Investigating the genetic structure of Atlantic Cod (*Gadus morhua*) spawning populations following the collapse of the Northern cod stock

by

Ty Gormley

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

Bachelor of Science with Honours in General Biology

Supervisor: Dr. Scott Pavey, Department of Biological Sciences

THE UNIVERSITY OF NEW BRUNSWICK

SAINT JOHN

April, 2019

ABSTRACT

The Atlantic cod (*Gadus morhua*), within the Northern cod stock, once served as the economic pillar for Newfoundland and Labrador. After years of overfishing and reduced ocean temperature, the stock population collapsed in 1992, driving many to question the genetic structure for its recovery. Currently, the metapopulation and isolation hypotheses have been proposed to describe the genetic structure of the stock, but there is much debate on which is reliable. The stocks genetic structure will be determined using fin clips obtained from 2001-2003, from various spawning populations, to indicate whether the stock exhibits panmixia. A double-digest restriction site associated DNA technique was used to identify 846 single nucleotide polymorphisms, and the population structure was determined using the Fixation Index (F_{ST}). The F_{ST} analysis found no significant differences, therefore, spawning populations of the stock are panmictic. Further research is necessary to obtain a more robust genetic structure for extensive ecological management.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Scott Pavey for the continuous support of my Honours study. I would like to also thank Dr. Greg Puncher, and my lab mates at the Canadian Rivers Institute Genomics Lab for providing guidance and assistance on this study. Thank you to Dr. Sherrylynn Rowe, Dr. George Rose, and Dr. John Nelson for providing the Atlantic cod samples. Finally, thank you to my parents, Chris Gormley and Theresa Keane-Gormley, and my friends for their constant support.

STATEMENT OF RESEARCH CONTRIBUTION

In June 2018, I began researching the literature about Atlantic cod (*Gadus morhua*) within the Northern cod stock. While researching the literature, I began the NGS pipeline to yield three libraries consisting of cod individuals from various spawning populations within the stock. The libraries were then sent to Génome Québec for sequenced, and the raw reads were obtained in November. In January 2019, I began bioinformatics to process and filter my data to isolate loci that are polymorphic among all the spawning populations within the stock. In conjunction with Dr. Scott Pavey, I ran my statistical analysis on the loci to determine whether cod of various spawning populations are panmictic.

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

F_{ST}: Fixation Index

NGS: Next Generation Sequencing

SNPs: Single nucleotide polymorphisms

DNA: Deoxyribonucleic acid

ddRAD: Double digested restriction site associated DNA

dsDNA: Double single deoxyribonucleic acid

1xTE: 10Mm Tris-HCl Tris pH buffer and EDTA [Ethylenediaminetetraacetic acid]

Tris-HCl: trisaminomethane-hydrochloric acid

PCR: Polymerase chain reaction

dNTPs: Deoxynucleotides

F_{IS}: Inbreeding coefficient

DAPC: discriminant analysis of principle components

INTRODUCTION

The Atlantic cod (*Gadus morhua*) is a pelagic species found throughout the northern Atlantic Ocean (Brander, 1994). It has supported fisheries that were both high abundance and high value, particularly within the Northern cod stock off the coast of Newfoundland, Canada (Rose *et al*., 2004; Lear, 1998). The cod industry, within the Northern cod stock, began in the late 15th century when European fishermen travelled to the waters of Newfoundland and Labrador to fish in one of the world's most productive cod stocks (Rose *et al*., 2004; Lear, 1998). However, it took until the 20th century for the cod production within the province to blossom into a world leader in the cod industry (Rose *et al*., 2004; Mason, 2002). The industry was recognized to be the single most important factor for fostering development within the province (William, 2005; Lear, 1998). In the 1960's while the industry peaked, the profits of the industry affected not only fishing communities, but the province at large (Myers *et al*., 1997; Lear, 1998). Taking advantage of rapid advances in technology, such as state-of-the-art boats, fishermen were catching millions of tons of cod each year in the stock (William & Noel, 2013; Lear, 1998). However, cod abundance within the Northern cod stock decreased drastically in the 1990's, resulting in the collapse of the stock and the formation of the Cod Moratorium which largely persist to date (Myers *et al*., 1997; Rose, 1994; Rochet, 1998). Due to the socio-economic importance of the cod industry, research has focused on determining possible causes for the stocks collapse (Rose *et al.*, 1994). Currently, researchers have credited the collapse of the stock to two key pressures: overfishing and a decrease in ocean temperature (Rose *et al*., 1994; Rochet, 1998).

Atlantic cod has supported large fisheries for centuries but overfishing within Northern cod stock contributed to the collapse of the stock (Lear & Myers, 1997). Prior to the crash of the

cod stock, cod catch by the fishing industry fluctuated between 400 to 700 thousand tons until the 1950's and gradually increased in the subsequent years exceeding 1 million tons (Lear & Myers 1997; Lear, 1998). In 1968, the industry peaked, catching 1.9 million tons of cod (Figure 1) (Lear, 1998). However, after centuries of intense exploitation, the cod industry began to reveal signs of weakness, noted by the reduced amount of cod catch during the 1970's (Lear, 1998). In 1980's the Canadian Government provided numerous cod abundance estimations within the Northern cod stock; however, the results were continuous over-estimations of the cod abundance (Lear, 1998). Nonetheless, the cod catch from the industry gradually increased in the subsequent years (Lear, 1998). Due to the subsequent increase in cod catch during the 1980's, the pressure exerted on the stock was amplified on an already reduced population (Lear, 1998). Therefore, the additional pressure has been recognized as a major factor in the collapse of the Northern cod stock, which led to the formation of the cod moratorium by the Canadian Government (Lear & Myers 1997; Lear, 1998).

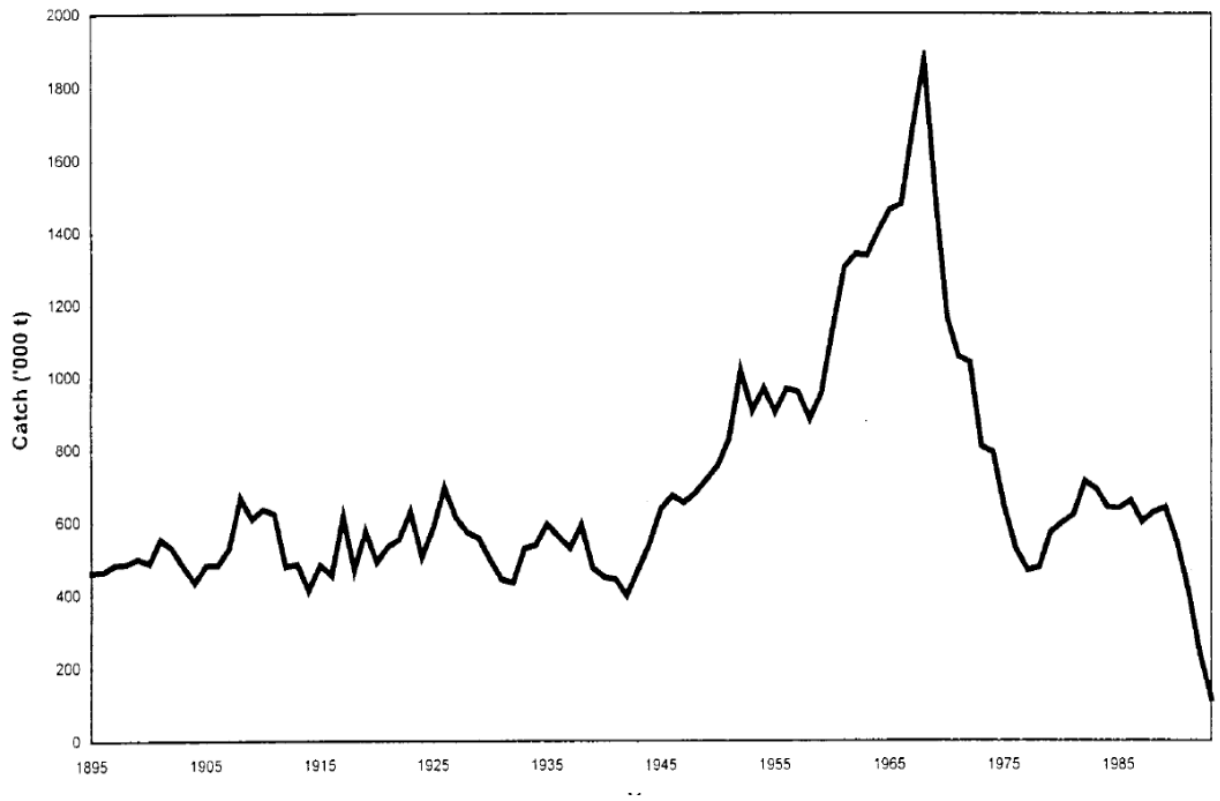

Figure 1: Nominal catches of Atlantic cod in the Northwest Atlantic, 1895-1993. Obtained from Lear, 1998, with approval.

Environmental factors also contributed to the collapse; two major cooling periods occurred in the waters off of Newfoundland and Labrador during the years leading up to the collapse of the stock, primarily within the Labrador Current which flows through the Northern cod stock (Rose *et al*., 1994; Petrie *et al.*, 2010; Petrie & Drinkwater, 1993). The periods of major ocean water cooling occurred during the years 1983-1985 and 1990-1992 (Drinkwater, 1996; Rose *et al*., 1994). During the former major cooling period, the ocean temperature decreased by approximately 1°C; however, the ocean temperature quickly rebounded in the subsequent years (Drinkwater, 1996; Rose *et al*., 1994). Following the first major cooling period, the ocean temperature within the Labrador Current continued to decrease leading up to the second major cooling period of 1990-1992 (Figure 2) (Drinkwater, 1996; Rose *et al*., 1994). As the second major cooling period occurred during the year 1990-1992, the decrease in ocean temperature exceeded 1°C (Figure 2) (Drinkwater, 1996; Rose *et al*., 1994). The decrease in ocean temperature for the latter major cooling period coincided with the decrease of cod population due to overfishing. Furthermore, the decrease in ocean temperature affected capelin, the preferred food choice of cod in the region, as the population shifted southward to warmer water during the 1990's (Lilly, 1994; Sherwood *et al*., 2007). The capelin-cod connection hypothesis highlights the importance of high capelin abundance to support sufficient cod abundance, therefore, given the reduced levels of capelin within the stock during the 1990's, it can be inferred that the limit food resources, caused by the decrease in ocean temperature, contributed to the decrease in cod abundance (Rose *et al*., 1994, Sherwood *et al*., 2007, Rose & O'Driscoll, 2002).

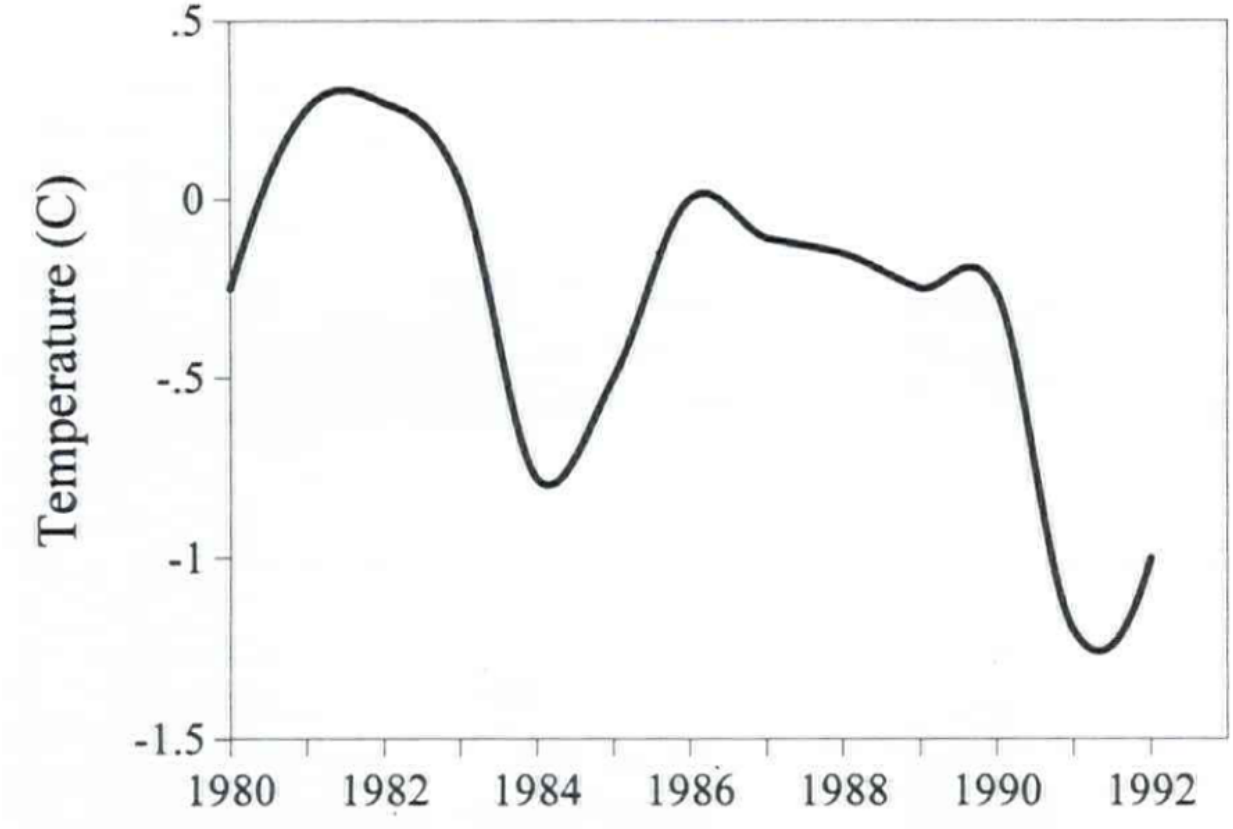

Figure 2: Depth-integrated annual temperature signal at Station 27 off St. John's, Newfoundland. Obtained from Rose *et al.*, 1994, with approval.

Prior to the collapse of the Northern cod stock morphological differences, primarily vertebra variations, among cod within different spawning population were documented (Lear & Wells, 1982). Pre-stock collapse inshore juvenile cod, which spawn from Notre-Dame Channel, have, on average, fewer vertebra when compared to cod of adjacent spawning populations (Lear & Wells, 1982). Furthermore, both inshore and offshore Labrador cod spawn within the same offshore spawning grounds, yet vertebral means of Labrador offshore cod are significantly lower than those of inshore cod (Templeman, 1979; Templeman, 1981) Considering that morphologic variations among pre-stock collapse cod of same and different spawning populations within the same stock have been documented, it's possible to conclude that genetic variations could also be present.

Previous research using microsatellites and single nuclear genes in post-stock collapse cod from offshore spawning populations, has indicated that cod from adjacent offshore spawning populations and inshore populations are genetically different (Rose *et al*., 2011; Ruzzante *et al*., 1996; Beacham *et al*., 2002). Pairwise F_{ST} showed significant, yet small, genetic differences between Hawke Channel and offshore spawning populations of Halibut Channel (F_{ST}=0.0076, *P*=0.0014), Bonavista Corridor (F_{ST}=0.0037, *P*=0.0051), and the inshore spawning population of Smith Sound (F_{ST}=0.0044, *P*=0.0040) (Rose *et al*., 2011). Furthermore, a small genetic difference between Smith Sound and Bonavista Corridor (F_{ST}=0.0013, *P*=0.0156), was also determined (Rose *et al*., 2011). Cod from inshore spawning populations, such as Trinity Bay, were also found to be genetically different than cod of all offshore spawning populations (Beacham *et al*., 2002). These findings contributed to earlier data which has indicated that offshore spawning populations, such as Hawke Channel, are heterogenous to adjacent spawning populations (Beacham *et al*., 2002). A possible reason for the genetic differences among

spawning populations within the stock is that inshore overwintering cod have higher levels of antifreeze within their blood, compared to offshore overwintering cod which have limited levels (Ruzzante *et al*., 1996).

Population structure within the Northern cod stock is a highly debated topic, as data from various studies has supported two conflicting hypotheses. As previously mentioned, data has suggested that cod of offshore spawning populations are genetically different than spawning populations of both adjacent offshore and inshore regions (Rose *et al*., 2011; Ruzzante *et al*., 1996; Beacham *et al*., 2002). Therefore, these data gave rise to the "Isolation Hypothesis", which describes small, individual spawning populations that are genetically distinct within the Northern cod stock (Ruzzante *et al*., 1996; Beacham *et al*., 2002). Furthermore, this hypothesis suggests that during the collapse of the stock, the isolated spawning populations genetically diverged and now that the stock has slowly increased in biomass, cod within the stock are genetically different (Ruzzante *et al*., 1996; Beacham *et al*., 2002). However, data from other colleagues in the field have indicated that cod within the spawning populations of the stock are not genetically different, thereby fueling the opposing argument, the "Metapopulation Hypothesis" (Smedbol & Wroblewski, 2002; Hanski & Gilpin, 1997; Hanski & Simberloff, 1997). The latter hypothesis states that as ocean temperatures decreased during the 1990's, cod migrated towards coastal areas for a period of several breeding seasons (Ruzzante *et al.*, 2000; Green & Wroblewski, 2000). While near the coast, cod biomass increased due to mass inbreeding from cod of various aggregated spawning populations (Ruzzante *et al*., 2000; Green & Wroblewski, 2000). As the ocean temperature rebounded, cod migrated back to their original spawning populations and are now genetically similar to cod of neighbouring spawning populations (Ruzzante *et al*., 1997; Smedbol & Wroblewski, 2002).

A new genetic sequencing approach, Next Generation Sequencing (NGS), has potential for much greater power and resolution to identify whether spawning populations within the stock exhibit panmixia (Pujolar *et al.*, 2014). The technique is capable of considering whole genomes by sequencing millions of small DNA (deoxyribonucleic acid) fragments that can be aligned with a reference genome (Behjati & Tarpey, 2013). Following the alignment, single nucleotide polymorphisms (SNPs) can be identified (Behjati & Tarpey, 2013). SNPs are a common genetic marker used in many NGS studies and are the substitution of a single base pair (e.g. adenine (A), thymine (T), cytosine (C), or guanine (G)) and exhibit sources of genetic variation within species (Paterson *et al*., 2012; Syvanen, 2001). It is key to note that these SNPs have the potential to exhibit direct effects on the expression or structure of certain genes and can code for specific proteins that have the ability to alter the phenotype of the individual (Bush, 2012). In order to produce the genotypes, a specific restriction digestion with the aid of enzymes is used to cut specific sites in DNA fragments (Davey, 2011). These regions are then sequenced with regions beside the cut sites as a comparison, to identify different DNA markers that represent a gene that is associated with a specific chromosome (Davey, 2011). Double digested restriction site associated DNA (ddRAD) sequencing is a specific NGS technique that uses two restriction enzymes to digest DNA, thereby increasing the confidence in the discovery of genetic markers (Peterson *et al.*, 2012).

The objective of this study is to use these high-resolution tools to determine whether spawning populations of Atlantic cod soon after the collapse of the Northern cod stock, are panmictic. The study will be done by genotyping 180 historic cod individuals from various spawning regions in the stock at thousands of SNPs. A genome-wide association study and the Fixation Index (F_{ST}) will be conducted utilizing the SNPs to determine whether the spawning

grounds within the stock are genetically similar (Szymczak *et al*., 2016). It is hypothesized that the historic cod, within the spawning populations, are significantly panmictic.

METHODS

Sampling

In January 2001 and 2002, fin clips were obtained from 791 wild caught cod from overwintering aggregations within spawning regions of the Northern cod stock (Rose *et al.*, 2011). The fin clips, from the wild caught cod, were obtained by the work of Dr. John Nelson, Dr. George Rose, and Dr. Sherrylynn Rowe. The spawning populations were identified using the Northwest Atlantic Fisheries Organization (NAFO) Divisions. Samples were obtained from Hawke Channel (2J), Funk Island (3K), Notre Dame Channel (3K), Bonavista Corridor (3L), Smith Sound (3L), Halibut Channel (3Ps), Clattice Harbour (3Ps), Western Cove (3Ps), and Eastern Channel (3Ps). Sample sites were then geographically grouped into 6 spawning populations: Hawke Channel, Notre-Dame Channel, Bonavista Corridor, Smith Sound, Halibut Channel, and Placentia Bay (Figure 3). A total of 180 cod fin clips, from the six populations were used in the study. Fin clips from approximately equal numbers of randomly selected fish from different size groups (e.g. 21-30 cm, 31-40 cm), spanning the length range of the fish caught, thereby limiting confounding variables. The fin clips were immersed in 95% ethanol for storage and kept at The University of Victoria prior to DNA extractions.

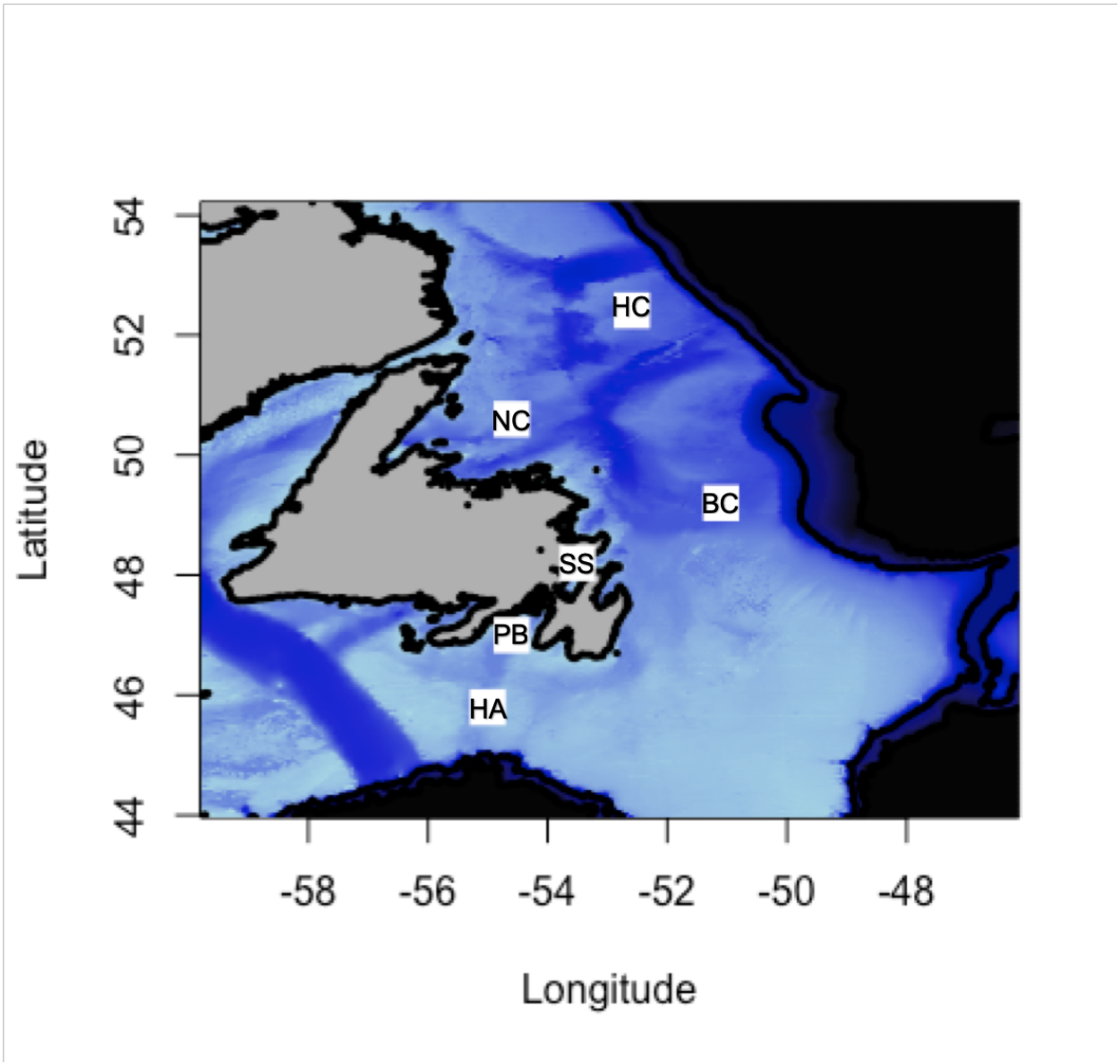

Figure 3: Bathymetric map indicating the genetic samples and places referred to in the article, off of the coast of eastern Newfoundland and Labrador, with the six spawning populations (i.e. Hawke Channel (HC), Notre-Dame Channel (NC), Bonavista Corridor (BC), Smith Sound (SS), Halibut Channel (HA), and Placentia Bay (PB)).

DNA Extractions

The first step of the Next Generation Sequencing (NGS) Pipeline is DNA Extractions. DNA extractions were performed on 180 cod fin clips using the QIAGEN DNeasy ® Blood & Tissue Kit (250) CAT: 69506 kit and its corresponding protocol (Table 1). Steps of the protocol were followed accordingly; however, two modifications were made. The first modification occurred during digestion phase of the extraction as the samples were shaken overnight, instead of the recommended 3-hour period, in a shaking incubator to produce a homogenous result. The second modification was performed as a volume of 75µL, instead of the recommended 100µL, was used to elute the DNA. In addition, the latter step was performed twice, yielding a final volume of 150µL, as the elusion step is critical to concentrate the DNA for an increased yield. Once extractions were completed, Nanodrop 2000 was used to quantify the DNA concentrations of each sample.

Normalization

Normalization is a technique used in the NGS pipeline to quantify the concentration of each DNA sample and reduce the concentration to a working stock value of 20ng/µL. All extracted DNA samples were made into a working stock solution by diluting them to 100-150ng/µL using the elution buffer, Buffer AE, from the QIAGEN Kit. Picogreen analysis was performed by plating a volume of 4µL of DNA from each sample and a volume of 196µL 1XTE (composed of 10Mm Tris-HCl Tris pH buffer and EDTA [Ethylenediaminetetraacetic acid]) into a 96-well plate. Standards were prepared in six microcentrifuge tubes, the first containing 343µL of 1XTE and 7µL of 2ng/µL DNA. A serial dilution was performed for the remaining 5 micro centrifuge tubes (each containing a volume of 375µL 1XTE) by pipetting 125µL from the former microcentrifuge tube to the latter microcentrifuge tube, proper mixing was ensured. Therefore,

the standards were prepared using a ratio of 1:4. A master mix of Quant-IT Picogreen dsDNA reagent was prepared by diluting the Picogreen reagent 1:199 with 1XTE. A 96-well Costar plate was aliquoted by pipetting 100µL of diluted sample, standards, and blanks (1XTE) into their pre-assigned wells. Furthermore, 100µL of 200x diluted Quant-IT Picogreen dsDNA reagent were placed into wells containing either sample, standards, or blanks. To finish the plate, wells that were not occupied by sample, standards, or blanks were filled with 200µL autoclaved Nanopore water, to ensure the empty wells do not reflect during analysis. Therefore, each well on the 96-well Costar plate contained a volume of 200µL. After pipetting was complete, the plate was set to incubate at room temperature, covered by a dark plate cover, in a dark cupboard, for a period of 5-minutes, prior to Picogreen analysis. Analysis was performed using the Gen5 program v. 2.09. Once concentrations of the dsDNA were obtained, samples were diluted to 20ng/µL with the elution buffer (10mM Tris-HCl; pH = 8) using the Eppendorf epMotion 5075 automated robot. Once completed, samples were placed in a -20^{o}C freezer.

Digestion & Ligation

Once each sample was diluted to the appropriate concentration (20ng/µL), digestion and ligation was performed on all samples. The digestion protocol consists of two restriction enzymes, PstI and MspI, that are used to cut the template DNA at specific sites. Once the template is cut, barcodes (adaptors 1 & 2) are then ligated to these newly cut regions. Simply put, adaptor 1 ligates to PstI cut sites, while adaptor 2 ligates to the MspI overhangs. The digestion procedure is performed by first creating a master mix that is composed of 1X NEB Buffer 4, 0.01M PstI, 0.01M MspI, 0.33M of normalized DNA, and autoclaved Nanopure water into a 1.5mL micro centrifuge tube. A volume of 20µL of master mix and 10µL of normalized DNA was added into an Axygen 96-well plate, creating a final volume of 30µL in each well of the

plate. The plate was sealed with caps and then inverted multiple times for thorough mixing. The plate was then centrifuged at 1000RPM for 1-minute to ensure the entire volume in each well was displaced at the bottom. Using an Agilent thermocycler, the plate was incubated for two hours at 37^{o}C for digestion to commence, followed by a period of 20 minutes at 80^{o}C to inactivate the enzymes in each sample.

Immediately following the digestion procedure, ligation occurred. Ligation attaches unique adapters (Y adapters) onto each sample for later identification. The procedure was performed by first creating a master mix containing 3.3X T4 DNA Ligase Rxn Buffer, 0.1M T4 DNA Ligase, and autoclaved Nanopore water into a 1.5mL microcentrifuge tube. A volume of 15µL from the established master mix was added to the 96-well Axygen plate which underwent digestion, increasing the volume in each well from 30 to 45 µL. Finally, a volume of 5µL from a pre-prepared plate of unique barcodes and Y adapters, was pipetted into the digestion plate with the master mix mixture and DNA. This increased the final volume of the plate to 50µL. Finally, once all wells containing the mixture of barcodes, adaptors, DNA, and master mix were added, each well on the plate was sealed and inverted multiple times for thorough mixing. The plate was once again centrifuged to 1000RPM for 1-minute and placed back in the thermocycler. The plate was incubated at 22^{o}C for 2 hours, then 65^{o}C for 20 minutes. Ligation products were then stored in the -20^{o}C freezer.

Pooling

The next procedure of the NGS pipeline, pooling, is performed only following the ligation of a unique barcode onto each sample is complete. Pooling is performed by combining all adapter-DNA samples together into multiple libraries. The volume of each sample DNA, within a library, must be equal, thereby eliminating individual bias during sequencing. The

pooling protocol was performed with Eppendorf epMotion, by placing the newly ligated Axygen plate containing the DNA with ligated adaptors. A volume of 6µL was transferred from each well into 8-well Axygen strip tubes. The strip tubes were then thoroughly mixed by vortex and then centrifuged 1000RPM for 15-seconds. Once spun down, a volume of 60µL from each well in the 8-well strip tube was transferred into a single 1.5mL microcentrifuge tube. Finally, the microcentrifuge tube was vortexed and centrifuged to ensure proper mixing. The final product was stored in a -20^{o}C freezer and can be used if a mistake is done during a latter part of the pipeline.

Bead Purification

Bead purification was performed for two reasons; (1) to purify the pooled DNA libraries (removing enzymes, ssDNA, and buffers), and (2) to reduce library volume to 40µL, which is required for size selection. The NucleoMag ® NGS Clean-up and Size Select beads were removed from the fridge to reach room temperature and inverted continuously until homogenous. A volume of 120µL of the bead mixture was pipetted into a 1.5mL microcentrifuge tube for each library. A volume of 120µL of pooled DNA was pipetted into the corresponding microcentrifuge tube with the bead mixture. The mixture of bead contents and pooled DNA was set to incubate at room temperature for a period of 15-minutes. The microcentrifuge tubes were then placed on a DynaMag 2 magnetic separator for a period of 5-minutes, to allow for the pellet containing DNA to be separated from the mixture. Without disturbing the pellet, the supernatant was removed. A volume of 200µL of 80% ethanol was used to rinse the pellet, followed by a short incubation period of 30-seconds at room temperature. The ethanol was then removed by pipette, followed by a second rinse of 200µL of 80% ethanol and incubation period of 30-seconds at room temperature. Once the ethanol was removed by pipette, the tubes were left open for trace ethanol

to be removed followed by a rehydration with 10µL of elution buffer (10mM Tris-HCl; pH = 8) by thoroughly mixing until homogeneous. Next, the mixture was placed into a new 1.5mL microcentrifuge tube at a volume of 40µL and placed back on the magnetic separator for a period of 5-minutes. Once the pellet formed, the remaining liquid, the supernatant, contained the purified DNA library. The supernatant was then transferred into a new 1.5mL microcentrifuge tube and stored in the +4°C fridge.

Size Selection

For size selection, the Sage Science ® Pippin Prep was utilized, and the provided protocol was followed to undergo the procedure. In addition, it is necessary to state that 1.5% DF Marker K was used when prompted with the Cassette folder. The size parameters were set to 500 base pairs. A volume of 30µL of DNA library along with 10µL of the loading solution (marker mix K) was loaded and the protocol was followed from there on. The finalized purified and size selected DNA was then pipetted into a 1.5mL micro centrifuge tube and kept in the +4°C fridge.

Polymerase Chain Reaction (PCR) Amplification

The goal of PCR Amplification is that the size selected fragments of DNA and attached adaptors are amplified with the help of a master mix of reagents. Therefore, the reaction will amplify all fragments with forward and reverse adapters attached. It is essential to use the correct number of cycles during this amplification, in this case 12, as too little will cause too few or will not create sufficient copies of size selected fragmented DNA and their corresponding adapters; while too many can create sequence errors. The protocol for the PCR Amplification consists of a master mix containing autoclaved Nanopore water, 5x Buffer Q5, GC Enhancer, dNTPs, forward and reverse primer Illumina, and Q5 HiFi Polymerase were mixed into a 1.5mL microcentrifuge tube with thorough mixing. A volume of 45µL of the established master mix was then transferred

to an Axygen 8-striptube with a volume of 5µL of the size selected DNA being added afterwards with sufficient mixing. A thermocycler was programmed with the correct conditions consisting of a period of 5-minutes at 72ºC, a period of 30-seconds at 98ºC. Followed by 12 cycles of three different temperature steps; 98ºC step for 10-seconds, a 65ºC step for 30-seconds, a 72ºC step for 30-seconds. Finally, a period of 5-minutes at a temperature of 72ºC. The final products were then stored in the +4ºC fridge.

Polymerase Chain Reaction (PCR) Purification

A second round of purification was performed after PCR Amplification to purify, concentrate, and decrease the final volume of the PCR amplified size selected DNA. The protocol for the procedure is the same as the Bead Purification protocol with a slight modification. The only modification of the Bead Purification protocol is that the DNA is contained in an Axygen 8-strip tube compartment, rather than a microcentrifuge tube. Using a pipette, the DNA from two wells were combined into one 1.5mL microcentrifuge tube, making a volume of 100µL, this is repeated 3 more times for a final product of four 1.5mL microcentrifuge tubes each with 100µL.

Bioanalyzer & Purity Assessment

An Agilent 2100 bioanalyzer was utilized to verify that the appropriate base pairs lengths were amplified. To assess the purity of DNA for each library, Nanodrop was utilized. More specifically, the 260/280 ratio of the Nanodrop results represents the absorbance at 260 nm and 280 nm and a value between 1.8-2.0 is generally accepted as pure. In addition, the 260/230 ratio is a secondary measure for DNA purity and are often higher than then former ratio, yielding results of 2.0-2.2. Once the libraries are confirmed to be pure, they were sequenced at 150 bp with an Illumina HiSeq 400 PE 150 at Génome Québec.

Alignment & Bioinformatics

Once raw reads were obtained from Génome Québec, adaptors were trimmed from the DNA libraries using Cutadapt version 1.13 (Martin 2011), and quality control was assessed by FASTQC v. 0.11.5 (Andrews 2010). Stacks 1.46, and modules, from Éric Normandeau's Stacks workflow scripts were utilized to trimmed, demultiplex, and filter the libraries, available on github (https://github.com/enormandeau/stacks_workflow, downloaded January 2019),
The Stacks program was implemented to process the genome through a series of computational inputs in a stepwise pipeline. The first step of the program is quality control and trimming, which filters the sequences based on their quality score and length. Following quality control, "process_radtags", which demultiplexes the data to obtain one file for each individual, was conducted (Catchen, 2013). A reference genome of the Atlantic cod was then used to align the reads accordingly. (Li, 2010). A reference genome is a nucleic acid sequence for a particular species that has been previously assembled by researchers (Li, 2010). There are many alignment programs available; however, for this project the Burrows-Wheeler aligner was implemented.

Following the alignment, "pstacks" groups the data from each individual into loci and identifies the polymorphic nucleotide sites, this is accomplished by combining all the reads and stacking them one on top of the other into segments which align to the same part of the genome (Catchen, 2013). It is key to note that removing individuals with low sequence reads is essential prior to the "pstacks" step, as they often represent sequencing errors rather than true polymorphisms. The program has the ability to detect the amount of stacks a particular individual at a particular locus contain (i.e. one individual could contain 8 stacks while another could contain 48) (Catchen, 2013).

The next step of the pipeline "cstacks" uses the information generated from "pstacks" to formulate a catalogue of loci from all the individual files. Loci were then grouped together across

each individual to construct a catalogue. For this study, loci containing a sequencing depth of six or more reads in at leant on individual contributed to the catalogue. The next step, “sstacks”, used loci from each individual and compared them to the catalogue to identify allelic states for every locus. In “rxstacks”, a log likelihood model is produced which detailed the statistical confidence of assigned nucleotide bases in each locus of the catalog (Catchen, 2013). The model represents the frequency that a log likelihood value is present within the obtained SNPs, and as the log likelihood values increase, SNP quality, or “confidence” increases as well (Catchen, 2013). Therefore, the model allows one to eliminate loci of certain confidence by establishing a threshold. Once the threshold is implemented, “rxstacks” was performed again to adjust the log likelihood model (Figure 4). In addition, “cstacks”, followed by “sstacks” were also performed a second time to formulate a new catalogue and re-genotype each individual. The final step of Stacks is “populations” which generates a genotype file with all genotypes for each individual in the library. It is essential that a filtering criterion is implemented in this step to ensure only genotypes of good quality are obtained. Loci were only retained if 60% of individuals were genotyped in 3 of the 4 populations. Loci were removed if their inbreeding coefficient (F_{IS}) values were below -0.3 and less than a minor allele frequency of 0.01. Each SNP possess two alleles, each with their own locus, therefore, the minor allele frequency is the frequency of the less common allele (Hohenlohe *et al*., 2011). F_{IS} is a measure of inbreeding and the -0.3-cut-off point was determined to eliminate SNPs with an excess number of heterozygotes, given they may be paralogs (Li & Horvitz, 1953; Hohenlohe *et al*., 2011). Paralogs are copies of the same gene that occur within a single set of chromosomes or on the same chromosome from one individual and are not useful for SNP discovery (Hohenlohe *et al.*, 2011).

To determine genetic variations among Northern cod stock spawning populations, Arlequin v. 3.5.2.2 was utilized (Excoffier and Lischer 2010). Genepop files, outputted by Stacks

were converted to Arlequin input files using PGDSpider v. 2.1.1.0 (Excoffier and Lischer 2010). Global F_{ST} was calculated to determine, and to what degree, if the spawning populations significantly exhibit panmixia. The F_{ST} values range from 0 to 1; 0 indicating that the cod spawning populations are genetically similar, therefore, not panmictic. However, an F_{ST} value of 1 indicates that the cod spawning populations are genetically different, therefore, panmictic.

RESULTS

The three lanes sent to Génome Québec for sequencing, received a Phrap quality score of 36-37, out of a total score of 40, indicating high quality for each lane. Lane 1 had a total of 359,478,641 fragments amplified, while Lane 2 and Lane 3 had 358,458,502 and 357,426,260, respectively. The total amount of nucleotide bases identified were 107,843,592,300, 107,537,550,600, and 107,227,878,000 for Lane 1, 2, and 3, respectively. The size of the sequences, ranged from 10M to 18G, indicating that sequencing for all samples was not uniform. A total of 155 samples were retained (Table 1), and 25 samples were excluded due to low sequence coverage during the Stacks filtering process. The retained samples were from Hawke Channel (*n*=31), Notre-Dame Channel (*n*=12), Bonavista Corridor (*n*=44), Smith Sound (*n*=43), Halibut Channel (*n*=7), and Placentia Bay (*n*=18) (Table 1).

The initial catalogue contained 456,000 loci (Table 2). Filtering through the Stacks pipeline removed loci based on log likelihood (145,005), which only kept log likelihoods of -40 and greater, low F_{IS} values and nonpolymorphic loci (9,726), minor allele frequencies (5,251), loci not present in all populations and a final Arlequin filtering (4,333). The remaining loci, 846, were polymorphic in all of the spawning populations (Table 2). Therefore, filtering removed a total of 455,154 (Table 2). Pairwise F_{ST} values were determined between each spawning

population; however, none of the F_{ST} were significant (P>0.05) (Table 3; Table 4). The only two spawning populations close to significance were Smith Sound and Bonavista Corridor (F_{ST}=0.004; P=0.06306) (Table 3; Table 4).

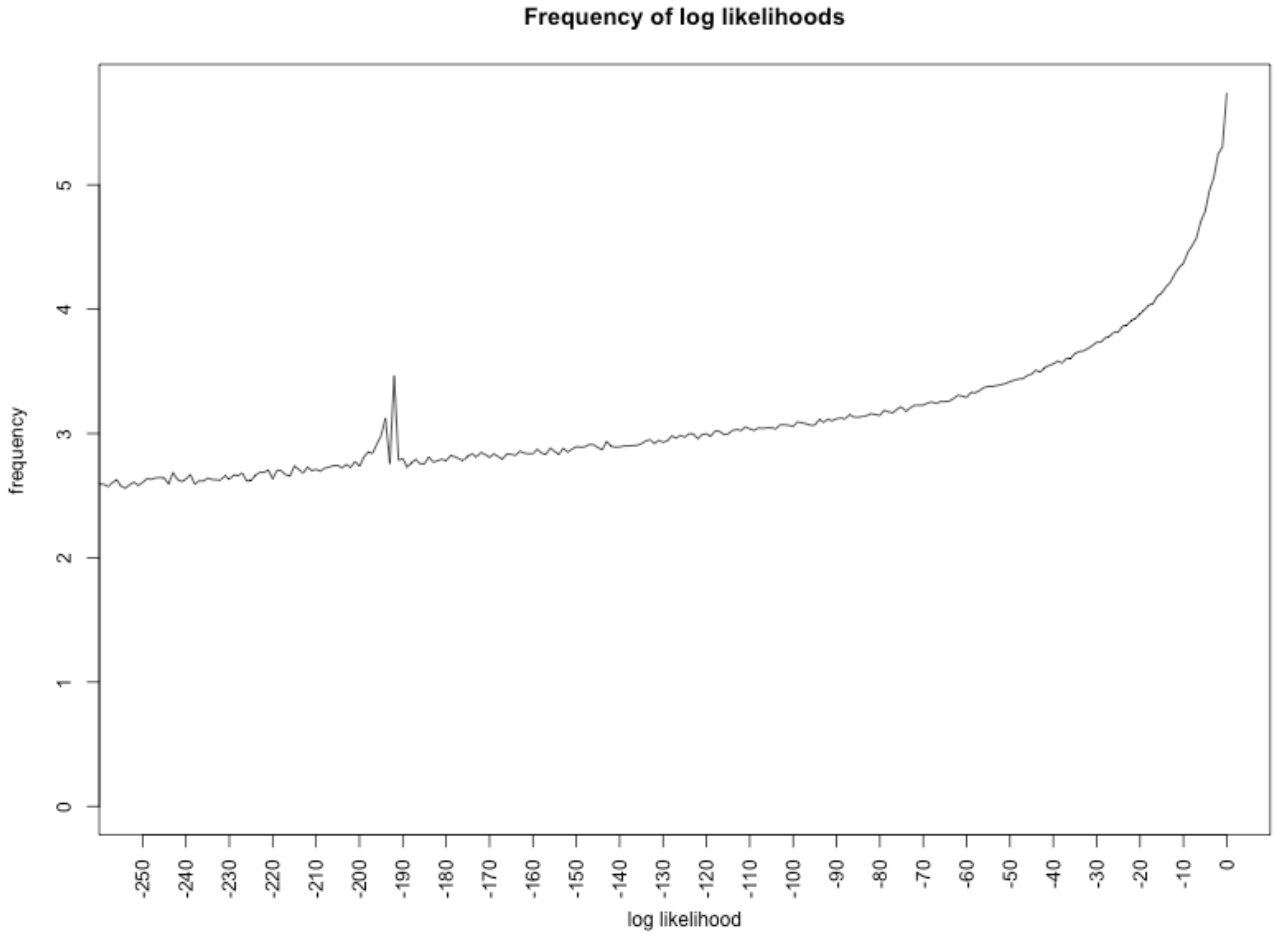

Figure 4: RXStacks log likelihood graph.

Table 1: Location and date of Atlantic cod sampled for genetic analysis and number of samples for retained for analysis.

Location	Date	*n* sampled	*n* analyzed
Hawke Channel	2001	23	21
Hawke Channel	2002	12	10
Notre-Dame Channel	2001	11	11
Notre-Dame Channel	2002	1	1
Bonavista Corridor	2001	17	13
Bonavista Corridor	2002	16	14
Bonavista Corridor	2003	17	17
Smith Sound	2001	19	16
Smith Sound	2002	14	10
Smith Sound	2003	18	17
Halibut Channel	2002	7	7
Placentia Bay	2001	25	18
Total		180	155

Table 2: Number of loci for each step in the Stacks pipeline

Stacks Step	Number of Loci
pstacks	456,000
cstacks	404,503
sstacks	347,666
rxstacks	210,812
rxstacks (with log likelihood threshold)	172,463
cstacks	27,458
sstacks	20,156
populations	846

Table 3: Pairwise F_{ST} among different samples of Atlantic cod from Hawke Channel (HWK), Notre-Dame Channel (NDC), Bonavista Corridor (BVC), Smith Sound (SS), Halibut Channel (HAL), and Placentia Bay (PB).

	HC	NC	BC	SS	HA	PB
HC	0.00000					
NC	-0.04307	0.00000				
BC	0.00205	-0.04288	0.00000			
SS	0.00155	-0.04541	0.00400	0.00000		
HA	0.00204	-0.06620	-0.00747	-0.01310	0.00000	
PB	-0.10538	-0.17867	-0.08879	-0.09638	-0.13798	0.00000

Table 4: Pairwise F_{ST} *P*-values among different samples of Atlantic cod from Hawke Channel (HWK), Notre-Dame Channel (NDC), Bonavista Corridor (BVC), Smith Sound (SS), Halibut Channel (HAL), and Placentia Bay (PB).

	HC	NC	BC	SS	HA	PB
HC	*					
NC	0.99099	*				
BC	0.16216	0.97297	*			
SS	0. 35135	0.97297	0.06306	*		
HA	0.45946	0.98198	0.76577	0.71171	*	
PB	0.99099	0.99099	0.99099	0.99099	0.97297	*

DISCUSSION

Atlantic cod fin clips were obtained from various spawning populations within the Northern cod stock and with the aid of the NGS pipeline, DNA from 180 individuals were isolated and sequenced with an Illumina HiSeq 4000 (Figure 3). The adaptors on the reads, obtained by Génome Québec, were removed and the reads were then trimmed to uniform length and aligned with a reference genome. Samples with low coverage were eliminated from the study, resulting in a total of 155 individuals being analyzed (Figure 3). An initial catalogue count of 456,000 was filtered, using the Stacks program, to 846 loci, which were polymorphic in all of the spawning populations (Table 2). The historic Atlantic cod from the Northern cod stock displayed panmixia among all of the spawning populations. The results from the study display a snapshot of genetic similarities among cod within the stock immediately following the collapse. These findings support the metapopulation hypothesis, suggesting that spawning populations within the stock are not isolated, rather linked through migration (Smedbol & Wroblewski, 2002; Neville, 2018). Therefore, the results match previous studies which depict the stock as a large, inter-connected population, primarily driven by migration (Smedbol & Wroblewski, 2002; Neville, 2018). Alternatively, these findings contrast many studies describing the genetic structure of the Northern cod stock as various isolated, genetically distinct spawning populations with insufficient to no migration (Ruzzante *et al.*, 1999; Beacham *et al*., 2002; Rose *et al.*, 2011; Rose & Rowe, 2015; Benzten *et al.*, 1996) However, the latter papers describe a subtle genetic structure with genetic differences among the spawning populations within the stock and interpret the structure as biologically significant, which it may not.

The factors which contribute to population genetic structure are: migration, genetic drift, selection, and mutation; however, migration and genetic drift play the primary roles, leading to the migration-drift equilibrium (Cockerham & Weir, 1987; Birky *et al.*, 1988; Hössjer *et al.*, 2016). Migration within a population or metapopulation, makes individual alleles more genetically similar (Cockerham & Weir, 1987) (Figure 5). Alternatively, genetic drift within a population and metapopulation, make the population more genetically different by removing alleles and contributing to the effective population size (Cockerham & Weir, 1987; Nosil *et al.*, 2009) (Figure 5). However, at equilibrium, neither of the two components is over-expressed (Figure 5). In context with the present study, it can be determined that migration between the spawning populations is substantial, therefore reducing the effect of genetic drift to make the cod within the stock more genetically similar. Due to the over-expression of migration, it can be inferred that in the present study and the metapopulation hypothesis at large, are not in migration-drift equilibrium (Smedbol & Wroblewski, 2002) (Figure 5). Given that the present study is determined to not be in migration-drift equilibrium, it is in accordance with other studies which describe migration as the primary factor determining the genetic structure of the stock (Smedbol & Wroblewski, 2002); however, more recent research has indicated that it is rather genetic drift that plays the primary role for determining the genetic structure of the Northern cod stock (Therkildsen *et al.*, 2013; Rose *et al.*, 2011; Rose & Rowe, 2015).

$$f_t = [\frac{1}{2N} + (1 - \frac{1}{2N}) f_{t-1}](1-m)^2$$

Figure 5: The Migration-Drift Equilibrium, where m=migration and N=effective population size.

The findings from the study support the overarching metapopulation hypothesis, depicted by Smedbol & Wroblewski (2002), given that the spawning populations within the Northern cod stock are genetically similar (Table 3; Table 4). The hypothesis details two predictions 1) fewer spawning populations would be occupied as the population decline, and 2) given the current moratorium on fishing within the stock, the remaining subpopulations would promote recolonization and accelerated recovery of the overall metapopulation (Smeldbol & Wroblewski, 2002). Given that the stock experienced a significant reduction in population, yet no spawning populations were eliminated, this would suggest a departure from the hypothesis. Furthermore, the findings from this study are premature, given the superficial statistical tests that were conducted and given that the study simply represents a snapshot of samples immediately following the stock collapse and not a longer temporal representation of the stock since its collapse.

The results contrast much of the modern research performed on the Northern cod stock, which recently has found significant physiological, genetic, and behavioural differences among cod within spawning populations of the stock (Templeman, 1979; Templeman, 1981; Lear & Wells, 1982; Ruzzante *et al*., 1996; Rose *et al*., 2011; Beacham *et al*., 2002; Rose & Rowe, 2015). Previous research has found an abundance of support of site-fidelity among Atlantic cod, therefore, suggesting that migration is not the primary factor determine genetic structure, rather genetic drift (Moen *et al.*, 2008; Rose *et al.*, 2011; Therkildsen *et al.*, 2013; Therkildsen *et al.*, 2013). If drift increases, the effect of migration would decrease, thereby eliminating the equilibrium and forcing the cod to be more genetically different (Figure 5). Research conducted on Atlantic cod in the eastern Atlantic, which experienced a small-scale depletion of abundance, has determined that cod following a population collapse become increasingly site-dependent, thereby reducing migration and intensifying the effects of genetic drift (Moen *et al.*, 2008).

Furthermore, the intensified genetic drift led to cod spawning populations along the coast of Norway to be genetically different (Moen *et al.*, 2008). The study by Moen *et al.* (2008) mirrors the stressors of the Northern cod stock, justifying that research is necessary to determine whether cod in the western Atlantic exhibit the same migratory behaviour and intensified effects of drift following a population collapse. Similar studies as Moen *et al.* (2008) have determined that following a reduction in stock population, cod of adjacent spawning populations become genetically different along the coast of Greenland, which is known for extreme population fluctuations, and along the overfished Scotian Shelf (Therkildsen *et al.*, 2013; Therkildsen *et al.*, 2013).

A recent study by Rose *et al.* (2011) depicted that cod within spawning populations, of the Northern cod stock, are genetically different. The study details that cod within Smith Sound and Bonavista Corridor are genetically distinct (F_{ST}=0.0013, *P*=0.027) and cod from Hawke Channel are genetically distinct from cod from Smith Sound (F_{ST}=0.0044, *P*=0.001), Bonavista Corridor (F_{ST}=0.0037, *P*=0.018), and Halibut Channel (F_{ST}=0.0076, *P*=0.009), whereas the present study found these spawning populations to be genetically similar (Rose *et al.*, 2011). The differences in results between Rose *et al.* (2011) and the present study, given that both studies used historical cod from the same regions, can be explained due to the unequal proportion of genotyped individuals in the present study. Thus, the non-uniform sequences yielded individuals with few genotypes, which could have resulted in loci missing from the reduce-genotyped individuals. In addition, the study by Rose *et al.* (2011) was conducted with microsatellites, which are small repetitive DNA motifs within a genome and can be identified to study the species of interest; whereas in the present study SNPs were utilized. Unlike microsatellites, SNP-based studies only have two alleles, thereby reducing the resolution per allele compared to microsatellites, which can have multiple alleles; however, using the SNP-based technique, one

can obtain many more SNPs (Fernández *et al.*, 2013). Therefore, compared to microsatellites, what SNP-based studies lack in power, is made up for in numbers of SNPs obtained, which correlates to better genome resolution (Fernández *et al.*, 2013). Furthermore, additional statistical tests should be conducted to achieve a more robust statistical analysis. For example, Structure files could be exported by Stacks and converted to an Arlequin file using PGDSpider to construct a mean ancestry coefficient should be performed, using LEA (Frichot & François, 2015). Mean ancestry coefficients can be used to estimate the ancestry among all individuals given a fixed number of ancestry populations, this can be used to details the origin of cod within each spawning population and compare the relatedness of a particular spawning population to another (Frichot *et al.*, 2014). Another statistical analysis that can be performed is a discriminant analysis of principle components (DAPC), which indicates the differences among groups (or populations) without prior knowledge of geographic locations (Jombart *et al.*, 2010). These applications would enhance the study by providing information regarding the origin of the individuals within the spawning populations and a visualization on how much the locations differ, without prior knowledge of where the individual was sampled (Jombart *et al.*, 2010; Frichot *et al.*, 2014).

Atlantic cod within the Northern cod stock should continue to be monitored, given their economic and ecological importance in the region. The heated debate regarding the genetic structure of the stock is justification alone that a better consensus is essential for the long-term preservation of the Northern cod stock. It is recommended that the stock be managed as a large interconnected population rather than small isolated populations. Large-scale stock models, similar to the dynamic model Frank & Brickman (2000), would better communicate the scientific research to implement sufficient stock management. Such models must include the genetic structure (e.g., isolated or metapopulation structure) of the stock with corroborating support for its justification (e.g., molecular or behavioural). In the case of a metapopulation genetic

structure, the model proposed by Frank & Brickman (2000) indicates that regions of the stock can be fished during alternating years, thereby preventing the over-exploitation of one region for subsequent years (Smedbol & Stephenson, 2001). Whereas, the model states that an isolated stock genetic structure should have reduced global fishing of the entire stock to ensure that the isolated populations are not depleted (Smedbol & Stephenson, 2001).

The following conceptual recommendations for stock management should be developed using an overarching dynamic model, similar to that of Frank & Brickman (2000). First and foremost, the genetic structure (if known) should be used to identify the populations of interest and to determine whether populations are genetically similar, thereby conceptualizing which populations vary. If it is not known, then it is imperative to obtain a complete spatial and temporal distribution of the population (Smedbol & Stephenson, 2001). Migration corridors should be preserved to eliminate the contraction of free-range throughout the stock, and extensive yearly biomass surveys should be deployed to ensure that extirpation is prevented. Finally, detailed goals, as depicted by Smedbol & Stephenson (2001) would serve as safe-guards for the large-scale recommendations for management units other than fish stocks; 1) preservation of spawning populations, 2) maintenance of spawning biomass, 3) maintenance of age composition within each spawning location, and 4) maintenance of migration between spawning populations.

Identifying the population structure of a stock is well-known to be of primary importance in the development of a healthy, self-sustaining population and stock management programs (Coyle, 1998; Ward, 2006). There are many techniques to identify the population structure of a specific species in a given area, such as morphology and habitat distribution between metapopulation links, but the most efficient approach is using a genetic approach (Coyle, 1998; Begg *et al.*, 1999; Ruzzante *et al.*, 2001; Ward, 2006). By determining the genetic structure of a

given population or stock, it provides a clear link towards the management of species in general (Ward, 2006). Given that the genetic structure identifies which individuals are within a population, it clearly indicates which individuals or areas need protection given their growth and survival rates (Ward, 2006). In context with the Northern cod stock, the identification of the genetic structure would allow management programs to identify which spawning populations, within the stock, experienced varying migration or genetic drift pressures following the population collapse.

A better understanding of the genetic structure for the Northern cod stock is essential for enhanced management of the stock. The findings of this study illustrate the complexity of the stock and depicts it as a large population with interconnected spawning populations. Therefore, the spawning populations within the Northern cod stock were determined to be genetically similar (i.e. panmictic) to other spawning populations within the stock. These results support the metapopulation hypothesis, depicted by Smedbol & Wroblewski (2002); however, additional research is necessary, with extensive loci filtering and statistical tests (e.g., Mean ancestry coefficients, DAPC) to better understand the genetic structure of Atlantic cod within spawning populations of the Northern cod stock. It is recommended that the stock is managed as a large interconnected population rather than small individually isolated populations.

REFERENCES

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ball A. D., Stapley J., Dawson D. A., Birkhead T. R., Burke T., & Slate J. (2010). A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (*Taeniopygia guttata*). *BMC Genomics*, *11*, 218-233.

Beacham T. D., Bratty 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*, 650-665.

Begg G. A., Friedland K. D., & Pearce J. B. (1999). Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries Research*, *43*, 1-8.

Behjati S., & Tarpey, P. (2013). What is next generation sequencing? Archives of Disease in Childhood. *Education and Practice Edition*, *98*, 236-238.

Birky C. W., Fuerst P., & Maruyama T. (1989). Organelle gene diversity under migration, mutation, and drift: Equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, *121*, 613-627.

Brander K. M. (1994). Patterns of distribution, spawning, and growth in North Atlantic cod: the utility of inter-regional comparison. *ICES Journal of Marine Science Symposia*, *198*, 406-413.

Bush W., Moore J. (2012). Chapter 11: Genome Wide Association Studies. *Computational Biology*, *8*, 1-11.

Catchen J., Hohenlohe, P. A., Bassham, S., Amores, S., Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124-31240.

Cockerham C. C., & Weir B. S. (1987). Correlations, descent measures: Drift with migration and mutation. *Evolution*, *84*, 8512-8514.

Coyle T. (1998). Stock identification and fisheries management: the importance of using several methods in a stock identification study. *Australian Society for Fish Biology*, Sydney pp.173-182.

Davey J. W., & Blaxter, M. (2011). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, *9*, 416-423.

Drinkwater K. F. (1996). Atmospheric and oceanic variability in the Northwest Atlantic during the 1980s and early 1990s. *Journal of Northwest Atlantic Fishery Science*, *18*, 77-97.

Excoffier L., & Lischer H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Fernández M. E., Goszcynski D. E., Lirón J. P., Villegas-Castagnasso E. E., Carino M. H., Ripoli M. V., Rogberg-Muñoz A., Rosik D. M., Peral-García P., & Giovambattista G. (2013). Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetic and Molecular Biology*, *36*, 185-191.

Frank K. T., & Brickman B. (2000). Allee effects and compensatory population dynamics within a stock complex. *Canadian Journal of Fisheries and Sciences*, *57*, 513-517.

Frichot E., Mathieu F., Trouillon T., Bouchard G., & François O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*, 973-983.

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*, 1077-1085.

Hanski J., & Gilpin M. E. (1997). Metapopulation biology: ecology, genetics and evolution. Academic Press, Edinburgh, Scotland.

Hanski I., & Simberloff D. (1997). The metapopulation approach, its history, conceptual domain, and application to conservation. In: Hanski, I., Gilpin, M.E. (Eds.), Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, Edinburgh. 5-26.

Hohenlohe P. A., Amish S. J., Catchen J. M., Allendorf F. W., & Luikart G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, *1*, 117-122.

Hössjer O., Laikre L., & Ryman N. (2016). Effective sizes and time to migration-drift equilibrium in geographically subdivided populations. *Theoretical Population Biology*, *112*, 139-156.

Lear W. H., & Wells R. (1982). Vertebral Averages of Juvenile Cod, *Gadus morhua,* from Coastal Waters of Eastern Newfoundland and Labrador as Indicators of Stock Origin. *Journal of Northwest Atlantic Fishery Science*, *5*, 23-31.

Lear W. H. (1998). History of Fisheries in the Northwest Atlantic: The 500-Year Perspective. *Journal of Northwest Atlantic Fishery Science*, *23*, 41-73.

Jombart T., Devillard S., & Balloux F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94-109.

Li C. C., & Horvitz D. G. (1953). Some methods for estimating inbreeding coefficient. *American Journal of Human Genetics*, *5*, 107-117.

Li H. & Durbin R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*, 589-595.

Lilly G. A. (1994). Predation by Atlantic cod on capelin on the southern Labrador and Northeast Newfoundland shelves during a period of changing spatial distributions. *ICES Journal of Marine Science Symposia*, *198*, 600-611.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* [online serial], *17*, 10–12.

Mason F. (2002). The Newfoundland Cod Stock Collapse: A Review and Analysis of Social Factors. *Electronic Green Journal*, *1*, 17-38.

Moen T., Haynes B., Nilsen F., Delghandi M., Fjalestand K. T., Fevolden S-E., Berg P. R., & Lien S. (2008). Identification and characterization of novel SNP markers in Atlantic cod: Evidence for directional selecton. *BMC Genetics*, *9*, 18-27.

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

Neville V., Rose G., Rowe S., Jamieson R., & Piercey G. (2018). Otolith chemistry and redistribution of Northern cod: evidence of Smith Sound – Bonavista Corridor connectivity. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*, 2302-2312.

Nosil P., Funk D. J., & Ortiz-Barrientos D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375-402.

Peterson B. K., Weber J. N., Kay E. H., Fisher H. S., & Hoekstra H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, *7*, 1-11.

Petrie B., & Drinkwater K. (1993). Temperature and Salinity Variability on the Scotian Shelf and in the Gulf of Maine 1945-1990. *Journal of Geophysical Research*, *98*, 20,079-20,089.

Petrie B., Loder J. W., Lazier J., & Akenhead S. (2010). Temperature and salinity variability on the eastern Newfoundland shelf: The residual field. *Atmosphere-Ocean*, *30*, 120-139.

Pujolar J. M., Jacobsen M. W., Als T. D., Frydenberg J., Munch K., Jónsson B., Jian J. B., Cheng L., Maes G. E., Bernatchez L., & Hansen M. M. (2014). Genome-wide single-generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, *23*, 2514-2528.

Rochet M. J. (1998). Short-term effects of fishing on life history traits of fishes. *ICES Journal of Marine Science Symposia*, *55*, 371–391.

Rose G. A., Atkinson B. A., Baird J., Bishop C. A., & Kulka D. W. (1994). Changes in distribution of Atlantic cod and thermal variations in Newfoundland waters, 1980-1992. *ICES Journal of Marine Science Symposia*, *198*, 542-552.

Rose G. A., & O'Driscoll R. L. (2002). Capelin are good for cod: can the northern stock rebuild without them? *ICES Journal of Marine Science*, *59*, 1018-1026.

Rose G. A., (2004). Reconciling overfishing and climate change with stock dynamics of Atlantic cod (*Gadus morhua*) over 500 years. *Canadian Journal of Fisheries and Aquatic Sciences*, *61*, 1553-1557.

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

Ruzzante D. E., Taggart C. T., Cook D., & Goddard S. (1996). Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: microsatellite DNA variation and antifreeze level. *Canadian Journal of Fisheries and Aquatic Sciences, 53*, 634-645.

Ruzzante D. E., Taggart C. T., Cook D., & Goddard S. V. (1997). Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: a test, and evidence of temporal stability. *Canadian Journal of Fisheries and Aquatic Sciences*, *54*, 2700-2708.

Ruzzante D. E., Wroblewski J. S., Taggart C. T., Smedbol R. K., Cook D., & Goddard, S. V. (2000). Differentiation between Atlantic cod (*Gadus morhua*) from Gilbert Bay, Labrador, Trinity Bay, Newfoundland, and offshore northern cod: evidence for bay-scale population structure. *Journal of Fish Biology*, *56*, 431-447.

Ruzzante D. E., Taggart C. T., Doyle R. W., & Cook D. (2001). Stability in the historical pattern of genetic structure of Newfoundland cod (*Gadus morhua*) despite the catastrophic decline in population size from 1969 to 1994. *Conservation Genetics*, *2*, 257-269.

Sherwood G. D., Rideout R. M., Fudge S. B., & Rose G. A. (2007). Influence of diet on growth, condition and reproductive capacity in Newfoundland and Labrador cod (*Gadus morhua*): Insights from stable carbon isotopes ($\boldsymbol{\delta}^{13}$C). *Deep. Sea. Research*, *54*, 2794-2809.

Smedbol R. K., & Stephenson R. (2001). The importance of managing within-species diversity in cod and herring fisheries of the north-western Atlantic. *Journal of Fish Biology*, *59*, 109-128.

Smedbol R. K., & Wroblewski J. S. (2002). Metapopulation theory and northern cod population structure: interdependency of subpopulations in recovery of a groundfish population. *Fisheries Research*, *55*, 161-174.

Syvanen A. C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Review Genetics*, *2*, 930-942.

Szymczak S., Holzinger E., Dasgupta A., Malley J. D., Molloy A. M., Mills J. L., Brody L. C., Stambolian D., & Bailey-Wilson J. E. (2016). r2VIM: A new variable selection method for random forests in genome-wide association studies. *BioData Mining*, 9, 1-15.

Templeman W. (1979). Vertebral number in Atlantic cod, *Gadus morhua,* of the Newfoundland and adjacent areas from tagging in 1962-1966. *ICNAF Research Bulletin, 14*, 5-50.

Templeman W. (1981). Vertebral numbers in Atlantic cod, *Gadus morhua,* of the Newfoundland and adjacent areas, 1947-1971, and their use of delineating cod stocks. *Journal of Northwest Atlantic Fishery Science*, *2*, 21-45.

Therkildsen N. O., Hemmer-Hansen J., Als T. D., Swain D. P., Morgan J. M., Trippel E. A., Palumbi S. R., Meldrup D., & Nielsen E. E. (2013). Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. *Molecular Ecology*, *22*, 2424-2440.

Therkildsen N. O., Hemmer-Hansen J., Hedeholm R. B., Wisz M. S., Pampoulie C., Meldrup D., Bonanomi S., Retzel A., Olsen S. M., & Neilsen E. E. (2013). Spatiotemporal SNP analysis reveals pronounces biocomplexity at the northern range margin of Atlantic cod (*Gadus morhua*). *Evolutionary Applications*, *6*, 690-705.

Ward R. D. (2006). The importance of identifying spatial population structure in restocking and stock enhancement programs. *Fisheries Research*, *80*, 9-18.

William E. S. (2005). The Newfoundland fishery: ten years after the moratorium. *Marine Policy*, *29*, 407-420.

William E. S., & Noel R. (2013). The Newfoundland Fishery and Economy Twenty Years after the Northern Cod Moratorium. *Marine Resource Economics, 28*, 397-413.

Appendix 1 Fixation Index Analysis

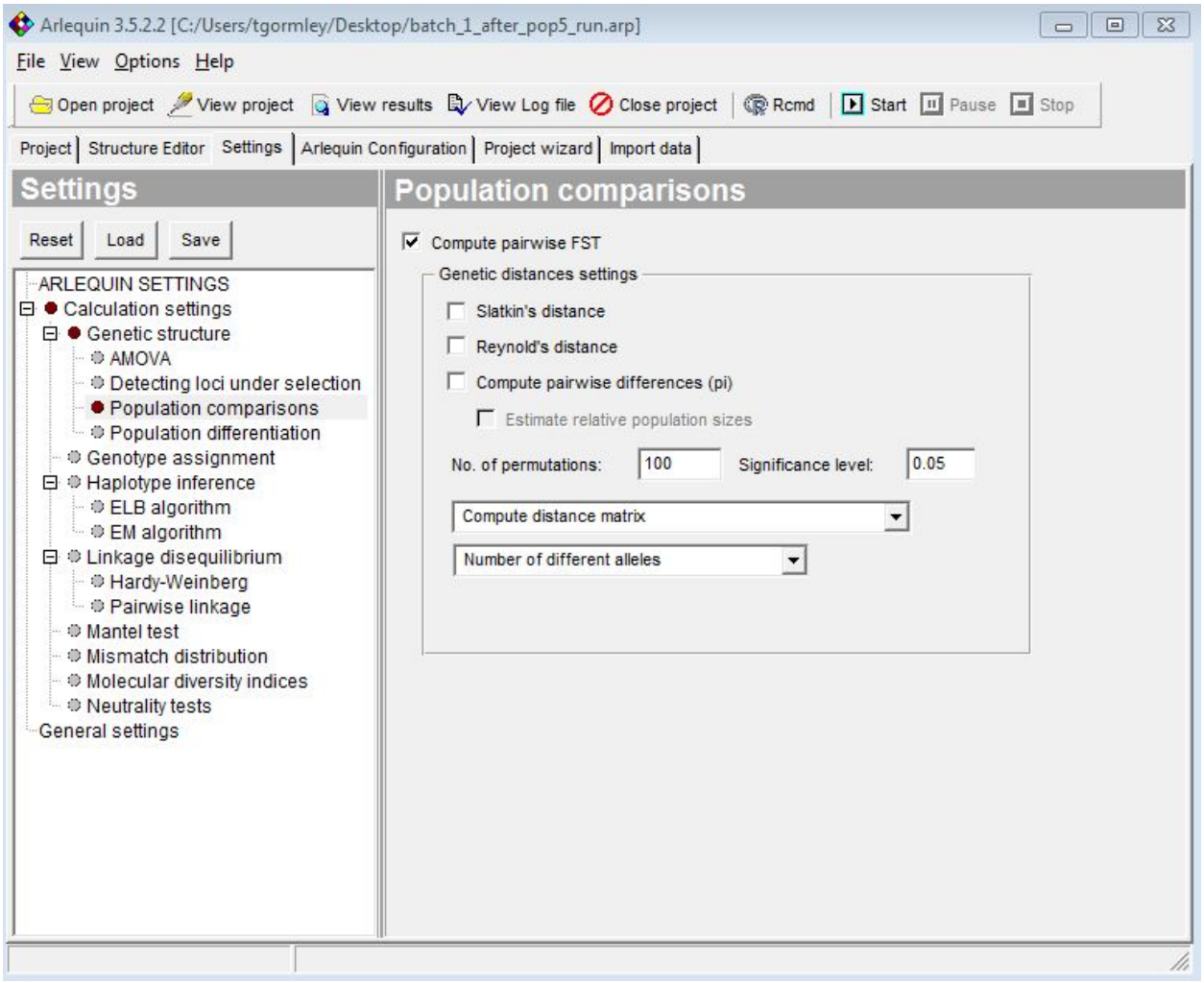

Figure 6: Arlequin v. 3.5.2.2 Fixation Index parameters.