Genetic Population Structure of the Waved Whelk (*Buccinum undatum*) in the western North Atlantic

by

William H. Sturch

Honors Bachelor of Science, University of Toronto, 2020

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

Master of Science

in the Graduate Academic Unit of Biology

Supervisor: Cassidy D'Aloia, Ph.D., Dept. of Biological Sciences, University of New Brunswick and Dept. of Biology, University of Toronto

Examining Board: Kimberley Davies, Ph.D., Dept. of Biological Sciences
Rémy Rochette, Ph.D., Dept. of Biological Sciences
Timothy Rawlings, Ph.D., Dept. of Biology, Cape Breton University

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

April 2022

Abstract

Delineating patterns of spatial genetic structure provides insight into the scale of connectivity and can inform effective management. Here, I characterized spatial genetic structure within the western North Atlantic lineage of the waved whelk (*Buccinum undatum*), a marine gastropod with limited dispersal capabilities. I genotyped 198 individuals from 9 sampling sites throughout Atlantic Canada using 1,052 single nucleotide polymorphisms obtained from double digest restriction-site associated DNA sequencing. *B. undatum* exhibits strong hierarchical genetic structuring throughout this region. There are two major clusters dividing eastern and western sites along with clear genetic substructure within each cluster. At the site level, pairwise genetic differentiation is strong, revealing limited connectivity. The exceptions are sites located on deep shelf habitat, suggesting relatively high gene flow in this environment. This study provides valuable information about the genetic structure of a direct-developing marine invertebrate and has implications for managing the *B. undatum* fishery across Atlantic Canada.

Acknowledgements

I am incredibly grateful for the opportunities and experiences granted to me during this project, and I would like to thank all of the people who helped facilitate this research. First and foremost, I would like to thank my supervisor, Cassidy D'Aloia, for never ceasing support and encouragement. In the same vein, thank you Taylor Naaykens and the entire D'Aloia lab for supporting me during this project, particularly for helping on some early morning field days. I would like to thank my supervisory committee, Kimberley Davies, Rémy Rochette, for their helpful feedback and Timothy Rawlings for serving as my external examiner. Moreover, I would like to thank Marie-Josée Maltais and Rémy Rochette for leading me through whelk physiology and for supporting me with finding study sites. Without the help of many people I would not have been able to collect samples for this project: Michael and Beth Brown, Kurt Simmons and Louisbourg Seafoods, Marie-Claude Cormier, Emily Suominen, Jamie Emberley, and Bernard Sainte-Marie. Additionally, I would like to thank Larissa Roehl for assisting me during DNA normalization, and Steven Bogdanowicz for assisting with library preparation, sequencing, and being willing to answer all of my many questions. Lastly, I would like to thank my friends and family for supporting me.

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

bp: Base pair

µl: Microliter

n: Sample size

SD: Standard deviation

DNA: Deoxyribonucleic acid

dsDNA: Double stranded DNA

minDP: Minimum genotype depth

MAS: Minimum number of individuals with a rare allele

F_{IS}: Inbreeding coefficient

F_{ST}: Fixation index

H_E: Expected heterozygosity

H_O: Observed heterozygosity

PCA: Principal component analysis

SNP: Single nucleotide polymorphism

ddRAD: Double digest restriction-site associated DNA sequencing

IBD: Isolation-by-distance

ESU: Evolutionarily significant unit

MU: Management unit

Genetic Population Structure of the Waved Whelk (*Buccinum undatum*) in the western North Atlantic

1. Introduction

A major goal of population genetics is to identify spatial genetic structure, defined as the non-random distribution of alleles displayed by a species over its range. Unravelling these patterns is a first step to formulating hypotheses about their underlying evolutionary mechanisms (Selkoe et al., 2016; Metivier et al., 2017), inferring contemporary rates of connectivity among metapopulations (Hellberg et al., 2002; Xuereb et al., 2018), and informing conservation strategies (Palumbi, 2003; Balbar & Metaxas, 2019). One of the strongest drivers of spatial genetic structure is natal dispersal, (i.e., the movement of an individual from its birth site to its breeding site (Nathan et al., 2008; Burgess et al., 2016)), as successful dispersal facilitates the exchange of alleles among populations (Bohonak, 1999; Kritzer & Sale, 2006).

In the marine environment, a species' dispersive capacity is expected to be strongly influenced by its life cycle (Pechenik, 1999; Burgess et al., 2016). Most benthic marine species experience a biphasic life cycle characterized by a pelagic larval developmental phase and a relatively sedentary adult stage (Todd 1998; Carrier et al., 2018). However, some 'aplanktonic' species never experience a pelagic phase, either developing directly without larvae or with larvae contained within demersal egg capsules laid on the seafloor (Carrier et al., 2018). Given the preponderance of biphasic marine species, the pelagic larval stage has traditionally been viewed as the primary dispersive period in the ocean. Thus, increasing larval durations are predicted to lead to greater

dispersal distances and decreased genetic structure between populations, due to the homogenizing effect of gene flow (Strathmann, 1986; Caley et al., 1996; Siegel et al., 2003). However, despite extensive research, the relationship between larval duration and dispersal capacity remains unclear. For example, some biophysical models predict a clear positive relationship between planktonic larval duration and dispersal capacity (Siegel et al., 2003), while meta-analyses have found weaker (Bradbury et al., 2008; Selkoe & Toonen, 2011) or no support (Weersing & Toonen, 2009; Riginos et al., 2011) for a relationship between planktonic larval duration and genetic estimates of dispersal. Instead, interactions between egg type, larval duration, seascape features, habitat continuity, and larval behavior have been hypothesized to be important drivers of realized (i.e., successful) dispersal (Todd, 1998; Nanninga & Berumen, 2014; Liggins et al., 2016).

Although there is some uncertainty regarding the relative influence of pelagic larval duration on connectivity, aplanktonic marine species are generally predicted to exhibit strong spatial genetic structure over small spatial scales. Indeed, various marine fishes (Riginos et al., 2011), echinoderms (Puritz et al., 2017), and gastropods (Hoskin, 1997; Cahill & Viard, 2014) that lack a pelagic larval phase exhibit strong genetic structure, particularly when compared to species with pelagic larval stages (Teske et al., 2007; Weersing & Toonen, 2009; Selkoe & Toonen, 2011). However, other marine species lacking a larval phase exhibit more complicated genetic structure (Kyle & Boulding, 2000; Nikula et al., 2010; Sotka, 2012). For example, Ayre et al., (2009) demonstrated that *within* biogeographic regions along the southeast Australian coast, planktonic developers displayed weaker genetic structure than aplanktonic developers. However, they also found that planktonic developers experienced more restricted gene

flow *across* the biogeographic barrier dividing these regions than two aplanktonic developers. Pearman et al. (2020) similarly found evidence for recent dispersal events across known biogeographic barriers and high amounts gene flow for a direct developing marine isopod. Collectively, these studies indicate that factors such as habitat specificity, seascape features, or alternative mechanisms of dispersal can strongly influence the genetic structure of some aplanktonic marine species.

Increasing evidence suggests alternative dispersal mechanisms may allow for greater gene flow than was previously predicted for marine species lacking pelagic larvae. The most commonly invoked mechanism is rafting, in which ocean currents carry marine organisms attached to floating substrata (Winston, 2012). If an individual ultimately settles into a new habitat post-rafting, successful (long-distance) dispersal occurs. Most marine phyla have been recorded on floating substrata (Thiel & Gustrow, 2005a,b; Thiel & Haye, 2006), and rare gene flow events from rafting can reduce the genetic structure exhibited by a species. This is most plausible for species with strong associations with floating substrata, such as subantarctic kelp-dwelling crustaceans where rafting is supported as the primary mechanism of dispersal (Nikula et al., 2010). However, data on the spatial extent and frequency of rafting events remains limited. Alternatively, a hypothesized mechanism for gene flow in aplanktonic species is intergenerational stepping-stone dispersal, in which organisms extend their distribution incrementally across suitable habitat (*sensu* "Hopping" and "Creeping", Winston, 2012). More generally, meta-analyses of the drivers of marine spatial genetic structure include only a small number of empirical studies which focus on species lacking pelagic larvae (Kelly & Palumbi, 2010; Selkoe & Toonen, 2011; Nanninga & Manica, 2018). To improve our

understanding of the drivers of spatial genetic structure in species lacking pelagic larvae, the number of empirical studies which investigate their genetic structure should be increased.

A species well suited for exploring the population genetic consequences of limited dispersal potential is *Buccinum undatum* (Linnaeus, 1758). This species is a neogastropod that develops from a veliger larva within benthic egg masses directly into a benthic carnivorous juvenile (Martel et al., 1986a; Martel, et al, 1986b). As such, it experiences an entirely benthic life cycle and has limited movement during its adult stage. Individuals have been observed to remain within an 8-km radius over a 3-year timespan, which is a small distance relative to the distances over which the species occurs (100s to 1000s of km; Himmelman, 1988). While the species' total lifespan is unknown, caught individuals are typically estimated to be under 10 years old (Hollyman et al., 2018).

B. undatum is widely distributed throughout the high latitude North Atlantic Ocean and, while primarily a subtidal species, can occupy a broad depth distribution (0 to greater than 400 meters; Magnúsdóttir et al., 2019), though abundance has been found to be highest from 40 to 70 meters (Borsetti et al., 2018). It can be found less commonly in the intertidal zone where it tends to occupy tide pools. In the North Atlantic Ocean, the species occurs as two monophyletic lineages that diverged in allopatry during the Pleistocene (Magnúsdóttir et al., 2019): a western Atlantic lineage (New Jersey to Labrador) and an eastern Atlantic lineage (Celtic Sea to the Baltic Sea). Recurring glaciation events during the Pleistocene Epoch are known to have shaped marine invertebrate phylogeography throughout the North Atlantic (Wares & Cunningham, 2001). Present-day species distributions are generally believed to reflect post-glacial expansion from either southern

populations, eastern Atlantic refugia, or isolated northern refugia. While the historical demographic patterns of *B. undatum* are not fully resolved, genetic data indicate that the two lineages diverged approximately 2.1 million years ago (Magnúsdóttir et al., 2019).

Within both monophyletic lineages, *B. undatum* exhibits high phenotypic variability. Significant differences in shell morphology, growth rates, and size at sexual maturity occur across short (~10 km) geographic distances (western North Atlantic: Kenchington & Glass, 1998; Borsetti et al., 2018; Ashfaq et al., 2019; eastern North Atlantic: Haig et al., 2015; McIntyre et al., 2015; Magnúsdóttir et al., 2018). Likewise, *B. undatum* exhibits morphological variability across depth zones: individuals in deeper locations generally have thinner, rounder shell spires and display increased colour diversity (Emmerson et al., 2018; Magnúsdóttir et al., 2018) While phenotypic variation may reflect plastic responses to local predators (Rochette & Himmelman, 1996; Rochette et al., 1999), it can also indicate fixed genetic differences between populations and has previously motivated studies of genetic connectivity (Goodall et al., 2021).

Several population genetic studies have been conducted on the eastern North Atlantic *B. undatum* lineage, where the species is the target of a valuable fishery (Fahy, 2001, 2008; de Vooys & van der Meer, 2010). Consistently, *B. undatum* populations separated by hundreds of kilometers or known oceanographic barriers have exhibited strong genetic structure, concordant with expectations of limited dispersal capacity (Ireland global F_{ST} = 0.019, Mariani et al., 2012; England global F_{ST} = 0.014, Weetman et al., 2006; Iceland global F_{ST} = 0.029, Goodall et al., 2021). Comparatively, over shorter spatial scales (~40-100 km) subtler genetic structure has been observed. These findings are consistent across both traditional microsatellite marker sets (e.g. Weetman etal., 2006;

Pálsson et al., 2014) and recent single nucleotide polymorphism (SNP) data sets (Morrissey, 2019; Goodall et al., 2021). Lower genetic structure between populations separated over shorter spatial scales has been attributed to either rare rafting events or stepping-stone connectivity across continuous habitat (Weetman et al., 2006; Mariani et al., 2012). This previous population genetic research, focused exclusively on the eastern North Atlantic *B. undatum* lineage, creates a baseline for comparison with the relatively understudied western North Atlantic *B. undatum* lineage.

Fishery interest in the western North Atlantic lineage of *B. undatum* has increased in recent years, and the species is recognized as an emerging commercial resource within Atlantic Canada (Kenchingon & Glass, 1998; Brulotte, 2019) and the Northeastern United States (Borsetti et al., 2018). In 2020, 2,323 metric tonnes of *B. undatum* individuals were landed in Québec and Newfoundland (DFO, 2020). While considered a Major Fish Stock in parts of Atlantic Canada, the species is considered data poor and thus the stock status is listed as "uncertain". The paucity of population genetic data within the western North Atlantic lineage represents a knowledge gap for understanding the species' genetic populations and patterns of connectivity between them. Data on spatial genetic structure are broadly considered valuable for effective species management (Palumbi, 2003; Funk et al., 2012; Crawford & Oleksiak, 2016). Using genetic data to delineate populations can inform management plans to prevent overexploitation and loss of genetic diversity (Kenchingon & Glass, 1998; Hohenlohe et al., 2021; Kardos et al., 2021). Such knowledge may be especially important for aplanktonic species such as *B. undatum* because connectivity is expected to be restricted across populations (Emmerson et al., 2018). Thus, documenting its spatial genetic patterns is valuable for informing fisheries management.

In this study, my primary goal is to describe the patterns of population genetic structure of *B. undatum* within the western North Atlantic. Specifically, I identify the major genetic groupings and infer patterns of genetic connectivity throughout the western North Atlantic, from the Gulf of Maine to the coast of Newfoundland.

2. Methods

2.1 Sample collection

B. undatum were collected between July 2020 and March 2021 from nine sites within the western North Atlantic (Figure 1a). Given travel and field work restrictions related to the COVID-19 pandemic, samples were obtained in several ways (Table A1). Individuals were hand collected from the intertidal zone during low spring tides at Green's Point, New Brunswick and Pubnico, Nova Scotia. Samples from other sites were collected at depth with traps by commercial fisheries (30 meters depth: Gannet Lighthouse, New Brunswick; Magdalen Islands, Québec; and Newfoundland; 60 meters depth: Scotian Shelf A and Scotian Shelf B, Nova Scotia) and with trawls by Fisheries and Oceans Canada (30 meters depth: Georges Bank A and Georges Bank B) (Table 1). Individuals were also collected from Sept-Îles, Québec but high-quality DNA could not be extracted, and sequencing failed, so this site was removed prior to genotype assembly. At each site I gathered between 17 and 54 individuals (Table 1; mean = 31 individuals ± 12.41 SD). Whole specimens were kept on ice upon collection then stored in a -80°C freezer until dissections took place. For all individuals, foot tissue was preserved in 95% ethanol at -20°C. Samples were transferred into fresh 95% ethanol after an initial ~24 hours.

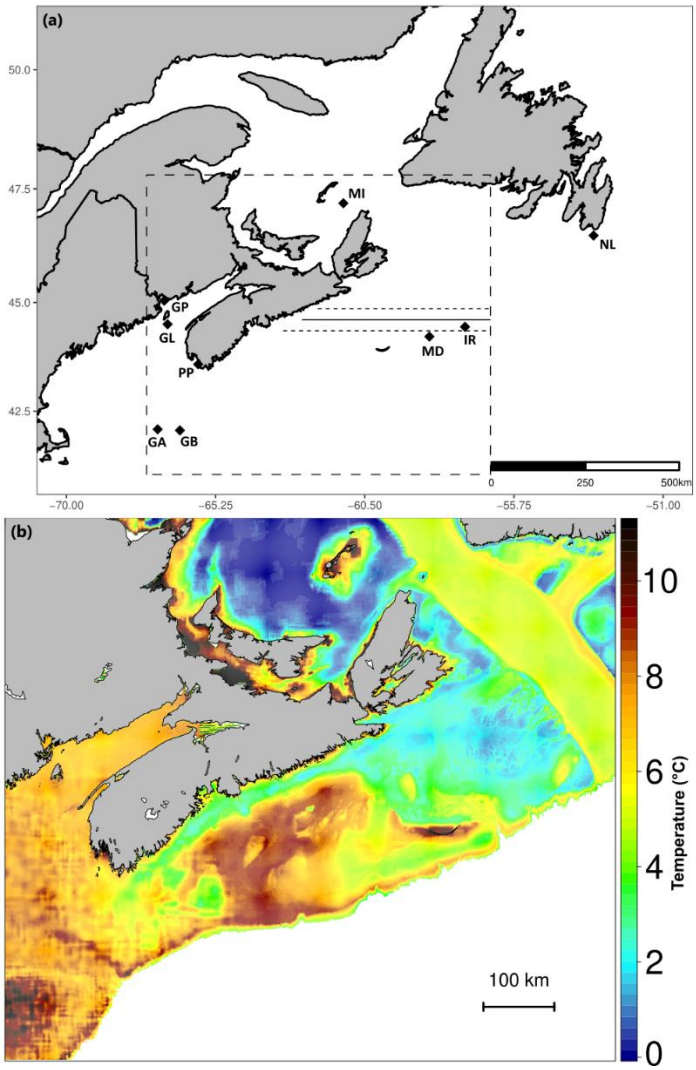

Figure 1 Map of the western North Atlantic covering the sampling region. (a) Spatial distribution of sampling locations indicated by abbreviations: GP, Green's Point; PP, Pubnico Point; GA, Georges Bank A; GB, Georges Bank B; GL, Gannet Lighthouse; NL, Newfoundland; IR, Scotian Shelf A; MD, Scotian Shelf B; MI, Magdalen Islands. Solid and dashed horizontal lines denote the location of the multi-species genetic cline (mean ± SD) identified by Stanley et al. (2018). Dashed box denotes the boundaries of Figure 1b. (b) Interpolated 2016 mean summer bottom temperature (°C) derived from Hubley et al. (2018).

Table 1 For each sampling site: sample geographic coordinates, collection date, depths, and individuals collected (N_C).

Location	Site ID	Latitude	Longitude	Date	Depth (m)	N_C
Gannet Lighthouse	GL	44.509	-66.779	03/2021	30	18
Green's Point	GP	45.04	-66.889	08/2020	0	34
Pubnico Point	PP	43.593	-65.806	09/2020	0	22
Scotian Shelf A	IR	44.389	-57.311	08/2020	63	43
Scotian Shelf B	MD	44.228	-58.444	08/2020	59	34
Magdalen Islands†	MI	47.191	-61.184	09/2020	30	54
Newfoundland†	NL	46.488	-53.217	09/2020	30	40
Georges Bank A	GA	42.074	-67.097	03/2021	34	17
Georges Bank B	GB	42.05	-66.388	03/2021	42	18

† For MI and NL, only the fishing region where collection took place was available. In lieu of an exact location, the location with the highest recorded *B. undatum* fishery effort within the region, and the midpoint of the fishing region, were used for MI and NL, respectfully. For an explanation of the coordinates associated with each site see Table A1.

2.2 DNA extraction, ddRAD library preparation, and sequencing

I extracted DNA from 17 to 35 individuals per site (mean = 28 ± 7.98 SD). For sites where I collected more than 35 individuals, n = 35 individuals were chosen randomly. For sites where I collected fewer than 35 individuals, all individuals were used. Tissue was air-dried briefly then DNA was extracted using the DNEasy Blood and Tissue kit (Qiagen). The manufacturer's protocol was followed with two modifications: first, I lysed tissue at 56°C overnight; second, I eluted samples in 60 µl AE buffer. Following extraction, DNA

quality was checked on 0.8% agarose gels. DNA concentration was measured using Quant-iT PicoGreen dsDNA assays with Gen 5 software and a Synergy HTX multi-mode microplate reader (Biotek). All samples were normalized to 10 ng/µl using an epMotion 5075vt automated liquid handling system (Eppendorf).

The double digest restriction-site associated DNA (ddRAD) sequencing library was constructed at the Cornell Evolutionary Genetics Core Facility (Ithaca, NY, USA) following the protocol of Peterson et al. (2012) with several modifications. Whole genomic DNA (150 ng) was digested using (0.2 µl) SbfI (High Fidelity) and (0.2 µl) MspI restriction enzymes, while P1 (Sbfl) and P2 (Mspl) adapters were simultaneously ligated using T4 DNA ligase. Plates were incubated within a thermal cycler at 37°C for 30 min, then at 20°C for one hour. Following digestion and ligation, 1.5 µl of each sample was amplified with 18 unique Illumina Truseq indexes for 35 PCR cycles (95°C for 40 s, 60°C for 45 s, and 68°C for 30 s). I used 16 unique P1 adaptors and 18 indexes to maximize complexity at the 5' end of reads. Then, I pooled samples and removed small fragments using an Ampure XP cleanup (0.7x). Next, I size selected 400 to 600 bp fragments using a Pippin Prep (Sage Science) and assessed the resulting size distribution on an Agilent 2100 Bioanalyzer. Because there was a small peak at 211 bp (likely representing adaptor dimers), I ran a second Ampure XP cleanup (0.8x) that removed these small fragments. Finally, the library was quantified on a Qubit fluorometer, diluted to 3 nanomolar, and submitted to the Cornell University BioResource Center for sequencing on an Illumina NextSeq2000 platform (single-end 230 bp), generating an average of 4,246,481± 3,014,427 SD raw reads per individual.

2.3 Bioinformatics

Raw reads were demultiplexed and quality filtered using the *process_radtags* program in *Stacks* v2.58 (Catchen et al., 2013). The purpose of demultiplexing is to organize raw sequence files using identifiable barcode sequences. I used several default options, including `--inline_null` (to indicate that barcodes were located within the sequence itself), `--barcode_dist_1 = 1` (number of allowed bp mismatches when rescuing barcodes), `w = 0.15` (sliding window size as fraction of read length), and `s = 10` (minimum average read quality score required within the sliding window). I also enabled the following arguments: `-c` (discard read with an uncalled base), `-q` (discard reads with a quality score less than 10), and `-r` (recover reads that had 1 bp mutation in barcode sequence). To ensure equal read lengths after demultiplexing, reads were truncated to 212 bp to account for the size of the largest barcode (12 bp + 6 bp cut site). Additionally, the first restriction enzyme was specified (`-e sbfI`), and the Illumina universal adaptor (`--adapter-1 [AGATCGGAAGAG] --adapter-mm [0]`) was filtered out. This initial demultiplexing resulted in an average of 3,575,960 reads per sample ± 2,748,077 SD (Figure A1).

To choose optimal parameters for a *de novo* alignment in Stacks, I explored parameter values on a subset of samples using the 'r80' method (Rochette & Catchen, 2017). This method tests input values for three key alignment parameters to assess the combination that maximizes the total number of SNPs: `m` controls the number of identical reads required to form a new putative allele (i.e., minimum stack depth), `M` controls the number of bp mismatches allowed between two alleles of an individual before being

considered separate loci (i.e., distance allowed between stacks), and `n` controls the number of bp mismatches allowed between any two alleles of a population for them to be merged together into one locus (i.e., distance allowed between catalog loci).

I ran the full Stacks pipeline to test m = 3 - 4 and M and n = 1 - 8, using a subset of 12 samples collected throughout my geographic sampling area. The number of assembled loci, polymorphic loci, and SNPs were compared across these runs (Figure A2). The number of loci and the largest addition of SNPs (Figure A3) plateaued at a value of m = 3, M = 4, and n = 4, thus these values were chosen for the full de novo assembly.

Each *Stacks* module was run separately. First, *Ustacks* was run to identify loci within each individual (`-m 3 -M 4`). Samples with fewer than 100,000 retained reads or 10× coverage were removed from subsequent steps. Second, *Cstacks* constructed a catalogue of loci and alleles in the population (`-n 4`). For this step, I used 10 samples from each of the 9 sampling locations. To exclude over- or under-represented loci, samples closest to the median value of retained reads were chosen. Third, I ran the following modules with default options enabled: *Sstacks* (putative loci constructed by *Ustacks* were matched against the catalogue created by *Cstacks* for individuals), *tsv2bam* (data were transposed to be oriented by locus instead of by individual), and *gstacks* (SNPs are identified for each locus and each individual is genotyped for each identified SNP). This pipeline yielded 313,711 loci, with an effective per-sample coverage of 118.8×. Then, *Populations* was run to generate a raw *.vcf* SNP file for additional filtering.

Following O'Leary et al. (2018) recommendations, I iteratively filtered SNPs (Table 2) using VCFtools (Danecek et al., 2011). To begin, I required genotypes to have a minimum depth of 5 (`minDP 5`) and then set the mean read depth to 10 (`min-meanDP`

10). Next, I iteratively increased missing data requirements, up to loci being called in ≥ 70% of sampled individuals and individuals having ≤ 30% missing data. To deal with possible multilocus contigs, remaining loci with depth > 400 were removed. Then, using scripts created by Eric Normandeau (https://github.com/enormandeau/stacks_workflow), I set the MAS = 3 (minimum number of samples with a rare allele) and filtered for linkage disequilibrium using a difference threshold of 0.5. Finally, I filtered for strong deviations from Hardy Weinberg Equilibrium by removing SNPs with a Hardy-Weinberg p-value < 0.001 in at least 25% of populations using a Perl script created by Chris Hollenbeck (https://github.com/jpuritz/dDocent/raw/master/scripts/filter_hwe_by_pop.pl). In total, I retained 1,052 SNPS and 198 individuals.

Table 2 Filtering protocol with number of remaining individuals and SNP markers corresponding to each step. Step 13 coincides with the final remaining SNPs and individuals used during population genetic analyses.

Step	Filter	SNPs	Individuals
0	Raw file	137,515	243
1	Min genotype depth (min DP) 5	137,515	243
2	Mean read depth (min-mean DP) 10	26,213	243
3	Remove loci with genotype call rate (max-missing) > 50%	8,552	243
4	Remove individuals with missing data > 90%	8,552	243
5	Remove loci with genotype call rate < 60	5,455	243
6	Remove individuals with missing data > 70%	5,455	238
7	Remove loci with genotype call rate < 70%	2,973	238
8	Remove individuals with missing data > 50%	2,973	224
9	Removed loci with greater than 400 depth.	2,962	224
10	Remove individuals with > 30% missing data	2,962	198
11	MAS value of 3	1,247	198
12	Linkage (difference threshold of 0.5)	1,067	198
13	HWE (p value of 0.001)	1,052	198

To test for the presence of SNPs putatively under selection, I used the *gl.outflank* function from the R package OutFLANK v0.2 (Whitlock & Lotterhos, 2015). This package infers the distribution of F_{ST} for neutral markers using likelihood on a trimmed

distribution of F_{ST} values, then assigns *q*-values to each marker to detect F_{ST} outliers. Whitlock & Lotterhos (2015) have shown this approach has lower false positive rates compared to other outlier detection methods. OutFLANK was run with default options enabled and a q-value threshold of 0.5, which is considered a 5% or lower false discovery rate.

2.4 Population Structure Analyses

I calculated basic genetic summary statistics including observed and expected heterozygosity and inbreeding coefficient (Table 3) within sites using the *basic.stats* function in the R package hierfstat (Goudet, 2005). F_{ST} was used as my primary estimate of genetic structure between sampling sites. I calculated pairwise F_{ST} following Weir & Cockerham (1984) using the *pairwise.WCfst* function from the R package hierfstat (Goudet, 2005) and visualized results using a heat map plotted with the R package ggplot2 (Wickham 2016). I performed a hierarchical cluster analysis on the pairwise F_{ST} matrix using the R package pvclust v.2.2-0 (Suzuki & Shimodaira, 2015) with 10,000 replications to obtain approximately unbiased p-values (*AU*), and visualized the results as a dendrogram. The *AU* values use bootstrap probabilities calculated at several scales (i.e., changing the sample size of data in bootstrap resampling), and represent a probability value from 0 to 100 of a given cluster in the dendrogram being true (Shimodaira, 2002).

To explore genetic clustering among all samples, I conducted principal component analysis (PCA) using the function *dudi.pca* in the package adegenet (Jombart & Ahmed, 2011). PCA is a nonparametric approach that can be used to visualize genetic structure by summarizing the main axes of variation in allele frequencies and then placing individuals

along these axes. Upon finding evidence of strong structuring into two major clusters (see Results), the data were subset into two subgroups (East Cluster, West Cluster) and the PCA was repeated to explore hierarchical structure within each group. Hierarchical analyses of population genetic data are sometimes appropriate because distinct patterns of genetic structure can arise at different spatial scales due to the combined effects of biological, geographic, historical, and oceanographic factors on realized genetic connectivity (Selkoe et al., 2016). To explore the effect of outlier SNPs on genetic structure, the PCA was repeated using the filtered neutral dataset derived from OutFLANK.

To investigate the proportion of individual genotypes assigned to mixed ancestry, admixture was estimated using the *snmf* function in the package LEA v3.6 (Frichot & François, 2015). Assuming K ancestral populations, the *snmf* function provides least-squares estimates of ancestry proportions using sparse non-negative matrix factorization algorithms. This analysis was initially run with 10 repetitions for K = 1 to 9 using all genotyped individuals. Cross-entropy criteria were calculated for each value and K = 2 was determined to be the optimal number of ancestral populations (Figure A4). Based on this, alongside the strong hierarchical structure demonstrated by the PCA (see Results), I subsequently repeated the admixture analysis on the two subset “Eastern” and “Western” clusters to explore substructure. For each cluster I visualized admixture with K = 2 to 3 ancestral populations.

2.5 Population Assignment Test

To test for the presence of recent immigrants among the sampled individuals, I conducted self-assignment tests using the R package assignPOP (Chen et al., 2018). I used k-fold cross-validation (k = 5) to split the data into training and test sets to avoid upward bias in assignment tests (Anderson, 2010) while ensuring that each individual was assigned. Using the function *assign.kfold* I applied a support vector machine model on the training set, which was then used to assign individuals in the test set. By default, assignPOP takes the top assignment probability for each individual to be the inferred source population. I applied a more stringent confidence criterion, requiring individuals to have an assignment probability ≥ 0.9 to be considered a high-confidence assignment.

Based on the strong hierarchical structure in *B. undatum* (see Results), I conducted two sets of assignment tests. I began treating the clearest genetic groupings as the three source populations (Western Cluster, Pubnico Point, Easter Cluster). I also explored whether individuals could be assigned back to their collection site. For this analysis, I grouped two pairs of sites with $F_{ST} = 0$ (see Figure 3 in Results), leading to a total of 7 potential source populations.

For each source population I calculated two metrics (*sensu* Bradbury et al., 2018; Layton et al., 2020). First, self-assignment accuracy: the percentage of individuals assigned to a source population (Western Cluster, Pubnico Point, or Easter Cluster) that were also collected from that population. This value reflects the percentage of residents relative to putative immigrants. Second, efficiency: the percentage of individuals collected from that population that were assigned to that same population. This value reflects the percentage of individuals that cannot be accurately assigned.

3. Results

3.1 SNP Filtering and Genetic Summary Statistics

Of the 248 individuals sequenced, five individuals were initially removed because they had <100,000 total reads. The remaining 243 individuals underwent genotype assembly in Stacks V2.58. Per sample effective coverage was high, ranging from 16× to 172× (mean = 87× ± 35 SD). A total of 198 individuals and 2,962 biallelic SNPs remained after iterative missing data filtering, further reduced to 1,052 SNPs after removing putative multilocus contigs, applying MAS, Hardy-Weinberg, and linkage disequilibrium filters (Table 2).

OutFLANK detected 5 outliers SNPs based on F_{ST} (Figure 2). Therefore, 1,047 SNPs were retained as a putatively "neutral" dataset. I present results using both the full SNP data set and this putatively neutral data set for the PCA analyses below (see Section 3.2).

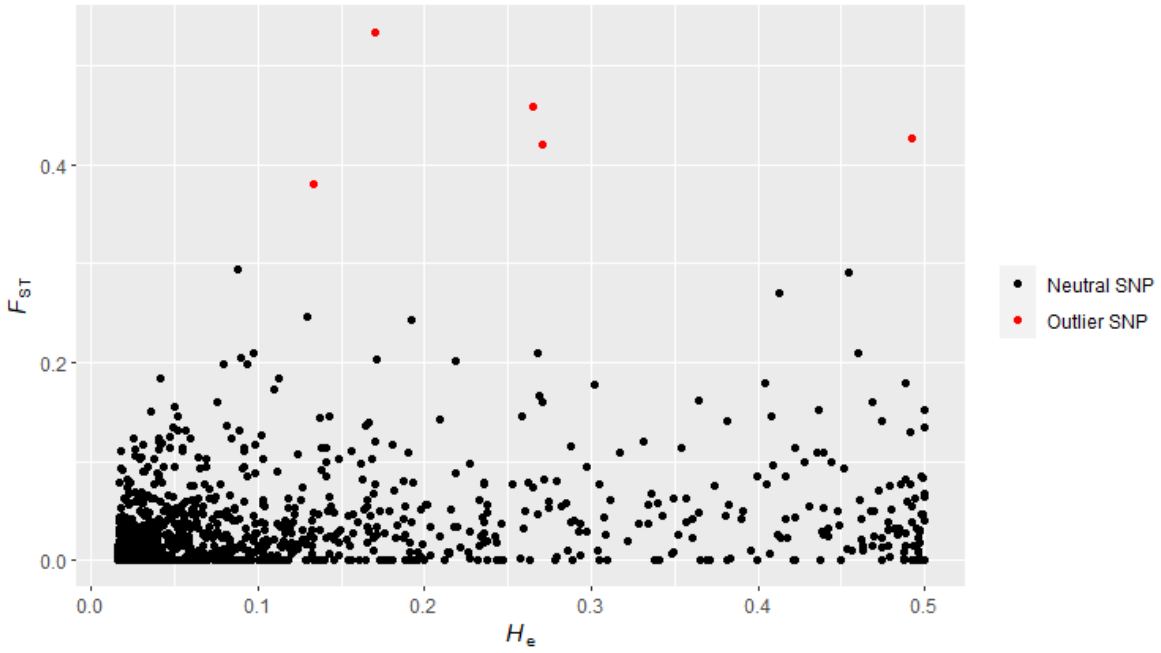

Figure 2 Results of OutFLANK outlier detection test. For each SNP, F_{ST} (fixation index) is plotted over H_E (expected heterozygosity). This package infers the distribution of F_{ST} for neutral markers to identify outliers. Five outlier SNPs (red) were identified (q value < 0.5).

In terms of genetic diversity, mean observed heterozygosity (H_O) and mean expected heterozygosity (H_E) were consistently low across all sites, ranging from 0.099 to 0.111 and 0.118 to 0.133, respectfully (Table 3). The inbreeding coefficient (F_{IS}), which measures the deficiency of heterozygotes among individuals within a site compared to HWE expectations under random mating, was similar across all sites, ranging from 0.075 to 0.15, but was highest at Georges Bank A. The most private alleles were found at intertidal sites (Green's Point = 18, Pubnico Point = 6) and the two most geographically isolated sites (Newfoundland and Magdalen Island 9). All other sites had one or zero private alleles (Table 3).

Table 3 For each sampling site: samples genotyped (N_g), samples retained post-filtering (N_r), mean observed and expected heterozygosity (H_o and H_e), inbreeding coefficient (F_{IS}) and the total number of unique private alleles.

Location	N_g	N_r	H_o	H_e	F_{IS}	Private alleles
Gannet Lighthouse	18	14	0.111	0.123	0.075	2
Green's Point	34	30	0.108	0.131	0.136	18
Pubnico Point	22	21	0.111	0.133	0.118	6
Scotian Shelf A	35	29	0.099	0.118	0.127	1
Scotian Shelf B	34	19	0.099	0.119	0.136	1
Magdalen Islands	35	33	0.103	0.118	0.111	9
Newfoundland	35	29	0.101	0.117	0.118	9
Georges Bank A	17	14	0.101	0.125	0.15	0
Georges Bank B	18	9	0.107	0.126	0.103	0

3.2 F-statistics

Pairwise F_{ST} estimates provided evidence of strong population genetic structure between most sites (Figure 3). Pairwise values ranged between 0 and 0.079 (mean = 0.045 ± 0.02 SD) and global F_{ST} was 0.041. Pairwise F_{ST} tended to increase with geographic distance, and, indeed, the highest value was observed between two of the most geographically distant pairs of sites: George's Bank B (GB) and the Magdalen Islands (MI), which are separated by approximately 710 km. I note, however, that I did not conduct a statistical test of genetic isolation-by-distance because I only had approximate geographic locations

for several sites (see Table A1). Qualitatively, I found that not all pairwise F_{ST} values conformed to this pattern. For example, differentiation observed between Newfoundland and the Magdalen Islands was relatively low ($F_{ST} = 0.016$) despite the sites being separated by >500 km. Several other interesting spatial patterns emerged from the F-statistics. The two pairs of sampling sites located on "deep" shelf habitat (defined here as 34-63 meters) exhibited no pairwise differentiation ($F_{ST} = 0$): the two Georges Bank sites (GA, GB) and, separately, the two Scotian Shelf sites (IR, MD). Finally, Pubnico Point, an intertidal site located on a peninsula along the south-west tip of Nova Scotia (Figure 1a), exhibited high differentiation with all other sites (pairwise $F_{ST} = 0.046 – 0.074$).

A dendrogram based on these pairwise F_{ST} values helped visualize hierarchical patterns of genetic structure (Figure 3). At the broadest scale, two major clusters were supported (approximately unbiased probability "*AU*" = 100): a "Western" cluster comprised of sites on Georges Bank and the Bay of Fundy, and an "Eastern" cluster comprised of sites on the Scotian Shelf and towards the Gulf of Saint Lawrence. A third unique cluster (Pubnico Point) split off from the rest of the Western sites, though this split was not strongly supported (*AU* = 66). Within the Western cluster, there was additional support for a cluster of Georges Bank sites (GA & GB; *AU* = 100) and a cluster of Bay of Fundy sites (GL & GP; *AU* = 95). Within the Eastern cluster, there was additional support for a cluster of Scotian Shelf sites (IR & MD; *AU* = 100) and a cluster of Magdalen Islands and Newfoundland sites (MI & NL; *AU* = 94). Notably, hierarchical cluster analysis functions iteratively: at each stage the two most similar clusters are joined until a single cluster remains (Suzuki & Shimodaira, 2006).

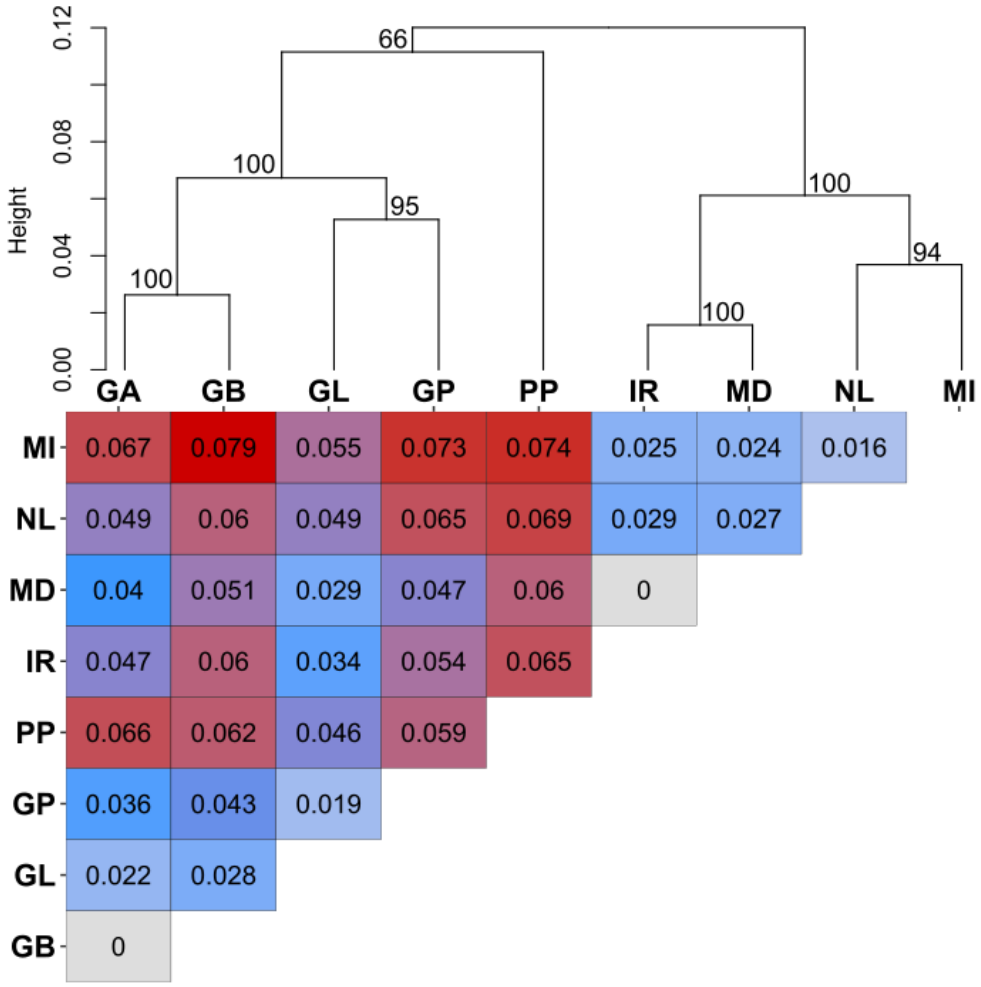

Figure 3 Dendrogram and heat map of pairwise genetic differentiation (F_{ST}) showing a split between Western and Eastern regions, along with strong hierarchical structure within regions. Numbers at dendrogram nodes represent approximately unbiased probabilities (AU values) based on multiscale bootstrap resampling. Dendrogram y-axis refers to assigned Euclidean distance between putative clusters. The location of cluster nodes corresponds to the degree of similarity between enclosed sites, with increasing height corresponding to decreased similarity, and where a distance of zero would indicate enclosed sites are identical.

3.3 Principal Component Analysis

Hierarchical principal component analyses (PCAs) were in agreement with F-statistics results. The PCA that included all individuals identified three clear genetic clusters. The first axis delineated Eastern and Western sites, identifying a longitudinal divide between the regions (Figure 4a). The second axis primarily delineated Pubnico Point from the rest of the Western sites. In total, the first and second axes explained 5.4% of the variation in the data set.

Given the strong hierarchical structure within these regions, I conducted a PCA on the putative Eastern and Western clusters separately. The second PCA on the Western cluster further delineated Pubnico Point as an independent grouping, separate from the other four sites within the cluster (Figure 4b). Of the remaining four Western sites, those originating from George's Bank (GA and GB) clustered together. The Bay of Fundy sites (GL and GP) were more clearly distinguishable as two non-overlapping clusters in accordance with the first PCA (Figure 4a) and F_{ST}-based dendrogram (Figure 3). The third PCA on the Eastern cluster further delineated the two Scotian Shelf sites (IR and MD) from the Newfoundland and Magdalen Island sites (Figure 4c). Similar to the two George's Bank sites, the two Scotian Shelf sites clustered together with individuals completely overlapping. Although Newfoundland and the Magdalen Islands appeared as two distinct genetic sites, there was some overlap of individual points between them, indicating potential admixture between these locations.

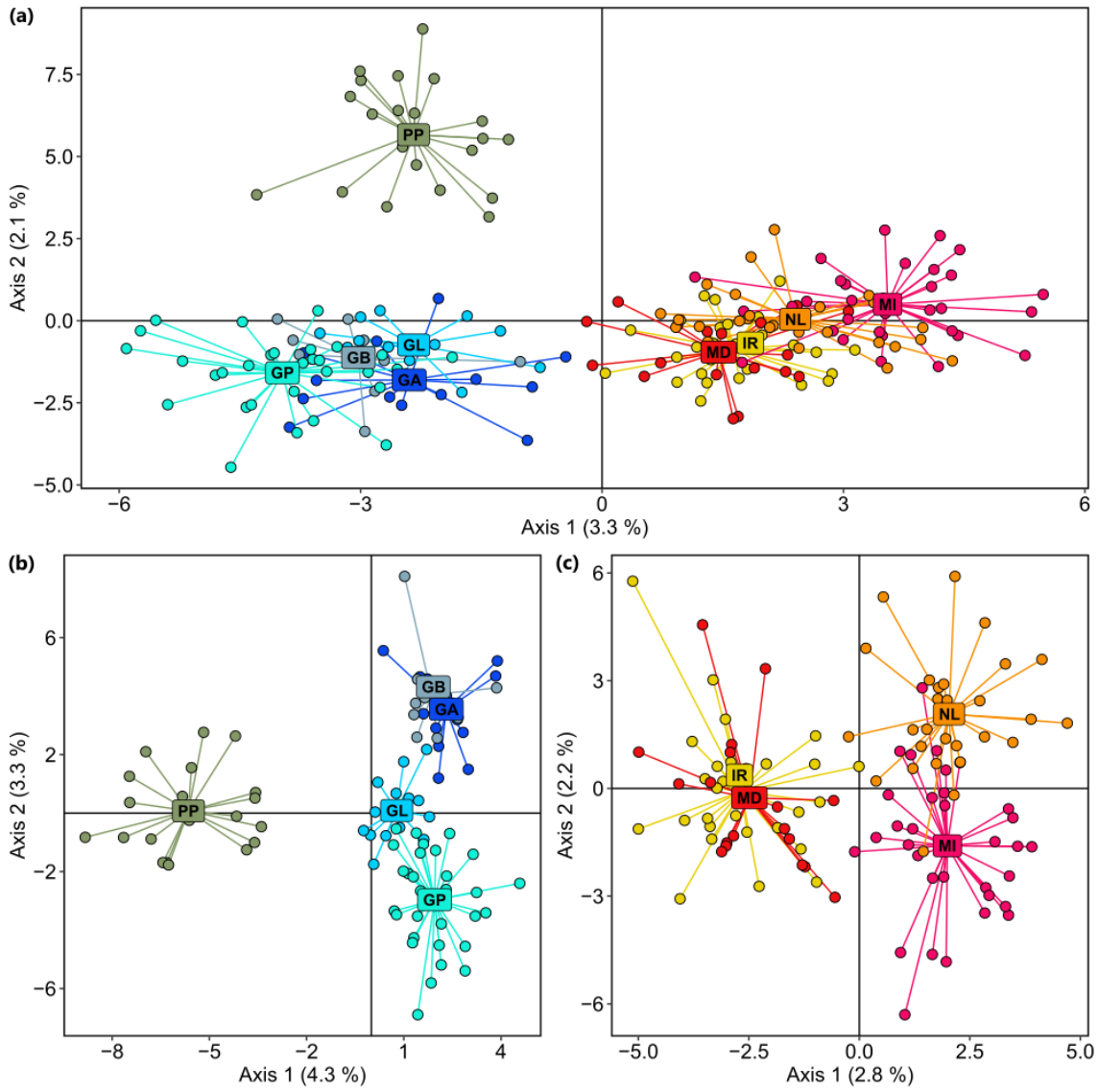

Figure 4 Principal components analysis (PCA): (a) all 1,052 SNPs and all 9 locations; (b) a subset dataset containing only sites from a putative "Western" cluster; (c) a subset dataset containing only sites from a putative "Eastern" cluster. Each point denotes an individual and each colour denotes the site where individuals originate. MD and GB labels were jittered to be legible within each figure.

A PCA conducted using the SNP dataset with 5 putatively adaptive markers removed uncovered these same patterns of spatial genetic structure as the full SNP data set (Figure 5). Thus, it was determined that the five outlier SNPs had no effect on the estimates of population structure, and I proceeded with the full dataset.

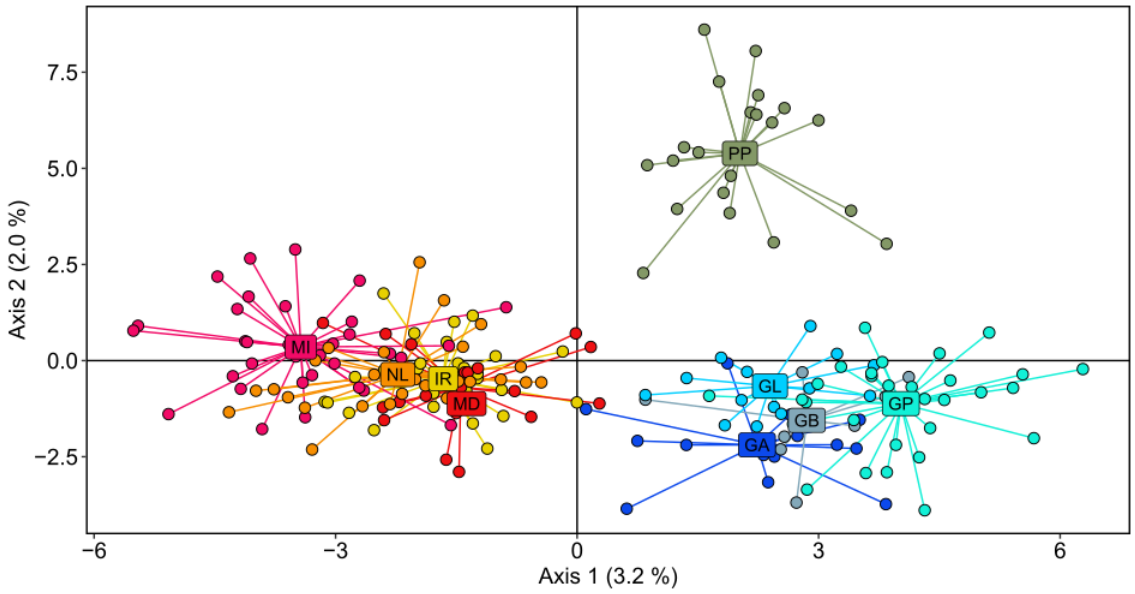

Figure 5 PCA using the SNP dataset with 5 putatively adaptive markers removed (n = 1,047 SNPs). MD identifier was jittered for visibility.

3.4 Admixture Analysis

In agreement with F_{ST} and PCA results, admixture analyses supported the demarcation of Western and Eastern sites into two genetic clusters. For the full dataset, cross-entropy criteria revealed a clear elbow pattern, with K = 2 ancestral genetic populations being optimal (Figure A4). Individual ancestry showed a clear longitudinal split forming Western and Eastern clusters (Figure 6a), corroborating evidence from the PCA (Figure 4a) and dendrogram (Figure 3). Notably, Green's Point (GP) individuals

displayed the lowest proportion of admixture from the Eastern cluster, and, similarly, Magdalen Island (MI) individuals displayed the lowest proportion of admixture from the Western cluster.

Additional admixture analyses conducted within each of these main clusters revealed hierarchical structure. At K = 2 within the Western cluster, Pubnico Point (PP) individuals displayed higher ancestry proportions to the ancestral cluster denoted in teal, whereas individuals from the rest of the region had high ancestry proportions to another cluster (shown in purple) (Figure 6b). Exploring further substructure within this Western region, at K = 3, differences in individual ancestry proportions became apparent between individuals from the Georges Bank (GA & GB), Green's Point (GP), and Pubnico Point (PP) sites (Figure 6d). Notably, Gannet Lighthouse (GL) individuals were highly admixed between the "grey" and "purple" ancestral clusters. In other words, Gannet's Lighthouse individuals tended to display intermediate genotypes relative to Georges Bank and Green's Point individuals, and the site is geographically located between these sites in the Bay of Fundy.

At K = 2 within the Eastern cluster, individual ancestry proportions distinguished the Scotian Shelf individuals (IR & MD; predominantly assigned to the "yellow" cluster) from the Magdalen Island and Newfoundland individuals (MI & NL; predominantly assigned to the "red" cluster) (Figure 6c). At K = 3, individual ancestry proportions further demarcate differences between individuals from the Magdalen Islands (MI; predominantly assigned to the "red" cluster) and Newfoundland (NL; predominantly assigned to the "brown" cluster). The Scotian Shelf individuals (IR & MD) largely exhibited high proportions to the "yellow" cluster (Figure 6e).

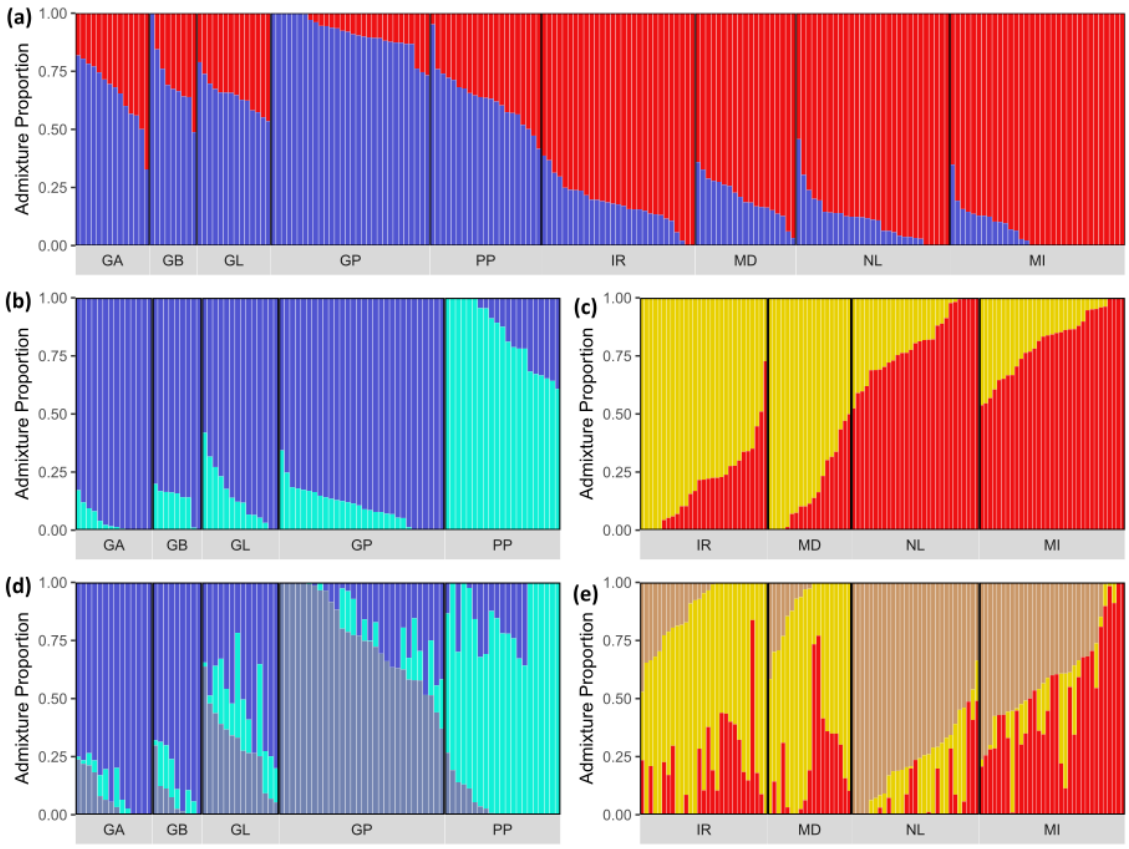

Figure 6 Plots of individual admixture coefficients estimated using the program LEA for (a) all nine *B. undatum* populations with K = 2; five sites from the Western cluster with (b) K = 2 and (d) K = 3 ancestral clusters; and four sites from the Eastern cluster with (c) K = 2 and (e) K = 3 ancestral clusters.

Taken as a whole, results of admixture analyses reveal clear broad-scale patterns of genetic structure, whereby individuals and sites separate into "Eastern" and "Western" groups based on their genotypes. Furthermore, there is genetic substructure within each of these major groups. Though my sampling design impairs me from demarcating exact population boundaries, of nine sites my results suggest seven distinct genetic populations exist. Pairs of sites along two deeper regions along the continental shelf (IR and MD; GA

and GB) each formed a single genetic population, and all other sampling sites represented genetically distinct populations.

3.5 Population Assignment Tests

Using a probability of assignment threshold ≥ 0.9 and treating the Western Cluster, Pubnico Point, and the Eastern Cluster as source populations, no recent immigrants were detected and self-assignment accuracy was 100%. Moreover, efficiency was high for each source population (mean = 81.46% ± 5.52% SD), meaning most individuals were assigned to a source population with high confidence. The unassigned individuals (Western Cluster = 14; Pubnico = 5; Eastern Cluster = 12) were admixed to varying degrees, in accordance with the admixture results (Figure 6).

Given the strong hierarchical structure observed (Figures 4, 5 and 6) I also explored population assignments to individual collection sites; however, I found these could not be used as source populations, as assignment probabilities to any given site were generally below the 0.9 threshold for high-confidence assignments.

4. Discussion

This study aimed to explore the population genetic consequences of limited dispersal capacity in a marine invertebrate. I focused on the western North Atlantic lineage of waved whelk (*Buccinum undatum*), a neogastropod which is expected to exhibit strong population structure due to its lack of a pelagic larval phase. Using 1,052 ddRAD derived SNP markers, I characterized the spatial genetic structure of *B. undatum* from the Gulf of Maine to the coast of Newfoundland, and into the Gulf of Saint Lawrence. My results demonstrate that *B. undatum* exhibits strong hierarchical structure throughout this region.

Consistently, my results establish two major genetically and geographically distinct *B. undatum* groupings separating sites into south-west and north-east groups. In addition, my results indicate significant genetic substructure exists within these regions, delineating seven distinct genetic populations across my nine sampling sites. This work provides one of the first descriptions of population genetic structure in an aplanktonic marine species across the western North Atlantic and adds to the growing literature documenting a major multi-species genetic cline in the region.

Broad-scale population structure

At the broadest spatial scale, my analyses of population structure, including hierarchical cluster analysis based on pairwise genetic differentiation, principal component analysis, and estimates of individual ancestry, consistently established two major *B. undatum* genetic groupings. Throughout this study, I have referred to these groupings as "Western" and "Eastern" clusters because my sampling sites could be divided by a putative longitudinal break that falls somewhere between 61° W and 65° W (Figure 1a). However, given the geography of the region, some ambiguity exists regarding whether a latitudinal or longitudinal divide best delineates these genetic groupings. In reality, sites are distributed in a south-west and north-east configuration relative to the eastern Nova Scotia coastline. I can gain insight into this major genetic break by comparing my results to other work on marine species in this region.

Population genetic studies have been conducted on multiple taxa in this region of the western North Atlantic, including fish (Bradbury et al., 2010), crustaceans (Benestan et al., 2015; Jorde et al., 2015; Jeffery et al., 2017), and molluscs (Chu et al., 2014; Van Wyngaarden et al., 2017; Lehnert et al., 2019), with many highlighting a biogeographic

break. For example, Stanley et al. (2018) examined clinal population structure in five planktonic developing species in this region, finding a multispecies latitudinal clinal transition around 44.61° N along the eastern coast of Nova Scotia. This pattern roughly aligns with the broad genetic break I detected for *B. undatum.* One exception in my study may be the Scotian Shelf sites (IR & MD), which cluster with the northeastern genetic group. These sites fall south of the 44.61° N line from Stanley et al. (2018); however, they lie within approximately 1 SD of the line (see dashed horizontal lines, Figure 1a). Therefore, similar mechanisms may underlie these patterns.

Several hypotheses have been invoked to explain major biogeographic breaks for co-distributed marine species in the western North Atlantic. First, historical vicariance events during the Pleistocene glaciation could potentially explain the genetic structure I observed. Historical glaciation patterns have been suggested to play a strong role in shaping phylogeographic patterns throughout *B. undatum's* range and, more broadly, there is evidence that glaciation has strongly influenced contemporary patterns of genetic structure in various North Atlantic marine species (Magnúsdóttir et al., 2019; Wares & Cunningham, 2001; Kenchington et al., 2009). A genetic consequence of postglacial recolonization is a loss of genetic diversity and genetic homogeneity, defined as the founder effect (Hewitt, 1996). For instance, Kenchington et al (2009) found that American lobster (*Homarus americanus*) populations in the Gulf of St. Lawrence exhibit relative genetic homogeneity compared to Gulf of Maine populations, supporting the possibility of a rapid northward range expansion and demarcating a north/south population divide similar to the "eastern" and "western" divide I observed in *B. undatum*. Though I did not observe high inbreeding coefficients in my Eastern sites (Table 3), which would be

expected for recently recolonized populations, I did observe relatively weak genetic structure in this region which could reflect relatively recent divergence following postglacial recolonization.

Second, ocean currents could similarly affect the direction and magnitude of species' dispersal and gene flow patterns, generating concordant spatial genetic patterns across multiple species (Weersing & Toonen, 2009; White et al., 2010). Third, the genetic cline may reflect adaptive divergence to gradients in seasonal ocean temperature minima (Wares, 2002; Stanley et al., 2018; Teske et al., 2019). Ocean temperature is a dominant selective force for other marine species in this region (Chu et al., 2014; Van Wyngaarden et al., 2018). Given the observations that (1) this genetic break occurs in species with a variety of life history traits, habitat requirements, and larval durations and (2) *B. undatum* is a direct developer whose connectivity patterns should not be strongly influenced by ocean currents, I suspect that contemporary divergence in response to an ocean temperature gradient could influence the major patterns of genetic structure observed.

Mean bottom temperature varies substantially throughout my study region (Figure 1b) and, in general, temperature strongly influences the biology of *B. undatum* with respect to growth rates and size of maturity (Emmerson et al., 2020; Borsetti et al., 2021), locomotion (Kideys & Hartnoll, 1991), and seasonal reproduction (Thatje et al., 2019; Borsetti et al., 2020). For example, timing of egg-laying varies spatially to coincide with region-specific minimum bottom temperature, so that hatching occurs during the warmest regional temperatures to promote rapid juvenile growth (Borsetti et al., 2020). With respect to genetic patterns, the major genetic break I observed corresponds with a steep gradient in minimum bottom temperature (Figure 1b; see also Stanley et al., 2018), with

warmer summer temperatures (~5-9°C) occurring in the south-west region and colder summer temperatures (~1–5°C) occurring in the north-east region. Spatial environmental heterogeneity can accentuate local adaptation alongside limited gene flow and contribute to genetic breaks such as the one I observed. For instance, it is possible that this temperature gradient has led to adaptations for heat stress among south-west populations, as has been observed in the aplanktonic intertidal dog whelk (*Nucella lapillus*) in this region (Chu et al., 2014). Interestingly, even in planktonic species, patterns of genetic structure can be influenced by ocean temperature, which may indicate local adaptation in the face of high gene flow. For instance, fine-scale patterns of genetic structure in the planktonic developing sea scallop (*Placopecten magellanicus*) in this region have been attributed to winter temperature driven juvenile mortality (Lehnert et al., 2019). To directly test the hypothesis that temperature influences spatial patterns of adaptive genetic variation in *B. undatum*, a larger genetic marker panel is needed to undertake genotype-environment association analyses (see Limitations and future directions).

Patterns of genetic sub-structure and potential depth-dependent connectivity

Our analyses of population structure revealed strong substructure within each of the two major *B. undatum* genetic groupings, delineating seven putative populations across the nine sampling sites. Aplanktonic benthic marine species are predicted to exhibit strong population structure because of their limited ability to disperse (Teske et al., 2007). Generally, my findings support this expectation for *B. undatum*. I found strong levels of differentiation, measured by pairwise F_{ST}, between most sampling sites (Figure 3), suggesting that the two major population clusters were each comprised of multiple genetically distinct populations reflecting limited genetic connectivity. Previous studies

on the *B. undatum* lineage in the eastern North Atlantic are largely congruent, having consistently found populations to be characterized by weak but significant genetic structure (Morrissey, 2019; Goodall et al., 2021).

Yet, genetic structure between some sites was weaker than might be expected given the hundreds of kilometers separating some locations and the lack of a long-distance dispersal strategy for *B. undatum*. For instance, Newfoundland and Magdalen Islands are separated by ~500 km and the Cabot Strait, yet I observed relatively weak pairwise genetic structure (F_{ST} = 0.016) between these sites. For the European populations of *B. undatum*, weak genetic structure has been attributed to large effective population sizes suppressing the effects of genetic drift (Mariani et al., 2012; Morrissey, 2019). Were this the case in the western North Atlantic, it could explain why genetic structure is weaker within the northeastern genetic cluster than the southwestern genetic cluster, as the Gulf of Maine and Bay of Fundy *B. undatum* populations are smaller and more fragmented (Kenchington & Glass, 1998; Borsetti et al., 2018). While total census population size is not a direct measure of effective population size, they are generally correlated (Waples et al., 2013). Alternatively, this weaker genetic structure within the northeastern cluster could be indicative of more recent population divergence following postglacial recolonization. Conversely, stepping stone dispersal (Mariani et al., 2012; Morrissey, 2019) or rare rafting events (Goodall et al., 2021) could facilitate inter- or intra- generational long-distance dispersal, respectively. While some aplanktonic species have been found to exhibit unique rafting dispersal strategies on floating substrata (Nikula et al., 2010; Pearman et al., 2020), my estimates of individual admixture proportions and the population assignment tests provided no evidence for long-distance dispersal among the sampled individuals.

Furthermore, to my knowledge, there is no direct evidence documenting a *B. undatum* rafting event.

Another difference between my findings and previous work in *B. undatum* is that studies of European *B. undatum* populations have consistently found evidence for genetic isolation by distance (IBD) patterns (Morrissey, 2019; Goodall et al., 2021). IBD refers to a pattern where genetic differences between populations increase with the geographic distance between them (Wright, 1943). My study does not have the sampling design to test IBD, but it is possible that a more intensive sampling design could demonstrate such a pattern in the Northwest Atlantic. For example, in the southwestern region of my study, where sampling was more spatially-intensive, sites seemingly followed an IBD pattern, with Gannet's Lighthouse individuals having intermediate admixture proportions relative to Georges Bank (to the South) and Green's Point (to the North). Although sites in the northeastern genetic region did not appear to exhibit an IBD trend, limited spatial coverage makes it difficult to draw definitive conclusions regarding the presence or absence of IBD for *B. undatum*.

Unexpectedly, I found evidence of potential depth-dependent genetic connectivity in *B. undatum*. Specifically, pairs of sampling sites within both the Scotian Shelf and Georges Bank exhibited no genetic structure ($F_{ST} = 0$) despite being separated by ~100 km and ~50 km, respectfully. Additionally, other pairs of sites in our study region, spaced at similar distances apart, did exhibit genetic structure (Figure 3). This finding indicates that gene flow may be elevated for *B. undatum* within these deeper shelf habitats. The effect of depth as a driver of genetic structure is not fully understood (reviewed by Taylor & Roterman, 2017) and previous work on European *B. undatum* populations has

documented conflicting patterns of structure between inshore and offshore sites. Weetman et al., (2006) and Mariani et al., (2012) found that, in the British Isles, shallower inshore populations experience more restricted gene flow compared to deeper offshore populations and suggested that offshore stepping-stone dispersal maintains higher rates of gene flow. In contrast, Goodall *et al*. (2021) found in Iceland that gene flow was stronger among shallower inshore sites, with deeper offshore sites exhibiting stronger genetic structure and unidirectional gene flow towards the coast. Reconciling my findings with this previous work and with *B. undatum's* limited movement ability (Himmelman, 1988) will require more extensive sampling across multiple depths.

Evidence for genetic isolation at Pubnico Point, NS

Pubnico Point, NS, one of my two intertidal sites, demonstrated unexpectedly strong genetic structure compared to all other sampling locations (Figure 3 and 4). While *B. undatum* is widely considered a subtidal species, it also has a patchy distribution within the intertidal zone. Populations that inhabit this zone experience distinct environmental conditions compared to those within the subtidal. Generally, heat is regarded as the strongest environmental stressor in the intertidal (Somero, 2002) and has been found to drive local adaptation in other intertidal marine invertebrates (Chu et al., 2014). Although temperature could plausibly influence the strong genetic structure observed at Pubnico Point, my neutral genetic data cannot test this hypothesis. To note, while the observed genetic is relatively strong, it does not exceed the proposed upper threshold ($F_{ST} > 0.35$) demarcating an inter-species comparison (Hey & Pinho, 2012). Furthermore, it is surprising that stronger genetic structure is not observed at my other intertidal site, Green's Point, NB, though I note that both intertidal sites exhibited a relatively high number of

private alleles. If ocean currents can facilitate occasional dispersal events in this species (i.e., via rafting), they may help explain the difference in genetic 'isolation' between Pubnico Point and Green's Point. Pubnico Point is sheltered on the western bank of a peninsula, whereas Green's Point experiences strong Bay of Fundy tidal currents that may facilitate gene flow with nearby subtidal populations. Kamel et al. (2014) found that physical transport processes explained genetic structure in another aplanktonic gastropod (*Solenosteira macrospira).* Finally, my sparse sampling in this region (Figure 1a) means I cannot rule out that Pubnico Point actually has little genetic structure with nearby intertidal and/or subtidal populations which I have not yet sampled from. To elucidate the patterns and mechanisms driving the putative genetic isolation at Pubnico Point, more intensive sampling, including paired intertidal-subtidal sites, is warranted.

Management implications

My results have implications for *B. undatum* fisheries in the western North Atlantic, as delineating patterns of spatial genetic structure is broadly considered valuable for effective species management (Palumbi, 2003; Funk et al., 2012; Crawford & Oleksiak, 2016). *B. undatum* is particularly vulnerable to overharvesting because of its slow reproductive cycle and aplanktonic life history (Borsetti et al., 2018; Ashfaq et al., 2019). While *B. undatum* represents a small, emerging fishery in the western North Atlantic (DFO, 2020), it continues to be a large fishery in Europe where populations are vulnerable to localized overexploitation (Emmerson et al., 2018). Therefore, the demarcation of independently evolving populations, or evolutionary significant units (ESUs) (Moritz, 1994), is valuable for identifying stock boundaries. My results show that at least two broad *B. undatum* ESUs exist within the western North Atlantic, partitioned into a southwestern ESU and a

northeastern ESU relative to the eastern Nova Scotia coastline. Moreover, the strong genetic structure within these major groups suggests that localized *B. undatum* ESUs are likely to exist both between and within Atlantic Canada provinces where there are several commercial fishing operations, primarily concentrated in Québec and Newfoundland.

In addition to identifying ESUs, the demarcation of management units (MUs), has management value. MUs refer to populations that experience limited demographic connectivity, such that they should be managed separately to prevent localized overexploitation (Palsboll et al., 2007). Notably, the lack of genetic structure observed in deeper *B. undatum* populations does not necessarily reflect strong demographic connectivity in deeper environments. Without data on the demographic independence of sampled populations (i.e., evidence that changes in local abundance are a function of birth and death rates alone, not dispersal rates; Cooper & Hurd, 2019) I cannot demarcate them into MUs. However, the species' high phenotypic variability (Borsetti et al., 2018; Ashfaq et al., 2019) alongside limited movement potential suggest that limited demographic connectivity is likely and is an added concern for local fishery management.

Limitations and future directions

In this study, the primary limitation was the sparse geographic sampling design, owing to travel and field work restrictions during the COVID-19 pandemic. For instance, I was unable to sample *B. undatum*'s entire distribution into the Gulf of St Lawrence, despite Québec representing the highest annual *B. undatum* fishery landings in Canada (Brulotte, 2019; DFO, 2020). Moreover, sparse sampling coverage limited my ability to precisely identify the location of the major genetic breaks. A finer-resolution sampling design would allow further investigation into the apparent lack of genetic structure on

deep shelf habitats, as well as further comparisons between intertidal and subtidal populations.

Likewise, my neutral SNP marker panel prevented me from conducting genotype-environment association analyses to address drivers of local adaptation in *B. undatum.* While the restriction enzymes I used resulted in high coverage, and thus high confidence genotypes, restriction enzymes which cut more frequently would provide a much larger SNP panel (though they would also require a concomitant increase in sequencing effort). Obtaining a larger SNP panel would likely increase my ability to detect outlier loci, which would allow me to test for associations between SNPs and key environmental variables (e.g., bottom temperature; ocean currents; depth; habitat continuity) to begin identifying drivers of local adaptation in *B. undatum*. The role environmental variation has in driving contemporary genetic structure within *B. undatum* is knowledge I currently lack. This information would have important implications for predicting how *B. undatum* will be affected by a changing seascape during climate change (Gerber et al., 2014), and would also provide a broader understanding of the suite of factors influencing contemporary patterns of genetic structure across aplanktonic species.

Finally, I note that the restriction enzymes chosen for this study were different than those used in the only other ddRADseq study conducted on *B. undatum* by Goodall et al. (2021). I was unaware of the manuscript by Goodall et al. (2021) when designing my own study. This difference in restriction enzymes and sequencing platform limited my ability to potentially integrate their Canadian samples from a site in the Gulf of St. Lawrence into my analysis. However, this difference did not prevent me from making comparisons about the strength of genetic structure at similar spatial scales between the two lineages of the

B. undatum, which was my primary objective in terms of comparing my results with previous genetic work on this species.

5. Conclusion

This study on the population genetic structure of *B. undatum* in the western North Atlantic reaffirms evidence from the eastern North Atlantic that *B. undatum* exhibits strong hierarchical population genetic structure across its range. My primary objective was to describe the patterns of population structure in *B. undatum*, and my findings of strong population structure reflect the expectation that aplanktonic benthic marine species possess a limited ability to disperse (Teske et al., 2007). Moreover, I found new evidence supporting a major genetic break along the eastern Nova Scotia coastline which seemingly follows a previously discovered multispecies genetic cline (Stanley et al., 2018). Unexpectedly, I found evidence of potential depth-dependent genetic connectivity patterns, whereby connectivity appears to be especially strong along sampled regions of deep continental shelves. Conversely, I found evidence that genetic structure is particularly strong for this species in at least one intertidal location. Despite my inability to directly address the drivers of these notable genetic patterns, my findings generate new questions regarding the strength of depth and local environmental conditions as drivers of connectivity and local adaptation.

In conclusion, this study provides valuable information about the spatial genetic structure of *B. undatum*, an aplanktonic marine invertebrate, and has implications for managing the *B. undatum* fishery across Atlantic Canada. I highlight the strong genetic structure both between and within regions where commercial *B. undatum* fishery

operations are ongoing. In the future, more population genetic data should be gathered on *B. undatum* using more spatially intensive sampling and a larger SNP marker panel. These data would allow us to begin elucidating the drivers of genetic structure and should be gathered alongside demographic information to best inform management practices.

Bibliography

Anderson, E. C. (2010). Assessing the power of informative subsets of loci for population assignment: Standard methods are upwardly biased. *Molecular Ecology Resources*, *10*(4), 701–710. https://doi.org/10.1111/j.1755-0998.2010.02846.x

Ashfaq, U., Mugridge, A., & Hatcher, B. G. (2019). Size at sexual maturity of waved whelk (*Buccinum undatum*) on the Eastern Scotian Shelf. *Fisheries Research*, *212*, 12–20. https://doi.org/10.1016/j.fishres.2018.11.025

Ayre, D. J., Minchinton, T. E., & Perrin, C. (2009). Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular Ecology*, *18*(9), 1887–1903. https://doi.org/10.1111/j.1365-294X.2009.04127.x

Balbar, A. C., & Metaxas, A. (2019). The current application of ecological connectivity in the design of marine protected areas. *Global Ecology and Conservation*, *17*, e00569. https://doi.org/10.1016/j.gecco.2019.e00569

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, *24*(13), 3299–3315. https://doi.org/10.1111/mec.13245

Borsetti, S., Hollyman, P. R., & Munroe, D. (2021). Using a sclerochronological approach to determine a climate-growth relationship for waved whelk, *Buccinum undatum*, in the U.S. Mid-Atlantic. *Estuarine, Coastal and Shelf Science*, *252*, 107255. https://doi.org/10.1016/j.ecss.2021.107255

Borsetti, S., Munroe, D., Rudders, D. B., Dobson, C., & Bochenek, E. A. (2018). Spatial variation in life history characteristics of waved whelk (*Buccinum undatum* L.) on the U.S. Mid-Atlantic continental shelf. *Fisheries Research*, *198*, 129–137. https://doi.org/10.1016/j.fishres.2017.10.006

Borsetti, S., Munroe, D., Rudders, D., & Chang, J.-H. (2020). Timing of the reproductive cycle of waved whelk, *Buccinum undatum*, on the U.S. Mid-Atlantic Bight. *Helgoland Marine Research*, *74*(1), 5. https://doi.org/10.1186/s10152-020-00537-6

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., & Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3725–3734. https://doi.org/10.1098/rspb.2010.0985

Bradbury, I. R., Laurel, B., Snelgrove, P. V. R., Bentzen, P., & Campana, S. E. (2008). Global patterns in marine dispersal estimates: The influence of geography, taxonomic category and life history. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1644), 1803–1809. https://doi.org/10.1098/rspb.2008.0216

Bradbury, I. R., Wringe, B. F., Watson, B., Paterson, I., Horne, J., Beiko, R., Lehnert, S. J., Clément, M., Anderson, E. C., Jeffery, N. W., Duffy, S., Sylvester, E., Robertson, M., & Bentzen, P. (2018). Genotyping-by-sequencing of genome-wide microsatellite loci reveals fine-scale harvest composition in a coastal Atlantic salmon fishery. *Evolutionary Applications*, *11*(6), 918–930. https://doi.org/10.1111/eva.12606

Brulotte, S. (2019). Assessment of the whelk fishery in Quebec's inshore waters – methodology and results. *DFO Can. Sci. Advis. Sec. Res. Doc.*, *2019/040*, xi + 76.

Burgess, S. C., Baskett, M. L., Grosberg, R. K., Morgan, S. G., & Strathmann, R. R. (2016). When is dispersal for dispersal? Unifying marine and terrestrial perspectives. *Biological Reviews*, *91*(3), 867–882. https://doi.org/10.1111/brv.12198

Cahill, A., & Viard, F. (2014). Genetic structure in native and non-native populations of the direct-developing gastropod *Crepidula convexa*. *Marine Biology*, *161*(10), 2433–2443. https://doi.org/10.1007/s00227-014-2519-2

Caley, M. J., Carr, M. H., Hixon, M. A., Hughes, T. P., Jones, G. P., & Menge, B. A. (1996). Recruitment and the local dynamics of open marine populations. *Annual Review of Ecology and Systematics*, *27*(1), 477–500. https://doi.org/10.1146/annurev.ecolsys.27.1.477

Carrier, T., Reitzel, A., & Heyland, A. (2018). Evolutionary Ecology of Marine Invertebrate Larvae (First Edition). Oxford University Press. https://doi.org/10.1093/oso/9780198786962.001.0001

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. https://doi.org/10.1111/mec.12354

Chen, K.-Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsin, S. A. (2018). assignPOP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution*, *9*(2), 439–446. https://doi.org/10.1111/2041-210X.12897

Chu, N. D., Kaluziak, S. T., Trussell, G. C., & Vollmer, S. V. (2014). Phylogenomic analyses reveal latitudinal population structure and polymorphisms in heat stress genes in the North Atlantic snail *Nucella lapillus*. *Molecular Ecology*, *23*(7), 1863–1873. https://doi.org/10.1111/mec.12681

Cooper, G. J., & Hurd, L. E. (2019). The House and the Household: Habitat, Demographic Independence, and Ecological Populations. *Philosophical Topics*, 47(1), 21–44. https://www.jstor.org/stable/26948090.

Crawford, D. L., & Oleksiak, M. F. (2016). Ecological population genomics in the marine environment. *Briefings in Functional Genomics*, *15*(5), 342–351. https://doi.org/10.1093/bfgp/elw008

de Vooys, C. G. N., & van der Meer, J. (2010). The whelk (*Buccinum undatum* L.) in the western Dutch Wadden Sea in the period 1946–1970: Assessment of population characteristics and fishery impact. *Journal of Sea Research*, *63*(1), 11–16. https://doi.org/10.1016/j.seares.2009.08.005

DFO. (2020). Commercial Fisheries: Landings. https://www.dfo-mpo.gc.ca/stats/commercial/land-debarq/sea-maritimes/s2020aq-eng.htm

Emmerson, J. A., Haig, J. A., Bloor, I. S. M., & Kaiser, M. J. (2018). The complexities and challenges of conserving common whelk (*Buccinum undatum* L.) fishery resources: Spatio-temporal study of variable population demographics within an environmental context. *Fisheries Research*, *204*, 125–136. https://doi.org/10.1016/j.fishres.2018.02.015

Emmerson, J. A., Hollyman, P. R., Bloor, I. S. M., & Jenkins, S. R. (2020). Effect of temperature on the growth of the commercially fished common whelk (*Buccinum undatum*, L.): A regional analysis within the Irish Sea. *Fisheries Research*, *223*, 105437. https://doi.org/10.1016/j.fishres.2019.105437

Fahy, E. (2001). Conflict between two inshore fisheries: For whelk (Buccinum undatum) and brown crab (*Cancer pagurus*), in the southwest Irish Sea. *Hydrobiologia*, *465*(1), 73–83. https://doi.org/10.1023/A:1014544925248

Fahy, E. (2008). Performance of an inshore fishery in the absence of regulatory enforcement. *Marine Policy*, *32*(6), 1037–1042. https://doi.org/10.1016/j.marpol.2008.02.010

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Frichot E, Francois O (2015). LEA: an R package for Landscape and Ecological Association studies. *Methods in Ecology and Evolution*, 6, 925–929 https://doi.org/10.1111/2041-210X.12382

Gerber, L. R., Mancha-Cisneros, M. D. M., O'Connor, M. I., & Selig, E. R. (2014). Climate change impacts on connectivity in the ocean: Implications for conservation. *Ecosphere*, *5*(3), art33. https://doi.org/10.1890/ES13-00336.1

Goodall, J., Westfall, K. M., Magnúsdóttir, H., Pálsson, S., Örnólfsdóttir, E. B., & Jónsson, Z. O. (2021). RAD sequencing of common whelk, *Buccinum undatum*, reveals fine-scale population structuring in Europe and cryptic speciation within the North Atlantic. *Ecology and Evolution*, 11:2616–2629. https://doi.org/10.1002/ece3.7219

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Haig, J. A., Pantin, J. R., Salomonsen, H., Murray, L. G., & Kaiser, M. J. (2015). Temporal and spatial variation in size at maturity of the common whelk (*Buccinum undatum*). *ICES Journal of Marine Science*, *72*(9), 2707–2719. https://doi.org/10.1093/icesjms/fsv128

Hellberg, M. E., Burton, R. S., Neigel, J. E., & Palumbi, S. R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science*, *70*(1), 18.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58(3), 247–276. https://doi.org/10.1006/bijl.1996.0035

Hey, J., & Pinho, C. (2012). Population Genetics and Objectivity in Species Diagnosis. *Evolution*, 66(5), 1413–1429. https://doi.org/10.1111/j.1558-5646.2011.01542.x

Himmelman, J. H. (1988). Movement of whelks (*Buccinum undatum*) towards a baited trap. *Marine Biology*, *97*(4), 521–531. https://doi.org/10.1007/BF00391048

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62–82. https://doi.org/10.1111/mec.15720

Hollyman, P. R., Chenery, S. R. N., Leng, M. J., Laptikhovsky, V. V., Colvin, C. N., & Richardson, C. A. (2018). Age and growth rate estimations of the commercially fished gastropod *Buccinum undatum*. *ICES Journal of Marine Science*, *75*(6), 2129–2144. https://doi.org/10.1093/icesjms/fsy100

Hoskin, M. g. (1997). Effects of contrasting modes of larval development on the genetic structures of populations of three species of prosobranch gastropods. *Marine Biology*, *127*(4), 647. https://doi.org/10.1007/s002270050055

Hubley, P.B., Zisserson, B.M., Cameron, B.J., and Choi, J.S. (2018). Assessment of Scotian Shelf Snow Crab in 2016. *DFO Can. Sci. Advis. Sec. Res. Doc.* 053. vii + 136

Jeffery, N. W., DiBacco, C., Van Wyngaarden, M., Hamilton, L. C., Stanley, R. R. E., Bernier, R., FitzGerald, J., Matheson, K., McKenzie, C. H., Nadukkalam Ravindran, P., Beiko, R., & Bradbury, I. R. (2017). RAD sequencing reveals genomewide divergence between independent invasions of the European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Ecology and Evolution*, *7*(8), 2513–2524. https://doi.org/10.1002/ece3.2872

Jorde, P. E., Søvik, G., Westgaard, J.-I., Albretsen, J., André, C., Hvingel, C., Johansen, T., Sandvik, A. D., Kingsley, M., & Jørstad, K. E. (2015). Genetically distinct populations of northern shrimp, *Pandalus borealis*, in the North Atlantic: Adaptation to different temperatures as an isolation factor. *Molecular Ecology*, *24*(8), 1742–1757. https://doi.org/10.1111/mec.13158

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences*, *118*(48), e2104642118. https://doi.org/10.1073/pnas.2104642118

Kelly, R. P., & Palumbi, S. R. (2010). Genetic Structure Among 50 Species of the Northeastern Pacific Rocky Intertidal Community. *PLOS ONE*, *5*(1), e8594. https://doi.org/10.1371/journal.pone.0008594

Kenchingon, E., & Glass, A. (1998). Local adaptation and sexual dimorphism in the Waved Whelk (*Buccinum undatum*) in Atlantic Nova Scotia with applications to fisheries management. Can. Tech. Rep. Fish. Aquat. Sci. 2237: iv + 43 p.

Kenchington, E. L., Harding, G. C., Jones, M. W., & Prodöhl, P. A. (2009). Pleistocene glaciation events shape genetic structure across the range of the American lobster, Homarus americanus. *Molecular Ecology*, 18(8), 1654–1667. https://doi.org/10.1111/j.1365-294X.2009.04118.x

Kideys, A. E., & Hartnoll, R. G. (1991). Energetics of mucus production in the common whelk *Buccinum undatum* L. *Journal of Experimental Marine Biology and Ecology*, *150*(1), 91–105. https://doi.org/10.1016/0022-0981(91)90108-9

Kritzer, J. P., & Sale, P. F. (2006). Marine metapopulations. Academic Press, Amsterdam. https://doi.org/10.1016/B978-0-12-088781-1.X5000-6

Kyle, C. J., & Boulding, E. G. (2000). Comparative population genetic structure of marine gastropods (*Littorina* spp.) with and without pelagic larval dispersal. *Marine Biology*, *137*(5), 835–845. https://doi.org/10.1007/s002270000412

Layton, K. K. S., Dempson, B., Snelgrove, P. V. R., Duffy, S. J., Messmer, A. M., Paterson, I. G., Jeffery, N. W., Kess, T., Horne, J. B., Salisbury, S. J., Ruzzante, D. E., Bentzen, P., Côté, D., Nugent, C. M., Ferguson, M. M., Leong, J. S., Koop, B. F., & Bradbury, I. R. (2020). Resolving fine-scale population structure and fishery exploitation using sequenced microsatellites in a northern fish. *Evolutionary Applications*, *13*(5), 1055–1068. https://doi.org/10.1111/eva.12922

Lehnert, S. J., DiBacco, C., Van Wyngaarden, M., Jeffery, N. W., Ben Lowen, J., Sylvester, E. V. A., Wringe, B. F., Stanley, R. R. E., Hamilton, L. C., & Bradbury, I. R. (2019). Fine-scale temperature-associated genetic structure between inshore and offshore populations of sea scallop (*Placopecten magellanicus*). *Heredity*, *122*(1), 69–80. https://doi.org/10.1038/s41437-018-0087-9

Liggins, L., Treml, E. A., Possingham, H. P., & Riginos, C. (2016). Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *Journal of Biogeography*, *43*(2), 256–267. https://doi.org/10.1111/jbi.12647

Magnúsdóttir, H., Pálsson, S., Westfall, K. M., Jónsson, Z. O., Goodall, J., & Örnólfsdóttir, E. B. (2019). Revised phylogeography of the common whelk *Buccinum undatum* (Gastropoda: Buccinidae) across the North Atlantic. *Biological Journal of the Linnean Society*, *127*(4), 890–899. https://doi.org/10.1093/biolinnean/blz060

Magnúsdóttir, H., Pálsson, S., Westfall, K. M., Jónsson, Z. O., & Örnólfsdóttir, E. B. (2018). Shell morphology and color of the subtidal whelk *Buccinum undatum* exhibit fine-scaled spatial patterns. *Ecology and Evolution*, *8*(9), 4552–4563. https://doi.org/10.1002/ece3.4015

Mariani, S., Peijnenburg, K., & Weetman, D. (2012). Independence of neutral and adaptive divergence in a low dispersal marine mollusc. *Marine Ecology Progress Series*, *446*, 173–187. https://doi.org/10.3354/meps09507

Martel, A., Larrivée, D. H., & Himmelman, J. H. (1986a). Behaviour and timing of copulation and egg-laying in the neogastropod *Buccinum undatum* L. *Journal of Experimental Marine Biology and Ecology*, *96*(1), 27–42. https://doi.org/10.1016/0022-0981(86)90011-0

Martel, A., Larrivée, D. H., Klein, K. R., & Himmelman, J. H. (1986b). Reproductive cycle and seasonal feeding activity of the neogastropod *Buccinum undatum*. *Marine Biology*, *92*(2), 211–221. https://doi.org/10.1007/BF00392838

McIntyre, R., Lawler, A., & Masefield, R. (2015). Size of maturity of the common whelk, *Buccinum undatum*: Is the minimum landing size in England too low? *Fisheries Research*, *162*, 53–57. https://doi.org/10.1016/j.fishres.2014.10.003

Metivier, S. L., Kim, J.-H., & Addison, J. A. (2017). Genotype by sequencing identifies natural selection as a driver of intraspecific divergence in Atlantic populations of the high dispersal marine invertebrate, *Macoma petalum*. *Ecology and Evolution*, *7*(19), 8058–8072. https://doi.org/10.1002/ece3.3332

Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, *9*(10), 373–375. https://doi.org/10.1016/0169-5347(94)90057-4

Morrissey, D. (2020). Population genomics of a marine gastropod with limited dispersal capabilities. Masters Thesis. Universidad do Algarve; UA01-Teses. http://hdl.handle.net/10400.1/14615.

Nanninga, & Berumen, M. L. (2014). The role of individual variation in marine larval dispersal. *Frontiers in Marine Science*, 1:71 https://doi.org/10.3389/fmars.2014.00071

Nanninga, G., & Manica, A. (2018). Larval swimming capacities affect genetic differentiation and range size in demersal marine fishes. *Marine Ecology Progress Series*, *589*, 1–12. https://doi.org/10.3354/meps12515

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*(49), 19052–19059. https://doi.org/10.1073/pnas.0800375105

Nikula, R., Fraser, C., Spencer, H., & Waters, J. (2010). Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans. *Marine Ecology Progress Series*, *405*, 221–230. https://doi.org/10.3354/meps08523

Palsboll, P., Berube, M., & Allendorf, F. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, *22*(1), 11–16. https://doi.org/10.1016/j.tree.2006.09.003

Pálsson, S., Magnúsdóttir, H., Reynisdóttir, S., Jónsson, Z. O., & Örnólfsdóttir, E. B. (2014). Divergence and molecular variation in common whelk *Buccinum undatum* (Gastropoda: Buccinidae) in Iceland: a trans-Atlantic comparison. *Biological Journal of the Linnean Society*, *111*(1), 145–159. https://doi.org/10.1111/bij.12191

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, *13*(sp1), 146–158. https://doi.org/10.1890/1051-0761(2003)013[0146:PGDCAT]2.0.CO;2

Pearman, W. S., Wells, S. J., Silander, O. K., Freed, N. E., & Dale, J. (2020). Concordant geographic and genetic structure revealed by genotyping-by-sequencing in a New Zealand marine isopod. *Ecology and Evolution*, 10(24), 13624–13639. https://doi.org/10.1002/ece3.6802

Pechenik, J. (1999). On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology Progress Series*, *177*, 269–297. https://doi.org/10.3354/meps177269

Puritz, J. B., Keever, C. C., Addison, J. A., Barbosa, S. S., Byrne, M., Hart, M. W., Grosberg, R. K., & Toonen, R. J. (2017). Life-history predicts past and present population connectivity in two sympatric sea stars. *Ecology and Evolution*, *7*(11), 3916–3930. https://doi.org/10.1002/ece3.2938

Riginos, C., Douglas, K. E., Jin, Y., Shanahan, D. F., & Treml, E. A. (2011). Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, *34*(4), 566–575. https://doi.org/10.1111/j.1600-0587.2010.06511.x

Rochette, R., & Himmelman, J. H. (1996). Does vulnerability influence trade-offs made by whelks between predation risk and feeding opportunities? *Animal Behaviour*, *52*(4), 783–794. https://doi.org/10.1006/anbe.1996.0223

Rochette, R., Maltais, M.-J., Dill, L. M., & Himmelman, J. H. (1999). Interpopulation and context-related differences in responses of a marine gastropod to predation risk. *Animal Behaviour*, *57*(4), 977–987. https://doi.org/10.1006/anbe.1998.1061

Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., Heyden, S. von der, & Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, *554*, 1–19. https://doi.org/10.3354/meps11792

Selkoe, K. A., & Toonen, R. J. (2011). Marine connectivity: A new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, *436*, 291–305. https://doi.org/10.3354/meps09238

Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, 51(3), 492–508. https://doi.org/10.1080/10635150290069913

Siegel, D., Kinlan, B., Gaylord, B., & Gaines, S. (2003). Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series*, *260*, 83–96. https://doi.org/10.3354/meps260083

Somero, G. N. (2002). Thermal physiology and vertical zonation of intertidal animals: optima, limits, and costs of living. *Integrative and Comparative Biology*, *42*(4), 780–789. https://doi.org/10.1093/icb/42.4.780

Sotka, E. E. (2012). Natural selection, larval dispersal, and the geography of phenotype in the sea. *Integrative and Comparative Biology*, *52*(4), 538–545. https://doi.org/10.1093/icb/ics084

Stanley, R. R. E., DiBacco, C., Lowen, B., Beiko, R. G., Jeffery, N. W., Van Wyngaarden, M., Bentzen, P., Brickman, D., Benestan, L., Bernatchez, L., Johnson, C., Snelgrove, P. V. R., Wang, Z., Wringe, B. F., & Bradbury, I. R. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, *4*(3), eaaq0929. https://doi.org/10.1126/sciadv.aaq0929

Strathmann, R. R. (1986). What controls the type of larval development? Summary statement for the evolution session. *Bulletin of Marine Science*, *39*(2), 616-622.

Taylor, M. L., & Roterman, C. N. (2017). Invertebrate population genetics across Earth's largest habitat: The deep-sea floor. *Molecular Ecology*, *26*(19), 4872–4896. https://doi.org/10.1111/mec.14237

Teske, P. R., Papadopoulos, I., Zardi, G. I., McQuaid, C. D., Edkins, M. T., Griffiths, C. L., & Barker, N. P. (2007). Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: Planktonic, abbreviated and direct development. *Marine Biology*, *152*(3), 697–711. https://doi.org/10.1007/s00227-007-0724-y

Teske, P. R., Sandoval-Castillo, J., Golla, T. R., Emami-Khoyi, A., Tine, M., von der Heyden, S., & Beheregaray, L. B. (2019). Thermal selection as a driver of marine ecological speciation. *Proceedings of the Royal Society B: Biological Sciences*, 286(1896), 20182023. https://doi.org/10.1098/rspb.2018.2023

Thatje, S., Dunbar, C. G., & Smith, K. E. (2019). Temperature-driven inter-annual variability in reproductive investment in the common whelk *Buccinum undatum*. *Journal of Sea Research*, *148–149*, 17–22. https://doi.org/10.1016/j.seares.2019.03.003

Thiel M, Gutow L. (2005a). The ecology of rafting in the marine environment. I. The floating substrata. *Oceanography and Marine Biology: An Annual Review*, 42, 181–264.

Thiel M, Gutow L. 2005b. The ecology of rafting in the marine environment. II. The rafting organisms and community. *Oceanography and Marine Biology: An Annual Review*, 43, 279–418.

Thiel M, Haye PA. 2006. The ecology of rafting in the marine environment. III. Biogeographical and evolutionary consequences. *Oceanography and Marine Biology: An Annual Review*, 44, 323–429.

Todd, C. D. (1998). Larval supply and recruitment of benthic invertebrates: Do larvae always disperse as much as we believe? *Hydrobiologia*, *375*(0), 1–21. https://doi.org/10.1023/A:1017007527490

Van Wyngaarden, M., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E., & Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RADseq-derived SNPs. *Evolutionary Applications*, *10*(1), 102–117. https://doi.org/10.1111/eva.12432

Van Wyngaarden, M., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Zhan, L., Beiko, R. G., & Bradbury, I. R. (2018). Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*. *Ecology and Evolution*, *8*(5), 2824–2841. https://doi.org/10.1002/ece3.3846

Waples, R. S., Luikart, G., Faulkner, J. R., & Tallmon, D. A. (2013). Simple life-history traits explain key effective population size ratios across diverse taxa. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1768), 20131339. https://doi.org/10.1098/rspb.2013.1339

Wares, J. P. (2002). Community genetics in the Northwestern Atlantic intertidal. *Molecular Ecology*, *11*(7), 1131–1144. https://doi.org/10.1046/j.1365-294X.2002.01510.x

Wares, J. P., & Cunningham, C. W. (2001). Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution*, *55*(12), 2455–2469. https://doi.org/10.1111/j.0014-3820.2001.tb00760.x

Weersing, K., & Toonen, R. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, *393*, 1–12. https://doi.org/10.3354/meps08287

Weetman, D., Hauser, L., Bayes, M., Ellis, J., & Shaw, P. (2006). Genetic population structure across a range of geographic scales in the commercially exploited marine gastropod *Buccinum undatum*. *Marine Ecology Progress Series*, *317*, 157–169. https://doi.org/10.3354/meps317157

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1688), 1685–1694. https://doi.org/10.1098/rspb.2009.2214

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. *The American Naturalist*, *186*(S1), S24–S36. https://doi.org/10.1086/682949

Winston, J. E. (2012). Dispersal in marine organisms without a pelagic larval phase. *Integrative and Comparative Biology*, *52*(4), 447–457. https://doi.org/10.1093/icb/ics040

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/genetics/28.2.114

Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M.-J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, *27*(10), 2347–2364. https://doi.org/10.1111/mec.14589

Appendix

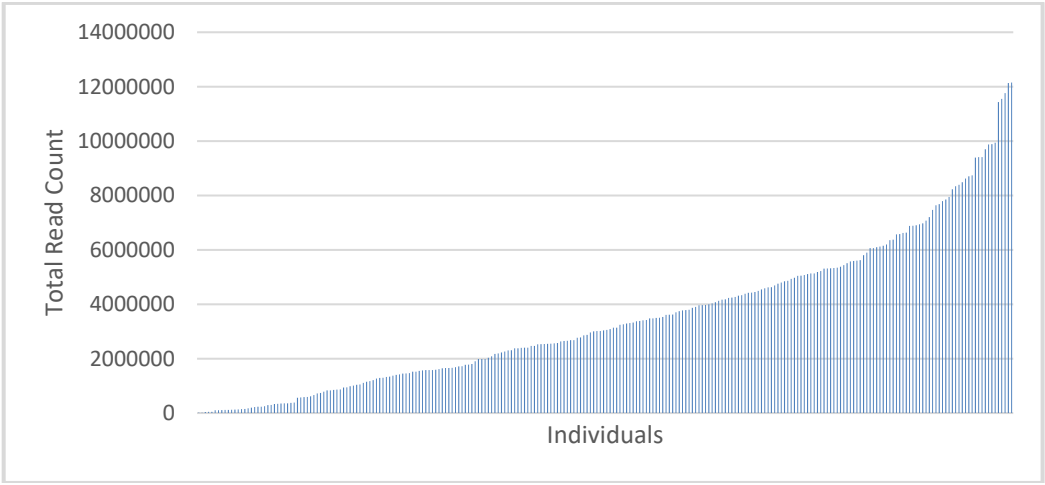

Figure A1 Distribution of retained reads for all 250 samples post quality control, pre-filtering.

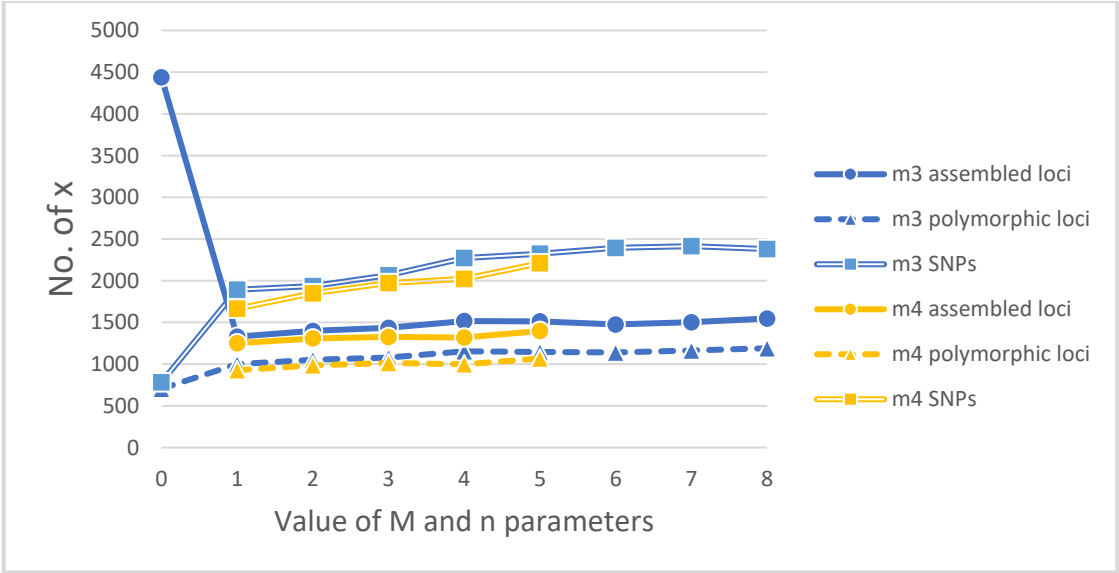

Figure A2 Number of loci shared by 80% of samples depending on m, M and n parameter values used during alignment.

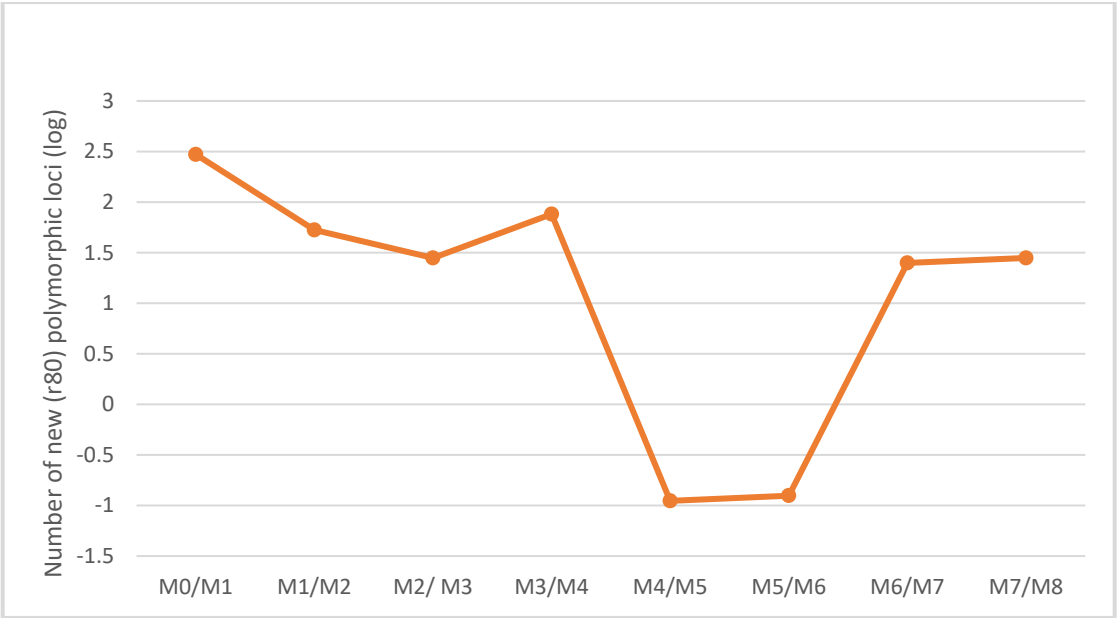

Figure A3 Log normalized (sign(X)*(LOG10(ABS(X)+1)) additional polymorphic loci per increase in the value of M. The number of new polymorphic loci plateaued at a value of M = 4.

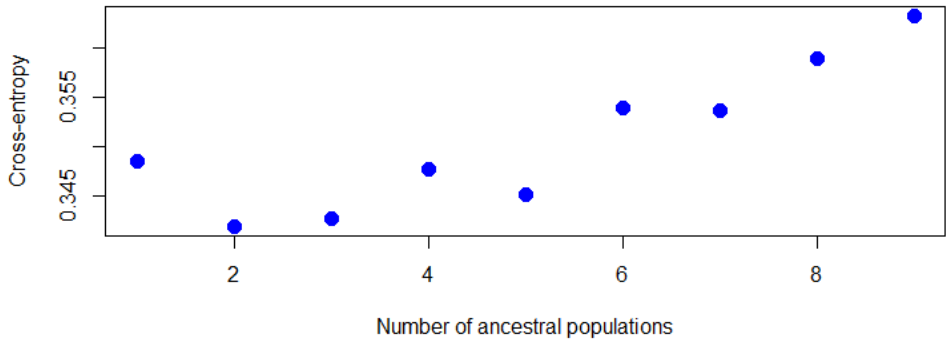

Figure A4 Cross-entropy criteria calculated to determine the optimal ancestry coefficient value for K = 1:9. Lower values indicate a more optimal value in terms of predictive capability. K = 2 is the optimal number of genetic clusters for the complete dataset.

Table A1 Methods used to acquire *B. undatum* coordinates and individuals for each site.

Site ID	Location explanation
GL	Individuals collected in a single trap by fishermen. Individuals and coordinates were provided by Michael Brown.
GP	Individuals collected by hand. Exact coordinates used.
PP	Individuals collected by hand. Exact coordinates used.
IR	Individuals collected in single trap on the vessel "Irish Rebel". Individuals and coordinates provided by Louisburg Fishery.
MD	Individuals collected in single trap on the vessel "Marie Dina". Individuals and coordinates provided by Louisburg Fishery.
MI	Individuals were collected within the *Buccinum* management zone 15 (See Brulotte 2019). Exact coordinates of collection were unknown, the location with highest recent recorded fishing effort for *B. undatum* were used (https://waves-vagues.dfo-mpo.gc.ca/Library/40854723.pdf).
NL	Individuals were collected by Green Seafoods within the snow crab management area 3L8A (See Southern Avalon; https://www.dfo-mpo.gc.ca/fisheries-peches/ifmp-gmp/snow-crab-neige/2019/index-eng.html. Exact coordinates of collection were unknown, so the midpoint of the fishing region and mean depth was used.
GA	Individuals collected during a 30-minute scientific bottom trawl. Samples and start and ending coordinates were and provided by Jamie Emberley, Department of Fisheries and Oceans. The midpoint of the trawl was used as the sampling location.
GB	Individuals collected during two 30-minute scientific bottom trawls. Samples and start and ending coordinates were and provided by Jamie Emberley, Department of Fisheries and Oceans. The trawls took place on the same day and at their farthest extent were 4.42 km apart. The midpoint of the combined trawl extent was used as the sampling location.

Curriculum Vitae

Candidate's full name: William Henry Sturch

Universities attended: University of Toronto, Honors Bachelor of Science, 2020.

Publications: NA

Conference Presentations:

Sturch, W., D'Aloia, C. 2021. Population genetic structure of a direct-developing marine gastropod. (Virtual Speaker). Canadian Society for Ecology and Evolution, Vancouver, Canada.