POPULATION GENOMICS AND ENVIRONMENTAL ADAPTATION IN THE SEA SCALLOP, *PLACOPECTEN MAGELLANICUS*, DETECTED USING RAD-SEQ DERIVED SNPS AND EXPERIMENTAL LARVAL REARING

by

© Mallory Van Wyngaarden

A thesis submitted to the

School of Graduate Studies

In partial fulfillment of the requirements for the degree of

Master of Science

Department of Biology

Memorial University of Newfoundland

January 2017

St. John's, Newfoundland and Labrador

Abstract

Understanding the scale of connectivity and adaptation among marine populations can inform fisheries conservation and management. We used a combination of advanced genomic techniques and experimental methods to determine the scale of connectivity and adaptation in the sea scallop, *Placopecten magellanicus*. Restriction-site Associated DNA sequencing genotyped 7163 SNPs in 245 individuals across 12 populations in the Northwest Atlantic. Subsequent analysis of these data identified a strong separation between populations north and south of Nova Scotia and identified an association between population structure and the coldest temperatures experienced by scallop populations. Common garden experiments on a northern and southern populations found that larvae from the north grew more quickly overall, potentially an adaptive strategy to the northern winter. These observations contribute to growing evidence of fine-scale population structure and adaptation in marine systems and support the hypothesis that a combination of limited dispersal and adaptive differentiation drives sea scallop population structure.

Acknowledgements

First and foremost, I would like to thank my supervisors, Drs. Ian Bradbury and Paul Snelgrove. Both of them provided extremely valuable guidance, advice, and support during my entire degree. No matter how busy, Paul always made time to provide edits and advice and my writing in particular is better for it. Ian tolerated my rapidly shifting deadlines with impressive patience and always had time for my questions. I would also like to thank the other member of my advisory committee, Dr. Claudio DiBacco. My larval rearing experiments would not have been possible without Claudio's valuable help and expertise, and his help and guidance helped me approach all parts of my thesis from a varied perspective.

I would also like to thank the members of all the labs I had the pleasure to work with. In the Bradbury Lab, Gwyn Mason in particular, but also Marion Sinclair-Waters and more recently post-docs Ryan Stanley, Brendan Wringe, and Nick Jeffery were helpful sounding boards for ideas and helped me work through challenges. In the Snelgrove lab, Christine Vickers, Rénald Belley, and Neus Campanyà Llovet (and Ryan Stanley again) were always there to help with advice on statistics, larval samples, and grad school in general. In the DiBacco lab, Jennifer Fitzgerald always made time to give me a hand when I needed it, and did so much to help me navigate lab work, experiments, and acquiring samples. Other staff at DFO in Dartmouth I would like to thank include Barry MacDonald (and his excellent larval rearing training) and Allan Reeves and Amy Glass (who helped with scallop collection and aging). There were also several scallop

harvesters who provided samples for our work, and without them this entire thesis would not have been possible.

Although I split my time between NL and NS for this thesis, a lot of staff at the OSC in St. John's made all my travelling impressively painless. JoAnn Greening in particular made it easy to manage all the paperwork that came along with attending conferences and research trips. Danielle Nichols and Margaret Shears both worked so hard to make the OSC a welcoming and friendly place, and I'm glad I had the chance to know them and work with them. I would of course like to thank the Friend Farm that developed from our cohort at MUN for always being helpful, supportive, and providing countless opportunities to help us all rescue our sanity.

Finally, thank you to my family for all their long-distance support over the past two years, and the never-ending support of my childhood dream of being a marine biologist.

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants to Drs. Paul Snelgrove and Ian Bradbury and the Genomic Resource Development Initiative from Fisheries and Oceans Canada. Travel to conferences, meetings, and workshops was provided by the Memorial University of Newfoundland (MUN) School of Graduate Studies, the MUN Department of Biology, the MUN Graduate Students' Union, and Genome Canada.

Table of Contents

Abstract ii
Acknowledgements iii
Table of Contents v
List of Tables viii
List of Figures xi
List of Common Terms, Abbreviations, and Symbols xv
List of Appendices xvii
Co-authorship Statement 1
Chapter 1: Introduction 2
1.1 Genomic data and studies of population connectivity 3
1.2 Environmental adaptation in the marine environment 4
1.3 Experimental evidence for local adaptation 5
1.4 The sea scallop, Placopecten magellanicus 5
1.5 Goals of this thesis 8
1.6 Thesis format 9
Chapter 2: Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus,* using RAD-seq derived SNPs 10
2.1 Abstract 10
2.2 Introduction 11
2.3 Methods 15
2.3.1 Sample collection 15
2.3.2 RAD-seq analysis 15
2.3.3 Summary statistics and outliers 17
2.3.4 Spatial structure 18
2.3.5 Estimates of dispersal and connectivity 19
2.4 Results 21
2.4.1 RAD-seq 21
2.4.2 Summary statistics, differentiation, and linkage 21
2.4.3 Spatial structure 22
2.4.4 Estimates of dispersal and connectivity 24
2.5 Discussion 25

2.5.1 RAD-seq and marine connectivity....26
2.5.2 Detection and influences of selection....27
2.5.3 Spatial population structure....29
2.5.4 Estimates of dispersal....33
2.5.5 Limitations....36
2.5.6 Conclusions....38
2.6 Tables....40
2.7 Figures....43
Chapter 3: Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*....51
3.1 Abstract....51
3.2 Introduction....52
3.3 Methods....54
3.3.1 Sample collection and RAD-seq....54
3.3.2 Environmental data collection and processing....55
3.3.3 Detection of outlier loci....57
3.3.4 Association between environmental factors and genetic variation....59
3.3.5 Gene ontology....62
3.4 Results....62
3.4.1 Sample collection and RAD-seq....62
3.4.2 Detection of outlier loci....62
3.4.3 Association between environmental factors and genetic variation....63
3.4.4 Gene Ontology....67
3.5 Discussion....68
3.5.1 Environmental variables driving adaptation....68
3.5.2 Mechanisms of adaptation....73
3.5.3 Challenges and limitations....75
3.5.4 Conclusions....78
3.6 Tables....80
3.7 Figures....86
Chapter 4: Countergradient variation in larval growth may be an adaptive strategy for sea scallops, *Placopecten magellanicus*, from northern cold-water populations....93
4.1 Abstract....93

4.2 Introduction 93
4.3 Methods 97
4.3.1 Study species..97
4.3.2 Experimental conditions 97
4.3.3 Adult scallop collection and conditioning 98
4.3.4 Spawning 100
4.3.5 Larval rearing..101
4.3.6 Larval sample analysis..102
4.4 Results 103
4.4.1 Growth 103
4.4.2 Survival..104
4.5 Discussion..105
4.5.1 Adaptation in marine environments..106
4.5.2 Thermal adaptation in the sea scallop..107
4.5.3 Limitations and implications 109
4.5.4 Conclusions..111
4.6 Tables 113
4.7 Figures 116
Chapter 5: Summary 122
5.1 Population Structure 122
5.2 Environmental adaptation..123
5.3 Larval evidence of adaptation..124
5.4 Implications 124
5.5 Conclusions..125
Bibliography 127
Appendices..141
Appendix S1: Supporting Information for Chapter 2: Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus,* using RAD-seq derived SNPs..141
Appendix S2: Supporting Information for Chapter 3: Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus* 160

List of Tables

Table 2.1. Site name, site code, coordinates, and the number of sequenced *P. magellanicus* from each of 12 collection sites in the Northwest Atlantic Ocean.

Table 2.2. Number of *P. magellanicus* individuals and number of SNP loci included in initial sequencing and final analysis following quality control (QC).

Table 2.3. Analysis of molecular variance (AMOVA) among 12 populations of *P. magellanicus*, among regional groups of populations identified by Structure analysis, and among individuals within populations using (A) all loci, (B) neutral loci, and (C) outlier loci.

Table 3.1. Site name, site code, coordinates, and the number of sequenced *P. magellanicus* from each of 12 collection sites in the Northwest Atlantic Ocean.

Table 3.2. Depth and surface ranges and bounding box coordinates used to select environmental data from databases for each of 12 *P. magellanicus* collection sites in the Northwest Atlantic Ocean.

Table 3.3. Data included in all multiple linear regression models used to determine the direction and magnitude of the effect of environmental variables on genetic variation among 12 populations of *P. magellanicus*.

Table 3.4. (A) Matrix of the number of outlier loci detected in *P. magellanicus* out of 7163 total loci by the methods BayEnv2 and LFMM using two environmental datasets, AllEnv and CST. The number of loci shared between different environmental datasets and programs are italicized. (B) Combined total number of loci detected from two methods, BayEnv2 and LFMM, using two environmental datasets, AllEnv and CST. The number of loci shared between the different environmental datasets is italicized.

Table 3.5. Cumulative Akaike Information Criterion model weights (Σ ωi) and model-averaged parameter estimates (full: variables were assumed to be present in all models but with a coefficient of 0 in some cases; subset: variables were only present in models where the coefficient was not 0) in models predicting whether genetic variation in outlier loci among populations of *P. magellanicus* is a function of environmental variation (A) Outlier loci were detected through correlations with an environmental dataset of 90 variables (AllEnv, n = 621 loci). Environmental were selected following the results of principal components analysis and redundancy analysis. (B) Outlier loci were detected through correlations with an environmental dataset of 36 variables (CST, n = 285 loci). Environmental variables were selected following the results of principal components analysis and redundancy analysis. (C) Outlier loci were detected through correlations with an environmental dataset of 90 variables (AllEnv, n = 621 loci). Environmental variables were selected following the results from (A) and (B). (D) Outlier loci were detected through correlations with an environmental

dataset containing a subset of all available environmental variables (CST, 36 variables, n=285 loci). Environmental were selected following the results from (A) and (B).

Table 4.1. Minimum, maximum, and average temperature, salinity, and pH experienced by *Placopecten magellanicus* larvae from two populations during a 21-day common garden larval growth experiment (average ± standard deviation in parentheses).

Table 4.2. ANOVA table for a linear model examining the effects of temperature, population, and their interaction on the Log10Growth of larval *P. magellanicus* from two populations.

Table 4.3. ANOVA table for a linear model examining the effects of temperature, population, and their interaction on the survival of larval *P. magellanicus* from two populations.

Table S1.1. (A) Parameter values for 8 sets tested using the program *stacks* (Catchen *et al.* 2011) (B) Initial catalog tags, catalog tags and SNPs following initial individual filtering, and catalog tags, SNPs, number of individuals, and number of populations included following detailed filtering for missing data and minor allele frequency for each parameter set tested. The final parameter map used was 4B.

Table S1.2. Ten datasets used in population structure analysis of 12 populations of *P. magellanicus* collected from the Northwest Atlantic.

Table S1.3. Comparison of outlier SNP loci from 12 populations of *P. magellanicus* determined using BAYESCAN (Bayesian method, 112 loci) and the 99% (157 loci) and 95% (184 loci) confidence intervals in ARLEQUIN (island model).
* Loci not present in ARLEQUIN 99%
¤ Loci not present in ARLEQUIN 95%

Table S1.4. Pairwise population-specific F_{ST} (above diagonal) and p-value (below diagonal) for (A) all loci, (B) neutral loci, and (C) outlier loci. Maximum F_{ST} values are highlighted in bold, minimum values are highlighted in bold and italicized. Significant p-values are italicized.

Table S1.5. Optimal number of genetic clusters (*K*) in 12 populations of *P. magellanicus* found by several analytical methods using 10 datasets listed in Table 2.3. K clusters indicated by * were suggested by the analysis, but no clear clustering could be observed.

Table S1.6. Estimates of the standard deviation of parent-offspring dispersal distance (in km) of *P. magellanicus* in (A) all populations, (B) north populations, and (C) south populations using all loci, neutral loci, and outlier loci, and low and high estimates of adult density calculated using the slope of the isolation-by-distance (IBD) relationship between pairwise population F_{ST} and pairwise population current-based distances and shortest ocean-based distances (*marmap*).

Significance of the IBD relationship is indicated as follows: p < 0.05 = *, p < 0.01 = **, p < 0.001 = ***.

Table S2.1. Number of loci and minor allele frequency (MAF) range included in each of five bins (based on global minor allele frequency) used by the program BayEnv2 to detect loci potentially under selection among 12 populations of *P. magellanicus*.

Table S2.2. Number of *P. magellanicus* individuals and number of SNP loci included in initial RAD-sequencing and final analysis following quality control (QC).

Table S2.3. Comparison of outlier SNP loci from 12 populations of *P. magellanicus* determined using BayEnv2 and a dataset of 90 environmental variables (AllEnv, 128 loci) and a subset of 36 environmental variables (CST, 72 loci). Shared loci are italicized.

Table S2.4. Comparison of outlier SNP loci from 12 populations of *P. magellanicus* determined using latent factor mixed models and a dataset of 90 environmental variables (AllEnv, 511 loci) and a subset of 36 environmental variables (CST, 218 loci). Shared loci are italicized.

Table S2.5. Blast2GO functional annotation of outlier SNP loci from 12 populations of *P. magellanicus*. Outliers were detected through environmental correlations with 90 environmental variables (AllEnv, 621 loci) or a subset of 26 environmental variables (CST, 285 loci).

List of Figures

Figure 2.1. Map of 12 sea scallop (*P. magellanicus*) collection locations from the Northwest Atlantic. Site MDA (Mid-Atlantic Bight) represents the middle of several nearby collection locations grouped as one population. Population codes are defined in Table 2.1.

Figure 2.2. Results from (A) the Bayesian test for selection completed using the program BayeScan and (B) the hierarchical island model test for selection completed using the program Arlequin for 7163 loci sequenced in 12 populations of *P. magellanicus*. BayeScan outliers are defined as all loci with a q-value higher than 0.05 (highlighted in red). Arlequin outliers are defined as the loci that fall above the simulated 1% quantile of F_{ST} vs Heterozygosity ($p <= 0.01$, highlighted in red).

Figure 2.3. Map of the proportion of each of the 12 *P. magellanicus* populations assigned to two population groups (blue and red) identified in the program Structure using outlier loci and the ΔK method to select the optimal number of genetic clusters in the data.

Figure 2.4. Plots of individual admixture for 12 populations of *P. magellanicus* at $K = 2$ determined using the program Structure and the ΔK method to select the optimal number of genetic clusters in the data. Results are presented for all populations at (A) all loci, (B) neutral loci, and (C) outlier loci as well as four north populations at (D) all loci, (E) neutral loci, and (F) outlier loci.

Figure 2.5. Principal components analysis plots for (A) all loci, (B) neutral loci, and (C) outlier loci in 12 populations of *P. magellanicus*.

Figure 2.6. Neighbour joining trees for Cavalli–Sforza and Edwards chord distance (D_c) between 12 populations of *P. magellanicus* for (A) neutral loci and (B) outlier loci. North populations are highlighted in grey, south populations in white, and bootstrap values greater than 50% are shown.

Figure 2.7. Isolation by distance plot of $F_{ST}/1\text{-}F_{ST}$ vs. population pairwise distance for 12 populations of *P. magellanicus* using (A) approximate current based distance and (B) least-cost distance for all loci (red squares), and outlier loci (blue triangles).

Figure 2.8. (A) Heat map of population specific standardized allele frequencies for 48 clinal outlier loci in 12 populations of *P. magellanicus*. (B) Plot of clines in allele frequency in 12 populations of *P. magellanicus* as a function of the distance in kilometers from the furthest north population (SUN) for clinal neutral loci ($n = 123$, 24.6% of tested loci, grey) and clinal outlier loci ($n = 48$, 42.9% of tested loci, black).

Figure 3.1. Map of sea scallop collection locations from the Northwest Atlantic. Site MDA (Mid-Atlantic Bight) represents the middle of several nearby collection locations grouped as one population. Population codes are defined in Table 3.1.

Figure 3.2. Principal components analysis plots for loci detected as potentially under selection through environmental correlation with (A) AllEnv (90 environmental variables, n= 621 loci), (B) CST (36 environmental variables, n = 285 loci) in 12 populations of *P. magellanicus*.

Figure 3.3. Principal components analysis plots for environmental variables used in the detection of adaptation among 12 populations of *P. magellanicus*. Environmental variables were separated into two datasets, (A) AllEnv (90 environmental variables) and (B) CST (36 environmental variables).

Figure 3.4. Linear regressions between the first principal component of PCA on population specific allele frequencies (Genetic PC1) and population specific environmental parameter values (Environmental PC1) for 12 populations of *P. magellanicus* for (A) AllEnv (90 environmental variables, 621 loci), and (B) CST (36 environmental variables, 285 loci).

Figure 3.5. Heat map of (A) standardized allele frequencies (AllEnvOutlier, 621 loci), (B) standardized environmental variable value (AllEnv, 90 variables), (C) standardized allele frequencies (CSTOutlier, 285 loci), and (D) standardized environmental variable value (CST, 36 variables) for 12 populations of *P. magellanicus*. Loci in (A) were selected as potentially under selection through correlation with environmental variables in (B). Loci in (C) were selected as potentially under selection through correlation with environmental variables in (D). SNPs in (A) and (C) are arrange in order of strongest to weakest differentiation pattern.

Figure 3.6. Redundancy analysis plots for loci detected as potentially under selection through environmental correlation with (A) AllEnv (90 environmental variables, n= 621 loci), (B) CST (36 environmental variables, n = 285 loci) in 12 populations of *P. magellanicus*. Explanatory variables were principal components axes from PCA on (A) AllEnv and (B) CST, retained following backwards stepwise variable selection.

Figure 3.7. Proportion of importance (average per variable importance / importance sum of all variables) for the top 15 environmental variables determined using Random Forest and (A) AllEnv and AllEnvOutlier, and (B) CST and CSTOutlier.

Figure 4.1. Map of collection locations for two populations of *P. magellanicus* used in the common-garden larval rearing experiment.

Figure 4.2. (A) Temperature, (B) salinity, and (C) pH experienced by adult *P. magellanicus* during conditioning. Horizontal lines indicate induced spawning periods for each population.

Figure 4.3. Example of typical D-stage scallop larval image indicating appropriate scale and shell height measurement.

Figure 4.4. (A, B, C) Temperature, (D, E, F) salinity, and (G, H, I) pH experienced by larvae in each temperature trial (10 °C – A, D, E; 13°C – B, E, H; 16 °C – C, F, I) over the duration of the three-week experiment. Dashed horizontal lines indicate the mean value over the duration of the trial. Solid horizontal line indicates duration of the experiment for each population; the timing of spawning for individual families within each population results in difference in length between NTS and BOF. Gaps are due to missing data.

Figure 4.5. (A) Average log growth ± SD of NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C). (B) Average survival ± SD of NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C).

Figure 4.6. (A) Log growth for individual families within NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C). (B) Survival for individual families within NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C).

Figure S1.1. Histogram of minor allele frequencies, observed heterozygosity, F_{ST} for 7163 SNP loci sequenced in 245 adult *P. magellanicus*.

Figure S1.2. Heatmap of pairwise linkage disequilibrium r^2 values calculated for 112 outlier loci detected among 12 populations of *P. magellanicus*.

Figure S1.3. Delta *K* (solid line) and Ln(*K*) determined by Bayesian clustering implemented in the program Structure for 12 populations of *P. magellanicus* using (A) all loci, (B) neutral loci, and (C) outlier loci for *K* = 1-15.

Figure S1.4. Delta *K* (solid line) and Ln(*K*) determined by Bayesian clustering implemented in the program Structure for 4 north populations of *P. magellanicus* using (A) all loci, (B) neutral loci, and (C) outlier loci for *K* = 1-5.

Figure S1.5. Cumulative variance explained by the principal components of a principal components analysis of 12 populations of *P. magellanicus* using (A) all loci, (B) neutral loci, and (C) outlier loci.

Figure S1.6. Bayesian Information Criterion values generated from *k*-means clustering of principal components of 12 populations of *P. magellanicus* for (A) all loci, (B) neutral loci, and (C) outlier loci.

Figure S1.7. Plot showing the individual-specific membership of 12 *P. magellanicus* populations in each of 4 genetic clusters identified by *k*-means clustering on the principal components analysis of the outlier loci. Square size corresponds to the number of individuals from each population assigned to a particular cluster.

Figure S2.1. Partial Redundancy analysis plots for loci detected as potentially under selection through environmental correlation with (A) AllEnv (90 environmental variables, n= 621 loci), (B) CST (36 environmental variables, n = 285 loci) in 12 populations of *P. magellanicus*. Explanatory variables used were principal components axes from PCA on (A) AllEnv and (B) CST, retained following backwards stepwise variable selection. The genetic matrix was conditioned on the distance between populations to reduce the effects of geographic separation between populations.

List of Common Terms, Abbreviations, and Symbols

Term	Description
AFLP	Amplified fragment length polymorphism
BOF	Sample site: Bay of Fundy
Chla	Chlorophyll a concentration, measurement of marine primary productivity
DFO	Fisheries and Oceans Canada
F_{ST}	Fixation index
GEO	Sample site: George’s Bank
GMI	Sample site: Gulf of Maine, Inshore
GMO	Sample site: Gulf of Maine, Offshore
IBD	Isolation-by-distance
LTB	Sample site: Little Bay, NL
MAF	Minor allele frequency
MDA	Sample site: Mid-Atlantic Bight
MGD	Sample site: Magdalen Islands
MUN	Memorial University of Newfoundland
Neutral	Locus putatively not under selection
NL	Newfoundland, Canada
NOAA	National Oceanic and Atmospheric Administration
NS	Nova Scotia, Canada
NSERC	Natural Sciences and Engineering Research Council of Canada
NTS	Sample site: Northumberland Strait
Outlier	Locus putatively under selection

PC	Principal component
PCA	Principal component analysis
PSB	Sample site: Passamaquoddy Bay, New Brunswick
RAD-seq	Restriction-site associated DNA sequencing
RDA	Redundancy analysis
Sea scallop	*Placopecten magellanicus* (Gmelin)
SNP	Single nucleotide polymorphism
SSB	Sample site: Scotian Shelf, Brown's Bank
SSM	Sample site: Scotian Shelf, Middle
SUN	Sample site: Sunnyside, NL

List of Appendices

Appendix S1: Supporting Information for Chapter 2: Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus,* using RAD-seq derived SNPs

Appendix S2: Supporting Information for Chapter 3: Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*

Co-authorship Statement

This work was completed and this thesis was written by Mallory Van Wyngaarden with help and guidance from Drs. Ian Bradbury, Paul Snelgrove, and Claudio DiBacco. RAD-seq sample collection was organized by Ian Bradbury and Claudio DiBacco and sequencing and bioinformatic processing were completed by Drs. Lorraine C. Hamilton and Naiara Rodríguez-Ezpeleta. Dr. Ryan RE Stanley assisted with figure preparation for Chapter 2, Dr. Robert Beiko and Luyao Zhan completed the random-forest analysis in Chapter 3, and Barry MacDonald assisted with experimental design and training for Chapter 4.

Chapter 2 was published in the journal Evolutionary Applications as: Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus*, using RADseq derived SNPs, authored by: Mallory Van Wyngaarden, Paul V. R. Snelgrove, Claudio DiBacco, Lorraine C. Hamilton, Naiara Rodríguez-Ezpeleta, Nicholas W. Jeffery, Ryan R.E. Stanley, and Ian R. Bradbury.

Chapter 3 has been submitted to the journal Molecular Ecology as: Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*, authored by: Mallory Van Wyngaarden, Paul V. R. Snelgrove, Claudio DiBacco, Lorraine C. Hamilton, Naiara Rodríguez-Ezpeleta, Luyao Zhan, Robert G. Beiko, and Ian R. Bradbury.

Chapter 1: Introduction

In any environment, an accurate understanding of population connectivity, dispersal, and adaptation can contribute significantly to successful species management and can help managers and harvesters ensure the continuation of sustainable management practices (Allendorf *et al.* 2010; Conover *et al.* 2006; Sale *et al.* 2005). The degree of connectivity between populations, influenced mainly by dispersal and the differential survival of dispersers resulting from adaptation, can affect population persistence, productivity, and response to exploitation (Cowen *et al.* 2006; Gaines *et al.* 2003; Hastings & Botsford 2006; Hellberg *et al.* 2002; Lowe & Allendorf 2010; Palumbi 2003; Waples 1998). In terrestrial systems, connectivity quantification methods include individual tracking and mark-recapture, but factors unique to the marine environment and to marine organisms complicate the measurement of connectivity. Many marine organisms, particularly sessile, benthic invertebrates, reproduce via broadcast spawning, releasing millions of larvae into the water column to disperse for varying periods of time before settling as juveniles. The high dispersal potential (Cowen & Sponaugle 2009; Hauser & Carvalho 2008; Neilsen & Kenchington 2001) that characterizes many marine organisms was thought to often produce a mixed pool of larvae that could easily migrate among different populations (Thorrold *et al.* 2002). These factors, along with a presumed lack of barriers to dispersal within the oceans, contributed to past assumptions of limited marine population structure and high levels of connectivity among all populations (Cowen *et al.* 2000).

Over the last few decades, however, accumulating phenotypic and genotypic evidence suggests that limited dispersal and low connectivity drive fine-scale population structure that may be more common than previously expected in marine environments (Hauser & Carvalho 2008; Hellberg 2009), potentially challenging current management paradigms in many exploited marine species. The scales at which this population structure exists may differ depending on the focal species or system. Studies suggest dispersal distances as low as 10 km per generation in some fish species (Buonaccorsi *et al.* 2004, Hauser & Carvalho 2008) although estimates can range widely. In some regions, dispersal and population structure on the scale of 1000s of kilometers is sufficient to contribute to significant genetic differentiation. Whether classified as large scale or fine scale, the most important conclusion from recent studies on marine population structure is that marine populations do not exhibit complete panmixia as previously assumed (Hauser & Carvalho 2008, Conover *et al.* 2006).

1.1 Genomic data and studies of population connectivity

In recent years, researchers have capitalized on improvements in technologies and methodologies for studying marine population structure and differentiation to enhance our understanding of marine population dynamics (Hauser & Carvalho 2008). Tracking the movement of individuals or groups of organisms in ocean systems is both challenging and time consuming, but key developments in genetic and genomic technologies have yielded much of the emerging evidence of limited dispersal and connectivity in marine species (Benestan *et al.* 2015; Bradbury & Bentzen 2007; Catchen *et al.* 2013; Hedgecock *et al.* 2007; Kinlan & Gaines 2003; Milano *et al.* 2014; Reitzel *et al.* 2013; Sotka & Palumbi

2006). These technologies have also enabled the identification of barriers to dispersal in the marine environment, whether physical or environmental.

High-throughput, next generation sequencing techniques in particular have dramatically altered marine population genomic inferences. They allow (relatively) inexpensive sequencing of loci across the entire genome of an organism in both coding and non-coding genomic regions. These techniques generate 1000s of markers for use in population and landscape genetics studies, even in species with limited existing data resources that made previous population genetic studies impossible. Past studies have demonstrated the effectiveness of one technique in particular, Restriction-site Associated DNA sequencing (RAD-seq) (Baird *et al.* 2008; Miller *et al.* 2007), in characterizing genetic diversity and differentiating marine species from fishes (Catchen *et al.* 2013; Hohenlohe *et al.* 2010) to invertebrates (Benestan *et al.* 2015; Reitzel *et al.* 2013).

1.2 Environmental adaptation in the marine environment

Another advantage of genomic studies (especially those using RAD-seq), apart from the generation of 1000s of markers in species with limited resources, is the detection of loci that may be under selection. This capacity allows researchers to target areas in the genome potentially associated with adaptive variation (Allendorf *et al.* 2010; Bradbury *et al.* 2010; Gagnaire *et al.* 2015; Hauser & Carvalho 2008; Jones *et al.* 2007). Large geographic ranges in some marine species expose different populations to multiple levels of environmental heterogeneity. This exposure, coupled with large population sizes, may lead to rapid divergence among populations and adaptation to particular environments. Putatively adaptive genetic loci have demonstrated genetic differentiation in a variety of

marine species (Bradbury *et al.* 2010; De Wit & Palumbi 2013; Lamichhaney *et al.* 2012), and the combination of environmental and genetic variation in landscape genomics studies have identified significant associations between climate and genetic structure in a variety of marine and anadromous species (Berg *et al.* 2015; Bradbury *et al.* 2014b; Bradbury *et al.* 2010; Hecht *et al.* 2015; Limborg *et al.* 2012; Milano *et al.* 2014; Pespeni & Palumbi 2013). This association has allowed researchers to pinpoint potential mechanisms of selection among marine populations and may also help researchers predict how species may react to a changing ocean climate in the future.

1.3 Experimental evidence for local adaptation

Although genetic differentiation and environmental associations with population structure provide important lines of evidence when assessing population management and predicting future population changes, the inclusion of experimental evidence along with genetic evidence offers the best strategy for verifying adaptation among populations (Rellstab *et al.* 2015). In combination with results from studies identifying potentially important environmental stressors, common garden experiments can help to identify environmental selective forces among populations. By rearing larvae or juveniles from different populations under identical conditions, a reaction norm can be identified (the relationship between a phenotypic trait and an environmental pressure). Selective pressures on populations can lead to detectable differences in reaction norms and provide further evidence for local adaptation to environmental stressors (De Jong 2005).

1.4 The sea scallop, Placopecten magellanicus

The sea scallop, *Placopecten magellanicus*, a dioecious bivalve mollusc, inhabits benthic environments in the Northwest Atlantic Ocean. The scallop reproduces via broadcast spawning through a multi-year life span (Posgay 1957), releasing eggs and sperm into the water column for external fertilization where scallop larvae live as free-floating plankton until settling to the benthos and metamorphosing into juveniles. In general, spawning occurs during autumn with a smaller spring spawn in some populations (e.g. DuPaul *et al.* 1989; Giguere *et al.* 1994; Langton *et al.* 1987; Naidu 1970; Posgay & Norman 1958; Schmitzer *et al.* 1991). The particularly long pelagic larval period (approximately 30 days) in sea scallops creates potential for long-distance dispersal among scallop populations (Naidu & Robert 2006).

Sea scallops span a large geographic range, from Newfoundland, Canada in the north to Cape Hatteras, USA in the south (Posgay 1957). A large gradient in temperature (resulting from the meeting of the cold Labrador Current and the warm Gulf Stream) and heterogeneity in other environmental factors characterizes this region of the Atlantic Ocean, and storm-related mixing along the coast and oceanographic properties of the major currents may influence all of these features (Townsend *et al.* 2006) and contribute to local adaptation among scallop populations. Several oceanographic barriers along the range may also influence larval movement and survival between populations (Townsend *et al.* 2006).

The fishery for sea scallop, one of the more economically important marine species in both Canada and the USA, extends back over 100 years (DFO 2016; Naidu & Robert 2006; NOAA 2016). In 2014, landed value from the sea scallop fishery comprised

7.4% of the total landing value for all Atlantic coast fisheries in Canada (4th most valuable, approximately 177 million CAD) (DFO 2016) and 7.7% of the total landed value in the USA (NOAA 2016, approximately 424 million USD).

Despite the high potential for population interconnectivity among sea scallop populations, past studies reported phenotypic differences among sea scallop populations over fine to moderate spatial scales (< 1000 km). When examining populations from St. Pierre Bank, the Bay of Fundy, and Georges Bank, Kenchington & Full (1994) identified significant differences in shell morphometry between some scallop beds, even when controlling for age and year class effects. Reproductive timing also varies. Populations in western Newfoundland spawn briefly in early summer followed by a protracted fall spawn, in contrast to a much more sudden fall spawn on Georges Bank, hundreds of kilometers to the south (Naidu 1970). Differences were also reported in the incidence of hermaphroditism, with greater occurrence in the northern Newfoundland populations than on Georges Bank (Naidu 1970). On a smaller scale, a study of shallow- and deep-water populations in the Gulf of Maine reported differences in fecundity, with fewer eggs released in deeper populations (although egg sizes were similar to those in the shallow population) (Barber *et al.* 1998). A small spring spawn was also observed in the shallow population, similar to that reported in western Newfoundland (Barber *et al.* 1998). Evidence of different larval behaviours has also been reported (Manuel *et al.* 1996b). As with many planktonic larvae, scallops exhibit diel vertical migration, however experimental evidence shows that larvae from different populations (both inshore and offshore) behave differently, potentially as a result of selective pressure caused by

different oceanographic environments at spawning locations. All of these reported differences are hypothesized to result from different environmental conditions across the species' range. In addition to the documented physiological, morphological, and behavioural differences, genetic studies examining portions of the scallop range have also detected population differentiation and structuring. These studies have used both microsatellites (Kenchington *et al.* 2006) and AFLPs (Owen & Rawson 2013), however, no studies have used genomic sampling along the entire range of the species to quantify population structure, connectivity, and adaptation.

1.5 Goals of this thesis

The aim of this work was threefold. First, using RAD-seq derived SNPs, we wanted to use range-wide samples of sea scallops to detect any existing structure and differentiation among scallop populations, in particular comparing and contrasting neutral genetic structure with that of markers within the genome potentially under selection. This work also allowed us to estimate dispersal distances for larval scallops, providing information that may be useful to fisheries management of such an important cross-border species. The second objective built on the results of the first. Using a large environmental dataset with data from both DFO and NOAA, we aimed to use landscape genetic techniques to determine what (if any) environmental parameters may contribute to population structure within sea scallops, providing evidence for local adaptation within its geographic range. The final objective was to identify experimentally the presence of local adaptations between two sea scallop populations through a common-garden larval growth experiment. The results from this experiment could corroborate the results from the first

two objectives, providing valuable information to researchers attempting to ensure successful management of the species and predict future response of the species to a changing ocean environment.

1.6 Thesis format

Five chapters comprise this thesis, including this introduction (Chapter 1) and a conclusion chapter (Chapter 5). Chapters 2-4 were prepared in manuscript format in preparation for publication, resulting in some overlap among them. Chapter 2 has been submitted to the journal Evolutionary Applications, Chapter 3 has been submitted to Molecular Ecology, and Chapter 4 is in preparation for submission to an as yet undetermined journal.

Chapter 2: Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus,* using RAD-seq derived SNPs

2.1 Abstract

Understanding patterns of dispersal and connectivity among marine populations can directly inform fisheries conservation and management. Advances in high-throughput sequencing offer new opportunities for estimating marine connectivity. We used Restriction-site Associated DNA sequencing to examine genetic structure among populations of the sea scallop *Placopecten magellanicus*, an economically important marine bivalve, and used this information to infer the role of dispersal, adaptation, and realized connectivity in their population dynamics. Based on 245 individuals sampled range-wide at 12 locations from Newfoundland to the Mid-Atlantic Bight we identified and genotyped 7163 Single Nucleotide Polymorphisms; 112 (1.6%) were identified as outliers potentially under directional selection. Bayesian clustering revealed a discontinuity between northern and southern samples and latitudinal clines in allele frequencies were observed in 42.9% of the outlier loci and in 24.6% of neutral loci. Dispersal estimates derived using these clines and estimates of linkage disequilibrium imply limited dispersal; 373.1 ± 407.0 km (mean ± SD) for outlier loci and 641.0 ± 544.6 km (mean ± SD) for neutral loci. Our analysis suggests restricted dispersal compared to the species range (>2000 km) and that dispersal and effective connectivity (the survival and subsequent reproduction of dispersers) differ. These observations support the hypothesis that limitations in effective dispersal structure scallop populations along eastern North America

and can help refine the appropriate scale of management and conservation in this commercially valuable species.

2.2 Introduction

Successful species management and conservation require an accurate understanding of population connectivity, including interbreeding and dispersal among populations (Allendorf *et al.* 2010). The degree of connectivity among adjacent populations can affect population persistence, productivity, and response to exploitation (Cowen *et al.* 2006; Gaines *et al.* 2003; Hastings & Botsford 2006; Hellberg *et al.* 2002; Lowe & Allendorf 2010; Palumbi 2003; Waples 1998). In terrestrial systems, connectivity quantification methods include individual tracking and mark-recapture, but many factors unique to the marine environment and to marine organisms complicate the measurement of connectivity. In many cases, the large effective sizes of temperate marine populations prevent genetic drift from promoting differentiation over short to moderate time-scales (10s to 1000s of generations), limiting the accumulation of neutral genomic divergence (Hauser & Carvalho 2008). In addition, most marine invertebrates, especially sessile, benthic species, reproduce via broadcast spawning, releasing millions of larvae into the water column that disperse for weeks to months or more before settling as juveniles. High larval dispersal potential characterizes these types of organisms, (Cowen & Sponaugle 2009; Hauser & Carvalho 2008; Neilsen & Kenchington 2001) potentially producing a mixed pool of larvae from different populations and from adults of different age cohorts (Thorrold *et al.* 2002), contributing to the assumption of limited marine population structure (Cowen *et al.* 2000). However, over the last few decades accumulating phenotypic and genetic evidence

suggests limited dispersal and low connectivity drive fine-scale population structure that may be more common than previously expected in marine environments (Hauser & Carvalho 2008; Hellberg 2009) potentially challenging current management paradigms in many exploited marine species.

Advances in genetic and genomic techniques drive much of the emerging evidence of limited dispersal and connectivity in marine species (Benestan *et al.* 2015; Bradbury & Bentzen 2007; Catchen *et al.* 2013; Hedgecock *et al.* 2007; Kinlan & Gaines 2003; Milano *et al.* 2014; Reitzel *et al.* 2013; Sotka & Palumbi 2006). High-throughput, next generation sequencing techniques in particular have dramatically increased the number and type of genetic loci (coding and non-coding) available to study in marine species, especially non-model species. The ability to survey genome-wide diversity and target loci possibly associated with adaptive variation has proven particularly informative in large marine populations where directional selection may drive rapid divergence and differentiation (Allendorf *et al.* 2010; Bradbury *et al.* 2010; Gagnaire *et al.* 2015; Hauser & Carvalho 2008; Jones *et al.* 2007). Examination of outlier loci (those putatively under selection) consistently demonstrates small-scale genetic differentiation in a variety of marine taxa including *Haliotis rufescens* (red abalone) (De Wit & Palumbi 2013), *Clupea harengus (*Atlantic herring) (Lamichhaney *et al.* 2012), and *Gadus morhua* (Atlantic cod) (Bradbury *et al.* 2010). The separate examination of neutral and outlier loci can provide important insights into the causes of genetic differentiation among marine populations. Different hypotheses can be tested using neutral and outlier loci, and the comparisons of neutral and outlier population structure patterns allows researchers to differentiate between structural

patterns driven by dispersal limitation (visible in neutral loci) and selective pressure (often identified by outlier loci presumed to be under selection). In addition, because large population sizes common in marine species can prevent genetic drift and limit the accumulation of neutral divergence (Hauser & Carvalho 2008), highly differentiated outlier loci can be useful in identifying small-scale differentiation. Large populations may also show weak levels of differentiation even when significantly diverged (Allendorf *et al.* 2010), and this divergent structure is often more detectable in loci under selection. The advent of Restriction-site Associated DNA sequencing (RAD-seq), (Baird *et al.* 2008; Miller *et al.* 2007) now permits genome-wide scans for outlier loci in model and non-model organisms and increases the characterization of genetic diversity and differentiating marine species from fishes (Catchen *et al.* 2013; Hohenlohe *et al.* 2010) to invertebrates (Benestan *et al.* 2015; Reitzel *et al.* 2013).

Placopecten magellanicus (Gmelin) (sea scallop), a dioecious bivalve, inhabits benthic environments in the Northwest Atlantic Ocean from Newfoundland, Canada in the north to Cape Hatteras, North Carolina, USA in the south (Posgay 1957). Sea scallops typically occur along the continental shelf at depths from approximately 10-100 m but as deep as 384 m (Naidu & Robert 2006). The sea scallop fishery extends back over 100 years, and currently represents one of the most economically important fisheries in North America in landed value on the east coast of the United States and Canada (Naidu & Robert 2006), in 2014 comprising 7.4% of the total landing value for all Atlantic coast fisheries in Canada (4th most valuable fishery) (DFO 2016) and 7.7% of the total landed value in the United States (NOAA 2016). High fecundity, broadcast spawning, and a long planktonic larval

period (30-35 days) in sea scallops all contribute to long distance dispersal potential among populations (Naidu & Robert 2006). Despite this high potential for population interconnectivity, phenotypic differences among sea scallop populations over fine to moderate spatial scales have been observed including differences in reproductive timing (Naidu 1970), population-specific fecundity (Barber *et al.* 1988), shell morphometry (Kenchington & Full 1994), larval behaviour (Manuel *et al.* 1996b), and growth (Naidu & Robert 2006). Ultimately the scale of dispersal and connectivity in this species remains unresolved and this knowledge could directly inform fisheries management and conservation efforts.

The objective of this study was to investigate sea scallop spatial population structure in the Northwest Atlantic using RAD-seq derived Single Nucleotide Polymorphisms (SNPs). We tested the alternate hypothesis that previously unidentified range-scale population structure exists in the sea scallop, and that the combined use of genome-wide neutral and outlier markers would provide a more powerful tool to detect finer structure than previous studies.

The objectives were to: (1) describe the spatial population structure of sea scallop in the Northwest Atlantic using RAD-seq derived SNPs, (2) contrast the structure present at multiple spatial scales and with outlier and non-outlier loci, and (3) estimate average dispersal distances among populations using the isolation by distance (IBD) relationship and clines in allele frequency. This work builds directly on previous scallop studies using both microsatellites (Kenchington *et al.* 2006) and AFLPs (Owen & Rawson 2013) to explore population structure and oceanographic influences in this region as well as previous

work on Northwest Atlantic cod which reported latitudinal clines in allele frequency in outlier loci (Bradbury *et al.* 2014a; Bradbury *et al.* 2010; Bradbury *et al.* 2013).

2.3 Methods

2.3.1 Sample collection

We collected 252 adult scallops by hand (SCUBA diving) or bottom trawl (both from commercial vessels and research cruises) from a total of 12 locations across the entire range of the species between 2011 and 2013 (Table 2.1, Figure 2.1). This sampling scheme yielded a minimum of 12 scallops per population (mean value ± SD of 20.4 ± 2.8 scallops), although we lacked age data. Tissue samples were collected and preserved in AllProtect (Qiagen, Toronto, ON, Canada) or 80% ethanol. DNA extraction and RAD-seq library preparation were performed by the Aquatic Biotechnology Lab at the Bedford Institute of Oceanography in Halifax, Nova Scotia. DNA was isolated from the tissue samples using DNeasy Blood and Tissue kit or DNeasy 96 Blood and Tissue kit (Qiagen) following the manufacturer's protocol, including the optional RNase A treatment. All DNA samples were quantified using the Qubit dsDNA HS Assay Kit (Life Technologies, Burlington, ON, Canada) with assays read on a Qubit v2.0 (Life Technologies) or using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies) with assays read on a FLUOStar OPTIMA fluorescence plate reader (BMG Labtech, Ortenberg, Germany). The DNA quality for all samples was verified by agarose gel electrophoresis of 100 ng of extracted DNA, visualized using SYBR Safe (Life Technologies), and documented using a Gel Logic 200 (Kodak).

2.3.2 RAD-seq analysis

One µg of DNA was used per individual for library preparation and sequencing. RAD-seq libraries were prepared as described by Etter *et al.* (2011b) (see also Etter *et al.* 2011a) with modifications. DNA samples from 22 individuals from the same geographical location comprised each library (with the exception of the library for SUN which consisted of only 20 individuals) with a different in-line barcode in the P1 adapter for each individual sample. With the exception of SSB, GEO, and SUN the P1 adapter in-line barcodes were all 6bp in length. For the SSB, GEO, and SUN libraries the P1 adapter in-line barcodes all ranged from 5bp-9bp in length and were chosen to ensure equal distribution of all nucleotides at each base position (including those that overlap with the restriction site) and to maximize the edit distance (Faircloth & Glenn 2012). Based on edittags analysis (Faircloth & Glenn 2012), the variable length barcodes edit distance ranged from 2-8 with a modal edit distance of 6. Gel size selection performed after sonication and PCR amplification was done on a Pippin Prep (Sage Science, Beverly, MA, USA) using the 2% agarose gel cassette with ethidium bromide (Sage Science) and size selection range of 300-500bp. PCR amplification used Q5 Hot Start Master Mix (New England Biolabs, Whitby, ON, Canada) for all libraries. Amplification cycles for all libraries were 98 °C for 30 seconds; *x* cycles of 98 °C for 30 seconds, 65 °C for 30 seconds, 72 °C for 30 seconds; 1 cycle of 72 °C for 5 minutes, where *x* was 18 for all libraries except for SSB, GEO, and SUN where *x* was 13. All libraries were sequenced on a HiSeq 2000 (Illumina) as 100bp paired end sequences with one library per lane. Sequencing was performed at the McGill University and Génome Québec Innovation Centre, Montréal, Canada.

SNPs were detected using the *de novo* pipeline in STACKS v.0.9999 (Catchen *et al.* 2011). Putative orthologous loci were assembled using *ustacks* with a minimum depth of coverage required to create a stack (m) of five and four maximum nucleotide mismatches (M) allowed between stacks. The catalog of loci was assembled using *cstacks* with a distance allowed between loci in the catalog (n) of six (several other parameter combinations were tested; see Table S1.1a). Using the *populations* module, only RADtags present in at least 75% of individuals were kept. The final dataset was filtered using PLINK v.1.07 (Purcell 2009; Purcell *et al.* 2007) to include SNPs present in at least 75% of individuals with a minor allele frequency greater than 5%. Furthermore, we excluded individuals with more than 20% missing loci from the analysis. Loci were filtered for Hardy-Weinberg Equilibrium using the program GENEPOP v.4 (Rousset 2008), excluding loci out of equilibrium in 6 or more populations from the analysis (<0.7% of all loci).

2.3.3 Summary statistics and outliers

We calculated allele frequencies and heterozygosities using the R (R Development Core Team 2012) package *gstudio* (Dyer 2014) and calculated locus specific F_{ST} using the program ARLEQUIN v.3.5 (Excoffier & Lischer 2010). To calculate pairwise linkage disequilibrium [r^2 (Hill & Robertson 1968)] between all loci, outlier loci, and neutral loci separately, we used PLINK v.1.07. Although a variety of methods have been developed to detect loci potentially under selection within a group of populations, individual methods vary in their ability to detect outliers (Narum & Hess 2011). We used a Bayesian method (Beaumont & Balding 2004) implemented in the program BAYESCAN v.2.1 (Foll & Gaggiotti 2008) and an island model implemented in the program ARLEQUIN v.3.5 to

determine a candidate list of outlier loci. We ran BAYESCAN with a burn-in period of 50 000 followed by 100 000 iterations, subsequently identifying outliers in R with a false discovery rate of 0.05. In ARLEQUIN, we ran 100 000 permutations using 500 demes, 50 groups, and a maximum expected heterozygosity of 0.5; ARLEQUIN outliers were determined with a *p*-value of 0.01.

2.3.4 Spatial structure

We examined population structure along the range of sea scallops using multiple methods. Hierarchical iterative clustering analysis was conducted using STRUCTURE v.2.2.4 (Pritchard *et al.* 2000) through the R package *parallelStructure* (Besnier & Glover 2013). Results from preliminary BAYESCAN and STRUCTURE analyses guided subsequent analyses (see *Results*), after which we analysed nine separate datasets [all loci, neutral loci, and outlier loci for each major sample group separately (see *Spatial Structure*), (Table S1.2)] to determine major population groups as well as any minor clusters.

We used Bayesian clustering in STRUCTURE to determine the number of distinct genetic clusters (K) present among the 12 sampled populations, running calculations with a burn-in period of either 50 000 repetitions followed by 200 000 repetitions, or 100 000 repetitions followed by 500 000 repetitions, until algorithm convergence was confirmed. We repeated all runs 3 times for each K, running datasets 1-3 for K = 1-15, datasets 3-6 for K = 1-5, and datasets 7-9 for K = 1-10. In order to determine the optimal K for each dataset we used the delta K method (Evanno *et al.* 2005) and processed results using STRUCTURE HARVESTER (Earl & vonHoldt 2012); runs were grouped and visually displayed using CLUMPAK (Kopelman *et al.* 2015). We also completed an analysis of molecular variance

(AMOVA) using ARLEQUIN with 25 000 permutations, defining genetic structure following the results from the STRUCTURE analysis. We conducted principal components analysis (PCA) followed by *k*-means clustering using the R package *adegenet* (Jombart 2008). This method determines the optimal number of clusters (*k*) in the PCA using the Bayesian Information Criterion (BIC). The lowest value of the BIC across each value of *k* indicates the number of clusters present in the data. Finally, we constructed neighbour-joining trees using the programs POPULATIONS (Langella 1999) and TREEVIEW (Page 1996) based on estimates of genetic distance among populations (Cavalli-Sforza and Edwards chord distance, D_c) with 1000 bootstrap replications on individuals.

2.3.5 Estimates of dispersal and connectivity

We explored two general approaches to estimate average per generation dispersal distance, both of which make different assumptions regarding the underlying model of gene flow. First, we used an IBD model which assumes a linear 1-dimensional stepping stone for gene flow (see Bradbury & Bentzen 2007). This approach used linear regression between pairwise population $F_{ST}/(1-F_{ST})$ and spatial distances based on two measures of geographic distance: approximate ocean distances following prevailing currents estimated in GOOGLE EARTH (Google 2013) following average current patterns in the Northwest Atlantic and least-cost geographic distance calculated using the R package *marmap* (Pante & Simon-Bouhet 2013), where distance was calculated excluding positive elevation (land). We calculated IBD separately using all loci, outlier loci, and neutral loci and for all sampled populations and each major sample group separately (see *Spatial Structure*). We performed Mantel tests to ascertain the significance of every IBD relationship using the R package

ade4 (Dray & Dufour 2007) and adjusted p-values for multiple comparisons using the Bonferroni method with *p.adjust* in R. Adult-offspring dispersal distance estimates were calculated following Rousset (1997) using the slope of the IBD relationship. We estimated adult density values required for the IBD methods from Mason *et al.* (2014), DuPaul & Rudders (2008), and Kelly (2007) for several areas within the study range and used them as density proxies along the entire species range. Furthermore, because census estimates of density likely differ from effective density, we explored the sensitivity of the dispersal estimate to a range of density values several orders of magnitude above and below the actual estimates used.

The second approach employed a clinal model of gene flow following Bartonand Gale (1993), Lenormand *et al.* (1998), and Sotka & Palumbi (2006). Here, clines in allele frequency for outlier loci and a random subset of 500 neutral loci were estimated using the R package *HZAR* (Derryberry *et al.* 2014) using 100 000 iterations following a 10 000 iteration burn-in period. We used population-specific allele frequencies for all loci tested, and estimated distances from the furthest north population (SUN) along a 1-dimensional transect that included all populations using GOOGLE EARTH (Google 2013). Four cline models and a null model were generated for each locus, and cline model selection used AICc criteria followed by a log-likelihood cutoff of -10. Models tested included fixed or free minimum and maximum allele frequency values and either no exponential cline tails or tails at both ends of the cline. We determined cline width from the best fit model and used cline width in estimates of adult-offspring dispersal distance. Here, adult-offspring dispersal distance estimates followed Sotka & Palumbi (2006) using cline width and

linkage disequilibrium to determine the standard deviation in parent-offspring distance. Differences between cline width and dispersal estimates in neutral and outlier loci were assessed using the Welch two-sample t-test.

2.4 Results

2.4.1 RAD-seq

Following filtering and quality control steps, we included 245 individual scallop samples in our analysis (97.2% of sequenced individuals), 19672 RADtags (14.9% of initial RADtags), and 7216 SNPs (4.2% of Initial SNPs) (Table 2.2). Applying alternative parameters sets produced similar SNP numbers (Table S1.1b). The 7163 SNPs in HWE that met all quality control standards were used in all subsequent analyses.

2.4.2 Summary statistics, differentiation, and linkage

For the final dataset, minor allele frequency (MAF) averaged 0.1855 ± 0.1253 (mean ± SD), expected heterozygosity averaged 0.2710 ± 0.1333 (mean ± SD), and locus-specific F_{ST} averaged 0.0066 ± 0.0198 (mean ± SD) (Figure S1.1). Of the final 7163 SNPs, 112 SNPs (1.6%) were identified as outliers by BAYESCAN, leaving 7051 (98.4%) in the neutral data set. Both approaches of outlier analysis yielded similar results with 91.1% of BAYESCAN-identified loci present in the outlier list produced via ARLEQUIN with the 99% confidence interval, and 93.8% were present with the 95% confidence interval (Figure 2.2, Table S1.3). Because the outlier lists were very similar, we focused on the BAYESCAN list from all subsequent analysis of outliers. Population specific F_{ST} calculated using ARLEQUIN was higher for outlier loci than either neutral loci or all loci

(Table S1.4). Using all loci, pairwise F_{ST} averaged 0.005 ± 0.006 (mean ± SD), with a maximum value of 0.018 between LTB and GEO and a minimum value of -0.004 between GMI and MDA. Using neutral loci, pair wise F_{ST} averaged 0.003 ± 0.005 (mean ± SD), with a maximum value of 0.014 between LTB and MDA and a minimum value of -0.004 between NTS and BOF, SSB and GMI, and GMI and GEO. Using outlier loci, pair wise F_{ST} averaged 0.094 ± 0.070 (mean ± SD), with a maximum value of 0.253 between LTB and SSB and a minimum value of 0.003 between GMO and MDA. In all cases, pairs of populations containing one north and one south population (see *Spatial Structure*) yielded maximum values with the highest differentiation. Average pairwise r^2 values indicating linkage disequilibrium were higher in outlier loci than neutral loci but even the outlier values remained low overall (outlier loci: 0.0258 ± 0.0829, neutral loci: 0.0044 ± 0.0098, all loci: 0.0044 ± 0.0098, mean ± SD). Within the outlier loci, a few small pockets of higher linkage seemed to drive the higher average r^2 value (Figure S1.2).

2.4.3 Spatial structure

We explored different methods of determining population structure however they generally produced similar results (Table S1.5). Bayesian population structure analysis in STRUCTURE clearly split north and south groups using three datasets; *K*=2 was best supported for all loci, neutral loci, and outlier loci (Figure S1.3). The north group consisted of four samples from Newfoundland and the Gulf of St. Lawrence, whereas the south group contained the remaining eight samples from south of the Scotian Shelf (Figure 2.3, 4ABC). Further hierarchical structure analysis on the north group revealed a split into two sample groups, however, the pattern of structure differed among the outlier and neutral loci (*K*=2

in all cases, Figure S1.4). When using all loci and only neutral loci, LTB clustered separately from all other north populations (Figure 2.4DE). When using outlier loci, LTB and the SUN sample clustered together separate from the Gulf of St. Lawrence samples (MGD and NTS) (Figure 2.4F). Structure analysis of the eight south populations revealed no clear clustering or evidence of differences among samples (Figure 2.4ABC). AMOVA, to explore the amount of variation explained by this subdivision, showed that the split between north and south sample groups explained a small percentage of total genetic variance in all loci and neutral loci (all loci = 0.58%, neutral loci 0.40%). However, AMOVA results for outlier loci differed from the other datasets with 11.0% of all variation explained by the split between north and south sample groups (Table 2.3).

In addition to the STRUCTURE analysis, we used principal components analysis (PCA) and Neighbour-joining trees (NJ) to explore spatial relationships in two dimensions. PCA on all sets of loci split north and south samples along the first principal component, similar to the division in the STRUCTURE analysis (Figure 2.5). This first principal component (PC) explained 0.97%, 0.78%, and 12.91% of the total variance explained by the analysis using all loci, neutral loci, and outlier loci, respectively (Figure S1.5). PCA using all loci and neutral loci further separated LTB from the other north populations along the second principal component, but this pattern was not seen in the outlier loci (comparable to the STRUCTURE results). *K*-means clustering only identified one genetic group when using all loci and the neutral loci (k=1), but with the outlier loci detected the same north-south split seen in the STRUCTURE results as well as further structuring within the regional sample groups (k=4, Figure S1.6, Figure S1.7). Neighbour-joining trees showed

the same north-south split seen in other analysis, however, only the outlier loci supported the split (Figure 2.6) as indicated by a bootstrap support value of 84%.

2.4.4 Estimates of dispersal and connectivity

We examined IBD relationships using 18 different combinations of samples, loci, and population distance measures. These included: using all populations, north populations, or south populations; using all loci, neutral loci, or outlier loci; and using current based pairwise population distance or least-cost pairwise population distance. After Bonferroni correction for multiple comparisons, only two of these 18 IBD relationships were significant. When using the least-cost pairwise population distance, we found a significant IBD relationship when using all 12 populations and all loci ($R^2 = 0.2609$, $p = 0.018$) or the outlier loci ($R^2 = 0.3363$, $p = 0.018$) (Figure 2.7). However, as the spatial analysis above clearly indicated the presence of two dominant clusters driving the IBD relationship, this pattern was not consistent with a one dimensional stepping stone framework assumed by the Malécot's lattice model (Malécot 1955) and the approach outlined by Rousset (1997) for estimating dispersal distance. As a result, dispersal estimates from IBD analysis were consistently unrealistically small (< 15km, Table S1.6) and not considered further.

We also used evidence of clinal trends in allele frequency to estimate average per generation dispersal distance. Of the 112 outlier loci tested, 48 (42.9%) showed significant clines (non-null model and log-likelihood > -10) (Figure 2.8A). Of the 48 clinal loci, 16 (33.3%) had fixed scaling and no exponential tails in allele frequencies and the remaining 32 (66.7%) had free scaling and no exponential tails. Average outlier cline width was 1157.0 ± 1268.6 (mean ± SD) km, with a minimum cline width of 14.4 km and a maximum

cline width of 4524.1 km. We next examined a randomly selected subset of 500 neutral loci for clinal patterns (7.09% of total neutral loci). Of these 500, 377 (75.4%) showed no cline in allele frequencies. Of the 123 loci showing clinal patterns, 68 (55.3%) had fixed scaling and no exponential tails, and 55 (44.7%) had free scaling and no exponential tails. Within these 123 loci (24.6% of the 500 tested), average cline width exceeded the outlier loci, at 2523.8 ± 2144.1 km (mean ± SD), with a minimum cline width of 3.4 km and a maximum cline width of 4529.8 km. Cline widths were found to be significantly different between outlier and neutral loci ($p < 0.001$) (Figure 2.8B). The estimated standard deviation of parent-offspring distance when using clines from outlier loci was 373.1 ± 407.0 km (mean ± SD). The estimated standard deviation of parent-offspring distance when using clines from neutral loci was higher than the outlier loci estimate, at 641.0 ± 544.6 km (mean ± SD). Both estimates are significantly lower than the maximum pairwise distance between our sample sites and are significantly different from one another ($p = 0.0007$).

2.5 Discussion

Successful management and conservation of exploited and threatened species requires an accurate understanding of population connectivity and dispersal patterns among populations and habitats (Allendorf *et al.* 2010). In marine species, estimates of dispersal and connectivity remain rare largely due to the difficulty in tracking relatively small pelagic larval stages to settlement (Bradbury *et al.* 2008c). Here we used RAD-seq derived SNPs to explore spatial patterns of connectivity and estimate dispersal in a commercially exploited marine bivalve, *P. magellanicus*. Our results show significant population differentiation and structure across the range of *P. magellanicus* despite high dispersal

potential during a pelagic larval stage. Our estimates of dispersal indicate geographically restricted connectivity, particularly when using outlier loci, suggesting a role for selection in determining realized connectivity and limiting gene flow. This work suggests significant cryptic intraspecific diversity in this species. Accurate knowledge of sources of larvae and dispersal patterns such as those revealed here can significantly influence population persistence into the future (Hastings & Botsford 2006); management strategies that incorporate results from studies mapping population structure and dispersal patterns may be among the most effective (Fogarty & Botsford 2007), and the use of genomic tools such as those used here can directly facilitate successful conservation and fisheries management (e.g. Miller *et al.* 2014; da Silva *et al.* 2015).

2.5.1 RAD-seq and marine connectivity

The use of RAD-seq in marine species has provided unprecedented access to measures of genome wide variation with obvious applications for marine management and conservation. SNPs generated using RAD-seq techniques have been used to identify historical phylogeography and phylogenetics (Herrera *et al.* 2015), study evolution and adaptation among and within species (Hohenlohe *et al.* 2010), provide recommendations for species conservation (Gruenthal *et al.* 2014; Hohenlohe *et al.* 2013; Ogden *et al.* 2013), generate genetic resources for future studies (Gonen *et al.* 2014; Kruck *et al.* 2013; Pujolar *et al.* 2013), and resolve contemporary population structure (Benestan *et al.* 2015; Catchen *et al.* 2013; Chu *et al.* 2014; Corander *et al.* 2013; Guo *et al.* 2015; Hess *et al.* 2013; Reitzel *et al.* 2013). Our results (i.e. number of SNPs and outliers) are consistent with previous work using RAD-seq, providing further support for the view that RAD-seq based genome

scans can generate 1000s of SNPs in non-model marine species with direct application to management and conservation needs. Our dataset of 7163 filtered SNPs is within the range presented for other RAD-seq studies in marine species [approximately 300 to > 40 000, (Benestan *et al.* 2015; Catchen *et al.* 2013; Chu *et al.* 2014; Combosch & Vollmer 2015; Corander *et al.* 2013; Gruenthal *et al.* 2014; Guo *et al.* 2015; Herrera *et al.* 2015; Hess *et al.* 2013; Hohenlohe *et al.* 2010; Hohenlohe *et al.* 2013; Reitzel *et al.* 2013)] and the number of loci from these studies is one to several orders of magnitude larger than the number of loci used in studies utilizing other markers such as microsatellites and AFLPs. In *P. magellanicus*, for example, two previous population genetic studies both used less than 10% of the loci used in our study [six microsatellites in Kenchington *et al.* (2006), 634 AFLPs in Owen & Rawson (2013)]. The sheer number of markers generated using RAD-seq, and their placement across the entire genome of an organism, are predicted to increase accuracy and power of statistical tests of differentiation and spatial patterns (Allendorf *et al.* 2010; Waples 1998).

2.5.2 Detection and influences of selection

The ability to detect loci potentially under directional selection offers a significant advantage to RAD-seq based genome scans over traditional approaches in marine population genetic studies (Gagnaire *et al.* 2015). Importantly, identifying markers potentially under selection can improve the accuracy of conclusions drawn from population genetic studies; failing to account for the effects of selection could lead to overestimation of neutral differentiation and underestimation of gene flow, highlighting the importance of separating neutral and outlier loci during analyses. Examining potential selection using

outlier loci may also increase spatial resolution (Hellberg 2009), providing opportunities to track individuals and predict adaptive differences (e.g. Therkildsen *et al.* 2013). Our separation of outlier and neutral loci allows us to disentangle the effects of selection and underlying neutral variation (and gene flow) within sea scallops, thereby generating a more complete picture of population connectivity in this species. Undoubtedly, each approach includes some identification error, but the combination of several outlier detection methods can help reduce rates of false positives (Gagnaire *et al.* 2015). We detected significant overlap in the loci identified as outliers using both BAYESCAN and ARLEQUIN, demonstrating that our list of outlier loci was largely robust to the assumptions of differing approaches and supporting the outlier status of these loci. In total we detected 112 outlier loci, representing approximately 1.6% of the loci examined. This value compares favourably with numbers of outlier loci detected and used in other studies of marine organisms Guo *et al.* (2015): 0.99% of identified SNPs; Milano *et al.* (2014): 4.59%; Hess *et al.* (2013): 3.65%; Bradbury *et al.* (2013): 5.2%; De Wit & Palumbi (2013): 3.2%; and Bourret *et al.* (2013b): 2.6%) and although low, still remains reasonably consistent with studies suggesting approximately 5-10% of a genome may show signatures of selection (Nosil *et al.* 2009; Strasburg *et al.* 2012).

Not surprisingly, all of our analyses found stronger genetic differentiation and population structure signals in outlier loci than when examining neutral loci alone or all loci together (neutral and outlier). This observation is consistent with other studies that detected higher levels of structure and increases in the spatial resolution of population structure with outlier loci (Bradbury *et al.* 2010; Hemmer-Hansen *et al.* 2014; Milano *et al.*

2014, but see Moore *et al.* 2014 for an exception). Separate use of neutral and outlier loci also detected differences in structural patterns, primarily in the clustering results for the north regional group. The apparent isolation of LTB, particularly in neutral loci, could reflect local larval retention and increased genetic drift. Circulation patterns near LTB can retain larvae from local populations and limit larval dispersal out of Placentia Bay (Bradbury *et al.* 2008d; Bradbury *et al.* 2000). Other causes may also have contributed to this pattern, including rare events and the sampling of different age classes. Ultimately, differences in spatial patterns between putative outliers and neutral markers may reflect influences of differing structuring forces such as selection and drift but attributing spatial variation to these factors will require additional study.

2.5.3 Spatial population structure

Many RAD-seq studies of marine species report fine-scale geographic structure (Benestan *et al.* 2015; Catchen *et al.* 2013; Reitzel *et al.* 2013). We observed significant population structure along the range of *P. magellanicus*, separating sampling locations into two distinct groups – north and south of Nova Scotia, Canada. These results mirror population structure detected in other marine species in the Northwest Atlantic, including *Homarus americanus* (Benestan *et al.* 2015) and *G. morhua* (Bradbury *et al.* 2014a; Bradbury *et al.* 2010; Bradbury *et al.* 2013) and build on smaller-scale levels of differentiation reported among scallop populations (Kenchington *et al.* 2006; Owen & Rawson 2013). The discontinuity between northern and southern population clusters was evident in all loci tested, but strongest using outlier loci. Further examination of the north and south groups independently detected no further structuring in the south, but different

patterns of population structure in the north emerged when using outlier and neutral loci separately. The homogeneity of the south group of populations was somewhat surprising given documented differences in reproductive timing and fecundity among different populations (e.g. Naidu 1970; Barber *et al.* 1988; Beninger 1987; DuPaul *et al.* 1989; Kirkley & Dupaul 1991), however in the south our results may be indicative of plasticity due to environmental variation in these traits or the importance of loci not detected in our analysis. Between the north and the south groups, especially considering the strength of the genetic break in outlier loci, our results (and the documented differences in phenology) may be indicative of adaptation to regional environmental conditions. In general, our results confirm patterns of genetic structure detected by Kenchington *et al.* (2006), who used microsatellites to identify three regional clusters and a putative barrier to gene flow that separates some of the western Scotian Shelf, Newfoundland, and the Gulf of St. Lawrence from all samples further south. This pattern mirrors our clustering of Newfoundland and the Gulf of St. Lawrence from everything south of Cape Breton, NS, Canada, however our results indicate higher over-all levels of differentiation among sampling sites. Kenchington *et al.* (2006) found further differentiation of the US portion of George's Bank as well as a sample from the Gaspé Peninsula, Canada, however modelling studies on George's Bank found little support for this differentiation (Davies *et al.* 2015; Davies *et al.* 2014; Gilbert *et al.* 2010; Tian *et al.* 2009a, b). Overall, the variance explained by the three regions found in Kenchington *et al.* (2006) (1.21%) was low compared to our outliers (11.0%) but similar to the amount explained by all of our loci combined (0.58%). Using AFLPs, Owen & Rawson (2013) found similar values of differentiation to our SNP data, however they sampled a smaller region in the Gulf of Maine at a finer scale, in contrast to the low

differentiation and limited spatial structuring present in our dataset in that region. Owen & Rawson (2013) also found significant inter-annual variation in population structure and recommend sampling multiple years and age classes when making population genomic inferences; our samples were collected over three years and contained multiple age classes, reducing the potential influence of age structure on our results.

Spatially, we observed significant isolation by distance in our samples (although the pattern was primarily driven by the north/south split rather than a true stepping-stone model), and clear latitudinal clines in allele frequency at both neutral and outlier loci. Our examination of IBD explored two estimates of pairwise distances between populations, the least-cost minimum distance between populations and an approximate ocean current based distance. Overall, the least-cost distance performed better than the ocean current distance supporting the hypothesis that dispersal alone is not the primary determinant of connectivity among sea scallop populations. Interestingly of the 18 IBD relationships tested, including different subsets of loci and locations (spatial subsets), only IBD relationships that included outlier loci were statistically significant. Furthermore, outlier loci produced much more pronounced clinal patterns in allele frequency than in neutral loci and outlier loci also produced lower average cline width than neutral loci. Processes other than selection can lead to the development of clinal patterns in loci (Vasemägi 2006), but the differences between clines seen in our outlier and neutral loci indicate that neutral forces are not the sole driver of clines among sea scallop populations. In light of these results, we conclude that the outlier loci used in our analyses predominantly drive observations of the

north/south split within sea scallops, comparable to patterns reported in Atlantic cod in the same locations (Bradbury *et al.* 2010; Bradbury *et al.* 2013).

For marine species with planktonic larval stages, larval dispersal is expected to contribute significantly to spatial population structuring (Bradbury *et al.* 2008c; Bradbury & Snelgrove 2001) though patterns of resultant connectivity can be complex. Planktonic larval duration (PLD) can last up to 40 days in sea scallops and creates the potential for extreme long-distance dispersal. Past studies show, however, that PLD is not necessarily a good proxy for gene flow, and can substantially overestimate levels of population connectivity (Bradbury *et al.* 2008c; Selkoe & Toonen 2011). Scallop larvae, as with other marine larvae, are also unlikely to occur uniformly through the water column (Manuel *et al.* 1996a; Tremblay & Sinclair 1990a, b), and may therefore encounter a range of currents induced by vertical shear that could influence dispersal distances and directions (Metaxas 2001). Similarly, changes in current patterns alter source populations of larvae (Kordos & Burton 1993). Geographic barriers in the Northwest Atlantic can also influence larval movement and survival; one barrier in particular, the deep Laurentian Channel, cuts between Newfoundland and Nova Scotia within the Cabot Strait (Townsend *et al.* 2006). Current outflow from the Gulf of St. Lawrence to the southeast may further hinder larval transit from Newfoundland southwards, contributing to the observed north/south sample split. Current patterns along the coast of the southern portion of the species range may also limit dispersal, given that the boundary between coastal currents and inshore waters could act as a barrier to larval movement (Tilburg *et al.* 2012). Scallop surveys show large scallop aggregations associated with particular habitat characteristics, including gravel substrate,

low predation, and a high proportion of filamentous organisms. These characteristics define appropriate settling substrate for scallop larvae with high settlement success and survival (Stokesbury & Himmelman 1995; Thouzeau *et al.* 1991); settlement of groups of larvae on different patches of appropriate habitat can further enhance genetic differentiation. The meeting of the cold Labrador Current and warm Gulf Stream off the coast of Nova Scotia, Canada, produces strong temperature gradients along the entire range of sea scallops (Townsend *et al.* 2006) and variation in temperature (and other environmental factors) between populations of sea scallops may contribute to adaptation within populations by way of within-generation selection (Pavey *et al.* 2015), further influencing the differential survival of dispersing and recently settled larvae as they presumably survive best in conditions for which they are adapted.

2.5.4 Estimates of dispersal

Estimating dispersal distance in marine species remains a significant challenge (Selkoe & Toonen 2011). Standard methods of estimating dispersal include (but are not limited to) drifter studies and biological-physical modelling, PLD, chemical tracking, direct observation, assignment tests, and use of natural or artificial markers (Bradbury *et al.* 2008c; Cowen & Sponaugle 2009; Hedgecock *et al.* 2007; Levin 2006; Saenz-Agudelo *et al.* 2009; Selkoe & Toonen 2011; Thorrold *et al.* 2002; Thorrold *et al.* 2007). Our estimates of effective dispersal here ranged from approximately 300-600 km per generation. These distances compare directly with estimates for other marine invertebrates and fish in eastern North America (Bradbury *et al.* 2008d; Kinlan & Gaines 2003), and particularly *P. magellanicus* on George's Bank and the Mid-Atlantic Bight where modelling studies of

scallop larvae identified some local retention and dispersal between adjacent populations (Davies *et al.* 2015; Davies *et al.* 2014; Gilbert *et al.* 2010; Tian *et al.* 2009a, b). Given a planktonic period of a month or more, some correlation with the direction and nature of coastal circulation patterns may be expected. In fact, previous work in sea scallops associated genetic structure with the dominant ocean currents, supporting larval dispersal as the main structuring agent. Kenchington *et al.* (2006) identified regional structure consistent with expected current patterns along the east coast of North America using a microsatellite panel and a simplified oceanographic model. Results indicated the potential for larval dispersal between populations on the Scotian Shelf and within the Gulf of Maine, however, patterns of larval movement differed based on the depth of model particles. Given larval *P. magellanicus* have been previously shown to exhibit diel behaviour (Tremblay & Sinclair 1990a), the assumption that surface currents approximate dispersal potential might be tenuous. Indeed, our observation that the least cost path distance was a better predictor of genetic spatial structure than current based geographic distance suggests our approximations of circulation may not capture the complexity of larval dispersal. This observation may also reflect the influence of variation in post-settlement processes (e.g. mortality) (Bradbury *et al.* 2008a; Clarke *et al.* 2010) associated with climatic variation expected across this range (Townsend *et al.* 2006) on the realized connectivity of the system.

Genetic methods of estimating and inferring dispersal patterns, like those employed in this study, can be very effective. However, they reflect effective dispersal (that is, the

subsequent survival and reproduction of dispersers) rather than strict movement among populations.

Comparison of dispersal estimates based on neutral and outlier loci may allow some inference of the roles that dispersal and selection play in regulating connectivity, because estimates from neutral loci may reflect absolute movement of larvae but those from outlier loci will reflect the survival and reproduction of dispersers. Our observation that the estimates of dispersal based on the outlier loci were smaller than those based on neutral loci supports a hypothesis that selection and differential survival may be important in limiting effective dispersal and connectivity. Other studies report similar observations for coastal fish species elsewhere, detecting genetic structure at smaller geographic scales than dispersal would suggest (Bradbury *et al.* 2008b; Clarke *et al.* 2010).

Both methods of estimating dispersal make inherent assumptions, raising concerns about dispersal estimates in both cases. With the clinal method, concerns include errors in LD calculations, equations that assume selection/dispersal balance, and violation of assumptions by long distance dispersal (Sotka & Palumbi 2006). Previous work demonstrates that IBD itself is robust to deviations from some model assumptions (Leblois *et al.* 2003; Leblois *et al.* 2004) with clear successes in estimating local dispersal (Broquet *et al.* 2006; Rousset 1997; Sumner *et al.* 2001), however, as discussed previously, the use of IBD to estimate dispersal distance in sea scallops may be inappropriate for the patterns of population structure detected in the system. A steep cline primarily drives the structure we identified within scallops in contrast to the patterns assumed in a gradual island or stepping-stone model. In this case cline-based estimates likely produce more accurate

dispersal values as reflected in our results. The IBD-based method produces estimates less than 15 kilometers, a distance far too small to ensure any connectivity among our sampled populations (the smallest pairwise distance between our samples is approximately 70 km). The distances calculated using the clinal method appear much more realistic, falling within the range of pairwise distances between our populations. Nonetheless, they yield much smaller estimates than the maximum pairwise distance found between our populations, indicating that limited dispersal may add significantly to population structure within the sea scallop.

2.5.5 Limitations

Although genetic methods, and RAD-seq in particular, offer great potential for measuring marine connectivity (Gagnaire *et al.* 2015), many limitations and caveats must also be considered. RAD-seq itself raises concerns about use of non-random missing data that may affect population genetic inferences and conclusions (Arnold *et al.* 2013; Gautier *et al.* 2013), however Arnold *et al.* (2013) found F_{ST} to be relatively robust to missing data compared to other differentiation estimates. Arnold *et al.* (2013) also recommend complete trimming of loci with missing data; we trimmed our loci to maximum 20% missing data, however, loci with missing data comprised a very small proportion of our total loci and likely had no substantial influence on our results. Each of our populations was sequenced in a single library on a single lane, which could lead to lane effects manifesting as population effects and biasing our results. Our strict filtering, however, helped to combat missing data effects. We also observed similar trends across multiple populations, lending support that these trends were not artefacts. The differentiation in the neutral loci for LTB

is an exception; however, given that previous studies have identified circulation patterns that limit larval dispersal out of Placentia Bay (Bradbury *et al.* 2000; Bradbury *et al.* 2008d), we do not believe this result reflects a sequencing bias.

As discussed previously, local adaptation and selection among different populations may influence the conclusions of population genetic studies by leading to an overestimation of the differentiation between populations. This may promote inaccurate estimates of migration and gene flow between populations if the possible influence of selection is not taken into consideration. We separated loci for analysis both to identify potential regions of local adaptation but also to generate a more conservative and potentially accurate pattern of dispersal and connectivity among sea scallop populations. We used a strong HWE filter when filtering our detected outlier and neutral loci. Although some outlier loci may have been expected to be out of HWE and thus removed during filtering, prioritizing true (and higher quality) loci furthered our aim for an accurate evaluation of population connectivity.

Genetic methods characterize effective connectivity, or only the contributions of dispersers that survive and reproduce. This approach may miss instances of larval movement without subsequent reproduction within the new population. For marine management, however, the effective movement and survival of dispersers and the contribution of dispersers to population stability generally represents the most important measure. The prevalence of rare events adds another concern when trying to draw management conclusions from genetic population connectivity results. These events (e.g. unusual currents, storms) may create temporary channels of dispersal. Larvae transported during these events that survive and reproduce may leave their genetic signature in the

population for some time. Depending on the size of the group of atypical dispersers, researchers may misinterpret patterns of dispersal within a group of populations. Similarly, yearly deviations from an average map of connections add further variation, emphasizing the importance of including a wide range of sample ages when inferring population connectivity from genetic data. Despite somewhat limited availability of samples, we included scallops spanning a wide range of ages wherever possible. In addition, repeated sampling, more individuals per population, and more detailed age structure analyses could help confirm the stability of patterns of population structure over time.

2.5.6 Conclusions

Using RAD-seq derived SNPs, we describe range-level population structure in sea scallops, building on work that detected smaller-scale differentiation using microsatellites (Kenchington *et al.* 2006) and AFLPs (Owen & Rawson 2013). Significant genetic differentiation between the northern and southern regions of the species distribution mirrored patterns in other Northwest Atlantic species. Estimates of dispersal using genomic clines, likely the most appropriate approach for our system, indicate moderate potential dispersal within sea scallops, however, variables other than larval transport may also drive population structure. Patterns in population structure differed when using neutral and outlier loci, indicating that selection and local adaptation may play a role in sea scallop population dynamics. The major population structure identified, as well as the potential for adaptation, offers valuable information for management of this economically important species. In particular, the strong division between northern and southern populations indicates that separate management strategies are likely appropriate for these regions.

Furthermore, the potential isolation of LTB warrants special attention. The same factors that structure sea scallop populations presumably affect other species in the region with similar life histories, and comparison of these species with associated environmental and oceanographic variation in the area may provide significant insights into prevalent factors influencing regional population differentiation and adaptation.

2.6 Tables

Table 2.1. Site name, site code, coordinates, and the number of sequenced *P. magellanicus* from each of 12 collection sites in the Northwest Atlantic Ocean.

Site Name	Site Code	Latitude	Longitude	Number of scallops used in analysis
Sunnyside, NL	SUN	47.824108	-53.869456	20
Little Bay, NL	LTB	47.1545	-55.10416667	21
Magdalen Islands	MGD	47.1143	-62.0243	21
Northumberland Strait	NTS	46.13383333	-63.77283333	22
Passamaquoddy Bay	PSB	45.06473333	-67.01663333	12
Bay of Fundy	BOF	44.67615	-66.07181667	22
Scotian Shelf - Middle	SSM	44.52066667	-60.635	19
Gulf of Maine Inshore	GMI	44.52	-67.0319	20
Browns Bank	SSB	42.83716667	-66.13583333	22
Gulf of Maine Offshore	GMO	42.44	-70.3874	22
George's Bank	GEO	41.61266667	-66.36216667	22
Mid Atlantic Bight*	MDA	38.82265936	-73.59895436	22

*several neighbouring sites sampled as one location

Table 2.2. Number of *P. magellanicus* individuals and number of SNP loci included in initial sequencing and final analysis following quality control (QC).

Parameter	Value
Individuals sequenced	252
Individuals following QC	*245 (97.2% of Individuals sequenced)*
Initial RAD tags	131897
RAD tags following QC	*19672 (14.9% of Initial RAD tags)*
Initial SNPs	173482
SNPs following QC	*7216 (4.2% of Initial SNPs)*
SNPs in HWE	7163 (99.3% of SNPs following QC)
Outlier SNPs	*112 (1.6% of SNPS in HWE)*
Neutral SNPs	*7051 (98.4% of SNPs in HWE)*

Table 2.3. Analysis of molecular variance (AMOVA) among 12 populations of *P. magellanicus*, among regional groups of populations identified by Structure analysis, and among individuals within populations using (A) all loci, (B) neutral loci, and (C) outlier loci.

(A)

Source of variation	df	Proportion of variation	p-value
Among groups	1	0.58	< 0.001
Among populations within groups	10	0.09	< 0.001
Among individuals within populations	233	5.43	< 0.001
Within individuals	245	93.9	< 0.001

(B)

Source of variation	df	Proportion of variation	p-value
Among groups	1	0.4	< 0.001
Among populations within groups	10	0.02	< 0.001
Among individuals within populations	233	5.5	< 0.001
Within individuals	245	94.08	< 0.001

(C)

Source of variation	df	Proportion of variation	p-value
Among groups	1	11.0	< 0.001
Among populations within groups	10	4.24	< 0.001
Among individuals within populations	233	1.71	< 0.001
Within individuals	245	83.06	< 0.001

2.7 Figures

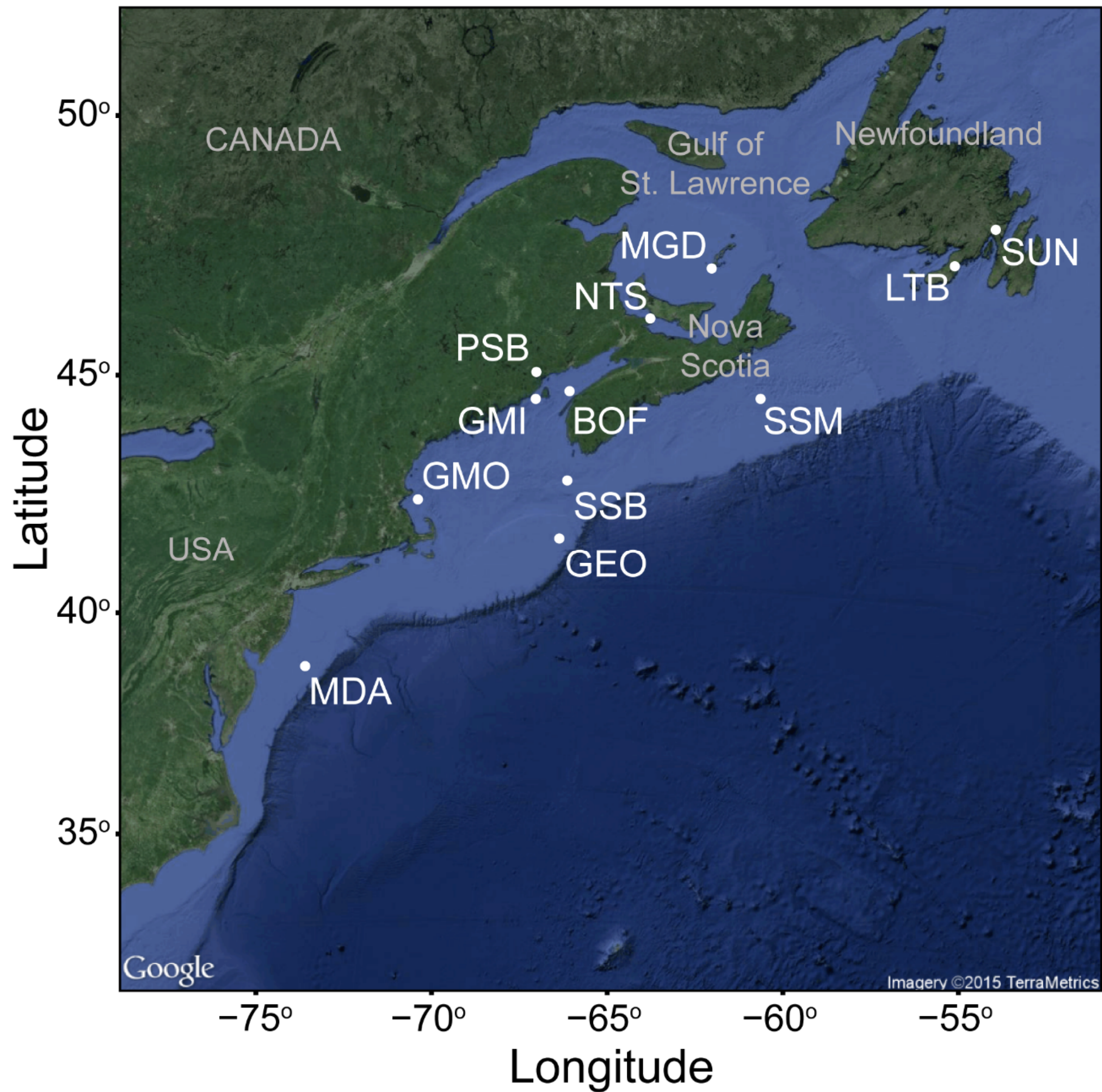

Figure 2.1. Map of 12 sea scallop (*P. magellanicus*) collection locations from the Northwest Atlantic. Site MDA (Mid-Atlantic Bight) represents the middle of several nearby collection locations grouped as one population. Population codes are defined in Table 2.1.

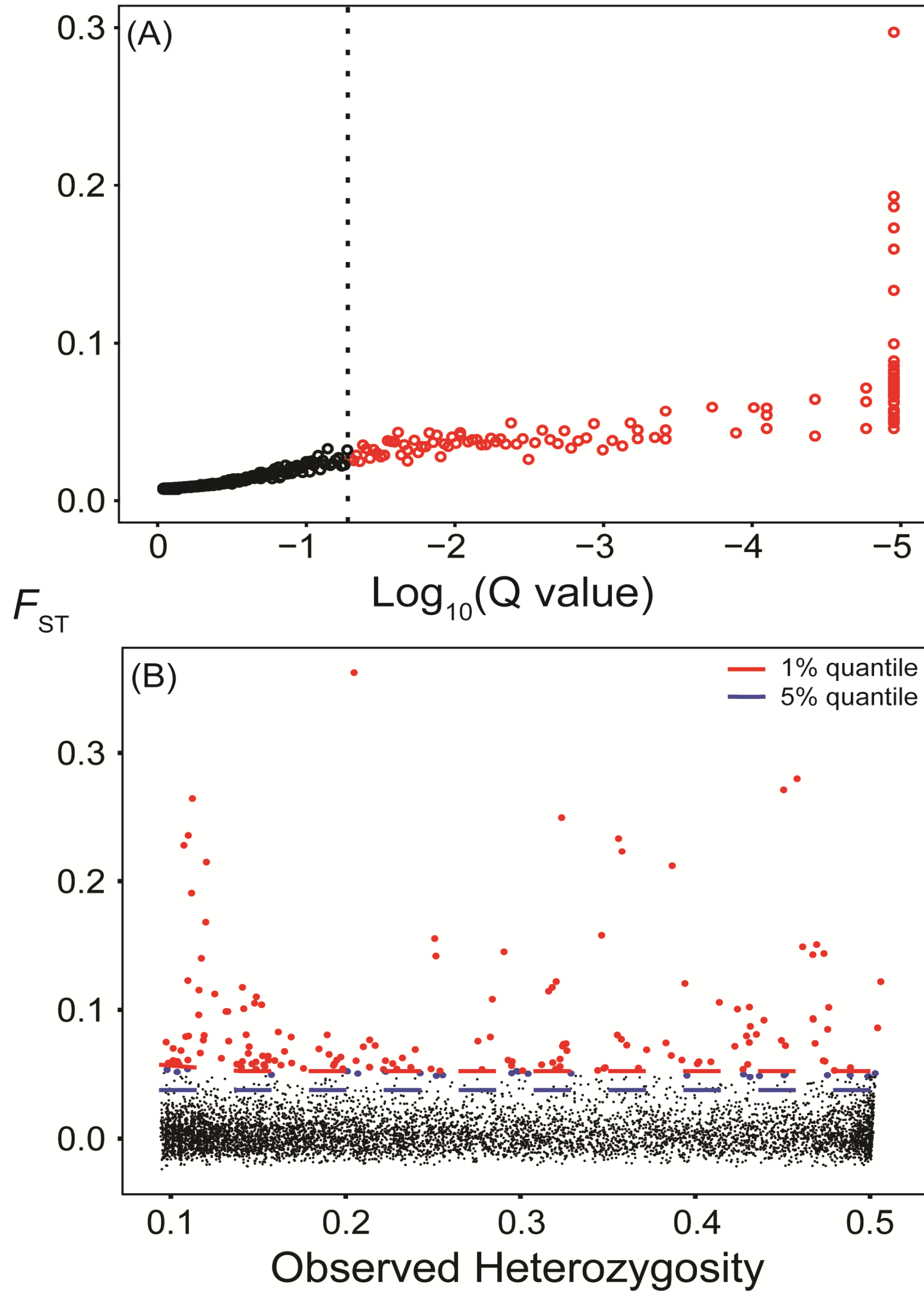

Figure 2.2. Results from (A) the Bayesian test for selection completed using the program BayeScan and (B) the hierarchical island model test for selection completed using the program Arlequin for 7163 loci sequenced in 12 populations of *P. magellanicus*. BayeScan outliers are defined as all loci with a q-value higher than 0.05 (highlighted in red). Arlequin outliers are defined as the loci that fall above the simulated 1% quantile of F_{ST} vs Heterozygosity ($p <= 0.01$, highlighted in red).

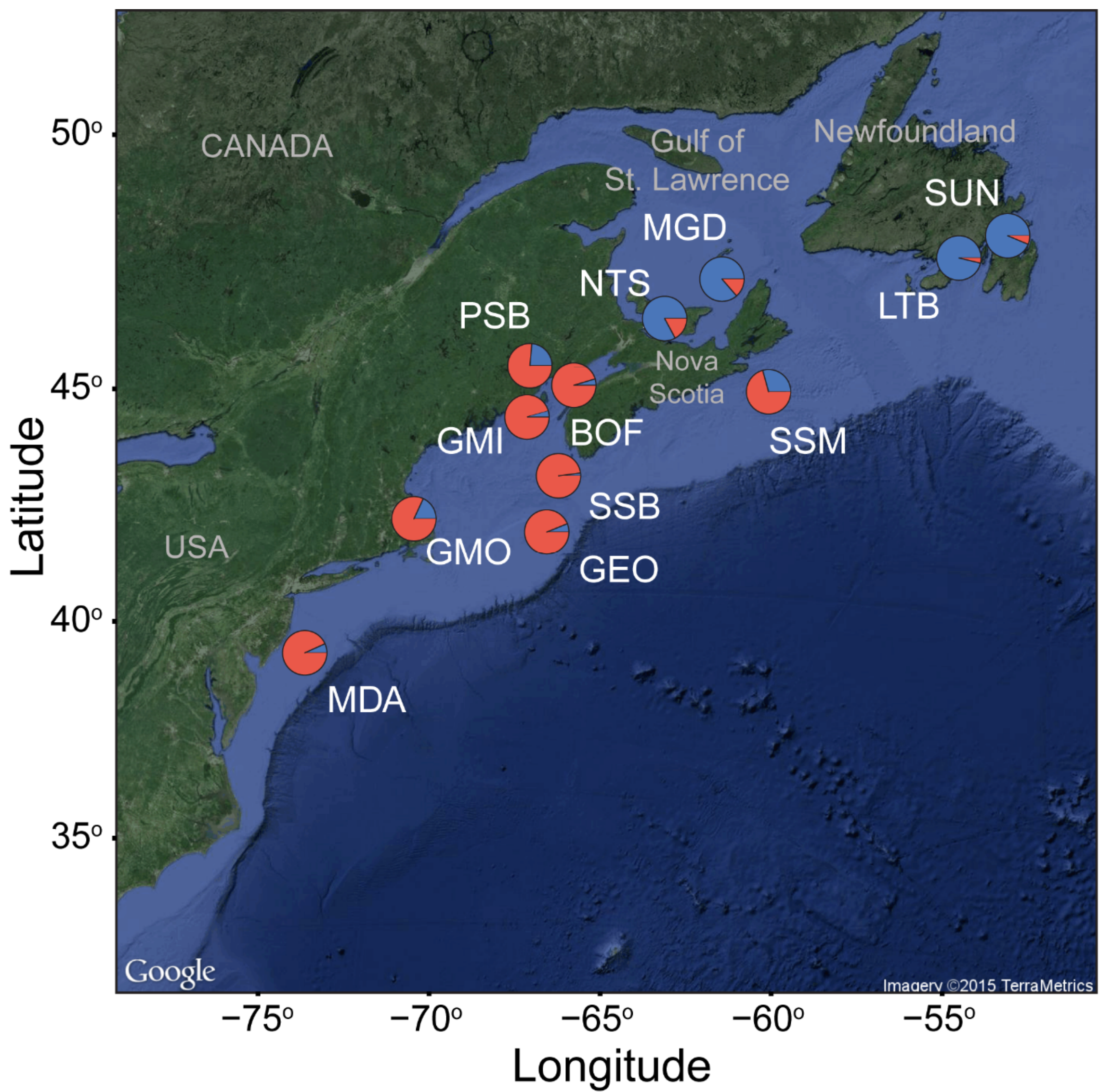

Figure 2.3. Map of the proportion of each of the 12 *P. magellanicus* populations assigned to two population groups (blue and red) identified in the program Structure using outlier loci and the ΔK method to select the optimal number of genetic clusters in the data.

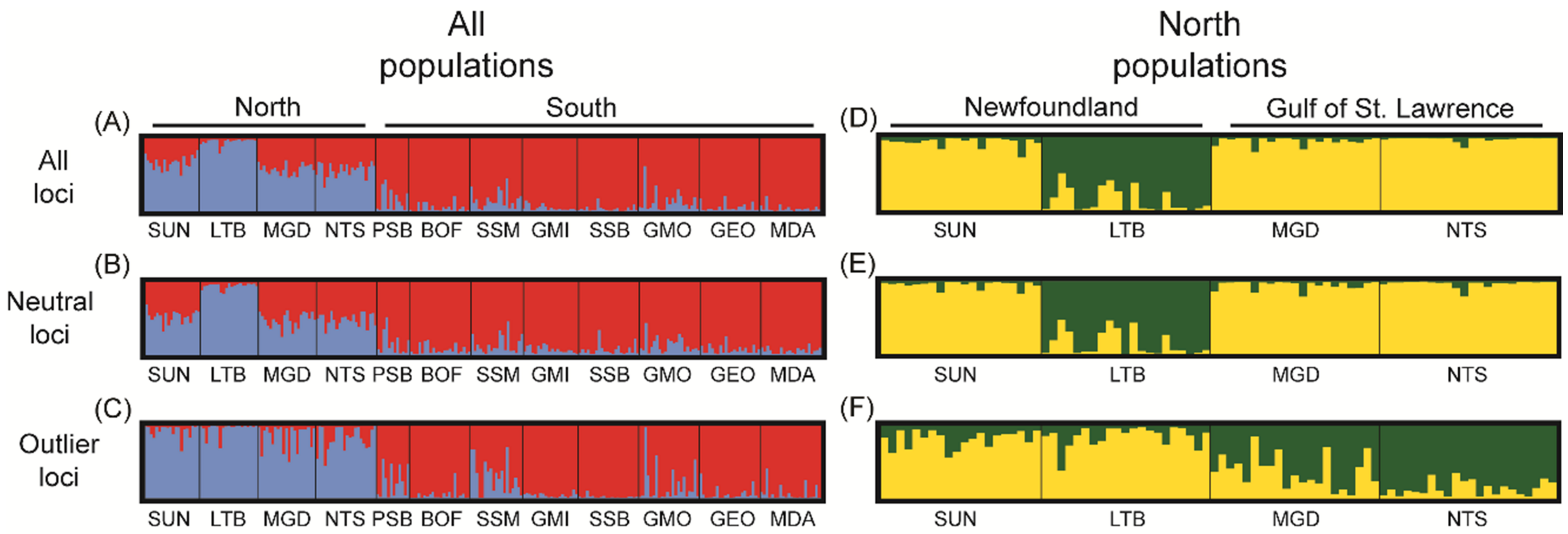

Figure 2.4. Plots of individual admixture for 12 populations of *P. magellanicus* at $K = 2$ determined using the program Structure and the ΔK method to select the optimal number of genetic clusters in the data. Results are presented for all populations at (A) all loci, (B) neutral loci, and (C) outlier loci as well as four north populations at (D) all loci, (E) neutral loci, and (F) outlier loci.

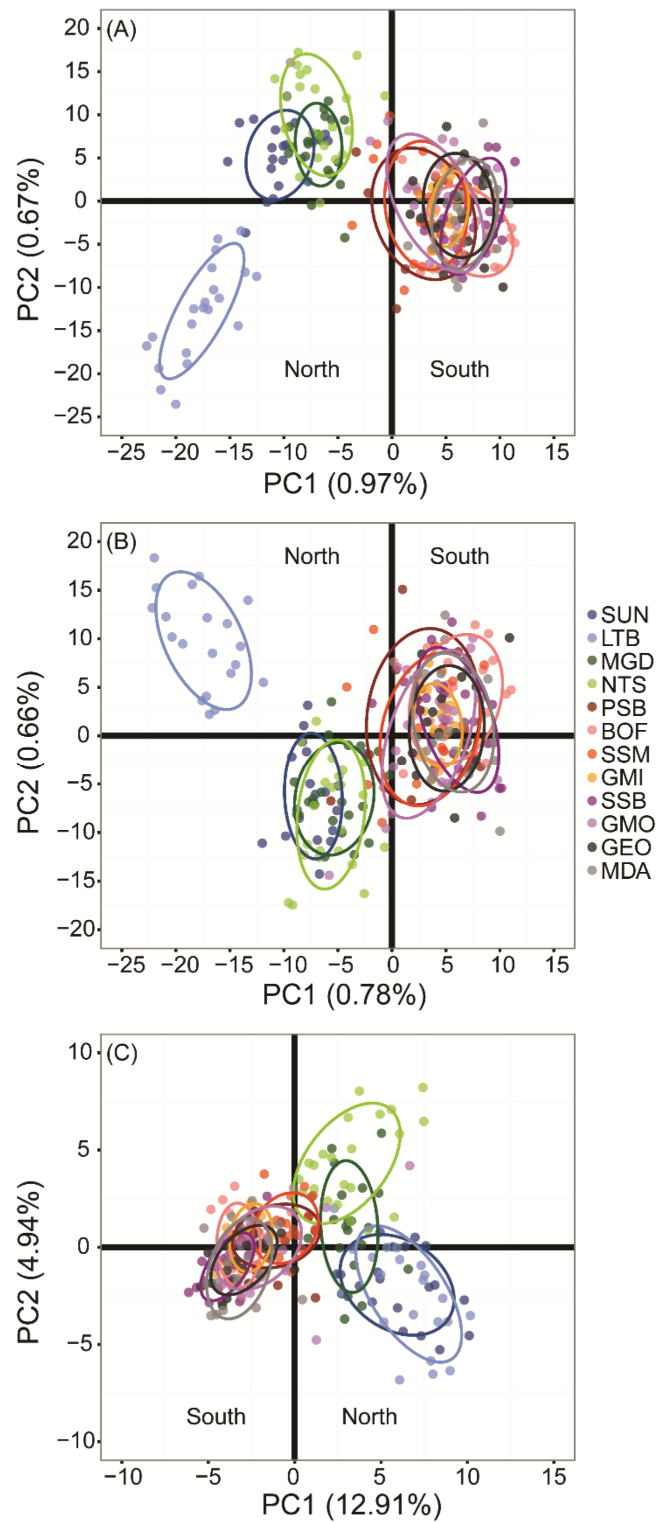

Figure 2.5. Principal components analysis plots for (A) all loci, (B) neutral loci, and (C) outlier loci in 12 populations of *P. magellanicus*.

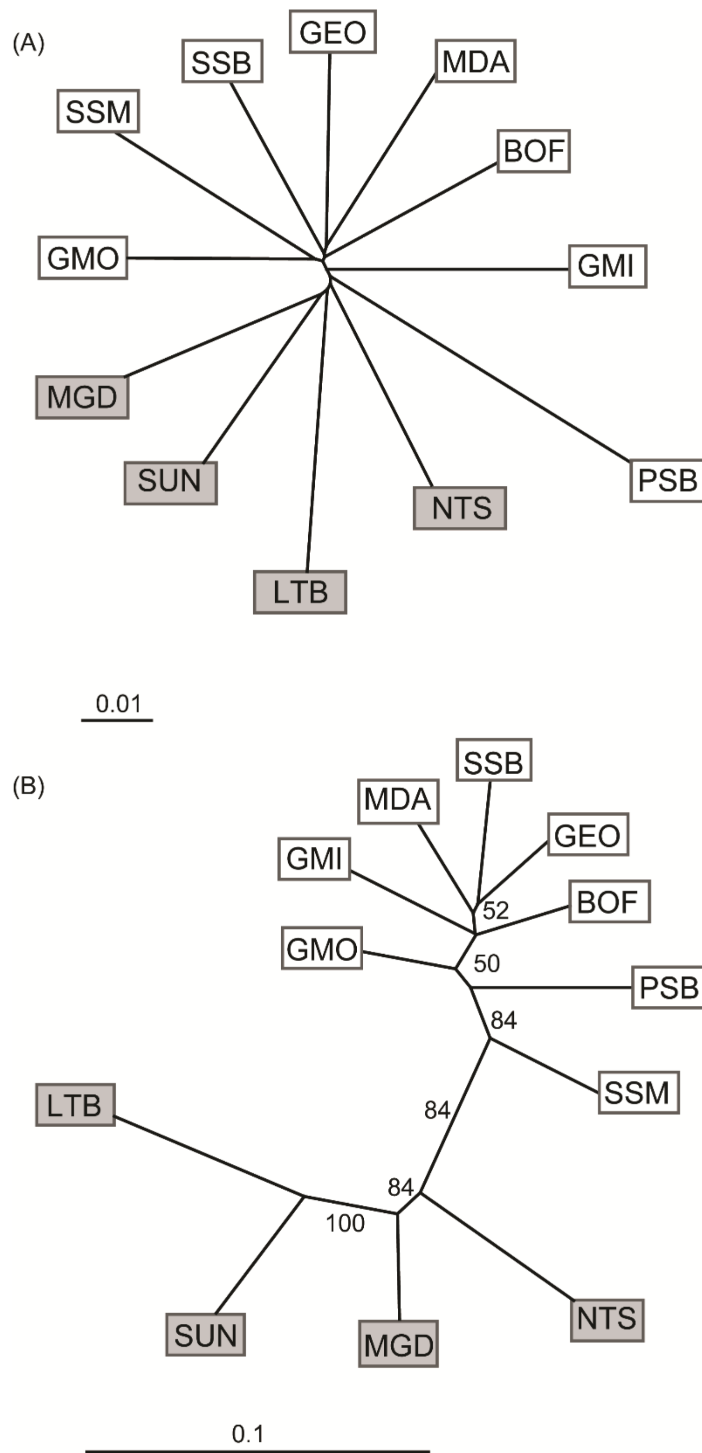

Figure 2.6. Neighbour joining trees for Cavalli–Sforza and Edwards chord distance (D_c) between 12 populations of *P. magellanicus* for (A) neutral loci and (B) outlier loci. North populations are highlighted in grey, south populations in white, and bootstrap values greater than 50% are shown.

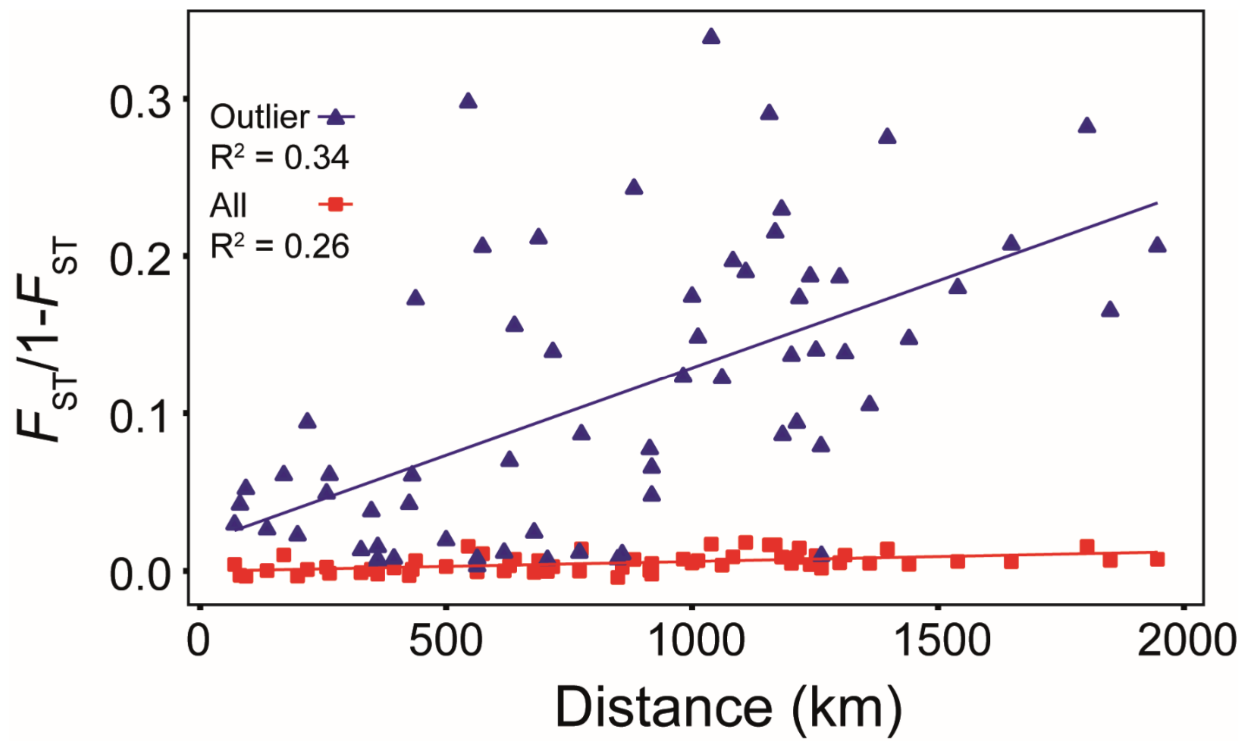

Figure 2.7. Isolation by distance plot of $F_{ST}/1-F_{ST}$ vs. population pairwise distance for 12 populations of *P. magellanicus* using least-cost distance for all loci (red squares), and outlier loci (blue triangles).

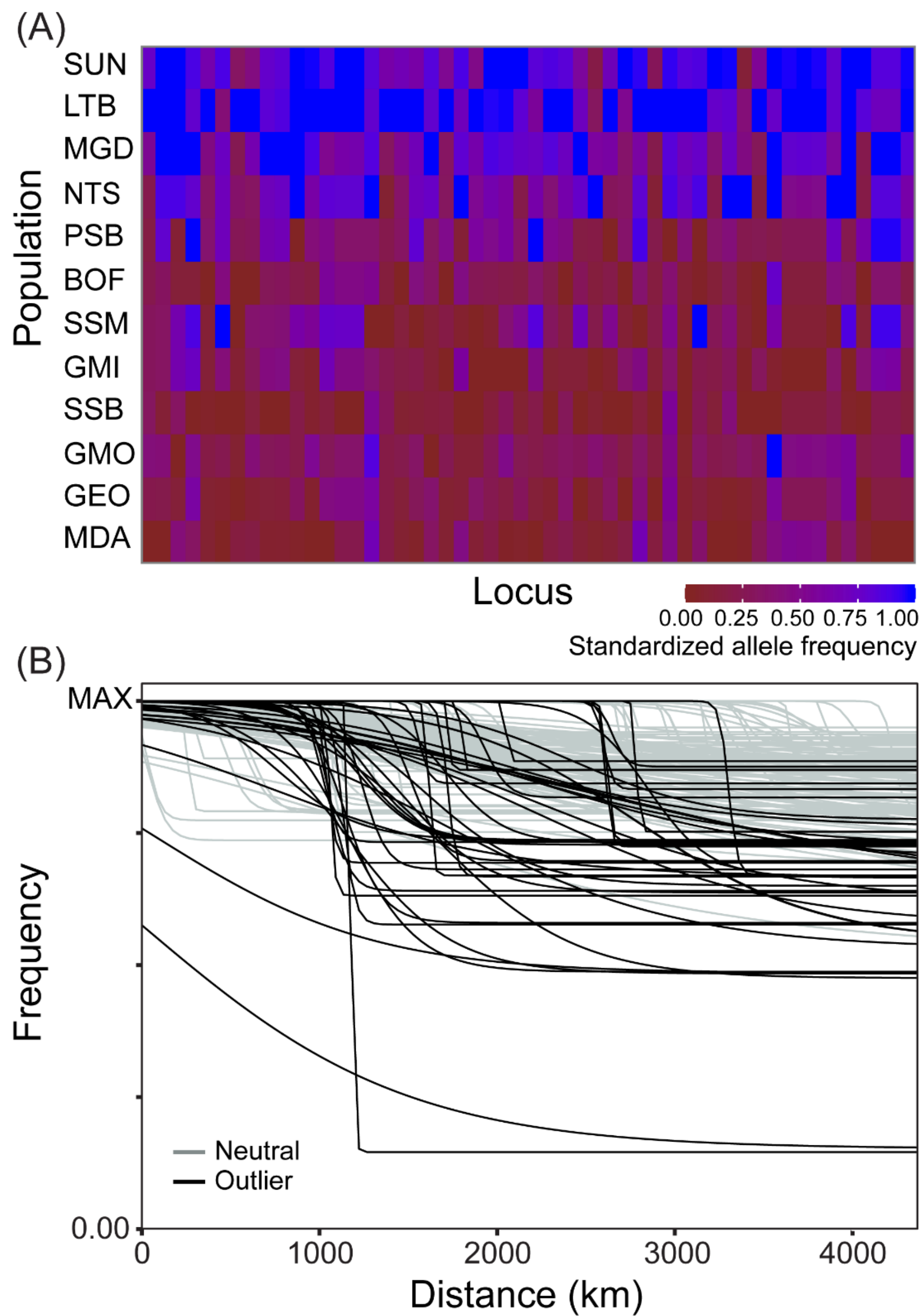

Figure 2.8. (A) Heat map of population specific standardized allele frequencies for 48 clinal outlier loci in 12 populations of *P. magellanicus*. (B) Plot of clines in allele frequency in 12 populations of *P. magellanicus* as a function of the distance in kilometers from the furthest north population (SUN) for clinal neutral loci (n = 123, 24.6% of tested loci, grey) and clinal outlier loci (n = 48, 42.9% of tested loci, black).

Chapter 3: Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*

3.1 Abstract

Environmental factors can influence diversity and population structure in marine species and an accurate understanding of this influence can both enhance fisheries management and help predict responses to environmental change. We used Restriction-Site Associated DNA sequencing to genotype 7163 SNPs in 245 individuals of the economically important sea scallop, *Placopecten magellanicus* to evaluate the correlations between oceanographic variation and a previously identified latitudinal genomic cline. Sea scallops span a broad latitudinal area (>10 degrees) and we hypothesized that climatic variation significantly drives clinal trends in allele frequency. Using a large environmental dataset, including temperature, salinity, chlorophyll a, and nutrient concentrations, we identified a suite of SNPs (285 - 621, depending on the analysis and environmental dataset) potentially under selection through correlations with environmental variation. Principal component analysis of the different outlier SNPs and environmental datasets revealed similar north and south clusters, with significant associations between the first axes of each (R^2_{adj} = 0.66 – 0.79). Multivariate redundancy analysis of outlier SNPs and the environmental principal components indicated that environmental factors explained more than 32% of the variance. Similarly, multiple linear regressions and random-forest analysis identified winter average and minimum ocean temperatures as significant parameters in the link between genetic and environmental variation. This work indicates that oceanographic variation is associated with the observed genomic cline in this species and that specifically,

seasonal periods of extreme cold restrict gene flow along a latitudinal gradient in this marine benthic bivalve. Incorporating this knowledge into management may improve accuracy of management strategies and future predictions.

3.2 Introduction

The application of population genomic-based approaches to the study of marine population structure is increasingly revealing higher levels of genetic differentiation and population structure in multiple marine species than previously recognized (e.g. Benestan *et al.* 2015; Bradbury *et al.* 2013; Corander *et al.* 2013; Milano *et al.* 2014; Moura *et al.* 2014). Recent observations of fine-scale differentiation are changing our view of marine connectivity and marine population dynamics (Hauser & Carvalho 2008). Limited dispersal may contribute to fine-scale population differentiation (e.g. Van Wyngaarden *et al.* 2017), but given large populations and large environmental gradients, selection may also contribute significantly to genetic differentiation among marine populations (Hauser & Carvalho 2008). As such, studies supporting a role for selection in regulating marine connectivity continue to accumulate (Bradbury *et al.* 2010; Clarke *et al.* 2010; Limborg *et al.* 2012; Milano *et al.* 2014; Sjöqvist *et al.* 2015). An accurate understanding of population structure and environmental influences can contribute to the identification of conservation units and allow prediction of a species' response to climate change, particularly for economically important species (Allendorf *et al.* 2010; Conover *et al.* 2006; Sale *et al.* 2005).

Genomic studies increasingly highlight a role for selection in regulating marine population structure (Berg *et al.* 2015; Bradbury *et al.* 2014b; Bradbury *et al.* 2010;

Gagnaire *et al.* 2015; Gaither *et al.* 2015; Hellberg 2009). Loci identified as putatively under selection repeatedly reflect small-scale genetic differentiation in multiple marine species (Bradbury *et al.* 2010; De Wit & Palumbi 2013; Lamichhaney *et al.* 2012). Marine landscape genomic studies combining traditional landscape approaches with large genomic datasets have identified significant associations between climate and genetic structure (genetic-environmental associations, GEA) in numerous marine (and anadromous) species, including Atlantic herring (*Clupea harengus*) (Limborg *et al.* 2012), Atlantic cod (*Gadus morhua*) (Berg *et al.* 2015; Bradbury *et al.* 2010), purple sea urchin (*Strongylocentrotus purpuratus*) (Pespeni & Palumbi 2013), Atlantic salmon (*Salmo salar*) (Bradbury *et al.* 2014b), European hake (*Merluccius merluccius*) (Milano *et al.* 2014), and Chinook salmon (*Oncorhynchus tshawytscha*) (Hecht *et al.* 2015). The pervasiveness of genetic-environmental associations across taxa and life histories supports the hypothesis that environmental associated selection may structure marine populations.

The sea scallop, *Placopecten magellanicus* (Gmelin) is an economically important benthic marine bivalve characterized by a planktonic period of development conducive with a potential for long distance dispersal among populations (Davies *et al.* 2014; Tian *et al.* 2009b). The scallop fishery in both the United States and Canada, one of the most economically important fisheries in the region, extends back over 100 years (DFO 2016; Naidu & Robert 2006; NOAA 2016). The sea scallop distribution extends from North Carolina, USA to Newfoundland, Canada (Posgay 1957). This region spans a vast latitudinal range where the cold Labrador Current meets the warm Gulf Stream, encompassing large gradients in ocean temperature and other environmental factors all of

which may be influenced by oceanographic properties of the major currents and storm-related mixing along the coast (Townsend *et al.* 2006). Several oceanographic barriers along the range may also influence larval dispersal and survival among populations (Townsend *et al.* 2006). Previous studies detected significant population structure among scallop populations (Kenchington *et al.* 2006; Owen & Rawson 2013; Van Wyngaarden *et al.* 2017) driven primarily by outlier loci, indicating that both limited dispersal and adaptation may play a role in structuring of scallop populations, particularly given potentially strong selective pressures along the species' range.

In light of the unique oceanographic features in the region, the large latitudinal range of the species, and previously identified clinal population structure, we hypothesize that directional selection and local adaptation drive sea scallop population structure and that ocean temperature likely contributes significantly to adaptation of the species to its local environment. Our specific objectives were to: (1) explore spatial variation in environmental variables across the range of the sea scallop, (2) use environmental correlation-based outlier detection methods to pinpoint potential targets of environment based selection across the genome of the sea scallop, and (3) identify potentially important environmental drivers of population structure and adaptation in scallops. We build directly on a previous study identifying latitudinal clinal trends in allele frequency across the range using 7163 RAD-seq derived SNPs (Van Wyngaarden *et al.* 2017) and extend that work by identifying environmental associations and possible mechanisms.

3.3 Methods

3.3.1 Sample collection and RAD-seq

Van Wyngaarden *et al.* (2017) describe sampling procedures and RAD-seq genotyping in detail. In brief, 252 adult scallops were collected by divers or bottom trawl from a total of 12 locations across the entire species range between 2011 and 2013 (Table 3.1, Figure 3.1) with a minimum of 12 scallops per population (mean value of 20.4 ± 2.8 scallops). Tissue samples were collected and preserved in AllProtect (Qiagen) or 80% ethanol. DNA extraction and RAD-seq library preparation were performed at the Aquatic Biotechnology Lab, Bedford Institute of Oceanography in Dartmouth, Nova Scotia. RAD-seq libraries were prepared as described by Etter *et al.* (2011b) (see also Etter *et al.* 2011a) with modifications. Sequencing was performed at the McGill University and Génome Québec Innovation Centre, Montréal, Canada. SNPs were detected using the *de novo* pipeline in STACKS v.0.9999 (Catchen *et al.* 2011). The final dataset was filtered using PLINK v.1.07 (Purcell 2009; Purcell *et al.* 2007) to include only RADtags present in 75% of individuals in SNP discovery and calling; all SNPs included in the analysis were present in 75% of individuals with a minor allele frequency greater than 5%. Furthermore, we excluded individuals with more than 20% missing loci from the analysis. Loci were filtered for Hardy-Weinberg Equilibrium using the program GENEPOP v.4 (Rousset 2008), excluding loci out of equilibrium in 6 or more populations from the analysis (<0.7% of all loci).

3.3.2 Environmental data collection and processing

We amalgamated environmental data from several databases; from Fisheries and Oceans Canada: Climate (Gregory 2004) (years 1970-2013), BioChem (DFO 2014) (years 2009-2014), and AZMP (DFO 2015), and from the National Oceanographic and

Atmospheric Administration in the United States of America (NOAA, years 1990-2010), and the MODIS satellite database (NASA Goddard Space Flight Center Ocean Ecology Laboratory 2014) (years 2002-2013). Data was averaged over multiple years available to remove the signatures of short-term variation in the marine environment. Measured variables included water temperature, salinity, sigma-t, and chlorophyll A, SiO_4^{4-}, NO_3^-, NO_2^-, and PO_4^{3-} concentrations.

We averaged data from all data sources within a bounding box of 1 square degree around each sample site to create site-specific averages for each data type used in the analysis. Data were separated into surface and depth values based on the collection depth for each sampling location. Where collection depth was unknown, we used Canadian Hydrographic Service charts, NOAA maps, and Google Earth to estimate depth based on collection GPS coordinates. Surface values encompassed depths between zero and 20 meters except for collection sites less than 20 m depth, where 10 m was used as the surface cut-off. We averaged values from a cut-off of approximately ten meters above a given collection depth to the collection depth for depth profiled variables. In cases where multiple sample collection depths were provided, depth cut-off parameters were altered to include the entire range of collection depths (Table 3.2).

Data validation and preparation were completed using R (R Development Core Team 2012). To address natural seasonal variation in the data, we calculated z-scores for each variable for each sample site per month and removed outliers where necessary. Variables with missing data for more than six sites were removed from subsequent analyses. For the remaining variables with missing data, we used single imputation using

neighbouring sites to estimate missing values (sites arranged by latitude, averaging sites directly north and south of the missing site). Following outlier removal and imputation, we standardized data to zero mean and unit variance by subtracting the mean and dividing by the standard deviation. We then identified site-specific maximum and minimum values as well as seasonal averages for each variable, basing seasons largely on equinoxes. Winter included January, February, and March, Spring included April, May, and June, Summer included July, August, and September, and Autumn included October, November, and December. The final dataset contained 90 variables spanning all available data types (hereby referred to as AllEnv). We repeated all analyses using only the temperature, salinity, and chlorophyll a variables (n=36 variables, henceforth CST), as we expected these to be the most likely associated with selection.

3.3.3 Detection of outlier loci

We used two separate methods to detect outlier loci using both environmental datasets (four tests in total). The first method used a Bayesian framework implemented in the program BAYENV2 (Coop *et al.* 2010; Guenther & Coop 2013). This method calculates a set of "standardized allele frequencies" which controls for population history and structure when detecting loci whose allele frequencies show significant associations with environmental variation. This method then calculates a Bayes Factor (BF), which measures the weight of evidence for a model in which the environmental variable affects the allele frequency of a locus versus a null model with no environmental variable effect. To calculate the "standardized allele frequencies", we randomly selected 700 loci (9.8% of total loci) identified as neutral (not under selection) in Van Wyngaarden *et al.* (2017). The

null model correlation matrix was estimated from these loci in three repetitions of 100 000 iterations. We visually compared correlation matrices from the final iterations of each run to each other and to an F_{ST} matrix of the neutral loci. The final matrix from the first run was selected as the neutral matrix for use in further analysis. The final analysis detected locus-specific deviations from the neutral allele frequencies using 100 000 iterations. BFs were calculated at every locus for each environmental variable separately. To assess the significance of each BF (and the likelihood of classifying a locus as an outlier), we created 5 bins of loci based on the global minor allele frequency, as recommended in Coop *et al.* (2010) and implemented in Hancock *et al.* (2010) (Table S2.1). We selected loci with BFs in the top 5% of the range of BFs for each bin as outliers.

Latent factor mixed models (LFMM) as described in Frichot *et al.* (2013) were implemented as the second method of outlier detection in the R package *LEA* (Frichot & François 2015). This method uses latent factors in a linear mixed model to control for population structure (the number of genetic clusters within a group of populations, K) while detecting correlations between environmental and genetic variation. Previous analysis using the program STRUCTURE v.2.2.4 (Pritchard *et al.* 2000) detected two genetic clusters (K=2), and the genomic inflation factor analysis (GIF) in *LEA* corroborated this result. The models were run for 3 repetitions, with a burn-in of 5000 followed by 15 000 iterations. We combined Z-scores from the 3 repetitions using the median, calculated adjusted p-values to correct for multiple testing, and produced a list of candidate outlier loci for each environmental variable (FDR = 0.05) following Frichot *et al.* (2015). To ensure we included any loci potentially under selection, we combined the list of detected

outliers from each method to create our final outlier lists (AllEnvOutlier for AllEnv and CSTOutlier for CST).

3.3.4 Association between environmental factors and genetic variation

We conducted principal components analysis (PCA) using the AllEnvOutlier and CSTOutlier loci using the R package *adegenet* (Jombart 2008) to examine population structure among the sampled populations at outlier loci. To examine the relationship between environmental and genetic variation among our collection sites, we calculated population specific allele frequencies for AllEnvOutlier and CSTOutlier using the R package *gstudio* (Dyer 2014). Next, we ran PCA on population specific allele frequencies for AllEnvOutlier and CSTOutlier (AllEnvOutlierPCA and CSTOutlierPCA), and the population specific environmental data in AllEnv and CST (AllEnvPCA and CSTPCA) using the R package *adegenet* (Jombart 2008). Linear regression was then performed between the first principal component (PC) from AllEnvOutlierPCA (AllEnvOutlierPC1) and the first PC from the PCA on AllEnv (AllEnvPC1) as well as the first PC from CSTOutlierPCA (CSTOutlierPC1) and the first PC from the PCA on CST (CSTPC1).

We then conducted redundancy analysis (RDA), a multivariate canonical correlation analysis, using the R package *vegan* (Oksanen *et al.* 2015) on population specific allele frequencies for AllEnvOutlier and CSTOutlier and selected PCs from AllEnvPCA and CSTPCA, respectively. Each PC that explained more than 5% of the total explainable variance in the AllEnvPCA (5 axes) and CSTPCA (4 axes) were selected as explanatory variables. Backwards stepwise variable selection using 1000 or 10000 iterations selected the most valuable environmental PCs within the model. In order to

determine the proportion of model variation attributable to climate vs. geographic distance between populations vs. combined effects, we next performed partial RDA, conditioning the genetic matrix on the distances from the furthest north population (SUN) along a 1-dimensional transect that included all populations [estimated using GOOGLE EARTH (2013)].

Multiple linear regressions quantified the direction and magnitude of the effect of environmental variables on genetic variation. We used results from RDA to select environmental variables used in the analyses. After examining weightings of the environmental variables on the important PCs selected during RDA, we selected the five most highly weighted variables from each PC for use as explanatory variables in linear mixed models. Based on results from the initial linear mixed models (see *Results*), we generated models focusing on measurements of water temperature at surface and at depth (Table 3.3). For each response variable (AllEnvOutlierPC1 and CSTOutlierPC1), we fitted a global multiple regression model with all selected environmental variables. We then used the R package *MuMIn* (Barton 2014) to run all possible configurations of the global model and pinpointed the best model fits with AIC_c model selection. We examined cumulative AIC_c model weights to rank each parameter in order of importance and estimated coefficients for each environmental parameter using model averaging (Arnold 2010).

We also used non-linear random-forest analysis (RF) to identify important environmental variables and then compared key drivers with those identified using multiple linear regressions. RF is a powerful machine-learning model that has been widely used in many disciplines since its introduction in 2001 (Breiman 2001) however it

has only recently been applied to genomic adaptation analysis (Brieuc *et al.* 2015). This analysis can take into account interaction between predictor variables and may manage the covariation between environmental variables more effectively than the multiple linear regression approach (Brieuc *et al.* 2015). This ensemble approach benefits from growing a large group of decision trees to improve overall performance. One key attribute of random forests is its automatic computation of variable importance. We used a method based on weighted k nearest neighbours (KNN) called KNNcatImpute (Schwender 2012) to impute the missing genotypes in our genetic SNP data using the *scrime* package in R (Schwender & Fritsch 2013). After imputation, the individual genotypes at each outlier SNP were transformed to categorical data. SNPs are a bi-allelic genetic marker and only two alleles and three genotypes can be present at each SNP; the built RF is thus a three-class classification model. Environmental variables were used as predictors of individual genotypes at each outlier SNP using 1,001 trees. We then used permutation importance, the variable importance function built in RF, to rank the relative roles of environmental variables. In order to obtain a reliable estimation of variable importance, we applied 10-fold cross-validation, dividing the entire data set into 10 subsets. Nine subsets trained the RF model and the other subset was used for validation; this process was repeated 10 times for each SNP genotype. In each of the 10 runs, we calculated a permutation importance array for all environmental variables. Noting that importance values can be negative, we first computed the exponential values of the importance array and then averaged each importance value over the total importance sum of all environmental variables to generate an importance proportion array. The importance proportions were averaged over the 10 runs to determine average importance proportions. For each SNP genotype output (621

for AllEnv and 285 for CST), an RF model was built to calculate an array of permutation importance proportions for all environmental variables. We calculated the overall average importance proportion for each environmental variable over all loci. All RF analyses were performed using *randomForest* package in R (Liaw & Wiener 2002).

3.3.5 Gene ontology

Using the full RADtags for all detected outlier loci, we performed gene ontology (GO) analysis and NCBI nucleotide BLAST searches on AllEnvOutlier and CSTOutlier in the program Blast2GO (Conesa *et al.* 2005) using the program default parameters and InterProScan to improve GO annotation quality.

3.4 Results

3.4.1 Sample collection and RAD-seq

Following filtering and quality control steps, we included 245 individual scallop samples in our analysis (97.2% of sequenced individuals), 19672 RADtags (14.9% of initial RADtags), and 7216 SNPs (4.2% of Initial SNPs) (Table S2.2). The 7163 SNPs in HWE that met all quality control standards were used in all subsequent analyses. Van Wyngaarden *et al.* (2017) describe further sequencing results in detail.

3.4.2 Detection of outlier loci

The neutral matrices calculated to generate "standardized allele frequencies" for BAYENV2 varied little within runs and when compared to the F_{ST} matrix calculated for the neutral loci; we therefore chose a single matrix for further calculations with BAYENV2. LFMM identified K=2 as the most supported number of clusters (and thus latent factors)

using GIF analysis, with values of 0.85 for AllEnv and 0.83 for CST [p-values calibrate correctly when GIF approaches 1 (Frichot & François 2015)].

Overall, LFMM identified more loci potentially under selection than BAYENV2. Combining the results from both programs, AllEnv identified 621 loci (8.7% of all loci) as under selection, whereas CST identified 285 loci (4.0% of all loci) as under selection. 250 loci were shared between the two datasets (Table 3.4). Using AllEnv, BAYENV2 detected 128 loci as putatively under selection (1.8% of total loci), whereas LFMM detected 511 (7.1% of total loci). The two sets of loci shared only 18 loci. Using CST, BAYENV2 detected 72 loci (1.0% of total loci) whereas LFMM detected 218 (3.0% of total loci), with only 5 loci shared between the two methods. Within the BAYENV2 results, the AllEnv outlier list and CST outlier list shared 37 loci (Table S2.3). The LFMM analysis of AllEnv and CST overlapped completely in loci identified (Table S2.4).

3.4.3 Association between environmental factors and genetic variation

PCA of all individuals and sets of outlier loci detected using AllEnv and CST both split north and south populations along the first PC, separating the populations into two clusters as seen in the BAYENV2 results and in Van Wyngaarden *et al.* (2017) (Figure 3.2). Using AllEnvOutlier, the first PC explained 2.38% of the total explainable variance in the model, and using CSTOutlier the first PC explained 3.33% of the total explainable variance. The PCA on the population specific allele frequencies for AllEnvOutlier and CSTOutlier (not shown) produced a similar clustering pattern, however, the first PC explained much more variance, with AllEnvOutlierPC1 explaining 26.47% of the total model variance and CSTOutlierPC1 explaining 31.93% of the total model variance.

The environmental data produced the same pattern of north/south population clustering for both datasets (AllEnv and CST, Figure 3.3). However, these PCAs further separated the southernmost population, Mid-Atlantic Bight (MDA), along the second PC. The first PC of the environmental data explained much more variance than in the genetic models, with AllEnvPC1 explaining 40.18% of the total model variance and CSTPC1 explaining 51.35%. Linear regressions between genetic and environmental data (i.e. AllEnvOutlierPC1 and AllEnvPC1 as well as CSTOutlierPC1 and CSTPC1) showed a strong and significant relationship (Figure 3.4), with adjusted R^2 values of 0.79 for AllEnv and 0.66 for CST, further indicating similar spatial patterns in genetic and environmental variation among our sample sites. The north/south population split can be seen in heat maps of standardized allele frequency and standardized values for environmental variables in AllEnv and AllEnvOutlier and CST and CSTOutlier (Figure 3.5), although only a subset of alleles show the strong clinal pattern driving the north/south split.

To examine the effects of climate vs. geography on the genetic variation within the outlier SNP loci, we selected 5 PCs from AllEnvPCA and 4 from CSTPCA for use as explanatory variables in RDA, each explaining more than 5% of the total variance in the PCA. In AllEnvPCA, the 5 selected axes explained 89.78% of the total model variance, and in CSTPCA, the 4 selected axes explained 88.96% of the total variance. Backwards stepwise variable selection on the RDA for AllEnv retained only AllEnvPC1 as an explanatory variable, whereas selection on the RDA for CST retained both CSTPC1 and CSTPC4 (Figure 3.6). Both models demonstrated significant relationships, despite low adjusted R^2 values (AllEnv, $R^2_{adj} = 0.15$, $p = 0.001$; CST, $R^2_{adj} = 0.23$, $p = 0.001$). Variance

partitioning showed that climate explained a significant component of the model variation in both cases, explaining 32.36% of model variation in AllEnv (compared to 30.37% explained by geography and 37.28% explained as joint effects) and 41.34% of model variation in CST (compared to 21.27% explained by geography and 37.39% explained by joint effects).

The RDAs for AllEnv and CST both separated north and south population groups. AllEnv retained only one environmental PC axis and we therefore show only one RDA axis in the plot (Figure 3.6A), however, this axis clearly divides the north and south populations. In CST, RDA1 divided north and south but added further division among sample sites along RDA2, including separation of populations from Newfoundland and the Gulf of St. Lawrence (Figure 3.6B). Partial RDA, following conditioning the genetic matrix on the distance between populations, no longer separated north and south populations once the effect of population separation distance was removed (Figure S2.1, AllEnv, $R^2_{adj} = 0.04$, $p = 0.06$; CST, $R^2_{adj} = 0.07$, $p = 0.03$). We expected this result given the strong relationship between environmental parameters and latitude in this region and the large latitudinal but small longitudinal span of the samples.

To choose environmental parameters to include in the multiple linear models, we examined variable weightings on the PC axes selected during RDA and retained the five most highly weighted variables from each axis. For all variables included in each global model we calculated cumulative Akaike Information Criterion weights and model-averaged parameter estimates (Table 3.5). Model selection using CSTEnv and all 10 selected environmental variables could not determine best fit models and provide accurate

estimates for parameter weights and coefficients because of overfitting of the model. Upon further examination of the RDA results, CSTPC1 appeared more important in driving the north/south population split. We repeated our multiple linear regressions and model averaging using only the five most highly weighted variables from CSTEnvPC1. In all cases, model weights averaged over all possible iterations of the models containing a particular variable indicated Surface Average Winter Temperature as the most important variable. Surface Minimum Temperature (occurred in winter) and Deep Average Winter Temperature also ranked highly, suggesting that the coldest temperatures encountered by both juvenile and adult scallops may play an important structuring role for scallop populations. Parameter estimates for all three variables were positive; increased minimum temperatures in the model corresponded to larger values of the first PC (higher PC values match the south population cluster).

Using RF, we calculated the importance proportion for all environmental variables using both AllEnvOutlier and CST Outlier (Figure 3.7). Using AllEnvOutlier, Deep Average Summer Salinity, Deep Minimum Salinity (occurred in Spring), and Deep Maximum Salinity (occurred in Autumn) ranked as the most important environmental variables. Surface Average Autumn Temperature, Deep Average Winter Temperature, and Deep Minimum Temperature (occurred in Winter) were also selected as important variables. CSTOutlier once again ranked salinity-associated variables as most important, however, Deep Average Winter Temperature and Deep Minimum Temperature ranked highly and the importance proportions for CSTOutlier exceeded those from AllEnvOutlier.

3.4.4 Gene Ontology

Blast2GO functionally annotated very few outlier loci. Using the CSTOutlier list, GO annotations were determined at only four loci (1.4% of loci), with a BLAST hit but no GO annotation at one further locus. In AllEnvOutlier only five loci (0.8% of total loci) were annotated, with a BLAST hit but no GO annotation in one further locus. The two lists of outliers shared three matches, with GO annotations split between molecular function (calcium ion and carbohydrate binding) and metabolic processes (regulation of transcription and steroid hormone mediated signalling). In the CSTOutlier list, GO annotation of the remaining locus identified a molecular function (oxidoreductase activity) and a metabolic process (oxidation-reduction process). In the AllEnvOutlier list, the GO annotations of the remaining two loci differed, one locus with molecular functions (oxidoreductase activity) and metabolic processes (oxidation-reduction process) and the other locus with several annotations (molecular function – catalytic activity, transferase activity, and folic acid binding, and metabolic processes – cellular metabolic processes) (Table S2.5). Examining the BLAST results for the outlier loci provided very little further information, with only two loci matching named genes in other bivalve species (the other loci only matched with predicted genes in a variety of species). Locus 12228_13 matched a retinoic acid reception mRNA from *Lumnaea stagnalis* (accession GU932671.1, similarity score of 44.2 with 75% matched nucleotides) with a role in embryonic development, while locus 20561_41 matched a CTL-9 mRNA from *Argopecten irradians* (accession JN166712.1, similarity score of 105 with 86% matches nucleotides) with a role

in cellular adhesion. The other loci only provided matches with predicted proteins in a variety of species.

3.5 Discussion

The identification of environmental factors regulating marine population structure can both inform fisheries management through the identification of management units and help predict species' responses to environmental change. Here we applied a landscape genomics approach using 7163 RAD-seq derived SNPs and 90 environmental variables to identify oceanographic factors associated with a latitudinal genomic cline in sea scallops in eastern North America. Our results support the hypothesis that seasonal periods of extreme cold restrict gene flow and influence population structure in this species. This work builds on previous studies on population structure in *P. magellanicus* (Kenchington *et al.* 2006; Owen & Rawson 2013), particularly the identification of a major genomic discontinuity separating the north and south of the species range (Van Wyngaarden *et al.* 2017). Our multivariate analysis using the outlier loci and environmental variables identified minimum and average winter temperatures as the most important variables describing genetic variation among populations of the scallop, indicating that over-winter survival may strongly influence structure of these populations. Overall the observed genomic and environmental correlations support the hypothesis of latitudinal structuring, driven predominantly by ocean temperature.

3.5.1 Environmental variables driving adaptation

Our results highlight ocean temperature as a critical environmental factor contributing to population structuring of the sea scallop. The sea scallop's distribution

spans almost 10° latitude encompassing an extremely large range of environmental conditions (approximately 5-10 °C difference in temperatures year-round between the northern and southern extremes of the range), primarily caused by prevailing currents (Townsend *et al.* 2006). The Labrador Current, a cold Arctic current, flows south from the coasts of northern Canada and Greenland, splitting around Newfoundland and circulating through the Gulf of St. Lawrence (Townsend *et al.* 2006). In contrast, the warm Gulf Stream moves north from the Gulf of Mexico along the east coast of North America. These two currents meet and move roughly offshore around Nova Scotia, exposing scallop populations to large differences in water temperature (and other oceanographic variables) in different areas of their range (Townsend *et al.* 2006). Our PCA based on environmental data clearly illustrates this variation and separation between two sections of the species range and clearly separates the north sampling locations in Newfoundland and the Gulf of St. Lawrence (exposed to the Labrador Current) from all of the more southern sampling locations (generally exposed to the Gulf Stream or its branches). Temperature variables primarily drive this separation and, in particular, the coldest temperatures (winter and minimum) differ most between the regions. In addition to cold temperatures, RF analysis also identified salinity as an important environmental variable, likely driven by very low salinity values at NTS and MGD in the Gulf of St. Lawrence. While salinity may exert some effect on these particular sites, the overall pattern of salinity variation in the environmental data differed from the north/south pattern detected in the population genetic structure analysis. However, when analysing outlier loci using only the 4 northern populations, Van Wyngaarden *et al.* (2017) nonetheless found a separation between the Gulf of St. Lawrence and the Newfoundland populations. Our RF analyses used allele

frequencies across all populations, however, by handling covariation between environmental data RF may have been able to detect smaller-scale variation associated with salinity that was masked by the strong temperature associations in the multiple linear regression analysis.

A strong genetic north and south discontinuity is supported here and in previous analysis (Van Wyngaarden *et al.* 2017). This split can clearly be seen in the first PC of the outlier PCAs, however, the first PC in both analyses explains little variance (2.38% and 3.33%). Van Wyngaarden *et al.* (2017) also performed PCA on the same genetic data used in this study, however they used 112 outlier SNPs detected using the program BayeScan (Foll & Gaggiotti 2008). Their results identified the same pattern in population structure that was identified using PCAs in the current study separating northern and southern populations along the first PC, however their PCA attributed more variance to the first PC (12.91%) than the PCAs in the current study. This difference in explained variance among the genetic PCs is likely due to the method used to detect outlier loci. BayeScan uses an F_{ST}-based method to detect outliers and generally selects the most divergent loci. Since the strongest structure seen among our populations is the split between northern and southern groups, it is likely that the most divergent loci will be the ones that follow this pattern. In the PCA, these loci load highly on the first PC and contribute their variance to the total variance explained by the first PC. In comparison, both BayEnv2 and LFMM use environmental correlations to detect outlier loci. Although these methods may also identify highly divergent loci as outliers, if a highly divergent locus (likely to be detected by BayeScan) does not correlate with the environmental variation captured in our

environmental dataset, it would not be included in the final outlier list. If this is the case, some of the most divergent loci that would explain a high proportion of the variance within the genetic data may not be included in the AllEnvOutlier and CSTOutlier lists and thus do not contribute their variance to that explained by the first PC. Since out intention was to determine loci putatively linked to adaptive variation, we feel the environmental association methods are the most appropriate despite the differences in variance explained by the different outlier lists.

The split between northern and southern populations can also be seen in the environmental PCAs, although the proportion of variance explained by the first PC in the environmental PCAs is much higher than that of the genetic PCAs (40.18% and 51.35% for AllEnv and CST, respectively). This is not surprising since the genetic data is inherently variable and each locus likely contributes a small amount to the total variation in the data (as expected in polygenic scenarios) rather than any one or a few loci solely driving the variation as opposed to the environmental data where a few variables (that load highly on the first PC) may be the sole drivers of the north/south discontinuity between populations. As may be expected by the similar patterns in the north/south discontinuity in both genetic and environmental data, we identified a strong relationship between spatial genetic and environmental variation. Previous analysis of the cline between northern and southern population groups (Van Wyngaarden *et al.* 2017) indicated that limited realized dispersal may be important in structuring sea scallop populations, however they also reported large differences in dispersal estimates calculated using outlier or neutral markers. This supports the hypothesis that selection and differential survival (evident in outlier loci) play a key

role in the genetic separation of north and south population groups. If selection and differential survival do play a role in structuring scallop populations, the patterns detected here may change with a changing ocean climate.

Increasingly the application of landscape genetic techniques in marine species provides evidence of local adaptation even with high levels of potential gene flow (Guo *et al.* 2015; Limborg *et al.* 2012; Sjöqvist *et al.* 2015). Studies consistently identify the importance of temperature to spatial structure and adaptive diversity. However, in many regions other environmental features often co-vary with temperature (e.g. salinity or ChlA) and in some analyses temperature may act as an unintentional proxy for the true selective force (a species may appear to adapt to temperature when in fact they are experiencing selection due to another variable such as ocean productivity). Although clear associations between genetic variation and temperature have been reported in several species, including Pacific invertebrates (Pespeni & Palumbi 2013), studies also demonstrate genomic adaptation to environmental gradients other than temperature, such as adaptation to salinity gradients in several Baltic Sea species (Berg *et al.* 2015; Limborg *et al.* 2012; Sjöqvist *et al.* 2015). In the North Atlantic, however, temperature variations (particularly with latitude) represent some of the strongest differences among regions (Townsend *et al.* 2006), and temperature may be the dominant selective force in this region. For example, strong correlations between genetic variation and ocean temperatures have been observed in many North Atlantic fish species including Atlantic cod, Atlantic herring, European hake, and Atlantic salmon (Berg *et al.* 2015; Bourret *et al.* 2013a; Bradbury *et al.* 2014b; Bradbury *et al.* 2010; Limborg *et al.* 2012)

3.5.2 Mechanisms of adaptation

The genomic associations with ocean temperature during periods of extreme cold (i.e. winter) suggest temperature associated mortality may significantly structure sea scallop populations. Sea scallops reproduce via broadcast spawning, generally in the autumn, although timing varies along their range. Given that scallops tend to spawn in the warmest water (Thompson 1977), generally between August and October (Beninger 1987; Langton *et al.* 1987; Naidu 1970), and they likely settle before December (Naidu & Robert 2006), a link between winter temperatures and larval mortality appears unlikely. Our analyses point to the overwinter survival of juvenile scallops as a potentially important structuring force limiting the effective dispersal of scallops between our northern and southern population groups, rather than selective mortality of planktonic larval scallops, and future experimental studies on larval and juvenile scallops may help to clarify this possibility. If true, this interpretation would point to the importance of bottom temperature, which is generally less readily available than surface temperature data. Although in some regions scallops may begin to spawn after their first winter (Naidu 1970), they may take a few seasons to reach reproductive maturity depending on food availability and other factors (Naidu & Robert 2006). Some evidence suggests that temperatures experienced by adults can help ensure a healthy larval year class (Dickie 1955; DuPaul *et al.* 1989; Kirkley & Dupaul 1991; Langton *et al.* 1987; Macdonald & Thompson 1985), further indicating that planktonic larval survival may not necessarily act as the main selective factor structuring sea scallop populations in the Northwest Atlantic. Interestingly, our study identified surface temperature rather than temperature at depth as the most important driver of selection,

contrary to expectations of juvenile scallop survival. One possible explanation is that deep temperature values are often estimated or provided as a range at collection sites, presumably reducing accuracy of those measurements relative to those for surface temperature. Our Blast2GO results identified possible genetic matches with several cellular processes which may highlight the potential mechanisms of thermal adaptation in the sea scallop. The lack of available genetic resources (i.e. reference genome) means these results are very preliminary and more study is needed before drawing conclusions from these results alone. However, as more genetic information becomes available for bivalves and scallops in particular, this may facilitate further detection of genes and pathways important to adaptation.

Despite the clear association observed with ocean temperature and population structure, the driving mechanism is still unclear. As described in a review by Bierne *et al.* (2011), local adaptation alone may not explain the genetic structure detected among populations or the geographic location of the strong break between population clusters. Tension zones (caused by endogenous barriers to gene flow) may have arisen independent of selection caused by environmental variation along the range of the species, potentially influencing the separation of population groups between the north and south of the species range. These tension zones may associate with environmental clines, and a combination of both endogenous and exogenous barriers (tension zones and selection) could contribute to the detected structure. This scenario could also reinforce local adaptation associated with environmental adaptation (Sexton *et al.* 2014; Shafer & Wolf 2013), furthering differentiation between regions. From a fisheries and population management perspective,

however, the initial cause of the genetic clustering becomes less important. The key conclusions here help in understanding how environmental variation influences clines in genetic structure and how climate change and harvesting pressures may interact to influence population dynamics in the future.

3.5.3 Challenges and limitations

Our environmental dataset spanned several decades and used 90 variables, a particularly large dataset compared to others used in landscape genomics studies. By utilizing available government data from both Canada and the United States (DFO 2014; Gregory 2004; NASA Goddard Space Flight Center Ocean Ecology Laboratory 2014), we compiled a dataset encompassing the entire range of our sample sites and the full latitudinal range of *P. magellanicus*. This data set allowed us to average values over many years and determine climatological values, reducing the influence of inter-annual variation on our results and focusing instead on long-term population level differences among our sample sites. Missing data required that we use simple imputation as necessary to complete the dataset. Although the use of single imputation to infer missing data can be problematic (Plaia & Bondi 2006), the majority of our data (particularly temperature and salinity data used in the final multiple linear models) required few or no imputations thus supporting the findings reported here.

This work identified outlier loci based on associations with environmental data using two techniques, BayEnv2 (Coop *et al.* 2010; Guenther & Coop 2013) and LFMM (Frichot & François 2015; Frichot *et al.* 2013). Both of these techniques account for neutral population structure when detecting outlier loci through environmental correlations,

however, the method differs between the two programs. BayEnv2 uses a subset of putatively neutral loci to first estimate neutral structure in the data, then tests for environmental correlations, whereas LFMM uses unobserved variables to account for population structure (due to population history and isolation-by-distance) within a mixed model. de Villemereuil *et al.* (2014) found that methods such as these that incorporate environmental data into outlier detection perform better than F_{ST}-based detection methods, and the number of loci detected by both methods (between 1.0% and 7.1%) is consistent with other studies examining putative selection on a genome wide scale (Nosil *et al.* 2009; Strasburg *et al.* 2012). In particular, de Villemereuil *et al.* (2014) found that BayEnv2 and LFMM performed better in highly structured hierarchical situations, however, in these situations BayEnv2 had low power to detect outliers compared to LFMM, especially when selection corresponded to an environmental gradient. Our results showed similarities to the de Villemereuil *et al.* (2014) analyses; LFMM detected more than twice as many loci as BayEnv2, potentially as a result of the putative lower power of BayEnv2. We also found little overlap in the results from both LFMM and BayEnv2, again mirroring the simulation results of de Villemereuil *et al.* (2014). Given the lack of overlap among approaches, we pooled outliers from both methods to increase our likelihood of detecting loci actually under selection. While this approach may decrease the power of some of our analyses to detect the strength of separation between the two population clusters, it increased the likelihood of capturing the important variables and genomic regions contributing to population structure in sea scallops.

Disentangling the influences of geography and climate remains a challenge in landscape genetics analyses, however strategies exist to help reduce this potential bias. Perhaps most importantly, many reviews on environmental association studies recommend removing the effects of neutral population structure to fully assess the effect of selection on population structure in natural systems (e.g. Rellstab *et al.* 2015). Previous researchers also recommended accounting for geographic distance and isolation-by-distance when examining potential isolation-by-ecology (e.g. Shafer & Wolf 2013). Because a single north/south population split characterizes our sample sites rather than a classic isolation-by-distance pattern (Van Wyngaarden *et al.* 2017), geographic distance among populations may not influence our results the way it would in a system characterized by a classic stepping-stone pattern. Our samples also align along the north/south axis of the population range and thus provide few opportunities to examine the effects of distance between samples without also removing the effects of latitude. The variation in environmental data primarily by latitude further complicates the question. With this in mind, we still attempted several methods to minimize the potential bias of neutral population structure on our results. We focused our analysis solely on outlier loci potentially under selection in the genome and using only these outlier loci likely made our analyses less prone to the confounding effects of neutral population structure. We also compared the results of RDA and pRDA, which controls for geographic distance among populations. Even when controlling for geographic distance, our results nonetheless suggest that climate can be a significant population structuring force, and although much of the variation in the data was attributed to the joint effect of climate and geography, the patterns in population structure are robust to the bias of distance among populations. Although we believe our results to be

robust to complications of neutral population structure and geographic distance, additional sampling (especially from populations at the same latitude) will help to separate more thoroughly the joint effects of climate and geography on scallop population structure.

Our analyses pinpointed potential environmental influences on sea scallop population structure, however, annotating the outlier SNPs of interest remains challenging. Although RAD-seq generates vast quantities of SNPs in organisms without reference genomes (Benestan *et al.* 2015; Catchen *et al.* 2013; Hohenlohe *et al.* 2012; Reitzel *et al.* 2013), the lack of more detailed genetic resources makes inference on the causal mechanisms contributing to local adaptations in sea scallops difficult, as noted by our lack of GO matches. Fortunately, with continued development of resources for *P. magellanicus* and related species, future studies will likely identify and study the features most important in characterizing sea scallop population structure. Our results also provide an important starting point for future studies. If temperature drives variation in the reproductive rates of scallops, then increasing water temperatures associated with global warming may alter scallop reproductive cycles and subsequent recruitment (Robinson *et al.* 2007). Genetic and genomic studies to examine further effects of selection on population structure in scallops in tandem with experimental studies to identify adaptations among scallop populations may be critical to predicting how the species will react to future climate change and harvesting pressures.

3.5.4 Conclusions

Our results show that ocean climate plays a role in structuring populations of sea scallops, in particular the influence of the coldest temperatures experienced. The

association with coldest temperatures points to the over winter survival of juvenile scallops as a structuring force rather than survival of larval scallops, contrary to what might be expected for broadcast spawning marine species. This work and similar landscape (or seascape) genetic studies highlight the possibility that local adaptation and the differential survival of dispersers (rather than solely limited dispersal) may have greater impact on the population structure of marine species than previously hypothesized. Our results can be useful in the effective management of *P. magellanicus* by helping managers in both Canada and the United States accurately determine geographic sources of larvae for exploited populations and predict the potential reactions of this species to a changing ocean climate, particularly with changes to the location and strength of dominant currents. Additionally, access to further genetic resources will continue to improve identification of the genes and pathways involved in adaptation and population structuring among sea scallop populations.

3.6 Tables

Table 3.1. Site name, site code, coordinates, and the number of sequenced *P. magellanicus* from each of 12 collection sites in the Northwest Atlantic Ocean.

Site Name	Site Code	Latitude	Longitude	Number of scallops used in analysis
Sunnyside, NL	SUN	47.824108	-53.869456	20
Little Bay, NL	LTB	47.1545	-55.10416667	21
Magdalen Islands	MGD	47.1143	-62.0243	21
Northumberland Strait	NTS	46.13383333	-63.77283333	22
Passamaquoddy Bay	PSB	45.06473333	-67.01663333	12
Bay of Fundy	BOF	44.67615	-66.07181667	22
Scotian Shelf - Middle	SSM	44.52066667	-60.635	19
Gulf of Maine Inshore	GMI	44.52	-67.0319	20
Browns Bank	SSB	42.83716667	-66.13583333	22
Gulf of Maine Offshore	GMO	42.44	-70.3874	22
George's Bank	GEO	41.61266667	-66.36216667	22
Mid Atlantic Bight*	MDA	38.82265936	-73.59895436	22

*several neighbouring sites sampled as one location

Table 3.2. Depth and surface ranges and bounding box coordinates used to select environmental data from databases for each of 12 *P. magellanicus* collection sites in the Northwest Atlantic Ocean.

Site Name	Site Code	Collection Depth (where available) (m)	Depth Range (m)	Surface Range (m)	Bounding Box Top Left		Bounding Box Bottom Right	
					Latitude	Longitude	Latitude	Longitude
Sunnyside, NL	SUN	12	-10 to -20	0 to -10	48.824108	-54.869456	46.824108	-52.2
Little Bay, NL	LTB		-30 to -40	0 to -20	48.1545	-56.10416667	46.1545	-54.10416667
Magdalen Islands	MGD		-35 to -45	0 to -20	48.1143	-63.0243	46.1143	-61.0243
Northumberland Strait	NTS		-15 to -25	0 to -10	47.13383333	-64.77283333	45.13383333	-62.77283333
Passamaquoddy Bay	PSB		-20 to -30	0 to -20	46.06473333	-68.01663333	44.06473333	-66.01663333
Bay of Fundy	BOF	30 to 137	-30 to -140	0 to -20	45.67615	-67.07181667	43.67615	-65.07181667
Scotian Shelf - Middle	SSM	38 to 49	-35 to -50	0 to -20	45.52066667	-61.635	43.52066667	-59.635
Gulf of Maine Inshore	GMI		-60 to -80	0 to -20	45.52	-68.0319	43.52	-66.0319
Browns Bank	SSB	45 to 123	-50 to -125	0 to -20	43.83716667	-67.13583333	41.83716667	-65.13583333
Gulf of Maine Offshore	GMO		-70 to -90	0 to -20	43.44	-71.3874	41.44	-69.3874
George's Bank	GEO	54 to 106	-50 to -100	0 to -20	42.61266667	-67.36216667	40.61266667	-65.36216667
Mid Atlantic Bight	MDA		-70 to -90	0 to -20	39.82265936	-74.59895436	37.82265936	-72.59895436

Table 3.3. Data included in all multiple linear regression models used to determine the direction and magnitude of the effect of environmental variables on genetic variation among 12 populations of *P. magellanicus.*

Method of Variable Selection	**Variables Included**	**Response Variable**
Most highly weighted variables from AllEnvPCs selected by RDA	Deep Average Autumn Salinity Deep Minimum SiO_4 Surface Average Autumn Salinity Surface Average Winter Temperature Surface Minimum Temperature	AllEnvOutlierPC1
Most highly weighted variables from CSTPCs selected by RDA	Deep Average Winter Temperature Deep Average Minimum Temperature Surface Average Winter Temperature Deep Maximum Salinity Deep Average Autumn Salinity Surface Maximum Chloropyll A Surface Average Spring Chloropyll A Surface Average Summer Chloropyll A Deep Minimum Chloropyll A Surface Minimum Chloropyll A	CSTOutlierPC1
Temperature variables selected following the results of initial linear mixed models	Deep Average Autumn Temperature Deep Average Spring Temperature Deep Average Summer Temperature Deep Average Winter Temperature Surface Average Autumn Temperature Surface Average Spring Temperature Surface Average Summer Temperature Surface Average Winter Temperature	AllEnvOutlierPC1, CSTOutlierPC1

Table 3.4. (A) Matrix of the number of outlier loci detected in *P. magellanicus* out of 7163 total loci by the methods BayEnv2 and LFMM using two environmental datasets, AllEnv and CST. The number of loci shared between different environmental datasets and programs are italicized. (B) Combined total number of loci detected from two methods, BayEnv2 and LFMM, using two environmental datasets, AllEnv and CST. The number of loci shared between the different environmental datasets is italicized.

(A)

		BayEnv2		**LFMM**	
		AllEnv	**CST**	**AllEnv**	**CST**
BayEnv2	**AllEnv**	128	*37*	*18*	
	CST		72		*5*
LFMM	**AllEnv**			511	*218*
	CST				218

(B)

	AllEnv	**CST**
AllEnv	621	*250*
CST		285

Table 3.5. Cumulative Akaike Information Criterion model weights ($\Sigma\ \omega_i$) and model-averaged parameter estimates (full: variables were assumed to be present in all models but with a coefficient of 0 in some cases; subset: variables were only present in models where the coefficient was not 0) in models predicting whether genetic variation in outlier loci among populations of *P. magellanicus* is a function of environmental variation. (A) Outlier loci were detected through correlations with an environmental dataset of 90 variables (AllEnv, n = 621 loci). Environmental were selected following the results of principal components analysis and redundancy analysis. (B) Outlier loci were detected through correlations with an environmental dataset of 36 variables (CST, n = 285 loci). Environmental variables were selected following the results of principal components analysis and redundancy analysis. (C) Outlier loci were detected through correlations with an environmental dataset of 90 variables (AllEnv, n = 621 loci). Environmental variables were selected following the results from (A) and (B). (D) Outlier loci were detected through correlations with an environmental dataset containing a subset of all available environmental variables (CST, 36 variables, n=285 loci). Environmental were selected following the results from (A) and (B).

			Model averaged parameter estimates	
	Parameter	**$\Sigma\ \omega_i$**	**Full**	**Subset**
(A)	SurfAvWinTemp	0.426	0.6767452	1.5889694
	SurfMinTemp	0.315	0.1673587	0.5337035
	SurfAvAutSal	0.272	0.2465678	0.9065121
	DepMinSiO_4	0.256	0.2009489	0.7841225
	DepAvAutSal	0.152	0.02467925	0.16288929
(B)	SurfAvWinTemp	0.437	0.5586436	1.2785748
	DepAvWinTemp	0.344	0.5687936	1.6505098
	DepMaxSal	0.343	-0.4611794	-1.3452456
	DepMinTemp	0.275	0.3441115	1.2523863
	DepAvAutSal	0.109	-0.001612218	-0.014871888
(B)	SurfAvWinTemp	0.643	1.031135	1.604291
	DepAvWinTemp	0.282	0.04939479	0.17471122
	SurfAvAutTemp	0.142	0.1584858	1.0963318
	SurfAvSumTemp	0.141	-0.09856687	-0.69285742
	DepAvSprTemp	0.103	0.04809931	0.45494143
	DepAvAutTemp	0.095	0.04477454	0.46824153
	DepAvSumTemp	0.083	0.01956854	0.23171602
	SurfAvSprTemp	0.080	-0.01290984	-0.1627252

(D)	SurfAvWinTemp	0.587	0.6509583	1.10358
	DepAvWinTemp	0.302	0.1932762	0.6457666
	DepAvSprTemp	0.131	0.004951037	0.067160279
	SurfAvAutTemp	0.121	0.08023971	0.5923327
	SurfAvSumTemp	0.095	0.01497057	0.16543456
	DepAvAutTemp	0.092	-0.02916119	-0.30036982
	DepAvSumTemp	0.089	0.02933703	0.312122
	SurfAvSprTemp	0.072	0.07923188	0.63938784

3.7 Figures

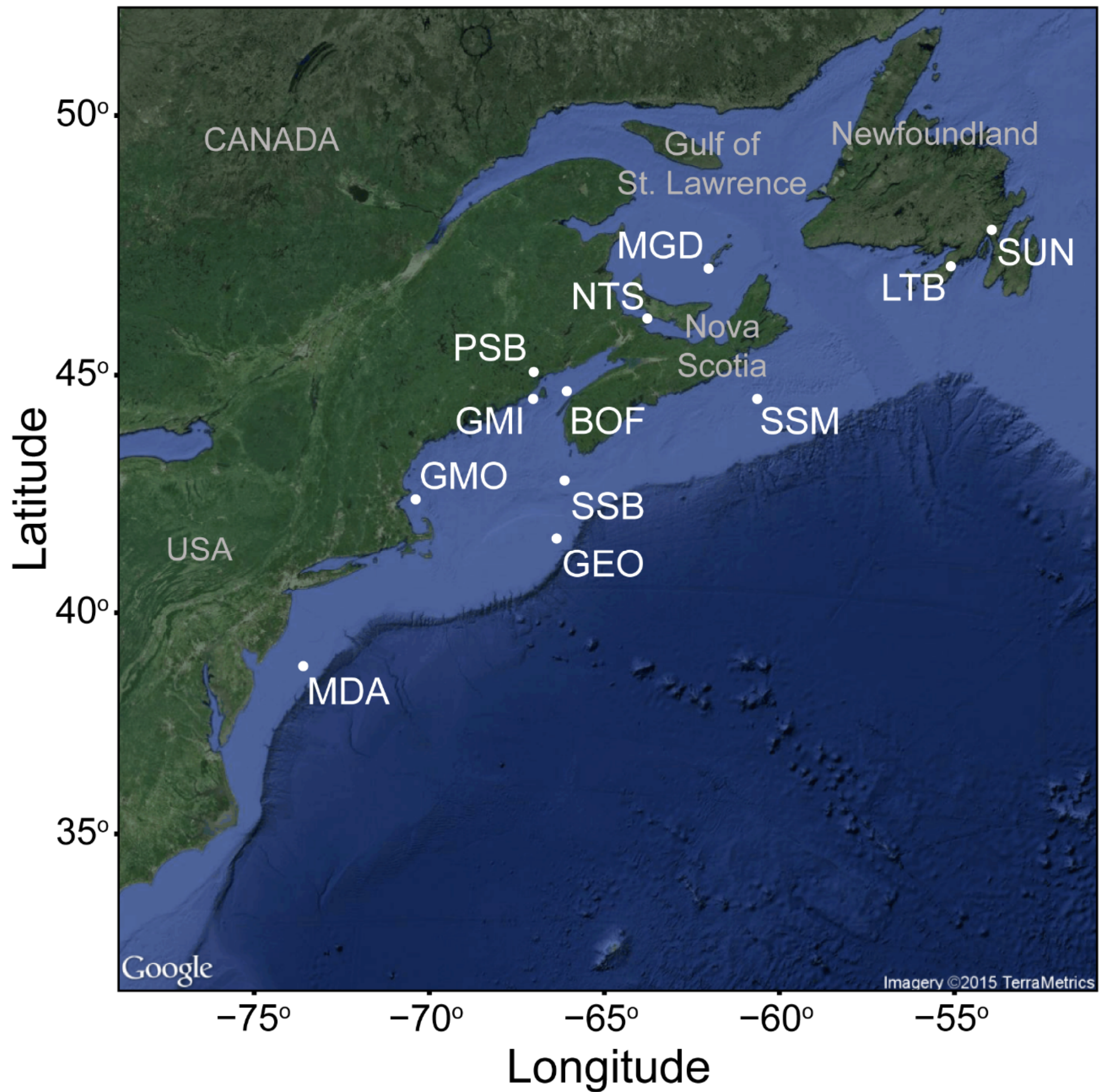

Figure 3.1. Map of sea scallop collection locations from the Northwest Atlantic. Site MDA (Mid-Atlantic Bight) represents the middle of several nearby collection locations grouped as one population. Population codes are defined in Table 3.1.

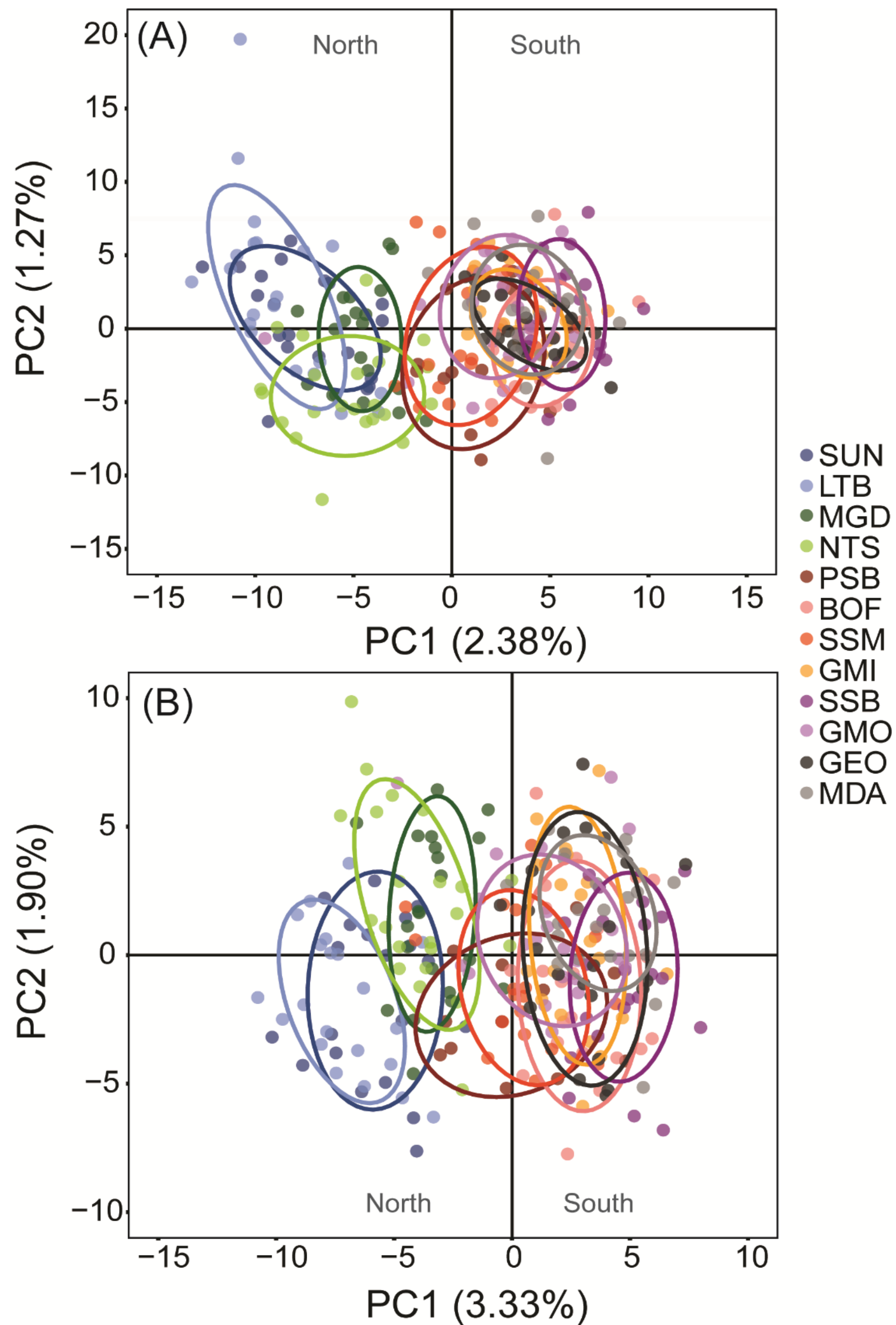

Figure 3.2. Principal components analysis plots for loci detected as potentially under selection through environmental correlation with (a) AllEnv (90 environmental variables, n= 621 loci), (b) CST (36 environmental variables, n = 285 loci) in 12 populations of *P. magellanicus*.

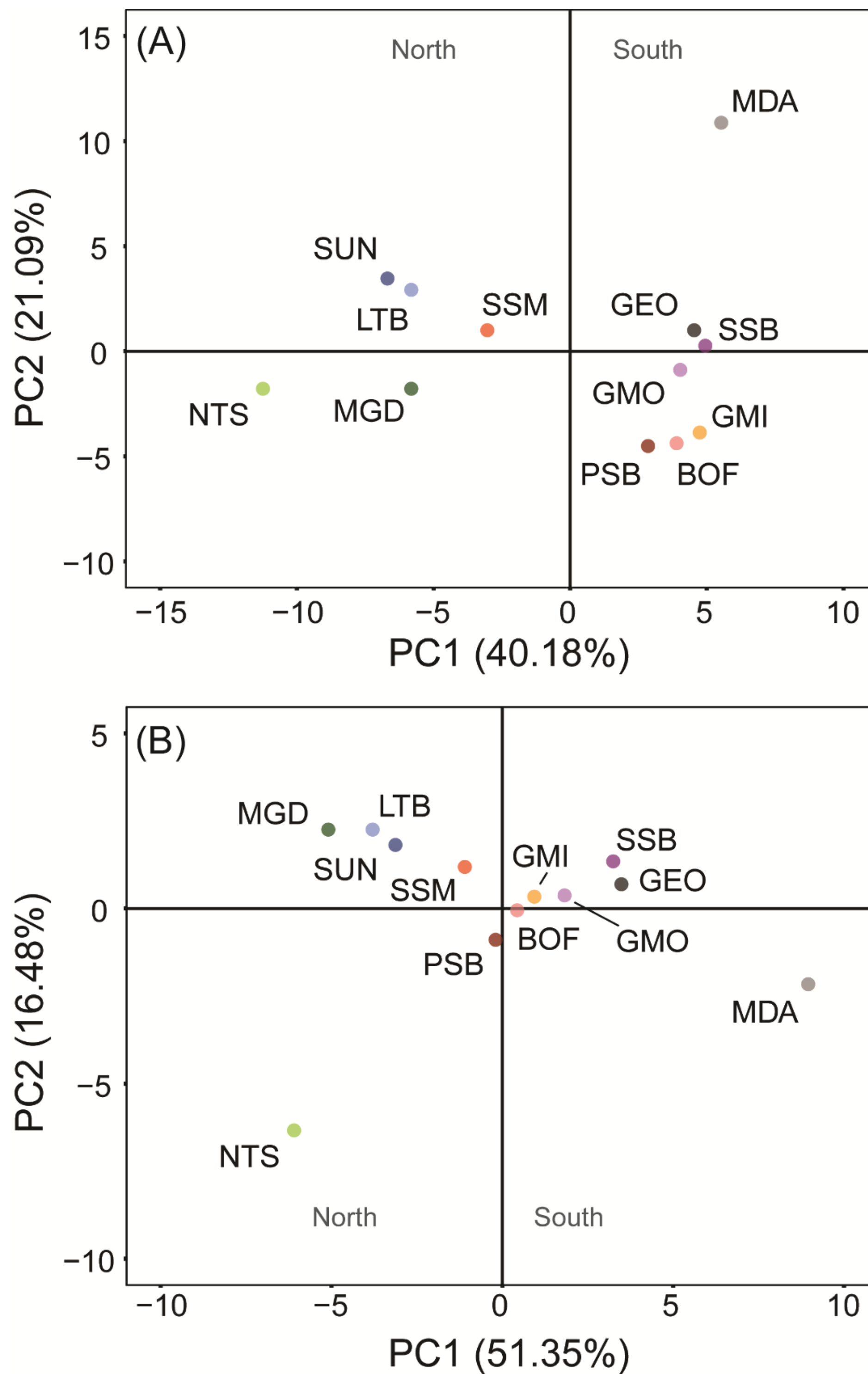

Figure 3.3. Principal components analysis plots for environmental variables used in the detection of adaptation among 12 populations of *P. magellanicus*. Environmental variables were separated into two datasets, (a) AllEnv (90 environmental variables) and (b) CST (36 environmental variables).

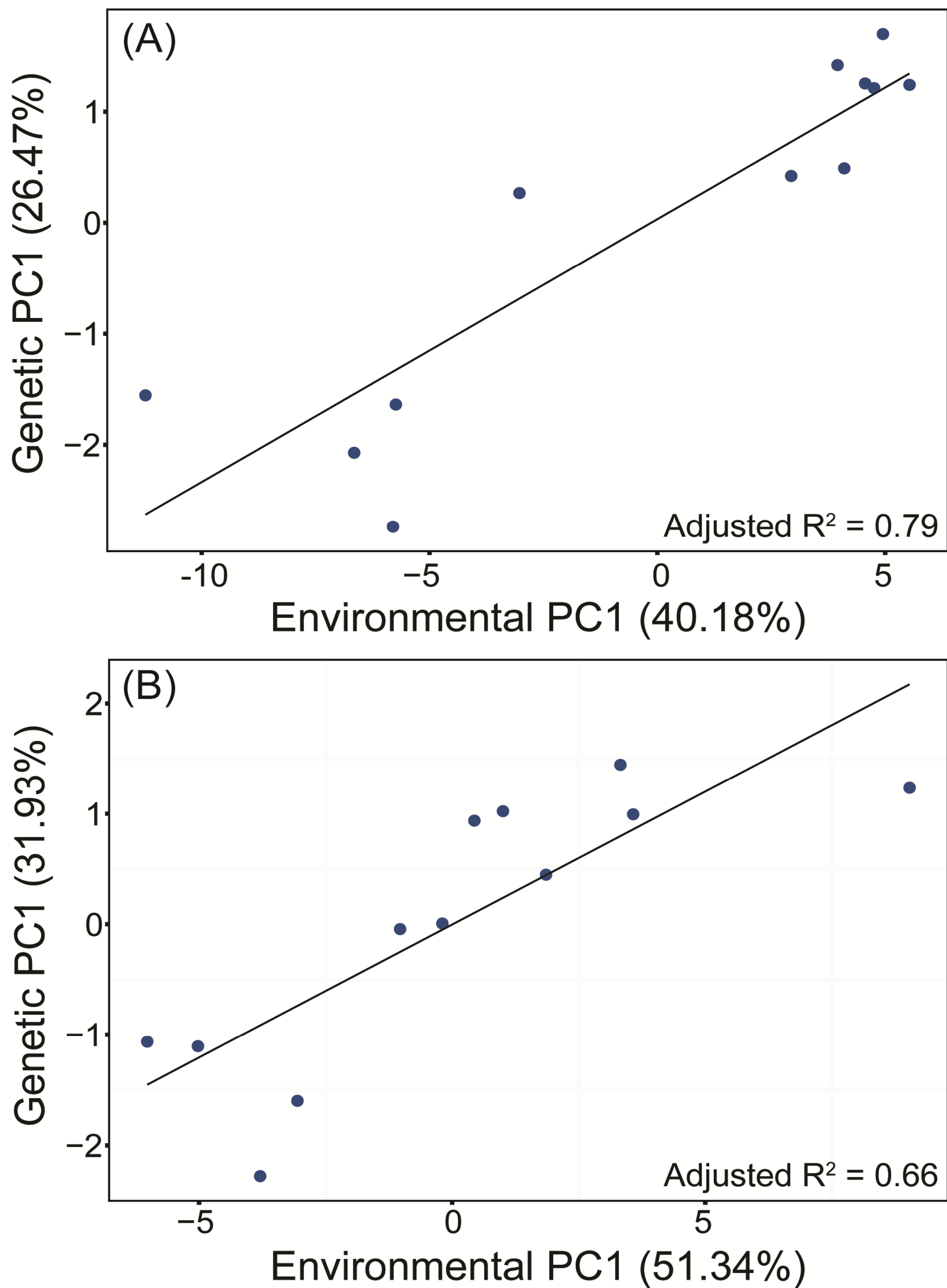

Figure 3.4. Linear regressions between the first principal component of PCA on population specific allele frequencies (Genetic PC1) and population specific environmental parameter values (Environmental PC1) for 12 populations of *P. magellanicus* for (a) AllEnv (90 environmental variables, 621 loci), and (b) CST (36 environmental variables, 285 loci).

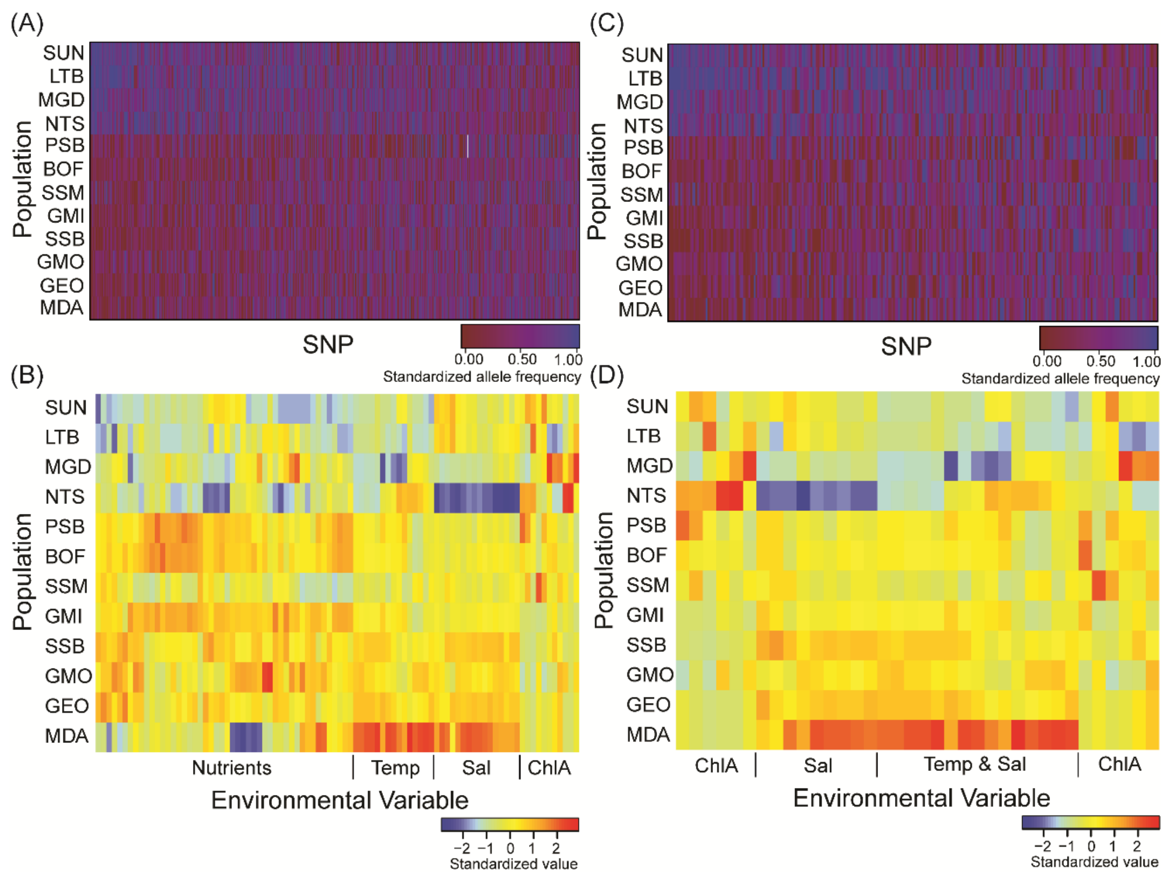

Figure 3.5. Heat map of (A) standardized allele frequencies (AllEnvOutlier, 621 loci), (B) standardized environmental variable value (AllEnv, 90 variables), (C) standardized allele frequencies (CSTOutlier, 285 loci), and (D) standardized environmental variable value (CST, 36 variables) for 12 populations of *P. magellanicus*. Loci in (A) were selected as potentially under selection through correlation with environmental variables in (B). Loci in (C) were selected as potentially under selection through correlation with environmental variables in (D). SNPs in (A) and (C) are arrange in order of strongest to weakest differentiation pattern.

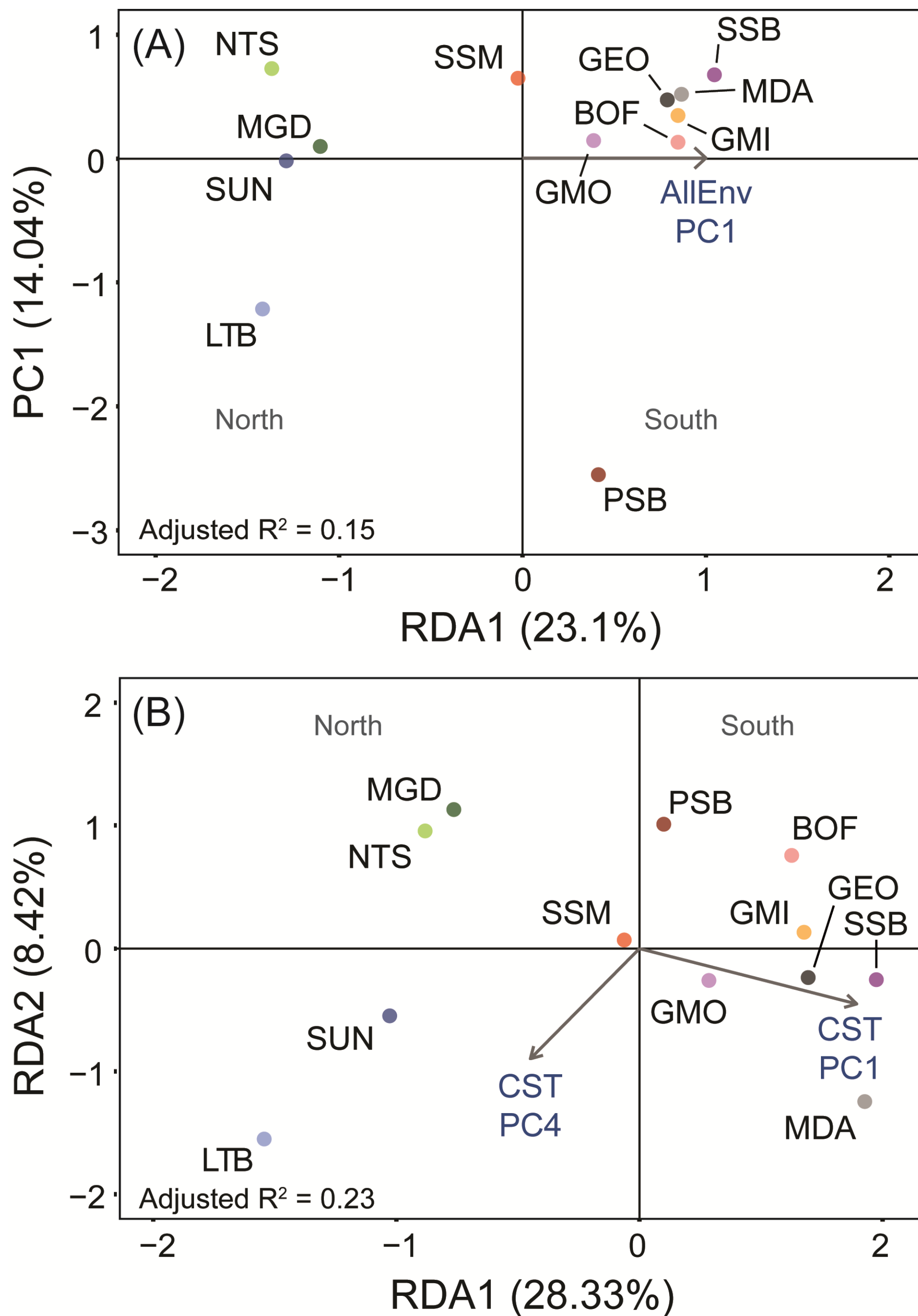

Figure 3.6. Redundancy analysis plots for loci detected as potentially under selection through environmental correlation with (a) AllEnv (90 environmental variables, n= 621 loci), (b) CST (36 environmental variables, n = 285 loci) in 12 populations of *P. magellanicus*. Explanatory variables were principal components axes from PCA on (a) AllEnv and (b) CST, retained following backwards stepwise variable selection.

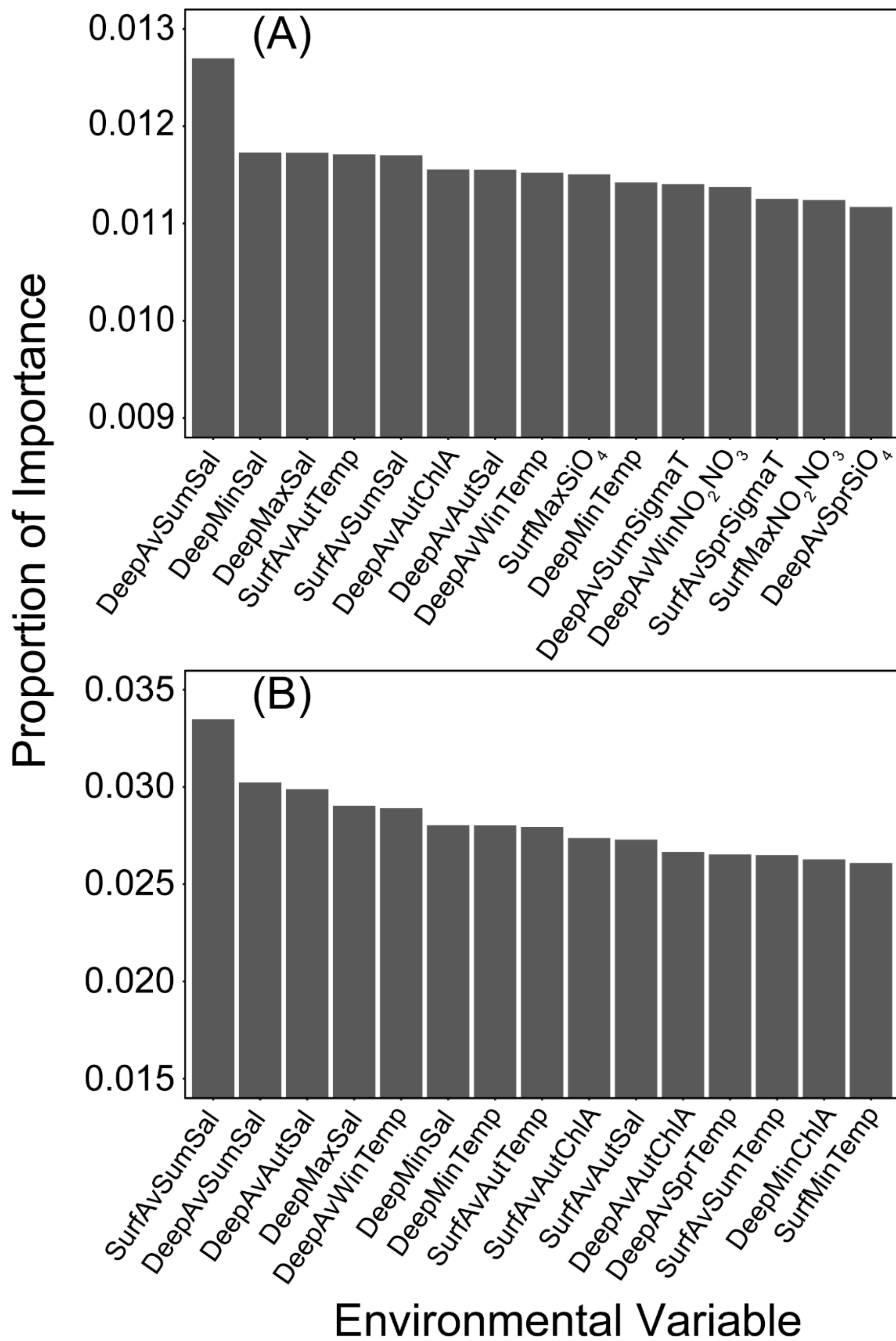

Figure 3.7. Proportion of importance (average per variable importance / importance sum of all variables) for the top 15 environmental variables determined using Random Forest and (A) AllEnv and AllEnvOutlier, and (B) CST and CSTOutlier.

Chapter 4: Countergradient variation in larval growth may be an adaptive strategy for sea scallops, *Placopecten magellanicus*, from northern cold-water populations

4.1 Abstract

Environmental heterogeneity can drive adaptive divergence among populations. In the marine environment, large geographic ranges increase the potential for differences in population-specific environmental exposure. We performed common garden larval rearing experiments comparing two populations of the sea scallop *Placopecten magellanicus*, a broadcast spawning marine bivalve that spans a large geographic range. These populations span over 1200 km of ocean characterized by an approximately 2 °C difference in average annual bottom temperature. ANCOVA on growth and survival versus experimental temperature indicated a higher overall growth rate in cold-water scallops ($P < 0.001$) and warm-water scallops exhibited higher overall survival ($P < 0.001$), however, we found no growth or survival responses to temperature in either population ($P > 0.05$). These results point to potential countergradient variation in response to cold winter temperatures experienced by the most northern population. This finding provides further support for previous studies of genomic differentiation in this species that found associations between genomic variation and ocean temperature. In conjunction with genomic data, these results provide evidence of temperature-associated adaptation over fine spatial scales between sea scallop populations and have implications for future management and conservation in sea scallops and other marine species.

4.2 Introduction

Gradients in environmental parameters over latitude, distance, or time often expose populations along a species' range to vastly different conditions, potentially leading to local adaptation. Environmental gradients can be especially prevalent in marine systems given the large geographic range of many marine species (Hauser & Carvalho 2008). Historical perspectives on the marine environment assumed limited adaptive differentiation resulting from large population sizes that would prevent genetic drift from promoting differentiation and presumed high levels of gene flow between marine populations (Hauser & Carvalho 2008), however, studies increasingly show evidence of fine-scale (<100 km) local adaptation (e.g. Clarke *et al.* 2010; Hutchings *et al.* 2007; Limborg *et al.* 2012; Sjöqvist *et al.* 2015). Understanding the patterns of adaptation in the marine environment can be particularly important in exploited marine species (Allendorf *et al.* 2010; Conover *et al.* 2006; Sale *et al.* 2005). By identifying adaptive divergence and the environmental pressures driving this divergence, researchers and managers can make informed decisions regarding management plans and begin to predict species response to changing environmental pressures and continued exploitation.

Recently, improvements in genetic and genomic technologies have enabled the detection of adaptation among populations through large-scale genetic and genomic studies, often utilizing landscape genetics techniques (Forester *et al.* 2015; Rellstab *et al.* 2015). Landscape genetic analyses have demonstrated environmental associated adaptation in many species, including ectothermic marine organisms (Berg *et al.* 2015; Bradbury *et al.* 2014b; Bradbury *et al.* 2010; Hecht *et al.* 2015; Limborg *et al.* 2012; Milano *et al.* 2014; Pespeni & Palumbi 2013). Evidence of adaptive diversity supports the

prediction of more frequent adaptation than expected in marine populations assuming high levels of gene flow in the marine environment (Cowen *et al.* 2000; Hauser & Carvalho 2008; Thorrold *et al.* 2002). Studies of adaptation using genetic methods can provide insight into the adaptive dynamics of marine populations, however, Rellstab *et al.* (2015) (and others) suggest combining both landscape genetic and more traditional experimental approaches to detect adaptation. This combination offers a robust strategy to detect adaptation among populations and identify the sources of adaptive pressure that may lead to genetic and phenotypic divergence.

Experimental methods of detecting adaptation focus on identifying population specific differences in reaction norms which can evolve in response to different selective influences (De Jong 2005). Experiments on reaction norms generally utilize common-garden rearing protocols, which allow comparison of different populations and families reared under identical conditions. Studies of genetic variation among populations and the link between genetic variation and the slope and intercept of the reaction norm (Gutteling *et al.* 2007; Jensen *et al.* 2008; Van Asch *et al.* 2007; Winterhalter & Mousseau 2007) validate the presence of local adaptation. The logistical challenges of raising experimental populations, especially marine organisms with small larvae, has focused much of the work on reaction norm analysis using common garden experiments on terrestrial organisms such as *Drosophila sp.* (Liefting *et al.* 2009). The last decade, however, has seen increased analyses on experimental reaction norms in a variety of marine organisms (Conover & Present 1990; Hutchings *et al.* 2007; Jensen *et al.* 2008; Oomen & Hutchings

2015; Yamahira *et al.* 2007), providing unique insight into the scale of marine adaptation (Cowen *et al.* 2000; Hauser & Carvalho 2008; Thorrold *et al.* 2002).

Among environmental variables, temperature often contributes significantly to local adaptation, particularly for ectothermic organisms in which temperature affects numerous morphological and developmental traits (Roff 2002) by influencing metabolic and physiological reactions at all levels (Yamahira *et al.* 2007). In marine systems, temperature variation with latitude often dominates environmental drivers, and the large ranges and population sizes of many species may result in different populations within the same species experiencing widely different environmental conditions (e.g. Bradbury *et al.* 2010; Van Wyngaarden *et al. in review*). Here we explored the presence of temperature associated adaptive differences among northern and southern populations of the sea scallop, *Placopecten magellanicus*, using common-garden experiments with larvae.

Previous work on populations from these regions identified two population clusters, one northern cluster (generally cold-water) and one southern cluster (generally warm-water) (Van Wyngaarden *et al.* 2017). Data on the temperature variation between these populations was collected for Van Wyngaarden *et al.* (*in review*) and demonstrates that although the yearly average temperature between our populations does not significantly differ, bottom temperatures vary more in our northern cluster population which also has a much colder winter depth temperature (roughly December to April). than the southern cluster population. Based on the genetic differences between scallop populations at the north and south of their range (Kenchington *et al.* 2006; Owen & Rawson 2013; Van Wyngaarden *et al.* 2017) , correlations between genetic differentiation

and environmental variation (Van Wyngaarden *et al. in review*), strong temperature gradients along the species range (Townsend *et al.* 2006), and the detection of thermal reaction-norm adaptation in other species in the same geographic region (Conover & Present 1990; Hutchings *et al.* 2007; Purchase & Brown 2000), we predicted population-level differences in the growth and survival thermal reaction norms of larval sea scallops.

4.3 Methods

4.3.1 Study species

The sea scallop occurs along the east coast of North America, from Newfoundland, Canada in the north to Cape Hatteras, USA in the south (Posgay 1957). It reproduces via broadcast spawning and, like many broadcast spawners, sea scallops exhibit extremely high early life mortality. The long planktonic larval stage (approximately 30 days) likely exacerbates mortality (Naidu & Robert 2006). Scallops typically begin spawning at approximately two years old, generally spawning when the water is warmest (Thompson 1977) between August and October (e.g. Beninger 1987; DiBacco *et al.* 1995; Langton *et al.* 1987; Naidu 1970; Posgay & Norman 1958), however, a smaller springtime spawning may also occur in several regions of the population range (Dadswell & Stokesbury 2007; Kirkley & Dupaul 1991; Macdonald & Thompson 1988; Schmitzer *et al.* 1991).

4.3.2 Experimental conditions

All experiments and algal culturing were completed at the Bedford Institute of Oceanography in Dartmouth, Nova Scotia, Canada between May and November 2014.We

conditioned all equipment in filtered seawater for several weeks prior to the beginning of the experiment. Prior to and during the experiment, all equipment was washed between uses with a 0.03% bleach solution, rinsed thoroughly with freshwater, and given a final rinse in filtered seawater before exposure to any animals to minimize contamination between experimental treatments and disease transmission between scallop families. Algae were cultured in sterile flasks (SF) and carboys (SC) as well as large, open-air, aerated tubes for larval and adult scallop consumption. Larval scallops were only fed algae from SF and SC in an effort to prevent contamination and disease transmission. We prepared and autoclaved filtered sea water with a mixture of required nutrients and silicates (for diatoms), in small and large flasks and carboys. In a clean fume hood using sterile pipettes, we transferred algae from an existing flask colony to a new SF or SC and grew them in a warm room with ample light and cropped them in log growth phase for feeding scallop larvae. The large open-air tubes were cropped daily to feed adult scallops during conditioning, and replenished with filtered, pasteurized seawater and nutrients. We verified algal cell concentration with a hemocytometer initially and then verified colour visually for subsequent feedings.

4.3.3 Adult scallop collection and conditioning

We collected adult sea scallops from the Northumberland Strait (NTS, June, northern population cluster) and the Bay of Fundy (BOF, July, southern population cluster) in the early summer of 2014 (Figure 4.1). Using data and following protocols from Van Wyngaarden *et al.* (*in review*), we estimated collection depths at approximately 20 m for NTS and between 30 m and 140 m for BOF. Average annual temperature at

depth did not significantly differ between these sites, however, the temperature range was greater at NTS than BOF (approximately 16 °C annually at NTS and 9 °C annually at BOF) and the winter temperatures (December to April) were significantly colder at NTS (-0.7 °C) than BOF (5.0 °C) (Welch two-sample t-test, $P < 0.001$).

Male and female scallops were identified and separated into 300L flow-through tanks supplied with filtered seawater from the Bedford Basin at ambient temperatures. We conditioned scallops over several weeks to prepare them for spawning, siphoning tanks every other day to remove feces and other detritus, monitoring temperature and salinity with a YSI 30 probe, and checking pH using a minilab pH meter. Temperature ranged from 4.0 to 11.9 °C, salinity ranged between 27.7 and 32.1 psu, and pH ranged between 7.6 and 8.2 over the course of conditioning for both populations (Figure 4.2). We adjusted the photoperiod for NTS scallops from the June average for the collection location (approximately 15h45m of daylight) to the late August average (approximately 13h20m of daylight) over a period of three weeks because photoperiod has been shown to affect the sea scallop reproductive cycle and influence conditioning for reproduction (Couturier and Aiken 1989). Photoperiod for BOF scallops was changed from the July average (approximately 15h0m of daylight) to the late-August average (approximately 13h20m of daylight) over one week (TheWeatherCompany 2016). Once scallops reached their final photoperiod, we held hours of daylight steady until spawning. Adult scallops were fed approximately 1L of *Tetraselmis sp., Pavlova lutheri,* and *Chaetoceros muelleri* algal culture/scallop/day at a concentration of approximately 300 000 algal cells/mL in the algal culture. Feeding took place over several sessions per day to allow scallops to

efficiently clear algal particles, with flow-through seawater temporarily stopped for approximately 30 minutes after algal addition. Overnight and on weekends, peristaltic pumps added algae to adult tanks at a constant rate. We filtered sea water for spawning and larval growth more thoroughly than for adult scallops by running incoming sea water through a series of filters (20µm, 5µm, and ceramic) and a UV sterilizer into a large 120 L container bubbled with filtered air for one hour. After bubbling, we filled a clean 0.5 L container to test temperature, salinity, and pH of the further filtered seawater (FS). An aquarium heater adjusted temperature to ± 1 °C of the experimental temperature as necessary (10 °C, 13 °C, and 16 °C).

4.3.4 Spawning

We induced adult scallops to spawn once gonads appeared ripe [approximately stage VII in Naidu (1970)]. NTS scallops spawned between 8 July and 16 July 2014, and BOF scallops spawned between 17 September and 19 October 2014. Adults were removed from their conditioning tanks, cleaned, and exposed to air for 15 minutes. Following air exposure, scallops were placed in a clean 4 L bucket and alternated between 4 °C and 16 °C seawater every 15 minutes for two hours or until spawning. We allowed an individual to spawn freely for up to 30 minutes once it began spawning, at which point we removed the scallop from the bucket, labelled it, and set it aside for sampling. Eggs were rinsed in FS through a 40 µm screen into a clean container. After the eggs were mixed well, we sampled to determine egg concentration by counting eggs under a microscope. Eggs were then added to a clean 4 L bucket at a concentration of ~500 eggs/mL. We strip spawned males that did not spawn naturally within 30 minutes

(strip spawning does not produce viable eggs in female scallops). Strip spawned sperm were allowed to activate in FS for 15 minutes, and we checked viability and counted sperm under a microscope. Sperm from one male was added to eggs from 1 female at a concentration of ~ 12:1 (Jones *et al.* 1996). Following fertilization, we split each family into three 4-L buckets and held them at 13 °C until day 3 post-spawn, the first appearance of D-stage larvae. Following spawning, adults were shucked to remove a small adductor muscle tissue sample for storage in 95% ethanol.

4.3.5 Larval rearing

At day 3 post-spawn, we confirmed the presence of D-stage larvae in all buckets. Larvae from a single family were then rinsed through a 40-µm sieve and combined into one container. Larvae were well mixed and then separated into three 4 L buckets, assigning each of these three buckets to a temperature treatment (10 °C, 13 °C, or 16 °C) and appropriate incubator, with each family replicated once in each temperature. We randomized bucket position on the incubator shelves, and placed quality control buckets on each shelf. Temperature, salinity, and pH were monitored daily from the quality control buckets which underwent the same cleaning and water changes as experimental buckets. Larval buckets alternated between a full clean and half clean every other day. To clean buckets, we filtered larvae from the bucket through an appropriately sized sieve (40-µm or 60-µm, depending on age and size of larvae) and placed the sieve in FS at the appropriate experimental temperature. The clean bucket was then refilled with 4 L of FS at the experimental temperature prior to gently rinsing in larvae. On half clean days, we removed only half of the water and refilled the bucket to 4 L with FS prior to gently

rinsing larvae back into the bucket from the sieve. Larvae were sampled every 2 days (every full clean day) for NTS and every 4 days (every other full clean day) for BOF. Once sieved from the bucket, larvae were rinsed into a clean 1 L container, and well mixed before we collected three 1 mL samples for each family and temperature treatment. Samples were fixed with buffered formalin on the day they were taken at a 9:1 water:formalin ratio. After sampling, we gently rinsed the remaining larvae back into their cleaned 4 L bucket with FS. Following daily cleaning, we immediately fed scallop larvae *Isochrysis galbana* starting on day three until day 13, and a mixture of *I. galbana* and *P. lutheri* from day 14 until the end of the experiment. We estimated algal concentrations from algal culture colour (consistently within the log growth phase) and added approximately 15 mL of algal solution from day three until day nine, and 25 mL of algal solution from day 10 until the end of the experiment to all scallop larval containers.

4.3.6 Larval sample analysis

The 1mL larval samples from day 3 and day 21 were mixed well with a pipette and placed under a microscope with attached camera. We collected images of each larva in each sample to count the number of larvae per sample and measure larval height using ImageJ (Rasband 1997-2015) (height defined as length of the larva perpendicular to the straight edge of the shell, Figure 4.3). We averaged larval height and number of larvae per sample over triplicate samples for day 3 and day 21 and calculated total growth (shell height at day 21 – shell height at day 3) and percent survival [(density at day 21 / density at day 3)*100] for each temperature and family. Partial samples provided larval size measurements but not survival estimates. Larval growth values were log transformed to

reduce heteroscedasticity in the data. We generated linear models in R (R Development Core Team 2012) separately for growth and survival, with Log_{10}Growth and Survival as response variables, and Temperature, Population, and the Temperature*Population interaction as fixed explanatory variables.

4.4 Results

All results are presented as mean ± standard deviation where available. During conditioning, we exposed adults from NTS to temperatures ranging from 4.0 to 10.4 °C (average 6.7 ± 1.3 °C), salinities ranging from 30.7 to 32.1 psu (average 31.3 ± 0.3 psu), and pH ranging from 7.6 to 8.2 (average 7.9 ± 0.2). Temperatures for BOF adults ranged from 5.8 to 11.9 °C (average 9.1 ± 1.5 °C), salinities from 27.7 to 31.7 psu (average 30.7 ± 0.6 psu), and pH from 7.7 to 7.9 (average 7.8 ± 0.1) (Figure 4.2).

During the common garden-like experiment, we exposed NTS larvae in the 10 °C temperature trial to an average temperature of 10.2 ± 1.9 °C, an average salinity of 30.9 ± 1.5 psu, and an average pH of 8.0 ± 0.1, whereas BOF larvae were exposed to 9.7 ± 1.2 °C, 31.0 ± 0.3 psu, and a pH of 8.0 ± 0.1. In the 13 °C trial, we exposed NTS larvae to an average temperature of 13.9 ± 1.8 °C, an average salinity of 30.9 ± 1.1 psu, and an average pH of 8.1 ± 0.1, in contrast to exposing BOF larvae to 13.2 ± 0.2 °C, 31.0 ± 0.2 psu, and a pH of 8.0 ± 0.1. In the 16 °C trial, we exposed NTS larvae to an average temperature of 16.2 ± 0.5 °C, an average salinity of 30.9 ± 1.4 psu, and an average pH of 8.0 ± 0.1, whereas BOF larvae were exposed to 16.7 ± 0.7 °C, 31.3 ± 0.2 psu, and a pH of 8.0 ± 0.1 (Table 4.1, Figure 4.4).

4.4.1 Growth

Final analysis of growth included six families from NTS (2, 3, 5, 7, 8, 9) and eight families from BOF (1, 2, 4, 5, 6, 7, 8, 9). The average number of larvae per sample measured was 12.9 ± 8.8. Average growth of BOF over the trial at 10 °C was 12.0 ± 7.1 µm, 12.1 ± 6.0 µm at 13 °C, and 11.6 ± 4.0 µm at 16 °C. Average growth of NTS over the trial was 19.6 ± 4.5 µm at 10 °C, 22.6 ± 6.3 µm at 13 °C was, and 24.0 ± 6.2 µm at 16 °C . ANOVA on the linear model of Log_{10}Growth found significant Population level differences in overall growth (F = 37.93, P = < 0.001) but showed no significant Temperature*Population interaction (P > 0.05, Table 4.2). Population specific estimates of reaction norm slope and intercept were not significantly different between NTS (intercept: 1.14 ± 0.34, slope: 0.01 ± 0.03) and BOF (intercept: 1.00 ± 0.37, slope: 0.00 ± 0.02) (Welch two-sample t-test, P > 0.05), and, in both cases, the slope estimate contained 0, indicating a limited interactive effect of temperature on larval growth (Figure 4.5A).

4.4.2 Survival

We could not analyze survival in several families that we analyzed for growth because of a sampling error, and therefore used six families from NTS (2, 3, 5, 6, 7, 8) and six from BOF (1, 5, 6, 7, 8, 9) in the final survival analysis. The average number of larvae per sample measured was 14.4 ± 9.4. Average survival of BOF over the trial was 74.4 ± 16.3% at 10 °C, 65.0 ± 8.0%, at 13 °C and 72.4 ± 21.5% at 16 °C . Average survival of NTS over the trial was 35.0 ± 21.0% at 10°C, 41.8 ± 17.3%, at 13 °C and 43.1 ± 4.8% at 16 °C. ANOVA on the linear model of survival identified Population level differences (F = 17.34, P = < 0.001), but showed no significant Temperature*Population interaction (P > 0.05, Table 4.3). Population specific estimates of survival reaction norm

slope and intercept did not differ significantly between NTS (intercept: 0.02 ± 1.03, slope: 0.03 ± 0.08) and BOF (intercept: 0.81 ± 0.66, slope: -0.01 ± 0.66) (Welch two-sample t-test, $P > 0.05$) and, as observed with the growth results, both slope estimates contained 0 indicating a limited effect of temperature on larval survival (Figure 4.5B).

Overall, families within each population varied greatly in both growth and survival, as can be seen by the within-Population Sum of Squares and in Figure 4.6, likely contributing to the lack of population level differences in the slopes of the reaction norms.

4.5 Discussion

Selection associated with variation in the environment experienced by different populations can drive local adaptation. Using sea scallops from populations previously shown to differ genetically (Van Wyngaarden *et al.* 2017), we performed a common-garden experiment examining larval growth and survival in response to temperature. We found no effect of temperature on larval survival or growth in either population, however, the more northern population showed overall higher growth rates whereas the more southern population showed overall higher survival. Although larval life stages did not respond to temperature variation, these results suggest evidence of local adaptation through possible countergradient variation (Conover & Present 1990), where higher latitude populations (or those from colder habitats) grow faster than lower latitude populations. Although further study is required to detect adaptive variation in early larval stages of the sea scallop, our results considered in tandem with previously identified genetic differences between northern and southern scallop population groups (Van Wyngaarden *et al.* 2017) and the likely influence of winter temperatures on effective

connectivity among scallop populations (Van Wyngaarden *et al. in review*) suggests adaptive differentiation occurs among scallop populations and should be incorporated into management strategies.

4.5.1 Adaptation in marine environments

Genetic and experimental studies continue to accumulate evidence of fine-scale adaptation among marine populations (e.g. Guo *et al.* 2015; Haugen & Vøllestad 2000; Hutchings *et al.* 2007; Limborg *et al.* 2012; Sjöqvist *et al.* 2015). Many regions exert a selective force that influences many species [for example, salinity gradients in the Baltic Sea (Berg *et al.* 2015; Sjöqvist *et al.* 2015) and the Mediterranean Sea (Milano *et al.* 2014)] and in the Northwest Atlantic, the dominant selective pressure appears to be temperature. Temperature varies particularly strongly with latitude in the Northwest Atlantic along the coast of the United States and Canada where the cold Labrador Current meets the warm Gulf Stream (Townsend *et al.* 2006). Many species ranges span this strong temperature gradient, and studies report both reaction norm variation (Conover & Present 1990; Hutchings *et al.* 2007; Purchase & Brown 2000) and genetic structure (Benestan *et al.* 2015; Bradbury *et al.* 2010; Van Wyngaarden *et al.* 2017) for species in the region.

In ectothermic organisms, temperature can drive adaptation through effects on metabolic and physiological reactions (Yamahira *et al.* 2007). Many studies on marine ectotherms have focused on thermal reaction norms and found evidence of adaptive phenotypic plasticity among populations, particularly in larval growth rates and survival (Conover & Present 1990; Haugen & Vøllestad 2000; Hutchings *et al.* 2007; Jensen *et al.*

2008; Oomen & Hutchings 2015; Yamahira *et al.* 2007). These adaptations generally fall into two categories as described in Conover & Present (1990) and Yamahira *et al.* (2007): populations from cooler-water regions may experience a shift in a trait maximum to lower temperatures (i.e. maximum values for a trait such as growth may occur at a lower temperature), populations from cooler-water regions may increase in trait performance overall (populations may exhibit higher growth potential at all temperatures), or populations may show a combination of the two strategies.

4.5.2 Thermal adaptation in the sea scallop

We found no evidence for variation in thermal response between NTS and BOF, however across all temperatures NTS had higher growth than BOF. This higher overall growth in NTS may exemplify countergradient variation (Conover & Present 1990) as an adaptive mechanism to compensate for shorter growing seasons (Yamahira *et al.* 2007) and a colder winter growth period in NTS than in BOF. Such patterns are reported in Atlantic silversides (Conover & Present 1990) and Atlantic cod (Purchase & Brown 2000) in the northwest Atlantic and the Pacific silversides in the northeast Pacific (Baumann & Conover 2011). There is evidence that scallop growth rates vary among locations (Claereboudt & Himmelman 1996; Harris & Stokesbury 2006), and although Macdonald & Thompson (1988) found similar growth rates in adult growth rates across latitudes, limited work has focused on larval scallops. Although previous work on this region has reported similar average autumn temperatures in both NTS and BOF and planktonic larval scallops from NTS likely experience temperatures no colder than those from BOF, NTS nonetheless shows greater variation in autumn temperatures with both

higher maximum temperatures and lower minimum temperatures (DFO 2015; Gregory 2004; Van Wyngaarden *et al. in review*). NTS also experiences much colder winter temperatures than BOF, so faster growth rate in larval scallops from NTS may represent a response to a short post-settlement pre-winter period rather than a colder growing season. Scallop larvae metamorphose and settle to the sea floor at approximately 200 µm (Culliney 1974), suggesting that a higher growth rate could shorten larval duration. A reduced larval duration at NTS may allow settled NTS scallops a longer period to build up energy stores to survive their first cold winter, compared to BOF scallops that experience less extreme winter temperatures. Alternatively, an earlier spawning season may allow NTS scallops a longer preparatory period, although reports on spawning time between scallop populations are conflicting likely due to high inter-annual variability (e.g. Beninger 1987; Macdonald & Thompson 1988). In other marine species in the Northwest Atlantic, over-winter mortality has been shown to vary with latitude, likely driven by salinity and temperature variations (Bauer & Miller 2010). Overall, these differences support recent landscape genetic studies on this species which highlighted winter temperatures as a possible structuring force in the strong split between northern and southern scallop populations (Van Wyngaarden *et al. in review*).

As with growth, our results show no evidence of a response to temperature in larval survival. There are overall population level differences, however, with higher survival in BOF larvae at all temperatures. Like broadcast spawning species, scallops invest energy in producing large numbers of gametes, and massive mortality and loss of numbers characterize the subsequent larval period (Naidu & Robert 2006; Thorson 1950).

Growth may be favoured at the expense of early larval survival in NTS to ensure that larvae settle out of the water column to become juveniles as quickly as possible. In contrast, BOF scallop larvae may be less constrained to grow quickly and scallops that allocate more energy to survival at a young larval stage may nonetheless survive until reproductive age. The lack of GEI detected in early larval survival in our experiment suggests that post-settlement mortality may structure sea scallop populations more than early life mortality, although further evidence is needed. This may mirror size-dependent winter mortality commonly seen in marine fish populations (e.g. Beamish *et al.* 2004; Schultz *et al.* 1998), and previous studies identify post-settlement mortality as a selective force in several other broadcast spawning marine invertebrates in the context of adaptation to salinity in *Mytilus edulis* (blue mussel) (Koehn *et al.* 1976; Koehn *et al.* 1980) and *Semibalanus balanoides* (acorn barnacle) (Schmidt *et al.* 2000).

4.5.3 Limitations and implications

High variability in growth rate among families clearly limited the power to detect differences among populations. Interestingly, we observed growth rates in two BOF families similar to those in NTS families, with greatly reduced growth rates in the remaining six BOF families. We also observed some overlap in the survival data, with survival rates in one NTS family similar to that of BOF, however, overall survival rates were more similar between NTS and BOF than growth rates. The variability observed suggests the possibility of experimental error in the analysis which can be seen in the large within population sum of squares in our ANOVA. Our experiment introduced several possible avenues for error. The experiment was logistically complicated,

particularly given the necessity of inducing natural spawning in conditioned females. Because of this constraint, the trials ran at slightly different times of the year. Although we made every effort to ensure stability in food and environmental conditions across adult conditioning and larval trials, our monitoring suggests slight variation during the trials. In addition, the effects of conditioning duration on adults, variation in collection time, and possible bacterial or parasitic infection of larval families may have also added sources of error. Other factors may also have contributed to the lack of GEI in both sea scallop populations. In the survival estimates in particular, the more frequent sampling of NTS than BOF may also have influenced results, although the small number of larvae removed during sampling procedures decreases the likelihood this factor influenced differences between survival estimates. Finally, previous landscape genetic work on sea scallops suggests that the coldest temperatures experienced by a population influence population structure (Van Wyngaarden *et al. in review*). If, as suggested here and elsewhere, winter temperatures are an important driving force in scallop population differentiation, the experimental temperatures used during our trials were potentially not low enough to elicit an effect in the scallop larvae. Temperatures were chosen within the range experienced in NTS and BOF during autumn, however, colder temperatures might have shown adaptive differences between the studied populations.

Even with experimental limitations, the results presented here provide further evidence of adaptive differentiation among sea scallop populations. Previous work on sea scallops showed differences in reproductive timing and fecundity among different populations (e.g. Barber *et al.* 1988; Beninger 1987; DuPaul *et al.* 1989; Naidu 1970, and

Kirkley & Dupaul 1991), evidence of genetic differentiation at large and small scales (Kenchington *et al.* 2006; Owen & Rawson 2013; Van Wyngaarden *et al.* 2017), and evidence of correlations between genetic structure and environmental differentiation (Van Wyngaarden *et al. in review*). Our results, building on previous experimental and observation examination of sea scallop populations, improve our understanding of the relationship between environmental variation, adaptation, and genetic population structure among Northwest Atlantic sea scallops.

4.5.4 Conclusions

Although we found no evidence of genotype-by-environment interaction in the reaction norms of larval growth and survival between the scallop populations studied, we nonetheless found evidence that countergradient variation occurs in the growth of larval sea scallops. NTS, our northern experimental population, experienced higher overall growth rates than our southern BOF population, which may indicate that NTS larvae are adapted for a shorter pelagic larval duration, possibly to allow settled scallops to prepare for the cold winters experienced by NTS. Experimental limitations may have prevented the detection of adaptive variation between sea scallop populations, however, the detected countergradient variation, in addition to previous work identifying adaptive genetic variation due to the coldest temperatures experienced by scallop populations (Van Wyngaarden *et al. in review*), suggests that post-settlement and overwinter mortality may play a stronger role than larval mortality in sea scallop population structure. These conclusions, and others from studies such as these, provide valuable information for

population management and can help managers effectively plan strategies and predict how species may respond to environmental change.

4.6 Tables

Table 4.1. Minimum, maximum, and average temperature, salinity, and pH experienced by *Placopecten magellanicus* larvae from two populations during a 21-day common garden larval growth experiment (average ± standard deviation in parentheses).

Population	Trial	Temperature (°C)	Salinity (ppt)	pH
NTS	10°C	2.6 - 12.2 (10.2 ± 1.9)	27.7 - 32.9 (30.9 ± 1.5)	7.9 - 8.3 (8.0 ± 0.1)
	13°C	12.9 - 21.0 (13.9 ± 1.8)	27.7 - 31.6 (30.9 ± 1.1)	7.9 - 8.4 (8.1 ± 0.1)
	16°C	14.3 - 16.8 (16.2 ± 0.5)	27.8 - 32.1 (30.9 ± 1.4)	7.9 - 8.3 (8.0 ± 0.1)
BOF	10°C	4.3 - 11.5 (9.7 ± 1.2)	30.4 - 31.5 (31.0 ± 0.3)	7.9 - 8.1 (8.0 ± 0.1)
	13°C	12.3 - 13.8 (13.2 ± 0.2)	30.5 - 32.0 (31.0 ± 0.2)	7.8 - 8.0 (8.0 ± 0.1)
	16°C	13.8 - 18.1 (16.7 ± 0.7)	30.8 - 32.0 (31.3 ± 0.2)	7.9 - 8.2 (8.0 ± 0.1)

Table 4.2. ANOVA table for a linear model examining the effects of temperature, population, and their interaction on the Log_{10}Growth of larval *P. magellanicus* from two populations.

Model Term	**DF**	**Sum of Squares**	**Mean of Squares**	**F**	**p-value**
Temperature	1	0.01605	0.01605	0.6917	0.4108
Population	1	0.87979	0.87979	37.9283	< 0.001*
Temperature x Population	1	0.0085	0.0085	0.3666	0.5485
Residual	38	0.88145	0.0232		

Table 4.3. ANOVA table for a linear model examining the effects of temperature, population, and their interaction on the survival of larval *P. magellanicus* from two populations.

Model Term	**DF**	**Sum of Squares**	**Mean of Squares**	**F**	**p-value**
Temperature	1	0.00569	0.00569	0.2249	0.6386
Population	1	0.84373	0.84373	33.3673	< 0.001 *
Temperature x Population	1	0.0153	0.0153	0.6051	0.4423
Residual	32	0.80916	0.02529		

4.7 Figures

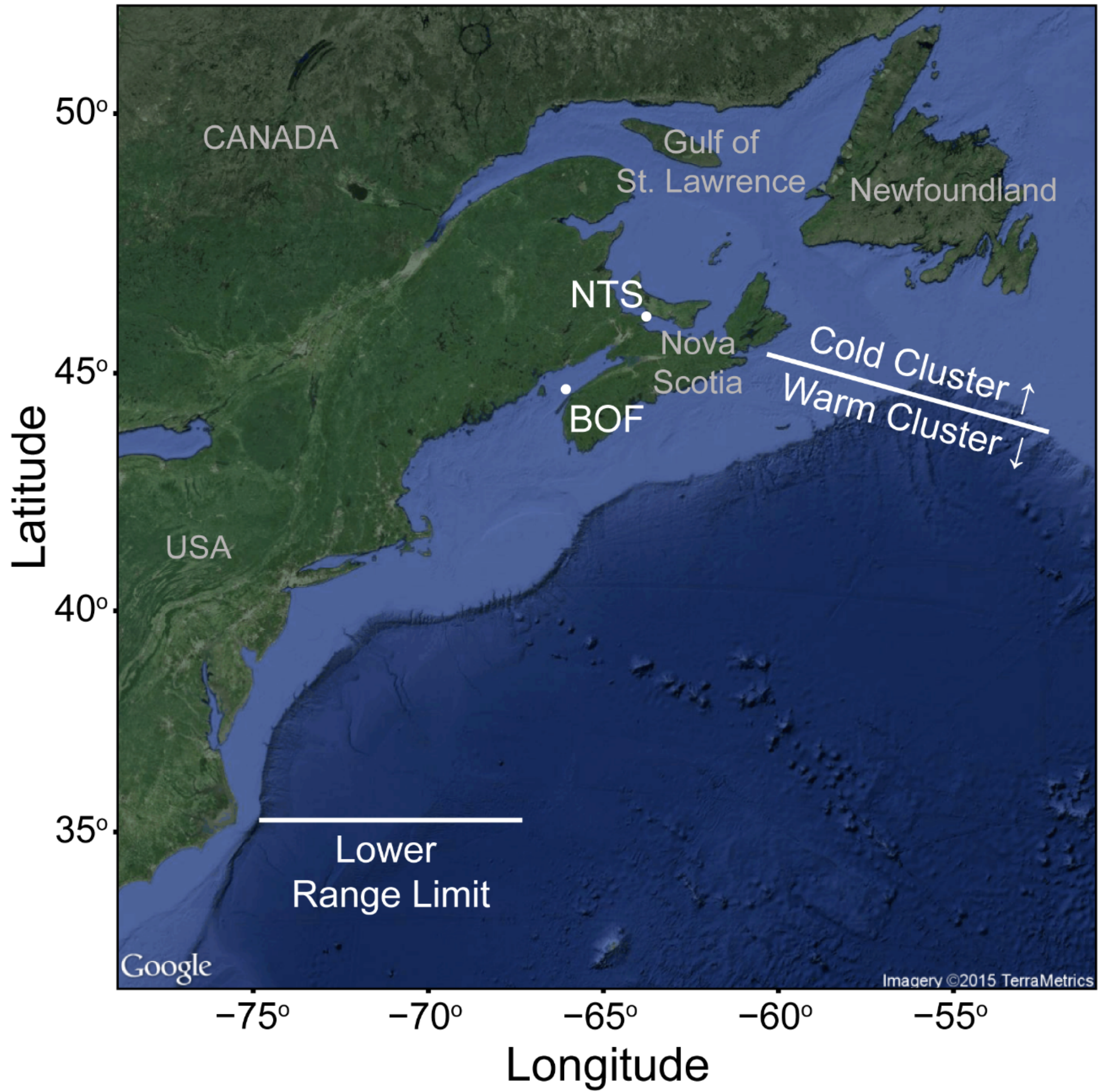

Figure 4.1. Map of collection locations for two populations of *P. magellanicus* used in the common-garden larval rearing experiment.

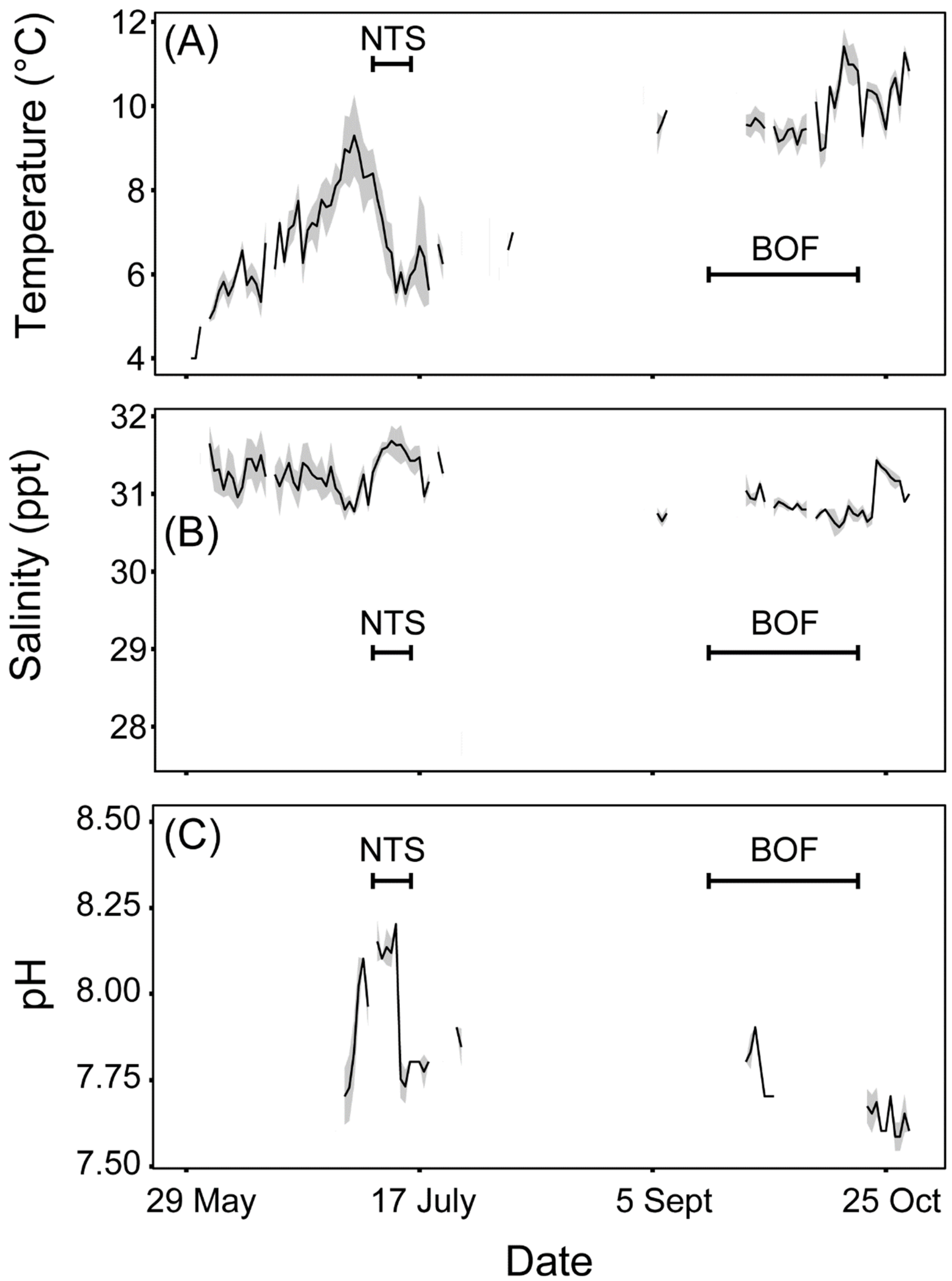

Figure 4.2. (A) Temperature, (B) salinity, and (C) pH experienced by adult *P. magellanicus* during conditioning. Horizontal lines indicate induced spawning periods for each population.

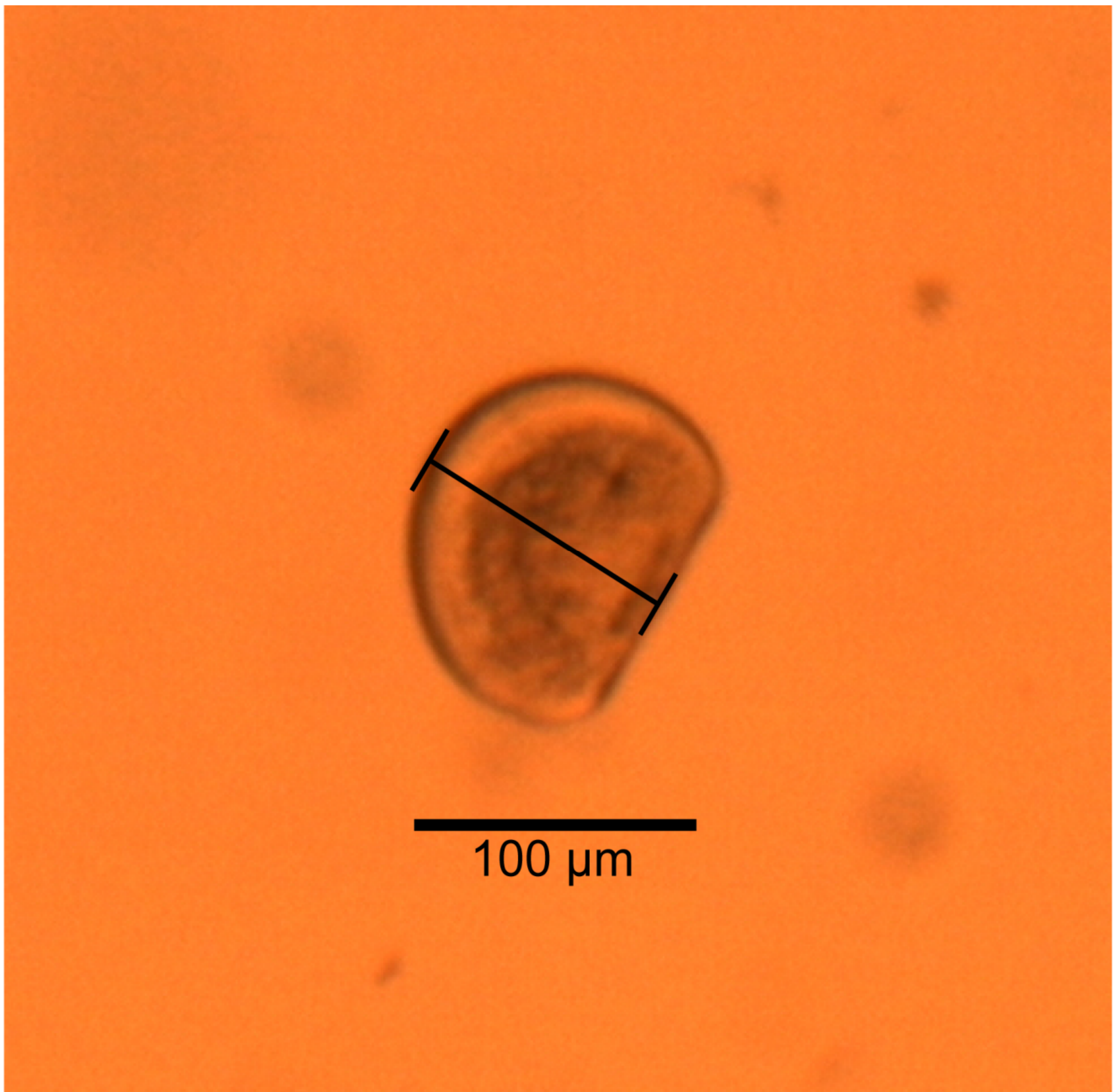

Figure 4.3. Example of typical D-stage scallop larval image indicating appropriate scale and shell height measurement.

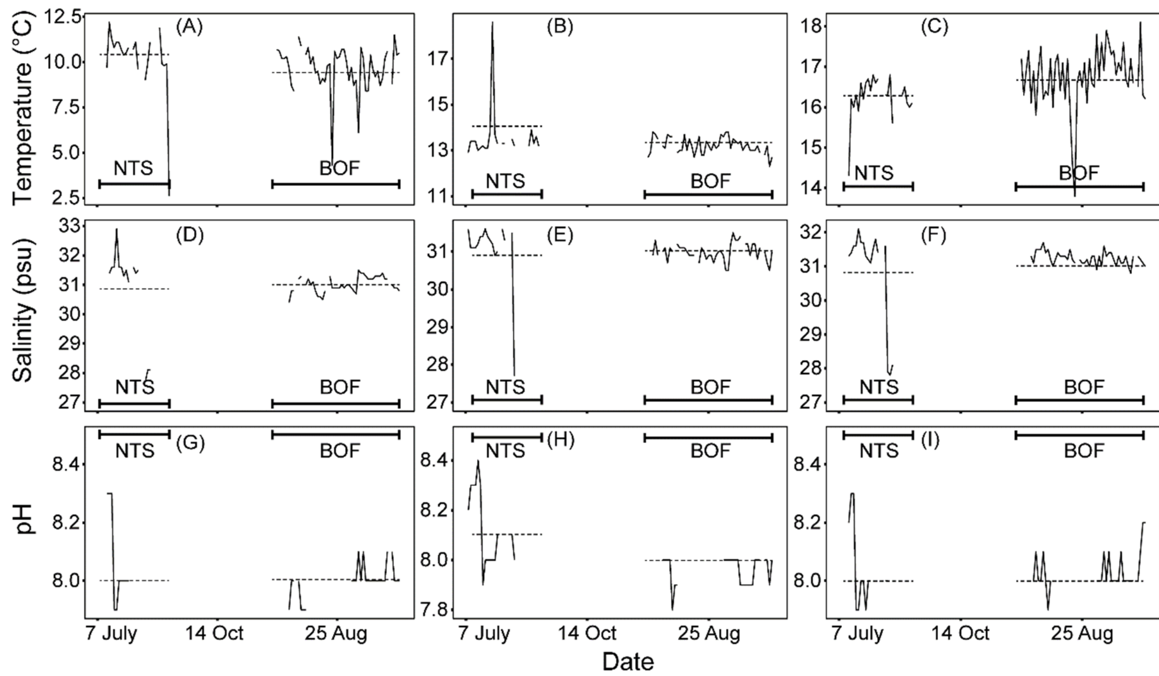

Figure 4.4. (A, B, C) Temperature, (D, E, F) salinity, and (G, H, I) pH experienced by larvae in each temperature trial (10 °C – A, D, E; 13°C – B, E, H; 16 °C – C, F, I) over the duration of the three-week experiment. Dashed horizontal lines indicate the mean value over the duration of the trial. Solid horizontal line indicates duration of the experiment for each population; the timing of spawning for individual families within each population results in difference in length between NTS and BOF. Gaps are due to missing data.

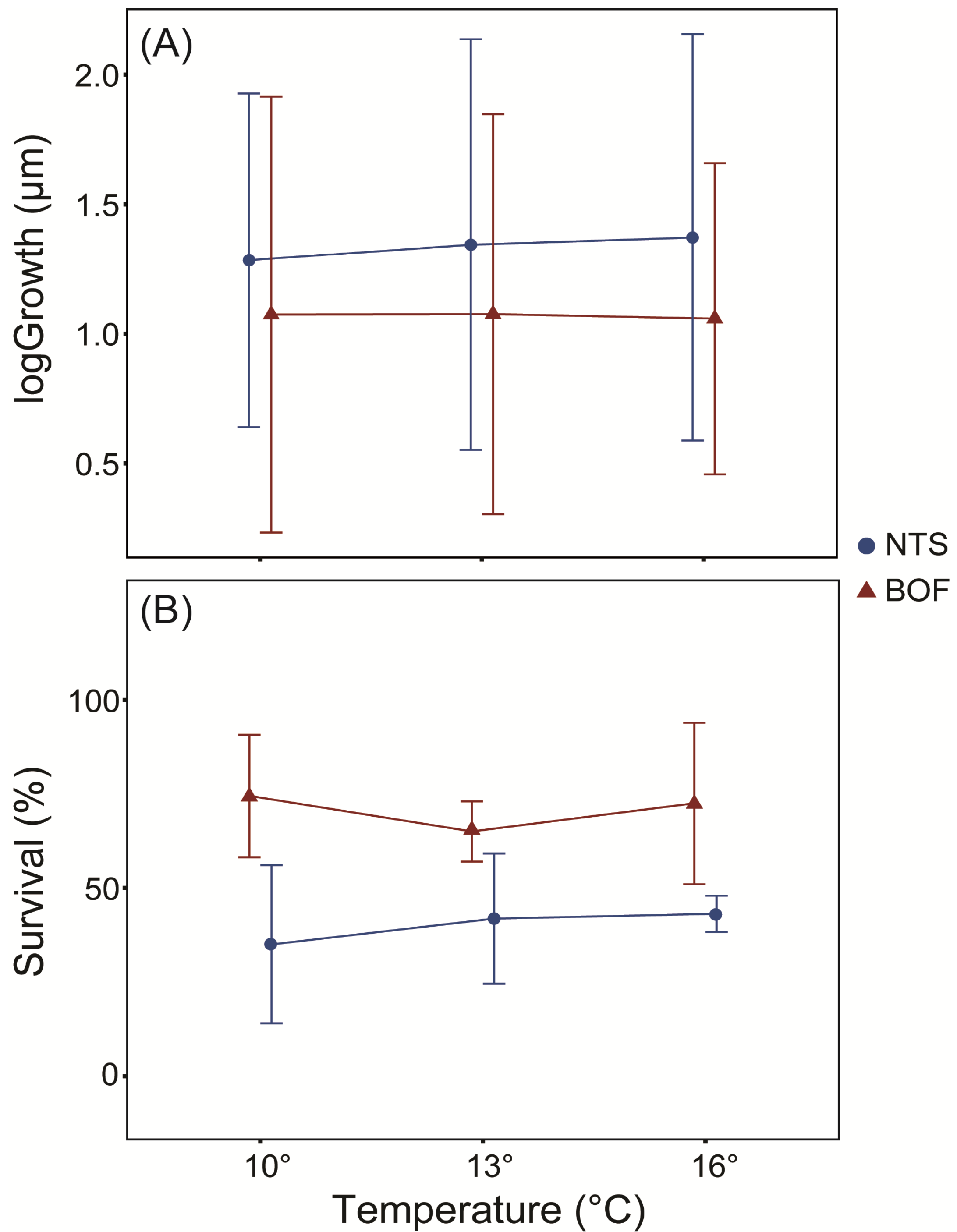

Figure 4.5. (A) Average log growth ± SD of NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C). (B) Average survival ± SD of NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C).

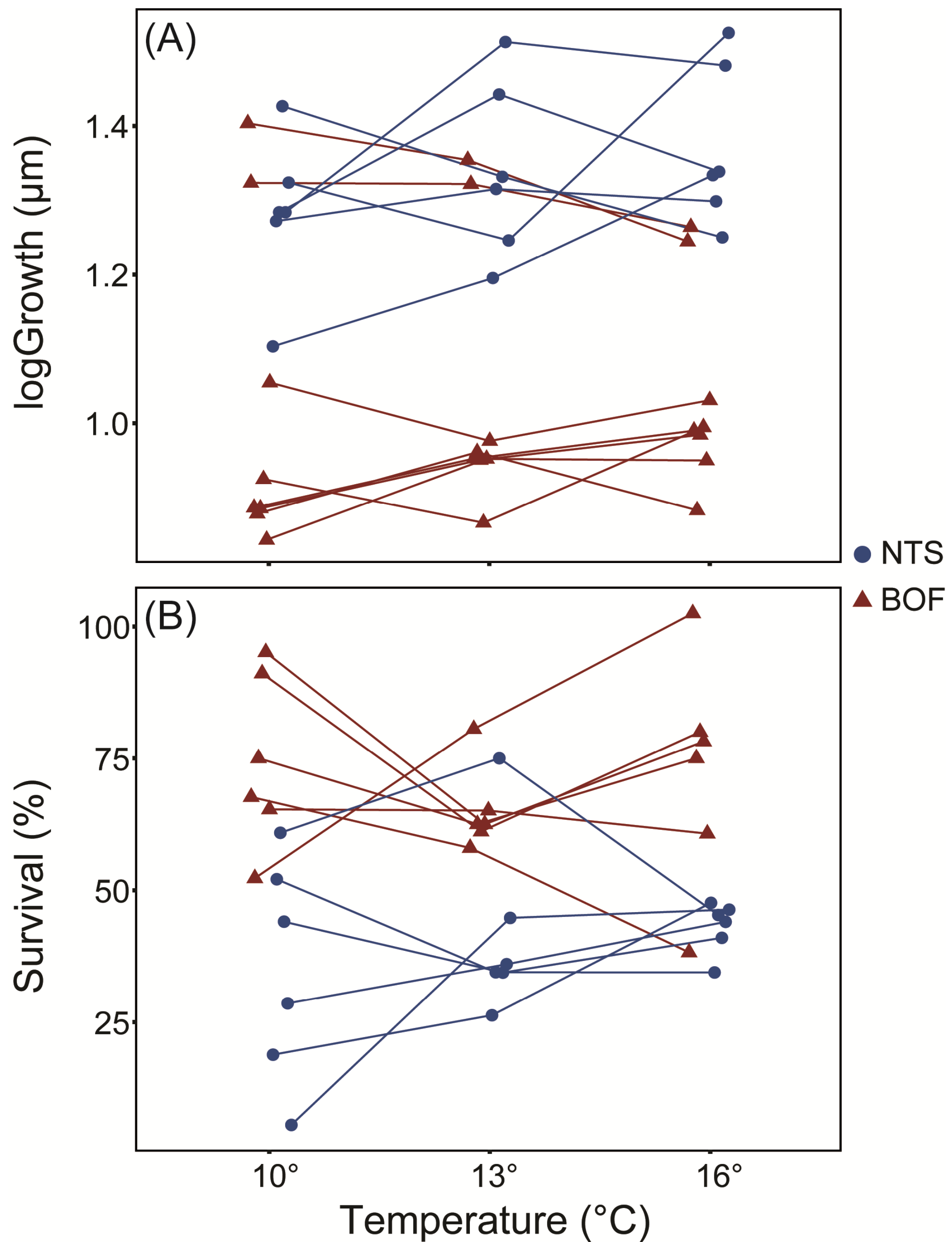

Figure 4.6. (A) Log growth for individual families within NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C). (B) Survival for individual families within NTS (blue circles) and BOF (red triangles) over a three-week period at three different temperatures (10 °C, 13 °C, 16 °C).

Chapter 5: Summary

Given documented phenotypic differences among sea scallop populations (e.g. Barber *et al.* 1988; Manuel *et al.* 1996b; Naidu 1970; Naidu & Robert 2006), the large range spanned by the species (Posgay 1957), the economic importance of the species, and previous work identifying genetic differences (Kenchington *et al.* 2006; Owen & Rawson 2013), we aimed to use modern genomic technologies and experimental approaches to determine range-wide population structure and potential drivers of adaptation in the sea scallop. RAD-seq genotyped 7163 SNPs in 245 individuals across 12 populations, and we used outlier analysis to identify approximately 1% of these SNPs putatively under selection in the scallop genome. Based on these markers and a complementary larval rearing experiment, our work identified previously unrecognized population structure, documented potential limited dispersal among populations, and highlighted the influence that environmental differences among populations may have on population structure and adaptation.

5.1 Population Structure

In Chapter 2, we characterized strong range-wide population structure in the sea scallop, separating populations into north and south clusters with the split located around Nova Scotia, Canada. This work builds on previous work in sea scallops that detected smaller-scale differentiation using other genetic markers (Kenchington *et al.* 2006; Owen & Rawson 2013), however ours is the first study to examine the entire range of the sea scallop. Using outlier and neutral markers separately, we found different patterns of genetic population structure in the north cluster, with outlier loci separately the Gulf of St.

Lawrence from Newfoundland and neutral markers isolating LTB alone. These differences indicate that neutral processes (such as limited dispersal) and potentially selective processes (such as local adaptation) interact to structure sea scallop populations, particularly in the north of their range. In the south cluster we found no evidence of further population structure. Estimates of dispersal generated using genomic clines were smaller than the average pairwise distance between sampled populations and our results identify limited realized dispersal as an important contributor to sea scallop population structure, although the differences between outlier and neutral population structure patterns indicate that dispersal alone cannot explain the strong genetic break between north and south.

5.2 Environmental adaptation

Building on the population structure detected in Chapter 2, Chapter 3 focused on the potential for adaptation among sea scallop populations and its potential effects on realized population structure. Using associations with environmental variation, we identified a subset of loci likely under environmental adaptation and used these loci to qualify the relationship between environmental and genomic differentiation among populations. Our analyses consistently identified cold winter temperatures as the most important variable in a model determining the environmental factors that contributed the most to patterns of genetic differentiation among scallop populations. Although larval scallops likely do not experience these temperatures, exposure to these extremes in overwintering juvenile scallops and the overwinter survival of young scallops may contribute to the population structure observed in Chapter 2. These results in particular

highlight the possibility that local adaptation and the differential of survival of dispersers may drive population connectivity of sea scallop populations in the Northwest Atlantic.

5.3 Larval evidence of adaptation

Although genetic and genomic analyses are powerful tools for identifying population structure and adaptation, multiple researchers have emphasized that the combination of genomics and experimental studies offers the most robust strategy to identify local adaptation among populations (e.g. Rellstab *et al.* 2015). Following the detection of strong genetic separation between north and south population clusters, we spawned and raised larval scallops from the north cluster (NTS) and the south cluster (BOF) in a common garden-like environment to ascertain population specific responses to temperature in larval growth and survival. Although we found no evidence of a larval response to temperature, northern scallops grew faster overall whereas southern scallops exhibited higher survival. This pattern suggests that northern scallops may be adapted to grow more quickly in response to slower growing seasons and prepare for the cold winters experienced by the north population cluster. Further experiments exposing larvae to colder temperatures than those used in our experiment may identify growth responses to temperature in sea scallop larvae.

5.4 Implications

The sea scallop range straddles the Canada and United States border. Noting the great economic importance of the fishery for this species in both countries, this study provides valuable information that can help with cross-border management. Our

conclusions, and suggested future work, particularly along the border and in the southern population cluster, may facilitate the creation of specific fishing areas, cooperative management, and quotas based on the effective connectivity between neighbouring populations.

The patterns of connectivity and putative adaptation detected in this research likely extend beyond sea scallops alone. Many species in the highly productive Northwest Atlantic share similar life histories as sea scallops. Recent work has begun to identify similar patterns of genetic connectivity and limited dispersal for other species elsewhere in the region (Benestan *et al.* 2015), with ongoing further exploration of many species in the Northwest Atlantic. Fishery management plans that work to incorporate this genetic connectivity for the sea scallop and other Northwest Atlantic species will promote sustainability of the species, individual populations, and fisheries.

5.5 Conclusions

The aim of this thesis and the work therein was to detect population structure among sea scallops and identify underlying causes, be it limited dispersal, adaptation and natural selection, or other causes. Our results show range-wide population structure in the sea scallop (separating north and south population clusters) caused by a combination of limited dispersal between populations and potential adaptation to cold winter temperatures in the north preventing the survival of maladapted dispersers. Noting the economic importance of the sea scallop in both Canada and the USA, integrating this evidence of local adaptation and limited dispersal into management decisions may help ensure this species continues to be a sustainable resource in the future. These results also

have implications for understanding response of sea scallop to future climate change; if the north cluster is indeed adapted to colder conditions, the geographic location of the separation between north and south clusters may shift in a warming ocean. Overall, our results provide further evidence for high levels of population structure in marine species, driven not only by oceanographic barriers to dispersal but also by adaptation to different local environments. Such knowledge can help guide the study and management of many marine species, particularly those in the Northwest Atlantic.

Bibliography

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology* **22**, 3179-3190.

Arnold TW (2010) Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management* **74**, 1175-1178.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* **3**.

Barber BJ, Getchell R, Shumway S, Schick D (1988) Reduced fecundity in a deep-water population of the giant scallop *Placopecten magellanicus* in the Gulf of Maine, USA. *Marine Ecology Progress Series* **42**, 207-212.

Barton K (2014) MuMIn: multi-model inference.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 13-45. Oxford University Press.

Bauer LJ, Miller TJ (2010) Spatial and interannual variability in winter mortality of the blue crab (*Callinectes sapidus*) in the Chesapeake Bay. *Estuaries and Coasts* **33**, 678-687.

Baumann H, Conover DO (2011) Adaptation to climate change: contrasting patterns of thermal-reaction-norm evolution in Pacific versus Atlantic silversides. *Proceedings of the Royal Society of London B: Biological Sciences* **278**, 2265-2273.

Beamish RJ, Mahnken C, Neville C (2004) Evidence that reduced early marine growth is associated with lower marine survival of coho salmon. *Transactions of the American Fisheries Society* **133**, 26-33.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Benestan L, Gosselin T, Perrier C*, et al.* (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**, 3299-3315.

Beninger PG (1987) A qualitative and quantitative study of the reproductive cycle of the giant scallop, *Placopecten magellanicus*, in the Bay of Fundy (New Brunswick, Canada). *Canadian Journal of Zoology* **65**, 495-498.

Berg PR, Jentoft S, Star B*, et al.* (2015) Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biology and Evolution* **7**, 1644-1663.

Besnier F, Glover KA (2013) ParallelStructure: A R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *Plos One* **8**.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072.
Bourret V, Dionne M, Kent MP, Lien S, Bernatchez L (2013a) Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution* **67**, 3469-3487.
Bourret V, Kent MP, Primmer CR*, et al.* (2013b) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **22**, 532-551.
Bradbury IR, Bentzen P (2007) Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. *Marine Ecology Progress Series* **340**, 245-257.
Bradbury IR, Bowman S, Borza T*, et al.* (2014a) Long distance linkage disequilibrium and limited hybridization suggest cryptic speciation in Atlantic cod. *Plos One* **9**.
Bradbury IR, Campana SE, Bentzen P (2008a) Estimating contemporary early life-history dispersal in an estuarine fish: integrating molecular and otolith elemental approaches. *Molecular Ecology* **17**, 1438-1450.
Bradbury IR, Campana SE, Bentzen P (2008b) Low genetic connectivity in an estuarine fish with pelagic larvae. *Canadian Journal of Fisheries and Aquatic Sciences* **65**, 147-158.
Bradbury IR, Hamilton LC, Robertson MJ*, et al.* (2014b) Landscape structure and climatic variation determine Atlantic salmon genetic connectivity in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences* **71**, 246-258.
Bradbury IR, Hubert S, Higgins B*, et al.* (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**, 3725-3734.
Bradbury IR, Hubert S, Higgins B*, et al.* (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications* **6**, 450-461.
Bradbury IR, Laurel B, Snelgrove PVR, Bentzen P, Campana SE (2008c) Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. *Proceedings of the Royal Society B-Biological Sciences* **275**, 1803-1809.
Bradbury IR, Laurel BJ, Robichaud D*, et al.* (2008d) Discrete spatial dynamics in a marine broadcast spawner: re-evaluating scales of connectivity and habitat associations in Atlantic cod (*Gadus morhua* L.) in coastal Newfoundland. *Fisheries Research* **91**, 299-309.
Bradbury IR, Snelgrove PVR (2001) Contrasting larval transport in demersal fish and benthic invertebrates: The roles of behaviour and advective processes in determining spatial pattern. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 811-823.
Bradbury IR, Snelgrove PVR, Fraser S (2000) Transport and development of eggs and larvae of Atlantic cod, *Gadus morhua*, in relation to spawning time and location in

coastal Newfoundland. *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 1761-1772.
Breiman L (2001) Random Forests. *Machine Learning* **45**, 5-32.
Brieuc MSO, Ono K, Drinan DP, Naish KA (2015) Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology* **24**, 2729-2746.
Broquet T, Ray N, Petit E, Fryxell JM, Burel F (2006) Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology* **21**, 877-889.
Buonaccorsi VP, Westerman M, Stannard J, Kimbrell C, Lynn E, Vetter RD (2004) Molecular genetic structure suggests limited larval dispersal in grass rockfish, *Sebastes rastrelliger*. *Marine Biology* **145**, 779-788.
Catchen J, Bassham S, Wilson T*, et al.* (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology* **22**, 2864-2883.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics* **1**, 171-182.
Chu ND, Kaluziak ST, Trussell GC, Vollmer SV (2014) Phylogenomic analyses reveal latitudinal population structure and polymorphisms in heat stress genes in the North Atlantic snail *Nucella lapillus*. *Molecular Ecology* **23**, 1863-1873.
Claereboudt M, Himmelman J (1996) Recruitment, growth and production of giant scallops (*Placopecten magellanicus*) along an environmental gradient in Baie des Chaleurs, eastern Canada. *Marine Biology* **124**, 661-670.
Clarke LM, Munch SB, Thorrold SR, Conover DO (2010) High connectivity among locally adapted populations of a marine fish (*Menidia menidia*). *Ecology* **91**, 3526-3537.
Combosch DJ, Vollmer SV (2015) Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific *Pocillopora* corals. *Molecular Phylogenetics and Evolution* **88**, 154-162.
Conesa A, Götz S, García-Gómez JM*, et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676.
Conover DO, Clarke LM, Munch SB, Wagner GN (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology* **69**, 21-47.
Conover DO, Present TMC (1990) Countergradient variation in growth rate: compensation for length of the growing season among Atlantic silversides from different latitudes. *Oecologia* **83**, 316-324.
Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411-1423.
Corander J, Majander KK, Cheng L, Merila J (2013) High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Molecular Ecology* **22**, 2931-2940.

Couturier C, Aiken DE (1989) Possible role of photoperiod in sea scallop reproduction. *Bulleting of the Aquaculture Association of Canada* **89**, 65-67.
Cowen RK, Lwiza KMM, Sponaugle S, Paris CB, Olson DB (2000) Connectivity of marine populations: Open or closed? *Science* **287**, 857-859.
Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. *Science* **311**, 522-527.
Cowen RK, Sponaugle S (2009) Larval dispersal and marine population connectivity. In: *Annual Review of Marine Science*, pp. 443-466.
Culliney JL (1974) Larval development of giant scallop *Placopecten magellanicus* (Gmelin). *Biological Bulletin* **147**, 321-332.
da Silva AG, Appleyard SA, Upston J (2015) Establishing the evolutionary compatibility of potential sources of colonizers for overfished stocks: a population genomics approach. *Molecular Ecology* **24**, 564-579.
Dadswell MJ, Stokesbury M (2007) Biannual spawning and spat settlement of sea scallops (*Placopecten magellanicus*) in Mahone Bay, Nova Scotia and the possible long term effect of a scallop farm on the recovery of the Mahone Bay sea scallop population. *Journal of Shellfish Research* **26**, 1336-1337.
Davies KT, Gentleman W, DiBacco C, Johnson C (2015) Fisheries closed areas strengthen scallop larval settlement and connectivity among closed areas and across international open fishing grounds: a model study. *Environmental management* **56**, 587-602.
Davies KTA, Gentleman WC, DiBacco C, Johnson CL (2014) Semi-annual spawning in marine scallops strengthens larval recruitment and connectivity on Georges Bank: a model study. *Marine Ecology Progress Series* **516**, 209-227.
De Jong G (2005) Evolution of phenotypic plasticity: patterns of plasticity and the emergence of ecotypes. *New Phytologist* **166**, 101-118.
de Villemereuil P, Frichot E, Bazin E, Francois O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology* **23**, 2006-2019.
De Wit P, Palumbi SR (2013) Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology* **22**, 2884-2897.
Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) hzar: hybrid zone analysis using an R software package. *Molecular Ecology Resources* **14**, 652-663.
Devine L, Kennedy MK, St-Pierre I*, et al.* (2014) BioChem: the Fisheries and Oceans Canada database for biological and chemical data. In: *Canadian Technical Report of Fisheries and Aquatic Sciences*. Fisheries and Oceans Canada.
DFO (2014) BioChem: database of biological and chemical oceanographic data. Fisheries and Oceans Canada, Canada.
DFO (2015) Atlantic Zone Monitoring Program. Fisheries and Oceans Canada, Canada.
DFO (2016) *Commercial Fisheries: Landings.* http://www.dfo-mpo.gc.ca/stats/commercial/land-debarq/sea-maritimes/s2014av-eng.htm
DiBacco C, Robert G, Grant J (1995) Reproductive cycle of the sea scallop, *Placopecten magellanicus* (Gmelin, 1791), on northwestern George's Bank. *Journal of Shellfish Research* **14**, 59-69.

Dickie LM (1955) Fluctuations in abundance of the giant scallop, *Placopecten magellanicus* (Gmelin), in the Digby area of the Bay of Fundy. *Journal of the Fisheries Research Board of Canada* **12**, 797-856.
Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**, 1-20.
DuPaul WD, Kirkley JE, Schmitzer AC (1989) Evidence of a semiannual reproductive cycle for the sea scallop, *Placopecten magellanicus* (Gmelin 1791) in the Mid-Atlantic region. *Journal of Shellfish Research* **8**, 173-178.
DuPaul WD, Rudders DB (2008) An assessment of sea scallop abundance and distribution in selected closed areas: Georges Bank area I and II, Nantucket Lightship and Elephant Trunk. In: *VIMS Marine Resource Report*.
Dyer RJ (2014) *An Introduction to the gstudio Package.* http://dyerlab.github.io/gstudio/
Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.
Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011a) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: *Molecular Methods for Evolutionary Genetics* (eds. Orgogozo V, Rockman MV), pp. 157-178. Springer.
Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011b) Local *De Novo* assembly of RAD paired-end contigs using short sequencing reads. *Plos One* **6**, 10.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.
Faircloth BC, Glenn TC (2012) Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *Plos One* **7**, 11.
Fogarty MJ, Botsford LW (2007) Population connectivity and spatial management of marine fisheries. *Oceanography* **20**, 112-123.
Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci sppropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.
Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2015) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology* **25**, 104-120.
Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* **6**, 925-929.
Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**, 1687-1699.

Frichot E, Schoville SD, de Villemereuil P, Gaggiotti OE, Francois O (2015) Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity* **115**, 22-28.

Gagnaire P-A, Broquet T, Aurelle D*, et al.* (2015) Using neutral, selected and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications* **8**, 769-786.

Gaines SD, Gaylord B, Largier JL (2003) Avoiding current oversights in marine reserve design. *Ecological Applications* **13**, S32-S46.

Gaither MR, Bernal MA, Coleman RR*, et al.* (2015) Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Molecular Ecology* **24**, 1543-1557.

Gautier M, Gharbi K, Cezard T*, et al.* (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology* **22**, 3165-3178.

Giguere M, Cliche G, Brulotte S (1994) Reproductive cycles of the sea scallop, *Placopecten magellanicus* (Gmelin), and the Iceland scallop, *Chlamys islandica* (Muller, O.F.), in Iles-de-la-Madeleine, Canada. *Journal of Shellfish Research* **13**, 31-36.

Gilbert CS, Gentleman WC, Johnson CL*, et al.* (2010) Modelling dispersal of sea scallop (*Placopecten magellanicus*) larvae on George's Bank: The influence of depth-distribution, planktonic duration and spawning seasonality. *Progress in Oceanography* **87**, 37-48.

Gonen S, Lowe NR, Cezard T*, et al.* (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* **15**, 166.

Google (2013) Google Earth. Google Inc., Mountain View, CA.

Gregory DN (2004) Climate: a database of temperature and salinity observations for the Northwest Atlantic. DFO, Canada.

Gruenthal KM, Witting DA, Ford T*, et al.* (2014) Development and application of genomic tools to the restoration of green abalone in southern California. *Conservation Genetics* **15**, 109-121.

Guenther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205.

Guo BC, DeFaveri J, Sotelo G, Nair A, Merila J (2015) Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *Bmc Biology* **13**, 19.

Gutteling E, Riksen J, Bakker J, Kammenga J (2007) Mapping phenotypic plasticity and genotype–environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* **98**, 28-37.

Hancock AM, Witonsky DB, Ehler E*, et al.* (2010) Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8924-8930.

Harris BP, Stokesbury KDE (2006) Shell growth of sea scallops (*Placopecten magellanicus*) in the southern and northern Great South Channel, USA. *Ices Journal of Marine Science* **63**, 811-821.

Hastings A, Botsford LW (2006) Persistence of spatial populations depends on returning home. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6067-6072.
Haugen T, Vøllestad L (2000) Population differences in early life-history traits in grayling. *Journal of Evolutionary Biology* **13**, 897-905.
Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**, 333-362.
Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology* **24**, 5573-5595.
Hedgecock D, Barber PH, Edmands S (2007) Genetic approaches to measuring connectivity. *Oceanography* **20**, 70-79.
Hellberg ME (2009) Gene flow and isolation among populations of marine animals. In: *Annual Review of Ecology Evolution and Systematics*, pp. 291-310.
Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70**, 273-290.
Hemmer-Hansen J, Therkildsen NO, Meldrup D, Nielsen EE (2014) Conserving marine biodiversity: insights from life-history trait candidate genes in Atlantic cod (*Gadus morhua*). *Conservation Genetics* **15**, 213-228.
Herrera S, Watanabe H, Shank TM (2015) Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents. *Molecular Ecology* **24**, 673-689.
Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2013) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology* **22**, 2898-2916.
Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-231.
Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics* **6**.
Hohenlohe PA, Catchen J, Cresko WA (2012) Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. *Methods in molecular biology (Clifton, N.J.)* **888**, 235-260.
Hohenlohe PA, Day MD, Amish SJ*, et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology* **22**, 3002-3013.
Hutchings JA, Swain DP, Rowe S*, et al.* (2007) Genetic variation in life-history reaction norms in a marine fish. *Proceedings of the Royal Society B-Biological Sciences* **274**, 1693-1699.
Jensen LF, Hansen MM, Pertoldi C*, et al.* (2008) Local adaptation in brown trout early life-history traits: implications for climate change adaptability. *Proceedings of the Royal Society of London B: Biological Sciences* **275**, 2859-2868.
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.
Jones GP, Srinivasan M, Almany GR (2007) Population connectivity and conservation of marine biodiversity. *Oceanography* **20**, 100-111.

Jones R, Bates JA, Innes DJ, Thompson RJ (1996) Quantitative genetic analysis of growth in larval scallops (*Placopecten magellanicus*). *Marine Biology* **124**, 671-677.

Kelly K (2007) Results from the 2006 Maine sea scallop survey. Maine Department of Marine Resources.

Kenchington EL, Full EW (1994) Fourier analysis of sea scallop (*Placopecten magellanicus*) shells in determining population structure. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 348-356.

Kenchington EL, Patwary MU, Zouros E, Bird CJ (2006) Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). *Molecular Ecology* **15**, 1781-1796.

Kinlan BP, Gaines SD (2003) Propagule dispersal in marine and terrestrial environments: A community perspective. *Ecology* **84**, 2007-2020.

Kirkley JE, Dupaul WD (1991) Temporal variation in spawning brehavior of sea scallops, *Placopecten magellanicus* (Gmelin 1791) in the Mid-Atlantic resource area. *Journal of Shellfish Research* **10**, 389-398.

Koehn RK, Milkman R, Mitton JB (1976) Population genetics of marine pelecypods. IV. Selection, migration and genetic differentiation in the blue mussel *Mytilus edulis*. *Evolution* **30**, 2-32.

Koehn RK, Newell R, Immermann F (1980) Maintenance of an aminopeptidase allele frequency cline by natural selection. *Proceedings of the National Academy of Sciences* **77**, 5385-5389.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Kordos LM, Burton RS (1993) Genetic differentiation of Texas gulf-coast populations of the blue crab *Callinectes sapidus*. *Marine Biology* **117**, 227-233.

Kruck NC, Innes DI, Ovenden JR (2013) New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). *Molecular Ecology Resources* **13**, 715-725.

Lamichhaney S, Barrio AM, Rafati N*, et al.* (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19345-19350.

Langella O (1999) Populations 1.2.30. *CNRS UPR9034*.

Langton RW, Robinson WE, Schick D (1987) Fecundity and reproductive effort of sea scallops *Placopecten magellanicus* from the Gulf of Maine. *Marine Ecology Progress Series* **37**, 19-25.

Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a "Continuous" population under isolation by distance. *Molecular Biology and Evolution* **20**, 491-502.

Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* **166**, 1081-1092.

Lenormand T, Guillemaud T, Bourguet D, Raymond M (1998) Evaluating gene flow using selected markers: A case study. *Genetics* **149**, 1383-1392.
Levin LA (2006) Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology* **46**, 282-297.
Liaw A, Wiener M (2002) Classification and regression by randomForest. *R news* **2**, 18-22.
Liefting M, Hoffmann AA, Ellers J (2009) Plasticity versus environmental canalization: population differences in thermal responses along a latitudinal gradient in *Drosophila serrata*. *Evolution* **63**, 1954-1963.
Limborg MT, Helyar SJ, De Bruyn M*, et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology* **21**, 3686-3703.
Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology* **19**, 3038-3051.
Macdonald BA, Thompson RJ (1985) Influence of temperature and food availability on the ecological energetics of the giant scallop *Placopecten magellanicus*. II. Reproductive output and total production. *Marine Ecology Progress Series* **25**, 295-304.
Macdonald BA, Thompson RJ (1988) Intraspecific variation in growth and reproduction in latitudinally differentiated populations of the giant scallop *Placopecten magellanicus* (Gmelin). *Biological Bulletin* **175**, 361-371.
Malécot G (1955) The decrease of relationship with distance. In: *Cold Spring Harbor Symposium on Quantitative Biology*, pp. 52-53.
Manuel JL, Burbridge S, Kenchington EL, Ball M, Odor RK (1996a) Veligers from two populations of scallop *Placopecten magellanicus* exhibit different vertical distributions in the same mesocosm. *Journal of Shellfish Research* **15**, 251-257.
Manuel JL, Gallager SM, Pearce CM, Manning DA, Odor RK (1996b) Veligers from different populations of sea scallop *Placopecten magellanicus* have different vertical migration patterns. *Marine Ecology Progress Series* **142**, 147-163.
Mason GE, Sameoto JA, Metaxas A (2014) In situ swimming characteristics of the sea scallop, *Placopecten magellanicus*, on German Bank, Gulf of Maine. *Journal of the Marine Biological Association of the United Kingdom* **94**, 1019-1026.
Metaxas A (2001) Behaviour in flow: Perspectives on the distribution and dispersion of meroplanktonic larvae in the water column. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 86-98.
Milano I, Babbucci M, Cariani A*, et al.* (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology* **23**, 118-135.
Miller KJ, Mundy CN, Mayfield S (2014) Molecular genetics to inform spatial management in benthic invertebrate fisheries: a case study using the Australian Greenlip Abalone. *Molecular Ecology* **23**, 4958-4975.
Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**, 240-248.

Moore J-S, Bourret V, Dionne M*, et al.* (2014) Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology* **23**, 5680-5697.
Moura AE, Kenny JG, Chaudhuri R*, et al.* (2014) Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Molecular Ecology* **23**, 5179-5192.
Naidu KS (1970) Reproduction and breeding cycle of giant scallop *Placopecten magellanicus* (Gmelin) in Port-Au-Port Bay, Newfoundland. *Canadian Journal of Zoology* **48**, 1003-&.
Naidu KS, Robert G (2006) Fisheries Sea Scallop, *Placopecten magellanicus*. In: *Scallops: Biology, Ecology and Aquaculture* (eds. Shumway S, Parsons GJ), pp. 869-905. Elsevier.
Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.
NASA Goddard Space Flight Center Ocean Ecology Laboratory OBPG (2014) MODIS-Aqua Ocean Color Data, NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group.
Neilsen EE, Kenchington E (2001) A new approach to prioritizing marine fish and shellfish populations for conservation. *Fish and Fisheries* **2**, 328-243.
NOAA (2016) Commercial Fisheries Statistics.
Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.
Ogden R, Gharbi K, Mugue N*, et al.* (2013) Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology* **22**, 3112-3123.
Oksanen J, Blanchet FG, Kindt R*, et al.* (2015) vegan: Community Ecology Package.
Oomen RA, Hutchings JA (2015) Variation in spawning time promotes genetic variability in population responses to environmental change in a marine fish. *Conservation Physiology* **3**.
Owen EF, Rawson PD (2013) Small-scale spatial and temporal genetic structure of the Atlantic sea scallop (*Placopecten magellanicus*) in the inshore Gulf of Maine revealed using AFLPs. *Marine Biology* **160**, 3015-3025.
Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.
Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, S146-S158.
Pante E, Simon-Bouhet B (2013) marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *Plos One* **8**, e73051.
Pavey SA, Gaudin J, Normandeau E, Dionne M, Castonguay M, Audet C, Bernatchez L (2015) RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic Americal eel. *Current Biology* **25**, 1666-1671.
Pespeni MH, Palumbi SR (2013) Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. *Molecular Ecology* **22**, 3580-3597.
Plaia A, Bondi AL (2006) Single imputation method of missing values in environmental pollution data sets. *Atmospheric Environment* **40**, 7316-7330.

Posgay JA (1957) The range of the sea scallop. *Nautilus* **71**, 55-57.
Posgay JA, Norman KD (1958) An observation on the spawning of the sea scallop, *Placopecten magellanicus* (Gmelin), on George's Bank. *Limnology and Oceanography* **3**, 478-478.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
Pujolar JM, Jacobsen MW, Frydenberg J*, et al.* (2013) A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. *Molecular Ecology Resources* **13**, 706-714.
Purcell S (2009) *PLINK 1.07.* http://pngu.mgh.harvard.edu/purcell/plink/
Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.
Purchase CF, Brown JA (2000) Interpopulation differences in growth rates and food conversion efficiencies of young Grand Banks and Gulf of Maine Atlantic cod (*Gadus morhua*). *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 2223-2229.
R Development Core Team (2012) *R: A language and environment for statistical computing.* http://www.R-project.org/.
Rasband WS (1997-2015) ImageJ. U. S. National Institutes of Health, Bethesda, Maryland, USA.
Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.
Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24**, 4348-4370.
Robinson S, Martin J, Chandler R, Parsons GJ (2007) Monitoring the reproduction of the giant sea scallop, *Placopecten magellanicus* in Passamaquoddy Bay, Bay of Fundy: will I live long enough to determine a long-term cycle? *Journal of Shellfish Research* **26**, 1336-1337.
Roff DADA (2002) *Life history evolution*.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.
Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology* **18**, 1765-1776.
Sale PF, Cowen RK, Danilowicz BS*, et al.* (2005) Critical science gaps impede use of no-take fishery reserves. *Trends in Ecology and Evolution* **20**, 74-80.
Schmidt PS, Bertness MD, Rand DM (2000) Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. *Proceedings of the Royal Society of London B: Biological Sciences* **267**, 379-384.

Schmitzer AC, Dupaul WD, Kirkley JE (1991) Gametogenic cycle of sea scallops *Placopecten magellanicus* (Gmelin, 1971) in the Mid-Atlantic Bight. *Journal of Shellfish Research* **10**, 221-228.
Schultz ET, Conover DO, Ehtisham A (1998) The dead of winter: size-dependent variation and genetic differences in seasonal mortality among Atlantic silverside (Atherinidae: *Menidia menidia*) from different latitudes. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 1149-1157.
Schwender H (2012) Imputing missing genotypes with weighted k nearest neighbors. *Journal of Toxicology and Environmental Health, Part A* **75**, 438-446.
Schwender H, Fritsch A (2013) *Scrime: Analysis of High-Dimensional Categorical Data such as SNP Data.* http://CRAN.R-project.org/package=scrime
Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* **436**, 291-305.
Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* **68**, 1-15.
Shafer A, Wolf JB (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters* **16**, 940-950.
Sjöqvist C, Godhe A, Jonsson PR, Sundqvist L, Kremp A (2015) Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea–Baltic Sea salinity gradient. *Molecular Ecology* **24**, 2871-2885.
Sotka EE, Palumbi SR (2006) The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology* **87**, 1094-1103.
Stokesbury KDE, Himmelman JH (1995) Biological and physical variables associated with aggregations of the giant scallop *Placopecten magellanicus*. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 743-753.
Strasburg JL, Sherman NA, Wright KM*, et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 364-373.
Sumner J, Rousset F, Estoup A, Moritz C (2001) 'Neighbourhood' size, dispersal and density estimates in the prickly forest skink (*Gnypetoscincus queenslandiae*) using individual genetic and demographic methods. *Molecular Ecology* **10**, 1917-1927.
Therkildsen NO, Hemmer-Hansen J, Hedeholm RB*, et al.* (2013) Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. *Evolutionary Applications* **6**, 690-705.
TheWeatherCompany (2016) *Weather Underground.* The Weather Company. https://www.wunderground.com/
Thompson RJ (1977) Blood chemistry, biochemical composition, and the annual reproductive cycle in the giant scallop, *Placopecten magellanicus*, from southeast Newfoundland *Journal of the Fisheries Research Board of Canada* **34**, 2104-2116.
Thorrold SR, Jones GP, Hellberg ME*, et al.* (2002) Quantifying larval retention and connectivity in marine populations with artificial and natural markers. *Bulletin of Marine Science* **70**, 291-308.

Thorrold SR, Zacherl DC, Levin LA (2007) Population connectivity and larval dispersal using geochemical signatures in calcified structures. *Oceanography* **20**, 80-89.
Thorson G (1950) Reproductive and larval ecology of marine bottom invertebrates. *Biological reviews* **25**, 1-45.
Thouzeau G, Robert G, Smith SJ (1991) Spatial variability in distribution and growth of juvenile and adult sea scallops *Placopecten magellanicus* (Gmelin) on eastern George's Bank (Northwest Atlantic). *Marine Ecology Progress Series* **74**, 205-218.
Tian RC, Chen C, Stokesbury KDE*, et al.* (2009a) Dispersal and settlement of sea scallop larvae spawned in the fishery closed areas on Georges Bank. *Ices Journal of Marine Science* **66**, 2155-2164.
Tian RC, Chen C, Stokesbury KDE*, et al.* (2009b) Modeling the connectivity between sea scallop populations in the Middle Atlantic Bight and over George's Bank. *Marine Ecology Progress Series* **380**, 147-160.
Tilburg CE, McCartney MA, Yund PO (2012) Across-shelf transport of bivalve larvae: can the interface between a coastal current and inshore waters act as an ecological barrier to larval dispersal? *Plos One* **7**.
Townsend DW, Thomas AC, Mayer LM, Thomas MA, Quinlan JA (2006) Oceanography of the Northwest Atlantic continental shelf (1,W). In: *The Sea: The Global Coastal Ocean: Interdisciplinary Regional Studies and Syntheses.* eds. Robinson AR, Brink KH), pp. 1-57. Harvard University Press, Cambridge, Massachusetts.
Tremblay MJ, Sinclair M (1990a) Diel vertical migration of sea scallop larvae *Placopecten magellanicus* in a shallow embayment. *Marine Ecology Progress Series* **67**, 19-25.
Tremblay MJ, Sinclair M (1990b) Sea scallop larvae (*Placopecten magellanicus*) on George's Bank: vertical distribution in relation to water column stratification and food. *Marine Ecology Progress Series* **61**, 1-16.
Van Asch M, Van Tienderen PH, Holleman LJM, Visser ME (2007) Predicting adaptation of phenology in response to climate change, an insect herbivore example. *Global Change Biology* **13**, 1596-1604.
Van Wyngaarden M, Snelgrove PVR, DiBacco C*, et al.* (*in review*) Oceanographic variation influences selection and population structure in the sea scallop, *Placopecten magellanicus*, identified using RADseq.
Van Wyngaarden M, Snelgrove PVR, DiBacco C*, et al.* (2017) Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus*, using clines in RADseq derived SNPs. *Evolutionary Applications* **10**, 102-117.
Vasemägi A (2006) The adaptive hypothesis of clinal variation revisited: single-locus clines as a results of spatially restricted gene flow. *Genetics* **173**, 2411-2414.
Waples RS (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**, 438-450.
Winterhalter WE, Mousseau TA (2007) Patterns of phenotypic and genetic variation for the plasticity of diapause incidence. *Evolution* **61**, 1520-1531.

Yamahira K, Kawajiri M, Takeshi K, Irie T (2007) Inter-and intrapopulation variation in thermal reaction norms for growth rate: evolution of latitudinal compensation in ectotherms with a genetic constraint. *Evolution* **61**, 1577-1589.

Appendices

Appendix S1: Supporting Information for **Chapter 2:** Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus,* using RAD-seq derived SNPs

Table S1.1. (A) Parameter values for 8 sets tested using the program *stacks* (Catchen et al. 2011) (B) Initial catalog tags, catalog tags and SNPs following initial individual filtering, and catalog tags, SNPs, number of individuals, and number of populations included following detailed filtering for missing data and minor allele frequency for each parameter set tested. The final parameter map used was 4B.

(A)

Stacks parameter set	*Min. stack depth*	*Differences between stacks (1 individual)*	*Differences between secondary reads and existing stacks*	*Block haplotypes from secondary reads*	*Remove repetitive stacks*	*Deleveraging*	*Distance allowed between catalog loci*	*SNP Model*
	m	**M**	**N**	**H**	**r**	**d**	**n**	
1	3	4	6	no	yes	on	2	snp
2	10	4	0	no	yes	off	2	snp
3	5	4	0	no	yes	on	6	snp
4	5	4	6	yes	yes	on	6	snp
4B	5	4	6	no	yes	on	6	snp
5	5	4	6	yes	yes	on	6	bounded
6	3	6	8	yes	yes	on	8	snp
7	3	4	6	no	yes	off	2	snp

(B)

Stacks parameter set	*Initial catalog*	*Tags present in 75% of individuals*		*Genotyped in 95% of individuals, MAF > 0.05, Missing loci per individual < 20%*		
	RAD Tags	**RAD Tags**	**SNPs**	**SNPs**	**Individuals**	**Populations**
1	236093	20366	176492	6194	255	12
2	111104	16634	131291	6456	214	11
3	132184	19672	171923	7705	247	12
4	132281	19659	169866	7208	245	12
4B	131897	19672	173482	7216	245	12
5	132425	19676	171673	7516	246	12
6	205367	19945	182112	6286	255	12
7	230352	19883	175547	7157	255	12

Table S1.2. Ten datasets used in population structure analysis of 12 populations of *P. magellanicus* collected from the Northwest Atlantic.

Dataset Number	Dataset Name	Loci Included	Populations Included
1	All loci, all	All loci	SUN, LTB, MGD, NTS, PSB, BOF, SSM, GMI, SSB, GMO, GEO, MDA
2	Outlier loci, all	Outlier loci	SUN, LTB, MGD, NTS, PSB, BOF, SSM, GMI, SSB, GMO, GEO, MDA
3	Neutral loci, all	Neutral loci	SUN, LTB, MGD, NTS, PSB, BOF, SSM, GMI, SSB, GMO, GEO, MDA
4	All loci, North	All loci	SUN, LTB, MGD, NTS
5	Outlier loci, North	Outlier loci	SUN, LTB, MGD, NTS
6	Neutral loci, North	Neutral loci	SUN, LTB, MGD, NTS
7	All loci, South	All loci	PSB, BOF, SSM, GMI, SSB, GMO, GEO, MDA
8	Outlier loci, South	Outlier loci	PSB, BOF, SSM, GMI, SSB, GMO, GEO, MDA
9	Neutral loci, South	Neutral loci	PSB, BOF, SSM, GMI, SSB, GMO, GEO, MDA
10	High F_{ST}, South	100 highest F_{ST}	PSB, BOF, SSM, GMI, SSB, GMO, GEO, MDA

Table S1.3. Comparison of outlier SNP loci from 12 populations of *P. magellanicus* determined using BAYESCAN (Bayesian method, 112 loci) and the 99% (157 loci) and 95% (184 loci) confidence intervals in ARLEQUIN (island model).
* Loci not present in ARLEQUIN 99%
¤ Loci not present in ARLEQUIN 95%

Locus Name		
FDR = 0.05	p = 0.01	
BayeScan	Arlequin 99%	Arlequin 95%
11_6	11_6	11_6
18_69	18_69	18_69
319_43	319_43	319_43
979_75	523_39	523_39
1089_23	523_83	523_83
1299_57	979_75	979_75
1467_19	1089_23	1089_23
3299_78	1299_57	1299_57
3350_55	1467_19	1467_19
3350_66	1659_63	1659_63
3497_39	2272_32	2272_32
3498_66	2816_12	2816_12
3619_79	3299_78	3286_90
3834_87	3350_55	3299_78
3929_9	3350_66	3350_55
3969_43	3497_39	3350_66
4484_19	3498_66	3497_39
4668_81	3619_79	3498_66
4847_58	3834_87	3619_79
4975_9	3929_9	3834_87
4975_23	3969_43	3929_9
4975_68	4063_7	3969_43
5246_66	4226_74	4063_7
5252_37	4373_68	4226_74
5439_27	4421_10	4373_68
5515_42	4484_19	4421_10
5750_64	4589_47	4484_19
5791_60*¤	4668_81	4589_47
6948_62	4678_65	4668_81

7115_30*	4847_58	4678_65
7115_35*	4975_9	4847_58
7203_45	4975_23	4975_9
7326_20	4975_68	4975_23
7396_78	5246_66	4975_68
7524_34	5252_37	5246_66
7740_53	5439_27	5252_37
8699_63	5515_42	5439_27
8782_41	5750_64	5515_42
9206_48	5924_59	5680_70
9554_49	6234_40	5750_64
9580_74	6234_86	5924_59
9976_51	6948_62	6234_40
9976_52	7203_45	6234_86
9978_28	7326_20	6948_62
10349_43	7396_78	7115_30
10366_19	7524_34	7115_35
10498_47	7740_53	7203_45
10832_30	8053_41	7216_63
10832_46	8699_63	7326_20
10832_74	8782_41	7396_78
10832_84	9114_34	7524_34
10964_88*¤	9206_48	7740_53
10964_89*¤	9554_49	7857_8
10964_90*¤	9580_74	8053_41
10987_36	9653_59	8699_63
10987_48	9653_60	8782_41
11110_18*¤	9976_51	9114_34
11162_70	9976_52	9206_48
11531_7	9978_28	9431_86
12073_7	10349_43	9554_49
12308_45	10366_19	9580_74
12767_68	10498_47	9653_59
13384_35	10832_30	9653_60
13891_65	10832_46	9976_51
14394_27	10832_74	9976_52
14571_64	10832_84	9978_28
14750_60	10893_84	10349_43
15099_34	10987_36	10366_19
15446_21	10987_48	10498_47
15645_89	11162_70	10832_30

15660_9	11235_69	10832_46
15821_29	11462_75	10832_74
16110_74	11531_7	10832_84
16228_63	11895_12	10863_83
16229_44	11895_17	10893_84
16309_22	11895_50	10987_36
16478_29	12073_7	10987_48
16894_73	12308_45	11162_70
17085_50	12767_68	11235_69
17567_80	12833_63	11462_75
18135_35	13061_52	11531_7
18391_26	13384_35	11718_14
18669_56	13761_22	11851_23
18848_34*¤	13891_65	11895_12
18848_38*¤	14069_58	11895_17
19165_26	14394_27	11895_50
19165_90*	14557_41	12073_7
20298_73	14571_64	12308_45
20400_32	14750_60	12767_68
20548_19	15099_34	12833_63
20633_88	15127_80	13061_52
20810_86	15245_72	13384_35
21297_65	15245_73	13761_22
21510_19	15446_21	13824_83
22068_17	15645_89	13889_83
22237_35	15660_9	13891_65
23947_78	15752_28	14069_58
24442_62	15821_29	14119_48
25322_42	16110_74	14394_27
25380_52	16228_63	14557_41
25380_58	16229_44	14571_64
25380_84	16309_22	14750_35
25405_78	16478_29	14750_60
25627_51	16794_23	15099_34
25748_78	16891_32	15127_80
25881_17	16894_73	15245_72
25888_33	17081_17	15245_73
25888_38	17085_50	15446_21
25962_11	17567_80	15645_89
26519_27	18135_35	15660_9
26611_88	18391_26	15752_28

26726_84	18656_70	15821_29
	18669_56	16110_74
	19165_26	16228_63
	19483_28	16229_44
	19716_73	16309_22
	20263_79	16363_34
	20298_73	16478_29
	20400_32	16794_23
	20464_35	16891_32
	20548_19	16894_73
	20633_88	17081_17
	20810_86	17085_50
	21297_65	17567_80
	21425_11	18135_35
	21510_19	18391_26
	21588_80	18656_70
	21600_9	18669_56
	21717_37	19165_26
	22068_17	19165_90
	22237_35	19483_28
	23339_64	19716_73
	23947_78	20263_79
	24310_65	20298_27
	24442_62	20298_73
	24880_8	20400_32
	25322_42	20464_35
	25380_52	20548_19
	25380_58	20633_88
	25380_84	20810_86
	25405_78	20942_64
	25433_34	21297_65
	25433_45	21311_7
	25627_51	21425_11
	25748_78	21510_19
	25879_86	21588_80
	25881_17	21600_9
	25888_33	21717_37
	25888_38	22066_26
	25962_11	22068_17
	25962_54	22237_35
	25972_88	23339_64

26519_27	23947_78
26588_15	24310_65
26611_88	24442_62
26726_84	24880_8
26781_56	25055_11
	25322_42
	25380_52
	25380_58
	25380_84
	25405_78
	25433_34
	25433_45
	25627_51
	25748_78
	25879_86
	25881_17
	25888_33
	25888_38
	25962_11
	25962_54
	25962_57
	25972_88
	26237_36
	26237_82
	26519_27
	26588_15
	26611_88
	26726_84
	26781_56
	26984_6
	27045_28
	27045_45

Table S1.4. Pairwise population-specific F_{ST} (above diagonal) and p-value (below diagonal) for (A) all loci, (B) neutral loci, and (C) outlier loci. Maximum F_{ST} values are highlighted in bold, minimum values are highlighted in bold and italicized. Significant p-values are italicized.

(A)	**SUN**	**LTB**	**MGD**	**NTS**	**PSB**	**BOF**	**SSM**	**GMI**	**SSB**	**GMO**	**GEO**	**MDA**
SUN		0.010	0.003	0.005	0.005	0.010	0.006	0.005	0.009	0.006	0.010	0.007
LTB	*0.000*		0.011	0.014	0.014	0.016	0.015	0.016	0.017	0.014	**0.018**	0.015
MGD	*0.001*	*0.000*		0.002	0.004	0.006	0.006	0.005	0.007	0.004	0.007	0.006
NTS	*0.000*	*0.000*	*0.035*		0.004	0.009	0.007	0.005	0.009	0.004	0.009	0.007
PSB	*0.002*	*0.000*	0.230	0.119		-0.003	-0.001	0.004	-0.002	-0.003	0.002	-0.002
BOF	*0.000*	*0.000*	*0.000*	*0.000*	1.000		0.003	-0.003	0.001	0.001	0.001	0.000
SSM	*0.000*	*0.000*	*0.000*	*0.000*	1.000	*0.004*		0.000	0.003	0.002	0.003	0.002
GMI	*0.000*	*0.000*	*0.000*	*0.000*	0.198	1.000	0.992		-0.003	-0.002	-0.001	***-0.004***
SSB	*0.000*	*0.000*	*0.000*	*0.000*	1.000	0.672	*0.020*	1.000		0.001	0.000	0.000
GMO	*0.000*	*0.000*	*0.000*	*0.000*	1.000	0.629	0.130	1.000	0.699		0.000	0.000
GEO	*0.000*	*0.000*	*0.000*	*0.000*	0.881	0.687	*0.012*	1.000	0.959	0.965		0.000
MDA	*0.000*	*0.000*	*0.000*	*0.000*	1.000	0.916	0.303	1.000	0.981	0.996	0.991	

(B)	**SUN**	**LTB**	**MGD**	**NTS**	**PSB**	**BOF**	**SSM**	**GMI**	**SSB**	**GMO**	**GEO**	**MDA**
SUN		0.009	***0.002***	0.003	0.003	0.007	0.005	***0.002***	0.005	0.004	0.005	0.007
LTB	*0.000*		0.010	0.012	0.011	0.012	0.013	0.012	0.012	0.011	0.012	**0.014**
MGD	*0.040*	*0.000*		0.003	0.002	0.004	0.005	0.003	0.005	0.003	0.004	0.005
NTS	0.073	*0.000*	0.640		0.002	***-0.004***	-0.002	0.004	-0.002	-0.003	-0.003	0.001
PSB	*0.001*	*0.000*	0.238	0.743		0.006	0.006	0.002	0.006	0.002	0.004	0.006
BOF	*0.000*	*0.000*	*0.000*	1.000	*0.000*		0.002	-0.003	0.001	0.001	0.000	0.001
SSM	*0.000*	*0.000*	*0.000*	1.000	*0.000*	0.090		-0.001	0.002	0.002	0.001	0.002
GMI	*0.006*	*0.000*	*0.007*	0.316	0.060	1.000	0.999		***-0.004***	-0.002	***-0.004***	-0.001
SSB	*0.000*	*0.000*	*0.000*	1.000	*0.000*	0.820	0.423	1.000		0.000	0.000	0.000
GMO	*0.000*	*0.000*	*0.009*	1.000	*0.016*	0.689	0.363	1.000	0.880		-0.001	0.000
GEO	*0.000*	*0.000*	*0.000*	1.000	*0.000*	0.954	0.775	1.000	0.987	0.998		0.000
MDA	*0.000*	*0.000*	*0.000*	0.968	*0.000*	0.753	0.199	1.000	0.977	0.979	0.993	

(C)	**SUN**	**LTB**	**MGD**	**NTS**	**PSB**	**BOF**	**SSM**	**GMI**	**SSB**	**GMO**	**GEO**	**MDA**
SUN		0.057	0.122	0.061	0.095	0.122	0.175	0.157	0.187	0.152	0.123	0.171
LTB	*0.000*		0.171	0.080	0.148	0.177	0.229	0.225	**0.253**	0.216	0.160	0.220
MGD	*0.000*	*0.000*		0.047	0.109	0.129	0.147	0.148	0.195	0.158	0.110	0.172
NTS	*0.000*	*0.000*	*0.000*		0.073	0.086	0.135	0.120	0.165	0.128	0.079	0.142
PSB	*0.000*	*0.000*	*0.000*	*0.000*		0.049	0.024	0.029	0.057	0.041	0.008	0.046
BOF	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*		0.065	0.040	0.086	0.057	0.037	0.072
SSM	*0.000*	*0.000*	*0.000*	*0.000*	*0.011*	*0.000*		0.012	0.019	0.011	0.008	0.010
GMI	*0.000*	*0.000*	*0.000*	*0.000*	*0.001*	*0.000*	*0.035*		0.022	0.015	0.013	0.007
SSB	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.003*	*0.001*		0.006	0.026	0.012
GMO	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.049*	*0.009*	0.160		0.007	***0.003***
GEO	*0.000*	*0.000*	*0.000*	*0.000*	0.221	*0.000*	0.115	*0.023*	*0.001*	0.115		0.007
MDA	*0.000*	*0.000*	*0.000*	*0.000*	*0.001*	*0.000*	*0.086*	0.127	*0.046*	0.334	0.152	

Table S1.5. Optimal number of genetic clusters (*K*) in 12 populations of *P. magellanicus* found by several analytical methods using 10 datasets listed in Table 3. K clusters indicated by * were suggested by the analysis, but no clear clustering could be observed.

Dataset Number	Dataset Description	Clustering Method	Number of clusters
1	All loci, all	Bayesian clustering, Structure	2
2	Outlier loci, all	Bayesian clustering, Structure	2
3	Neutral loci, all	Bayesian clustering, Structure	2
4	All loci, North	Bayesian clustering, Structure	2
5	Outlier loci, North	Bayesian clustering, Structure	2
6	Neutral loci, North	Bayesian clustering, Structure	2
7	All loci, South	Bayesian clustering, Structure	6*
8	Outlier loci, South	Bayesian clustering, Structure	2*
9	Neutral loci, South	Bayesian clustering, Structure	4*
10	High F_{ST}, South	Bayesian clustering, Structure	2*
1	All loci, all	k-means clustering, Principal components analysis, adegenet	1
2	Outlier loci, all	k-means clustering, Principal components analysis, adegenet	4
3	Neutral loci, all	k-means clustering, Principal components analysis, adegenet	1

Table S1.6. Estimates of the standard deviation of parent-offspring dispersal distance (in km) of *P. magellanicus* in (A) all populations, (B) north populations, and (C) south populations using all loci, neutral loci, and outlier loci, and low and high estimates of adult density calculated using the slope of the isolation-by-distance (IBD) relationship between pairwise population F_{ST} and pairwise population current-based distances and shortest ocean-based distances (*marmap*). Significance of the IBD relationship is indicated as follows: $p < 0.05$ = *, $p < 0.01$ = **, $p < 0.001$ = ***.

(A)

Loci included	Current		Least Cost	
	Low adult density	High adult density	Low adult density	High adult density
All loci	3.8111 ± 3.1294 *	1.6084 ± 0.7374 *	2.8502 ± 2.3404 ***	2.1506 ± 0.9860 ***
Neutral loci	5.5712 ± 4.5746	2.3993 ± 1.1000	4.2519 ± 3.4913	3.1438 ± 1.4412
Outlier loci	0.9506 ± 0.7805 **	0.3807 ± 0.1746 **	0.6747 ± 0.5540 ***	0.5364 ± 0.2459 ***

(B)

Loci included	Current		Least Cost	
	Low adult density	High adult density	Low adult density	High adult density
All loci	12.9331 ± 10.6196	7.2980 ± 3.3459	3.0842 ± 2.5325	1.7404 ± 0.7979
Neutral loci	10.9543 ± 8.9948	6.1814 ± 2.8339	3.6747 ± 3.0174	2.0736 ± 0.9507
Outlier loci	1.2141 ± 0.9969	0.6851 ± 0.3141	0.5597 ± 0.4596*	0.3158 ± 0.1448*

(C)

Loci included	Current		Least Cost	
	Low adult density	High adult density	Low adult density	High adult density
All loci	3.8111 ± 3.2194	2.1506 ± 0.9860	7.2589 ± 5.9604	4.0961 ± 1.8779
Neutral loci	5.5712 ± 4.5746	3.1438 ± 1.4413	5.6093 ± 4.6059	3.1652 ± 1.4511
Outlier loci	0.9506 ± 0.7805	0.5364 ± 0.2459	1.7271 ± 1.4182	0.9746 ± 0.4468

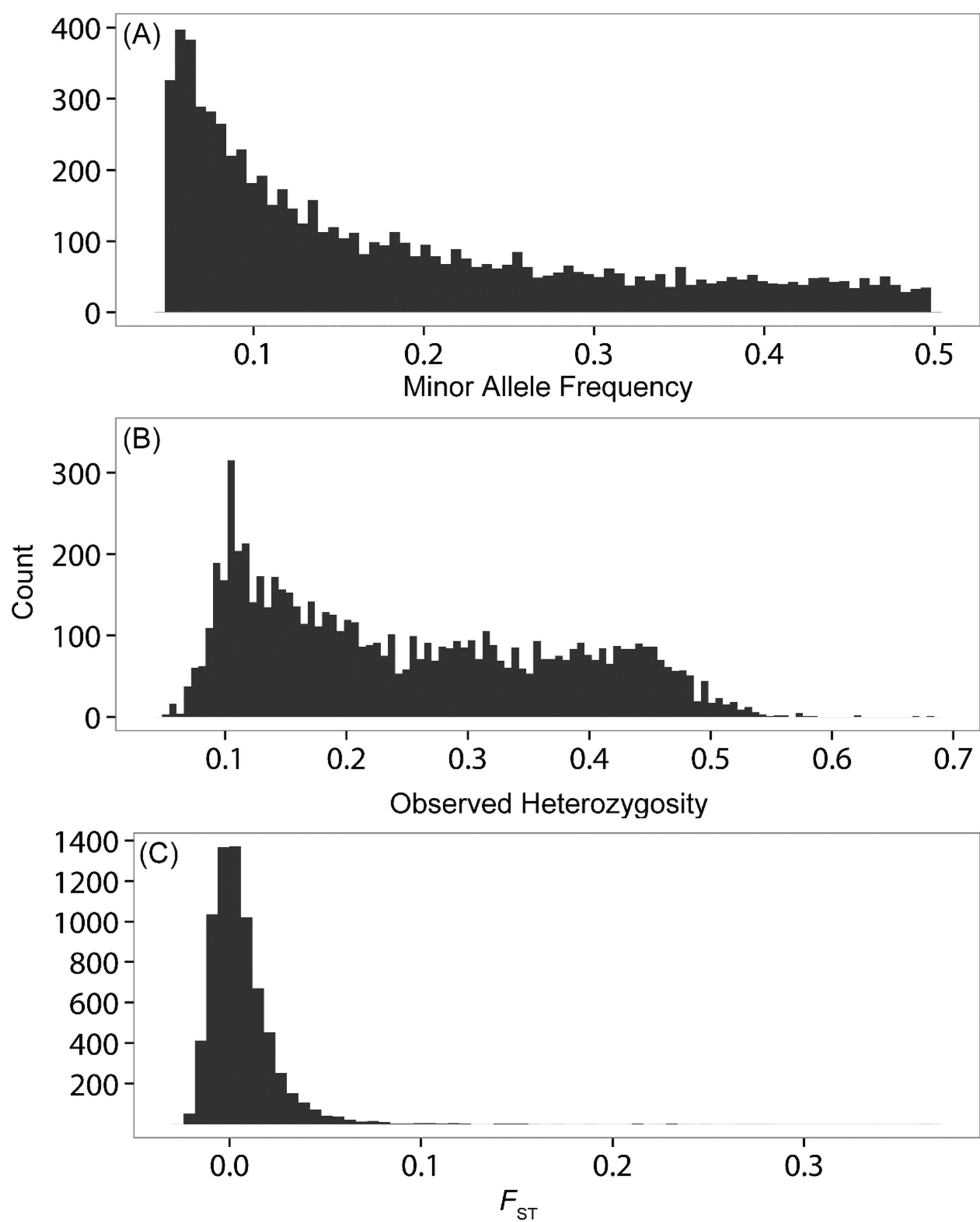

Figure S1.1. Histogram of minor allele frequencies, observed heterozygosity, F_{ST} for 7163 SNP loci sequenced in 245 adult *P. magellanicus*.

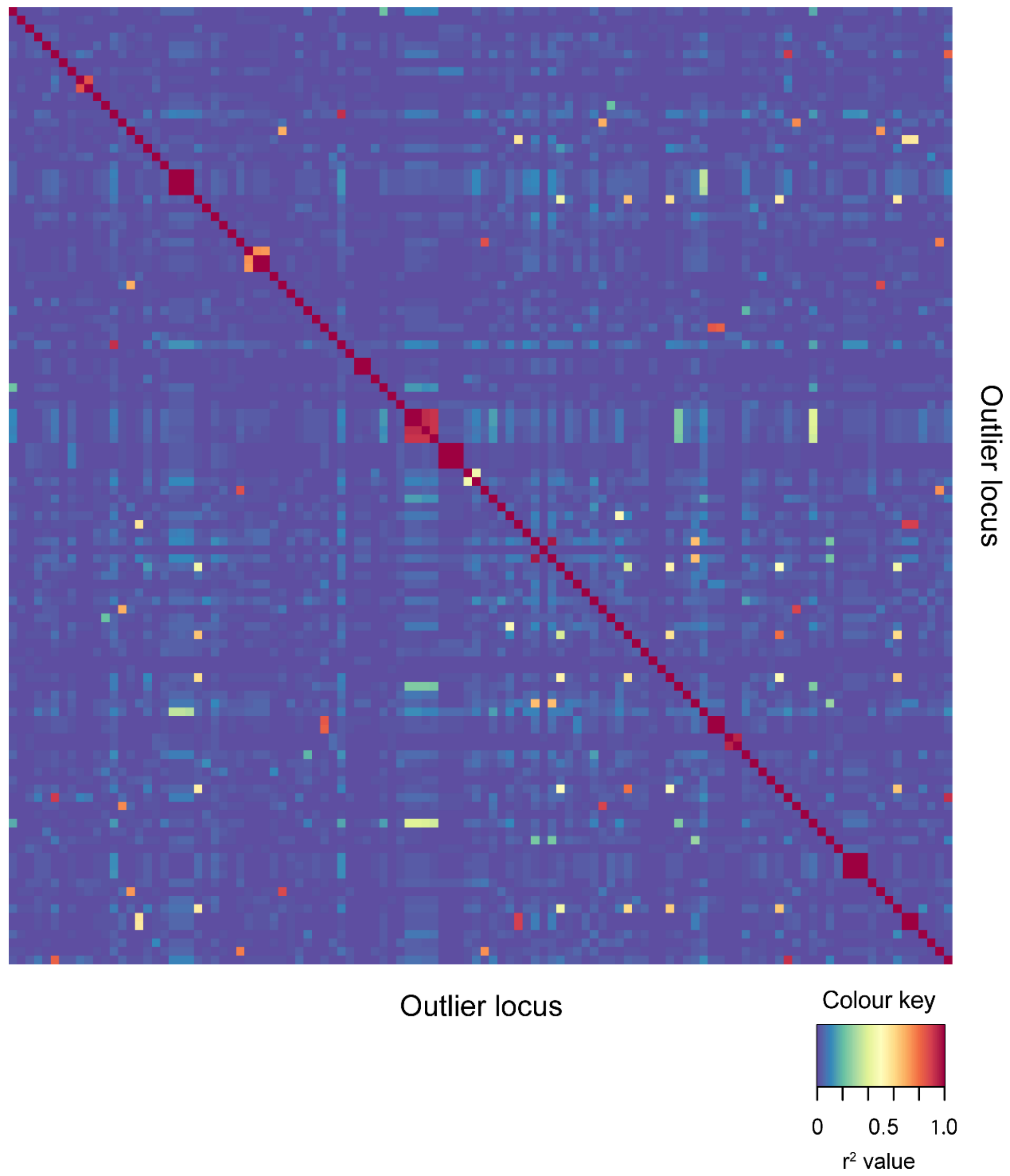

Figure S1.2. Heatmap of pairwise linkage disequilibrium r^2 values calculated for 112 outlier loci detected among 12 populations of *P. magellanicus*.

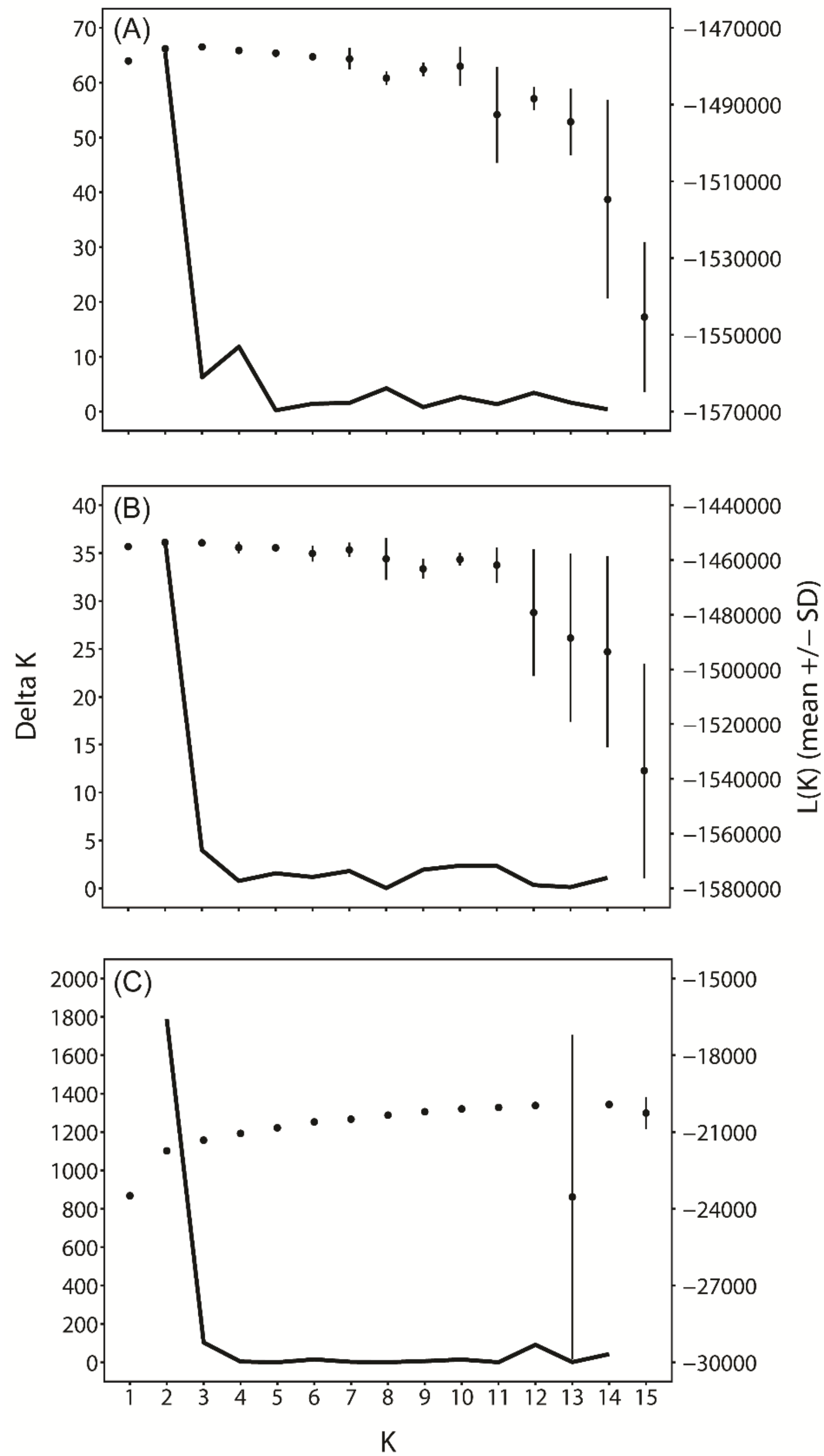

Figure S1.3. Delta *K* (solid line) and Ln(*K*) determined by Bayesian clustering implemented in the program Structure for 12 populations of *P. magellanicus* using (a) all loci, (b) neutral loci, and (c) outlier loci for *K* = 1-15.

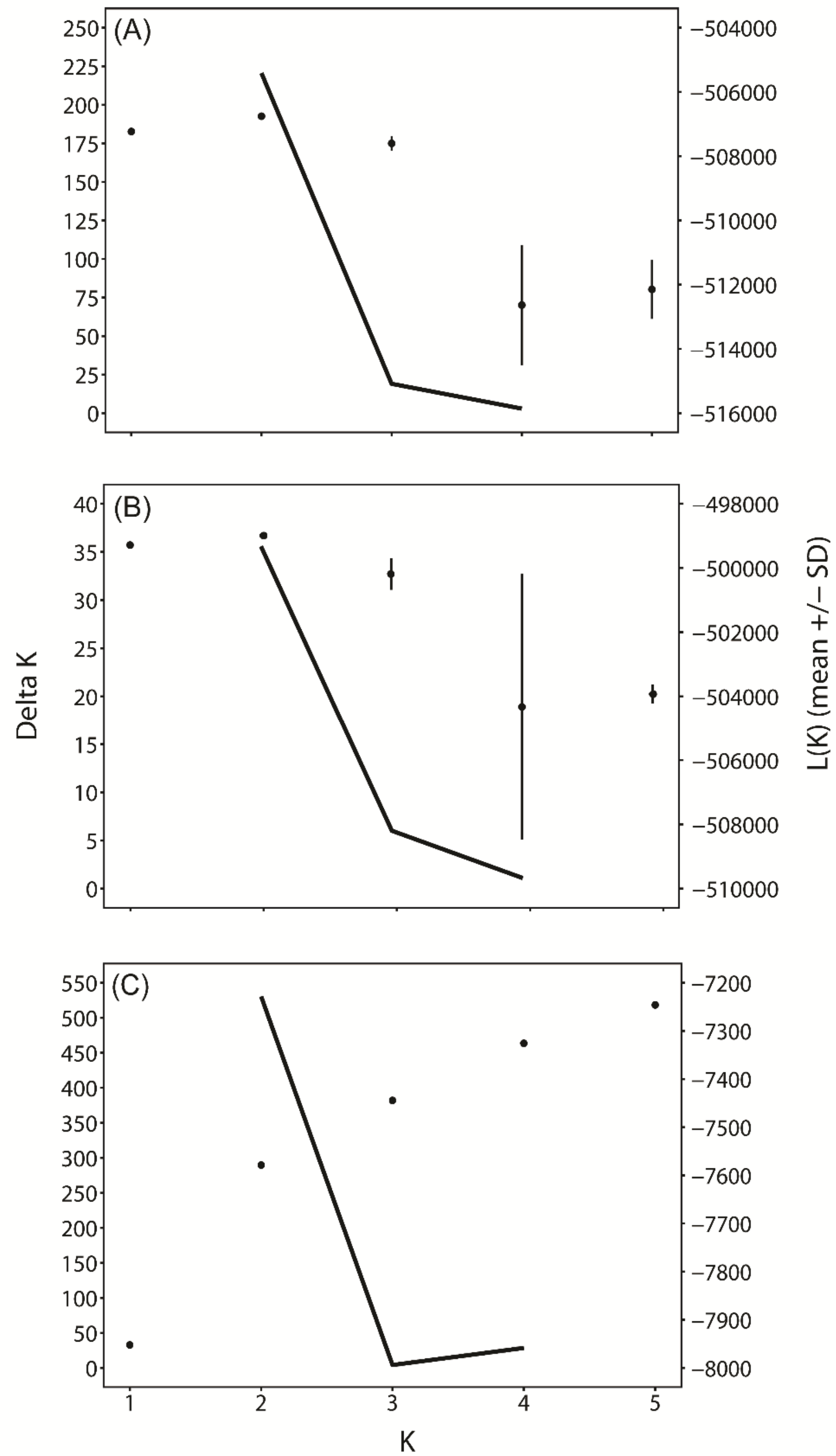

Figure S1.4. Delta *K* (solid line) and Ln(*K*) determined by Bayesian clustering implemented in the program Structure for 4 north populations of *P. magellanicus* using (a) all loci, (b) neutral loci, and (c) outlier loci for *K* = 1-5.

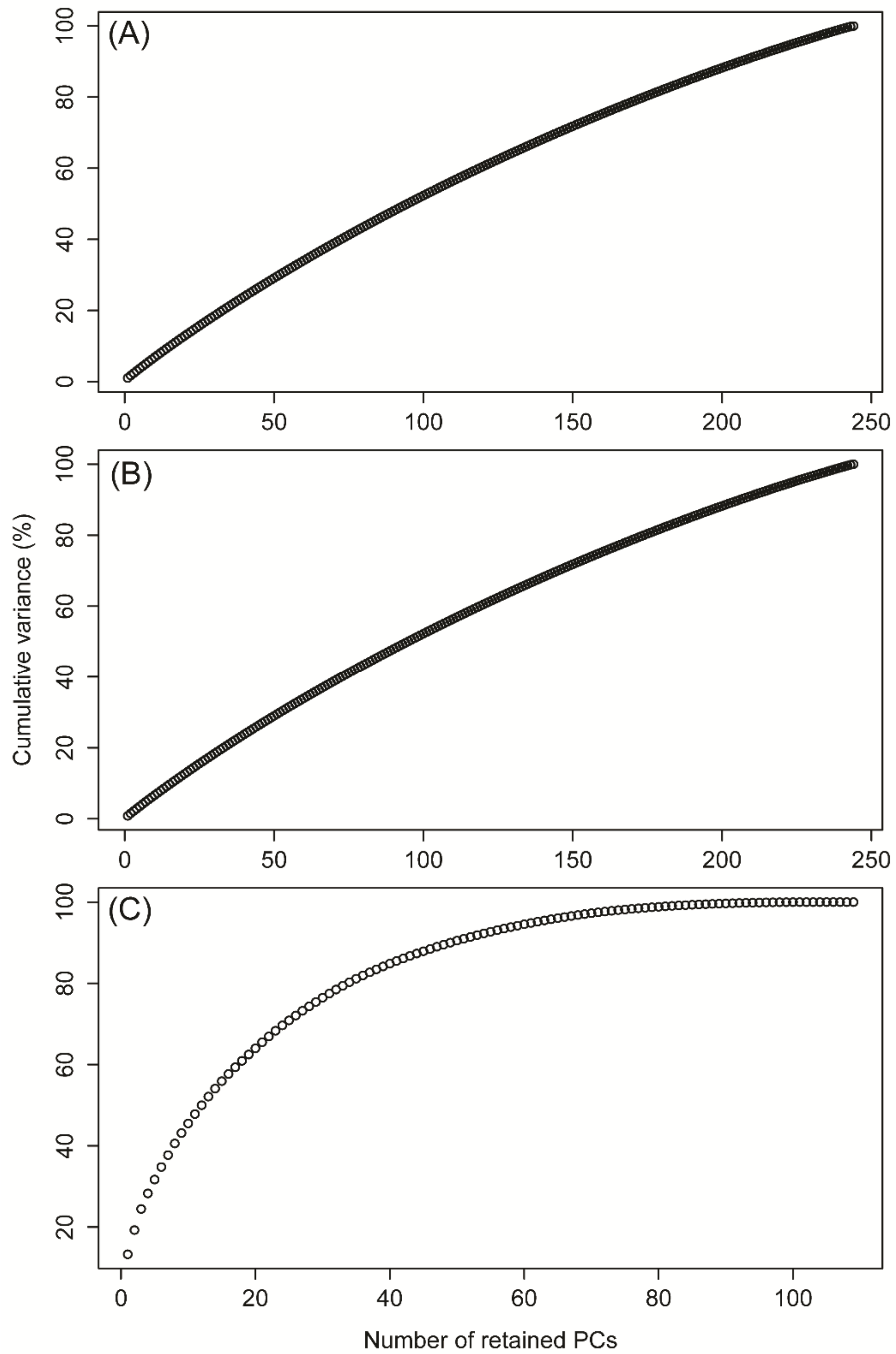

Figure S1.5. Cumulative variance explained by the principal components of a principal components analysis of 12 populations of *P. magellanicus* using (a) all loci, (b) neutral loci, and (c) outlier loci.

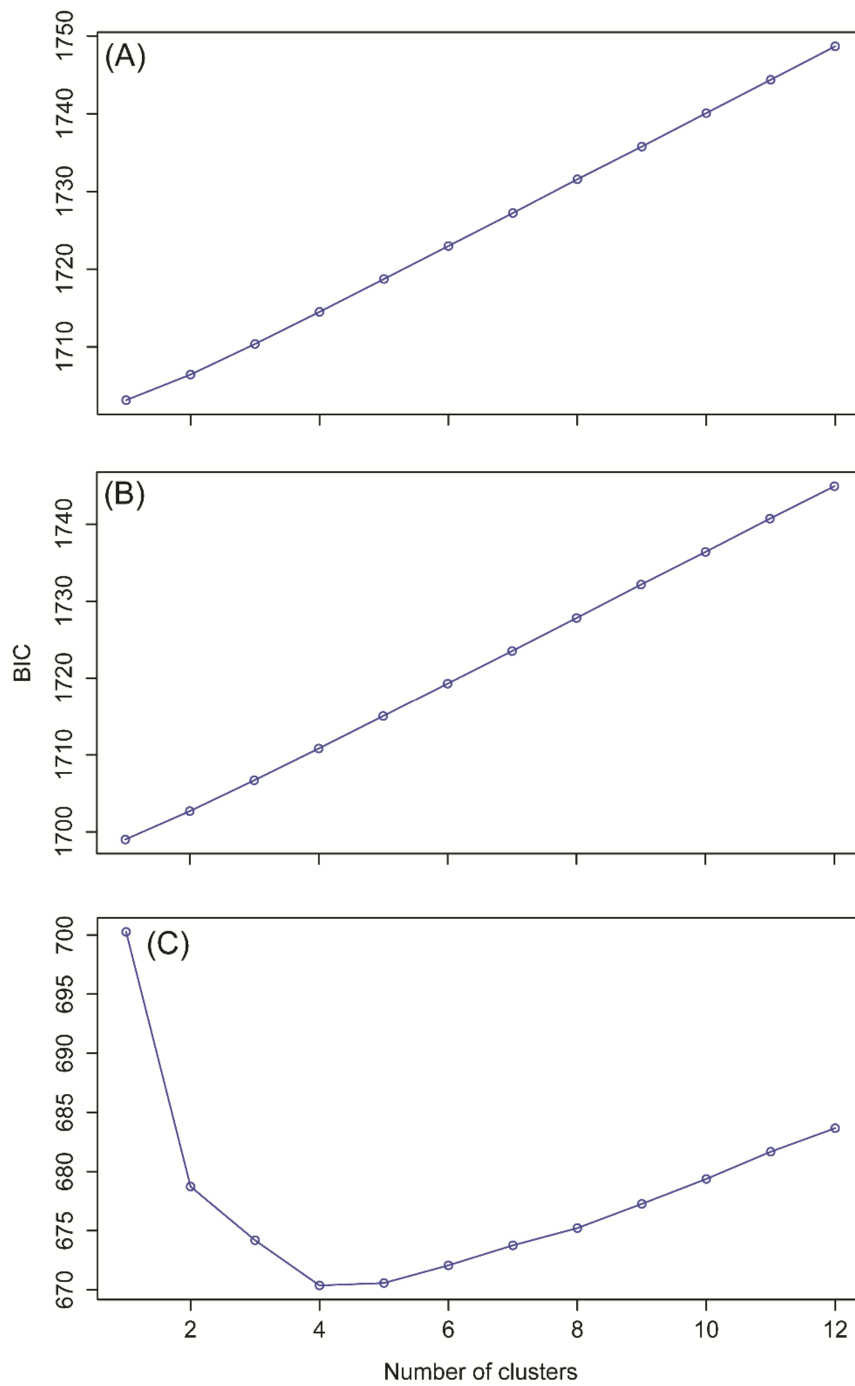

Figure S1.6. Bayesian Information Criterion values generated from *k*-means clustering of principal components of 12 populations of *P. magellanicus* for (a) all loci, (b) neutral loci, and (c) outlier loci.

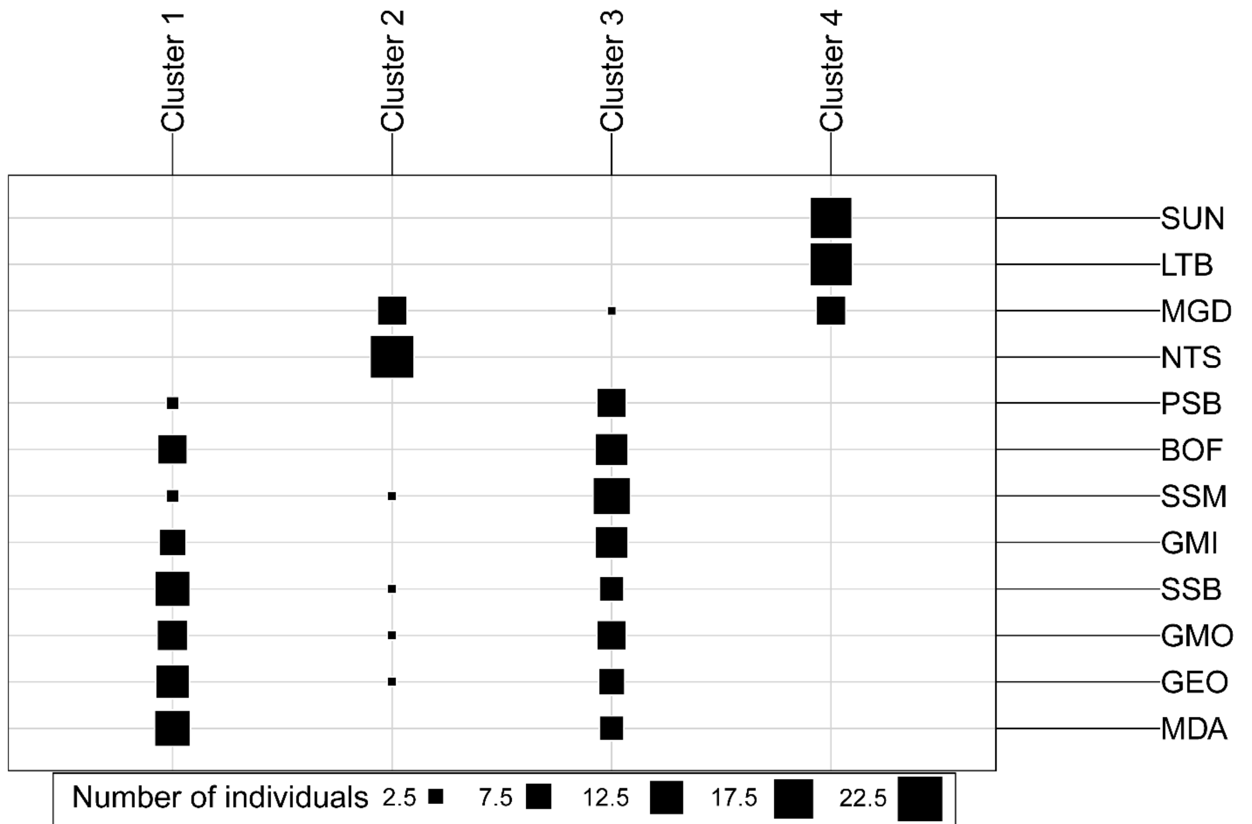

Figure S1.7. Plot showing the individual-specific membership of 12 *P. magellanicus* populations in each of 4 genetic clusters identified by *k*-means clustering on the principal components analysis of the outlier loci. Square size corresponds to the number of individuals from each population assigned to a particular cluster.

Appendix S2: Supporting Information for **Chapter 3:** Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*

Table S2.1. Number of loci and minor allele frequency (MAF) range included in each of five bins (based on global minor allele frequency) used by the program BayEnv2 to detect loci potentially under selection among 12 populations of *P. magellanicus*.

Bin	**Number of Loci**	**MAF range**
A	3566	0.05 - 0.139
B	1390	0.14 - 0.229
C	908	0.23 - 0.319
D	691	0.32 - 0.409
E	608	0.41 - 0.5

Table S2.2. Number of *P. magellanicus* individuals and number of SNP loci included in initial RAD-sequencing and final analysis following quality control (QC).

Parameter	Value
Individuals sequenced	252
Individuals following QC	*245* *(97.2% of Individuals sequenced)*
Initial RAD tags	131897
RAD tags following QC	*19672* *(14.9% of Initial RAD tags)*
Initial SNPs	173482
SNPs following QC	*7216* *(4.2% of Initial SNPs)*
SNPs in HWE	7163 (99.3% of SNPs following QC)

Table S2.3. Comparison of outlier SNP loci from 12 populations of *P. magellanicus* determined using BayEnv2 and a dataset of 90 environmental variables (AllEnv, 128 loci) and a subset of 36 environmental variables (CST, 72 loci). Shared loci are italicized.

AllEnv	**CST**
319_43	479_44
479_15	545_13
1089_23	*1089_23*
1299_57	*1299_57*
1320_61	*1320_61*
2272_32	*2272_32*
2966_52	2305_45
3061_77	2694_52
3069_13	*3240_12*
3199_76	3240_55
3240_12	*3240_56*
3240_56	*3929_9*
3299_78	4484_19
3350_55	*4668_81*
3420_33	4769_32
3498_66	5062_61
3619_79	5353_70
3834_87	*5439_27*
3929_9	6968_50
4063_7	7115_35
4383_42	7325_7
4668_81	7326_20
4847_58	8388_15
4975_23	8664_7
5045_48	*9554_49*
5084_54	9881_90
5084_64	*10070_67*
5161_82	*10832_84*
5247_15	*11531_7*
5420_31	12021_80

5439_27	*12155_52*
6805_6	*12314_60*
7120_90	*12767_68*
7203_45	*13384_35*
7343_38	13561_17
7361_21	13561_70
7882_64	13761_22
7997_36	13805_16
8325_27	*13824_83*
8655_19	*13891_65*
8782_41	14003_84
8922_62	*14197_90*
9203_74	*14394_27*
9431_86	15121_78
9554_49	15446_21
9653_60	*15645_89*
9976_51	*15821_29*
10070_67	*16110_74*
10159_36	16229_44
10366_19	17320_31
10498_47	17856_41
10529_34	*18391_26*
10832_74	*18439_32*
10832_84	18645_34
10917_59	*18669_56*
10987_36	19330_41
11265_39	19880_7
11531_7	*20263_79*
11748_62	*20298_27*
12014_82	*20298_73*
12155_52	20311_14
12314_60	20633_88
12767_68	20810_86
13260_51	*21297_65*
13384_35	*22123_81*
13438_71	*22237_35*
13744_54	25114_12
13824_83	*25405_78*
13891_65	*25627_51*
14076_10	25646_16
14197_86	*25748_78*

14197_90	*26519_27*
14394_27	
14421_56	
14557_41	
14571_64	
15171_47	
15245_72	
15459_62	
15645_89	
15660_9	
15722_73	
15821_29	
15876_15	
16110_74	
16161_6	
16206_40	
16309_22	
16393_64	
16517_82	
16943_74	
17480_65	
17720_35	
17853_46	
18135_35	
18238_55	
18391_26	
18439_32	
18669_56	
19353_39	
19487_24	
19960_23	
20047_88	
20263_79	
20298_27	
20298_73	
20548_19	
20596_34	
20910_86	
21297_65	
21600_62	
21679_27	

21717_37
22123_81
22237_35
23535_28
23969_73
24257_63
24939_35
25067_48
25405_78
25405_86
25627_51
25630_84
25748_78
26519_27
26726_84
27045_28

Table S2.4. Comparison of outlier SNP loci from 12 populations of *P. magellanicus* determined using latent factor mixed models and a dataset of 90 environmental variables (AllEnv, 511 loci) and a subset of 36 environmental variables (CST, 218 loci). Shared loci are italicized.

AllEnv	**CST**
85_6	*370_51*
367_35	*370_60*
367_88	*618_44*
370_51	*620_25*
370_60	*1115_68*
418_86	*1122_21*
450_37	*1463_75*
618_44	*1463_88*
620_25	*1471_75*
620_53	*1484_21*
874_57	*1714_38*
1115_23	*1867_71*
1115_68	*2450_62*
1122_21	*2670_59*
1129_54	*2809_37*
1143_46	*3216_12*
1383_82	*3303_73*
1459_43	*3895_8*
1463_75	*3895_88*
1463_88	*4066_45*
1471_75	*4186_46*
1484_21	*4546_80*
1566_34	*4613_35*
1683_20	*4665_62*
1689_82	*4668_42*
1714_38	*4668_81*
1742_90	*4844_25*
1743_28	*4956_31*
1847_71	*4970_90*
1856_80	*4975_9*

1856_82	*4975_68*
1867_71	*5170_73*
1974_44	*5204_27*
2334_90	*5354_34*
2395_23	*5419_44*
2450_62	*5522_80*
2670_59	*5842_16*
2809_37	*5978_53*
3000_6	*6026_30*
3028_9	*6026_53*
3124_7	*6190_47*
3124_74	*6491_67*
3124_8	*6579_25*
3216_12	*6591_67*
3286_90	*6645_6*
3303_73	*6688_49*
3381_74	*6976_68*
3420_33	*7042_35*
3895_8	*7376_36*
3895_88	*7408_7*
3922_74	*7648_39*
3922_84	*7724_27*
3936_13	*7846_50*
3945_18	*7941_58*
3975_33	*7999_74*
4004_43	*8380_74*
4066_45	*8520_42*
4186_46	*8575_6*
4320_27	*8575_39*
4546_80	*8794_9*
4613_35	*8797_34*
4665_62	*8836_15*
4667_77	*8847_48*
4668_42	*8849_28*
4668_66	*9076_28*
4668_81	*9076_54*
4729_44	*9098_83*
4832_89	*9169_48*
4844_25	*9196_22*
4956_31	*9196_23*
4970_90	*9196_87*

4975_23
4975_68
4975_9
5070_52
5170_73
5204_27
5298_76
5354_34
5363_56
5394_86
5419_44
5443_24
5443_32
5443_6
5443_67
5515_42
5522_80
5807_51
5842_16
5978_53
6026_30
6026_53
6186_46
6190_47
6191_10
6203_48
6228_45
6228_84
6236_60
6288_14
6346_68
6491_67
6579_25
6591_67
6604_68
6645_6
6688_49
6759_11
6810_7
6810_79
6887_79

9275_68
9275_76
9275_79
9275_81
9508_72
10464_33
10464_52
10681_10
10987_75
11131_89
11217_9
11315_68
11634_72
11909_85
11989_13
12014_82
12228_13
12236_9
12314_60
12363_39
12580_62
12639_88
12651_47
12680_26
12880_57
12884_30
13012_81
13075_87
14197_88
14197_90
14316_32
14326_62
14415_20
14532_68
14564_90
15508_47
15638_70
15759_32
15759_54
15903_58
16388_50

6888_15	*16442_35*
6888_38	*16442_83*
6945_90	*16442_86*
6976_68	*16470_49*
7042_35	*16691_70*
7114_22	*16771_68*
7318_31	*16961_32*
7376_36	*17189_71*
7408_7	*17191_81*
7423_59	*17259_54*
7423_71	*17259_55*
7540_65	*17369_30*
7626_32	*17377_63*
7637_73	*17406_14*
7648_39	*17691_29*
7700_76	*17793_42*
7724_27	*17865_54*
7724_45	*17939_43*
7737_36	*18011_51*
7846_50	*18013_68*
7857_24	*18084_42*
7857_32	*18084_67*
7857_74	*18175_82*
7857_8	*18180_41*
7941_58	*18180_85*
7999_74	*18180_90*
8380_74	*18272_54*
8520_42	*18286_51*
8540_90	*18411_69*
8563_90	*18473_63*
8575_39	*18555_47*
8575_6	*18669_56*
8635_12	*18692_21*
8735_68	*18692_22*
8741_86	*18800_64*
8794_56	*18832_29*
8794_9	*18968_21*
8797_34	*19146_65*
8797_89	*19197_77*
8836_15	*19256_31*
8847_48	*19284_7*

8849_28	*19284_8*
9076_28	*19478_77*
9076_31	*19738_6*
9076_54	*19738_49*
9098_83	*19752_32*
9163_31	*19844_50*
9169_48	*20070_71*
9196_22	*20140_47*
9196_23	*20314_14*
9196_87	*20314_46*
9211_22	*20556_7*
9275_68	*20556_14*
9275_76	*20556_31*
9275_79	*20559_46*
9275_81	*20561_41*
9281_51	*20721_17*
9387_50	*20868_9*
9387_7	*20868_67*
9431_43	*20885_25*
9431_86	*20933_88*
9471_77	*21167_6*
9508_72	*21362_67*
9996_36	*21381_83*
10141_55	*21413_20*
10156_85	*21413_21*
10361_56	*21413_48*
10464_33	*21503_20*
10464_52	*21696_25*
10485_37	*22000_20*
10498_47	*22000_86*
10527_56	*22011_10*
10587_84	*22011_82*
10681_10	*23849_13*
10810_32	*24019_69*
10832_30	*24336_71*
10832_46	*24406_47*
10832_74	*24509_20*
10832_84	*24636_63*
10873_33	*24673_67*
10987_74	*24892_12*
10987_75	*24908_10*

10987_8
10993_65
11126_44
11131_89
11217_9
11315_68
11407_80
11431_85
11634_72
11696_20
11696_43
11784_88
11810_49
11909_85
11966_39
11989_13
12014_82
12031_71
12041_65
12228_13
12236_9
12314_60
12363_39
12370_76
12370_81
12377_67
12419_58
12580_62
12639_88
12651_47
12651_75
12674_8
12680_26
12880_57
12884_30
12936_75
12969_14
12969_15
12969_59
13012_81
13020_90

25107_50
25246_11
25380_52
25380_84
25405_6
25405_9
25433_29
25433_34
25433_45
25433_82
25627_51
25993_87
26222_72
26293_14
26340_40
26431_8
26431_77
26625_75
26721_7
26721_89
26781_56
26880_82
26952_89
27011_89

13075_87
13109_21
13197_42
13270_21
13286_17
13303_49
13303_64
13303_79
13374_27
13438_71
13586_43
13713_14
13727_6
13727_69
13758_31
13966_68
13983_54
14018_57
14050_63
14117_19
14117_7
14139_22
14139_68
14185_21
14197_88
14197_90
14287_61
14310_37
14316_11
14316_32
14326_62
14415_20
14532_68
14564_90
14674_22
14967_65
15022_55
15090_47
15203_19
15205_46
15205_67

15261_65
15301_34
15508_47
15638_54
15638_70
15687_33
15759_32
15759_54
15900_50
15903_35
15903_58
15903_76
15903_78
16087_68
16161_58
16161_6
16161_63
16161_64
16232_32
16232_64
16300_20
16339_61
16356_41
16388_50
16396_57
16442_35
16442_83
16442_86
16466_34
16470_49
16691_70
16771_68
16809_40
16837_37
16869_84
16891_32
16961_32
17189_71
17191_81
17259_54
17259_55

17369_30
17377_63
17382_16
17406_14
17466_32
17581_60
17581_65
17632_52
17691_29
17731_31
17793_42
17819_48
17865_54
17929_79
17929_9
17939_43
17998_24
18011_51
18013_68
18084_42
18084_67
18175_82
18178_75
18180_41
18180_85
18180_90
18238_55
18238_56
18272_54
18286_51
18296_75
18411_69
18444_28
18473_63
18481_72
18555_47
18642_16
18669_56
18692_21
18692_22
18701_58

18701_60
18701_83
18800_64
18832_29
18865_67
18968_21
18972_74
19146_65
19197_77
19218_80
19252_8
19256_31
19284_24
19284_32
19284_7
19284_8
19302_50
19330_39
19478_77
19654_11
19738_49
19738_6
19752_32
19819_35
19844_50
19898_23
19898_6
19898_61
20015_23
20047_71
20047_79
20047_87
20062_71
20070_71
20106_11
20117_82
20140_47
20314_14
20314_46
20527_29
20556_14

20556_31
20556_59
20556_7
20559_46
20561_41
20641_46
20721_17
20868_56
20868_67
20868_9
20885_25
20929_40
20933_88
20942_64
20955_8
21028_28
21125_75
21138_65
21167_6
21362_67
21381_83
21413_20
21413_21
21413_48
21418_49
21468_29
21468_89
21503_20
21600_62
21696_25
21746_21
21762_13
21853_20
22000_20
22000_86
22011_10
22011_82
22100_34
22122_62
23183_26
23296_73

23306_43
23346_81
23361_35
23849_13
24019_69
24230_82
24336_71
24373_14
24380_74
24384_24
24406_47
24509_20
24587_47
24636_63
24673_67
24804_65
24882_77
24892_12
24892_33
24908_10
24949_72
24954_38
25064_55
25107_50
25246_11
25310_43
25310_48
25366_26
25380_52
25380_58
25380_84
25405_6
25405_86
25405_9
25433_29
25433_34
25433_45
25433_82
25464_47
25562_55
25580_62

25599_88
25627_51
25630_84
25652_55
25993_87
26091_42
26091_9
26141_15
26141_84
26222_72
26293_14
26293_68
26312_54
26337_72
26340_40
26371_65
26391_35
26391_71
26431_77
26431_8
26550_72
26570_14
26625_75
26721_7
26721_89
26781_56
26880_82
26937_69
26952_89
27011_89

Table S2.5. Blast2GO functional annotation and BLAST gene matches for outlier SNP loci from 12 populations of *P. magellanicus*. Outliers were detected through environmental correlations with 90 environmental variables (AllEnv, 621 loci) or a subset of 26 environmental variables (CST, 285 loci).

Environmental Data	Locus Name	GO Name	Best Gene Match
AllEnv	16087_68	F: catalytic activity; P: metabolic process; F: transferase activity; F: folic acid binding; P: cellular metabolic processes	Predicted formimidoyltransferase cyclodeaminase (various species)
AllEnv	24384_24	F: oxidoreductase activity; P: metabolic process; P: oxidation-reduction process	Predicted 2,4-dienoyl CoA reductase 1, mitochondrial (DECR1), mRNA (various species)
AllEnv CST	12228_13	P: regulation of transcription; P: steroid hormone mediated signaling pathway	GU932671.1: *Lymnaea stagnalis* retinoic acid receptor (RAR) mRNA, complete cds
AllEnv CST	20561_41	F: carbohydrate binding	JN166712.1: *Argopecten irradians* CTL-9 mRNA, complete cds
AllEnv CST	25748_78	F: calcium ion binding	Various predicted scaffold proteins (various species)
CST	15446_21	F: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced pteridine as one donor, and incorporation of one atom of oxygen; P: oxidation-reduction process	Various scaffold and other predicted proteins (various species)

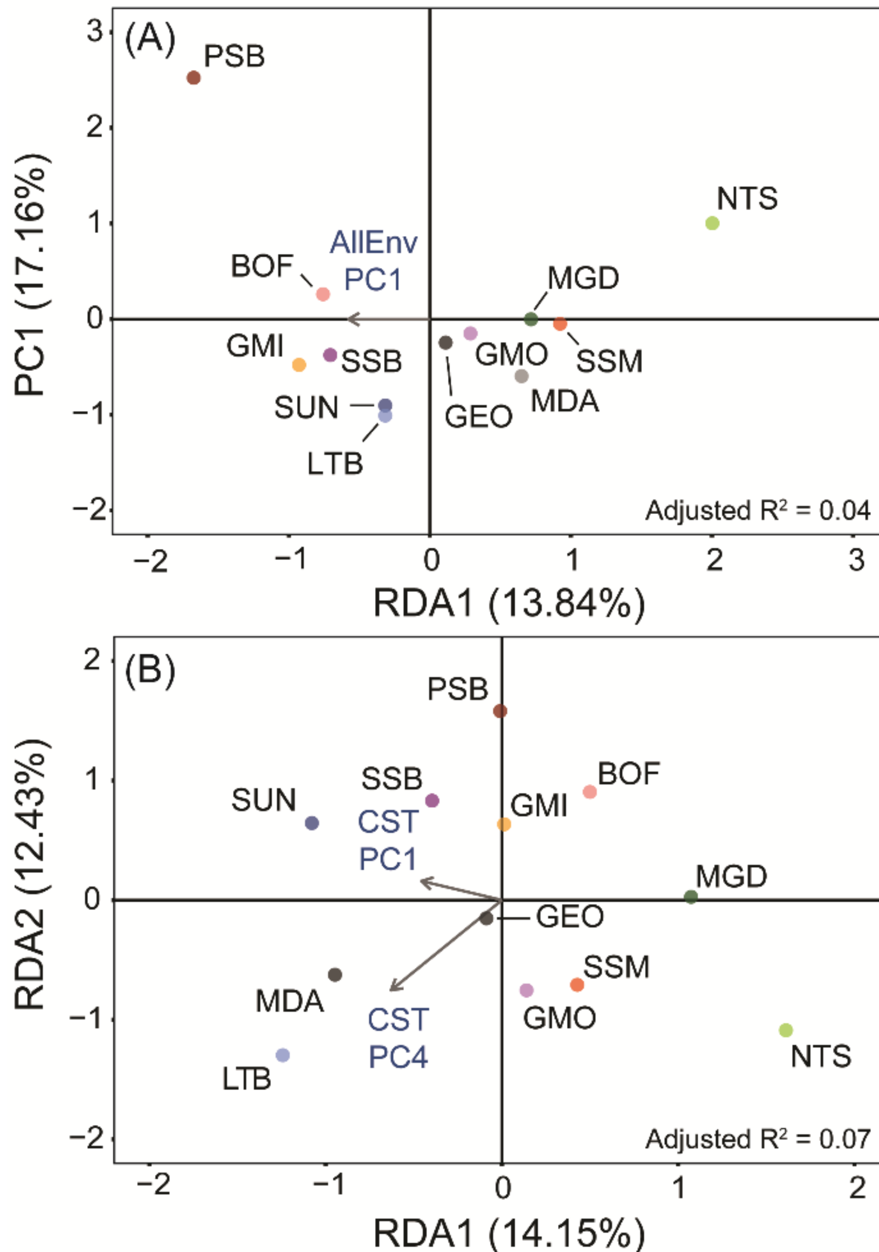

Figure S2.1. Partial Redundancy analysis plots for loci detected as potentially under selection through environmental correlation with (a) AllEnv (90 environmental variables, n= 621 loci), (b) CST (36 environmental variables, n = 285 loci) in 12 populations of *P. magellanicus*. Explanatory variables used were principal components axes from PCA on (a) AllEnv and (b) CST, retained following backwards stepwise variable selection. The genetic matrix was conditioned on the distance between populations to reduce the effects of geographic separation between populations.