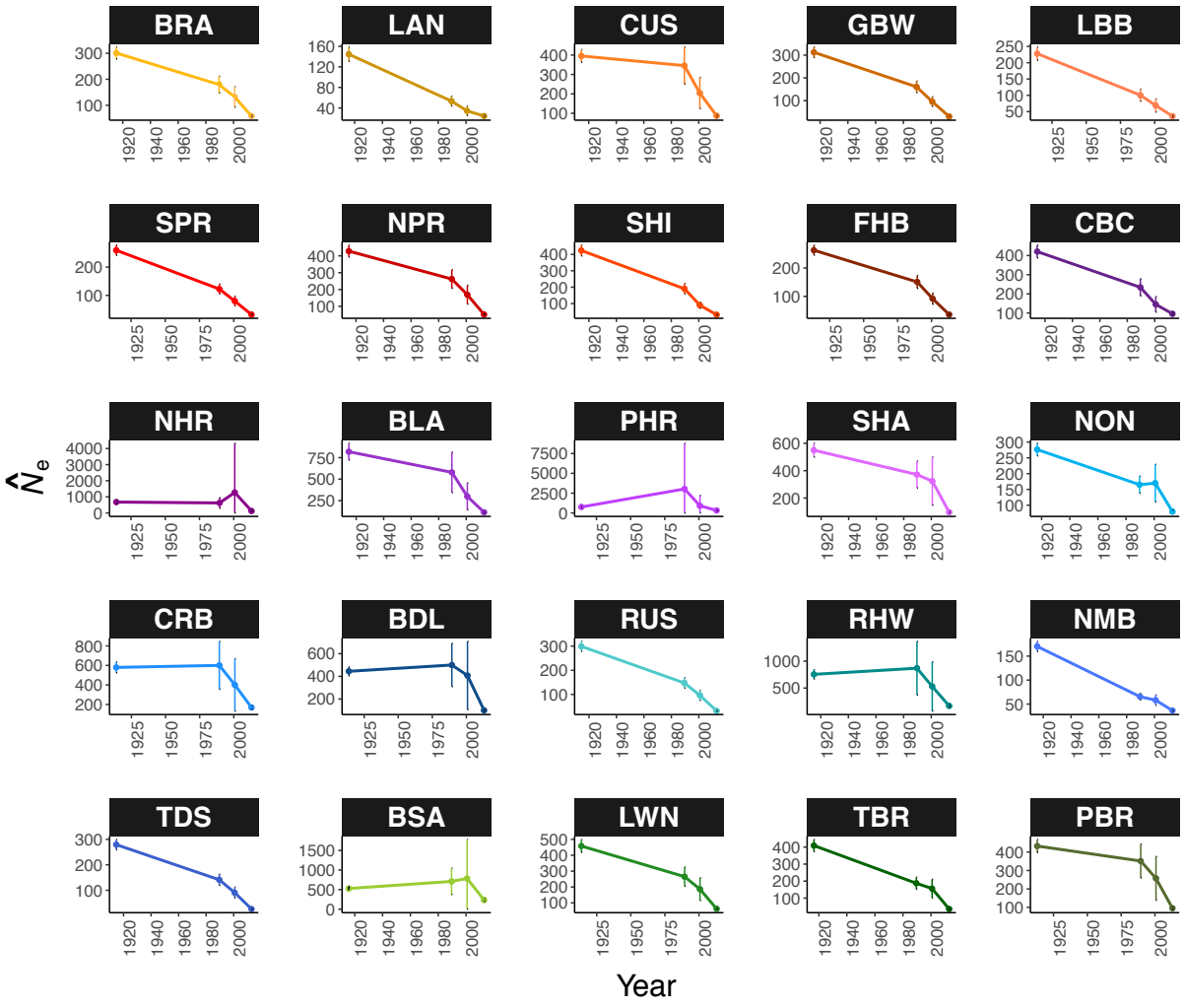

Figure S8 Recent effective population size ($\widehat{N}_e$) relative to maximum $\widehat{N}_e$ for Atlantic Salmon sampled from 25 rivers in Placentia Bay, Newfoundland, Canada in 2017 calculated using LINKN$_e$ (Hollenbeck et al., 2016).

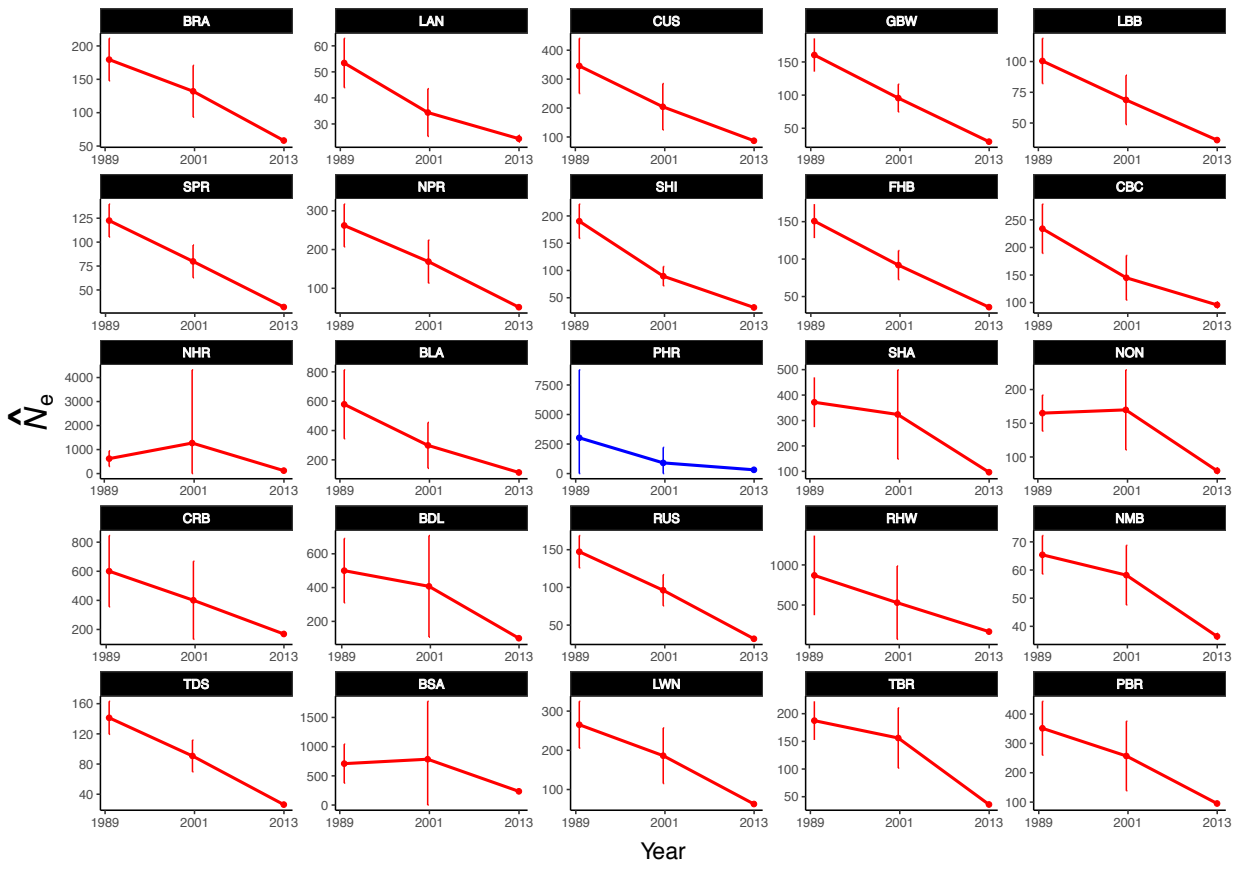

Figure S9 Change in effective population size ($\widehat{N}_e$), calculated using LINKN$_e$ (Hollenbeck et al., 2016), over the last ~ 6 generations for Atlantic Salmon sampled from 25 rivers in Placentia Bay, Newfoundland, Canada in 2017. Significant change (red), non-overlapping confidence intervals, and non-significant change (blue).

Appendix C – Supplementary Tables: Chapter 3

Table S5 Climatic and habitat variables used to identify potential drivers of translocation frequency using a partial redundancy analysis (RDA). Description of each measure is provided with the years measured, unit, and data source.

Variable	Description	Unit	Year	Source
BIO1	Annual mean temperature	∘C	1970-2000	WorldClim
BIO2	Mean diurnal range	∘C	1970-2000	WorldClim
BIO3	Isothermality	∘C	1970-2000	WorldClim
BIO4	Temperature seasonality	∘C	1970-2000	WorldClim
BIO5	Max temperature of warmest month	∘C	1970-2000	WorldClim
BIO6	Min temperature of coldest month	∘C	1970-2000	WorldClim
BIO7	Temperature annual range	∘C	1970-2000	WorldClim
BIO8	Mean temperature wettest quarter	∘C	1970-2000	WorldClim
BIO9	Mean temperature of driest quarter	∘C	1970-2000	WorldClim
BIO10	Mean temperature warmest quarter	∘C	1970-2000	WorldClim
BIO11	Mean temperature coldest quarter	∘C	1970-2000	WorldClim
BIO12	Annual precipitation	mm	1970-2000	WorldClim
BIO13	Precipitation of wettest month	mm	1970-2000	WorldClim
BIO14	Precipitation of driest month	mm	1970-2000	WorldClim
BIO15	Precipitation seasonality	mm	1970-2000	WorldClim
BIO16	Precipitation of wettest quarter	mm	1970-2000	WorldClim
BIO17	Precipitation of driest quarter	mm	1970-2000	WorldClim
BIO18	Precipitation of warmest quarter	mm	1970-2000	WorldClim
BIO19	Precipitation of coldest quarter	mm	1970-2000	WorldClim
Human density	Human population density		2000	NASA NEO
Axial length	Length of river along down-valley axis	km	1985	Inventory
Obstructions	Number of obstructions		1985	Inventory
Relief	Difference in elevation between the highest and lowest point of the basin	m	1985	Inventory
FW_dem_max	Maximum elevation	m	2000	HydroSHEDS

Table S6 Outlier block regions associated with a known chromosomal translocation (Ssa01p/Ssa23) in Atlantic Salmon identified here using pcadapt across rivers in Placentia Bay, Newfoundland, Canada. Outlier block regions were approximated by visual inspection of q-values. Boundaries of and number of significant (q-value < 0.05) single nucleotide polymorphisms (SNPs) in each region reported.

Chromosome	**Approximate boundaries of outlier blocks**		**Number of outlier SNPs**
	Start (bp)	**End (bp)**	
Ssa01	44,000,000	53,000,000	480
Ssa23	0	9,500,000	407

Table S7 Principal component, PC1 and PC2, loadings for each categorical PCA used in redundancy analyses (RDA) with code of climatic variable.

Climatic variable	Loading	
	PC1	PC2
Temperature		
BIO1	-0.19	-0.62
BIO2	0.34	-0.11
BIO3	0.20	-0.26
BIO4	0.36	-0.013
BIO5	0.34	-0.078
BIO6	-0.36	-0.036
BIO7	0.36	0.001
BIO8	-0.17	-0.54
BIO9	0.25	-0.23
BIO10	0.30	-0.38
BIO11	-0.35	-0.2
Precipitation		
BIO12	-0.39	0.048
BIO13	-0.38	0.18
BIO14	-0.37	-0.20
BIO15	-0.17	0.81
BIO16	-0.38	0.19
BIO17	-0.35	-0.42
BIO18	-0.37	-0.24
BIO19	-0.37	0.01

Table S8 Proportion of karyotypes for the Ssa01p/Ssa23 chromosomal translocation in Placentia Bay, Newfoundland, Canada. Karyotype assigned based on outlier SNPs (n = 887) from the outlier block regions on Ssa01p and Ssa23. Rivers are ordered geographically, east to west around the bay.

RiverID	2017			2018		
	Homo NA (Trans)	Hetero	Homo EU (No Trans)	Homo NA (Trans)	Hetero	Homo EU (No Trans)
BRA	0.000	0.233	0.767	0.000	0.250	0.750
CUS	0.667	0.267	0.067	0.267	0.533	0.200
GBW	0.111	0.556	0.333	0.073	0.341	0.585
SPR	0.000	0.111	0.889	0.000	0.133	0.867
NPR	0.138	0.517	0.345	0.118	0.294	0.588
SHI	0.045	0.409	0.545	0.000	0.438	0.563
FHB	0.533	0.400	0.067	0.320	0.600	0.080
CBC	0.533	0.400	0.067	0.421	0.526	0.053
NHR	0.667	0.292	0.042	0.310	0.552	0.138
BLA	0.542	0.458	0.000	0.655	0.276	0.069
PHR	0.310	0.655	0.034	0.567	0.367	0.067
SHA	0.667	0.333	0.000	0.889	0.111	0.000
NON	0.704	0.222	0.074	0.700	0.300	0.000
CRB	0.517	0.414	0.069	0.778	0.222	0.000
BDL	0.433	0.467	0.100	0.313	0.500	0.188
RUS	0.720	0.280	0.000	0.742	0.258	0.000
RHW	0.630	0.333	0.037	0.625	0.250	0.125
NMB	0.222	0.444	0.333	0.645	0.355	0.000
TDS	0.200	0.533	0.267	0.375	0.333	0.292
BSA	0.633	0.367	0.000	0.250	0.625	0.125
LWN	0.667	0.333	0.000	0.591	0.318	0.091
TBR	0.556	0.333	0.111	0.500	0.400	0.100
PBR	0.133	0.600	0.267	0.294	0.588	0.118

Table S9 Pairwise comparison of translocation and karyotype frequencies within river between years (2017 and 2018) sampled. Bold p-values indicate significantly different frequencies, calculated using a Fisher Exact Test. Rivers ordered geographically, east to west, around Placentia Bay, Newfoundland.

RiverID	Translocation frequency				Karyotype frequency			
	Estimate	lower CI	upper CI	p	Estimate	lower CI	upper CI	p
BRA	1.08	0.32	3.77	1.00	0.00	0.00	Inf	1.00
CUS	0.29	0.12	0.69	**0.00**	7.04	0.99	85.86	**0.04**
GBW	0.51	0.20	1.29	0.13	2.58	0.18	28.50	0.57
SPR	1.21	0.19	8.68	1.00	0.00	0.00	Inf	1.00
NPR	0.55	0.19	1.50	0.26	1.95	0.22	26.24	0.65
SHI	0.84	0.31	2.32	0.82	Inf	0.04	Inf	0.42
FHB	0.60	0.24	1.44	0.22	1.95	0.12	31.65	0.60
CBC	0.79	0.30	2.14	0.65	1.00	0.02	22.06	1.00
NHR	0.33	0.12	0.86	**0.02**	6.66	0.55	370.64	0.14
BLA	1.14	0.40	3.18	0.82	Inf	0.12	Inf	0.51
PHR	1.69	0.72	4.08	0.23	1.06	0.05	69.29	1.00
SHA	3.36	0.42	156.00	0.44	0.00	0.00	Inf	1.00
BDL	0.65	0.29	1.42	0.27	2.52	0.41	19.57	0.43
NON	1.28	0.38	4.75	0.78	0.00	0.00	8.01	0.51
CRB	3.02	1.01	10.31	**0.03**	0.00	0.00	4.25	0.19
RUS	1.10	0.31	3.78	1.00	0.00	0.00	Inf	1.00
RHW	0.77	0.27	2.16	0.64	3.29	0.24	188.57	0.60
NMB	5.70	2.32	14.89	**0.00**	0.00	0.00	0.22	**0.00**
TDS	1.35	0.49	3.73	0.64	0.60	0.06	4.89	0.67
BSA	0.29	0.11	0.74	**0.01**	Inf	0.99	Inf	**0.03**
LWN	0.60	0.20	1.81	0.33	Inf	0.23	Inf	0.20
TBR	0.90	0.23	3.68	1.00	1.00	0.01	23.98	1.00
PBR	1.86	0.73	4.80	0.20	0.22	0.01	2.11	0.17

Table S10 Gene ontology (GO) enrichment, using topGO, for single nucleotide polymorphisms (SNPs) identified as outliers in a genotype-environment analysis (redundancy analysis; RDA) and located within the outlier block regions on chromosomes Ssa01 and Ssa23. GO IDs in bold indicate the term was significant (p-value < 0.01) in both discrete years sampled.

GO ID	Description	2017				2018			
		N_{anno}	N_{sig}	N_{exp}	p-value	N_{anno}	N_{sig}	N_{exp}	p-value
GO:0045887	positive regulation of synaptic growth at neuromuscular junction	19	3	0.13	0.0003	19	3	0.12	0.00
GO:2000541	positive regulation of protein geranylgeranylation	6	2	0.040	0.0007	6	2	0.04	0.00
GO:0008380	RNA splicing	747	9	5.05	0.0016	747	9	4.8	0.0013
GO:0006729	tetrahydrobiopterin biosynthetic process	9	2	0.060	0.0016	9	2	0.06	0.0014
GO:0045075	regulation of interleukin-12 biosynthetic process	28	4	0.19	0.0002	28	3	0.18	0.0046
GO:0007346	regulation of mitotic cell cycle	1292	18	8.74	0.0028	1292	18	8.31	0.0023
GO:0031622	positive regulation of fever generation	12	2	0.080	0.0029	12	2	0.08	0.0026
GO:0031394	positive regulation of prostaglandin biosynthetic process	12	2	0.080	0.0029	12	2	0.08	0.0026
GO:0021912	regulation of transcription from RNA polymerase II promoter involved in spinal cord motor neuron fate specification	13	2	0.090	0.0034	13	2	0.08	0.0031
GO:0006979	response to oxidative stress	1072	13	7.25	0.0043	1072	13	6.89	0.0032
GO:0002091	negative regulation of receptor internalization	15	2	0.10	0.0045	15	2	0.1	0.0041
GO:1900025	negative regulation of substrate adhesion-dependent cell spreading	16	2	0.11	0.0051	16	2	0.1	0.0047

GO ID	Description	2017				2018			
		N_{anno}	N_{sig}	N_{exp}	p-value	N_{anno}	N_{sig}	N_{exp}	p-value
GO:0021913	regulation of transcription from RNA polymerase II promoter involved in ventral spinal cord interneuron specification	19	2	0.13	0.0072	13	2	0.08	0.0031
GO:0008152	metabolic process	19535	141	132.10	0.0022	19535	134	125.64	0.0084
GO:0014850	response to muscle activity	53	3	0.36	0.0056	53	3	0.34	0.0048
GO:0048515	spermatid differentiation	330	4	2.23	0.0007	330	3	2.12	0.0122
GO:0090076	relaxation of skeletal muscle	18	2	0.12	0.0065	18	2	0.12	0.0059
GO:0034612	response to tumor necrosis factor	321	7	2.17	0.0079	321	7	2.06	0.0069
GO:0071603	endothelial cell-cell adhesion	20	2	0.14	0.0080	20	2	0.13	0.0072
GO:0046326	positive regulation of glucose import	140	5	0.95	0.0027	140	4	0.9	0.0129
GO:0043201	response to leucine	20	2	0.14	0.0080	20	2	0.13	0.0072
GO:1900028	negative regulation of ruffle assembly	21	2	0.14	0.0088	21	2	0.14	0.008
GO:0010629	negative regulation of gene expression	4286	35	28.98	0.0104	4286	35	27.57	0.0079
GO:0032496	response to lipopolysaccharide	727	10	4.92	0.0053	727	9	4.68	0.0135
GO:0048535	lymph node development	65	3	0.44	0.0098	65	3	0.42	0.0085
GO:0071340	skeletal muscle acetylcholine-gated channel clustering	23	2	0.16	0.0105	23	2	0.15	0.0095
GO:0021918	regulation of transcription from RNA polymerase II promoter involved in somatic motor neuron fate commitment	7	1	0.050	0.0464	13	2	0.08	0.0031
GO:0048378	regulation of lateral mesodermal cell fate specification	12	2	0.080	0.0029	12	1	0.08	0.0745
GO:0003256	regulation of transcription from RNA polymerase II promoter involved in myocardial precursor cell differentiation	39	1	0.26	1.0000	13	2	0.08	0.0031

GO ID	Description	2017				2018			
		N_{anno}	N_{sig}	N_{exp}	p-value	N_{anno}	N_{sig}	N_{exp}	p-value
GO:1901213	regulation of transcription from RNA polymerase II promoter involved in heart development	62	1	0.42	1.0000	13	2	0.08	0.0031
GO:0000430	regulation of transcription from RNA polymerase II promoter by glucose	8	0	0.050	1.0000	13	2	0.08	0.0031
GO:0000431	regulation of transcription from RNA polymerase II promoter by galactose	8	0	0.050	1.0000	13	2	0.08	0.0031
GO:0010767	regulation of transcription from RNA polymerase II promoter in response to UV-induced DNA damage	5	0	0.030	1.0000	13	2	0.08	0.0031
GO:0021882	regulation of transcription from RNA polymerase II promoter involved in forebrain neuron fate commitment	9	0	0.060	1.0000	13	2	0.08	0.0031
GO:0021920	regulation of transcription from RNA polymerase II promoter involved in spinal cord association neuron specification	5	0	0.030	1.0000	13	2	0.08	0.0031
GO:0043618	regulation of transcription from RNA polymerase II promoter in response to stress	115	0	0.78	1.0000	13	2	0.08	0.0031
GO:0043619	regulation of transcription from RNA polymerase II promoter in response to oxidative stress	36	0	0.24	1.0000	13	2	0.08	0.0031
GO:0061418	regulation of transcription from RNA polymerase II promoter in response to hypoxia	74	0	0.50	1.0000	13	2	0.08	0.0031
GO:1900094	regulation of transcription from RNA polymerase II promoter involved in determination of left/right symmetry	38	0	0.26	1.0000	13	2	0.08	0.0031
GO:0046060	dATP metabolic process	19	2	0.13	0.0072	19	0	0.12	1

Appendix D – Supplementary Figures: Chapter 3

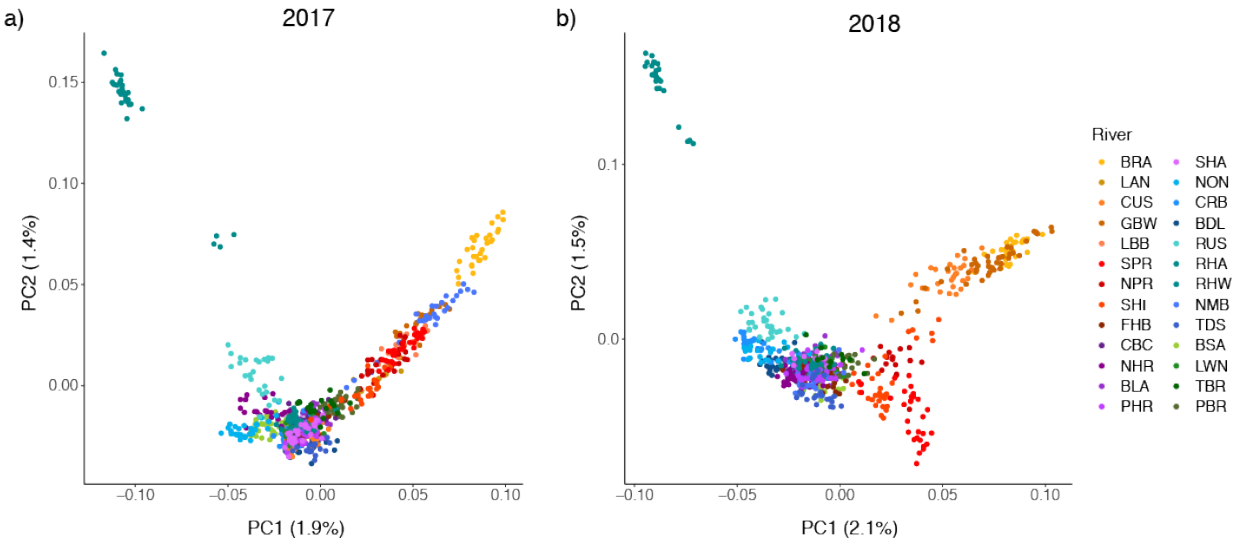

Figure S10 Population structure of Atlantic Salmon in Placentia Bay, Newfoundland in (a) 2017 and (b) 2018. Red Harbour East (RHA) was found to be genetically differentiated from all other rivers sampled based on the first two principal component (PC) axes from pcadapt (Luu et al., 2017) using 139,038 SNPs. Rivers coloured east (yellow-red) to west (green-blue) with head of the bay (purple).

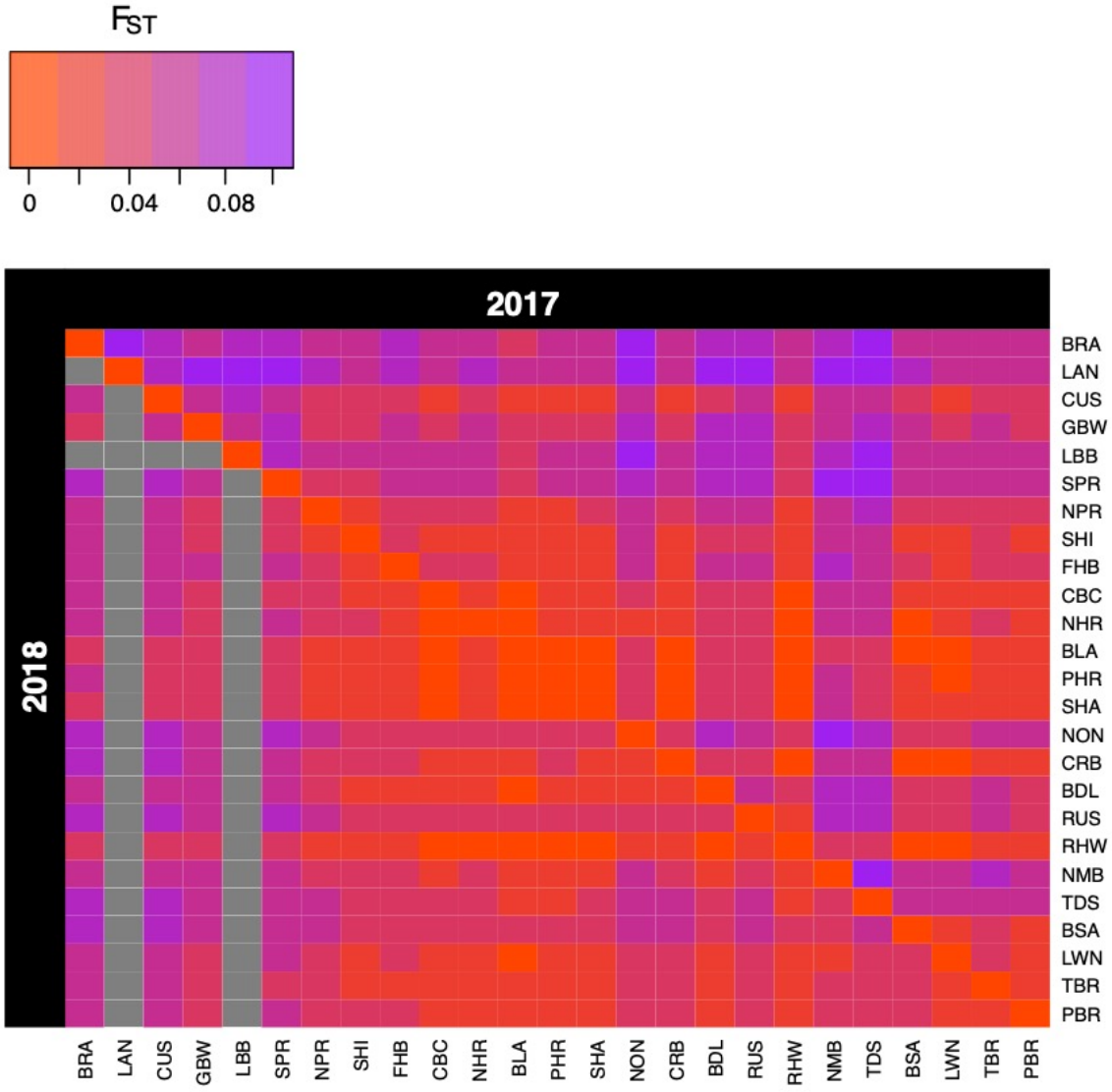

Figure S11 Heatmap of pairwise F_{ST} for Atlantic Salmon rivers (n = 25) in Placentia Bay, Newfoundland, Canada. The upper matrix represents 2017 and the lower matrix 2018. Rivers are ordered geographically from east to west. Lance (LAN) and Little Barasway (LBB) were not sampled in 2018.

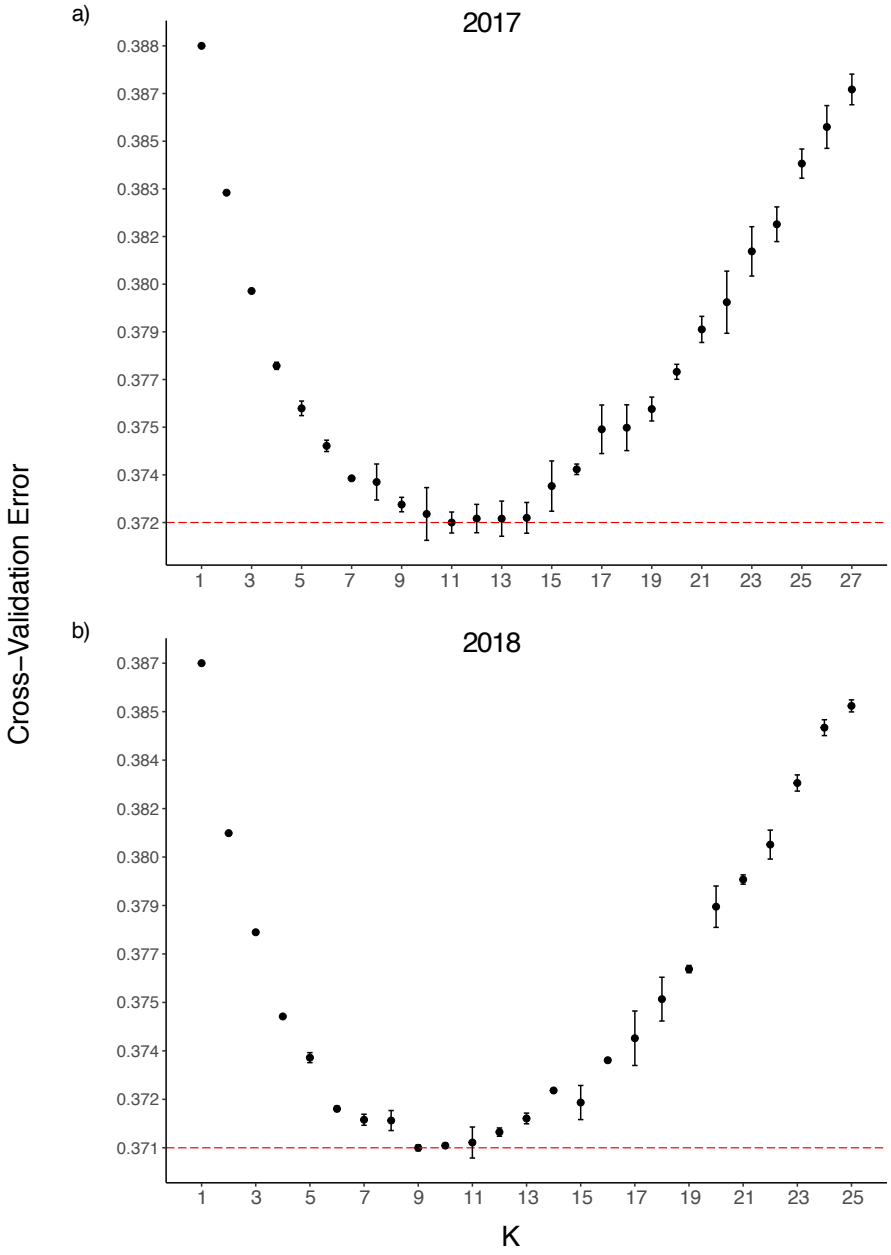

Figure S12 Cross-validation (CV) error means and standard deviations from three ADMIXTURE runs using different random number seeds. Standard deviations which overlap the dashed red line, indicating lowest mean CV error represent a reasonable range of K. (a) 2017, K = 10 through K = 14 and (b) 2018, K = 9 through K = 11.

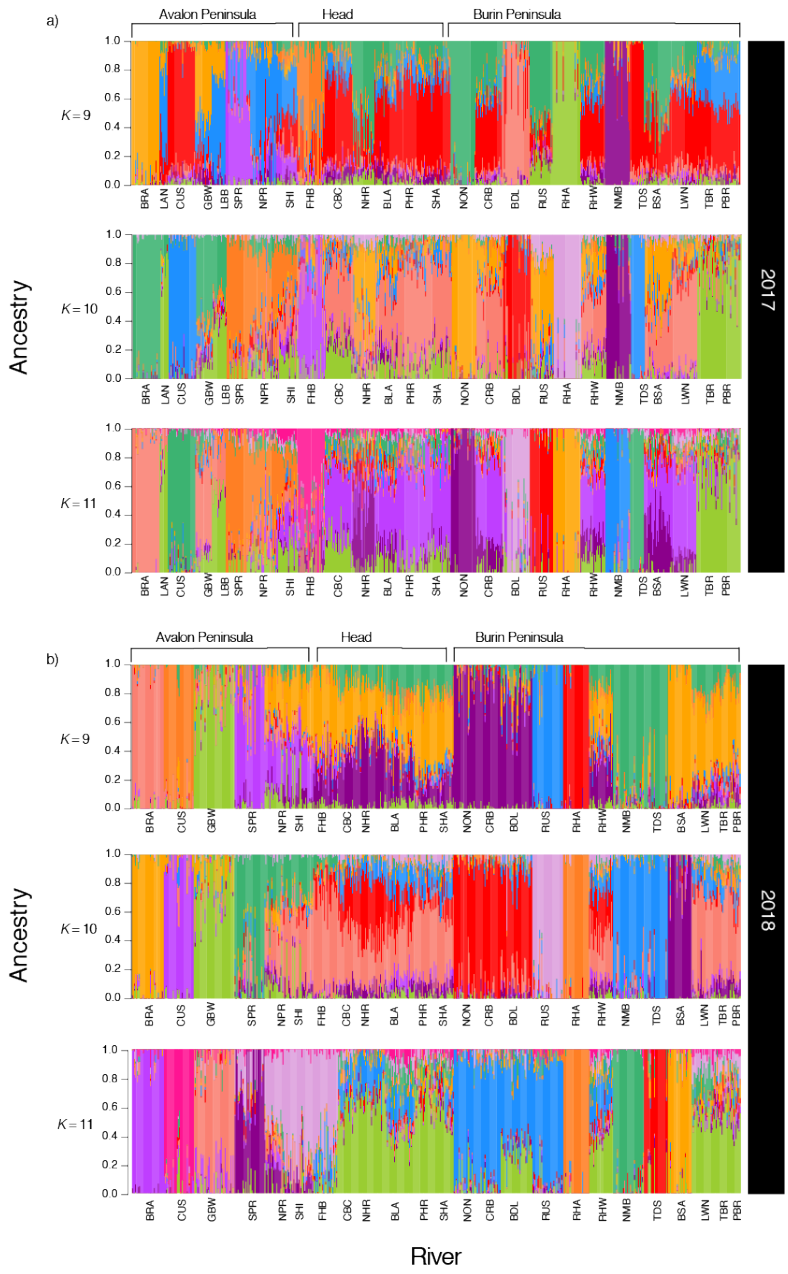

Figure S13 ADMIXTURE (K = 9 – 11) results for Atlantic Salmon (*Salmo salar*) sampled in Placentia Bay, Newfoundland, Canada in (a) 2017 (n = 662) and (b) 2018 (n = 611). Rivers ordered geographically, east to west, around the bay. Lance (LAN) and Little Barasway (LBB) were not sampled in 2018 due to limited sample size in 2017.

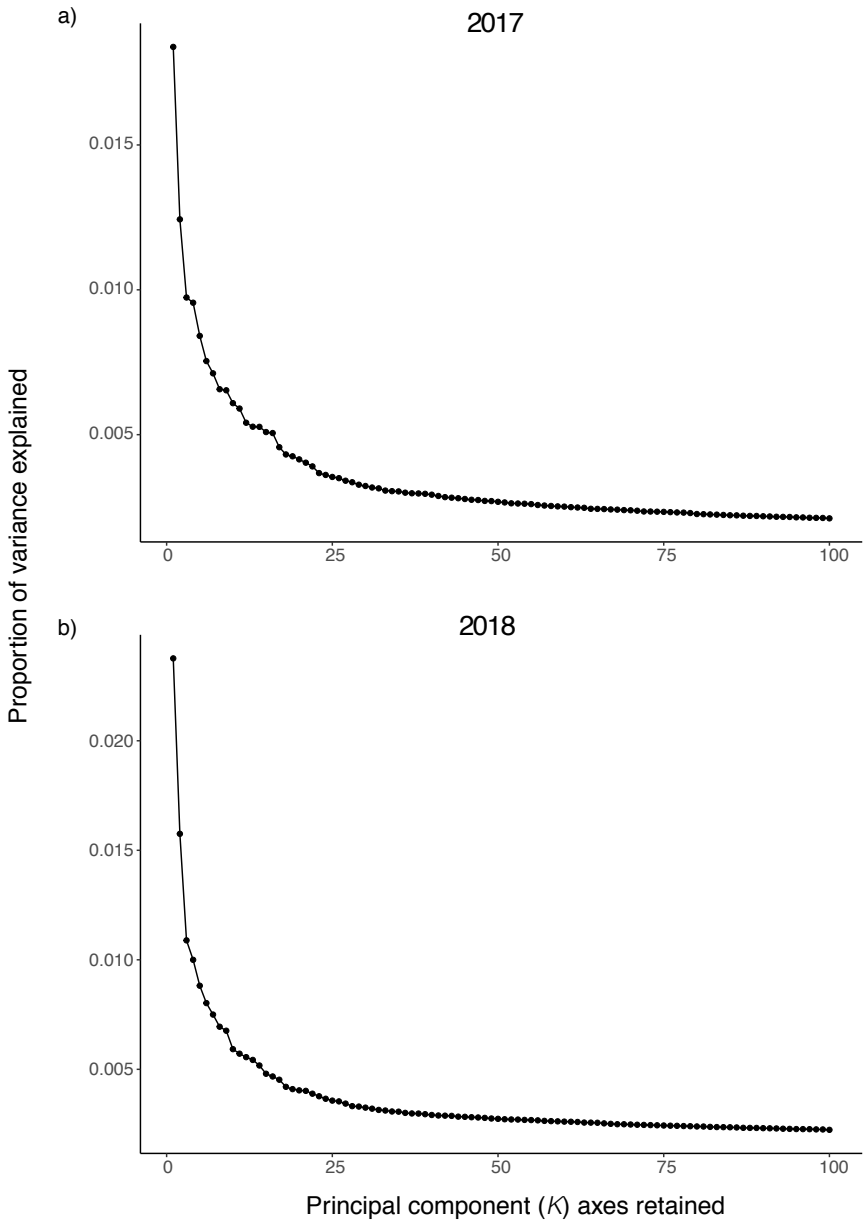

Figure S14 Scree plots generated in pcadapt showing proportion of variance explained as increasing principal component (K) axes are retained for a) 2017 and b) 2018. Results represent genetic variation in Atlantic Salmon (*Salmo salar*) across Placentia Bay, Newfoundland, Canada.

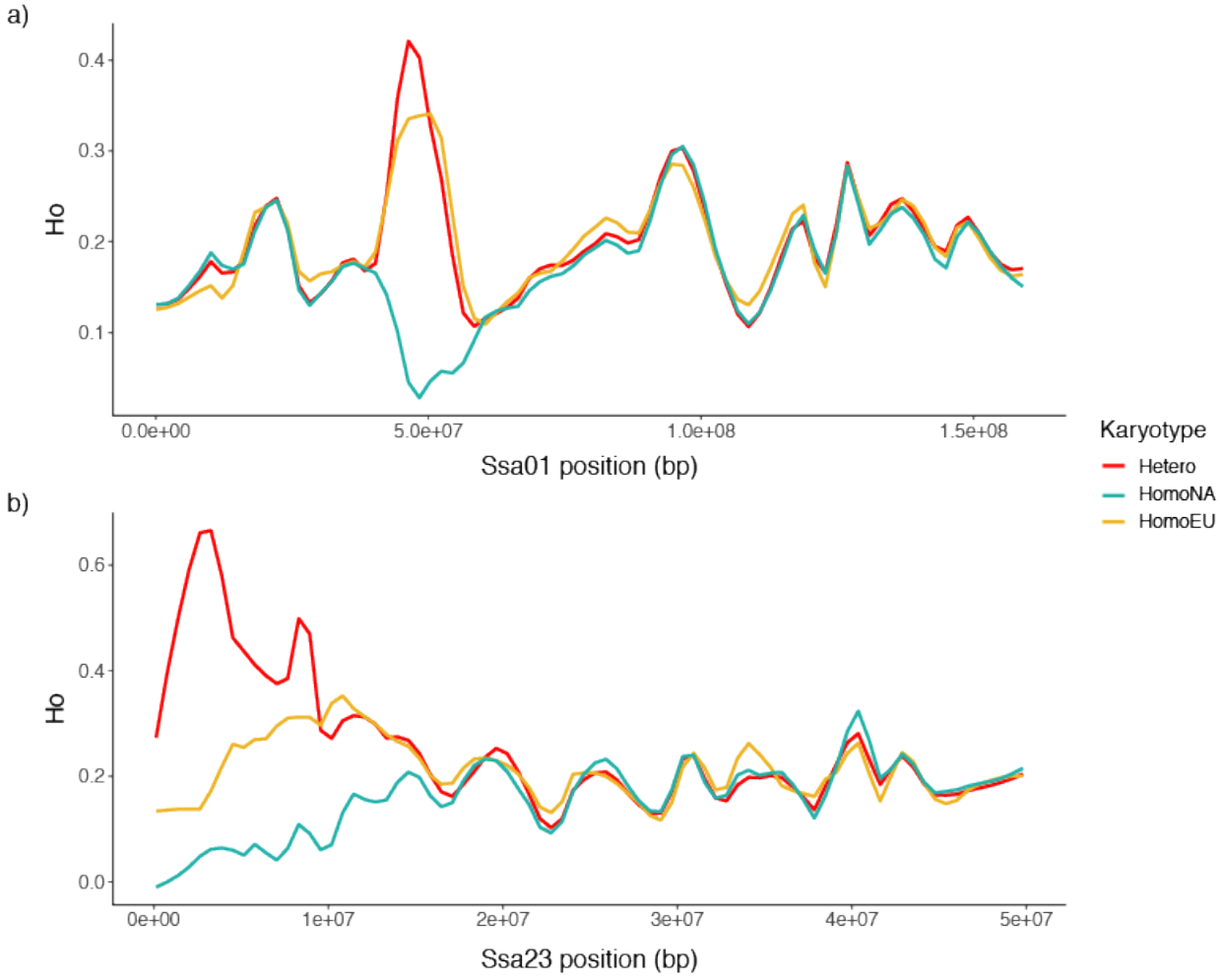

Figure S15 Observed heterozygosity (Ho) for each karyotype. Karyotypes assigned using kmeans for three clusters; standard European (homozygous non-translocated; Ssa01p/q and Ssa23), standard North American (homozygous translocated; Ssa01p/Ssa23 and Ssa01q), and heterozygous. All values calculated in plink v1.9. Lines represent smoothed values for a span of 0.1 using ggplot2.

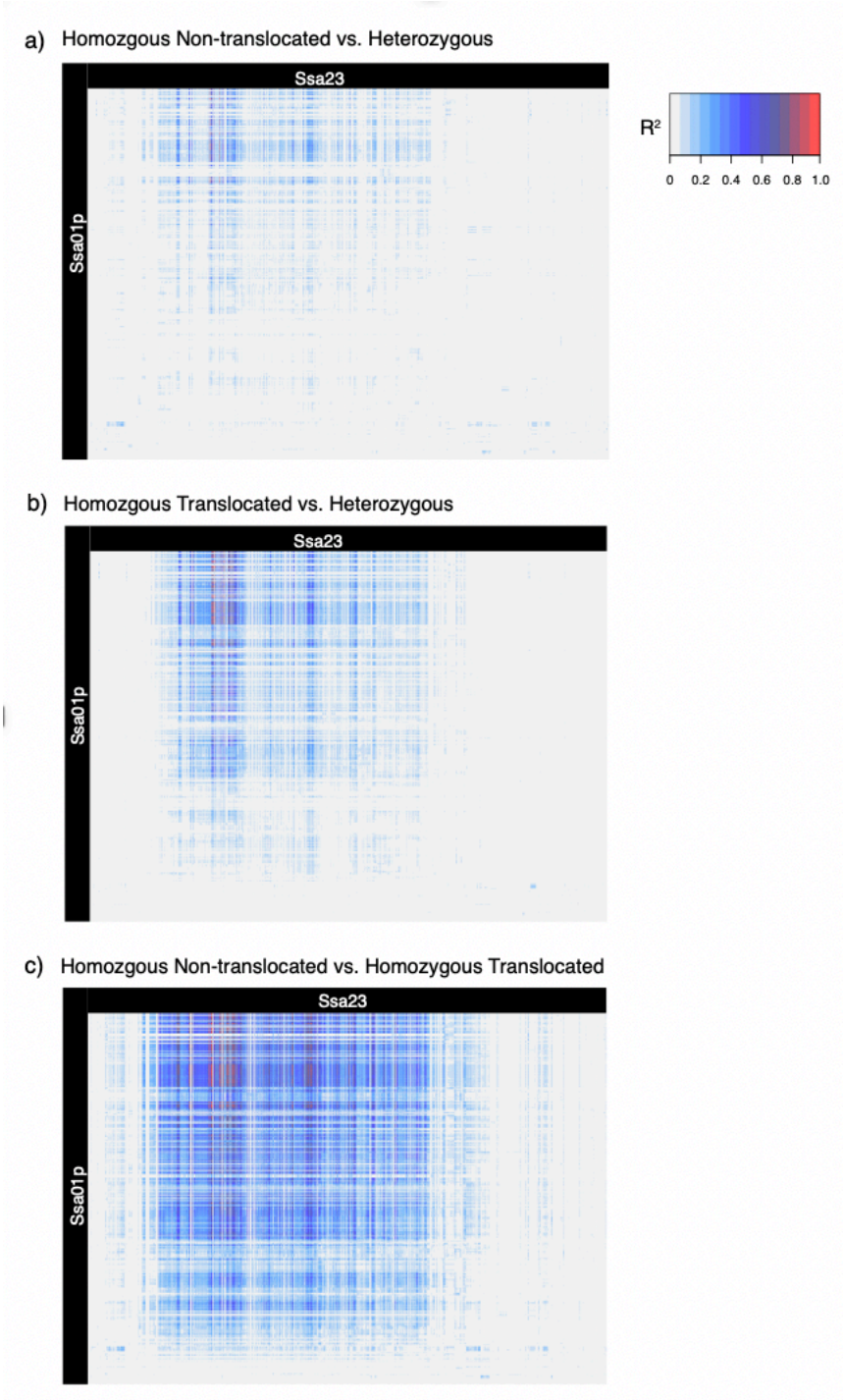

Figure S16 Heatmaps of linkage disequilibrium (LD) between outlier SNPs (n = 1913) within outlier block regions on chromosomes (Ssa01p and Ssa23). Pairwise LD (R^2) for (a) homozygous non-translocated vs. heterozygous, (b) homozygous translocated vs. heterozygous, and (c) homozygous non-translocated vs. homozygous translocated in 2017.

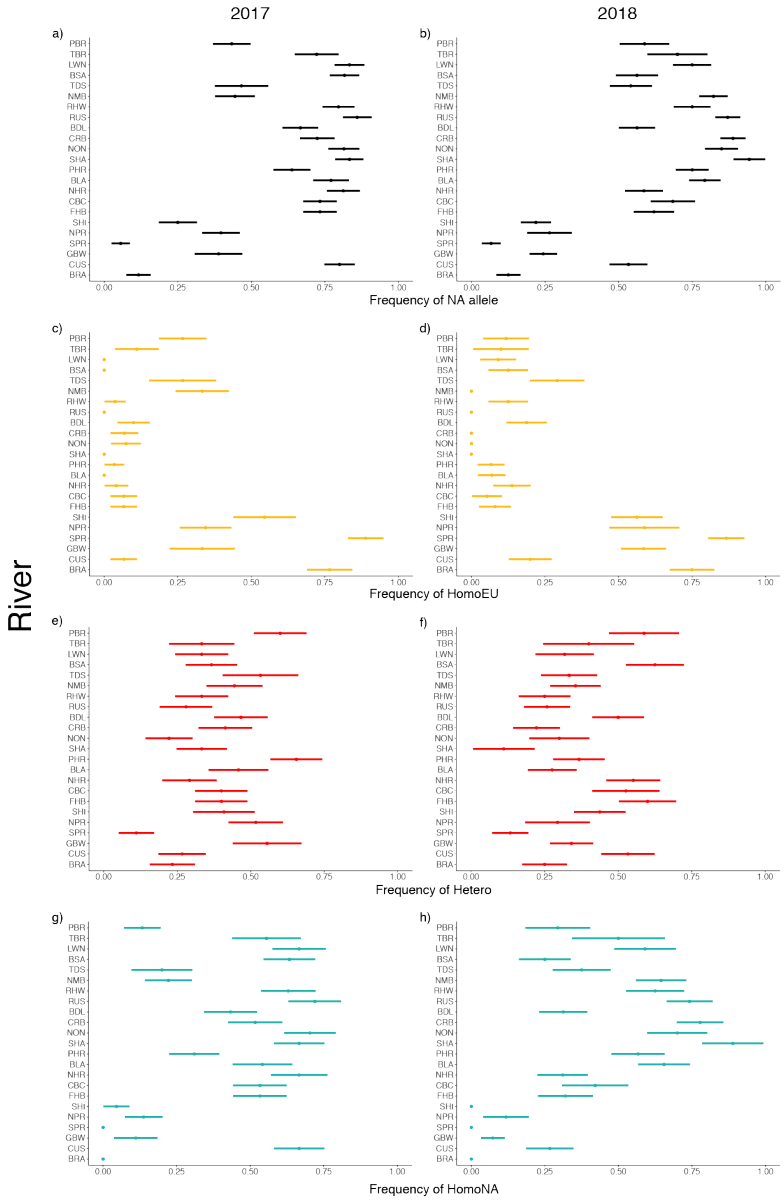

Figure S17 Heterogeneity of (a, b) translocation (c - h) and karyotype frequencies between rivers within year, (a, c, e, and g) 2017 and (b, d, f, and h) 2018 sampled. Bars represent confidence intervals. Rivers ordered geographically, east to west, around Placentia Bay, Newfoundland. Colours indicate karyotype; homozygous non-translocated (yellow), heterozygous (red), and homozygous translocated (blue-green).

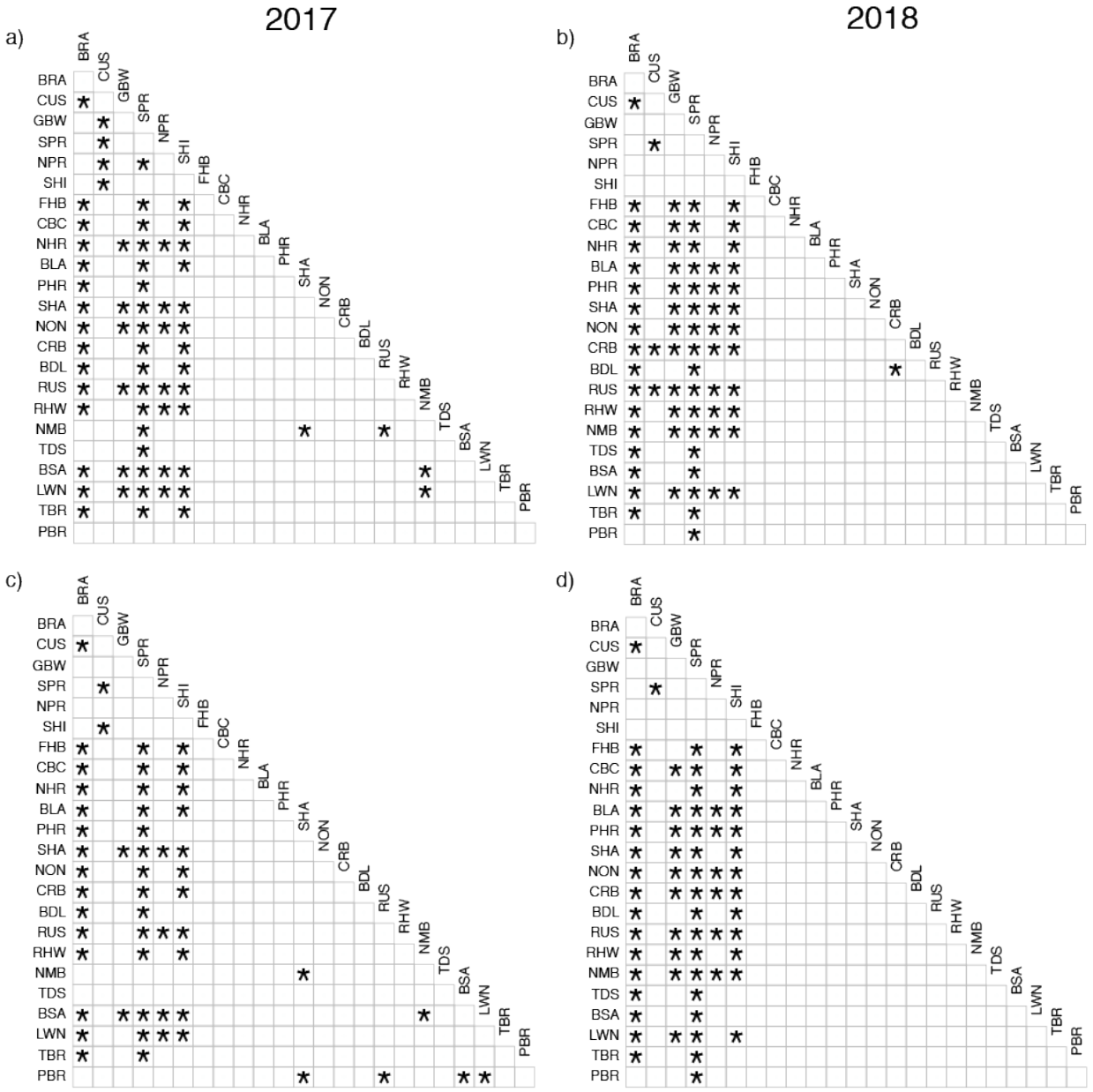

Figure S18 Pairwise comparison of (a, b) translocation (c, d) and karyotype frequencies between rivers within year, (a, c) 2017 (b, d) and 2018 sampled. Stars indicate significantly different frequencies, calculated using a Fisher Exact Test adjusted for multiple comparisons. Rivers ordered geographically, east to west, around Placentia Bay, Newfoundland.

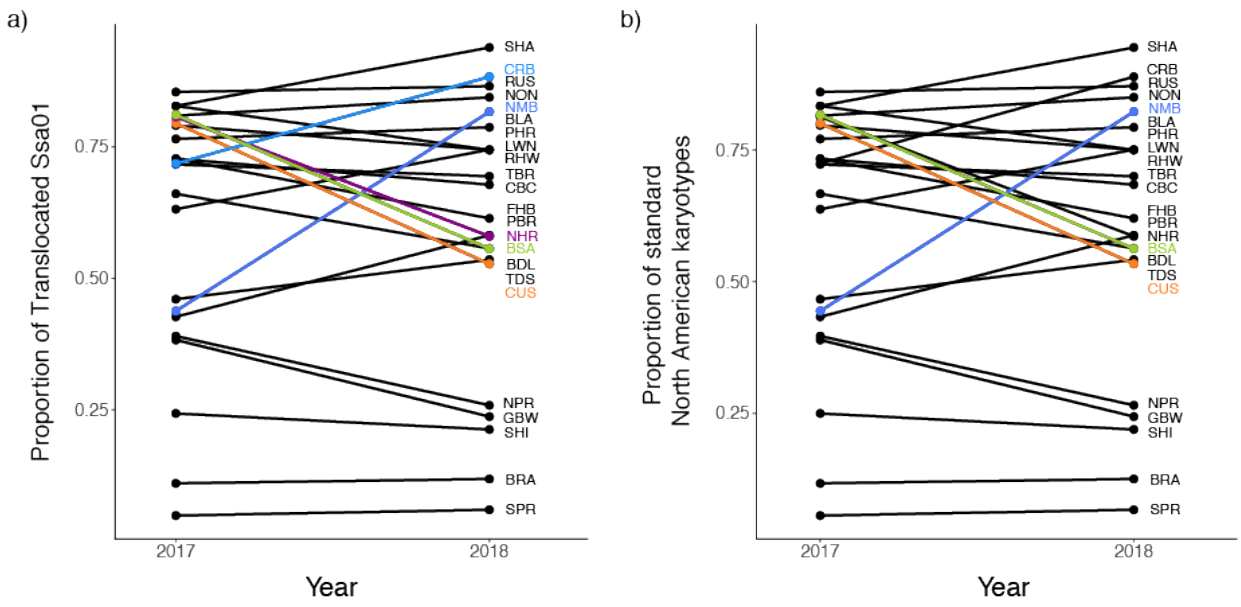

Figure S19 Temporal stability of (a) translocation (b) and karyotype frequencies within river between years sampled (2017 and 2018). Rivers that exhibited a significant change in translocation or karyotype frequency between 2017 and 2018 highlighted.

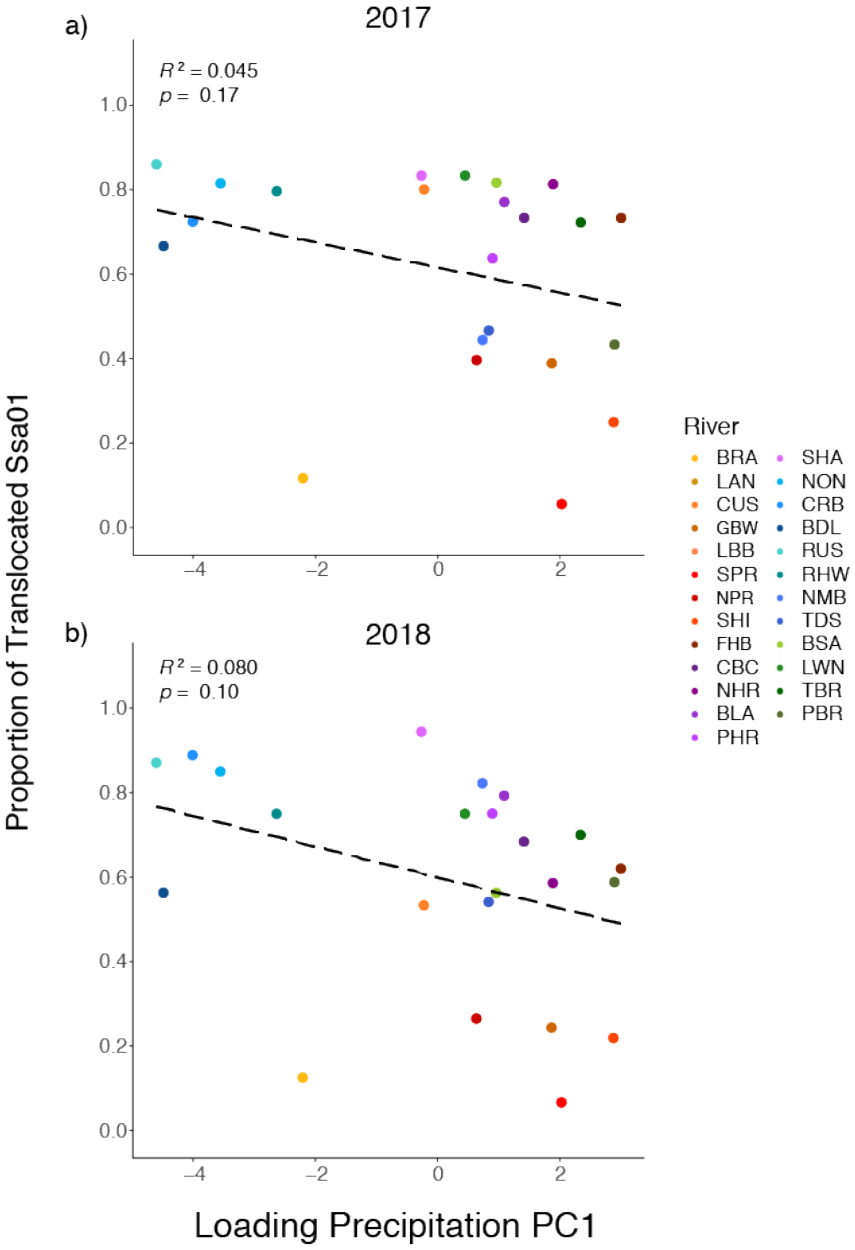

Figure S20 Correlation between precipitation and the Ssa01p/Ssa23 chromosomal translocation in Atlantic Salmon within Placentia Bay, Newfoundland, Canada. Linear regression of the first principal component (PC) of a PCA based on 9 precipitation variables (BIOCLIM) and proportion of non-translocated Ssa01p/q sampled in (a) 2017 and (b) 2018.

Appendix E – Matrix of Total Decline per Generation

Table S11 Total decline at each generation (up to a maximum of 20 generations as determined by the per generation rate of decline; 4 years per generation) across all modeled per generation rates of decline (0 – 60%).

							Generation												
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-5.0	-9.8	-14.3	-18.5	-22.6	-26.5	-30.2	-33.7	-37.0	-40.1	-43.1	-46.0	-48.7	-51.2	-53.7	-56.0	-58.2	-60.3	-62.3	-64.2
-10.0	-19.0	-27.1	-34.4	-41.0	-46.9	-52.2	-57.0	-61.3	-65.1	-68.6	-71.8	-74.6	-77.1	-79.4	-81.5	-83.3	-85.0	-86.5	-87.8
-20.0	-36.0	-48.8	-59.0	-67.2	-73.8	-79.0	-83.2	-86.6	-89.3	-91.4	-93.1								
-30.0	-51.0	-65.7	-76.0	-83.2	-88.2	-91.8	-94.2	-96.0	-97.2	-98.0	-98.6								
-40.0	-64.0	-78.4	-87.0	-92.2	-95.3	-97.2	-98.3												
-50.0	-75.0	-87.5	-93.8																
-60.0	-84.0	-93.6	-97.4																

Appendix F – Microsatellite Primers

Table S12 Primers with repeat motif, motif type, and linkage group for 101 microsatellite loci used to genotype Atlantic Salmon (*Salmo salar*) from Placentia Bay, Newfoundland, Canada.

Locus	LG	Accession Number	Left Oligo	Right Oligo	Repeat Motif
NGS-SSsp2210	12	AY081808.1	CACATTCACTGCAAAATAAAGCT	TGGGATTCAATAAAGGTAAGTAAGT	AGTT
NGS-SsaD486	16	AF525208.1	TGCAGTCCAATAATATCCCCGT	CCCTGCATGACTCGGATAAC	AGAT
Ssa-1.10	1	gi|925169083|ref|NW_012332498.1|	TGGATGACAACCTCCGTTAAAC	CGGGAAGCCTGGTGAAGATC	AAG
Ssa-1.11	1	gi|925169075|ref|NW_012332506.1|	CTCATCAACGCTATCCTCTTCC	GTCTTTCATCTGTCCGCGTG	ATC
Ssa-1.14	1	gi|925168868|ref|NW_012332713.1|	TCGTATTTGTCAAGGATGTGCC	AGATGCCCATTGTATTGCCC	AGT
Ssa-1.5	1	gi|925168832|ref|NW_012332749.1|	GCGTTATGTGCTTGCATGC	ACCACCGTACTCAGCTTATCC	ATT
Ssa-1.7	1	gi|925169069|ref|NW_012332512.1|	AGAACACAACAGAACCAGGTAC	CTCGAACACACTTCCAACCC	GAT
Ssa-1.8	1	gi|925168775|ref|NW_012332806.1|	AGGCCAAAGAAATCCTGCAC	ACTGACCCAAACACGCAAATAG	ATC
Ssa-1.9	1	gi|925169130|ref|NW_012332451.1|	CTGAGGAGCACAAAGGACAG	GTGTTGCTGGCTGTGTTCTC	AGG
Ssa-10.1	10	gi|925165007|ref|NW_012336574.1|	GGTCCTCCAGTACCTCCAAC	AATCTGGTGAGTTCGTCCGG	GTT
Ssa-10.2	10	gi|925164993|ref|NW_012336588.1|	TGATCCTCTTCACCACCCTG	CTGAAGACTCCTCCCTCACC	AAT
Ssa-10.4	10	gi|925164986|ref|NW_012336595.1|	GGTGAAATGTAGCCTGCATG	ACACACTGCTATATGTGTGG	AAT
Ssa-11.11	11	gi|925164498|ref|NW_012337083.1|	CGGCATATACCTTTAACGTTGG	GAAGAAGCGATGCGAGAGG	AAG
Ssa-11.12	11	gi|925164426|ref|NW_012337155.1|	CGTTAGCACACATGGCAAATC	GGTGCTGTTTGGGATGCATC	ACT
Ssa-11.2	11	gi|925164888|ref|NW_012336693.1|	AAAGTTTGTTTGTGGACCGC	CGGACAGTTTCTTGGACTTC	AAG
Ssa-11.3	11	gi|925164793|ref|NW_012336788.1|	AGCGTGTGTGTCGTTCAATAC	ATGTTTCACCTCTGCGTCAC	AAC
Ssa-11.5	11	gi|925164446|ref|NW_012337135.1|	GTGTGCCGTTCTATCGCTG	CCTAAAGAAATGCCAGAGTCCG	GAT
Ssa-11.6	11	gi|925164433|ref|NW_012337148.1|	TTAACCTGCTCTACCTCTCG	ACATCACCACACCTATCTTC	TGA
Ssa-11.8	11	gi|925164901|ref|NW_012336680.1|	AAAGGACCCAGAACGTACAG	ACCACACAGTACCCTCAATG	ATA
Ssa-12.12	12	gi|925164059|ref|NW_012337522.1|	TTGCTGCTGGTTTGTGCTC	GGGACAGTGAAGTGGTATTGC	GAT
Ssa-12.13	12	gi|925164020|ref|NW_012337561.1|	ATCAGGCTCAGAGGTGGAAC	ACACAGTGGAGGTAGAGATAGC	CCT
Ssa-12.5	12	gi|925164069|ref|NW_012337512.1|	TCTCCTTCCTCGATCAGCTC	AATGTGTCGCCTTCCCACC	ATC

Locus	LG	Accession Number	Left Oligo	Right Oligo	Repeat Motif
Ssa-13.10	13	gi|925163859|ref|NW_012337722.1|	TGAAAGTTGGCTGCAATCCG	GGAACCTGTCTGCCCACAC	AAG
Ssa-13.12	13	gi|925163769|ref|NW_012337812.1|	AGTTTGGCGTAGTCTGGGAC	TCCCATCATCCTCCTCTGGG	AGG
Ssa-13.2	13	gi|925163922|ref|NW_012337659.1|	CTACACCAAGAGTCCAGTGTC	ACAATTTGTCTCCCTGTTGTTG	AAT
Ssa-13.8	13	gi|925163870|ref|NW_012337711.1|	TGACGAGACAAGATTCAGGTTG	GACCTATGCAACCACCAACG	GTT
Ssa-14.10	14	gi|925163717|ref|NW_012337864.1|	GGGAACGTGTGGAAGATTCAC	AAGGTATGGAGGGTGATGCC	ATC
Ssa-14.2	14	gi|925163728|ref|NW_012337853.1|	GGGCATGATCTCGACACC	AGGAATGAGTAAGCTGGCTAAG	ATC
Ssa-14.3	14	gi|925163708|ref|NW_012337873.1|	TCAACCTAAACCCTCTGCCC	AATCATCACATTCCACAGCAAC	AAT
Ssa-14.5	14	gi|925163651|ref|NW_012337930.1|	CCAGGAGGCCTTCACATG	CCTCCTGGCAATGCTGTATAG	AAT
Ssa-14.6	14	gi|925163718|ref|NW_012337863.1|	AGTCAAGAAAGTCACTGCCC	GGAATGGCAAACAGAAAGGG	ATT
Ssa-14.8	14	gi|925163650|ref|NW_012337931.1|	AAACATTGATTTGGCTCTGTC	TATTGCACCATCCCGTTCTC	TAT
Ssa-14.9	14	gi|925163731|ref|NW_012337850.1|	CCATAATGGCACTGCTTCTTC	GTGTTGCTTCATTACACTCCG	AAT
Ssa-15.1	15	gi|925163276|ref|NW_012338305.1|	TTTCTTTGTGTGTTGTGCCC	CAGCTGTGGTTCCTCTGGG	CCT
Ssa-15.3	15	gi|925163318|ref|NW_012338263.1|	GCTAACGAATGACAGCTTGC	CATTAGTAAGACTGGCAGCAG	TTG
Ssa-15.7	15	gi|925163309|ref|NW_012338272.1|	GATGTGATGGCAGTGCTATG	CAGCAACAAGGTCAATCTCC	TGA
Ssa-16.5	16	gi|925163075|ref|NW_012338506.1|	CCGCTGGATTCCTCATTATGTC	GGACTGACAGGAAGAGAGACC	TTA
Ssa-17.1	17	gi|925162405|ref|NW_012339176.1|	CATCTTCCGGTTCGCTCAAC	GTCATGACCTGTGCAACCAG	ATT
Ssa-17.2	17	gi|925162379|ref|NW_012339202.1|	ACCCATAGAATTACTGCACTGG	GTCGTACTGGCATAATGTCAAC	ATT
Ssa-18.7	18	gi|925161808|ref|NW_012339773.1|	TGCAGGTTGTGGTCATGTTG	CACATTCTGTCCATTCGGCC	TTA
Ssa-19.1	19	gi|925161618|ref|NW_012339963.1|	TGTGCAAACGCCATGATACC	CCATGACAGCTCCATCCGG	GTA
Ssa-19.2	19	gi|925161617|ref|NW_012339964.1|	GTGACCCAAAGTGCTGCTG	CTCCAGACACCAGCACCTC	GCT
Ssa-19.3	19	gi|925161703|ref|NW_012339878.1|	ACGTCCTGACAGTTATCCTTG	GTCTTGTCATGGCTGTGCTC	TTC
Ssa-19.7	19	gi|925161695|ref|NW_012339886.1|	CTCCTTCACACAACCACC	AAGTGCAGACCTACCTTGTG	AGC
Ssa-19.9	19	gi|925161703|ref|NW_012339878.1|	TCTGGTGCTGACGATGAGAG	GAAATCAGAGGTCATTGGCCC	GAT
Ssa-2.1	2	gi|925168396|ref|NW_012333185.1|	AGACTCCACCTGCCTTGTTC	CTCACTGTCAGAGCATGCG	ATT
Ssa-2.12	2	gi|925168396|ref|NW_012333185.1|	CAGTACAGAAGCAGTCATCGC	ATTGTTTGCGGACGGTCATG	CTT
Ssa-2.13	2	gi|925168345|ref|NW_012333236.1|	GCTCAGATCGCAACCTTGAC	TCTAAACCGACCAGACCGAG	CCT

Locus	LG	Accession Number	Left Oligo	Right Oligo	Repeat Motif
Ssa-2.2	2	gi|925168422|ref|NW_012333159.1|	TGGCCATTCTCCAGAGCTAG	CCACCAAAGGAGAGTACGTG	CTT
Ssa-2.7	2	gi|925168409|ref|NW_012333172.1|	CCCAGACTTCCCACTCTCTATG	GGACACAGAACCTTGAACGG	CTG
Ssa-20.2	20	gi|925161392|ref|NW_012340189.1|	TCTTCCCTCTTCTGCAGCAG	AGCTCTGGACACCACACTG	GTG
Ssa-20.d56	20	gi|925161350|ref|NW_012340231.1|	GAGGTCAAGGTTTCCACTGG	TAGCTGCTCTCTGTTCTGGG	AG
Ssa-21.10	21	gi|925161196|ref|NW_012340385.1|	ACTGCTTAGCTAGATTTGGCC	TCTACAGACAGGTGAACATGC	ATT
Ssa-21.2	21	gi|925161196|ref|NW_012340385.1|	CTGTCCAAATTGCAGGCTTG	GCCTAATTTGCCTACTCCTGTC	TAT
Ssa-21.3	21	gi|925161206|ref|NW_012340375.1|	TTGAACCTGAACTGGAATCCC	ACCGGCCAGTCTGAAACAG	ATC
Ssa-21.5	21	gi|925161233|ref|NW_012340348.1|	CACTCCCTAACTCCATGGTC	TCATGGATGTCGTCACTGTG	ATG
Ssa-22.2	22	gi|925161105|ref|NW_012340476.1|	AGTGGTTGCTTTGGTTCTCC	GGATAAAGCGGACCAAGACG	AAT
Ssa-22.5	22	gi|925161127|ref|NW_012340454.1|	GTGACGTCTGGAATTGTGAC	GATCCAATCAACACCGGTAG	AAT
Ssa-22.9	22	gi|925161127|ref|NW_012340454.1|	CAAATGCCACACGACCTGAC	GGTCAACCGCTCTGCATATAG	ATT
Ssa-22.d31	22	gi|925161077|ref|NW_012340504.1|	AGTTTAGTAGGGCCTGCGTG	ACATTCTTCTGTCACAGCCTG	GT
Ssa-22.d40	22	gi|925161053|ref|NW_012340528.1|	GCACAGAGGTAAGAGTTCAGC	CTCTGCTGCTGTGGGTGG	AC
Ssa-22.d41	22	gi|925161051|ref|NW_012340530.1|	CTCTGTGGTCTGGGTCCTC	ACCTCGTACCCATGCACATC	CT
Ssa-22.d44	22	gi|925161034|ref|NW_012340547.1|	GTACCTTTGAACATGCACACG	CATCTCCACATGATAACGTTGC	AC
Ssa-23.10	23	gi|925161012|ref|NW_012340569.1|	TGATTGTGAACGGCTTTGGG	ACAAGCAAGCACCCTTTGTC	GAT
Ssa-23.2	23	gi|925161018|ref|NW_012340563.1|	GGTGGTTGTTTCTAGTGAGGG	GCACCTCTAAAGCACCATGG	CTT
Ssa-23.3	23	gi|925160997|ref|NW_012340584.1|	GGAGAAGTGATTATGGTTGTGC	GGACAACGGGTTCTACATGG	AAC
Ssa-23.9	23	gi|925161018|ref|NW_012340563.1|	ACGGATACAGAGAGACGCAC	ACAGCGAGGAGGACAAAGTC	ATC
Ssa-24.9	24	gi|925160923|ref|NW_012340658.1|	CACTCCATCTATCATCTGTGCC	GATGAGGAGCAGAAGAGGCC	CTT
Ssa-24.d09	24	gi|925160918|ref|NW_012340663.1|	ACCGTAAGCAGCATCACTTTAG	GTTTGGGCTGTCTGGTACTG	AC
Ssa-24.d24	24	gi|925160864|ref|NW_012340717.1|	CTGCCAACACACACTGCC	TTTGACTCTTCCTGTATGTCGG	CT
Ssa-25.11	25	gi|925160727|ref|NW_012340854.1|	GGGTCCATGAGAAAGGCAAC	TGGGATCCACACCTGACAAC	ATT
Ssa-25.2	25	gi|925160771|ref|NW_012340810.1|	TGCAGGAAGACTCTGAAAGG	AGGTGGGTGTTGTACATCAG	GAT
Ssa-25.3	25	gi|925160672|ref|NW_012340909.1|	TTCCCACTGGCCAAGAACTG	GACATTCCCTTGTGTTGATGAC	AAC
Ssa-26.1	26	gi|925160353|ref|NW_012341228.1|	TCACGCATAACCTTAGACAACC	AATGCCAACCCTGTTACAGC	ATT

Locus	LG	Accession Number	Left Oligo	Right Oligo	Repeat Motif
Ssa-26.d06	26	gi|925160645|ref|NW_012340936.1|	CATAATCACCTTGCATGACACC	CCTGCTGCACCGCTAAATAC	AC
Ssa-27.7	27	gi|925160107|ref|NW_012341474.1|	TCATCAGTGTGGAGGGAATC	TCTATCTTCCTCTGGCCTGG	TTG
Ssa-27.d46	27	gi|925160006|ref|NW_012341575.1|	TGGCTGGTGGTTATAGGAGC	ACCATGCCAAGACAGTGATG	AC
Ssa-28.d01	28	gi|925159997|ref|NW_012341584.1|	ATTACTGCCCTATCGCCATG	TCACCTTCTTCACACACGATG	AC
Ssa-29.d18	29	gi|925159852|ref|NW_012341729.1|	AGCTACCTATTCCTGGAGCG	AGAGATGTTAGCGGGTCAGG	GT
Ssa-29.d33	29	gi|925159809|ref|NW_012341772.1|	TAACTGCTGAGCCGTGTGTC	GCAGTGAATTCTATCTTCGTCG	AG
Ssa-3.10	3	gi|925168028|ref|NW_012333553.1|	GACTGCAACTAACTGAATGACG	TCCATCATCCCTTTCAGCTG	ATC
Ssa-3.2	3	gi|925168028|ref|NW_012333553.1|	GTCACCAATACCACGTCACC	TCGTCAAGGGATGTGGTCAC	ATC
Ssa-3.9	3	gi|925168037|ref|NW_012333544.1|	CACCTCCAACTGCTCAATTAGG	GAGGCCCGTGTTTCTCAAC	AGT
Ssa-4.d44	4	gi|925167334|ref|NW_012334247.1|	TTGGGTCTTAATGGCACCTG	GCTTTGGTTCCCTGAGAGTG	AC
Ssa-5.11	5	gi|925167048|ref|NW_012334533.1|	CAACCGCCGTTAAACATCATC	CATCATTGTGAATTGGAAGGCC	AAG
Ssa-5.2	5	gi|925167062|ref|NW_012334519.1|	AACTTGCGTGATGATGTGGC	GCTGGCCATGTTCTTCTGTG	CTT
Ssa-5.6	5	gi|925167074|ref|NW_012334507.1|	GTGCAGCTGTTCCTCACTTC	GGGACAGGCGTAGAAATCG	TAT
Ssa-5.8	5	gi|925167158|ref|NW_012334423.1|	ACACAGCTCTTATTTAACCGTC	GAAGGAATCTCACTCGTCTAAG	AAC
Ssa-6.11	6	gi|925166307|ref|NW_012335274.1|	CCGTGGAAAGCACTTAACATG	GAACGCATGTCATGGCCTC	ATT
Ssa-6.2	6	gi|925166267|ref|NW_012335314.1|	GGAGAAGAGGAGATGGAACTTG	ACACCTGACAATACCACACC	TGA
Ssa-6.7	6	gi|925166258|ref|NW_012335323.1|	GCAAATCAGCATTCAGGGC	CAGCTGATCGAACTGAATGGG	TAA
Ssa-7.1	7	gi|925166059|ref|NW_012335522.1|	CCACTCCCACGAATGATGTTC	GGAGGCCACATTGCAGTC	AAC
Ssa-7.12	7	gi|925166045|ref|NW_012335536.1|	CACTCCCTGACACGTTAACAC	CACTTCCTGACAAACATGCAC	ACT
Ssa-7.d33	7	gi|925166053|ref|NW_012335528.1|	AGCATAGCATAGGAACAGACAC	AGCACATCCTGACCTCATCG	AC
Ssa-7.d47	7	gi|925166033|ref|NW_012335548.1|	TGGAATTGGGTCAGCAGTTC	AGGACAGGGTTGAGATCAGC	GT
Ssa-8.d04	8	gi|925165795|ref|NW_012335786.1|	ACTGTGTGGACTGGGAGATC	CAGCAGCGTTGTCTTGTACC	GT
Ssa-8.d07	8	gi|925165792|ref|NW_012335789.1|	GGGTGTGAGGGAGGACTTAAC	TGCTAGCTACACTCCTGTCC	AC
Ssa-9.10	9	gi|925165582|ref|NW_012335999.1|	TCCATTGTTCCCTCAGACCC	GGTAACATGAAGGAGAGCTGG	ATT
Ssa-9.13	9	gi|925165161|ref|NW_012336420.1|	ATCCACACCTCTCTTGCCAC	GATCACCATCGTTACCATCCC	AGG
Ssa-9.3	9	gi|925165300|ref|NW_012336281.1|	GCCAACCACCGTTAAACCTC	TCAGCAGTTCCCAATATTTCCC	AAG

Ssa-9.8	9	gi|925165505|ref|NW_012336076.1|	GCGTCGACTGCCATTCAAC	TGTCCTTGCTTTCTCCGTGG	AAG

Appendix G – Summary Statistics: Microsatellites

Table 13 Summary of genetic diversity for Atlantic Salmon (*Salmo salar*) from 26 rivers in Placentia Bay, Newfoundland, Canada. Rivers are ordered geographically, east to west around the bay. Number of samples (N), mean observed (H_o) and expected (H_e) heterozygosity, and fixation index (F) calculated across 101 microsatellite loci using GENALEX.

RiverID	N	*Ho*	*He*	*F*
BRA	52.6	0.47	0.48	0.01
LAN	30.6	0.51	0.48	-0.06
CUS	54.8	0.48	0.48	0.00
GBW	57.8	0.52	0.51	-0.01
LBB	9.7	0.53	0.54	0.01
SPR	55.8	0.54	0.54	0.00
NPR	55.8	0.49	0.49	0.01
SHI	51.4	0.54	0.53	0.01
FHB	54.4	0.55	0.52	-0.05
CBC	56.5	0.53	0.52	-0.01
NHR	55.5	0.52	0.57	0.11
BLA	57.5	0.51	0.50	-0.01
PHR	52.1	0.54	0.57	0.07
SHA	47.9	0.52	0.52	0.01
NON	56.5	0.54	0.53	-0.02
CRB	56.2	0.53	0.54	0.04
BDL	55.9	0.53	0.52	-0.02
RUS	56.1	0.51	0.50	-0.01
RHA	54.9	0.47	0.45	-0.04
RHW	57.5	0.54	0.54	0.00
NMB	57.4	0.52	0.51	-0.02
TDS	56.4	0.50	0.50	0.02
BSA	56.7	0.55	0.55	0.00
LWN	53.6	0.54	0.53	-0.01
TBR	39.6	0.52	0.51	-0.02
PBR	53.8	0.54	0.53	-0.02

Appendix H – Gene Annotations

Table S14 Gene annotations with gene IDs, position, and description for single nucleotide polymorphisms (SNPs) identified as outliers within the outlier block regions on chromosomes Ssa01 and Ssa23 in both 2017 and 2018.

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87301894	1	44027302	100196073	sptlc2	serine palmitoyltransferase long chain base subunit 2
AX-87031939	1	44047970	100196073	sptlc2	serine palmitoyltransferase long chain base subunit 2
AX-87000133	1	44073867	106603015	LOC106603015	isthmin-2-like
AX-87112657	1	44094342	106603015	LOC106603015	isthmin-2-like
AX-87525318	1	44102747	106603015	LOC106603015	isthmin-2-like
AX-86978638	1	44127633	unknown	unknown	unknown
AX-87198976	1	44137333	106603014	LOC106603014	zinc finger protein 585A-like
AX-87669190	1	44140594	106603014	LOC106603014	zinc finger protein 585A-like
AX-87585106	1	44168773	106603014	LOC106603014	zinc finger protein 585A-like
AX-87827704	1	44202646	unknown	unknown	unknown
AX-86920941	1	44267246	unknown	unknown	unknown
AX-87109218	1	44278327	unknown	unknown	unknown
AX-87077941	1	44338122	unknown	unknown	unknown
AX-87163493	1	44338423	unknown	unknown	unknown
AX-87828321	1	44338572	unknown	unknown	unknown
AX-87014637	1	44341000	unknown	unknown	unknown
AX-87540882	1	44359836	unknown	unknown	unknown
AX-87484087	1	44371924	unknown	unknown	unknown
AX-87828481	1	44376516	unknown	unknown	unknown
AX-87851697	1	44378204	unknown	unknown	unknown
AX-87854532	1	44520582	106602991	LOC106602991	prostaglandin E2 receptor EP2 subtype-like
AX-87248197	1	44532294	106602991	LOC106602991	prostaglandin E2 receptor EP2 subtype-like
AX-87366989	1	44553756	unknown	unknown	unknown
AX-87093431	1	44555170	unknown	unknown	unknown
AX-87814747	1	44572331	unknown	unknown	unknown
AX-87246674	1	44621966	106602988	txndc16	thioredoxin domain containing 16
AX-86923463	1	44714935	106602981	LOC106602981	integral membrane protein GPR137B-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87485893	1	44717972	106602981	LOC106602981	integral membrane protein GPR137B-like
AX-87056913	1	44718557	106602981	LOC106602981	integral membrane protein GPR137B-like
AX-87232987	1	44766025	106602955	LOC106602955	protein Smaug homolog 1-like
AX-87082190	1	44766956	106602955	LOC106602955	protein Smaug homolog 1-like
AX-87061268	1	44813027	106602955	LOC106602955	protein Smaug homolog 1-like
AX-87703405	1	44818603	106602955	LOC106602955	protein Smaug homolog 1-like
AX-87321322	1	44861657	106602955	LOC106602955	protein Smaug homolog 1-like
AX-87460740	1	44892893	106602973	LOC106602973	GTP cyclohydrolase 1-like
AX-87066482	1	44892947	106602973	LOC106602973	GTP cyclohydrolase 1-like
AX-87577138	1	44895199	106602973	LOC106602973	GTP cyclohydrolase 1-like
AX-87207137	1	44900574	106602944	wdhd1	WD repeat and HMG-box Dunknown binding protein 1
AX-87439414	1	44905362	106602944	wdhd1	WD repeat and HMG-box Dunknown binding protein 1
AX-87035765	1	44907561	106602944	wdhd1	WD repeat and HMG-box Dunknown binding protein 1
AX-87024496	1	44951346	106602944	wdhd1	WD repeat and HMG-box Dunknown binding protein 1
AX-87798888	1	44953032	106602944	wdhd1	WD repeat and HMG-box Dunknown binding protein 1
AX-87685882	1	44988595	unknown	unknown	unknown
AX-87402125	1	45001057	106602923	LOC106602923	serine/threonine/tyrosine-interacting protein A-like
AX-87383404	1	45010395	106602923	LOC106602923	serine/threonine/tyrosine-interacting protein A-like
AX-87337216	1	45029241	106602923	LOC106602923	serine/threonine/tyrosine-interacting protein A-like
AX-86960675	1	45043263	106602906	LOC106602906	fos-related antigen 2-like
AX-87382934	1	45063862	106602899	LOC106602899	BRCA1-A complex subunit BRE-like
AX-87366360	1	45127391	106602899	LOC106602899	BRCA1-A complex subunit BRE-like
AX-87507040	1	45140746	106602899	LOC106602899	BRCA1-A complex subunit BRE-like
AX-87211747	1	45173535	106602899	LOC106602899	BRCA1-A complex subunit BRE-like
AX-87314541	1	45224612	106602899	LOC106602899	BRCA1-A complex subunit BRE-like
AX-87724078	1	45227079	106602899	LOC106602899	BRCA1-A complex subunit BRE-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87518572	1	45383995	106602852	LOC106602852	ribokiunknownse-like
AX-87780483	1	45386529	106602852	LOC106602852	ribokiunknownse-like
AX-87129687	1	45488112	106602886	LOC106602886	uncharacterized LOC106602886
AX-87538852	1	45794290	100196602	ap3m1	adaptor-related protein complex 3, mu 1 subunit
AX-87665933	1	45856239	106604120	LOC106604120	iunknownctive tyrosine-protein kiunknownse 7-like
AX-87633031	1	45858524	106604120	LOC106604120	iunknownctive tyrosine-protein kiunknownse 7-like
AX-87181742	1	45875647	106604182	LOC106604182	serum response factor-like
AX-87862056	1	45881901	106604182	LOC106604182	serum response factor-like
AX-87766838	1	45887730	106604182	LOC106604182	serum response factor-like
AX-87566215	1	45948175	106604182	LOC106604182	serum response factor-like
AX-87706082	1	45992431	106604219	LOC106604219	protein FAM160B1-like
AX-87197550	1	46021016	100196372	znf511	zinc finger protein 511
AX-87697616	1	46021057	100196372	znf511	zinc finger protein 511
AX-87606718	1	46021856	100196372	znf511	zinc finger protein 511
AX-87018974	1	46326660	106604129	LOC106604129	uncharacterized protein K02A2.6-like
AX-87588916	1	46462826	unknown	unknown	unknown
AX-86998955	1	46493383	106604231	LOC106604231	V-set and transmembrane domain-containing protein 4-like
AX-87742227	1	46512467	106604231	LOC106604231	V-set and transmembrane domain-containing protein 4-like
AX-87577764	1	46546646	106604247	wdfy4	WDFY family member 4
AX-86958517	1	46730283	106604247	wdfy4	WDFY family member 4
AX-87611269	1	46783740	106604271	LOC106604271	palmitoyltransferase akr1-like
AX-87051061	1	46949714	106604144	frmpd2	FERM and PDZ domain containing 2
AX-87249213	1	47048340	106604144	frmpd2	FERM and PDZ domain containing 2
AX-87666161	1	47116978	106605247	LOC106605247	uncharacterized LOC106605247
AX-87792105	1	47195293	106605253	LOC106605253	uncharacterized LOC106605253
AX-87378411	1	47220624	106605260	LOC106605260	ubiquitin domain-containing protein 1-like
AX-87072994	1	47267601	106604325	LOC106604325	probable palmitoyltransferase ZDHHC16
AX-87489849	1	47442649	106604352	oga	O-Glcunknowncase

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87131548	1	47496073	106604361	pprc1	peroxisome proliferator-activated receptor gamma, coactivator-related 1
AX-87235059	1	47501122	106604361	pprc1	peroxisome proliferator-activated receptor gamma, coactivator-related 1
AX-87411350	1	47503136	106604361	pprc1	peroxisome proliferator-activated receptor gamma, coactivator-related 1
AX-87711620	1	47504102	106604361	pprc1	peroxisome proliferator-activated receptor gamma, coactivator-related 1
AX-87409540	1	47505746	106604361	pprc1	peroxisome proliferator-activated receptor gamma, coactivator-related 1
AX-87190186	1	47506173	106604361	pprc1	peroxisome proliferator-activated receptor gamma, coactivator-related 1
AX-87862191	1	47513277	106604361	pprc1	peroxisome proliferator-activated receptor gamma, coactivator-related 1
AX-87002972	1	47829953	unknown	unknown	unknown
AX-87852407	1	47921484	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-87141381	1	47945880	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-86992511	1	48035929	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-87270787	1	48066560	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-87191078	1	48075882	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-87044398	1	48077805	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-87660926	1	48111481	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-87466867	1	48181298	106604421	kcnip2	potassium voltage-gated channel interacting protein 2
AX-87566793	1	48220496	106604458	LOC106604458	cytochrome c oxidase assembly protein COX15 homolog
AX-87630277	1	48295154	106604156	LOC106604156	leucine-rich repeat flightless-interacting protein 2-like
AX-87311187	1	48300318	106604156	LOC106604156	leucine-rich repeat flightless-interacting protein 2-like
AX-87538130	1	48307858	unknown	unknown	unknown

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87841790	1	48312017	106604487	LOC106604487	protein FAM83H-like
AX-87859879	1	48313032	106604487	LOC106604487	protein FAM83H-like
AX-86955577	1	48313321	106604487	LOC106604487	protein FAM83H-like
AX-87780910	1	48313559	106604487	LOC106604487	protein FAM83H-like
AX-87277662	1	48352848	106604487	LOC106604487	protein FAM83H-like
AX-87482156	1	48352912	106604487	LOC106604487	protein FAM83H-like
AX-87742618	1	48353378	106604487	LOC106604487	protein FAM83H-like
AX-87383255	1	48356688	106604487	LOC106604487	protein FAM83H-like
AX-87454829	1	48370707	106604506	LOC106604506	uncharacterized LOC106604506
AX-87383494	1	48370715	106604506	LOC106604506	uncharacterized LOC106604506
AX-87708945	1	48372430	106604506	LOC106604506	uncharacterized LOC106604506
AX-86970997	1	48372885	106604506	LOC106604506	uncharacterized LOC106604506
AX-87399290	1	48373522	106604506	LOC106604506	uncharacterized LOC106604506
AX-86930986	1	48375502	106604506	LOC106604506	uncharacterized LOC106604506
AX-87105803	1	48423083	106604510	LOC106604510	unknownD-dependent protein deacetylase sirtuin-1-like
AX-87192184	1	48441244	106604523	LOC106604523	CDK2-associated and cullin domain-containing protein 1-like
AX-87073730	1	48442760	106604523	LOC106604523	CDK2-associated and cullin domain-containing protein 1-like
AX-86913002	1	48444035	106604523	LOC106604523	CDK2-associated and cullin domain-containing protein 1-like
AX-87155100	1	48457529	106604523	LOC106604523	CDK2-associated and cullin domain-containing protein 1-like
AX-87672259	1	48460607	106604523	LOC106604523	CDK2-associated and cullin domain-containing protein 1-like
AX-87734876	1	48498689	106604517	LOC106604517	myopalladin-like
AX-87791526	1	48505736	106604517	LOC106604517	myopalladin-like
AX-87179595	1	48512316	106604517	LOC106604517	myopalladin-like
AX-87323592	1	48527551	100195378	tiar	Nucleolysin TIAR
AX-87292389	1	48560282	106604528	LOC106604528	BAG family molecular chaperone regulator 3-like
AX-87136071	1	48578249	106604534	LOC106604534	phosphatidylinositide phosphatase SAC2-like
AX-87793695	1	48592565	106604534	LOC106604534	phosphatidylinositide phosphatase SAC2-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87846116	1	48593271	106604534	LOC106604534	phosphatidylinositide phosphatase SAC2-like
AX-87753209	1	48608175	106604174	LOC106604174	mini-chromosome mainteunknownnce complex-binding protein-like
AX-87502413	1	48627411	106604541	LOC106604541	SEC23-interacting protein-like
AX-87591888	1	48628085	106604541	LOC106604541	SEC23-interacting protein-like
AX-87336933	1	48628601	106604541	LOC106604541	SEC23-interacting protein-like
AX-87041072	1	48628682	106604541	LOC106604541	SEC23-interacting protein-like
AX-87484496	1	48641373	106604550	LOC106604550	chondroitin sulfate N-acetylgalactosaminyltransferase 2-like
AX-87382269	1	48644092	106604550	LOC106604550	chondroitin sulfate N-acetylgalactosaminyltransferase 2-like
AX-87710406	1	48649283	106604550	LOC106604550	chondroitin sulfate N-acetylgalactosaminyltransferase 2-like
AX-87190225	1	48649784	106604550	LOC106604550	chondroitin sulfate N-acetylgalactosaminyltransferase 2-like
AX-87504051	1	48654386	106604566	LOC106604566	proto-oncogene tyrosine-protein kiunknownse receptor Ret-like
AX-87715625	1	48655822	106604566	LOC106604566	proto-oncogene tyrosine-protein kiunknownse receptor Ret-like
AX-87437394	1	48657171	106604566	LOC106604566	proto-oncogene tyrosine-protein kiunknownse receptor Ret-like
AX-87216261	1	48666917	106604566	LOC106604566	proto-oncogene tyrosine-protein kiunknownse receptor Ret-like
AX-87137509	1	48668258	106604566	LOC106604566	proto-oncogene tyrosine-protein kiunknownse receptor Ret-like
AX-87354044	1	48668964	106604566	LOC106604566	proto-oncogene tyrosine-protein kiunknownse receptor Ret-like
AX-86935330	1	48687968	106605279	LOC106605279	centromere protein O-like
AX-87520867	1	48704988	106605279	LOC106605279	centromere protein O-like
AX-87019853	1	48706665	106605279	LOC106605279	centromere protein O-like
AX-87695934	1	48706810	106605279	LOC106605279	centromere protein O-like
AX-86999786	1	48727169	106604574	LOC106604574	uncharacterized LOC106604574
AX-87356844	1	48731829	106604574	LOC106604574	uncharacterized LOC106604574
AX-87256823	1	48739735	106604574	LOC106604574	uncharacterized LOC106604574
AX-87191737	1	48740644	106604574	LOC106604574	uncharacterized LOC106604574
AX-87079514	1	48748596	100286615	phs	Pterin-4-alpha-carbinolamine dehydratase

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87508384	1	48757030	106604578	LOC106604578	sphingosine-1-phosphate lyase 1-like
AX-87010609	1	48759452	106604578	LOC106604578	sphingosine-1-phosphate lyase 1-like
AX-87047572	1	48759681	106604578	LOC106604578	sphingosine-1-phosphate lyase 1-like
AX-87720904	1	48789079	unknown	unknown	unknown
AX-87045154	1	48823294	106604587	LOC106604587	PDZ and LIM domain protein 1-like
AX-87070309	1	48829331	106604587	LOC106604587	PDZ and LIM domain protein 1-like
AX-87731886	1	48829980	106604587	LOC106604587	PDZ and LIM domain protein 1-like
AX-87732698	1	48864579	106604599	LOC106604599	sorbin and SH3 domain-containing protein 1-like
AX-87391063	1	48877890	106604599	LOC106604599	sorbin and SH3 domain-containing protein 1-like
AX-87621055	1	48882168	106604599	LOC106604599	sorbin and SH3 domain-containing protein 1-like
AX-87010253	1	48887213	106604599	LOC106604599	sorbin and SH3 domain-containing protein 1-like
AX-87479232	1	48887450	106604599	LOC106604599	sorbin and SH3 domain-containing protein 1-like
AX-87323552	1	48942442	106604619	LOC106604619	phosphoinositide 3-kiunknownse adapter protein 1-like
AX-87285681	1	48990407	106604631	fam204a	family with sequence similarity 204 member A
AX-86959643	1	48997853	106604631	fam204a	family with sequence similarity 204 member A
AX-87462801	1	49030704	unknown	unknown	unknown
AX-87403012	1	49037588	unknown	unknown	unknown
AX-86967883	1	49039114	unknown	unknown	unknown
AX-87747643	1	49040483	unknown	unknown	unknown
AX-87458069	1	49066162	unknown	unknown	unknown
AX-87672638	1	49118128	106604674	rab11fip2	RAB11 family interacting protein 2
AX-86918576	1	49118152	106604674	rab11fip2	RAB11 family interacting protein 2
AX-87341401	1	49120419	106604674	rab11fip2	RAB11 family interacting protein 2
AX-87833129	1	49122217	106604674	rab11fip2	RAB11 family interacting protein 2
AX-87510648	1	49158402	unknown	unknown	unknown
AX-86948426	1	49210162	unknown	unknown	unknown
AX-87723237	1	49217410	unknown	unknown	unknown
AX-87239262	1	49254649	unknown	unknown	unknown

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87618258	1	49254955	unknown	unknown	unknown
AX-87131034	1	49256562	unknown	unknown	unknown
AX-86978019	1	49256763	unknown	unknown	unknown
AX-86929664	1	49288584	unknown	unknown	unknown
AX-87276333	1	49291893	unknown	unknown	unknown
AX-87556671	1	49309982	unknown	unknown	unknown
AX-87269284	1	49317397	100196704	emx2	empty spiracles homeobox 2
AX-87027805	1	49318851	100196704	emx2	empty spiracles homeobox 2
AX-87079676	1	49319181	100196704	emx2	empty spiracles homeobox 2
AX-87826315	1	49321592	100196704	emx2	empty spiracles homeobox 2
AX-87691363	1	49324976	100196704	emx2	empty spiracles homeobox 2
AX-87528991	1	49326117	100196704	emx2	empty spiracles homeobox 2
AX-87505311	1	49330788	100196704	emx2	empty spiracles homeobox 2
AX-87418125	1	49358964	106604689	eif3a	eukaryotic translation initiation factor 3 subunit A
AX-87262905	1	49365075	106604689	eif3a	eukaryotic translation initiation factor 3 subunit A
AX-87231287	1	49404244	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87564325	1	49405828	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87511068	1	49407304	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87853604	1	49408532	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87137786	1	49410273	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87503746	1	49410301	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87805964	1	49412476	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87417936	1	49416518	100196706	tm9s3	Transmembrane 9 superfamily member 3
AX-87264266	1	49438177	106605283	LOC106605283	uncharacterized LOC106605283
AX-87503203	1	49455689	106604730	LOC106604730	molybdenum cofactor sulfurase-like
AX-87625202	1	49546635	106605298	spata48	spermatogenesis associated 48
AX-87452395	1	49556786	106605298	spata48	spermatogenesis associated 48
AX-87720407	1	49564295	unknown	unknown	unknown
AX-87293248	1	49575898	100380614	ikzf1	IKAROS family zinc finger 1
AX-87313757	1	49593818	100380614	ikzf1	IKAROS family zinc finger 1
AX-87335532	1	49596163	100380614	ikzf1	IKAROS family zinc finger 1

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87273504	1	49596329	100380614	ikzf1	IKAROS family zinc finger 1
AX-87369499	1	49599989	100380614	ikzf1	IKAROS family zinc finger 1
AX-87255413	1	49600225	100380614	ikzf1	IKAROS family zinc finger 1
AX-86957335	1	49600305	100380614	ikzf1	IKAROS family zinc finger 1
AX-87671772	1	49600643	100380614	ikzf1	IKAROS family zinc finger 1
AX-87395630	1	49602974	100380614	ikzf1	IKAROS family zinc finger 1
AX-87577862	1	49603548	100380614	ikzf1	IKAROS family zinc finger 1
AX-87818304	1	49618181	106604764	fignl1	fidgetin like 1
AX-87100472	1	49637857	106604770	LOC106604770	tubulin--tyrosine ligase-like
AX-87550522	1	49664275	106604770	LOC106604770	tubulin--tyrosine ligase-like
AX-87431404	1	49679304	106604799	unknown	N-acetylneuraminic acid phosphatase
AX-87462400	1	49683479	unknown	unknown	unknown
AX-87362966	1	49683485	unknown	unknown	unknown
AX-87709697	1	49683814	unknown	unknown	unknown
AX-87313561	1	49690768	unknown	unknown	unknown
AX-87728385	1	49690870	unknown	unknown	unknown
AX-87317995	1	49691301	unknown	unknown	unknown
AX-87637743	1	49696309	unknown	unknown	unknown
AX-87771759	1	49697680	unknown	unknown	unknown
AX-87857866	1	49698014	unknown	unknown	unknown
AX-86955176	1	49698029	unknown	unknown	unknown
AX-87188235	1	49702074	106604807	LOC106604807	cytochrome P450 1B1-like
AX-87601000	1	49703954	106604807	LOC106604807	cytochrome P450 1B1-like
AX-87426384	1	49733099	106604807	LOC106604807	cytochrome P450 1B1-like
AX-87334866	1	49736375	106604807	LOC106604807	cytochrome P450 1B1-like
AX-87853126	1	49737044	106604820	LOC106604820	regulator of microtubule dyunknownmics protein 2-like
AX-87730551	1	49737945	106604820	LOC106604820	regulator of microtubule dyunknownmics protein 2-like
AX-87576202	1	49753100	106604820	LOC106604820	regulator of microtubule dyunknownmics protein 2-like
AX-87042393	1	49753626	106604820	LOC106604820	regulator of microtubule dyunknownmics protein 2-like
AX-86915328	1	49753808	106604820	LOC106604820	regulator of microtubule dyunknownmics protein 2-like
AX-87595785	1	49754498	106604820	LOC106604820	regulator of microtubule dyunknownmics protein 2-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87083660	1	49871938	unknown	unknown	unknown
AX-87589665	1	49873312	unknown	unknown	unknown
AX-87811957	1	49906012	106604835	LOC106604835	cdc42 effector protein 3-like
AX-87861562	1	49907745	106604835	LOC106604835	cdc42 effector protein 3-like
AX-86985188	1	49909537	106604835	LOC106604835	cdc42 effector protein 3-like
AX-87764739	1	49925476	unknown	unknown	unknown
AX-86972623	1	49943773	unknown	unknown	unknown
AX-86946779	1	49949891	unknown	unknown	unknown
AX-87516840	1	49962520	unknown	unknown	unknown
AX-86952759	1	49966134	unknown	unknown	unknown
AX-87506376	1	49990007	unknown	unknown	unknown
AX-87479837	1	49990755	unknown	unknown	unknown
AX-87481281	1	49991361	unknown	unknown	unknown
AX-86965447	1	49991624	unknown	unknown	unknown
AX-87495673	1	50044944	106604861	LOC106604861	kynurenine 3-monooxygeunknownse-like
AX-87754508	1	50052340	106604861	LOC106604861	kynurenine 3-monooxygeunknownse-like
AX-87117428	1	50066517	106604861	LOC106604861	kynurenine 3-monooxygeunknownse-like
AX-87429031	1	50066850	106604861	LOC106604861	kynurenine 3-monooxygeunknownse-like
AX-87617726	1	50071517	106604847	LOC106604847	insulin-degrading enzyme-like
AX-87764200	1	50072413	106604847	LOC106604847	insulin-degrading enzyme-like
AX-87198100	1	50093203	106604847	LOC106604847	insulin-degrading enzyme-like
AX-87240289	1	50109846	106604847	LOC106604847	insulin-degrading enzyme-like
AX-87129395	1	50125662	100195005	marh5	E3 ubiquitin-protein ligase MARCH5
AX-87238139	1	50141828	100195005	marh5	E3 ubiquitin-protein ligase MARCH5
AX-87586818	1	50144250	100195005	marh5	E3 ubiquitin-protein ligase MARCH5
AX-87600824	1	50151573	100195005	marh5	E3 ubiquitin-protein ligase MARCH5
AX-87609167	1	50156043	100195005	marh5	E3 ubiquitin-protein ligase MARCH5
AX-87141329	1	50193615	106604870	LOC106604870	cytoplasmic polyadenylation element-binding protein 3-like
AX-86998876	1	50206231	106604870	LOC106604870	cytoplasmic polyadenylation element-binding protein 3-like
AX-87014301	1	50207120	106604870	LOC106604870	cytoplasmic polyadenylation element-binding protein 3-like
AX-87627830	1	50207319	106604870	LOC106604870	cytoplasmic polyadenylation element-binding protein 3-like
AX-87690898	1	50211489	106604870	LOC106604870	cytoplasmic polyadenylation element-binding protein 3-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87858389	1	50239695	106604870	LOC106604870	cytoplasmic polyadenylation element-binding protein 3-like
AX-87075595	1	50266319	106604898	LOC106604898	TATA-binding protein-associated factor 172-like
AX-87179048	1	50293398	106604898	LOC106604898	TATA-binding protein-associated factor 172-like
AX-87400020	1	50311348	106604898	LOC106604898	TATA-binding protein-associated factor 172-like
AX-87571052	1	50388677	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87623562	1	50403133	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87229463	1	50433933	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87435707	1	50435234	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87643919	1	50446982	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87159923	1	50470122	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87581092	1	50481682	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87003641	1	50482152	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87479374	1	50505977	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87009219	1	50535255	106605322	LOC106605322	unknown nucleotidylexotransferase
AX-87866920	1	50600243	unknown	unknown	unknown
AX-87306066	1	50600253	unknown	unknown	unknown
AX-87011129	1	50612057	unknown	unknown	unknown
AX-86997915	1	50623127	unknown	unknown	unknown
AX-87204831	1	50660143	106604920	LOC106604920	heparan sulfate glucosamine 3-O-sulfotransferase 1-like
AX-87011816	1	50661390	106604920	LOC106604920	heparan sulfate glucosamine 3-O-sulfotransferase 1-like
AX-87195337	1	50703791	106604920	LOC106604920	heparan sulfate glucosamine 3-O-sulfotransferase 1-like
AX-87633850	1	50704617	106604920	LOC106604920	heparan sulfate glucosamine 3-O-sulfotransferase 1-like
AX-87679429	1	50733671	106605336	LOC106605336	uncharacterized LOC106605336
AX-87329264	1	50741326	106605336	LOC106605336	uncharacterized LOC106605336
AX-87544662	1	50781778	106604928	kndc1	kiunknownse non-catalytic C-lobe domain containing 1

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87734970	1	50785388	106604928	kndc1	kiunknownse non-catalytic C-lobe domain containing 1
AX-87663325	1	50787973	106604928	kndc1	kiunknownse non-catalytic C-lobe domain containing 1
AX-87179029	1	50864461	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87543264	1	50899430	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87233148	1	50901058	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87271720	1	50903366	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87344424	1	50918589	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87058951	1	50923922	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87629790	1	50969514	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-86977119	1	50993963	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87119374	1	50994313	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87072505	1	51062615	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87540230	1	51072941	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87477161	1	51073938	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87082091	1	51088646	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87345641	1	51145694	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87727357	1	51146064	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87367709	1	51146474	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87786884	1	51146683	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87682162	1	51151496	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87479277	1	51154704	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87355720	1	51154743	106604940	LOC106604940	adhesion G protein-coupled receptor A3-like
AX-87866183	1	51173533	unknown	unknown	unknown
AX-86962616	1	51191372	unknown	unknown	unknown
AX-87023609	1	51215190	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-86960332	1	51215708	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87640192	1	51216013	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87682443	1	51216249	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87011370	1	51217017	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87323723	1	51221719	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87794276	1	51235513	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87268697	1	51235692	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87821798	1	51235697	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87433791	1	51247144	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87476454	1	51256550	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87284797	1	51300442	106604947	LOC106604947	gamma-aminobutyric acid receptor subunit pi-like
AX-87339774	1	51338863	unknown	unknown	unknown
AX-87330607	1	51364044	unknown	unknown	unknown

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87310681	1	51379224	unknown	unknown	unknown
AX-86976975	1	51389895	106604960	cfap46	cilia and flagella associated protein 46
AX-87449071	1	51402261	106604960	cfap46	cilia and flagella associated protein 46
AX-87871177	1	51410757	106604960	cfap46	cilia and flagella associated protein 46
AX-87107583	1	51435389	106604960	cfap46	cilia and flagella associated protein 46
AX-87033624	1	51435526	106604960	cfap46	cilia and flagella associated protein 46
AX-87141449	1	51490070	106604960	cfap46	cilia and flagella associated protein 46
AX-87372258	1	51508754	unknown	unknown	unknown
AX-87564525	1	51569087	106604996	LOC106604996	vertebrate ancient opsin-like
AX-87544721	1	51594613	106604996	LOC106604996	vertebrate ancient opsin-like
AX-87183870	1	51642763	106605004	LOC106605004	homeobox protein Nkx-6.2-like
AX-87018674	1	51681338	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87366364	1	51710786	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87136475	1	51715845	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87093619	1	51720606	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87786996	1	51721116	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87319611	1	51732998	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87869588	1	51736852	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87810466	1	51737789	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87472926	1	51743778	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-86986868	1	51807705	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87153723	1	51807958	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87536984	1	51829193	106605012	LOC106605012	type I inositol 1,4,5-trisphosphate 5-phosphatase-like
AX-87591616	1	51880378	unknown	unknown	unknown
AX-87499735	1	51906720	106605019	LOC106605019	cysteine sulfinic acid decarboxylase-like
AX-87205442	1	51907801	106605019	LOC106605019	cysteine sulfinic acid decarboxylase-like
AX-86965298	1	51924392	106605019	LOC106605019	cysteine sulfinic acid decarboxylase-like
AX-87710990	1	51948575	106605046	pwwp2b	PWWP domain containing 2B
AX-87307914	1	51951939	106605046	pwwp2b	PWWP domain containing 2B
AX-87382255	1	51953625	106605046	pwwp2b	PWWP domain containing 2B
AX-87423299	1	51957640	106605046	pwwp2b	PWWP domain containing 2B
AX-87617649	1	51959376	106605046	pwwp2b	PWWP domain containing 2B
AX-87561026	1	51959423	106605046	pwwp2b	PWWP domain containing 2B
AX-87461582	1	51965164	106605046	pwwp2b	PWWP domain containing 2B
AX-87385710	1	51965724	106605046	pwwp2b	PWWP domain containing 2B
AX-87507045	1	51975596	106605343	LOC106605343	uncharacterized LOC106605343
AX-87052619	1	51976030	106605343	LOC106605343	uncharacterized LOC106605343
AX-87207362	1	51976625	106605343	LOC106605343	uncharacterized LOC106605343
AX-86930251	1	51986091	106605062	LOC106605062	serine/threonine-protein kiunknownse 32C-like
AX-87371428	1	52038843	106605062	LOC106605062	serine/threonine-protein kiunknownse 32C-like
AX-87335763	1	52063857	106605062	LOC106605062	serine/threonine-protein kiunknownse 32C-like
AX-87169031	1	52068317	106605062	LOC106605062	serine/threonine-protein kiunknownse 32C-like
AX-87290215	1	52107301	106605075	LOC106605075	serine racemase-like
AX-86935107	1	52128743	100286448	cj057	CJ057 protein
AX-87128145	1	52129282	100286448	cj057	CJ057 protein
AX-87313264	1	52130046	100286448	cj057	CJ057 protein

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87429422	1	52137559	100286448	cj057	CJ057 protein
AX-87073732	1	52137934	100286448	cj057	CJ057 protein
AX-87809928	1	52138536	100286448	cj057	CJ057 protein
AX-87269122	1	52141621	106605349	LOC106605349	homeobox protein OTX2-A-like
AX-87809620	1	52142891	106605349	LOC106605349	homeobox protein OTX2-A-like
AX-87313838	1	52142896	106605349	LOC106605349	homeobox protein OTX2-A-like
AX-87024309	1	52150240	100306744	rhoq	ras homolog gene family, member Q
AX-87148783	1	52243526	unknown	unknown	unknown
AX-87269320	1	52272605	106605358	LOC106605358	homeobox protein prophet of Pit-1-like
AX-87355995	1	52287450	106605097	LOC106605097	tetratricopeptide repeat protein 7A-like
AX-87272450	1	52308693	106605097	LOC106605097	tetratricopeptide repeat protein 7A-like
AX-87607257	1	52313026	106605097	LOC106605097	tetratricopeptide repeat protein 7A-like
AX-87674317	1	52335036	106605143	LOC106605143	calmodulin
AX-87315972	1	52335664	106605143	LOC106605143	calmodulin
AX-87041957	1	52344535	106605143	LOC106605143	calmodulin
AX-87296269	1	52365775	106605150	LOC106605150	uncharacterized LOC106605150
AX-87787930	1	52384363	106605174	LOC106605174	adhesion G protein-coupled receptor E2-like
AX-86941285	1	52429633	106604177	thada	THADA, armadillo repeat containing
AX-87724717	1	52437489	106604177	thada	THADA, armadillo repeat containing
AX-87259478	1	52450843	106604177	thada	THADA, armadillo repeat containing
AX-87285835	1	52451599	106604177	thada	THADA, armadillo repeat containing
AX-87713056	1	52458994	106604177	thada	THADA, armadillo repeat containing
AX-87560483	1	52459302	106604177	thada	THADA, armadillo repeat containing

SNP	Chr	Pos	GeneID	Symbol	Description
AX-86908740	1	52459396	106604177	thada	THADA, armadillo repeat containing
AX-87051785	1	52468702	106604177	thada	THADA, armadillo repeat containing
AX-87268019	1	52474219	106604177	thada	THADA, armadillo repeat containing
AX-87550278	1	52513445	unknown	unknown	unknown
AX-87227772	1	52516610	unknown	unknown	unknown
AX-87296126	1	52523760	unknown	unknown	unknown
AX-87069245	1	52529963	106605192	LOC106605192	zinc finger protein 36, C3H1 type-like 2-A
AX-87457406	1	52535262	106605192	LOC106605192	zinc finger protein 36, C3H1 type-like 2-A
AX-87215184	1	52535841	106605192	LOC106605192	zinc finger protein 36, C3H1 type-like 2-A
AX-87648379	1	52553290	unknown	unknown	unknown
AX-86924049	1	52604105	unknown	unknown	unknown
AX-87857208	1	52605990	unknown	unknown	unknown
AX-86975092	1	52665258	unknown	unknown	unknown
AX-87487372	1	52704138	unknown	unknown	unknown
AX-87295048	1	52704649	unknown	unknown	unknown
AX-87561168	1	52704715	unknown	unknown	unknown
AX-87620160	1	52718537	unknown	unknown	unknown
AX-87296127	1	52719121	unknown	unknown	unknown
AX-87394149	1	52720499	unknown	unknown	unknown
AX-87221833	1	52734696	unknown	unknown	unknown
AX-87269286	1	52760676	unknown	unknown	unknown
AX-87830893	1	52807120	unknown	unknown	unknown
AX-87290063	1	52881702	106605202	oxer1	oxoeicosanoid receptor 1
AX-87390158	1	52920219	unknown	unknown	unknown
AX-87322772	1	53017952	106605225	LOC106605225	cytochrome c oxidase subunit 7A-related protein, mitochondrial
AX-87200462	1	53017999	106605225	LOC106605225	cytochrome c oxidase subunit 7A-related protein, mitochondrial
AX-87465920	1	53026347	unknown	unknown	unknown
AX-87589210	1	53026652	unknown	unknown	unknown

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87164338	1	53032262	unknown	unknown	unknown
AX-86961211	1	53032853	unknown	unknown	unknown
AX-87693435	1	53033873	unknown	unknown	unknown
AX-87049326	1	53045991	unknown	unknown	unknown
AX-87350307	1	53050709	unknown	unknown	unknown
AX-87397804	1	53051022	unknown	unknown	unknown
AX-87425653	1	53051484	unknown	unknown	unknown
AX-87433537	1	53052383	unknown	unknown	unknown
AX-87250539	1	53063130	unknown	unknown	unknown
AX-87379010	1	53063542	unknown	unknown	unknown
AX-87426543	1	53063954	unknown	unknown	unknown
AX-87016163	1	53064737	unknown	unknown	unknown
AX-87263698	1	53068660	unknown	unknown	unknown
AX-87652608	1	53078100	unknown	unknown	unknown
AX-87136459	1	53085295	unknown	unknown	unknown
AX-87094720	1	53096484	unknown	unknown	unknown
AX-87135844	1	53165221	106605363	LOC106605363	nuclear factor 7, brain-like
AX-87102017	1	53166730	106605363	LOC106605363	nuclear factor 7, brain-like
AX-87002442	1	53173961	106605363	LOC106605363	nuclear factor 7, brain-like
AX-87715527	1	53198965	106605369	LOC106605369	uncharacterized LOC106605369
AX-87047896	1	53214973	106605231	eml4	echinoderm microtubule associated protein like 4
AX-87141554	1	53220670	106605231	eml4	echinoderm microtubule associated protein like 4
AX-87222267	1	53221602	106605231	eml4	echinoderm microtubule associated protein like 4
AX-87122091	1	53240797	106605231	eml4	echinoderm microtubule associated protein like 4
AX-87708703	1	53313434	unknown	unknown	unknown
AX-87812594	1	53336864	106608628	LOC106608628	uncharacterized LOC106608628
AX-87333357	1	53342700	106608628	LOC106608628	uncharacterized LOC106608628
AX-87089197	1	53361295	106608624	LOC106608624	uncharacterized LOC106608624
AX-87004842	1	53435951	unknown	unknown	unknown
AX-87101006	1	53477854	106605748	pkdcc	protein kiunknownse domain containing, cytoplasmic

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87860651	1	53478495	106605748	pkdcc	protein kiunknownse domain containing, cytoplasmic
AX-87719143	1	53479687	106605748	pkdcc	protein kiunknownse domain containing, cytoplasmic
AX-87009135	1	53484180	106605748	pkdcc	protein kiunknownse domain containing, cytoplasmic
AX-86965814	1	53488451	106605748	pkdcc	protein kiunknownse domain containing, cytoplasmic
AX-87169277	1	53522676	unknown	unknown	unknown
AX-87088607	1	53549467	unknown	unknown	unknown
AX-87436571	1	53646460	106605759	LOC106605759	E3 ubiquitin-protein ligase UBR2-like
AX-86952454	1	53659042	106605759	LOC106605759	E3 ubiquitin-protein ligase UBR2-like
AX-87106564	1	53659437	106605759	LOC106605759	E3 ubiquitin-protein ligase UBR2-like
AX-87600286	1	53663872	106605759	LOC106605759	E3 ubiquitin-protein ligase UBR2-like
AX-87202749	1	53667186	106605759	LOC106605759	E3 ubiquitin-protein ligase UBR2-like
AX-87412094	1	53678638	106605759	LOC106605759	E3 ubiquitin-protein ligase UBR2-like
AX-87333774	1	53688552	unknown	unknown	unknown
AX-87305275	1	53690253	unknown	unknown	unknown
AX-87799539	1	53691921	unknown	unknown	unknown
AX-87155720	1	53721259	unknown	unknown	unknown
AX-87630316	1	53750917	unknown	unknown	unknown
AX-87692156	1	53751369	unknown	unknown	unknown
AX-87584272	1	53756166	unknown	unknown	unknown
AX-87017100	1	53756217	unknown	unknown	unknown
AX-87015925	1	53794610	unknown	unknown	unknown
AX-87410434	1	53836578	106605778	LOC106605778	uncharacterized LOC106605778
AX-87409400	1	53839703	106605778	LOC106605778	uncharacterized LOC106605778
AX-86982104	1	53843537	106605778	LOC106605778	uncharacterized LOC106605778
AX-86980072	1	53869878	unknown	unknown	unknown
AX-87072445	1	53872013	unknown	unknown	unknown

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87848750	1	53872121	unknown	unknown	unknown
AX-87239730	1	53879658	unknown	unknown	unknown
AX-87640864	1	53880834	unknown	unknown	unknown
AX-87749388	1	53885696	unknown	unknown	unknown
AX-87248852	1	53885853	unknown	unknown	unknown
AX-87849003	1	53885913	unknown	unknown	unknown
AX-87578525	1	53885977	unknown	unknown	unknown
AX-86960945	1	53886950	unknown	unknown	unknown
AX-87590797	1	54413652	106605925	LOC106605925	uncharacterized LOC106605925
AX-87043685	1	54697058	106605994	LOC106605994	ninein-like protein
AX-87585898	1	54698409	106605994	LOC106605994	ninein-like protein
AX-87247215	1	54740857	106605568	LOC106605568	apoptosis-inducing factor 2-like
AX-87710564	1	54743253	106605568	LOC106605568	apoptosis-inducing factor 2-like
AX-87359293	1	54764000	106606050	LOC106606050	PH and SEC7 domain-containing protein 1-like
AX-87428840	1	54804095	106606050	LOC106606050	PH and SEC7 domain-containing protein 1-like
AX-87367919	1	54805483	106606050	LOC106606050	PH and SEC7 domain-containing protein 1-like
AX-87407468	1	54807217	106606050	LOC106606050	PH and SEC7 domain-containing protein 1-like
AX-87677526	1	54808286	106606050	LOC106606050	PH and SEC7 domain-containing protein 1-like
AX-86916127	1	54846725	106606070	LOC106606070	nuclear factor NF-kappa-B p100 subunit-like
AX-87606722	1	54969438	106606107	LOC106606107	multimerin-2-like
AX-87256974	1	54970216	106606107	LOC106606107	multimerin-2-like
AX-87204574	1	54977686	106606107	LOC106606107	multimerin-2-like
AX-87867918	1	54982368	106606107	LOC106606107	multimerin-2-like
AX-87539788	23	135206	106583870	LOC106583870	thyroid hormone receptor alpha
AX-87219041	23	542514	106583880	LOC106583880	vesicle-fusing ATPase-like
AX-87168852	23	585089	106583882	LOC106583882	26S proteasome non-ATPase regulatory subunit 11-like
AX-87678387	23	673290	unknown	unknown	unknown
AX-87402415	23	708176	unknown	unknown	unknown
AX-87707887	23	746105	106583889	LOC106583889	heme-binding protein 2-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87649060	23	1059399	106583883	LOC106583883	exportin-6-like
AX-87728904	23	1123112	106583886	LOC106583886	V-type proton ATPase 16 kDa proteolipid subunit
AX-87619830	23	1196542	106583929	LOC106583929	C-Jun-amino-termiunknownl kiunknownse-interacting protein 3-like
AX-87035916	23	1226976	106583929	LOC106583929	C-Jun-amino-termiunknownl kiunknownse-interacting protein 3-like
AX-87165759	23	1292593	100136929	hn1l	hematological and neurological expressed 1-like
AX-87279488	23	1292713	100136929	hn1l	hematological and neurological expressed 1-like
AX-87405554	23	1332471	106583928	LOC106583928	protein cramped-like
AX-87477347	23	1342474	106583928	LOC106583928	protein cramped-like
AX-87392189	23	1364105	106583928	LOC106583928	protein cramped-like
AX-87540092	23	1366373	106583928	LOC106583928	protein cramped-like
AX-87580282	23	1368432	106583928	LOC106583928	protein cramped-like
AX-87128349	23	1368750	106583928	LOC106583928	protein cramped-like
AX-87363946	23	1400897	106583926	LOC106583926	transmembrane protein 11, mitochondrial-like
AX-87003597	23	1402364	106583926	LOC106583926	transmembrane protein 11, mitochondrial-like
AX-87661903	23	1404244	106583926	LOC106583926	transmembrane protein 11, mitochondrial-like
AX-87426966	23	1406100	106583926	LOC106583926	transmembrane protein 11, mitochondrial-like
AX-87621978	23	1411446	106583926	LOC106583926	transmembrane protein 11, mitochondrial-like
AX-86963217	23	1425481	106583899	LOC106583899	uncharacterized LOC106583899
AX-87675354	23	1470237	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87282336	23	1470602	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87449636	23	1471573	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87196115	23	1471673	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87396853	23	1475496	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87843876	23	1476063	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87662799	23	1476922	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87149676	23	1477351	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87167834	23	1477420	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87315947	23	1481283	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87667920	23	1481364	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87690031	23	1481775	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87357712	23	1510834	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87167267	23	1510963	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-86971256	23	1512544	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87745867	23	1512805	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87660855	23	1513271	106583924	LOC106583924	ATP-sensitive inward rectifier potassium channel 12-like
AX-87674983	23	1524047	unknown	unknown	unknown
AX-87243846	23	1526853	unknown	unknown	unknown
AX-87172839	23	1545947	unknown	unknown	unknown
AX-87141856	23	1563339	unknown	unknown	unknown
AX-87024889	23	1599119	106583923	LOC106583923	U1 small nuclear ribonucleoprotein 70 kDa-like
AX-87564765	23	1599455	106583923	LOC106583923	U1 small nuclear ribonucleoprotein 70 kDa-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87524433	23	1600185	106583923	LOC106583923	U1 small nuclear ribonucleoprotein 70 kDa-like
AX-87141910	23	1609880	106583923	LOC106583923	U1 small nuclear ribonucleoprotein 70 kDa-like
AX-87267858	23	1615028	106583923	LOC106583923	U1 small nuclear ribonucleoprotein 70 kDa-like
AX-87573377	23	1616646	106583923	LOC106583923	U1 small nuclear ribonucleoprotein 70 kDa-like
AX-86946958	23	1666131	106583922	LOC106583922	protein arginine N-methyltransferase 1-like
AX-87630158	23	1667278	106583922	LOC106583922	protein arginine N-methyltransferase 1-like
AX-87858946	23	1700815	106583920	LOC106583920	AP-2 complex subunit alpha-2-like
AX-87265651	23	1722292	106583920	LOC106583920	AP-2 complex subunit alpha-2-like
AX-87257484	23	1722633	106583920	LOC106583920	AP-2 complex subunit alpha-2-like
AX-87021802	23	1766429	106583920	LOC106583920	AP-2 complex subunit alpha-2-like
AX-87250774	23	1794908	106583918	LOC106583918	ras-related protein R-Ras-like
AX-87803273	23	1818796	106583898	LOC106583898	transmembrane gamma-carboxyglutamic acid protein 2-like
AX-87791752	23	1871276	106583894	LOC106583894	glutamine and serine-rich protein 1-like
AX-87082790	23	1872756	106583894	LOC106583894	glutamine and serine-rich protein 1-like
AX-87854671	23	1924492	106583916	LOC106583916	splicing factor, arginine/serine-rich 19-like
AX-87810107	23	1937279	106583917	LOC106583917	uncharacterized LOC106583917
AX-87434124	23	1962796	106583893	LOC106583893	reticulocalbin-1-like
AX-87743413	23	1963291	106583913	LOC106583913	paired box protein Pax-6-like
AX-87232377	23	1977908	106583913	LOC106583913	paired box protein Pax-6-like
AX-87761720	23	1982031	unknown	unknown	unknown
AX-87094060	23	1998134	106583914	LOC106583914	tetraspanin-4-like
AX-87547695	23	2010085	106583914	LOC106583914	tetraspanin-4-like
AX-87657903	23	2010936	106583914	LOC106583914	tetraspanin-4-like
AX-87183862	23	2012936	106583914	LOC106583914	tetraspanin-4-like
AX-87247549	23	2014808	106583914	LOC106583914	tetraspanin-4-like
AX-87058348	23	2022223	106583912	LOC106583912	mucin-2-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87530118	23	2045671	106583911	LOC106583911	serine/threonine-protein phosphatase 4 catalytic subunit B
AX-87236129	23	2088897	106583908	LOC106583908	fructose-bisphosphate aldolase A-like
AX-87109089	23	2141917	106583907	LOC106583907	histone acetyltransferase KAT8-like
AX-87788904	23	2159507	106583907	LOC106583907	histone acetyltransferase KAT8-like
AX-87826390	23	2184177	100196413	thoc6	THO complex 6 homolog (Drosophila)
AX-87871624	23	2242030	106583905	LOC106583905	protein yippee-like 3
AX-87103270	23	2255827	106583905	LOC106583905	protein yippee-like 3
AX-87561113	23	2284630	106583904	LOC106583904	glycerophosphodiester phosphodiesterase domain-containing protein 3-like
AX-86983773	23	2284898	106583904	LOC106583904	glycerophosphodiester phosphodiesterase domain-containing protein 3-like
AX-87291109	23	2286570	106583904	LOC106583904	glycerophosphodiester phosphodiesterase domain-containing protein 3-like
AX-87421127	23	2295447	100380511	mk03	Mitogen-activated protein kiunknownse 3
AX-87197291	23	2337139	106583903	LOC106583903	uncharacterized LOC106583903
AX-86963524	23	2387584	106583902	LOC106583902	sarcoplasmic/endoplasmic reticulum calcium ATPase 1-like
AX-86956770	23	2401123	106583902	LOC106583902	sarcoplasmic/endoplasmic reticulum calcium ATPase 1-like
AX-87171179	23	2402515	106583902	LOC106583902	sarcoplasmic/endoplasmic reticulum calcium ATPase 1-like
AX-87852082	23	2403113	106583902	LOC106583902	sarcoplasmic/endoplasmic reticulum calcium ATPase 1-like
AX-87707929	23	2403800	106583902	LOC106583902	sarcoplasmic/endoplasmic reticulum calcium ATPase 1-like
AX-86963560	23	2423148	106583892	LOC106583892	uncharacterized LOC106583892
AX-87873005	23	2528489	106583936	LOC106583936	ataxin-2-like protein
AX-87394917	23	2562791	106583937	LOC106583937	protein spinster homolog 1-like
AX-87162229	23	2585819	106583937	LOC106583937	protein spinster homolog 1-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-86936045	23	2597934	106583937	LOC106583937	protein spinster homolog 1-like
AX-86988739	23	2600250	106583937	LOC106583937	protein spinster homolog 1-like
AX-87235620	23	2609764	106583937	LOC106583937	protein spinster homolog 1-like
AX-87840770	23	2616856	106583938	LOC106583938	SAGA-associated factor 29 homolog
AX-87856359	23	2623875	106583938	LOC106583938	SAGA-associated factor 29 homolog
AX-87681765	23	2695942	106583942	LOC106583942	tumor suppressor candidate 5 homolog
AX-86997998	23	2696235	106583942	LOC106583942	tumor suppressor candidate 5 homolog
AX-87734331	23	2696318	106583942	LOC106583942	tumor suppressor candidate 5 homolog
AX-87730744	23	2702164	106583942	LOC106583942	tumor suppressor candidate 5 homolog
AX-86983703	23	2735201	106600750	mvp	major vault protein
AX-87588621	23	2840885	106583946	LOC106583946	serine/threonine-protein kiunknownse TAO2-like
AX-87424133	23	2842289	106583946	LOC106583946	serine/threonine-protein kiunknownse TAO2-like
AX-86929741	23	2858347	106583946	LOC106583946	serine/threonine-protein kiunknownse TAO2-like
AX-87715672	23	2971259	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-87829317	23	2971764	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-86938612	23	2973012	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-87565281	23	2973121	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-87345844	23	2981584	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-87687901	23	3047638	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-87453570	23	3056396	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-86979919	23	3086157	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-87832206	23	3091438	106583948	LOC106583948	potassium voltage-gated channel subfamily C member 1-like
AX-87448518	23	3127337	106583950	LOC106583950	insulin receptor-like
AX-87303014	23	3147015	106583950	LOC106583950	insulin receptor-like
AX-87747144	23	3165792	106583950	LOC106583950	insulin receptor-like
AX-87482527	23	3271539	106583950	LOC106583950	insulin receptor-like
AX-87203511	23	3274942	106583950	LOC106583950	insulin receptor-like
AX-86980819	23	3307239	unknown	unknown	unknown
AX-87291944	23	3319262	unknown	unknown	unknown
AX-87699261	23	3323010	unknown	unknown	unknown
AX-87253127	23	3338754	unknown	unknown	unknown
AX-86954291	23	3417691	100136426	LOC100136426	vitellogenin
AX-86943695	23	3459425	106583931	LOC106583931	vitellogenin-like
AX-87709750	23	3498142	100380288	adgrl4	adhesion G protein-coupled receptor L4
AX-87139791	23	3517717	unknown	unknown	unknown
AX-86944514	23	3529429	100194783	zranb2	zinc finger RANBP2-type containing 2
AX-87668471	23	3529451	100194783	zranb2	zinc finger RANBP2-type containing 2
AX-87391909	23	3532197	100194783	zranb2	zinc finger RANBP2-type containing 2
AX-87620961	23	3576121	106583932	LOC106583932	rho guanine nucleotide exchange factor 18-like
AX-87563417	23	3706719	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87172133	23	3723113	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87292409	23	3723463	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87774963	23	3738783	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87421317	23	3746372	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-86948981	23	3769460	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87385859	23	3775926	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87528941	23	3779240	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87622440	23	3779339	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87311526	23	3779383	106583953	LOC106583953	WW domain-containing transcription regulator protein 1-like
AX-87032514	23	3820164	106583956	LOC106583956	putative gamma-glutamyltransferase YwrD
AX-87816187	23	3869003	106583958	LOC106583958	Hermansky-Pudlak syndrome 3 protein-like
AX-86934746	23	3878694	106583958	LOC106583958	Hermansky-Pudlak syndrome 3 protein-like
AX-87837058	23	3904497	106583959	LOC106583959	uncharacterized LOC106583959
AX-86906539	23	3961281	106583961	LOC106583961	discs large homolog 1-like protein
AX-87122104	23	3977274	106583961	LOC106583961	discs large homolog 1-like protein
AX-87734218	23	3982646	106583961	LOC106583961	discs large homolog 1-like protein
AX-87363159	23	3985738	106583961	LOC106583961	discs large homolog 1-like protein
AX-87692213	23	4014110	106583961	LOC106583961	discs large homolog 1-like protein
AX-87059097	23	4025505	106583961	LOC106583961	discs large homolog 1-like protein
AX-87133712	23	4031417	106583961	LOC106583961	discs large homolog 1-like protein
AX-87332387	23	4041871	106583961	LOC106583961	discs large homolog 1-like protein

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87575850	23	4050966	106583961	LOC106583961	discs large homolog 1-like protein
AX-87077911	23	4063772	106583961	LOC106583961	discs large homolog 1-like protein
AX-87843923	23	4080589	106583961	LOC106583961	discs large homolog 1-like protein
AX-87296787	23	4128208	106583961	LOC106583961	discs large homolog 1-like protein
AX-87180073	23	4132841	106583961	LOC106583961	discs large homolog 1-like protein
AX-86963912	23	4135460	106583961	LOC106583961	discs large homolog 1-like protein
AX-87626935	23	4148421	106583961	LOC106583961	discs large homolog 1-like protein
AX-87143526	23	4148734	106583961	LOC106583961	discs large homolog 1-like protein
AX-87586368	23	4199558	unknown	unknown	unknown
AX-86915111	23	4258938	106583965	LOC106583965	uncharacterized LOC106583965
AX-87707383	23	4320638	unknown	unknown	unknown
AX-86948789	23	4332754	100196375	phb2	prohibitin 2
AX-87486464	23	4340045	100196375	phb2	prohibitin 2
AX-87276963	23	4355130	100196375	phb2	prohibitin 2
AX-87721185	23	4424987	106583966	LOC106583966	anoctamin-8-like
AX-87219159	23	4425069	106583966	LOC106583966	anoctamin-8-like
AX-87099596	23	4425100	106583966	LOC106583966	anoctamin-8-like
AX-87194948	23	4427969	106583966	LOC106583966	anoctamin-8-like
AX-86950404	23	4429075	106583966	LOC106583966	anoctamin-8-like
AX-87650128	23	4464961	106583966	LOC106583966	anoctamin-8-like
AX-87500873	23	4465619	106583966	LOC106583966	anoctamin-8-like
AX-87144373	23	4474106	106583966	LOC106583966	anoctamin-8-like
AX-87269535	23	4575398	106583933	LOC106583933	DENN domain-containing protein 1A-like
AX-87292208	23	4577937	106583933	LOC106583933	DENN domain-containing protein 1A-like
AX-87715906	23	4581015	106583933	LOC106583933	DENN domain-containing protein 1A-like
AX-87672630	23	4613644	106583933	LOC106583933	DENN domain-containing protein 1A-like
AX-87000708	23	4619794	106583933	LOC106583933	DENN domain-containing protein 1A-like
AX-87831158	23	4646200	106583933	LOC106583933	DENN domain-containing protein 1A-like
AX-87841264	23	4692269	106583972	LOC106583972	cytosolic phospholipase A2-like
AX-86991035	23	4716702	106583972	LOC106583972	cytosolic phospholipase A2-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87046911	23	4741442	106583972	LOC106583972	cytosolic phospholipase A2-like
AX-87098467	23	4895587	unknown	unknown	unknown
AX-86904160	23	4976214	unknown	unknown	unknown
AX-87024138	23	5012051	unknown	unknown	unknown
AX-87512347	23	5021262	unknown	unknown	unknown
AX-87761868	23	5030740	unknown	unknown	unknown
AX-86936885	23	5067365	unknown	unknown	unknown
AX-87640402	23	5135402	unknown	unknown	unknown
AX-87342180	23	5135592	unknown	unknown	unknown
AX-87475301	23	5210426	unknown	unknown	unknown
AX-87378078	23	5274947	unknown	unknown	unknown
AX-87688496	23	5275719	unknown	unknown	unknown
AX-87860488	23	5276938	unknown	unknown	unknown
AX-87722525	23	5311117	unknown	unknown	unknown
AX-87098328	23	5339200	unknown	unknown	unknown
AX-87711662	23	5339257	unknown	unknown	unknown
AX-87043005	23	5361181	unknown	unknown	unknown
AX-86979596	23	5363795	unknown	unknown	unknown
AX-87735530	23	5373671	unknown	unknown	unknown
AX-87698333	23	5423378	unknown	unknown	unknown
AX-87514931	23	5467615	unknown	unknown	unknown
AX-87641924	23	5482049	unknown	unknown	unknown
AX-87322481	23	5593417	106583979	LOC106583979	protein crumbs homolog 1-like
AX-87061095	23	5593714	106583979	LOC106583979	protein crumbs homolog 1-like
AX-87255609	23	5606420	106583979	LOC106583979	protein crumbs homolog 1-like
AX-87569255	23	5614228	106583979	LOC106583979	protein crumbs homolog 1-like
AX-87754027	23	5637563	unknown	unknown	unknown
AX-87859525	23	5645115	unknown	unknown	unknown
AX-87317696	23	5659372	unknown	unknown	unknown
AX-86950577	23	5667430	100196385	nek7	NIMA related kiunknownse 7
AX-87405186	23	5668111	100196385	nek7	NIMA related kiunknownse 7
AX-87030070	23	5701999	100196385	nek7	NIMA related kiunknownse 7
AX-87033678	23	5771537	100196385	nek7	NIMA related kiunknownse 7
AX-87180902	23	5788310	100196385	nek7	NIMA related kiunknownse 7
AX-87640000	23	5843612	106583981	LOC106583981	LIM/homeobox protein Lhx9
AX-87228764	23	5848711	106583981	LOC106583981	LIM/homeobox protein Lhx9

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87781097	23	5849480	106583981	LOC106583981	LIM/homeobox protein Lhx9
AX-87420938	23	5849999	106583981	LOC106583981	LIM/homeobox protein Lhx9
AX-87181719	23	5869180	unknown	unknown	unknown
AX-87058306	23	5904813	unknown	unknown	unknown
AX-87775154	23	6016501	106583984	LOC106583984	beta-crystallin B1-like
AX-87229426	23	6025414	unknown	unknown	unknown
AX-87815151	23	6025979	unknown	unknown	unknown
AX-87510291	23	6026231	unknown	unknown	unknown
AX-87060100	23	6082057	unknown	unknown	unknown
AX-87577142	23	6084376	unknown	unknown	unknown
AX-87378314	23	6128507	106584002	LOC106584002	influenza virus NS1A-binding protein homolog A-like
AX-87355526	23	6188162	106584004	LOC106584004	uncharacterized LOC106584004
AX-87563205	23	6192303	106584004	LOC106584004	uncharacterized LOC106584004
AX-87382035	23	6268909	106584005	LOC106584005	protein Niban-like
AX-87358530	23	6290012	106584005	LOC106584005	protein Niban-like
AX-86992860	23	6292832	106584005	LOC106584005	protein Niban-like
AX-87257842	23	6294224	106584005	LOC106584005	protein Niban-like
AX-87046754	23	6296066	106584006	LOC106584006	kinesin-associated protein 3-like
AX-87020095	23	6308385	106584006	LOC106584006	kinesin-associated protein 3-like
AX-87355589	23	6361999	unknown	unknown	unknown
AX-87351005	23	6403286	unknown	unknown	unknown
AX-87275938	23	6430215	106583996	LOC106583996	uncharacterized LOC106583996
AX-86935810	23	6432811	106583996	LOC106583996	uncharacterized LOC106583996
AX-87020735	23	6439226	106583996	LOC106583996	uncharacterized LOC106583996
AX-87031221	23	6443787	106583996	LOC106583996	uncharacterized LOC106583996
AX-87823385	23	6449921	106584008	LOC106584008	paired mesoderm homeobox protein 1-like
AX-87202934	23	6476532	106584009	LOC106584009	myomegalin-like
AX-87133177	23	6481617	106584009	LOC106584009	myomegalin-like
AX-87835135	23	6482160	106584009	LOC106584009	myomegalin-like
AX-87718688	23	6482620	106584009	LOC106584009	myomegalin-like
AX-87372314	23	6487936	106584009	LOC106584009	myomegalin-like
AX-87095055	23	6488050	106584009	LOC106584009	myomegalin-like
AX-87773440	23	6488420	106584009	LOC106584009	myomegalin-like
AX-87322355	23	6541919	106584009	LOC106584009	myomegalin-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87067066	23	6569594	106584009	LOC106584009	myomegalin-like
AX-87499911	23	6578184	106584009	LOC106584009	myomegalin-like
AX-87764909	23	6624028	106583985	LOC106583985	uncharacterized LOC106583985
AX-87167429	23	6627387	106583985	LOC106583985	uncharacterized LOC106583985
AX-87321979	23	6673076	106583985	LOC106583985	uncharacterized LOC106583985
AX-87258995	23	6723863	106583985	LOC106583985	uncharacterized LOC106583985
AX-87341158	23	6845526	106583986	LOC106583986	microtubule-associated serine/threonine-protein kiunknownse 2-like
AX-87121416	23	6871572	106584010	LOC106584010	microtubule-associated serine/threonine-protein kiunknownse 2-like
AX-87858967	23	6881307	106584010	LOC106584010	microtubule-associated serine/threonine-protein kiunknownse 2-like
AX-87065267	23	6979396	unknown	unknown	unknown
AX-86919962	23	7022501	unknown	unknown	unknown
AX-87446529	23	7097593	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87290518	23	7097638	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87382864	23	7097771	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87567044	23	7140240	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-86946566	23	7200700	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87081392	23	7212631	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87660694	23	7275291	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87610776	23	7276303	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87466266	23	7325209	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87001250	23	7347807	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87542166	23	7355302	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87099584	23	7355967	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87102736	23	7373853	106584012	LOC106584012	phosphatidylinositol 3-kiunknownse regulatory subunit gamma
AX-87080789	23	7532221	unknown	unknown	unknown
AX-87275779	23	7565841	unknown	unknown	unknown
AX-87205031	23	7592209	106584015	pomgnt1	protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)
AX-86994688	23	7618583	106584015	pomgnt1	protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)
AX-87479239	23	7626251	106584015	pomgnt1	protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)
AX-87216830	23	7647999	100196736	pr38a	Pre-mRunknown-splicing factor 38A
AX-86952354	23	7651548	100196736	pr38a	Pre-mRunknown-splicing factor 38A
AX-87768363	23	7692319	106584018	LOC106584018	tyrosine-protein kiunknownse receptor Tie-1-like
AX-87634372	23	7708732	106584020	LOC106584020	leucine-rich repeat-containing protein 19-like
AX-87588997	23	7717330	106584020	LOC106584020	leucine-rich repeat-containing protein 19-like
AX-87303229	23	7718306	106584020	LOC106584020	leucine-rich repeat-containing protein 19-like
AX-87647406	23	7769735	106584021	LOC106584021	transmembrane protein 125-like
AX-87380720	23	7770501	106584021	LOC106584021	transmembrane protein 125-like
AX-87014585	23	7772607	106584021	LOC106584021	transmembrane protein 125-like
AX-87267205	23	7786900	106584022	LOC106584022	cilia- and flagella-associated protein 57-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87857558	23	7811908	100196728	ebp2	Probable rRunknown-processing protein EBP2
AX-87822321	23	7819275	100196728	ebp2	Probable rRunknown-processing protein EBP2
AX-87426009	23	7820606	106583998	fam183a	family with sequence similarity 183 member A
AX-86939073	23	7887557	unknown	unknown	unknown
AX-87155373	23	7896594	unknown	unknown	unknown
AX-86952061	23	7896930	unknown	unknown	unknown
AX-86983043	23	7900670	unknown	unknown	unknown
AX-87717911	23	7924840	unknown	unknown	unknown
AX-86976037	23	7925329	unknown	unknown	unknown
AX-87851810	23	7943113	unknown	unknown	unknown
AX-87176721	23	7958963	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87370700	23	7959363	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87246249	23	7961853	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87096043	23	7961883	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87686971	23	8011286	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87138861	23	8011329	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87654961	23	8022415	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87720497	23	8025569	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87834589	23	8028870	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87257464	23	8030941	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87373134	23	8031592	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-86905343	23	8071752	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87646081	23	8073598	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87520918	23	8075879	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87655916	23	8076158	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87497659	23	8117115	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87450172	23	8155612	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87496994	23	8206898	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87649580	23	8209668	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87291870	23	8210055	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-86938787	23	8216750	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87193372	23	8234131	106584023	LOC106584023	low-density lipoprotein receptor-related protein 8-like
AX-87134632	23	8244601	106584025	LOC106584025	KN motif and ankyrin repeat domain-containing protein 4-like
AX-87159858	23	8249242	106584025	LOC106584025	KN motif and ankyrin repeat domain-containing protein 4-like
AX-87077958	23	8302809	106584025	LOC106584025	KN motif and ankyrin repeat domain-containing protein 4-like
AX-87472417	23	8335671	106584025	LOC106584025	KN motif and ankyrin repeat domain-containing protein 4-like
AX-87465907	23	8362243	unknown	unknown	unknown
AX-87672154	23	8400201	106584026	LOC106584026	dedicator of cytokinesis protein 7
AX-87537431	23	8416415	106584026	LOC106584026	dedicator of cytokinesis protein 7
AX-87785210	23	8417289	106584026	LOC106584026	dedicator of cytokinesis protein 7
AX-87138762	23	8456810	106584026	LOC106584026	dedicator of cytokinesis protein 7
AX-87855450	23	8457144	106584026	LOC106584026	dedicator of cytokinesis protein 7

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87160816	23	8457615	106584026	LOC106584026	dedicator of cytokinesis protein 7
AX-87326614	23	8459414	106584026	LOC106584026	dedicator of cytokinesis protein 7
AX-86946835	23	8475884	106584026	LOC106584026	dedicator of cytokinesis protein 7
AX-87729540	23	8529397	unknown	unknown	unknown
AX-87020967	23	8540281	106583988	LOC106583988	cysteine protease ATG4C-like
AX-87741605	23	8546429	106583988	LOC106583988	cysteine protease ATG4C-like
AX-87075500	23	8589775	unknown	unknown	unknown
AX-87568664	23	8590688	unknown	unknown	unknown
AX-87279660	23	8594831	unknown	unknown	unknown
AX-86996867	23	8595305	unknown	unknown	unknown
AX-87478923	23	8596703	unknown	unknown	unknown
AX-87601259	23	8605786	unknown	unknown	unknown
AX-87820411	23	8607433	unknown	unknown	unknown
AX-87746782	23	8608167	unknown	unknown	unknown
AX-87562938	23	8651628	unknown	unknown	unknown
AX-87289236	23	8673446	unknown	unknown	unknown
AX-87578649	23	8676078	unknown	unknown	unknown
AX-87828605	23	8692466	unknown	unknown	unknown
AX-87623786	23	8728332	unknown	unknown	unknown
AX-87669098	23	8734942	unknown	unknown	unknown
AX-87350342	23	8745200	unknown	unknown	unknown
AX-87722151	23	8788481	106584029	LOC106584029	forkhead box protein D3-B-like
AX-87159931	23	8796386	106584029	LOC106584029	forkhead box protein D3-B-like
AX-87423357	23	8803104	106584029	LOC106584029	forkhead box protein D3-B-like
AX-87364900	23	8803121	106584029	LOC106584029	forkhead box protein D3-B-like
AX-87104073	23	8808338	106583989	LOC106583989	dolichyl pyrophosphate Man9Glcunknownc2 alpha-1,3-glucosyltransferase-like
AX-87048770	23	8815863	106583989	LOC106583989	dolichyl pyrophosphate Man9Glcunknownc2 alpha-1,3-glucosyltransferase-like
AX-87674457	23	8815923	106583989	LOC106583989	dolichyl pyrophosphate Man9Glcunknownc2 alpha-1,3-glucosyltransferase-like

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87380133	23	8864955	106583989	LOC106583989	dolichyl pyrophosphate Man9Glcunknownc2 alpha-1,3-glucosyltransferase-like
AX-87765392	23	8865101	106583989	LOC106583989	dolichyl pyrophosphate Man9Glcunknownc2 alpha-1,3-glucosyltransferase-like
AX-87084327	23	8892813	106583989	LOC106583989	dolichyl pyrophosphate Man9Glcunknownc2 alpha-1,3-glucosyltransferase-like
AX-87452258	23	8897344	100194572	pgm1	phosphoglucomutase 1
AX-87821095	23	8897500	100194572	pgm1	phosphoglucomutase 1
AX-87442701	23	8897912	100194572	pgm1	phosphoglucomutase 1
AX-87203583	23	8903091	100194572	pgm1	phosphoglucomutase 1
AX-86965558	23	8904057	100194572	pgm1	phosphoglucomutase 1
AX-87460091	23	8925831	106584030	LOC106584030	tyrosine-protein kiunknownse transmembrane receptor ROR1-like
AX-86932938	23	8942915	106584030	LOC106584030	tyrosine-protein kiunknownse transmembrane receptor ROR1-like
AX-87640991	23	8954328	106584030	LOC106584030	tyrosine-protein kiunknownse transmembrane receptor ROR1-like
AX-87132031	23	8992797	106584030	LOC106584030	tyrosine-protein kiunknownse transmembrane receptor ROR1-like
AX-87157553	23	9061272	106584030	LOC106584030	tyrosine-protein kiunknownse transmembrane receptor ROR1-like
AX-86964847	23	9067562	106584030	LOC106584030	tyrosine-protein kiunknownse transmembrane receptor ROR1-like
AX-87222244	23	9118667	106584030	LOC106584030	tyrosine-protein kiunknownse transmembrane receptor ROR1-like
AX-87052583	23	9150250	unknown	unknown	unknown
AX-87611731	23	9181175	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-87601958	23	9181641	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-86985295	23	9182360	106584031	LOC106584031	VWFA and cache domain-containing protein 1

SNP	Chr	Pos	GeneID	Symbol	Description
AX-87141919	23	9186127	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-87720742	23	9237683	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-87429173	23	9240944	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-87238836	23	9244174	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-87526284	23	9252352	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-86917011	23	9260192	106584031	LOC106584031	VWFA and cache domain-containing protein 1
AX-87406555	23	9289476	106584032	LOC106584032	ribonucleoprotein PTB-binding 2-like
AX-87076476	23	9289501	106584032	LOC106584032	ribonucleoprotein PTB-binding 2-like
AX-86969316	23	9290852	106584032	LOC106584032	ribonucleoprotein PTB-binding 2-like
AX-87026869	23	9324965	106584032	LOC106584032	ribonucleoprotein PTB-binding 2-like
AX-87695297	23	9331327	106584032	LOC106584032	ribonucleoprotein PTB-binding 2-like
AX-87500541	23	9334293	106584032	LOC106584032	ribonucleoprotein PTB-binding 2-like
AX-87846743	23	9420549	106584035	LOC106584035	putative tyrosine-protein phosphatase auxilin
AX-87322747	23	9420844	106584035	LOC106584035	putative tyrosine-protein phosphatase auxilin
AX-87344479	23	9422569	106584035	LOC106584035	putative tyrosine-protein phosphatase auxilin
AX-87568882	23	9432443	106584035	LOC106584035	putative tyrosine-protein phosphatase auxilin
AX-87782899	23	9434644	106584035	LOC106584035	putative tyrosine-protein phosphatase auxilin

Appendix I – Identity by Descent

Table S15 Preliminary analysis of within river relatedness among Atlantic Salmon (*Salmo salar*) in 26 rivers in Placentia Bay, Newfoundland, Canada. Identity by descent (IBD) calculated using PLINK v 1.9 (Chang et al., 2015). Family ID (FID), individual ID (IID), relationship type (RT), IBD sharing value (EZ), proportion of IBD (PI_HAT) for each year (2017 and 2018) sampled.

FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2	PI_HAT	YEAR
BDL	BDL_1_1_14	BDL	BDL_1_1_26	OT	0	0.2495	0.3572	0.3933	0.5719	2017
BRA	BRA_2_1_18	BRA	BRA_2_1_27	OT	0	0.2908	0.4198	0.2895	0.4993	2017
CBC	CBC_1_1_29	CBC	CBC_1_1_8	OT	0	0.4128	0.1975	0.3898	0.4885	2017
CUS	CUS_1_1_10	CUS	CUS_1_1_22	OT	0	0.3258	0.2407	0.4335	0.5539	2017
FHB	FHB_1_1_13	FHB	FHB_1_1_20	OT	0	0.137	0.4369	0.4262	0.6446	2017
FHB	FHB_1_1_13	FHB	FHB_1_1_3	OT	0	0.1737	0.5483	0.278	0.5521	2017
FHB	FHB_1_1_13	FHB	FHB_1_1_6	OT	0	0.216	0.4585	0.3255	0.5547	2017
FHB	FHB_1_1_15	FHB	FHB_1_1_17	OT	0	0.2907	0.3207	0.3887	0.549	2017
FHB	FHB_1_1_20	FHB	FHB_1_1_3	OT	0	0.3163	0.3601	0.3236	0.5036	2017
FHB	FHB_1_1_20	FHB	FHB_1_1_6	OT	0	0.1837	0.4818	0.3345	0.5754	2017
FHB	FHB_1_1_3	FHB	FHB_1_1_6	OT	0	0.2105	0.58	0.2095	0.4995	2017
FHB	FHB_1_1_4	FHB	FHB_1_1_8	OT	0	0.2696	0.536	0.1943	0.4624	2017
GBW	GBW_1_YOY_14	GBW	GBW_1_YOY_6	OT	0	0.2558	0.4609	0.2833	0.5138	2017
LBB	LBB_1_2_1	LBB	LBB_1_2_8	OT	0	0.2618	0.3769	0.3613	0.5497	2017
LWN	LWN_1_1_10	LWN	LWN_1_1_4	OT	0	0.1762	0.5126	0.3113	0.5675	2017
LWN	LWN_1_1_11	LWN	LWN_1_1_12	OT	0	0.3056	0.3621	0.3323	0.5133	2017
LWN	LWN_1_1_16	LWN	LWN_1_1_6	OT	0	0.2411	0.4735	0.2853	0.5221	2017
LWN	LWN_1_1_3	LWN	LWN_1_1_32	OT	0	0.1551	0.431	0.4139	0.6294	2017
NMB	NMB_11_3	NMB	NMB_11_9	OT	0	0.2719	0.4739	0.2542	0.4912	2017
NON	NON_1_1_11	NON	NON_1_1_25	OT	0	0.2927	0.3573	0.35	0.5287	2017
NPR	NPR_1_1_13	NPR	NPR_1_1_16	OT	0	0.3198	0.3998	0.2805	0.4803	2017
PBR	PBR_1_1_29	PBR	PBR_1_1_9	OT	0	0.2418	0.4018	0.3564	0.5573	2017

FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2	PI_HAT	YEAR
RHW	RHW_1_1_15	RHW	RHW_1_1_3	OT	0	0.2402	0.5123	0.2475	0.5037	2017
SHI	SHI_1_1_10	SHI	SHI_1_1_12	OT	0	0.2273	0.5602	0.2126	0.4926	2017
SHI	SHI_1_1_25	SHI	SHI_1_1_26	OT	0	0.1906	0.5336	0.2758	0.5426	2017
SPR	SPR_1_1_1	SPR	SPR_1_1_20	OT	0	0.1873	0.473	0.3397	0.5762	2017
SPR	SPR_1_1_15	SPR	SPR_1_1_18	OT	0	0.2418	0.493	0.2652	0.5117	2017
SPR	SPR_1_1_15	SPR	SPR_1_1_25	OT	0	0.244	0.4721	0.2839	0.5199	2017
SPR	SPR_1_1_18	SPR	SPR_1_1_25	OT	0	0.2541	0.4616	0.2843	0.5151	2017
SPR	SPR_1_1_21	SPR	SPR_1_1_26	OT	0	0.1883	0.538	0.2737	0.5427	2017
SPR	SPR_1_1_21	SPR	SPR_1_1_28	OT	0	0.232	0.5125	0.2554	0.5117	2017
SPR	SPR_1_1_21	SPR	SPR_1_1_4	OT	0	0.2195	0.506	0.2745	0.5275	2017
SPR	SPR_1_1_26	SPR	SPR_1_1_28	OT	0	0.2454	0.4305	0.3241	0.5394	2017
SPR	SPR_1_1_26	SPR	SPR_1_1_4	OT	0	0.2729	0.544	0.183	0.455	2017
SPR	SPR_1_1_28	SPR	SPR_1_1_4	OT	0	0.2979	0.4858	0.2164	0.4593	2017
BDL	BDL_1_1_9	BDL	BDL_1_1_30	OT	0	0.2134	0.4805	0.3061	0.5463	2018
BLA	BLA_1_1_8	BLA	BLA_1_1_25	OT	0	0.2972	0.4207	0.2821	0.4925	2018
BRA	BRA_2_1_14	BRA	BRA_2_1_32	OT	0	0.2576	0.4572	0.2851	0.5138	2018
BRA	BRA_2_1_16	BRA	BRA_2_1_34	OT	0	0.2489	0.5426	0.2085	0.4798	2018
CUS	CUS_1_2_11	CUS	CUS_1_2_17	OT	0	0.2419	0.4129	0.3452	0.5517	2018
FHB	FHB_1_2_20	FHB	FHB_1_2_25	OT	0	0.2577	0.4605	0.2818	0.512	2018
GBW	GBW_1_1_4	GBW	GBW_1_1_12	OT	0	0.3011	0.3362	0.3627	0.5308	2018
GBW	GBW_1_1_5	GBW	GBW_1_1_7	OT	0	0.2489	0.451	0.3001	0.5256	2018
GBW	GBW_1_1_8	GBW	GBW_1_1_13	OT	0	0.1943	0.5295	0.2761	0.5409	2018
GBW	GBW_1_1_8	GBW	GBW_1_1_24	OT	0	0.0944	0.5386	0.367	0.6363	2018
GBW	GBW_1_1_8	GBW	GBW_1_1_26	OT	0	0.3486	0.366	0.2854	0.4684	2018
GBW	GBW_1_1_13	GBW	GBW_1_1_24	OT	0	0.2209	0.3804	0.3987	0.5889	2018
GBW	GBW_1_1_13	GBW	GBW_1_1_26	OT	0	0.1431	0.6176	0.2393	0.5481	2018
GBW	GBW_1_1_13	GBW	GBW_1_1_28	OT	0	0.3207	0.3842	0.295	0.4872	2018
GBW	GBW_1_1_24	GBW	GBW_1_1_26	OT	0	0.1709	0.5005	0.3286	0.5789	2018
GBW	GBW_1_1_26	GBW	GBW_1_1_28	OT	0	0.2874	0.4419	0.2708	0.4917	2018
GBW	GBW_1_2_5	GBW	GBW_1_2_16	OT	0	0.3716	0.2778	0.3506	0.4895	2018

FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2	PI_HAT	YEAR
LWN	LWN_1_1_7	LWN	LWN_1_2_4	OT	0	0.2443	0.4404	0.3153	0.5355	2018
NHR	NHR_1_2_9	NHR	NHR_1_2_14	OT	0	0.1384	0.476	0.3855	0.6235	2018
NMB	NMB_1_1_10	NMB	NMB_1_1_20	OT	0	0.2496	0.4506	0.2998	0.5251	2018
NON	NON_1_1_6	NON	NON_1_1_17	OT	0	0.1458	0.612	0.2422	0.5482	2018
NPR	NPR_1_1_23	NPR	NPR_1_2_1	OT	0	0.2751	0.4574	0.2675	0.4962	2018
SPR	SPR_1_1_4	SPR	SPR_1_2_11	OT	0	0.2939	0.4979	0.2082	0.4571	2018
SPR	SPR_1_1_4	SPR	SPR_2_2_1	OT	0	0.2585	0.4771	0.2644	0.503	2018
SPR	SPR_1_1_4	SPR	SPR_2_2_7	OT	0	0.243	0.3603	0.3968	0.5769	2018
SPR	SPR_1_2_11	SPR	SPR_2_1_9	OT	0	0.1531	0.4998	0.3471	0.597	2018
SPR	SPR_1_2_11	SPR	SPR_2_2_2	OT	0	0.1808	0.4864	0.3328	0.576	2018
SPR	SPR_2_1_4	SPR	SPR_2_2_2	OT	0	0.2918	0.4816	0.2266	0.4674	2018
SPR	SPR_2_1_4	SPR	SPR_2_2_7	OT	0	0.2397	0.5301	0.2303	0.4953	2018
SPR	SPR_2_1_9	SPR	SPR_2_2_2	OT	0	0.1764	0.5248	0.2988	0.5612	2018
SPR	SPR_2_1_12	SPR	SPR_2_2_3	OT	0	0.1598	0.4856	0.3545	0.5973	2018
SPR	SPR_2_1_14	SPR	SPR_2_2_4	OT	0	0.303	0.3999	0.2971	0.4971	2018
SPR	SPR_2_2_1	SPR	SPR_2_2_7	OT	0	0.2042	0.5364	0.2595	0.5277	2018