Measurement of Population Structure of Cod (*Gadus morhua*) in NAFO Subdivision 3Ps off Southern Newfoundland by Means of Next-Generation Sequencing and Genomic Analysis

by

Sarah Babaei

HBSc Molecular Biology and Genetics, McMaster University, 2022

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the Graduate Academic Unit of Biology

Supervisors: Scott Pavey, Ph.D, Dept. of Biological Sciences

Cassidy D'Aloia, Ph.D, Dept. of Ecology and Evolutionary Biology, University of Toronto

Examining Board: Rémy Rochette, Ph.D, Dept. of Biological Sciences
Divya Varkey, Ph.D, Fisheries and Oceans Canada
Steven M. Carr, Ph.D, Dept. of Biology, Memorial University of Newfoundland

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

March 2024

ABSTRACT

Once Canada's biggest fishery, Atlantic Cod (*Gadus morhua)* stocks in the Northwest Atlantic collapsed in the 1990s, with recovery strategies proving unsuccessful. I focus on NAFO subdivision 3Ps off Newfoundland's south coast. Tagging and genetic studies suggest 3Ps is a mixed stock of unknown composition. Greater power is required to see how 3Ps cod fits within Northwest Atlantic cod, which we address using Single Nucleotide Polymorphisms (SNPs). I compared 238 fish in 3Ps to three genetic clusters (Northern (NAFO 2J3KL), Saint Lawrence (NAFO 4SRT), Southern populations (NAFO 4X, 5YZ)) at 25,458 neutral SNPs. Supergene allele frequencies in juveniles and adults were also compared. I provide evidence for genetic similarity between 3Ps and Northern cod and evidence of two potential Northern cod subpopulations. Lastly, I observed decreasing supergene allele frequencies as 3Ps cod age. For species like cod with complex population structure, genomics helps understand population structure to inform recovery strategies.

ACKNOWLEDGEMENTS

Thank you to my supervisory committee Dr. Scott Pavey, Dr. Cassidy D'Aloia, Dr. Divya Varkey, and Dr. Rémy Rochette for their support, advice, and guidance. Thank you to my external examiner Dr. Steven Carr. Thank you to Dr. Nathalie LeBlanc for training me, advising me, and helping me be a better researcher throughout this degree.

I would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), the University of New Brunswick, The Canada Research Chairs (CRC), Fisheries and Oceans Canada, and the New Brunswick Innovation Foundation for funding.

Thank you to Fisheries and Oceans Canada, Dr. Sherrylynn Rowe, and Dr. Greg Puncher for providing samples for this project and to Genome Quebec for sequencing. Thank you to the cod that provided their scales.

Thank you to other members of the Pavey lab, including Larissa Roehl, Shawn Kroetsch, Jagger Watters-Gray, Quinn Carvey, Rina Guxholli, and Jake Reicker for their advice with presentations and help with lab work. Thank you to Dr. Aaron Adamack and Rick Rideout for support and advice with analyzing and interpreting my genetic and sampling data.

Thank you to my partner Benny for your endless support.

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

NAFO: North Atlantic Fisheries Organization

NGS: Next-Generation Sequencing

SNPs: Single Nucleotide Polymorphisms

LG: Linkage Group

DNA: Deoxyribonucleic Acid

ddRADseq: Double-Digest Restriction Site-Associated DNA Sequencing

TL: Lysis Buffer

BWA: Burrows-Wheeler Aligner

F_{IS}: Inbreeding Coefficient

minDP: Minimum Read Depth

maf: Minor Allele Frequency

PCA: Principal Components Analysis

DAPC: Discriminant Analysis of Principal Components

F_{ST}: Fixation Index

HWP: Hardy-Weinberg Proportions

BIC: Bayesian Information Criterion

AIC: Akaike's Information Criterion

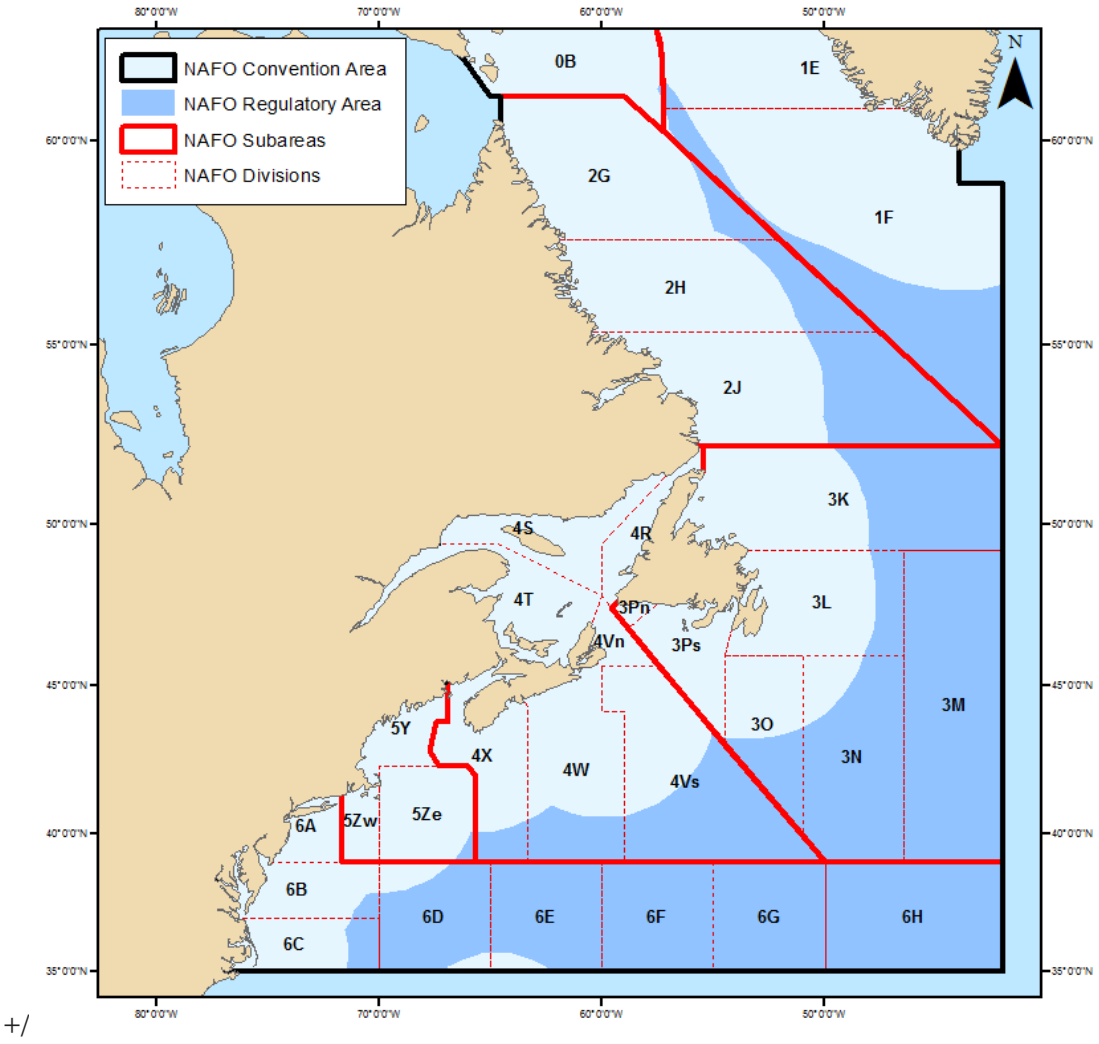
CI: Confidence Interval

1 Measurement of population structure of cod (*Gadus morhua*) in NAFO Subdivision 3Ps off southern Newfoundland by means of Next-Generation Sequencing and genomic analysis

Introduction

1.1 Canadian Cod Stocks in the North-West Atlantic

Atlantic cod *(Gadus morhua)* are groundfish with a population distribution across the North Atlantic (Lilly et al. 2008). Cod in Canada are managed in 12 divisions designated by the North Atlantic Fisheries Organization (NAFO) that extend from coastal waters off northern Labrador to southern New Brunswick, including the Gulf of Saint Lawrence (Rose & Rowe 2015; DFO 2021; Figure 1).



+/

Figure 1: NAFO Areas. Source: ©Northwest Atlantic Fisheries Organization and Fisheries and Oceans Canada. Northern Cod comprise fish in are 2J, 3K, and 3L. Gulf of Saint Lawrence fish include 4S, 4R, and 4T. Areas south of the Laurentian Channel include Bay of Fundy (4X) and Gulf of Maine (5YZ).

Northern Cod, located off the coast of Newfoundland and Labrador, at one time supported an abundant and important fishery resource in Canada before its collapse leading to closure in 1992 (Rose & Rowe 2015). Canadian cod were discovered as a food

source in the 1500s by Europeans. Once one of the world's largest marine fisheries, overexploitation, resource mismanagement, and environmental changes led to a depletion of the natural resource (Rose & Rowe 2015; Sobel 1996). Changes included smaller size at age and maturity at younger ages (Chen & Mello 1999; Brattey & Morgan 1996).

Cod have been important culturally. Cod were an important part of the diets of Indigenous peoples such as the Innu, Beothuk, and Mi'kmaw peoples long before European discovery (Castañeda et al. 2020)., The abundance of cod led to an influx of immigrants to Newfoundland and Labrador, many of whom made their living fishing and selling cod. Cod became ingrained into Newfoundland and Labradorian culture as fishermen passed down their professions to future generations. The decline and collapse of the Northern Cod stock in the 1980 ~ 90s resulted in the largest layoff in Canadian history: as much as 10% of Newfoundland's population emigrated from the province (Bavington 2011). Cod remain a key cultural aspect of Canada's east coast (Rose 2007).

The Federal government placed a moratorium on commercial cod that was in place from 1992-1999, where spawning stock biomass in 3Ps was estimated to be 66kt (DFO 2024). However, stocks continued to decline (Lilly et al. 2008). After some apparent recovery in 1997 (spawning stock biomass of <100kt) subdivision 3Ps, found off the southern coast of Newfoundland, was the first area to be reopened, but it remains in the critical zone to this day (2024 estimated spawning stock biomass is 35.5kt) (Lilly et al. 2008; DFO 2022; DFO 2024).

Two reasons have been suggested to explain the collapse: overfishing and changes in climate, specifically freezing on the Northwest Atlantic and decrease in important prey (capelin; Lilly et al. 2008; Myers & Hutchings 1997). Overfishing caused

a decline in the average age and size at age (Lilly et al. 2008). However, despite bans on fishing, the southern Newfoundland cod – including those in 3Ps - did not recover as hoped (Lilly et al. 2008). Thus, overfishing may not have been the sole factor behind the collapse. Secondly, environmental drivers in of the North Atlantic are proposed to be part of the reason the stocks collapsed (Lilly et al. 2008). Lower ocean temperatures and subsequent ice cover in the Northwest Atlantic from the 1960-90s caused decreased productivity in the Northern cod stock. Decrease in prey may have also played a role, since capelin numbers also dwindled (Drinkwater 2005; Lilly et al. 2008).

1.2 Atlantic Cod Spawning

Cod numbers are still dwindling. only recently have Northern cod been declared out of the critical zone (Kennedy 2023). This may be because we continue to deplete cod from their spawning locations (Bui et al. 2011). Atlantic cod spawn by gathering in large groups once or twice a year , for as long as 15-20 years (Brander 1994; Hutchings & Rangeley 2011). When spawning, female cod release batches of eggs every 4-7 days.

Identification of locations of cod spawning grounds is an active area of research, the most common methods being molecular, tagging, or egg density/distribution techniques. Years of tagging data suggest that individual cod may return to a particular spawning location year after year, a behaviour called ‘homing’ (Robichaud & Rose 2001). For example, Ouellet et al. (1997) used acoustic and bottom trawl surveys in combination with ichthyoplankton distribution surveys to determine whether spawning occurs in the Northern Gulf of Saint Lawrence. From 1993-1995, they found the majority

of fish were spawning or spent in late March to early April. Egg abundance, sampled from the Northern Gulf using net sampling techniques, also peaked during this time, confirming that cod were spawning.

Other locations off the Newfoundland coast have also been described as potential spawning locations, including its Southern coast. For example, within division 3Ps, spawning has been observed in Placentia Bay, Fortune Bay, Burgeo Bank, Halibut Channel, and on the St. Pierre Bank (Robichaud & Rose 2001). Northern cod spawn in other areas as well, including the Grand Banks (NAFO 3L), Belle Isle Banks (NAFO 3K), and Hamilton Bank (NAFO 2J) (Myers et al. 1993).

1.3 Population structure of Northern Cod

Elucidation of the genetic population structure of Northern Cod is an active area of research, with much discussion of the alternative 'isolation' or 'metapopulation' hypotheses. The 'isolation' hypothesis is that the Northern stocks do not interbreed and thus are genetically isolated from each other, (Rose et al. 2011). Isolationist views also suggest the existence of 'bay stocks', where inshore cod in various deep-water bays remain localized and thus are genetically isolated from one another (Beacham et al. 2002; Wroblewski et al. 2005). An example of 'Bay Stock' is Gilbert Bay cod, whose existence has been studied by tracking data, preliminary genetic studies, and modern large-scale genomic studies (Green & Wroblewski 2000; Beacham et al. 2002; Sinclair-Waters et al. 2018). These findings and 'isolationist' views pushed the independent management of 3Ps cod (Lawson & Rose 2000).

In contrast, the ‘metapopulation’ hypothesis argues for the presence of multiple population components that mix and interbreed depending on season or other environmental factors, making overall recovery more likely (Hanski & Simberloff 1997; Rose et al. 2011). Modern studies have found more support for the ‘metapopulation’ hypothesis in the Northern cod stocks, especially with large-scale genomic technology (Rose et al. 2011). Today, most literature focuses on the ‘metapopulation’ hypothesis, with some papers touching on Gilbert Bay cod as a rather unusual example of an isolated genetic population.

1.4 Atlantic Cod Connectivity

Studies that aim to elucidate broad-scale patterns of population structure within Canadian cod stocks have adopted various methods. Prior to the development of modern genetic and genomic tools, tag-and-recapture was a popular method for studying individual movement and population structure. Templeman (1974) investigated year-round movement of Atlantic cod in the Newfoundland and Labrador area with tagging and recapture. Cod tagged in southern Strait of Belle Isle (NAFO 4R), migrated in winter through the Gulf of Saint Lawrence to the Cabot Strait (NAFO 4Vn) and eastward autumn migration went through the Strait of Belle Isle into the Hamilton Bank (NAFO 2J). Winter migration occurred with most individuals moving south to the Cabot Strait. Overall, cod tagged on Newfoundland’s east coast (NAFO 3L & 3K), were found to mix with individuals off the coast of Southern Labrador (NAFO 4S) and subdivision 3Ps. Tagged individuals within 3Ps were recaptured during spring and summer in Northern

and Southern Grand Bank (NAFO 3LNO), as well as St. Pierre Bank (NAFO 3Ps). Overall, Templeman's tagging study (1974) displays strong evidence for intermingling of cod in the Gulf of Saint Lawrence, Eastern Newfoundland, and Southern Newfoundland waters (NAFO 4R, 3KL, and 3Ps, respectively). This study also showed limited mixing between the Gulf of Saint Lawrence-Eastern Newfoundland-Southern Newfoundland cod and cod south of the Laurentian channel, potentially due to the channel's deep waters.

Previous genetic studies have also described connectivity among cod, using far less genetic markers than modern genomic studies.Ruzzante et al. (2000) conducted a mixed-stock analysis of pre-, post-, and spawning Gulf of Saint Lawrence cod with six microsatellite loci to measure the proportional contribution of each geographical region to the overwintering cod in the region. Highest contributions were from waters off southern Gulf of Saint Lawrence (NAFO 4T), southeast and central Newfoundland, Cape Breton Island region, and 3Ps bays (Fortune and Placentia Bays), respectively. Genetic distances among groups were also investigated, with the highest among southern Gulf of Saint Lawrence (NAFO 4RT) and southern Newfoundland waters (NAFO 4Vn, 3Pn, 3Ps). This study suggested that changes in population dynamics and structure could be influenced by seasonal difference in spawning and migration. Genetic differences among fish sampled were marginal and perhaps due to low sample size (for example, only three fish were collected from 3Ps).

Beacham et al. (2002) employed seven new neutral microsatellite loci and one nuclear gene to examine genetic distinctness between Northern cod stocks. With mature pre-, post-, and spawning fish, they found the Gilbert Bay (NAFO 2J) to be the most genetically "distinctive" inshore sites, whereas no significant differences were found

among cod from Notre Dame Bay, Bonavista Corridor, Trinity Bay, and Conception Bay (NAFO 3KL; Beacham et al. 2002). The offshore, oceanographically distinct Flemish Cap site (NAFO 3M) was the most genetically distinctive, followed by Northern Cod in division 2J, likely due to the distance from other sampled areas. No genetic differences were found among different areas of the northern stock (including among bays, against the 'bay stock' hypothesis) and between subdivision 3Ps and Northern cod (Beacham et al. 2002). However, the genetic distinctiveness (Beacham et al. 2002; Bradbury et al. 2013), homing behaviour, and year-round residency (as reported by Green & Wroblewski 2000) of cod in Gilbert Bay was taken to suggest the possibility of a genetically distinct spawning population there.

Puncher et al. (2019) examined population structure and genotypic diversity with Next Generation Sequencing (NGS), Samples north of 45 ^{o} N (Northern Cod and St. Anns Bank) were panmictic and distinguishable from those further south (Bay of Fundy & Browns Bank) (Puncher et al. 2019; 2021). Follow-up found no population structure among Northern Cod areas (NAFO 2J+3KL) (Puncher et al. 2020). Three genetic clusters were found in this study with a neutral SNP dataset: (1) cod north of the Laurentian channel (Bonavista Corridor (3KL), Hawke Channel (2J), Notre Dame Channel (3KL), St. Pierre and Green Bank (3Ps)), (2) cod in the Gulf of Saint Lawrence (Burgeo Bank (3n), northeastern Gulf of Saint Lawrence (4R), St. Anns Bank (4Vn), southeastern Gulf of Saint Lawrence (4T)), and (3) cod south of the Laurentian channel (Bay of Fundy (4X), Browns Bank (4X), southern Scotian Shelf (4X)) (Puncher et al. 2020).

1.5 The 3Ps Cod Stock

Atlantic cod in NAFO subdivision 3Ps have a complex structure (DFO 2022) (Figure 1). Tagging studies have shown that 3Ps cod move through other divisions near 3Ps and coastal areas along the Avalon and Burin peninsulas of Newfoundland (DFO 2021; Lawson & Rose 2000).

Spawning grounds occur in 3Ps. Bar Haven, an area inside Placentia Bay, is the best known and most used spawning ground for Newfoundland cod, making 3Ps an important division for spawning (Lawson & Rose 2000). Robichaud & Rose (2001) investigated homing behaviour in Bar Haven with tag-and recapture of 48 cod tracked for two years. Spawning fish were tagged in April, prior to summer migration. 2/3rds of tagged fish were relocated in subsequent spawning seasons, all within at least 10 km from Bar Haven. No tagged fish were relocated in Bar Haven in summer months (Robichaud & Rose 2001). These fish are known to move out of 3Ps during the summer, since previous tagging studies show aggregations fluctuate between spring and summer (Robichaud & Rose 2001; Lawson & Rose 2000). In a similar experiment for all of Placentia Bay most tagged individuals were recaptured at the bay after summer migration, suggesting migration back to the spawning grounds (Lawson & Rose 2000). Due to homing behaviour and returning to specific 3Ps spawning grounds after migration, there is a chance for one or more genetically distinct spawning populations within 3Ps (Robichaud & Rose 2001).

Older tagging studies reported more connectivity between 3Ps and nearby populations. About 5% of fish tagged outside 3Ps were recaptured within 3Ps, (Thompson 1943; Templeman & Fleming 1962; Templeman 1974; Templeman 1979;

Lear 1998; Moguedet 1994; Taggart et al. 1995). The main contributors to 3Ps to are Northern Cod (NAFO 2J3KL), Burgeo Bank (NAFO 3Pn), and Southern Grand Banks (NAFO 3NO) (Figure 1; Brattey 1997). Cod from south of the Laurentian Channel, there seems to be little or no evidence that fish tagged south of the Laurentian Channel move north of the Laurentian Channel into 3Ps; Templeman 1974). Preliminary genetic analyses, based on small microsatellite panels, agree that 3Ps is likely connected to nearby cod populations, as no significant genetic differentiation was found between 3Ps and inshore Northern Cod bays (Conception and Trinity Bay) or between sampling locations within 3Ps (Beacham et al. 2002).

These studies have provided some insight into the contribution of other areas to 3Ps, but little about the reciprocal contribution of 3Ps to nearby areas. Here, I examine the contribution of Atlantic Cod in 3Ps to those in the Gulf of St Lawrence and the Northern Cod complex.

1.6 Atlantic Cod 'Supergene' Inversion Complexes

Supergenes are clusters of genes inherited together due to reduced recombination on a chromosome (Dobzhansky et al. 1966; Joron et al. 2011). They typically encode for complex phenotypes and are present across taxa (Schwander et al. 2014). For example, in birds and butterflies, supergenes are responsible for the development of colouring and patterns, like Batesian mimicry in *Heliconius* butterflies (Schwander et al. 2014). Atlantic cod display four supergene complexes in the form of chromosomal inversions: Linkage Groups 01, 02, 07, and 12, found on Chromosomes 1, 2, 7, and 12, respectively.

The Northern Cod stock shows Mendelian segregation of the LG01 supergene, which has a size of 18.5 Mb and includes 736 genes (Berg et al. 2017). These genes have been linked to migratory behaviour in the Northeast Atlantic (Sodeland et al. 2022). The other complexes (LG02, LG7, and LG12) contain genes involved in phenotype expression potentially subject to selection, such as temperature regulation, protein production, and response to water salinity and oxygen levels (Puncher et al. 2019; Berg et al. 2015). Clucas et al. (2019) found that inversions LG07 and LG12 also vary, with potential response to selection by environmental factors such as ocean temperature.

Among Atlantic Cod off the coasts of Norway, these linkage groups likely drive population structure. LG01 contains genes responsible for swim bladder pressure regulation, skeletal muscle organisation for migratory behaviour, and much more, is responsible for maintaining migratory and stationary ecotypes in Norwegian cod, where the migratory ecotype is a homokaryon "homozygous" for the inversion (Kirubakaran et al. 2016; Sinclair-Waters et al. 2017; Puncher et al. 2019). Hybridization between the two ecotypes does occur, where hybrids (heterokaryons) display an intermediate migratory behaviour (Hemmer-Hansen et al. 2013). However, population structure in Northeast Atlantic cod is not only driven by inversions. Prezygotic reproductive isolation has also led to genetic differentiation of Baltic Sea ecotypes, offshore "North Sea" ecotypes, and inshore "fjord" ecotypes (Knutsen et al. 2018; Barth et al. 2019).

In the Northwest Atlantic, LG01 has the same genes and may contribute to ecotype divergence as well. A total of 141 cod, including migratory fish taken off the coast of Labrador (NAFO 2J), and inshore fish from Gilbert Bay, were sequenced at a panel of 11,000 SNPs (Sinclair-Waters et al. 2017). The populations displayed variation

in the LG01 inversion, which suggests offshore migratory and inshore stationary ecotypes were associated with the LG01 inversion and that it plays a role in population divergence (Sinclair-Waters et al. 2017). Looking at Northern cod collectively, LG01 follows Mendelian inheritance, and both alleles are found in HWP (Puncher et al. 2019). In cod from the Gulf of Maine (NAFO 5Y & 5Z), LG01 is only found in heterokaryons for the inversion (Puncher et al. 2019). LG02 and LG07 however, display the highest variation (F_{ST} values) in the South, suggesting they contribute most of the genetic differentiation in the Gulf of Maine (Puncher et al. 2021).

In Northern cod (offshore NAFO 2J3KL), proportions of LG01 homozygotes vary between age groups (Puncher et al. 2019; 2020). Mature cod display a reduction of the LG01 inverted homozygote, which suggests genotype-dependent habitat selection as LG01 is correlated with temperature as well as migratory and stationary ecotypes, allowing individuals with the inversion to take advantage of ecological niches that other fish may be sensitive to (Puncher et al. 2020). Due to the patterns described for LG01 in Northern Cod, I focus on this inversion in the main text. Analyses for other linkage groups can be found in supplement.

1.7 Next Generation Sequencing (NGS): ddRAD-Sequencing

Double-digest restriction site-associated DNA sequencing (ddRADseq) is a form of reduced representation genome sequencing that allows for good sequence coverage of short loci for non-model organisms (Andrews et al. 2016). ddRADseq can be used to determine thousands of genetic markers, called SNPs, across the genome, in a cost-

effective manner (Andrews et al. 2016). This technique has proven to be important in ecological studies, especially in fisheries science and conservation genomics (Kumar & Kocour 2017). Use of NGS techniques and SNPs in population genetics has the advantage over other markers such as microsatellites or mitochondrial DNA (mtDNA) of a rapid, cost-effective way of collecting tens of thousands of markers from hundreds of individuals (Morin et al. 2004). The maximises the amount of information available from extracted DNA, and allows for a more accurate representation of the full genome and population of the study species (DeFaveri et al. 2013).

1.8 Objectives

Here, I use ddRADseq to compare (1) genetic variation in Atlantic Cod from NAFO 3Ps with that in other geographic populations in the Northwest Atlantic, and (2) differences between juvenile and mature fish in NAFO 3Ps.

2 Methods

2.1 Sampling and DNA Extraction

Samples for DNA analysis were obtained from the annual Fisheries and Oceans Canada Spring multispecies survey. Scale samples were collected from a total of 734 adult and juvenile fish from random stratified depth trawls throughout all of NAFO 3Ps in the Spring and Summer of 2021 and 2022 (Appendix II). After sampling, scales were

stored between paper slips in envelopes at room temperature. See Figure 2 for 3Ps sampling map highlighting sets represented in the final datasets.

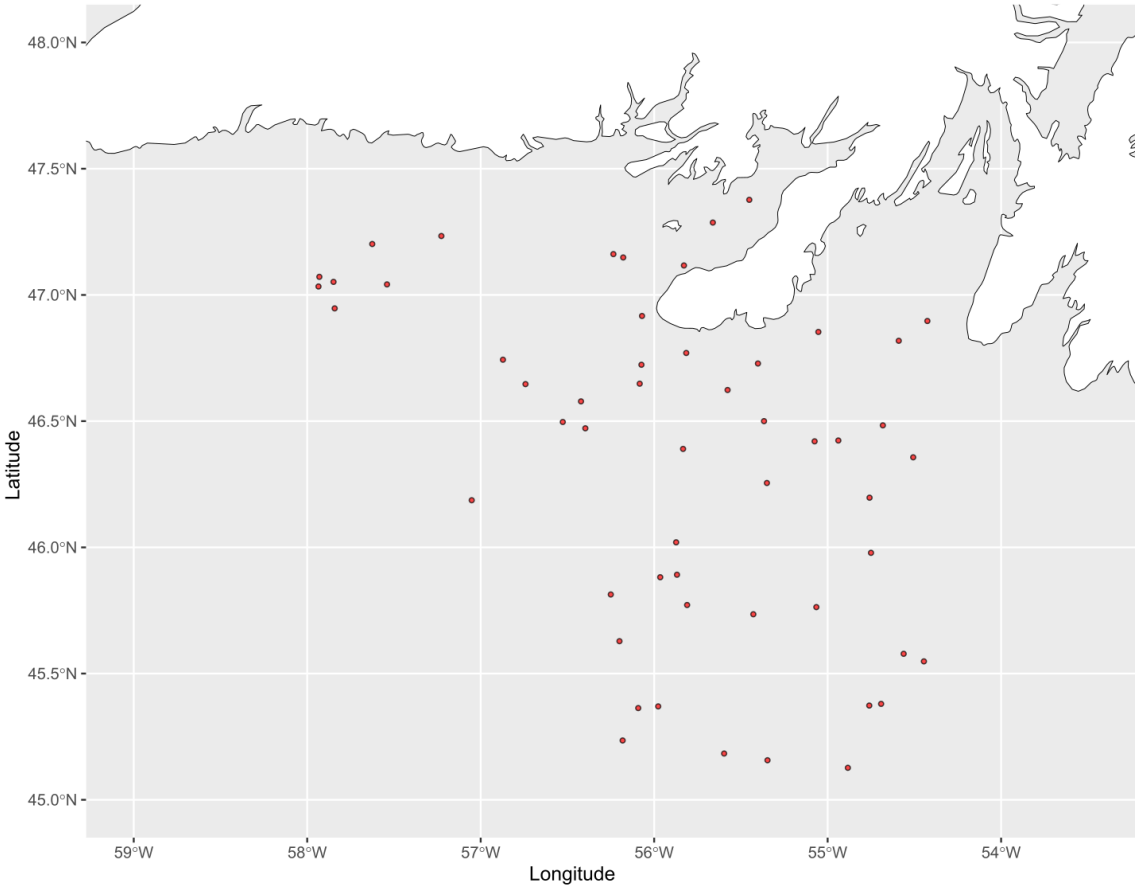


Figure 2: 3Ps Trawl Scale sampling locations collected in Spring-Summer 2021 & 2022. Points from sets represented in the final, filtered full dataset are plotted in red.

DNA was isolated from scale samples with the Omega E.Z.N.A.® Tissue DNA Kit or Omega E-Z 96™ Tissue DNA Kit (Omega Bio-Tek). The manufacturer's protocol was followed, with modifications when required. Samples with small numbers of scales and/or scales too small to reliably remove using forceps were extracted off of paper

following a modified protocol where 2X the amount of TL buffer is added for lysis. Second, the elution volume and number of elution rounds were modified depending on the sample, with smaller samples having one elution step with 30ul elution buffer and larger samples having one or two elution steps with 100ul total elution buffer. DNA concentrations were determined using ThermoFisher Scientific® Picogreen kits and normalised to 20ng/ul. If the sample had a Picogreen concentration of <20ng/ul and did not have excess scales to extract from, it was concentrated with NucleoMag NGS Clean-up and Size Select beads.

2.2 ddRAD library preparation

A total of 683 3Ps fish with high-quality DNA extractions were prepared for sequencing. I also had sequence data from 383 samples collected from divisions other than 3Ps (280 in final filtered dataset: 66 from Northern cod, 83 from Gulf of Saint Lawrence , 51 from Bay of Fundy, and 80 from Gulf of Maine See Appendix I for more information on each sample. Because reference data from other cod populations were sequenced on a different sequencing platform, Illumina® HiSeq™ at Genome Quebec (San Diego, U.S.A.) (Puncher et al. 2019; 2020; 2021), there was concern about potential lane/batch effects. A lane effect occurs when individuals are clustered due to similarities in sequencing chemistry rather than true genetic variation. To identify such effects, I selected 48 of the previously sequenced samples from outside 3Ps and re-sequenced them with our 3Ps samples, and sequenced 37 3Ps individuals on more than one lane. These reference samples were: 15 Gulf of Maine, 9 Bay of Fundy, 15 Northern, and 9 Gulf of

Saint Lawrence samples (Appendix I). The ddRAD protocol was used from Puncher et al. 2019. Amplicons were purified with NucleoMag NGS Clean-up and Size Select beads. Quality control was done with an Agilent® 2100 BioAnalyzer (Waldbronn, Germany) to confirm fragment size prior to paired-end sequencing at 150bp with an Illumina® NovaSeq™ (San Diego, U.S.A.).

2.3 STACKS Pipeline

Sequence data were filtered through to ensure high quality and throughput using standard bioinformatic pipelines. The STACKS workflow developed by Éric Normandeau (available on github: https://github.com/enormandeau/stacks_workflow) was used to call and filter SNPs after sequencing based on quality and read coverage. CutAdapt and Process_Radtags STACKS v2.5.2 (Rochette et al. 2019; Catchen et al., 2013) were used to trim adaptor sequences and demultiplex samples, respectively. Sequences were trimmed to 110 bp and aligned to the Atlantic cod reference genome gadMor3.0 (NCBI accession ID: GCF_902167405.1) with the Burrows-Wheeler Aligner at 90% alignment accuracy to ensure the sequences come from cod DNA (BWA) (Li & Durbin, 2009). gstacks from STACKS2 v2.5.2 (Rochette et al. 2019) was used to compile sequencing information into genotypes, which were then filtered with the Populations module (removed FIS < -0.3, heterozygosity > 0.6). Filtering thresholds were based on recommendations from Shafer et al. (2017) and previous filtering methods from Puncher et al. (2019; 2020; 2021). VCFTools was used for further and iterative filtering, including minor allele frequency (maf 0.05), minimum read depth (minDP 5, iteratively filtered at

10), maximum number of alleles (max-alleles 2), maximum percent missing data by locus (max-missing 0.6, 60% missing data), maximum percent missing data by individual (max-missing 0.3, iteratively filtered at 0.2: 20% missing data), related individuals (cutoff 0.5, siblings), and linkage equilibrium for the neutral dataset (--min-r2 0.2) (Table 1).

Table 1: Stepwise filtering details for full dataset (containing inversions) and neutral dataset.

Step	SNPs or Individuals Filtered Out
FIS < -0.3 Heterozygosity > 0.6	Filtered out 37,515 SNPs
Minor Allele Frequency (maf) 0.05	Kept 1103/1103 Individuals (720 3Ps (37 individuals sequenced on multiple lanes), 95 GoM, 70 BoF, 94 GSL, 76 Northern, 48 duplicate individuals) Kept 34579/551596 SNPs
Minimum Read Depth (minDP) 5	
Maximum Number of Alleles (max-alleles 2)	
Maximum Missing Data (max-missing 0.6)	
Maximum Missing Data by Individual (max-missing 0.3)	Kept 779/1103 Individuals (402 3Ps, 95 GoM, 68 BoF, 94 GSL, 74 Northern, 47 duplicate individuals) Kept 34579/34579 SNPs
Related Individuals 0.5	Identified 2 pairs of siblings, removed 1 from each pair. Removed some mislabeled individuals (3Ps sequenced on multiple lanes) Kept 721/779 Individuals (380 3Ps, 81 GoM, 60 BoF, 87 GSL, 66 Northern, 47 duplicate individuals) Kept 34579/34579 SNPs

Linkage Equilibrium for the Neutral Dataset ONLY (--min-r2 0.2)	Kept 721/721 Individuals Kept 26049/34579 SNPs
Minimum Read Depth (minDP) 10	No Changes
Maximum Missing Data by Individual: Full Dataset (max-missing 0.2)	Kept 646/721 Individuals (320 3Ps, 80 GoM, 51 BoF, 83 GSL, 65 Northern, 47 duplicate individuals) Kept 34579/34579 SNPs
Maximum Missing Data by Individual: Neutral Dataset (max-missing 0.2)	Kept 642/721 Individuals (316 3Ps, 80 GoM, 51 BoF, 83 GSL, 65 Northern, 47 duplicate individuals) Kept 26049/26049 SNPs
Lane Effect Loci and Individuals	36 individuals removed from Neutral (606 individuals: 280 3Ps, 80 GoM, 51 BoF, 83 GSL, 65 Northern, 47 duplicate individuals) 173 Lane Effect SNPs removed from Neutral (25876 SNPs) 173 Lane Effect SNPs removed from Full Dataset (34407 SNPs)
Outliers (Neutral ONLY)	25,714 SNPs left
Non-Nuclear Loci	FULL: 323 SNPs removed; 34084 SNPs left
	NEUTRAL: 256 SNPs removed; 25,458 SNPs left
FINAL DATASETS	FULL: 34,084 SNPs, 646 Total Individuals: 320 from 3Ps, 47 duplicate from reference populations
	NEUTRAL: 25,458 SNPs, 606 Total Individuals: 280 from 3Ps, 47 duplicate from reference populations

2.4 Detecting Lane Effect SNPs

To determine if a lane effect was present, a Principal Components Analysis (PCA) was run on the neutral dataset with the dudi.pca function in the R package *adegenet* (Jombart 2008). In the absence of a lane effect, duplicate samples from outside

3Ps were expected to cluster with their original groups. Instead, all samples sequenced on NovaSeq clustered together, which indicates the presence of a lane effect along PC1 and PC2 (Figure 2).

To remove the lane effect, BayeScan v.2.1 (Foll & Gaggioti 2008) was run on the combined neutral dataset as well as the previously sequenced neutral datasets to detect outliers (including 3Ps). A total of 173 consensus outlier SNPs (<0.7% of SNPs) between the two datasets were discarded. This was done to remove SNPs that may have been affected by erroneous polyG tails known to occur in data sequenced by NovaSeq, thus causing the lane effect when the dataset was combined with HighSeq-sequenced data, without erroneous polyG tails (De-Kayne et al. 2020). Second, following similar methods done by Crotti et al. 2020 and Leek et al 2010, the function snpgdsPCACorr from the R package *SNPRELATE* v.1.28.0 (Zheng et al., 2012) was used to determine which SNPs drove the most variation between the clusters on PC1 and PC2. *SNPRELATE* is used for Genome-Wide Association Studies (GWAS) and subsequent analyses with genomic data, including PCAs. The final 173 lane-effect SNPs were consensus SNPs between the non-discarded SNPs from BayeScan (850) and the *snpgdsPCACorr* SNPs with a correlation coefficient > 0.3. These SNPs were removed from all datasets prior to downstream analysis with the *genepopedit* R package, used to edit Genepop files (Stanley et al. 2017).

Since I sequenced samples on two lanes, a second lane effect was discovered only along PC2 (separation of groups based on sequencing differences) in the neutral dataset of 36 3Ps individuals from the first sequenced lane. These individuals were removed from only the neutral dataset since linkage groups overpowered any variation it caused in other datasets.

2.5 Locating Linkage Groups

To identify SNPs on linkage groups, I followed methods from Puncher et al. 2019. A Manhattan plot was made with these F_{ST} values and the R package ggplot2 v.3.0.0 (Wickham, 2011). Chromosome RefSeq codes were identified in the gadMor3.0 reference genome using NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_902167405.1/). 322 SNPs that were not located on the 23 nuclear chromosomes were removed from downstream analyses. SNPs with $F_{ST} > 0.15$ on chromosomes 1, 2, 7, & 12 were identified as linkage group SNPs for their respective chromosomes.

2.6 Making the Neutral Dataset

After removing linkage disequilibrium SNPs, lane effect SNPs, and non-nuclear SNPs, outliers were detected from the five populations with BayeScan and the Bonferroni correction function from the *PCAdapt* R package (Luu et al. 2017). Neutral SNPs were retained and used for all downstream analyses. A total of 154 consensus outliers, were removed from downstream analysis with *genepopedit* (Stanley et al. 2017) and compared to the cod genome with ENSEMBL (Zerbino et al. 2018). Gene ontology of any locus of interest was investigated with UniProt (www.uniprot.org). 25,458 neutral SNPs were retained from 606 individuals, 280 of which were from 3Ps.

2.7 Population Structure

F-statistics are a means of resolving population structure among individuals and sub-populations within a total population sample (Wright 1951; Weir & Cockerham 1984). This includes the inbreeding coefficient (F_{IS}), total heterozygosity (F_{IT}), and fixation index (F_{ST}). F_{IS} is a measure of observed vs expected heterozygosity within a subpopulation. is F_{ST} measures subpopulation structure by comparing heterozygosity of two populations separately versus pooled (pairwise F_{ST}; Meirmans & Hedrick 2011). Finally F_{IT} is the total measure of heterozygosity in a system (Wright 1951). Values for F_{ST} can range from 0.0 - 1.0, where 0.0 indicates identical expected heterozygosity in both measured populations. Although other statistics exist, such as G_{ST} and R_{ST}, these may be better suited for multi-allelic microsatellite markers (G_{ST}) and may not be a better estimate than F_{ST}, especially for bi-allelic markers like SNPs (Meirmans & Hedrick 2011). Thus, use of F_{ST} for SNP data here is appropriate. F_{ST} was used by Puncher et al. (2019), such that our results are comparable. Pairwise F_{ST} and 95% confidence intervals with 100 bootstraps were run with the *StAMPP* package in R (Pembleton et al. 2013). Allele frequencies were calculated with the basic.stats function in the *hierfstat* package in R (Goudet 2005).

In population genetics, it is important to use multiple statistical methods to evaluate potential population structure, to avoid potential bias (Patterson et al. 2006). Alongside PCAs, Admixture analysis is another tool used to detect ancestral populations, which uses Bayesian clustering and assign percents to the genetic makeup of each individual to form a full ancestry (Lawson et al. 2018). This method is employed by most major human genetic ancestry testing companies, such as AncestryDNA and 23andMe

(Kirkpatrick & Rashkin 2017). Although useful in determination of the presence of population structure, the accuracy of this method depends on prior probabilities estimated from its reference populations (Lawson et al. 2018). It also requires an estimated number of ancestral populations, known as 'K', which can be estimated using likelihood-of-model estimates such as Akaike's or Bayesian information criterion (AIC and BIC, respectively; Verity & Nichols 2016). Bayesian clustering and admixture analyses were run with the R package *TESS* 2.3.1 (Chen et al. 2007) with 100 iterations of the Markov Chain Monte Carlo (MCMC) chain and 10 repetitions for each simulation of K with no priors given. A plot was generated for K=2, 3, 4 for comparison, using the repetition with the lowest cross-entropy.

2.8 Supergene Frequencies for 3Ps Maturity Groups

To determine whether LG01 frequencies changed between juvenile and mature fish, as described in Puncher et al. (2019) for Northern cod, 3Ps fish were classified as juvenile or mature based on maturity codes collected during sampling (codes >100 for males and 500 for females were considered 'mature') and ages were determined using otoliths. Individuals were genotyped as homozygous for the inversion, heterozygous, or homozygous for the non-inverted chromosome based on PCA groupings of a dataset that contained only SNPs from LG01. Hardy-Weinberg proportions, Chi square (χ^2) goodness of fit tests, and Fisher's Exact Tests were calculated to determine if the linkage group is under selection between maturity and age groups. Fisher's exact tests were used for sample sizes <20. Bonferroni corrections for p-values were done in the *stats* R package.

2.9 Sequencing and Filtering

Out of the 734 extracted scale samples, 587 had high-quality DNA and sufficient PicoGreen concentrations (>20 ng/ul) to be sent for sequencing. Post-sequencing reads had an average quality of 35, and an average of 2,760,999 reads per individual. After SNP filtering, there were 646 individuals and 34,084 SNPs in the full dataset and 606 individuals and 25,458 SNPs in the neutral dataset. See Table 1 for filtering details.

Out of the 174 consensus outliers identified by BayeScan and *PCAdapt* Bonferroni correction, 103 are located in annotated protein-coding genes and 4 in lncRNA genes. The remaining SNPs are located very close to annotated genes. See supplemental file 2 for more information about each outlier.

3 Results

3.1 Lane Effect Mitigation

When all 3Ps and duplicate samples were added to the larger reference dataset. they clustered and fell out in a separate group in the neutral data set (Figure 3A). After removal of SNPs responsible for the lane effect, the lane effect was no longer visually present in the PCA and all duplicate samples clustered with their original populations (Figure 3B).

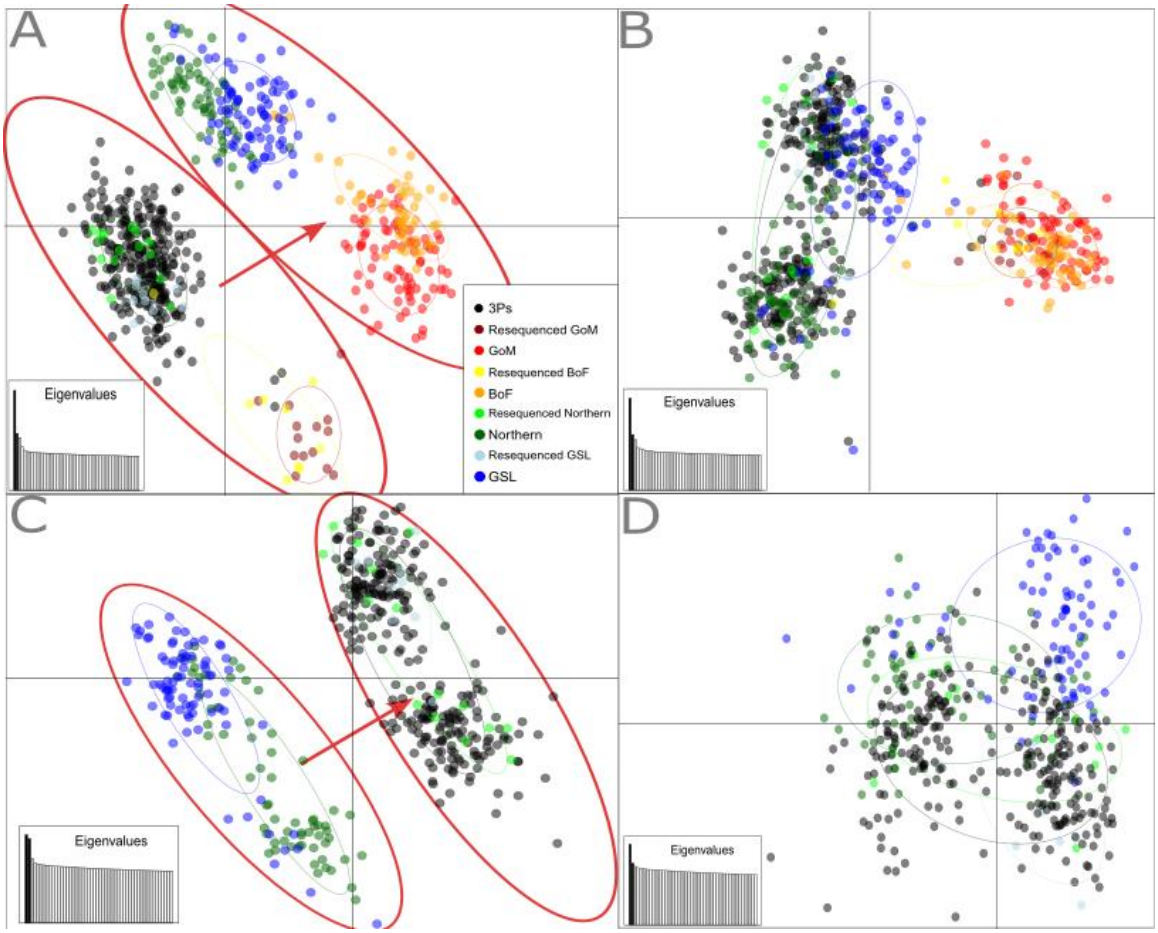


Figure 3: Neutral PCAs Before and After Lane Effect Removal. A) Before lane effect SNPs removal. Large red ovals show samples that were sequenced on the same machine. Lane effect is observed between samples sequenced with NovaSeq (bottom oval) *versus* HiSeq (top oval). Duplicate samples do not cluster with their original populations and instead cluster with the samples they were sequenced with (ex, duplicate GoM samples should cluster with GoM samples). Red arrow depicts the direction the lane effect affected samples, which is along the PC1 and PC2 axes. B) After lane effect SNPs removal. Duplicate samples now cluster with their original populations. C) Same as A without southern samples (GoM and BoF). D) Same as B without southern individuals. Inertia ellipses depict where the majority of individuals cluster for each group in the

legend, with a set size of 1.5. Eigenvalue plot shows the amount of variance each principal component contains.

Removal of the lane effect via higher filtering thresholds was also explored. The lane effect could be removed with strict sequencing depth threshold of 20 (Appendix III). However, use of such a stringent filter inhibited detection of fine-scale neutral genetic variation: all populations clustered together. Therefore, the data set with lane effects removed as in Figure 3 are analyzed here.

3.2 Population Structure

Outlier SNPs have been removed in neutral datasets unless specified.

3.2.1 *Whole Dataset (Including Inversion SNPs)*

PCA of 34,084 SNPs from 646 individuals, identified three clusters (Figure 4A). Individuals from NAFO 3Ps, Gulf of Saint Lawrence, and Northern Cod were found in all three, whereas only two clusters contain individuals from southerly populations (Gulf of Maine and Bay of Fundy). When run with no prior grouping information, the BIC plot shows three clusters (Appendix IV).

A Manhattan Plot showed areas within Chromosomes 1, 2, 7, and 12 had most of the variation, corresponding to the locations of LG01, 02, 07 and 12 and suggesting these linkage groups drive most of the overall genetic variation between populations (Figure 4B).

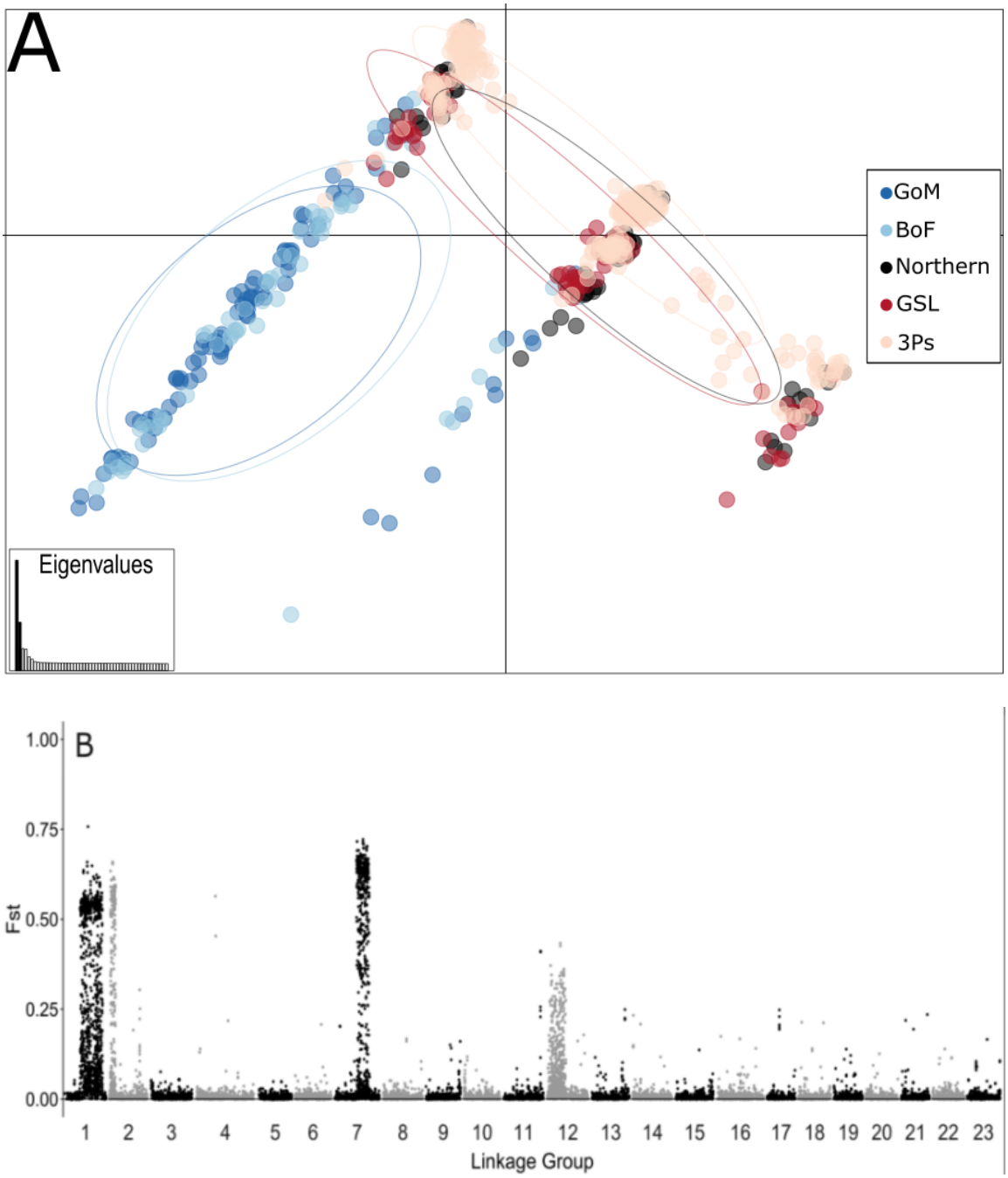


Figure 4: A) PCA and B) Manhattan Plot with Whole Dataset (34,084 SNPs) and all Populations. The greatest variation occurs in LG01, LG02, LG07, LG12. GoM=Gulf of Maine, BoF=Bay of Fundy, GSL=Gulf of Saint Lawrence. Inertia ellipses (default size 1.5) show where the majority of individuals cluster for each group.

Bayesian clustering by the cross-entropy criterion suggests between three and four ancestral populations (Figure S2). TESS plots using K=3 and 4 show similar patterns, with most 3Ps individuals having similar genetic profiles to Northern and Gulf of Saint Lawrence (Figure 5BC). In both TESS plots, a few fish from Gulf of Maine/Bay of Fundy are genetically similar to those in the Northern/Gulf of Saint Lawrence populations.

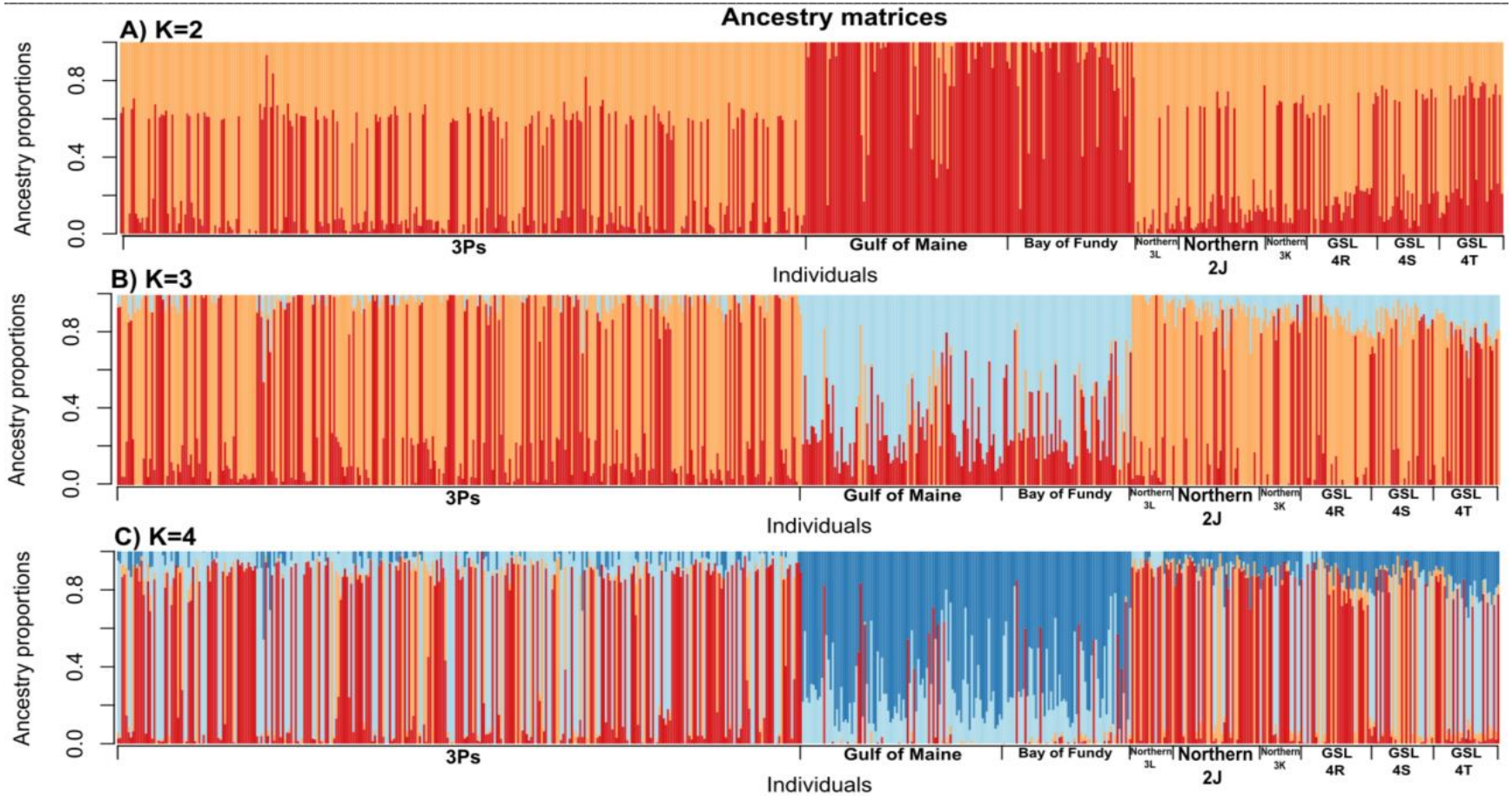


Figure 5: TESS plots of ancestry proportions with K= a)2 b)3 c)4 for the full dataset (including inversions).

Pairwise F_{ST} values between populations, are low but differ significantly from zero. They range from 0.0024 to 0.0363 (Table 2). Since these F_{ST} values are significantly different from zero, they indicate potential differences between these populations.

Table 2: Pairwise F_{ST} values using the whole dataset for each sampling location. Outliers have been removed. 95% confidence intervals are present (CI).

	3Ps	Northern	Gulf of Saint Lawrence	Gulf of Maine
Northern	F_{ST}: 0.0059 CI: 0.0052-0.0065			
Gulf of Saint Lawrence	F_{ST}: 0.0070 CI: 0.0062-0.0078	F_{ST}: 0.0024 CI: 0.0023-0.0027		
Gulf of Maine	F_{ST}: 0.0363 CI: 0.0349-0.0377	F_{ST}: 0.0363 CI: 0.0347-0.0380	F_{ST}: 0.0317 CI: 0.0304-0.0332	
Bay of Fundy	F_{ST}: 0.0353 CI: 0.0338-0.0367	F_{ST}: 0.0356 CI: 0.0342-0.0372	F_{ST}: 0.0295 CI: 0.0281-0.0308	F_{ST}: 0.0025 CI: 0.0023-0.0027

3.2.2 *Neutral Dataset (Noninversion)*

A PCA with 25,458 neutral SNPs for 606 individuals (280 from 3Ps) identified three clusters, one with primarily trans-Laurentian individuals (Gulf of Maine and Bay of Fundy), one containing individuals from the Gulf of Saint Lawrence, Northern Cod, and 3Ps, and another primarily individuals from Northern Cod and 3Ps (Figure 6).

Considering the Gulf of Saint Lawrence fish (southern Gulf (NAFO 4T), northern Gulf (NAFO 4RS)), we observed individuals from division 4T cluster farther away from Northern Cod compared to divisions 4R and 4S, apart from some mixing between 4T and 4R (Figure 6). NAFO 4T has the least overlap with 3Ps and Northern Cod, followed by NAFO 4R and 4S, but stretches closest towards the southern populations (NAFO 4X and 5Z). Similar patterns were observed in a PCA done with only spawning individuals (Appendix V).

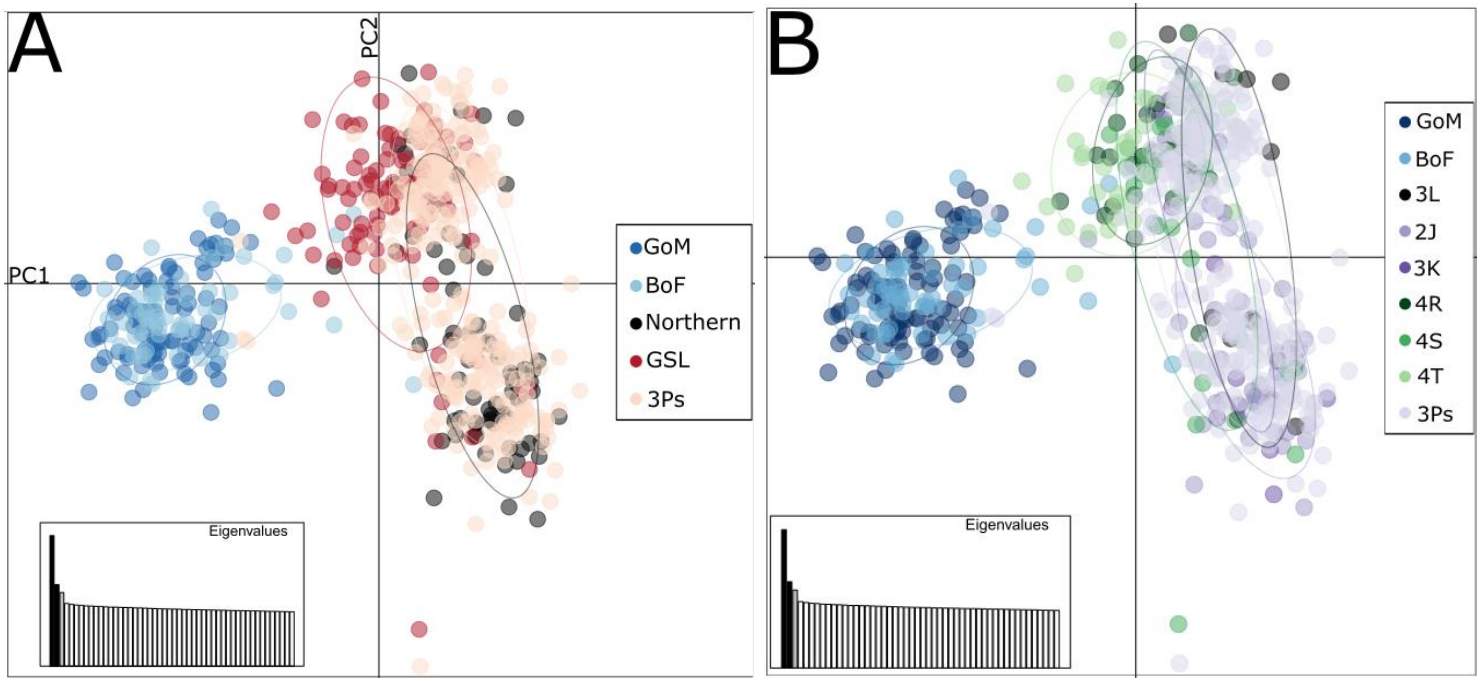


Figure 6: Neutral PCA with 25,458 SNPs and 606 individuals from all populations. A) by geographic origin and b) by NAFO Division. SNPs from Linkage Disequilibrium, inversions, lane effect analyses, and outlier analyses removed. Inertia ellipses (default size 1.5) show where the majority of individuals cluster for each sampling unit. Three clusters are present, with one on the left containing mostly Southern individuals (Gulf of Maine and Bay of Fundy), one near the top containing some Northern Cod, 3Ps, and most Gulf of Saint Lawrence individuals, and one in the bottom right containing Northern and 3Ps with some Gulf of Saint Lawrence

individuals. Northern=Northern Cod (2J, 3K, 3L), GoM=Gulf of Maine (5Z), BoF=Bay of Fundy (4X), GSL=Gulf of Saint Lawrence (4R, 4S, 4T).

When the neutral dataset was run with no prior grouping information, the BIC plot does not show a clear indication for numerous ancestral populations (Appendix VI). TESS Bayesian clustering detected two ancestral populations using cross-entropy criterion (Appendix VI).

Clustering with K=2 shows both southern populations (GOM and BOF) south differ from all more northerly locations (Figure 7A). Because strong structure can sometimes mask more subtle genetic structure between more geographically proximate populations, we explore adding another ancestral population (K=3). Clustering with K = 3 showed a more subtle genetic differentiation among 3Ps, Northern Cod, and Gulf of Saint Lawrence (Figure 7B). Gulf of Saint Lawrence individuals display higher overall ancestry proportions similar to southerly individuals, also seen in the upper left quadrant of Figure 6A. Some individuals in 3Ps and the Northern Cod area display this as well, but not as frequently as Gulf of Saint Lawrence. Northern and 3Ps individuals have higher overall ancestry proportions for the 'navy' population than Gulf of Saint Lawrence. In all TESS plots, 3Ps individuals have genetic profiles closer to Northern cod than to Gulf of Saint Lawrence fish, who show more fine genetic similarities to Southern populations as K increases. Among Neutral SNPs, population pairwise F_{ST} show the smallest difference between 3Ps and Northern Cod and the largest between Bay of Fundy and Northern Cod (Table 3).

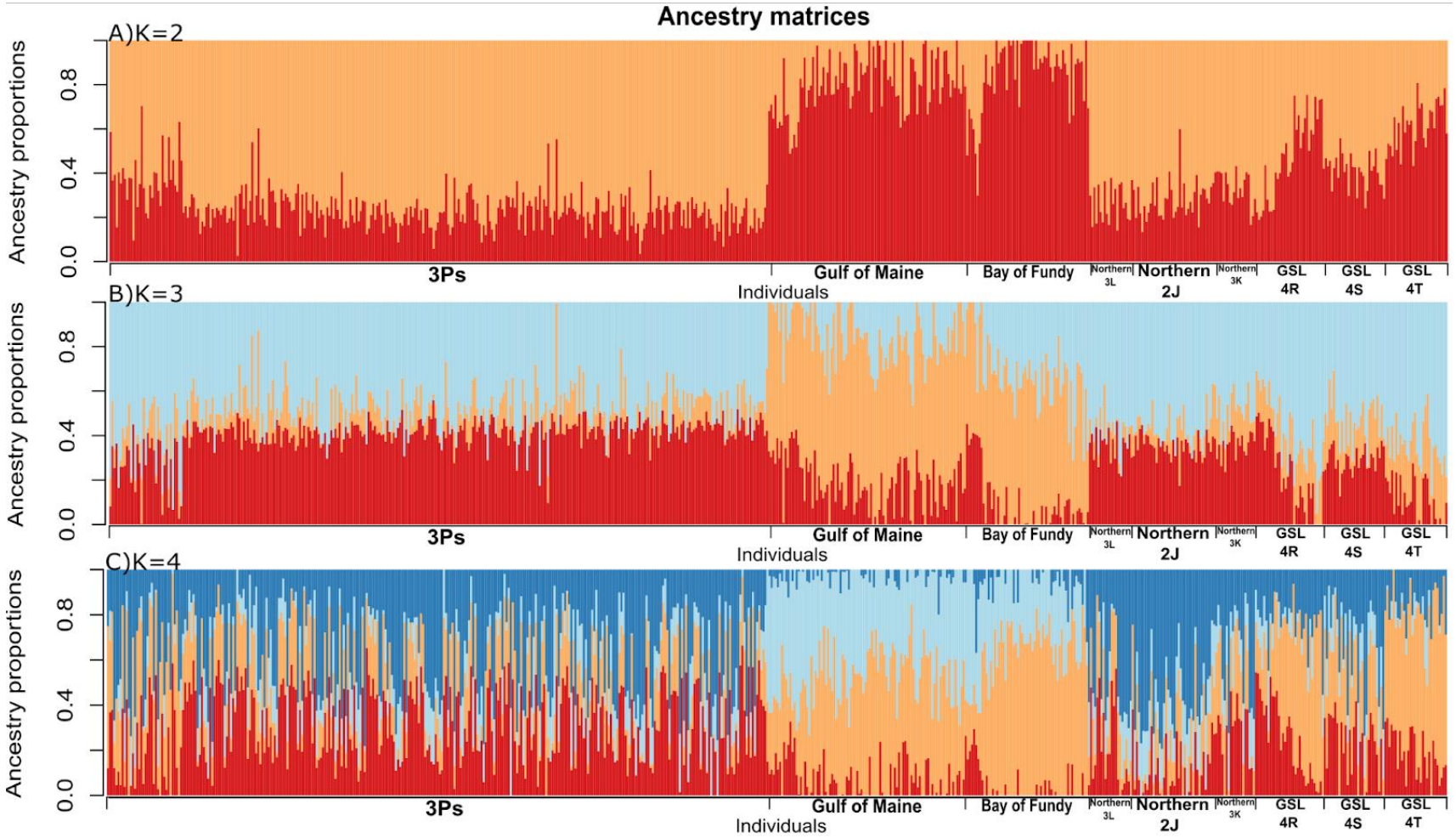


Figure 7: TESS plots using K= a)2 b)3 c)4 for Neutral SNPs only. Outliers have been removed.

Table 3: Pairwise F_{ST} values for Neutral dataset for each sampling location. Outliers have been removed. 95% confidence intervals are present (CI).

	3Ps	Northern	Gulf of Saint Lawrence	Bay of Fundy
Northern	F_{ST}: 0.0011 CI: 0.0009-0.0012			
Gulf of Saint Lawrence	F_{ST}: 0.0017 CI: 0.0015-0.0018	F_{ST}: 0.0016 CI: 0.0014-0.0018		
Bay of Fundy	F_{ST}: 0.0057 CI: 0.0054-0.0060	F_{ST}: 0.0064 CI: 0.0060-0.0067	F_{ST}: 0.0041 CI: 0.0038-0.0043	
Gulf of Maine	F_{ST}: 0.0056 CI: 0.0053-0.0058	F_{ST}: 0.0061 CI: 0.0057-0.0064	F_{ST}: 0.0049 CI: 0.0047-0.0052	F_{ST}: 0.0014 CI: 0.0012-0.0016

To further investigate the neutral PCA clusters containing, specifically the ones without trans-Laurentian fish, I separated the fish based on which PCA cluster they were found in. These clusters show a pairwise F_{ST} of approximately 0.0025 with and without Gulf of Saint Lawrence (CI 0.0024-0.0027). Furthermore, with K-means clustering, AIC plot suggests the presence of 2 clusters (Appendix VII).

PCA Loading plots with and without the presence of GSL showed that SNPs responsible for the variation between these two clusters, separating along PC1, are scattered across the genome (Appendix VIII). Two clusters remain when in the absence and presence of outlier SNPs (Figure 8). The clusters remain when the top 150 loadings are removed, but merge when the top 1000 loadings are removed, revealing this as

genome wide neutral genetic differentiation (Appendix IX). This same result is observed in the presence and absence of Gulf of Saint Lawrence individuals (Appendix IX).

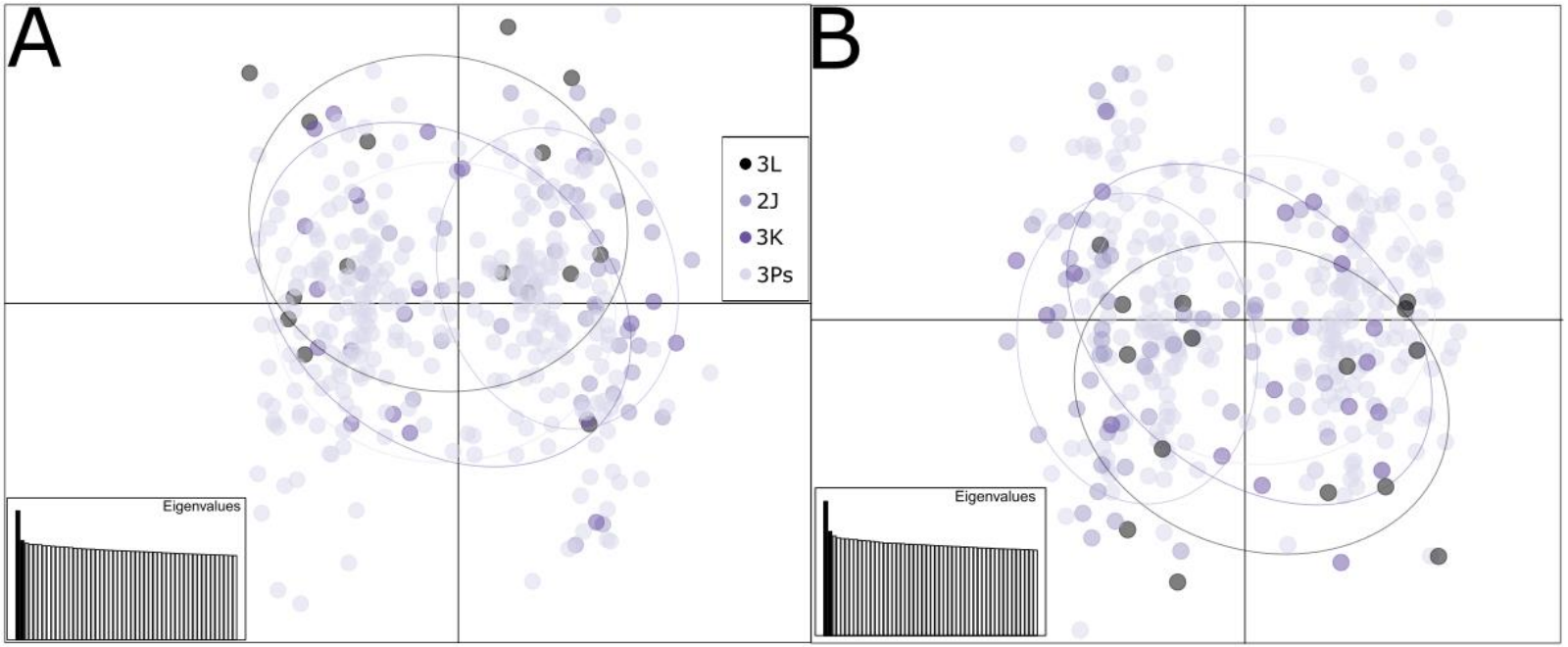


Figure 8: Neutral PCA of Northern Cod (NAFO 2J3KL) and 3Ps cod with outliers (A) and without (B). Gulf of Saint Lawrence individuals were removed to determine if the two clusters would still be present. Legends are the same for both panels.

The loading plot for the PCA without outliers indicates that the five most divergent SNPs (with the highest loading scores) between the two groups are located on Chromosomes 21 (two loci), 18, 23, and 13 (Appendix VIII). The SNP on Chromosome 21 contributes the most variation between the two groups. It occurs between two protein-coding genes, estrogen receptor 2a (*esr2a*) and jagged canonical Notch ligand 2b (*jag2b*). Another SNP is also located in *jag2b*. Esr2a is located in the nucleus and plays a role in cellular response to estrogen by binding to DNA to allow for transcription (GO:0008270). Jag2b is a transmembrane protein that binds calcium ions in the Notch pathway, which plays a role in cell and organ development (GO:0005509). The third SNP is on Chromosome 18 in the gene encoding Neurexin 1a, a transmembrane protein that plays a role in anatomical structure development (GO:0016020). The fourth SNP is on Chromosome 23 occurs close to a gene that encodes bile acid-CoA:amino acid N-acyltransferase-like (locus LOC115536838; GO:0016790). This enzyme is found in the cell's cytoplasm and plays a role in acyl-CoA metabolic processes. The final SNP is located on Chromosome 13 next to the R-spondin (*rspo4*) 4 gene and an RNA-coding gene. The rspo4 protein plays a role in the Wnt signalling pathway and sensory transduction by binding to heparin (GO:0050896). Allele frequencies for each of these SNPs tend to show general north-to-south geographic clines, but none are smooth (Table 4). Similar patterns are seen with allele frequencies by cluster, with and without individuals from GSL.

Table 4: Allele Frequencies for five SNPs that contribute to variation between the Two Putative Neutral Northern Stock Clusters. Cells are colors from the highest (dark blue) to the lowest (lightest blue) allele frequency within each row. N(3Ps)=297, n(2J)=45, n(3K)=20, n(3L)=15, n(4R)=26, n(4S)=30, n(4T)=36.

SNP	2J	3K	3L	3Ps	4S	4R	4T
(1)1671013:104:+ CHR21: between *esr2a* & *jag2b*	0.41	0.45	0.27	0.30	0.30	0.22	0.23
(2) 1444689:237:- CHR18: inside *fshr*	0.43	0.14	0.30	0.21	0.15	0.06	0.09
(3) 1818645:207:+ CHR23: near bile acid-CoA: amino acid N-acyltransferase like gene	0.41	0.14	0.23	0.29	0.17	0.12	0.11
(4)1671048:222:- CHR21: in *jag2b*	0.95	0.84	0.70	0.81	0.83	0.78	0.79
(5)1051858:203:+ CHR13: near *rspo4*	0.88	0.69	0.77	0.78	0.67	0.84	0.53

Allele Frequencies in the Putative Northern/3Ps Clusters (with GSL, without GSL)		
Locus	Cluster 1 (more 2J)	Cluster 2 (more 4T)
(1)1671013:104:+ CHR21: btwn *esr2a* & *jag2b*	0.50, 0.49	0.85, 0.86
(2) 1444689:237:- CHR18: inside *fshr*	0.39, 0.40	0.07, 0.07
(3) 1818645:207:+ CHR23: near bile acid-CoA: amino acid N-acyltransferase like gene	0.44, 0.45	0.13, 0.13

(4)1671048:222:- CHR21: in *jag2b*	0.98, 0.97	0.69, 0.66
(5)1051858:203:+ CHR13: near *rspo4*	0.92, 0.92	0.65, 0.66

3.3 Inversion SNPs

3.3.1 *LG01*

A PCA with 680 LG01 SNPs identified three clusters along PC1, similar to the full dataset PCA (Figure 9). All clusters contain 3Ps, Northern Cod, and Gulf of Saint Lawrence individuals, whereas only two contain southern individuals (Gulf of Maine & Bay of Fundy). Southern individuals are more differentiated along the PC2 axis. A Bayesian cross-entropy plot and cross-entropy plots suggest two ancestral populations (Appendix X). For LG02, 07, & 12, all PCAs showed a similar pattern with three clusters (Appendix XI).

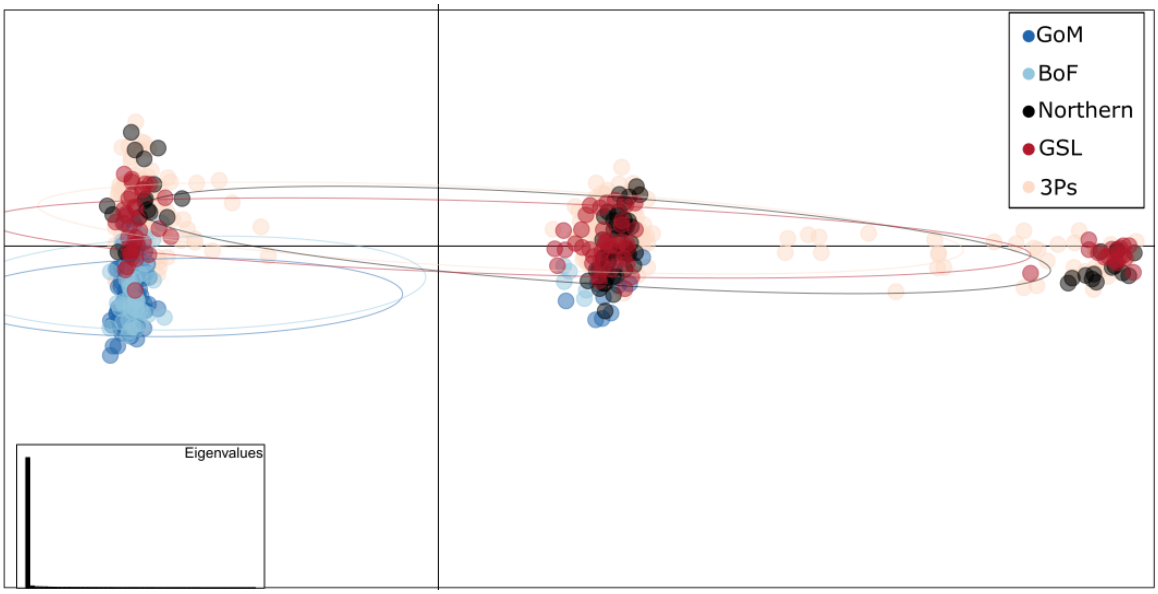


Figure 9: PCA of 680 LG01 SNPs. Inertia ellipses (default size 1.5) show where the majority of individuals cluster for each group.

3.3.2 *LG01 Inversion Frequencies Among Juvenile and Adult 3Ps Cod*

Presence or absence of the LG01 inversion on Chromosome 01 can be treated as alternative alleles (Puncher et al. 2019). Then, expected proportions of the inverted and non-inverted homokaryon and their heterokaryon classes can be calculated from Hardy-Weinberg considerations. Older fish (>5 years) have a lower proportion of the LG01 inversion homozygotes than younger fish (Figure 10; Appendix XII). The χ^2 values of LG01 was 7.1 for juveniles (Bonferroni adjusted p-value=0.056) and 30.7 for adults (Bonferroni adjusted p-value<0.0001). Thus, mature individuals are not in HWE, whereas juveniles are. Fisher's Exact Test on each otolith age class to smaller sample sizes

showed no significant departures from HW expectations (all Bonferroni adjusted p-values >0.05; Appendix XIII).

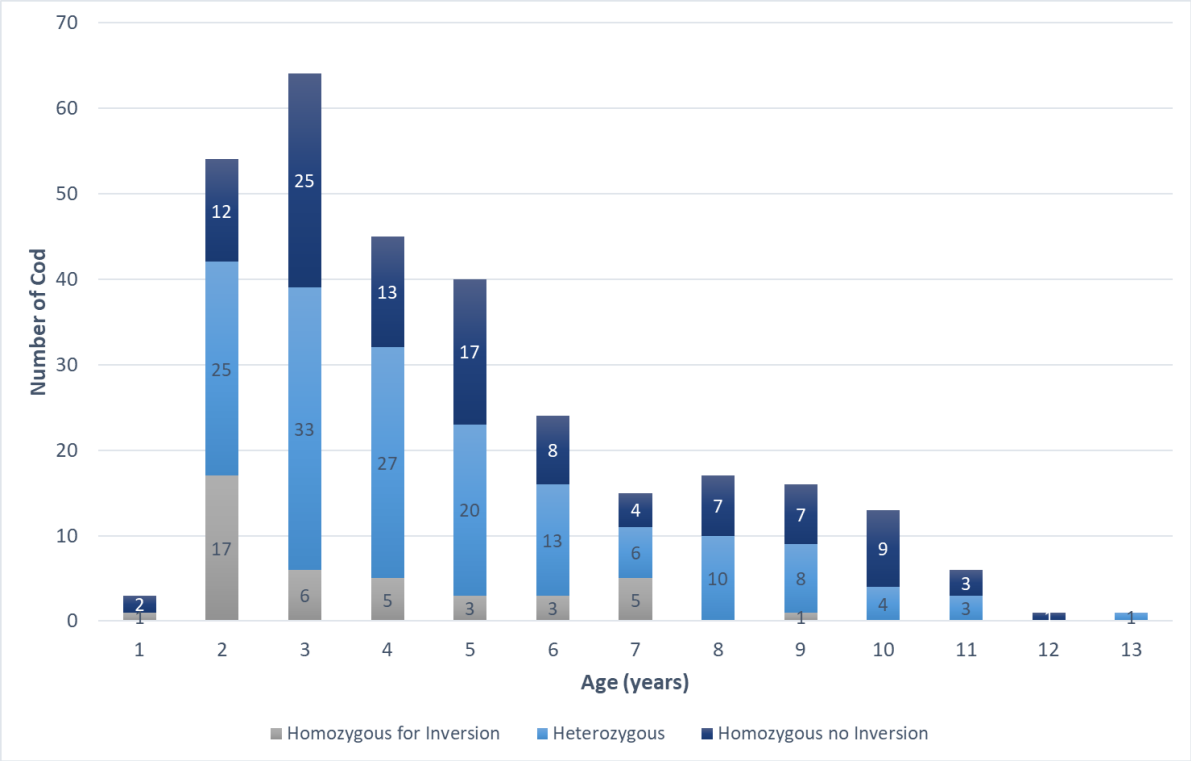


Figure 10: Stacked bar chart of LG01 Genotypes in 3Ps by age class. Pie chart depicts the proportions of each age in years for the sampled 3Ps population. 55% of the sample is between the ages of 2-4. A decrease in homozygotes for the LG01 inversion as cod age is also visualized.

4 Discussion

Fish in NAFO 3Ps off the southern coast of Newfoundland, is the least understood in terms of its genetic composition and population structure. Previous tagging and genetic studies suggested that cod from other management zones, mainly Northern Cod (NAFO 2J3KL) and the Gulf of Saint Lawrence (NAFO 4RST), mix with and move through subdivision 3Ps, (Templeman 1974; Beacham et al. 2002). However, no previous large-scale study has attempted to determine the contributions of various cod population to subdivision 3Ps. Here, next-generation sequencing data and whole-genome analysis indicate that 3Ps cod are not genetically distinguishable from Northern Cod, but that they each include two groups separated by fine-scale, genome-wide neutral genetic variation. There is also evidence that fish homozygous for the LG01 inversion, which has previously been shown to decrease in older Northern cod cohorts, also decreases in older 3Ps cod.

4.1 Genomic Connectivity of 3Ps Cod

Collectively, the data here show that 3Ps and Northern Cod are genomically indistinguishable. Juveniles and spawning cod from 3Ps cluster with Northern cod in both neutral and non-neutral datasets, and 3Ps individuals have similar genetic profiles to Northern cod. Partition of genomic variance by F_{ST} confirm the absence of genetic differentiation between 3Ps and Northern cod at neutral SNPs.

Consistent with the metapopulation hypothesis (Rose & Rowe 2015), Northern Cod are expected to be found in 3Ps, as tagging studies recaptured cod tagged in

Northern cod stocks within 3Ps (Templeman 1974). Previous research has described the Northern stock as comprising multiple genetic lineages across NAFO 2J3KL that mix and interbreed, whereas the isolationist hypothesis describes its components as genetically isolated from one another (Rose et al. 2011). Result here favour the metapopulation hypothesis, (Lawson & Rose 2000; Robichaud & Rose 2001). Large-scale genomic results expand on more restricted genetic analyses with seven microsatellite loci (Beacham et al. 2002), which detected no differences among Northern Cod in north-eastern Newfoundland deep-water bays (NAFO 3K inshore). They also account for the observation that subdivision 3Ps cod tagged in Placentia Bay have also been recaptured in NAFO 3L during the summer (Beacham et al. 2002.

4.2 Neutral Genetic Population Structure

4.2.1 *Two Northern Cod Groups*

I found evidence for three genetically identifiable cod populations in the Northern Cod, Gulf of Saint Lawrence, and more southerly divisions, consistent with the conclusions by Puncher et al. (2021). The addition of data from NAFO 3Ps off the southern coast of Newfoundland north of the Laurentian Channel shows a more complex population structure and connectivity, with evidence of more subtle north-to-south spatial genetic differences. There is evidence for two genetically distinguishable groups of Northern cod, one in NAFO 2J and another that clusters closer to the Gulf of Saint Lawrence stocks. Individuals from NAFO 3KL and 3Ps occur equally in both clusters.

This structure has not been previously detected. The clustering is not associated with inversion genotypes, but is detected by genome-wide, neutral genetic variation.

Based on the loading plot for the PCA based on the neutral data set and including Gulf of Saint Lawrence individuals, the five SNPs that display highest loading scores between the two groups are located on chromosomes 21 (two), 18, 23, and 13. The pattern of percent allele frequencies for these five SNPs follow a spatial pattern. Due to the spatial patterns in their allele frequencies between the stocks and two Northern Cod clusters, the separation of these clusters may be associated with latitudinal differences for the northern (NAFO 2J) and southern stocks (NAFO 4RST).

4.2.2 *Connectivity of Gulf of Saint Lawrence and Northern Cod Stocks*

The results suggest greater genetic mixing between Northern Cod and northern Gulf of Saint Lawrence divisions 4R and 4S) than with the more southerly division 4T. More specifically, we observed similar neutral genetic profiles between Northern cod stocks and northern GSL stocks (NAFO 4R and S), whereas individuals from the southern GSL (NAFO 4T) clustered closer to the southern populations (BoF & GoM). These patterns clarify previous studies. For example, by use of trace element composition of otoliths, Méthot et al. (2005) found some seasonal variability in the origins of cod in 3Ps. In winter and spring, there were more northerly cod (especially in eastern NAFO 3Ps), and in late April more cod from the northern Gulf of Saint Lawrence (NAFO 4RS) were found on the Burgeo Bank (western 3Ps). These data combined suggest that 4S and

4R have more mixing with Northern cod than NAFO 4T, but this may vary seasonally as most of our samples were collected during spring-fall.

The similarity between NAFO 4T and southern populations (Bay of Fundy and Gulf of Maine) shown in PCA and TESS plots (K=3), as well as pairwise F_{ST} could be explained by migration. Little migration has been found across the 45th -degree of latitude (Martin & Jean 1964; Comeau et al. 2002; Puncher et al. 2019). The summer migration of Gulf of Saint Lawrence cod through the Strait of Belle Isle into the Northern cod stocks and around Newfoundland's northern peninsula has been part of fish harvesters' ecological knowledge (FEK) as well as tagging data for many years (Murray et al. 2008). Cod tagged in the northern Gulf (NAFO 4RS) are typically recaptured near their tagging area or to the east 3Ps, and very few are recaptured to the west (Martin & Jean 1964; Murray et al. 2008). However, since the stocks are close in proximity, some individuals do mix between NAFO 4R, 4S and 4T, as evident in our TESS genetic profiles. as well as tagging data and FEK (Martin & Jean 1964; Murray et al. 2008). Furthermore, 4T cod are known to migrate southward and overwinter in 4Vn, returning to the Gulf in the Spring (Campana et al. 1995). Tagging studies also suggest that movement of NAFO 4T cod is restricted by the depth of the Laurentian Channel, such that they move through shallower waters into NAFO 4Vn and even as far as 4W (Martin & Jean 1964; Campana et al. 1995). This might explain why NAFO 4T fish cluster genetically closer to more southerly population, whereas 4RS fish cluster with NAFO 2J3KL and 3Ps.

4.3 LG01 Inversion Frequencies Decrease in 3Ps Adults versus Juveniles

The inverted and non-inverted alleles of LG01 alleles in subdivision 3Ps were found to be out of expected Hardy-Weinberg proportions in juvenile and adult 3Ps individuals, and the frequency of the inversion decline between juveniles and adults. Similar trends were observed in NAFO 2J3KL fish (Puncher et al. 2021).

The relative decline of LG01 homokaryons in adult fish has been explained as a result of either increased mortality or genotype-dependent habitat selection (Heikkinen et al. 1999; Puncher et al. 2021). “Homozygosity” of the LG01 inversion may result in a loss of fitness for adults in 3Ps, causing higher mortality and thus decreasing the frequency of adults homozygous for LG01. However, LG01 has not yet been linked to a trait that may decrease fitness, so it is not currently possible to determine whether this is the case (Puncher et al. 2021). The second and more favored hypothesis is that genotype-dependent habitat selection may be the cause of lower LG01 homozygosity in adult individuals compared to juveniles (Puncher et al. 2021). Genotype-dependent habitat selection occurs when the genotype of an individual affects the habitat it chooses (Bolnick & Otto 2013). For example, the genotypes at 101 loci in the American eel (*Anguilla rostrata*) determine whether an individual becomes a saltwater or freshwater ecotype (Pavey et al. 2015). The LG01 inversion homokaryon has been correlated with a migratory ecotype and the non-inverted homokaryon with a stationary type , (Kirubakaran et al. 2016). Genes found in the LG01 inversion include those responsible for swim bladder regulation and skeletal muscle organisation, selectively advantageous alleles of which could be important for long-distance migrations (Kirubakaran et al. 2016; Puncher et al. 2021). Thus, the LG01 inversion may allow migratory individuals to

exploit other niches via migration, such that they leave 3Ps at maturity. The areas these individuals migrate to may have not been sampled in my study. Juveniles typically remain in their natal territory for some time after hatching (Robichaud & Rose 2000), leading to less of a skew in genotype frequencies compared to adults.

During the Spring-Summer 2021-2022) sampling season, the majority of sampled 3Ps fish were juveniles. Spawning season in subdivision 3Ps reaches its peak in April-May, which would leave more juveniles and immature cod in the stock once adults are spent and leave in the summer months (Methot et al. 2005; Myers et al. 1993; Lawson & Rose 2000). Juveniles in Placentia Bay (3Ps), have consistently high tag-and-recapture rates (87%), which may explain the relatively high proportion of older juveniles (1-2 years old) during the spring-summer sampling season (Lawson & Rose 2000).

4.4 Lane Effects and Mitigation Strategies

To mitigate the lane effect while retaining signals of fine-scale neutral genetic variation, we used a consensus approach to remove 173 lane-effect SNPs. Similar methods were used by Crotti et al. (2021), where a lane effect was detected visually using a PCA. One drawback is that the potential biological significance of these discarded SNPs is lost, may be impacting downstream clustering (Lazar et al. 2013). Another issue with this method is that the mitigation is done visually and is specific to this PCA. Thus, the lane effect may still be present in other methods, such as Bayesian clustering in TESS or making a PCA with a subset of the data.

Combining new sequencing data with previously sequenced datasets often creates a lane effect (Leigh et al. 2018; De-Kayne et al. 2020). Reasons for this include changes in read length, which can cause misalignments and false SNP calls, differences in sequencing chemistry, difference in sequencing depths, and varying levels of DNA quality (Leigh et al. 2018; Lou & Therkildsen 2021). Specific to this project, older datasets sequenced on Illumina HiSeq machinery were combined with newer datasets sequenced using Illumina NovaSeq. The sequencing chemistry of these differs, as HiSeq uses a four-channel system and NovaSeq employs a two-channel system (De-Kayne et al. 2020; Stoler & Nekrutenko 2021). Two-channel systems are prone to erroneous poly-G tails in both paired-end sequences, which are able to pass through bioinformatic filtering and remain in final datasets. Thus, sequencing chemistry can add variation to a dataset, which may be picked up by downstream analyses and treated as genetic variation (De-Kayne et al 2020). As library preparation and sequencing technologies are constantly changing, problems like these are common. In the case of this study, more than just sequencing chemistry could have played a role in the observed lane effect, to varying degrees. Two different sequencing technologies were used in the combined dataset, which differ in sequencing depth and chemistry. 3Ps samples were sequenced at 150 bp fragments and trimmed to 110 bp, whereas reference samples were sequenced at 125 bp then trimmed to 110 bp. Since the reads were trimmed before alignment, the chance of false SNP calls decreases but is still not zero (Leigh et al. 2018; Lou & Therkildsen 2021). Further, although the lane effect was not observed in all PCAs with non-neutral datasets, it may still add erroneous variation that decreases statistical power of detecting [?] any biologically important results (Leek & Storey 2007; Leigh et al. 2018). RAD-Seq

techniques are overall more susceptible to lane effects since a single SNP is used to represent a section of the genome (Leigh et al. 2018).

Other ways of removing lane effects exist, such as use of more stringent SNP filters and a different SNP caller (Lou & Therkildsen 2021; De-Kayne et al. 2020). Although use of more stringent SNP filters, such as increasing the depth coverage to 20, did successfully remove the lane effect, it also removed almost all informative SNPs. , Use of a different SNP caller for recalibration, such as GATK (Poplin et al. 2018) is recommended to mitigate lane effects. STACKS v2's gstacks used in this study is a newer method that has been shown to outperform GATK in terms of SNP recall and genotype precision at higher coverages (Rochette et al. 2019). GATK and other recalibration methods are also not readily available for most non-model organisms, including Atlantic cod, and require high coverage sequencing data (Lou & Therkildsen 2021). That said, the use of more than one SNP-calling method and cross-validation of the results may aid in improved recognition and exclusion false SNPs.

4.5 *Quo vadis* 3Ps cod?

Despite its collapse in the early 1990s, Atlantic cod remains an important Canadian natural resource (Rose & Rowe 2015). Management efforts are put in place to protect Atlantic cod and encourage rebuilding. However, there has been little increase in the cod biomass almost 30 years post-collapse, with Northern cod only recently coming out of the critical zone (Lilly et al. 2008; Rose & Rowe 2015; Kennedy 2023). Thus, management efforts may benefit from better understanding Atlantic cod population

structure. This first large-scale genomic study shows that 3Ps cod are genetically more similar to Northern Cod than to any other management unit, and that the measured difference is so small as to make them indistinguishable. As future sequencing efforts confirm and extend these findings, we may expect an improved scientific basis for the management of the 3Ps stock with respect to nearby cod stocks and populations within the Northern cod stock complex and adjacent waters, including the Southern Grand Bank (NAFO 3NO) and the Flemish Cap (NAFO 3M).

4.6 Conclusions and Future Directions

This study has identified two genetically identifiable groups within the Northern cod population. The data are subject to lane effects, which have been corrected according to the latest standards. Future studies should sequence all individuals using the same methods and avoid combining datasets. By doing this, we can be more confident of the results being due to genetic variation and not a sequencing artefact/result of lane effect mitigation.

Furthermore, a more spatially and seasonally extensive sampling effort would be beneficial. This means including more individuals from reference populations throughout the Northwest Atlantic, as well as taking into account migration and spawning patterns for different cod aggregates. For example, to measure neutral, fine-scale population structure in the Northern cod stocks, more individuals from NAFO 2J3KL, as well as NAFO 3NO and even the Flemish Cap (NAFO 3M) should be included in analyses. This will help make results more robust and applicable to the whole cod population. Studies

should also aim to include as many spawning individuals as possible to increase confidence in the populations the cod belong to and consider seasonal variability.

Our study aimed to determine which population 3Ps is most similar to and investigate whether 3Ps cod had a change in LG01 allele frequencies as they age. We found 3Ps to be genetically most similar to Northern cod, and we observed two potentially genetically identifiable groups within Northern cod. Lastly, we concluded that adult 3Ps cod displayed a decrease in LG01 allele frequencies compared to juveniles, similar to patterns previously observed in Northern cod stocks (NAFO 2K3KL).

Literature Cited

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), 81-92.

Bavington, D. (2011). Managed annihilation: an unnatural history of the Newfoundland cod collapse. UBC press.

Beacham, T. D., Brattey, J., Miller, K. M., Le, K. D., & Withler, R. E. (2002). Multiple stock structure of Atlantic cod (Gadus morhua) off Newfoundland and Labrador determined from genetic variation. ICES Journal of Marine Science, 59(4), 650-665.

Beacham, T. D. (1983). Growth and maturity of Atlantic cod (Gadus morhua) in the southern Gulf of St. Lawrence. Department of Fisheries and Oceans, Fisheries Research Branch, Pacific Biological Station.

Bolnick, D. I., & Otto, S. P. (2013). The magnitude of local adaptation under genotype-dependent dispersal. Ecology and evolution, 3(14), 4722-4735.

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., ... & Bentzen, P. (2013). Genomic islands of divergence and their consequences for the

resolution of spatial structure in an exploited marine fish. Evolutionary Applications, 6(3), 450-461.

Brander, K. M. (1994). The location and timing of cod spawning around the British Isles. ICES Journal of Marine Science, 51(1), 71-89.

Brattey, J. (1997). Biological characteristics of Atlantic cod (Gadus morhua) from three inshore areas of northeastern Newfoundland. NAFO Scientific Council Studies, 29, 31-42.

Bui, A. O., Castonguay, M., Ouellet, P., & Sévigny, J. M. (2011). Searching for Atlantic cod (Gadus morhua) spawning sites in the northwest Gulf of St Lawrence (Canada) using molecular techniques. ICES Journal of Marine Science, 68(5), 911-918.

Campana, S., Fanning, P., Fowler, M., Frank, K., Halliday, R., Lambert, T., ... & Sinclair, A. (1995). Report of the 4Vn cod working group on the scientific value of a 4Vn cod (May-Oct) stock assessment. Fisheries and Oceans.

Campana, S. E., Chouinard, G. A., Hanson, J. M., & Fréchet, A. (1999). Mixing and migration of overwintering Atlantic cod (Gadus morhua) stocks near the mouth of the Gulf of St. Lawrence. Canadian Journal of Fisheries and Aquatic Sciences, 56(10), 1873-1881.

Castañeda, R. A., Burliuk, C. M., Casselman, J. M., Cooke, S. J., Dunmall, K. M., Forbes, L. S., ... & Mandrak, N. E. (2020). A brief history of fisheries in Canada. Fisheries, 45(6), 303-318.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular ecology, 22(11), 3124-3140.

Chen, C., Durand, E., Forbes, F., & François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes, 7(5), 747-756.

Chen, Y., & Mello, L. G. S. (1999). Growth and maturation of cod (Gadus morhua) of different year classes in the Northwest Atlantic, NAFO subdivision 3Ps. Fisheries Research, 42(1-2), 87-101.

Clucas, G. V., Kerr, L. A., Cadrin, S. X., Zemeckis, D. R., Sherwood, G. D., Goethel, D., ... & Kovach, A. I. (2019). Adaptive genetic variation underlies biocomplexity of Atlantic Cod in the Gulf of Maine and on Georges Bank. PLoS One, 14(5), e0216992.

Comeau, L. A., Campana, S. E., & Chouinard, G. A. (2002). Timing of Atlantic cod (Gadus morhua L.) seasonal migrations in the southern Gulf of St Lawrence:

interannual variability and proximate control. ICES Journal of Marine Science, 59(2), 333-351.

Crotti, M., Yohannes, E., Winfield, I. J., Lyle, A. A., Adams, C. E., & Elmer, K. R. (2021). Rapid adaptation through genomic and epigenomic responses following translocations in an endangered salmonid. Evolutionary Applications, 14(10), 2470-2489.

DeFaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNP s. Molecular ecology resources, 13(3), 377-392.

De-Kayne, R., Frei, D., Greenway, R., Mendes, S. L., Retel, C., & Feulner, P. G. (2021). Sequencing platform shifts provide opportunities but pose challenges for combining genomic data sets.

DFO. (2021). Stock assessment of northern cod (NAFO divisions 2J3KL) in 2021. Canadian Science Advisory Secretariat Science Advisory Report, 2021, 1-20.

DFO. (2022). Stock Assessment of NAFO subdivision 3Ps cod. Can Sci Advis Sec Sci Advis Rep. 2022.

DFO. (2024). NAFO SUBDIVISION 3PS ATLANTIC COD (GADUS MORHUA) STOCK ASSESSMENT IN 2023. Can Sci Advis Sec Sci Advis Rep. 2023.

Dobzhansky, T., Anderson, W. W., & Pavlovsky, O. (1966). Genetics of natural populations. XXXVIII. Continuity and change in populations of Drosophila pseudoobscura in western United States. Evolution, 418-427.

Drinkwater, K. F. (2005). The response of Atlantic cod (Gadus morhua) to future climate change. Ices journal of marine science, 62(7), 1327-1337.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, 180(2), 977-993.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular ecology notes, 5(1), 184-186.

Green, J. M., & Wroblewski, J. S. (2000). Movement patterns of Atlantic cod in Gilbert Bay, Labrador: evidence for bay residency and spawning site fidelity. Journal of the Marine Biological Association of the United Kingdom, 80(6), 1077-1085.

Hanski, I., & Simberloff, D. (1997). The metapopulation approach, its history, conceptual domain, and application to conservation. Metapopulation biology, 5-26.

Heikkinen, S., Pietilä, M., Halmekytö, M., Suppola, S., Pirinen, E., Deeb, S. S., ... & Laakso, M. (1999). Hexokinase II-deficient mice: prenatal death of homozygotes without disturbances in glucose tolerance in heterozygotes. Journal of Biological Chemistry, 274(32), 22517-22523.

Hemmer-Hansen, J., Nielsen, E. E., Therkildsen, N. O., Taylor, M. I., Ogden, R., Geffen, A. J., ... & Carvalho, G. R. (2013). A genomic island linked to ecotype divergence in A tlantic cod. Molecular Ecology, 22(10), 2653-2667.

Hutchings, J. A., & Rangeley, R. W. (2011). Correlates of recovery for Canadian Atlantic cod (Gadus morhua). Canadian Journal of Zoology, 89(5), 386-400.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403-1405.

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., ... & French-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature, 477(7363), 203-206.

Kennedy, A. (2023, October 25). New DFO Northern COD assessment model shows stock out of critical zone | CBC News. CBCnews. https://www.cbc.ca/news/canada/newfoundland-labrador/2j3kl-new-stock-assessment-model-1.7007858

Kirkpatrick, B. E., & Rashkin, M. D. (2017). Ancestry testing and the practice of genetic counseling. Journal of Genetic Counseling, 26, 6-20.

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., ... & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. Molecular ecology, 25(10), 2130-2143.

Kumar, G., & Kocour, M. (2017). Applications of next-generation sequencing in fisheries research: A review. Fisheries Research, 186, 11-22.

Lawson, G. L., & Rose, G. A. (2000). Seasonal distribution and movements of coastal cod (Gadus morhua L.) in Placentia Bay, Newfoundland. Fisheries Research, 49(1), 61-75.

Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature communications, 9(1), 3258.

Lazar, C., Meganck, S., Taminau, J., Steenhoff, D., Coletta, A., Molter, C., ... & Nowé, A. (2013). Batch effect removal methods for microarray gene expression data integration: a survey. Briefings in bioinformatics, 14(4), 469-490.

Lear, W. H. (1998). History of fisheries in the Northwest Atlantic: the 500-year perspective. Journal of Northwest Atlantic Fishery Science, 23.

Leek, J. T., & Storey, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS genetics, 3(9), e161.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., ... & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics, 11(10), 733-739.

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. Molecular Ecology Resources, 18(4), 778-788.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25(14), 1754-1760.

Lilly, G. R., Wieland, K., Rothschild, B. J., Sundby, S., Drinkwater, K., Brander, K., ... & Vázquez, A. (2008). Decline and recovery of Atlantic cod (Gadus morhua) stocks throughout the North Atlantic.

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

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(1), 67-77.

Martin, W. R., & Jean, Y. (1964). Winter cod taggings off Cape Breton and on offshore Nova Scotia banks, 1959–62. Journal of the Fisheries Board of Canada, 21(2), 215-238.

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. Molecular ecology resources, 11(1), 5-18.

Méthot, R., Audet, M. C. Y. L. C., & Campana, S. E. (2005). Spatio-temporal distribution of spawning and stock mixing of Atlantic cod from the northern Gulf of St. Lawrence and southern Newfoundland stocks on Burgeo Bank as revealed by maturity and trace elements of otoliths. Journal of Northwest Atlantic Fishery Science, 36.

Moguedet, P. (1994). Cod (Gadus morhua) migrations in the Gulf of St. Lawrence and areas south of Newfoundland. NAFO Scientific Council Studies, (22), 71-84.

Morgan, M. J., & Brattey, J. (1996). Maturity of Female Cod in Divisions 2J+ 3KL with. Scientific Council Studies, 26, 61.

Morin, P. A., Luikart, G., Wayne, R. K., & SNP Workshop Group. (2004). SNPs in ecology, evolution and conservation. Trends in ecology & evolution, 19(4), 208-216.

Murray, G., Neis, B., Palmer, C. T., & Schneider, D. C. (2008). Mapping cod: fisheries science, fish harvesters' ecological knowledge and cod migrations in the Northern Gulf of St. Lawrence. Human Ecology, 36, 581-598.

Myers, R. A., Hutchings, J. A., & Barrowman, N. J. (1997). Why do fish stocks collapse? The example of cod in Atlantic Canada. Ecological applications, 7(1), 91-106.

Myers, R. A., Mertz, G., & Bishop, C. A. (1993). Cod spawning in relation to physical and biological cycles of the northern North–west Atlantic. Fisheries Oceanography, 2(3-4), 154-165.

Ouellet, P., Lambert, Y., & Castonguay, M. (1997). Spawning of Atlantic cod (Gadus morhua) in the northern Gulf of St. Lawrence: a study of adult and egg distributions and characteristics. Canadian Journal of Fisheries and Aquatic Sciences, 54(1), 198-210.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. PLoS genetics, 2(12), e190.

Pavey, S. A., Gaudin, J., Normandeau, E., Dionne, M., Castonguay, M., Audet, C., & Bernatchez, L. (2015). RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. Current Biology, 25(12), 1666-1671.

Pembleton, L. W., Cogan, N. O., & Forster, J. W. (2013). St AMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular ecology resources, 13(5), 946-952.

Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., ... & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nature biotechnology, 36(10), 983-987.

Puncher, G. N., Rowe, S., Rose, G. A., Leblanc, N. M., Parent, G. J., Wang, Y., & Pavey, S. A. (2019). Chromosomal inversions in the Atlantic cod genome: Implications for management of Canada's Northern cod stock. Fisheries research, 216, 29-40.

Puncher, G. N., Rowe, S., Rose, G. A., Parent, G. J., Wang, Y., & Pavey, S. A. (2020). Life-stage-dependent supergene haplotype frequencies and metapopulation

neutral genetic patterns of Atlantic cod, Gadus morhua, from Canada's Northern cod stock region and adjacent areas. Journal of Fish Biology, 98(3), 817-828.

Puncher, G. N., Wang, Y., Martin, R., DeCelles, G., Cadrin, S. X., Zemeckis, D., ... & Pavey, S. A. (2021). Transborder Gene Flow between Canada and the USA and Fine-Scale Population Structure of Atlantic Cod in the Broader Gulf of Maine Region. Transactions of the American Fisheries Society, 150(5), 560-577.

Robichaud, D., & Rose, G. A. (2001). Multiyear homing of Atlantic cod to a spawning ground. Canadian Journal of Fisheries and Aquatic Sciences, 58(12), 2325-2329.

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Molecular ecology, 28(21), 4737-4754.

Rose, G. A., & Rowe, S. (2015). Northern cod comeback. Canadian Journal of Fisheries and Aquatic Sciences, 72(12), 1789-1798.

Rose, G. A., Nelson, R. J., & Mello, L. G. (2011). Isolation or metapopulation: whence and whither the Smith Sound cod?. Canadian Journal of Fisheries and Aquatic Sciences, 68(1), 152-169.

Rose, G. A. (2007). Cod: the ecological history of the North Atlantic fishery. St. John's (NL): Breakwater Books, 32.

Ruzzante, D. E., Taggart, C. T., Lang, S., & Cook, D. (2000). Mixed-stock analysis of Atlantic cod near the Gulf of St. Lawrence based on microsatellite DNA. Ecological Applications, 10(4), 1090-1109.

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. Current Biology, 24(7), R288-R294.

Sinclair-Waters, M., Bentzen, P., Morris, C. J., Ruzzante, D. E., Kent, M. P., Lien, S., & Bradbury, I. R. (2018). Genomic tools for management and conservation of Atlantic cod in a coastal marine protected area. Canadian Journal of Fisheries and Aquatic Sciences, 75(11), 1915-1925.

Sinclair-Waters, M. (2017). Genomic Perspectives for Conservation and Management of Atlantic Cod in Coastal Labrador (Doctoral dissertation).

Sobel, J. (1996). IUCN red list of threatened species: Gadus morhua. Retrieved July, 11, 2018.

Sodeland, M., Jentoft, S., Jorde, P. E., Mattingsdal, M., Albretsen, J., Kleiven, A. R., ... & Knutsen, H. (2022). Stabilizing selection on Atlantic cod supergenes through a

millennium of extensive exploitation. Proceedings of the National Academy of Sciences, 119(8), e2114904119.

Stanley, R. R., Jeffery, N. W., Wringe, B. F., DiBacco, C., & Bradbury, I. R. (2017). genepopedit: A simple and flexible tool for manipulating multilocus molecular data in R. Molecular Ecology Resources, 17(1), 12-18.

Stoler, N., & Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing instruments. NAR genomics and bioinformatics, 3(1), lqab019.

TAGGART, C. T., P. PENNEY, N. BARROWMAN, and GEORGE. 1995. The 1954–1993 Newfoundland cod tagging database: statistical summaries and spatial temporal distributions. Can. Tech. Rep. Fish. Aquat. C. Sci., 2042.

Templeman, W., & Fleming, A. M. (1962). Cod tagging in the Newfoundland area during 1947 and 1948. Journal of the Fisheries Board of Canada, 19(3), 445-487.

Templeman, W. (1974). Migrations and intermingling of Atlantic cod (Gadus morhua) stocks of the Newfoundland area. Journal of the Fisheries Board of Canada, 31(6), 1073-1092.

Templeman, W. (1979). Migration and intermingling of stocks of Atlantic cod, Gadus morhua, of the Newfoundland and adjacent areas from tagging in 1962–66. ICNAF Res. Bull, 14, 5-50.

Thompson, H. (1943). biological and economic study of cod (Gadus callarias, L.) in the Newfoundland area, including Labrador. Research bulletin (Fisheries)-Dept. of Natural Resources (Newfoundland) eng.

Verity, R., & Nichols, R. A. (2016). Estimating the number of subpopulations (K) in structured populations. Genetics, 203(4), 1827-1839.

Walters, C., & Maguire, J. J. (1996). Lessons for stock assessment from the northern cod collapse. Reviews in fish biology and fisheries, 6, 125-137.

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

Wickham, H. (2011). ggplot2. Wiley interdisciplinary reviews: computational statistics, 3(2), 180-185.

Wright, S. (1951). The genetical structure of populations. Annals of eugenics, 15(1), 323-354.

Wroblewski, J., Neis, B., & Gosse, K. (2005). Inshore stocks of Atlantic cod are important for rebuilding the East Coast fishery. Coastal Management, 33(4), 411-432.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., ... & Flicek, P. (2018). Ensembl 2018. Nucleic acids research, 46(D1), D754-D761.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28(24), 3326-3328.

Appendix I

Sample Name	Population (abbrev)	Year	Maturity Stage	Age Group	Age
1_21_19	3Ps	2022	immature	immature	4
1_21_23	3Ps	2022	immature	immature	4
1_21_29	3Ps	2022	immature	immature	2
1_21_5	3Ps	2022	immature	immature	3
1_218_13	3Ps	2021	immature	immature	2
1_218_19	3Ps	2021	immature	immature	2
1_218_24	3Ps	2021	immature	immature	9
1_218_6	3Ps	2021	immature	immature	2
1_218_71	3Ps	2021	immature	immature	4
1_219_12	3Ps	2021	immature	immature	3
1_219_33	3Ps	2021	immature	immature	2
1_219_53	3Ps	2021	immature	immature	4
1_219_61	3Ps	2021	immature	immature	3
1_219_79	3Ps	2021	immature	immature	2
1_22_11	3Ps	2022	immature	immature	5
1_22_29	3Ps	2022	immature	immature	5
1_22_48	3Ps	2022	immature	immature	2
1_22_55	3Ps	2022	immature	immature	3

1_22_7	3Ps	2022	immature	immature	4
1_22_30	3Ps	2022	immature	immature	5
1_23_35	3Ps	2022	immature	immature	
1_23_38	3Ps	2022	immature	immature	
1_23_42	3Ps	2022	immature	immature	
1_23_52	3Ps	2022	immature	immature	4
1_24_14	3Ps	2022	immature	immature	5
1_24_17	3Ps	2022	immature	immature	
1_24_6	3Ps	2022	immature	immature	5
10_23_46	3Ps	2022	immature	immature	3
10_23_52	3Ps	2022	immature	immature	3
11_219_12	3Ps	2021	immature	immature	2
11_22_8	3Ps	2022	immature	immature	
11_24_6	3Ps	2022	immature	immature	4
113_24_13	3Ps	2022	immature	immature	2
12_24_6	3Ps	2022	immature	immature	4
124_218_54	3Ps	2021	immature	immature	3
13_21_8	3Ps	2022	immature	immature	3
13_219_12	3Ps	2021	immature	immature	3
13_22_32	3Ps	2022	immature	immature	5
13_22_7	3Ps	2022	immature	immature	3
13_23_46	3Ps	2022	immature	immature	3

14_23_46	3Ps	2022	immature	immature	3
146_218_54	3Ps	2021	immature	immature	3
15_218_33	3Ps	2021	immature	immature	4
15_22_6	3Ps	2022	immature	immature	3
15_23_46	3Ps	2022	immature	immature	2
16_22_4	3Ps	2022	immature	immature	3
16_22_6	3Ps	2022	immature	immature	3
16_23_46	3Ps	2022	immature	immature	4
163_218_54	3Ps	2021	immature	immature	2
17_218_14	3Ps	2021	immature	immature	2
17_219_12	3Ps	2021	immature	immature	3
17_22_19	3Ps	2022	immature	immature	2
17_23_46	3Ps	2022	immature	immature	3
18_219_83	3Ps	2021	immature	immature	4
18_23_52	3Ps	2022	immature	immature	3
189_218_54	3Ps	2021	immature	immature	2
19_218_13	3Ps	2021	immature	immature	2
19_219_12	3Ps	2021	immature	immature	2
19_23_52	3Ps	2022	immature	immature	3
2_21_17	3Ps	2022	immature	immature	3
2_21_22	3Ps	2022	immature	immature	3
2_21_25	3Ps	2022	immature	immature	2

2_21_36	3Ps	2022	immature	immature	
2_21_38	3Ps	2022	immature	immature	
2_218_13	3Ps	2021	immature	immature	2
2_218_20	3Ps	2021	immature	immature	3
2_219_12	3Ps	2021	immature	immature	3
2_219_22	3Ps	2021	immature	immature	2
2_219_33	3Ps	2021	immature	immature	2
2_219_53	3Ps	2021	immature	immature	4
2_219_61	3Ps	2021	immature	immature	2
2_219_7	3Ps	2021	immature	immature	3
2_219_70	3Ps	2021	immature	immature	2
2_219_71	3Ps	2021	immature	immature	3
2_22_26	3Ps	2022	immature	immature	4
2_22_48	3Ps	2022	immature	immature	2
2_22_5	3Ps	2022	immature	immature	5
2_22_55	3Ps	2022	immature	immature	4
2_22_6	3Ps	2022	immature	immature	4
2_22_9	3Ps	2022	immature	immature	3
2_23_35	3Ps	2022	immature	immature	
2_23_42	3Ps	2022	immature	immature	
2_23_46	3Ps	2022	immature	immature	5
2_23_52	3Ps	2022	immature	immature	5

2_24_13	3Ps	2022	immature	immature	5
2_24_7	3Ps	2022	immature	immature	5
2_24_9	3Ps	2022	immature	immature	3
20_219_83	3Ps	2021	immature	immature	3
22_219_83	3Ps	2021	immature	immature	4
23_219_83	3Ps	2021	immature	immature	4
23_23_52	3Ps	2022	immature	immature	3
24_219_83	3Ps	2021	immature	immature	4
26_219_43	3Ps	2021	immature	immature	2
27_23_52	3Ps	2022	immature	immature	4
28_24_6	3Ps	2022	immature	immature	2
29_219_83	3Ps	2021	immature	immature	5
29_23_52	3Ps	2022	immature	immature	3
29_24_12	3Ps	2022	immature	immature	1
3_21_25	3Ps	2022	immature	immature	3
3_218_20	3Ps	2021	immature	immature	2
3_218_24	3Ps	2021	immature	immature	3
3_218_71	3Ps	2021	immature	immature	3
3_219_10	3Ps	2021	immature	immature	2
3_219_12	3Ps	2021	immature	immature	2
3_219_17	3Ps	2021	immature	immature	4
3_219_22	3Ps	2021	immature	immature	2

3_219_33	3Ps	2021	immature	immature	2
3_219_63	3Ps	2021	immature	immature	3
3_219_70	3Ps	2021	immature	immature	1
3_22_4	3Ps	2022	immature	immature	5
3_23_35	3Ps	2022	immature	immature	
3_23_37	3Ps	2022	immature	immature	
3_23_52	3Ps	2022	immature	immature	3
3_24_6	3Ps	2022	immature	immature	4
3_24_7	3Ps	2022	immature	immature	3
33_219_83	3Ps	2021	immature	immature	4
39_218_26	3Ps	2021	immature	immature	4
4_21_23	3Ps	2022	immature	immature	2
4_219_12	3Ps	2021	immature	immature	2
4_219_7	3Ps	2021	immature	immature	3
4_22_11	3Ps	2022	immature	immature	4
4_22_54	3Ps	2022	immature	immature	2
4_22_9	3Ps	2022	immature	immature	5
4_23_37	3Ps	2022	immature	immature	
4_23_46	3Ps	2022	immature	immature	3
4_23_52	3Ps	2022	immature	immature	4
40_23_52	3Ps	2022	immature	immature	3
41_218_54	3Ps	2021	immature	immature	6

43_22_32	3Ps	2022	immature	immature	3
46_218_54	3Ps	2021	immature	immature	3
47_218_28	3Ps	2021	immature	immature	5
5_218_71	3Ps	2021	immature	immature	2
5_219_63	3Ps	2021	immature	immature	2
5_219_7	3Ps	2021	immature	immature	3
5_22_5	3Ps	2022	immature	immature	3
5_23_42	3Ps	2022	immature	immature	
5_23_46	3Ps	2022	immature	immature	3
50_218_54	3Ps	2021	immature	immature	3
6_21_8	3Ps	2022	immature	immature	3
6_218_20	3Ps	2021	immature	immature	2
6_218_22	3Ps	2021	immature	immature	4
6_219_33	3Ps	2021	immature	immature	1
6_219_6	3Ps	2021	immature	immature	3
6_219_7	3Ps	2021	immature	immature	3
6_22_30	3Ps	2022	immature	immature	4
6_22_5	3Ps	2022	immature	immature	2
6_24_12	3Ps	2022	immature	immature	2
60_22_4	3Ps	2022	immature	immature	2
7_218_26	3Ps	2021	immature	immature	3
7_219_19	3Ps	2021	immature	immature	3

7_219_56	3Ps	2021	immature	immature	5
7_219_6	3Ps	2021	immature	immature	2
7_23_42	3Ps	2022	immature	immature	
7_23_46	3Ps	2022	immature	immature	3
8_21_23	3Ps	2022	immature	immature	2
8_218_71	3Ps	2021	immature	immature	2
8_219_12	3Ps	2021	immature	immature	3
8_219_43	3Ps	2021	immature	immature	4
8_219_56	3Ps	2021	immature	immature	5
8_22_30	3Ps	2022	immature	immature	4
8_23_46	3Ps	2022	immature	immature	3
82_218_22	3Ps	2021	immature	immature	3
89_218_22	3Ps	2021	immature	immature	2
9_218_15	3Ps	2021	immature	immature	2
9_219_12	3Ps	2021	immature	immature	3
9_219_19	3Ps	2021	immature	immature	2
9_219_56	3Ps	2021	immature	immature	4
9_219_83	3Ps	2021	immature	immature	3
9_22_30	3Ps	2022	immature	immature	4
1_21_36	3Ps	2022	immature	immature	
1_218_5	3Ps	2021	immature	immature	3
1_219_10	3Ps	2021	immature	immature	4

1_219_13	3Ps	2021	immature	immature	2
1_219_15	3Ps	2021	immature	immature	3
1_219_67	3Ps	2021	immature	immature	4
1_219_77	3Ps	2021	immature	immature	2
15_218_54	3Ps	2021	immature	immature	4
2_219_13	3Ps	2021	immature	immature	2
2_219_72	3Ps	2021	immature	immature	3
20_218_54	3Ps	2021	immature	immature	4
27_219_43	3Ps	2021	immature	immature	2
32_218_25	3Ps	2021	immature	immature	2
37_218_54	3Ps	2021	immature	immature	3
4_218_21	3Ps	2021	immature	immature	4
4_218_71	3Ps	2021	immature	immature	2
4_219_22	3Ps	2021	immature	immature	2
4_219_67	3Ps	2021	immature	immature	3
6_218_23	3Ps	2021	immature	immature	2
88_218_54	3Ps	2021	immature	immature	2
9_219_43	3Ps	2021	immature	immature	3
1_218_22	3Ps	2021	mature	mature	9
1_218_25	3Ps	2021	mature	mature	10
1_219_16	3Ps	2021	mature	mature	10
1_219_42	3Ps	2021	mature	mature	11

1_219_6	3Ps	2021	mature	mature	11
1_219_7	3Ps	2021	mature	mature	5
1_219_83	3Ps	2021	mature	mature	6
1_22_1	3Ps	2022	mature	mature	7
1_22_21	3Ps	2022	mature	mature	5
1_22_25	3Ps	2022	mature	mature	6
1_22_32	3Ps	2022	mature	mature	7
1_22_4	3Ps	2022	mature	mature	6
1_22_5	3Ps	2022	mature	mature	7
1_24_3	3Ps	2022	mature	mature	8
1_24_7	3Ps	2022	mature	mature	6
10_22_54	3Ps	2022	mature	mature	6
10_24_6	3Ps	2022	mature	mature	5
11_218_24	3Ps	2021	mature	mature	7
11_22_30	3Ps	2022	mature	mature	10
12_219_83	3Ps	2021	mature	mature	4
13_219_19	3Ps	2021	mature	mature	15
13_219_83	3Ps	2021	mature	mature	5
13_24_3	3Ps	2022	mature	mature	5
14_218_24	3Ps	2021	mature	mature	6
14_22_11	3Ps	2022	mature	mature	11
15_219_83	3Ps	2021	mature	mature	4

155_218_34	3Ps	2021	mature	mature	10
16_219_83	3Ps	2021	mature	mature	5
16_22_11	3Ps	2022	mature	mature	6
16_22_7	3Ps	2022	mature	mature	8
17_22_11(only inNeutral)	3Ps	2022	mature	mature	8
19_218_25	3Ps	2021	mature	mature	9
19_219_83	3Ps	2021	mature	mature	6
2_219_42	3Ps	2021	mature	mature	7
2_219_44	3Ps	2021	mature	mature	10
2_219_6	3Ps	2021	mature	mature	9
2_219_83	3Ps	2021	mature	mature	6
2_22_11	3Ps	2022	mature	mature	5
2_22_13	3Ps	2022	mature	mature	6
2_22_19	3Ps	2022	mature	mature	11
2_22_54	3Ps	2022	mature	mature	9
2_24_6	3Ps	2022	mature	mature	6
20_218_23	3Ps	2021	mature	mature	
22_218_25	3Ps	2021	mature	mature	11
23_218_28	3Ps	2021	mature	mature	7
27_219_77	3Ps	2021	mature	mature	9
28_218_25	3Ps	2021	mature	mature	5

3_218_22	3Ps	2021	mature	mature	5
3_218_53	3Ps	2021	mature	mature	10
3_219_19	3Ps	2021	mature	mature	6
3_22_9	3Ps	2022	mature	mature	7
36_218_25	3Ps	2021	mature	mature	6
4_219_42	3Ps	2021	mature	mature	10
4_219_76	3Ps	2021	mature	mature	12
4_22_8	3Ps	2022	mature	mature	7
4_24_13	3Ps	2022	mature	mature	5
45_219_16	3Ps	2021	mature	mature	9
5_22_11	3Ps	2022	mature	mature	7
5_22_8	3Ps	2022	mature	mature	5
52_218_33	3Ps	2021	mature	mature	8
6_219_83	3Ps	2021	mature	mature	8
7_218_24	3Ps	2021	mature	mature	8
8_219_16	3Ps	2021	mature	mature	9
8_22_11	3Ps	2022	mature	mature	6
8_24_3	3Ps	2022	mature	mature	8
1_218_11	3Ps	2021	mature	mature	6
1_218_14	3Ps	2021	mature	mature	4
15_218_22	3Ps	2021	mature	mature	5
2_218_11	3Ps	2021	mature	mature	4

2_218_71	3Ps	2021	mature	mature	4
2_219_19	3Ps	2021	mature	mature	8
26_219_77	3Ps	2021	mature	mature	10
4_219_43	3Ps	2021	mature	mature	7
7_219_16	3Ps	2021	mature	mature	9
9_218_24	3Ps	2021	mature	mature	6
1_218_33	3Ps	2021	ripe/running	mature	8
1_22_13	3Ps	2022	ripe/running	mature	7
10_22_32	3Ps	2022	ripe/running	mature	10
16_218_33	3Ps	2021	ripe/running	mature	10
16_219_16	3Ps	2021	ripe/running	mature	4
2_219_16	3Ps	2021	ripe/running	mature	6
20_22_4	3Ps	2022	ripe/running	mature	5
206_218_34	3Ps	2021	ripe/running	mature	6
3_219_76	3Ps	2021	ripe/running	mature	10
3_24_14	3Ps	2022	ripe/running	mature	4
4_218_11	3Ps	2021	ripe/running	mature	4
4_219_83	3Ps	2021	ripe/running	mature	8
4_24_6	3Ps	2022	ripe/running	mature	4
5_22_54	3Ps	2022	ripe/running	mature	9
67_22_32	3Ps	2022	ripe/running	mature	9
7_219_83	3Ps	2021	ripe/running	mature	8

1_218_12	3Ps	2021	ripe/running	mature	7
1_219_72	3Ps	2021	ripe/running	mature	10
2_219_58	3Ps	2021	ripe/running	mature	5
4_218_22	3Ps	2021	ripe/running	mature	5
1_21_25	3Ps	2022	spent	mature	6
1_21_31	3Ps	2022	spent	mature	6
1_218_26	3Ps	2021	spent	mature	9
1_219_56	3Ps	2021	spent	mature	5
1_22_36	3Ps	2022	spent	mature	5
10_219_16	3Ps	2021	spent	mature	8
10_219_83	3Ps	2021	spent	mature	11
10_22_30	3Ps	2022	spent	mature	6
12_22_30	3Ps	2022	spent	mature	8
16_24_3	3Ps	2022	spent	mature	6
2_219_56	3Ps	2021	spent	mature	6
2_22_30	3Ps	2022	spent	mature	5
2_22_7	3Ps	2022	spent	mature	7
2_24_14	3Ps	2022	spent	mature	6
25_218_28	3Ps	2021	spent	mature	8
29_218_45	3Ps	2021	spent	mature	9
3_218_54	3Ps	2021	spent	mature	8
3_219_83	3Ps	2021	spent	mature	7

35_22_30	3Ps	2022	spent	mature	9
4_218__	3Ps	2021	spent	mature	
49_219_83	3Ps	2021	spent	mature	9
5_219_19	3Ps	2021	spent	mature	8
5_219_43	3Ps	2021	spent	mature	5
5_219_83	3Ps	2021	spent	mature	10
5_22_30	3Ps	2022	spent	mature	9
6_219_56	3Ps	2021	spent	mature	5
8_218_54	3Ps	2021	spent	mature	7
9_24_6	3Ps	2022	spent	mature	7
1_218_20	3Ps	2021	spent	mature	5
11_22_54	3Ps	2022	spent	mature	4
22_218_33	3Ps	2021	spent	mature	8
3_219_56	3Ps	2021	spent	mature	5
5_219_56	3Ps	2021	spent	mature	5
6_219_43	3Ps	2021	spent	mature	3
96_218_54	3Ps	2021	spent	mature	8
2_23_37	3Ps	2022			
GM2J15L005	Northern (2J)	2015			
GM2J15L007	Northern (2J)	2015			

GM2J15L008	Northern (2J)	2015			
GM2J15L012	Northern (2J)	2015			
GM2J15L029	Northern (2J)	2015			
GM2J15L048	Northern (2J)	2015			
GM2J15L050	Northern (2J)	2015			
GM2J15L059	Northern (2J)	2015			
GM2J15L091	Northern (2J)	2015			
GM2J15L092	Northern (2J)	2015			
GM2J15L093	Northern (2J)	2015			
GM2J15M002	Northern (2J)	2015			
GM2J15M003	Northern (2J)	2015			

GM2J15M004	Northern (2J)	2015			
GM2J15M011	Northern (2J)	2015			
GM2J15M025	Northern (2J)	2015			
GM2J15M026	Northern (2J)	2015			
GM2J15M027	Northern (2J)	2015			
GM2J15M056	Northern (2J)	2015			
GM2J15M060	Northern (2J)	2015			
GM2J15M085	Northern (2J)	2015			
GM2J15M088	Northern (2J)	2015			
GM2J15M101	Northern (2J)	2015			
GM2J15M108	Northern (2J)	2015			

GM2J15M109	Northern (2J)	2015			
GM2J15S001	Northern (2J)	2015			
GM2J15S006	Northern (2J)	2015			
GM2J15S009	Northern (2J)	2015			
GM2J15S010	Northern (2J)	2015			
GM2J15S030	Northern (2J)	2015			
GM2J15S031	Northern (2J)	2015			
GM2J15S045	Northern (2J)	2015			
GM2J15S047	Northern (2J)	2015			
GM2J15S052	Northern (2J)	2015			
GM2J15S057	Northern (2J)	2015			

GM2J15S058	Northern (2J)	2015			
GM2J15S061	Northern (2J)	2015			
GM2J15S062	Northern (2J)	2015			
GM2J15S063	Northern (2J)	2015			
GM2J15S064	Northern (2J)	2015			
GM2J15S087	Northern (2J)	2015			
GM2J15S089	Northern (2J)	2015			
GM2J15S096	Northern (2J)	2015			
GM2J15S098	Northern (2J)	2015			
GM2J15S099	Northern (2J)	2015			
GM3K15L156	Northern (3K)	2015			

GM3K15L157	Northern (3K)	2015			
GM3K15L158	Northern (3K)	2015			
GM3K15L169	Northern (3K)	2015			
GM3K15L177	Northern (3K)	2015			
GM3K15L178	Northern (3K)	2015			
GM3K15M154	Northern (3K)	2015			
GM3K15M159	Northern (3K)	2015			
GM3K15M161	Northern (3K)	2015			
GM3K15M164	Northern (3K)	2015			
GM3K15M166	Northern (3K)	2015			
GM3K15M170	Northern (3K)	2015			

GM3K15M181	Northern (3K)	2015			
GM3K15M182	Northern (3K)	2015			
GM3K15M187	Northern (3K)	2015			
GM3K15M188	Northern (3K)	2015			
GM3K15S152	Northern (3K)	2015			
GM3K15S155	Northern (3K)	2015			
GM3K15S167	Northern (3K)	2015			
GM3K15S176	Northern (3K)	2015			
GM4R17L1594	GSL(4R)	2017			
GM4R17M1414	GSL(4R)	2017			
GM4R17M1420	GSL(4R)	2017			

GM4R17M1426	GSL(4R)	2017			
GM4R17M1427	GSL(4R)	2017			
GM4R17M1428	GSL(4R)	2017			
GM4R17M1436	GSL(4R)	2017			
GM4R17M1592	GSL(4R)	2017			
GM4R17S1417	GSL(4R)	2017			
GM4R18L2313	GSL(4R)	2018			
GM4R18M2301	GSL(4R)	2018			
GM4R18M2302	GSL(4R)	2018			
GM4R18M2304	GSL(4R)	2018			
GM4R18M2305	GSL(4R)	2018			

GM4R18M2306	GSL(4R)	2018			
GM4R18M2312	GSL(4R)	2018			
GM4R18S2300	GSL(4R)	2018			
GM4R18S2303	GSL(4R)	2018			
GM4R18S2307	GSL(4R)	2018			
GM4R18S2308	GSL(4R)	2018			
GM4R18S2309	GSL(4R)	2018			
GM4R18S2310	GSL(4R)	2018			
GM4R18S2311	GSL(4R)	2018			
GM4S05M2453	GSL(4S)	2005			
GM4S05S2365	GSL(4S)	2005			

GM4S05S2391	GSL(4S)	2005			
GM4S17L1207	GSL(4S)	2017			
GM4S17L1210	GSL(4S)	2017			
GM4S17M1180	GSL(4S)	2017			
GM4S17M1187	GSL(4S)	2017			
GM4S17M1192	GSL(4S)	2017			
GM4S17M1194	GSL(4S)	2017			
GM4S17M1208	GSL(4S)	2017			
GM4S17M1213	GSL(4S)	2017			
GM4S17M1215	GSL(4S)	2017			
GM4S17M1216	GSL(4S)	2017			

GM4S17M1518	GSL(4S)	2017			
GM4S17M1534	GSL(4S)	2017			
GM4S17M1544	GSL(4S)	2017			
GM4S17S1520	GSL(4S)	2017			
GM4S17S1521	GSL(4S)	2017			
GM4S17S1522	GSL(4S)	2017			
GM4S17S1523	GSL(4S)	2017			
GM4S17S1524	GSL(4S)	2017			
GM4S17S1533	GSL(4S)	2017			
GM4S17S1541	GSL(4S)	2017			
GM4S17S1542	GSL(4S)	2017			

GM4S17S1550	GSL(4S)	2017			
GM4S17S1554	GSL(4S)	2017			
GM4S17S1555	GSL(4S)	2017			
GM4S17S1557	GSL(4S)	2017			
GM4S17S1559	GSL(4S)	2017			
GM4S17S1560	GSL(4S)	2017			
GM4T17L867	GSL(4T)	2017			
GM4T17L914	GSL(4T)	2017			
GM4T17L918	GSL(4T)	2017			
GM4T17L927	GSL(4T)	2017			
GM4T17M875	GSL(4T)	2017			
GM4T17M877	GSL(4T)	2017			
GM4T17M888	GSL(4T)	2017			
GM4T17M889	GSL(4T)	2017			
GM4T17M906	GSL(4T)	2017			
GM4T17M921	GSL(4T)	2017			

GM4T17M924	GSL(4T)	2017			
GM4T17S880	GSL(4T)	2017			
GM4T17S885	GSL(4T)	2017			
GM4T17S886	GSL(4T)	2017			
GM4T17S890	GSL(4T)	2017			
GM4T17S892	GSL(4T)	2017			
GM4T17S895	GSL(4T)	2017			
GM4T17S899	GSL(4T)	2017			
GM4T17S907	GSL(4T)	2017			
GM4T17S910	GSL(4T)	2017			
GM4T17S911	GSL(4T)	2017			
GM4T17S912	GSL(4T)	2017			
GM4T17S913	GSL(4T)	2017			
GM4T17S915	GSL(4T)	2017			
GM4T17S916	GSL(4T)	2017			
GM4T17S919	GSL(4T)	2017			
GM4T17S920	GSL(4T)	2017			
GM4T17S922	GSL(4T)	2017			
GM4T17S925	GSL(4T)	2017			
GM4T17S926	GSL(4T)	2017			
GM4X17L559	BoF	2017			
GM4X17L565	BoF	2017			

GM4X17L567	BoF	2017			
GM4X17L579	BoF	2017			
GM4X17L580	BoF	2017			
GM4X17L581	BoF	2017			
GM4X17L585	BoF	2017			
GM4X17L587	BoF	2017			
GM4X17L591	BoF	2017			
GM4X17L592	BoF	2017			
GM4X17L593	BoF	2017			
GM4X17M558	BoF	2017			
GM4X17M560	BoF	2017			
GM4X17M561	BoF	2017			
GM4X17M562	BoF	2017			
GM4X17M563	BoF	2017			
GM4X17M564	BoF	2017			

GM4X17M566	BoF	2017			
GM4X17M568	BoF	2017			
GM4X17M569	BoF	2017			
GM4X17M570	BoF	2017			
GM4X17M571	BoF	2017			
GM4X17M573	BoF	2017			
GM4X17M574	BoF	2017			
GM4X17M575	BoF	2017			
GM4X17M576	BoF	2017			
GM4X17M577	BoF	2017			
GM4X17M578	BoF	2017			

GM4X17M582	BoF	2017			
GM4X17M583	BoF	2017			
GM4X17M584	BoF	2017			
GM4X17M586	BoF	2017			
GM4X17M588	BoF	2017			
GM4X17M589	BoF	2017			
GM4X17M590	BoF	2017			
GM4X17M594	BoF	2017			
GM4X18_2445	BoF	NoData			
GM4X18L2120	BoF	2018			
GM4X18L2122	BoF	2018			

GM4X18L2128	BoF	2018			
GM4X18M2118	BoF	2018			
GM4X18M2119	BoF	2018			
GM4X18M2121	BoF	2018			
GM4X18M2123	BoF	2018			
GM4X18M2124	BoF	2018			
GM4X18M2125	BoF	2018			
GM4X18M2126	BoF	2018			
GM4X18M2127	BoF	2018			
GM4X18M2129	BoF	2018			
GM4X18M2130	BoF	2018			

GM4X18M2131	BoF	2018			
GM5Y17L1740	GoM	2017			
GM5Y17L1745	GoM	2017			
GM5Y17L1746	GoM	2017			
GM5Y17L1749	GoM	2017			
GM5Y17L1750	GoM	2017			
GM5Y17L1752	GoM	2017			
GM5Y17L1759	GoM	2017			
GM5Y17L1760	GoM	2017			
GM5Y17L1764	GoM	2017			
GM5Y17L804	GoM	2017			
GM5Y17L810	GoM	2017			

GM5Y17L814	GoM	2017			
GM5Y17L816	GoM	2017			
GM5Y17L822	GoM	2017			
GM5Y17L829	GoM	2017			
GM5Y17L830	GoM	2017			
GM5Y17L832	GoM	2017			
GM5Y17L833	GoM	2017			
GM5Y17L835	GoM	2017			
GM5Y17M1734	GoM	2017			
GM5Y17M1735	GoM	2017			
GM5Y17M1736	GoM	2017			
GM5Y17M1738	GoM	2017			
GM5Y17M1739	GoM	2017			
GM5Y17M1743	GoM	2017			
GM5Y17M1744	GoM	2017			

GM5Y17M1748	GoM	2017			
GM5Y17M1751	GoM	2017			
GM5Y17M1753	GoM	2017			
GM5Y17M1754	GoM	2017			
GM5Y17M1755	GoM	2017			
GM5Y17M1756	GoM	2017			
GM5Y17M1757	GoM	2017			
GM5Y17M1758	GoM	2017			
GM5Y17M1761	GoM	2017			
GM5Y17M1763	GoM	2017			
GM5Y17M806	GoM	2017			

GM5Y17M807	GoM	2017			
GM5Y17M808	GoM	2017			
GM5Y17M809	GoM	2017			
GM5Y17M811	GoM	2017			
GM5Y17M812	GoM	2017			
GM5Y17M813	GoM	2017			
GM5Y17M815	GoM	2017			
GM5Y17M817	GoM	2017			
GM5Y17M818	GoM	2017			
GM5Y17M821	GoM	2017			
GM5Y17M823	GoM	2017			

GM5Y17M824	GoM	2017			
GM5Y17M825	GoM	2017			
GM5Y17M828	GoM	2017			
GM5Y17M831	GoM	2017			
GM5Y17M834	GoM	2017			
GM5Y17M836	GoM	2017			
GM5Y17M837	GoM	2017			
GM5Y17M839	GoM	2017			
GM5Y17M841	GoM	2017			
GM5Y17S1650	GoM	2017			
GM5Y17S1737	GoM	2017			

GM5Y17S1741	GoM	2017			
GM5Y17S805	GoM	2017			
GM5Y17S819	GoM	2017			
GM5Y17S820	GoM	2017			
GM5Y17S840	GoM	2017			
GM5Y18L1766	GoM	2018			
GM5Y18L1768	GoM	2018			
GM5Y18L2188	GoM	2018			
GM5Y18L2189	GoM	2018			
GM5Y18L2190	GoM	2018			
GM5Y18L2191	GoM	2018			
GM5Y18L2192	GoM	2018			
GM5Y18M1765	GoM	2018			

GM5Y18M1767	GoM	2018			
GM5Y18M2187	GoM	2018			
GM5Y18M2195	GoM	2018			
GM5Y18S2193	GoM	2018			
GM5Y18S2194	GoM	2018			
GM5Z17L1820	GoM	2017			
GM5Z17M1821	GoM	2017			
GM5Z18M2186	GoM	2018			
Greg_BoF_783 (originalGM5Y17J783)	BoF	2017			
Greg_BoF_848	BoF	2017			

(originalGM4 X17J848)					
Greg_BoF_850 (originalGM4 X17J850)	BoF	2017			
Greg_BoF_851 (originalGM4 X17J851)	BoF	2017			
Greg_BoF_852 (originalGM4 X17J852)	BoF	2017			
Greg_BoF_853 (originalGM4 X17J853)	BoF	2017			
Greg_BoF_854 (originalGM4 X17J854)	BoF	2017			

Greg_BoF_855 (originalGM4X17J855)	BoF	2017			
Greg_BoF_857 (originalGM4X17J857)	BoF	2017			
Greg_GoM_609 (originalGM5Y17Y609)	BoF	2017			
Greg_GoM_618 (originalGM5Y17Y618)	GoM	2017			
Greg_GoM_619 (originalGM5Y17Y618)	GoM	2017			
Greg_GoM_623	GoM	2017			

(originalGM5 Y17Y623)					
Greg_GoM_628 (originalGM5 Y17Y628)	GoM	2017			
Greg_GoM_629 (originalGM5 Y17Y629)	GoM	2017			
Greg_GoM_636 (originalGM5 Y17Y636)	GoM	2017			
Greg_GoM_640 (originalGM5 Y17Y640)	GoM	2017			
Greg_GoM_641 (originalGM5 Y17Y641)	GoM	2017			

Greg_GoM_650 (originalGM5Y17Y650)	GoM	2017			
Greg_GoM_826 (originalGM5Y17S826)	GoM	2017			
Greg_GoM_827 (originalGM5Y17M827)	GoM	2017			
Greg_GoM_838 (originalGM5Y17L838)	GoM	2017			
Greg_GoM_842 (originalGM5Y17L842)	GoM	2017			
Greg_North3L_1002	Northern (3L)	2017			

(originalGM3L17_2071)					
Greg_North3L_953 (originalGM3L17_2022)	Northern (3L)	2017			
Greg_North3L_959 (originalGM3L17_2028)	Northern (3L)	2017			
Greg_North3L_976 (originalGM3L17_2045)	Northern (3L)	2017			
Greg_North3L_978 (originalGM3L17_2047)	Northern (3L)	2017			
Greg_North3L_980 (originalGM3L17_2049)	Northern (3L)	2017			

Greg_North3L_981 (originalGM3L17_2050)	Northern (3L)	2017			
Greg_North3L_982 (originalGM3L17_2051)	Northern (3L)	2017			
Greg_North3L_983 (originalGM3L17_2052)	Northern (3L)	2017			
Greg_North3L_984 (originalGM3L17_2053)	Northern (3L)	2017			
Greg_North3L_988 (originalGM3L17_2057)	Northern (3L)	2017			
Greg_North3L_990	Northern (3L)	2017			

(originalGM3L17_2059)					
Greg_North3L_991 (originalGM3L17_2060)	Northern (3L)	2017			
Greg_North3L_995 (originalGM3L17_2064)	Northern (3L)	2017			
Greg_North3L_998 (originalGM3L17_2067)	Northern (3L)	2017			
Greg_SL4Rc_258 (originalGM4R17M1124)	GSL(4R)	2017			
Greg_SL4Rc_262 (originalGM4R17L1128)	GSL(4R)	2017			

Greg_SL4Rc_267 (originalGM4R17L1133)	GSL(4R)	2017			
Greg_SL4T_746 (originalGM4T17S1611)	GSL(4R)	2017			
Greg_SL4T_750 (originalGM4T17S1615)	GSL(4R)	2017			
Greg_SL4T_751 (originalGM4T17S1616)	GSL(4R)	2017			
Greg_SL4T_754 (originalGM4T17M1619)	GSL(4R)	2017			
Greg_SL4T_756	GSL(4R)	2017			

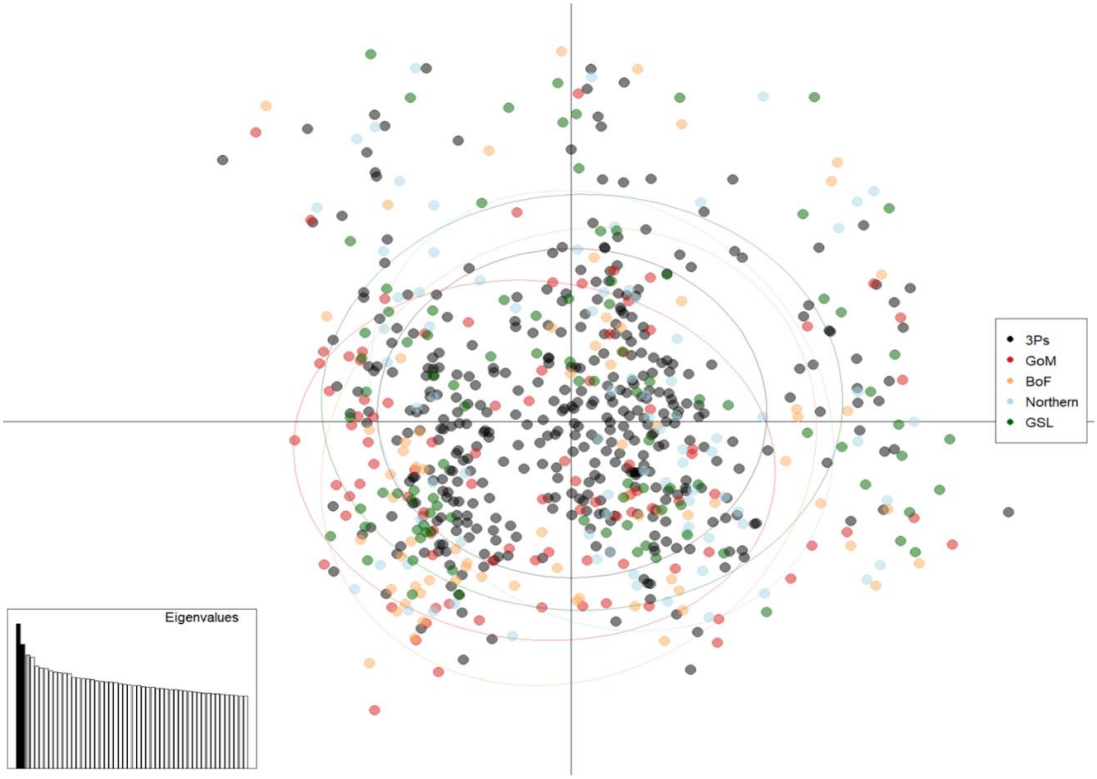
(originalGM4T17M1621)					
Greg_SL4T_765 (originalGM4T17S1630)	GSL(4R)	2017			

Appendix II

Datasets used and the loci they include/exclude.

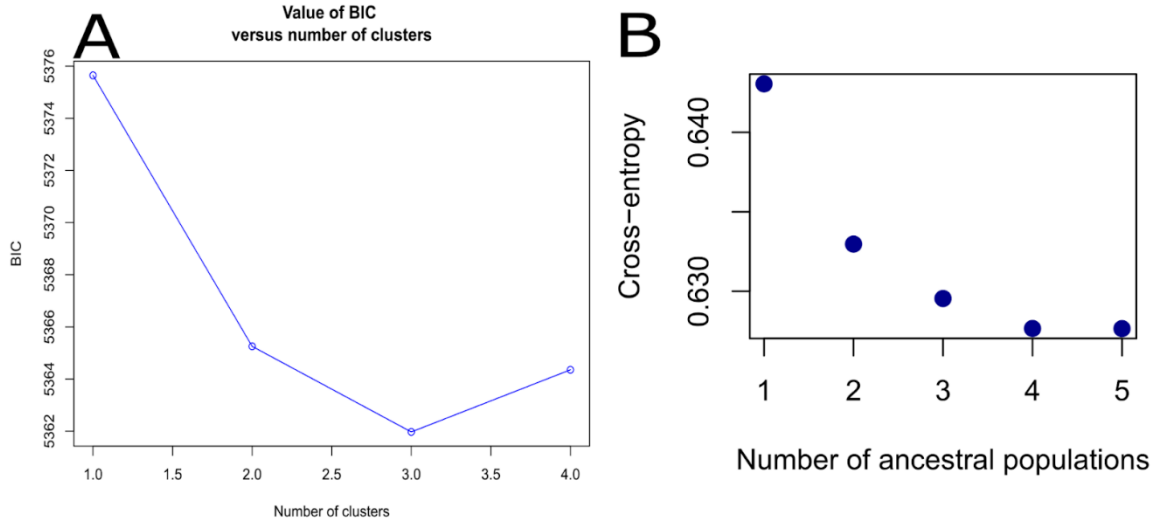
Dataset Name	Includes	Excludes
Whole Dataset	Neutral Loci, Inversions, all Loci under selection	Lane Effect Loci, Outliers
Neutral Dataset	Neutral Loci	Lane Effect Loci, Outliers, Loci under selection, Inversions

Appendix III



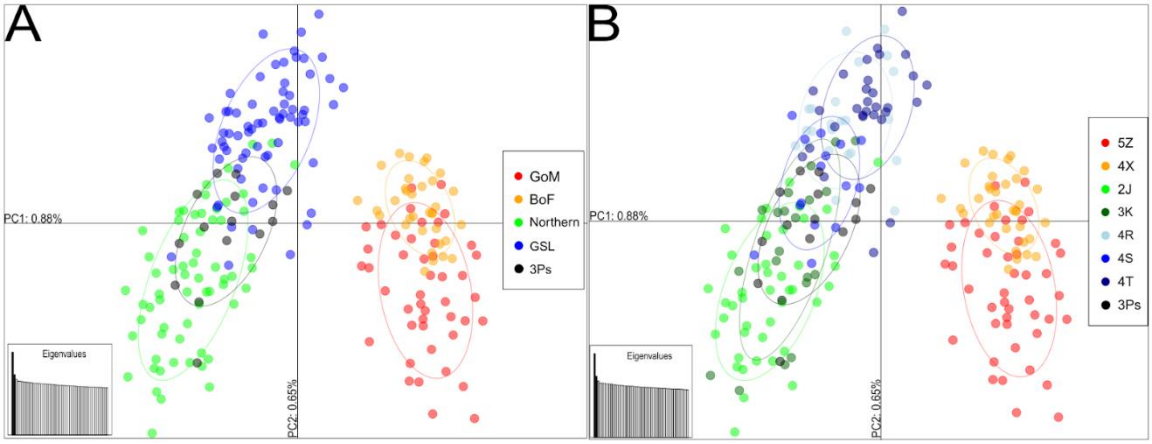
DP20 Neutral dataset with 1968 loci. Dataset filtered at read depth of 20 as opposed to original 10. Lane effect is completely removed. No population structure is seen, even between Northern and Southern populations. Inertia ellipses (default size 1.5) show where the majority of individuals cluster for each group.

Appendix IV



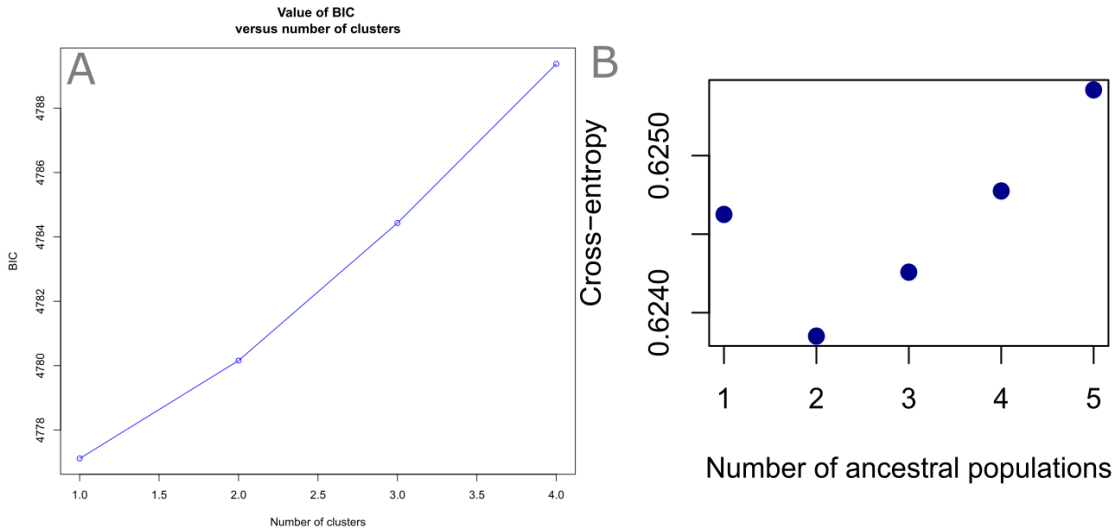
A) Bayesian Information Criterion (BIC) and B) Cross-Entropy Plots for Whole Dataset. The BIC plot suggests 3 clusters as it has a dip in BIC values at 3, whereas the cross-entropy plot suggests around 4 ancestral populations.

Appendix V



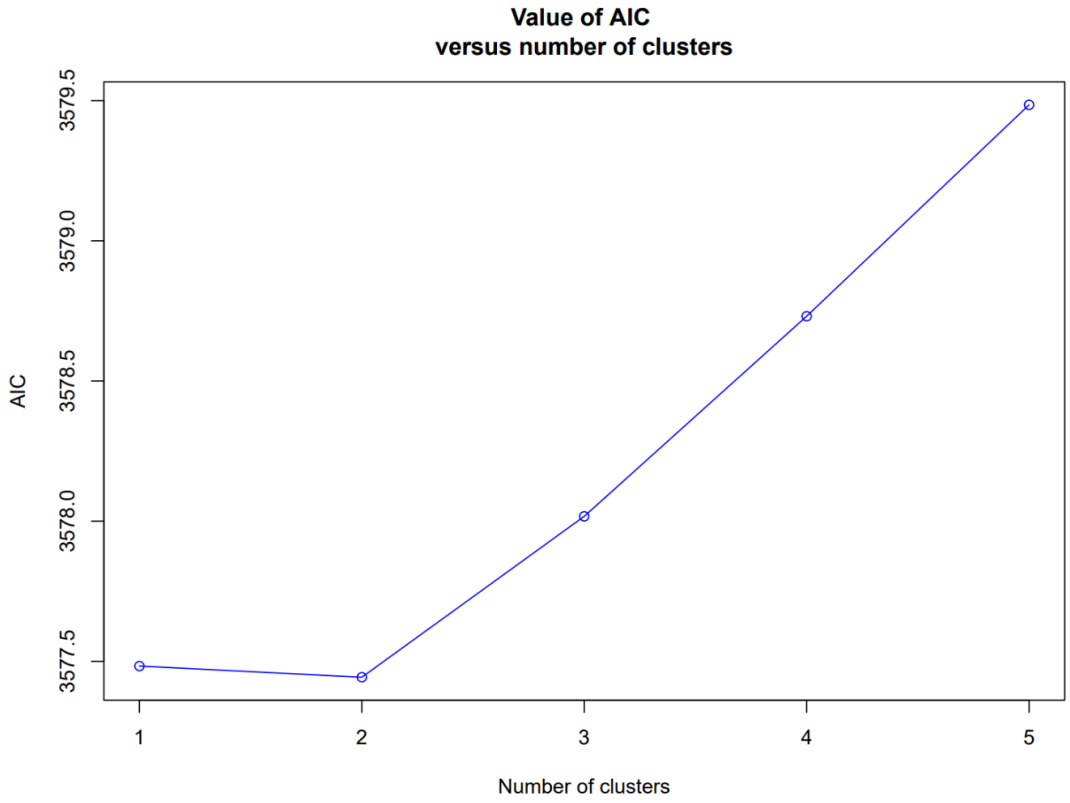
Neutral Dataset PCAs with only spawning individuals labelled A)by population and B)by fishery.

Appendix VI



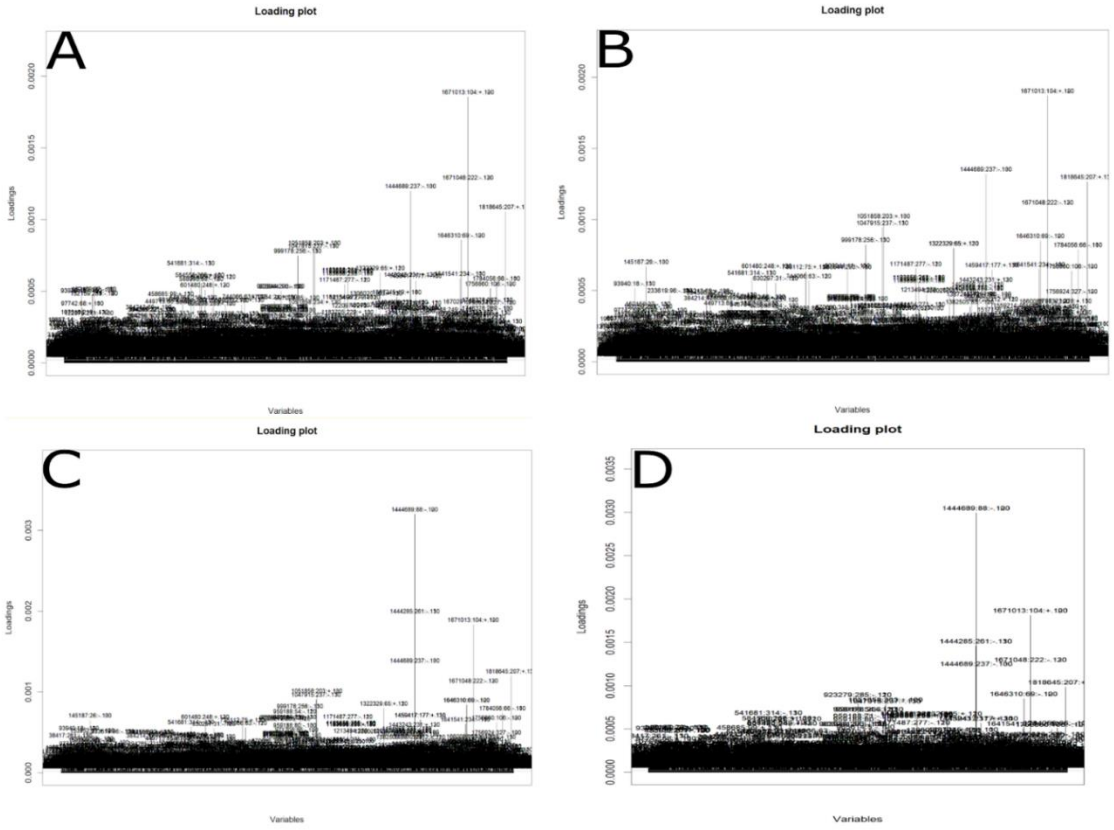
A) Bayesian Information Criterion (BIC) and B) Cross-Entropy Plots for Neutral Dataset. Loci from Linkage Disequilibrium, inversions, lane effect analyses, and outlier analyses removed.

Appendix VII



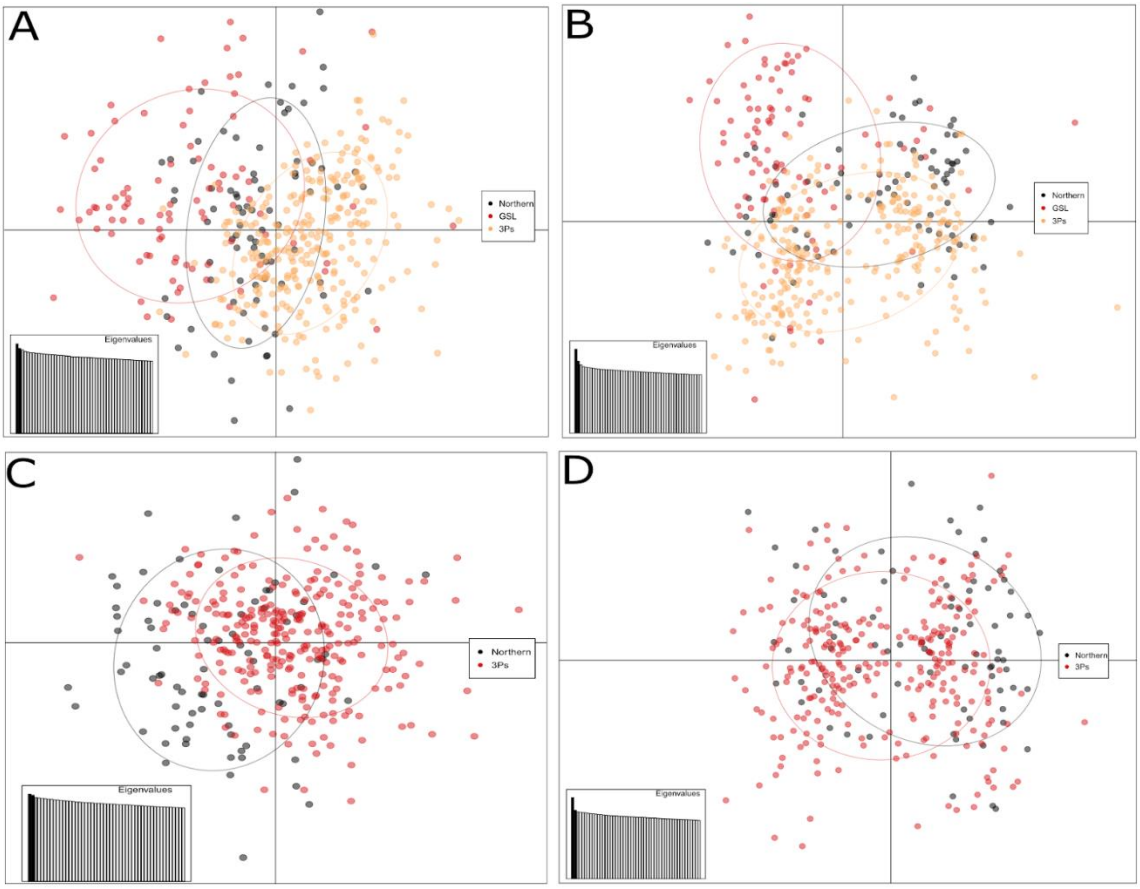
Akaike's Information Criterion Plot from K-means clustering for two neutral clusters including Northern, GSL, and 3Ps populations. Suggests 2 clusters.

Appendix VIII



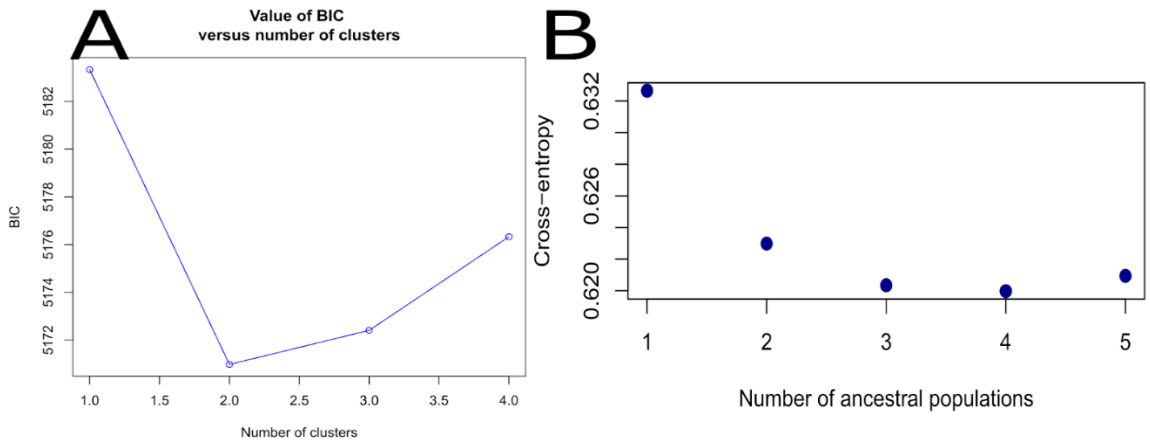
Loading Plots for 2 Neutral Northern clusters. A) outliers and GSL removed. B) outliers removed, with GSL. C) with outliers, with GSL. D) with outliers, GSL removed. Overall showing genome-wide neutral variation in all.

Appendix IX



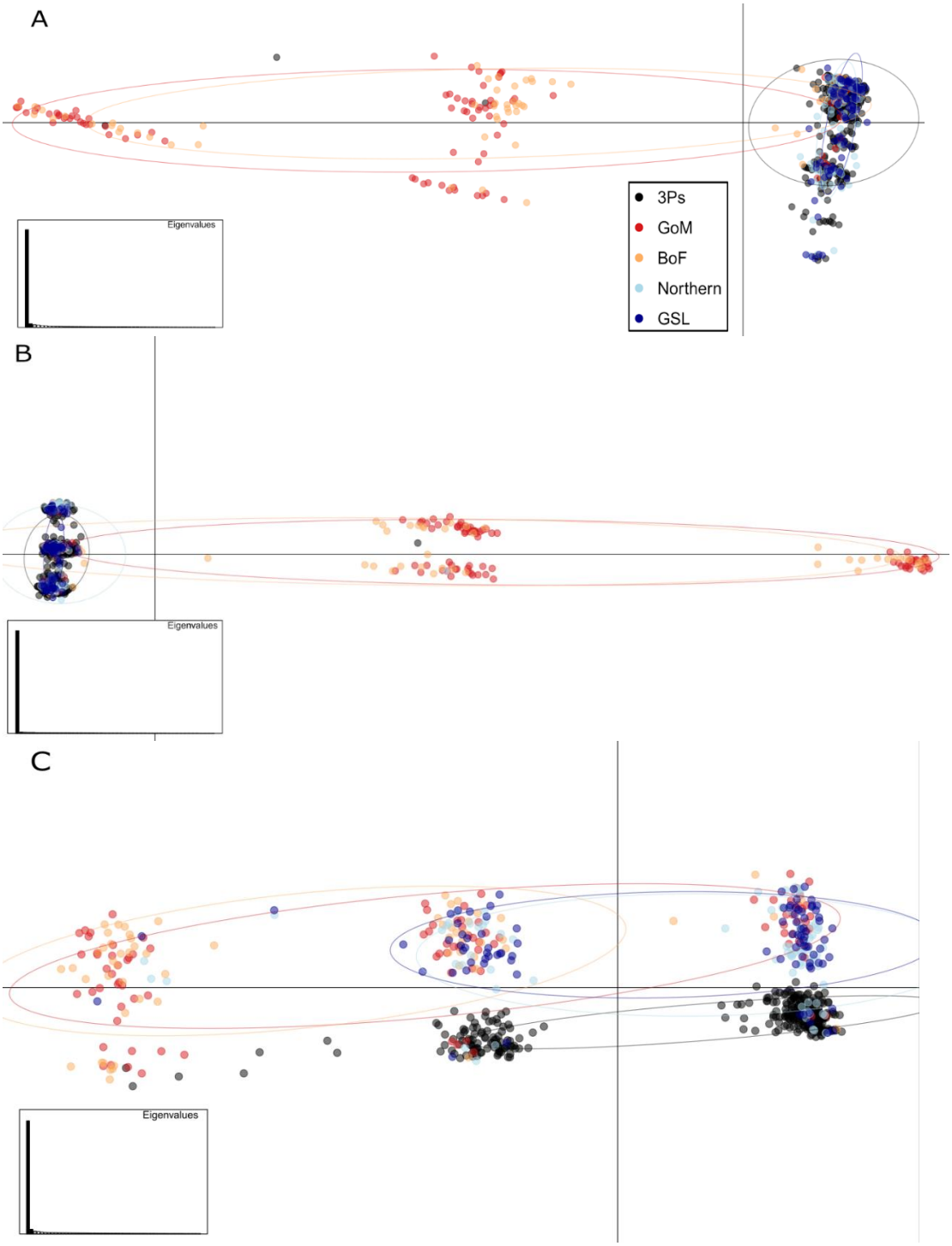
Loadings removed from the 2 neutral Northern/3Ps/GSL clusters. A) Top 1000 loadings removed, 2 clusters no longer present. B) Top 150 loadings removed, 2 clusters still present. C) Top 1000 loadings removed, no GSL, 2 clusters still not present. D) Top 150 loadings removed, no GSL, 2 clusters still present. Inertia ellipses (default size 1.5) show where the majority of individuals cluster for each group.

Appendix X



680 LG01 loci A)BIC and B)Cross-Entropy Plot. BIC suggests 2 clusters, cross-entropy suggests 2 ancestral pops.

Appendix XI



PCAs of *A)*174 LG02 loci, *B)*448 LG07 loci *C)*147 LG12 loci. Legend for all panels is the same as panel A.

Appendix XII

χ^2 Test for Hardy-Weinberg Equilibrium (HWE) LG01 Allele Frequencies between Adult and Juvenile 3Ps Cod. LG01 inversion was treated as one allele. LG01 was in HWE for juveniles but not mature individuals. Bonferroni adjusted p-values are present.

Total Individuals:	317				
Number of Juveniles:	190				
Number of Mature:	127				
Juvenile Heterozygous	94	**HWE Frequencies**			
Juvenile Homozygous Right	35	**p (right)**	82	0.431579	**Chi^2 = 7.14**
Juvenile Homozygous Left	61	**q (left)**	108	0.568421	**p-value=0.028**
Total Juvenile:	190	**Total**	190	1	**Adjusted p-value=0.056**
					For Juveniles: LG01 in HWE
1:2:1 Expected Ratio	**Observed**	**Expected**			
AA	35	47.5			
AB	94	95			
BB	61	47.5			
(O-E)^2/E	(O-E)^2/E	(O-E)^2/E	SUM	df	critical value
3.289473684	0.010526	3.836842	**7.136842**	2	5.991
Mature Heterozygous:	61	**HWE Frequencies**			
Mature Homozygous Right:	11	**p (right)**	41.5	0.326772	**Chi^2 = 30.69**
Mature Homozygous Left:	55	**q (left)**	85.5	0.673228	**p-value<0.0001**
Total Mature	127	**Total**	127	1	**Adjusted p-value<0.0001**
					For Mature: LG01 not in HWE
1:2:1 Expected Ratio	**Observed**	**Expected**			
AA	11	31.75			
AB	61	63.5			
BB	55	31.75			
(O-E)^2/E	(O-E)^2/E	(O-E)^2/E	SUM	df	critical value
13.56102362	0.098425	17.02559	**30.68504**	2	5.991

Appendix XIII

Fisher's Exact Test for Hardy-Weinberg Equilibrium for LG01 genotypes for each age group.

Age Group	Number of Individuals	Observed:Expected			Bonferroni Adjusted P-value
1	3	AA	1	0.75	p=1
		AB	0	1.5	In HWE
		BB	2	0.75	
2	54	AA	17	13.5	p=1
		AB	25	27	In HWE
		BB	12	13.5	
3	64	AA	6	16	p=0.46
		AB	33	32	in HWE
		BB	25	16	
4	45	AA	5	11.25	p=1
		AB	27	22.5	in HWE
		BB	13	11.25	
5	40	AA	3	10	p=0.74
		AB	20	20	in HWE
		BB	17	10	
6	25	AA	4	6.25	p=1
		AB	13	12.5	in HWE

		BB	8	6.25	
7	15	AA	5	4.5	p=1
		AB	6	9	in HWE
		BB	7	4.5	
8	17	AA	0	4.25	p=1
		AB	10	8.5	in HWE
		BB	7	4.25	
9	16	AA	1	4	p=1
		AB	8	8	in HWE
		BB	7	4	
10	13	AA	0	3.25	p=0.572
		AB	4	6.5	in HWE
		BB	9	3.25	
11	6	AA	0	1.5	p=1
		AB	3	3	in HWE
		BB	3	1.5	
12	1	AA	0	0.25	p=1
		AB	0	0.5	in HWE
		BB	1	0.25	
15	1	AA	0	0.25	p=1
		AB	1	0.5	in HWE
		BB	0	0.25	

Curriculum Vitae

Candidate's full name: Sarah Babaei

Universities attended (with dates and degrees obtained):

McMaster University (BSc Hons Molecular Biology and Genetics, June 2022)

Publications:

S Babaei, LA Grieves, BJ Evans, JS Quinn. Multiplex PCR reveals population structure in an inbred communal bird. *Plos One*. Submitted.

Conference Presentations:

S Babaei, NM LeBlanc, SA Pavey, CC D'Aloia, D Varkey, A Adamack. 2023. Determining stock structure of cod in a Newfoundland subdivision using next-generation sequencing. Science Atlantic Aquaculture & Fisheries (A&F) and Biology Conference (In-Person, University of New Brunswick Saint John)

S Babaei, LA Grieves, BJ Evans, JS Quinn. 2022. Optimization of multiplex PCR to determine population structure and heterozygosity in an inbreeding communal bird. Society of Canadian Ornithologists Annual Conference (Virtual).

S Babaei, LA Grieves, BJ Evans, JS Quinn. 2022. Optimization of multiplex PCR to determine population structure and heterozygosity in an inbreeding communal bird. Biology Undergraduate Symposium (In-Person, McMaster University)