Molecular Divergence and Gene Flow in Cold Ocean Tunicates: Insights from *Cytochrome Oxidase 1*

by

©Ainsley R Latour, B.Sc., B.Ed.

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Master of Science

Department of Biology, Faculty of Science
Memorial University of Newfoundland

September 2021
St. John's, Newfoundland and Labrador

ABSTRACT

Gaps in our ability to identify species and understand factors influencing the evolution of species ranges, limit the extent to which we can diagnose changes to ocean ecosystems that occur in response to anthropogenic stressors. DNA barcodes offer one method to facilitate our assessment and understanding of diversity. Tunicates are a group of marine invertebrates for which researchers have used the typical animal barcoding region of the *cytochrome oxidase 1* (*CO1*) gene to identify cryptic species complexes and elucidate patterns of gene flow in non-indigenous ascidians. Waters surrounding Newfoundland contain several species of invasive ascidians and native appendicularian tunicates. My research utilized the *CO1* animal barcoding region to explore the genetic structure of two cold ocean tunicates in Newfoundland waters, the native appendicularian tunicate, *Oikopleura vanhoeffeni* and the non-native ascidian *Botryllus schlosseri.* Specifically, I provide the first report of a partial *CO1* sequence for *O. vanhoeffeni and report on gene flow of B. schlosseri* collected from Newfoundland waters. The *CO1* gene proved unsuitable for further evaluation as a marker for species level identification in *O. vanhoeffeni*, because the presence of tandem repeat regions would require multiple primer pairs to amplify the *CO1* barcoding region. Across Newfoundland harbours, considerable genetic structure among the 7 *Botryllus schlosseri* populations sampled indicates increasing genetic diversity from west to east. All of the *B. schlosseri CO1* haplotypes identified in Newfoundland waters were recorded previously in Nova Scotian harbours, and a few in the Mediterranean Sea. Thus, regional boating traffic from mainland Atlantic Canada likely provided an important vector of introduction for this species in Newfoundland, but transoceanic and/or local boating traffic may play a role.

ACKNOWLEDGEMENTS

First, I thank my supervisors Dr. D. Deibel and Dr. H. Dawn Marshall for all their support throughout my project.

I thank my supervisory committee members, Dr. I. Bradbury (Department of Fisheries and Oceans/Memorial University) and Dr. P. Snelgrove (Memorial University) for their insightful feedback throughout my project. I also thank Dr. C. McKenzie and Dr. B. Lowen (Department of Fisheries and Oceans Canada) for their collaboration in this project.

Additionally, I acknowledge the staff of the Field and Laboratory service units of the Ocean Sciences Centre, including R. O'Donnell, M. O'Flaherty, and A. Perry for their assistance in the field. I wish to thank Dr. J. Hall, B. Pilgrim and L. Tao of the GaP (Genomics and Proteomics) Laboratory (Memorial University of Newfoundland). For the actual generation of sequences from PCR products, I thank Genome Quebec.

I acknowledge fellow graduate students Dr. K. Carew and Dr. B. Langille.

I am deeply grateful for the support of my parents and extended Newfoundland family. Finally, I dedicate this work to my husband, Dr. M. Sukhai, for his unwavering love.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND SYMBOLS

AMOVA: analysis of molecular variance

bp: base pair

°C: Degrees Celsius

CO1: cytochrome c oxidase I

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleotide triphosphate

DOM: dissolved organic matter

F: forward primer

F_{CT} or $F_{CT:}$ Proportion of total variation among populations

F_{SC} or F_{SC}: Proportion of total variation among subpopulations

F_{ST} or F_{ST}: Proportion of total variation within subpopulations

GaP: Genomics and Proteomics

h or *h:* Haplotype diversity

ITD2: internal transcribed spacer 2

mtDNA: Mitochondrial Deoxyribonucleic acid

n or *n:* Sample size

ND2: dehydrogenase subunit 2

NL: Newfoundland and Labrador

NS: Nova Scotia

PCR: Polymerase chain reaction

R: Reverse primer

RNA: Ribonucleic acid

s: Seconds

T_a: Annealing temperature

Taq: Thermus aquaticus

µL: Microlitre

π: Nucleotide diversity

CHAPTER 1: GENERAL INTRODUCTION

Changes in sea ice cover, ocean temperatures, salinity, and the introduction of non-indigenous species affect the distribution of marine species by altering the distribution of suitable habitat and changing dispersal pathways (Hardy et al. 2011). Human activity, including shipping, influences many of these variables, and can thus be considered anthropogenic stressors. Gaps in our current knowledge of species diversity and range limits, as well as the ecological and evolutionary processes facilitating range expansion in marine biota, limit current capacity to diagnose changes to ocean ecosystems that occur in response to anthropogenic stressors.

Increasingly, modern research methods of research use deoxyribonucleic acid (DNA) barcodes or short, species specific sequences of DNA in addressing ecological, evolutionary, and biogeographical questions with regards to marine invertebrates. DNA barcodes have helped identify prey in gut contents and juvenile or damaged zooplankton specimens (Bucklin et al. 2010b), and to provide early warning of presence of non-indigenous species by detection of eggs or larvae in water samples (Briski et al. 2011). A growing number of DNA barcode sequences have also been used in population and phylogenetic analyses to delineate patterns of taxonomically significant variation within and among species across their distributions (Bucklin et al 2010a). The application of barcodes to delimit species boundaries (i.e. DNA taxonomy) has also resulted in the identification of hybrids and new or cryptic species (e.g. Brasier et al. 2016).

A fragment of DNA that is associated with a certain location within the genome is referred to as a molecular marker (Halliburton 2004). Molecular markers from both nuclear (e.g., *ITS2*) and mitochondrial genomes (e.g. *CO1, 12S, 16S*, and *ND2*) have been used as DNA barcodes (Goetze 2010; Geller et al. 2013; Vella et al. 2017; Milan et al. 2020). A molecular

marker must meet certain criteria to be used as a DNA barcode for phylogenetic and population studies. Specifically, the molecular marker must exhibit sufficient variability within and among species (Hebert et al. 2003). The potential of a DNA barcode to assign reliably a specimen to a given species depends on the extent of, and separation between, intraspecific variation and interspecific divergence in the selected marker (Meyer and Paulay 2005). Ideally, there is a lack of overlap between intra and interspecific variation that is referred to as the "barcode gap" (Meyer and Paulay 2005). In addition, where the goal is to design primers that will detect and distinguish most of the species in the studied group of taxa, then the presence of conserved flanking sites for universal polymerase chain reaction (PCR) primers in the molecular marker is required. For many marine taxa, the use of mitochondrial genes compared to nuclear sequences as DNA barcodes is advantageous. Unlike nuclear genes, mitochondrial genes belong to a haploid genome, lack introns, exhibit limited recombination and are generally maternally inherited (Bucklin et al. 2010b).

For animals, a 650 base pair (bp) region of the 5' end of the cytochrome oxidase 1 gene (*CO1*) is used often as a DNA barcode because this region exhibits sufficient variability within and among species (Hebert et al. 2003). This region of *CO1* in animals contains conserved sites, where universal primers can be used to amplify this sequence in a broad spectrum of phyla (Folmer et al. 1994). While many studies have used other mitochondrial genes as molecular makers for DNA barcoding, some of these genes have the disadvantage in the presence of insertions and deletions that make sequence alignment and comparative analyses more challenging (Bucklin et al. 2010b). The *CO1* fragment has been used as molecular marker for DNA barcoding for species identification across the 14 metazoan phyla and 20+ invertebrate classes occurring in the pelagic environment (Bucklin et al. 2010a,c).

Despite the utility of the *CO1* fragment as a molecular marker for the identification of many marine species, the presence of introns and/or very high rates of molecular evolution may limit the use of this region as a DNA barcode in some marine metazoans (Bucklin et al. 2010b). Species in which the presence of non-coding regions or non-conserved universal primer sites interrupt the *CO1* sequence may necessitate the identification of an alternative barcoding region or gene, because the use of multiple or species-specific primers does not facilitate fast and inexpensive efforts to identify species in environmental DNA (Rach et al. 2017). Examples of groups of marine invertebrates with species in which intron(s) interrupt the *CO1* sequence include the Cnidarians (Beagley et al. 1996, Fukamio et al. 2007), Placozoans (Burger et al. 2009), and Porifera (Rot et al. 2006, Erpenbeck et al. 2015).

The *cytochrome oxidase 1* gene is also of limited use as a molecular marker for species delimitation where either a slow rate of nucleotide substitution or relatively recent divergence of the species results in similar or identical DNA barcodes (Bucklin et al. 2010b). *CO1* haplotypes can be shared between two distinct species of marine invertebrates (Cardenas et al. 2009). Incomplete lineage sorting or introgression can result in very similar or identical barcodes in distinct but sympatric species (Kemppainen et al. 2009; Steinke et al. 2009). In sponges, the *CO1* barcode cannot adequately discriminate between some closely related species because of an evolutionary rate of mitochondrial DNA change 10 to 20 times lower than in other metazoans (Shearer et al. 2002; Huang et al. 2008). Thus, in some species, the molecular marker *(CO1)* typically used for animal DNA barcoding may be ineffective to resolve interspecific diversity in some groups.

In the past, some authors have used other single genes as markers to elucidate the taxonomic relationships among marine zooplankton, such as *ITS2* and*16S* (Goetze 2010) and *18S* (Pearman et al. 2014). In other studies, the use of these single gene markers as a DNA barcode has been problematic (Hawlitschek et al. 2016). Recently diverged taxa may require additional examination of variation across the mitochondrial genome to effectively characterize interspectific diversity in some tunicate groups (Zhan et al. 2010). One of the groups where mitochondrial sequencing has been used to study intraspecific and interspecific diversity is tunicates.

Recent sequencing of nuclear and mitochondrial genomes in tunicates, a group of chordates (classes Ascidicea, Thalicea and Appendicularia) (Berrill, 1950), have revealed peculiar patterns of genome organization and high rates of molecular evolution relative to other metazoans (Denoeud et al. 2010; Berná L and Alvarez-Valin 2014; Griggio 2014; Jue et al. 2016). The genomes of tunicates have attracted significant interest because of their phylogenetic position as the closest living invertebrate relatives to vertebrates (Berrill 1928; Bourlat et al. 2006; Delsuc et al. 2006, 2008; Dunn et al. 2008). Despite their high rates of molecular evolution, researchers have used the typical animal *CO1* barcoding region in ascidians to produce DNA barcodes and differentiate morphologically similar ascidian tunicates (Nishikawa et al. 2014; Jaffarali et al. 2017). However, DNA barcodes for appendicularian tunicates are currently unavailable (Bucklin et al. 2010b), and the appropriateness of the *CO1* sequence as a marker for DNA barcoding in appendicularians remains unclear. As a result, our current understanding of appendicularian diversity, species diversity, species-specific spatial distributions, and population dynamics in coastal Newfoundland (NL) and Arctic waters is limited due to taxonomic uncertainties (Choe and Deibel 2008). This is surprising, as the *CO1*

barcoding region exhibits levels of intra- and inter-species variability favorable for accurate species identification in pelagic zooplankton (e.g. Bucklin et al. 2010a,b; Bucklin et al. 2007; Bucklin et al. 2003) and other metazoans (Costa et al. 2007).

The *CO1* sequence has been used to elucidate patterns of gene flow in invasive and non-indigenous tunicates (Lopez-Legentil et al. 2006; Perez-Portela and Turon 2008). Over the past ten years, invasive ascidian tunicates (*Styela clava, Ciona intestinalis, Botryllus schlosseri and Botrylloides violaceus*) have reached high levels of abundance at aquaculture sites in Nova Scotia and Prince Edward Island (Carver et al. 2003; Locke et al. 2007). Recently, some of these ascidians have been identified in NL (Callahan et al. 2010). The presence of these ascidians may pose a threat to the NL aquaculture industry as they foul nets and lines. The identification of the connectivity of established populations in Newfoundland harbours, and the source locations will aid in the development of an accurate risk assessment, and informed management and/or mitigation strategies for this species and other invasive ascidians in Newfoundland waters. Moreover, studying these invasive ascidians, where Newfoundland waters may define their northern range limit, can facilitate understanding of the genetic consequences of range expansion (Excoffier et al. 2009).

OBJECTIVES

In this thesis, we explore patterns of genetic diversity and differentiation among cold ocean tunicates in insular Newfoundland harbours. As appendicularian tunicates are often difficult to identify based on their morphology, we aimed develop a molecular identification tool using *CO1* sequences for the pelagic tunicate *Oikopleura vanhoeffeni.* Chapter 2 assesses the utility of using the typical animal barcoding region (*CO1*) as a species-specific marker in the appendicularian *Oikopleura vanhoeffeni*. In this chapter, we provide the details of several primer pairs we developed to obtain partial *CO1* sequences from this species. Finally, we provide the first report of genetic differentiation in this ecologically important tunicate. Conversely, reports of genetic variation in another tunicate, *Botryllus schlosseri* are common from its geographic range outside of Newfoundland waters. As little is known of this non-native species in Newfoundland, in chapter three we use the 5' barcoding region of *CO1* to explore the genetic characteristics of range expansion in the non-indigenous ascidian tunicate *Botryllus schlosseri*. Specifically, samples from 7 populations were collected and sequenced. From the data, we compute tests of selection, diversity indices, and utilize phylogenetic methods to infer source population(s) and pathways of spread for this non-native ascidian in Newfoundland harbours. Collectively, developing more reliable identification tools for zooplankton species and a greater understanding of their geographic ranges will aid in our ability to mitigate anthropogenic stressors in the ocean.

Chapter 2: Variation and Partial Characterization of the *CO1* gene in the appendicularian tunicate *Oikopleura vanhoeffeni* with implications for DNA Barcoding

INTRODUCTION

Pelagic Tunicates

Appendicularians are among the most abundant zooplankton taxa in the world, occupying several ecologically important roles in marine ecosystems (Choe and Deibel 2008). These pelagic tunicates feed on a wide variety of particles ranging in size from dissolved organic matter (DOM) to large diatoms (Flood et al. 1992; Gorsky and Fenaux 1998). Invertebrates and fishes prey on appendicularions (Purcell et al. 2005), providing an efficient energy transfer mechanism from microbes to higher trophic levels, thus short-circuiting intermediate trophic links (Gorsky and Fenaux 1998). Additionally, the sinking of fecal pellets and filtration apparatus carries a substantial portion of primary production both to deeper layers of the ocean (Sato et al. 2003; Alldredge 2005; Dagg and Brown 2005) and to benthic organisms (Alldredge 2005). As a consequence, appendicularians play an important role in the transport of carbon in the ocean (Alldredge 2005).

Within the common appendicularian genus *Oikopleura*, two species occur in Newfoundland coastal waters. Taxonomic differentiation between specimens of *Oikopleura vanhoeffeni* and *O. labradoriensis* adults is possible using patterns of inclusion bodies in the house rudiment and subchordal cells in the tail (Shiga 1993; Galt and Flood 1998). However, morphological identification is difficult for juveniles and for specimens damaged during collection and/or preservation (Choe and Deibel, 2008). Mistakes in identification can obscure species-specific differences in life history characteristics and population dynamics (Choe and

Deibel 2008). In addition, morphological-based taxonomy is not useful as a tool to identify cryptic taxa, which are common in pelagic zooplankton groups (Bucklin et al. 2010b). Second, morphological characters used for species delineation may have intermediate phenotypes in hybrid individuals. Field observations near northern Greenland suggested the presence of hybrid individuals within the genus *Oikopleura* (D. Deibel, pers. comm.). Thus, the development of a diagnostic nucleotide sequence to distinguish among oikopleurid species could facilitate understanding of their ecology, evolution, and systematics.

DNA Barcoding using *CO1* in Oikopleurids

For animal taxa, geneticists commonly designate the region of *CO1* flanked by the "universal invertebrate" primers LCO1490 and HCO2198 (Folmer et al. 1994) as the DNA barcode region (Hebert et al. 2003). Within and among many zooplankton taxa, the 5' end of the *CO1* gene provides sufficient variation for systematic analyses (Bucklin et al. 2010a). Despite extensive citation of Folmer et al. (1994), the primers have not been used successfully to produce a *CO1* sequence from all invertebrate taxa, despite attempts at optimization (Geller at al 2014). The experience of our research group and informal exchanges with colleagues indicate that the Folmer primers have not been successful in attempts to sequence the *CO1* gene in oikopleurids, noting the availability of one putative *CO1* sequence (Genbank accession: AY116611) from *Oikopleura dioica* (Stach and Turbeville 2002). We believe this sequence may not be representative of *Oikopleura dioica*, based on the attempts of members of our research group to repeat the methods of Stach et al. (2007), and no additional published *CO1* sequences for any oikopleurid since 2007 to 2013. Researchers have attempted to improve the performance of the Folmer primers using a variety of strategies, including introducing degenerate bases (Leray et al. 2013) and introduction of inosine nucleotides (Geller et al. 2014). The inosine nucleotides replaced bases in the Folmer primers at the most degenerate positions based on an alignment of

CO1 sequences from major groups of marine invertebrates. The resulting primers containing the inosine bases were renamed jgLCO1490 and jgHCO2198 (Geller et al. 2014).

Difficulties with the amplification of the *CO1* gene from species within the genus *Oikopleura* may reflect mismatches at the targeted annealing sites created by a high rate of molecular evolution among tunicates. Where complete genomes are available, *Oikopleura dioica* has the highest rate of protein evolution of any metazoan (Denoeud et al. 2010). Mitochondrial genes exhibit particularly high rates of protein evolution (Denoeud et al. 2010).

In addition, the presence of unusual gene structure may account for difficulty in amplification of the *CO1* gene in oikopleurids using a single set of primers. Denoeud et al. (2010) used both a gDNA and cDNA approach to assemble the mitochondrial genome of *Oikopleura dioica,* and documented the presence of homopolymeric, non-coding regions. Repeat regions were marked by dT repeats, but not all repeats marked the location of a non-coding region (Figure S4 in Denoeud et al. 2010). A total of are 6 dT repeat regions occur within the *CO1* gene of *Oikopleura dioica*, including 3 non-coding sequences (Figure S4 in Denoeud et al. 2010).

OBJECTIVES

The initial goal of this chapter was to sequence *CO1* in *Oikopleura vanhoeffeni* using the Folmer primers (Folmer et al. 1994) in order to assess the utility of *CO1* as a molecular marker for DNA barcoding in this species. A second objective was to utilize this data to generate the first report of population genetic structure in this fast- evolving pelagic tunicate. However, the difficulty of amplifying *CO1* in *O. vanhoeffeni* forced us to revise the objectives of this study to i) characterize the typical animal barcoding region of *CO1* sequence of *Oikopleura vanhoeffeni,* and ii) provide the first report on the genetic structure of this species in Newfoundland waters.

MATERIALS and METHODSSample Collection and Extraction of Nucleic Acids

SCUBA divers collected *Oikopleura vanhoeffeni* from three harbours on the Avalon Peninsula of insular Newfoundland (Figure 2.1). In the field, live individuals were placed in glass jars containing 100% sea water for transfer to the laboratory. At the time of field work, we had not planned to extract ribonucleic acid (RNA) from these individuals, and therefore, the specimens were stored in ~95% ethanol at 4℃ for DNA preservation.

For each animal, we extracted total genomic DNA from either tails or the whole animal using the DNeasy Blood and Tissue Kit (Qiagen Inc., Mississauga, Ontario). RNA extraction from the trunks of the specimens used the RNeasy Fibrous Tissue Mini Kit (Qiagen Inc.). In order to ensure that putative introns contained within genomic DNA did not contaminate amplification of RNA sequences, we added a genomic DNA digest in the RNA isolation procedure by adding a mixture of 70 µL of buffer RDD (Qiagen Inc.) and 10 µL of DNAase 1 (Qiagen Inc.) to each sample prior to RNA elution from the spin column. Rinsing samples in nuclease-free water removed residual ethanol prior to all extractions.

Determination of the Structure of the *CO1 gene* in *Oikopleura vanhoeffeni*

Initial attempts to amplify a partial region of the *CO1* gene in *Oikopleura vanhoeffeni* utilizing the Folmer primers (Folmer et al. 1994) following the method of Stach et al. (2007) were unsuccessful. We therefore designed species-specific primers from a 1250 bp cDNA copy of *CO1* of *Oikopleura vanhoeffeni* collected in Newfoundland harbours in 2007 (J. Hall, pers. communication). The cDNA sequence contained the 3' end of *CO1* but was missing approximately 300 bp on the 5' end, relative to the full *CO1* transcript of *Ciona intestinalis* (Genbank Accession: AJ517314). In this study, we designed all primers except for OvanjgLCO from J. Hall's cDNA sequence. PCR products generated from genomic DNA and preliminary primer pairs designed based on the cDNA sequence were significantly larger than expected (data

not shown). Homopolymeric regions (data not shown) precluded Sanger sequencing of initial purified PCR products in both directions. Once we identified the locations of all the homopolymeric regions, we designed primers to amplify the homopolymeric regions and the sequences flanking these putative introns.

Amplification of exons and putative introns by polymerase chain reaction (PCR) was carried out for 31 samples in reactions with a final volume of 25 µL containing: 1X TopTaq PCR buffer (Qiagen Inc.), augmented with $MgCl_2$ to a final concentration of 2.5 mM, 200 µM of deoxyribonucleotide triphosphates (dNTPs) (Qiagen Inc.), 0.4 µM of each of forward/reverse primers, 1.25 Units of TopTaq DNA polymerase (Qiagen Inc.), and 80-100 ng of gDNA. For each primer pair, the thermal cycling profile consisted of an initial denaturation step at 94℃ for 3 min, followed by 42 cycles of 94℃ for 50 sec, annealing temperature (Table 2.1) for 50 sec, and 72℃ for 60 sec with a final extension at 72℃ for 5 min. We purified amplified PCR products obtained for putative exons 1-3 using the QIAquick PCR purification kit (Qiagen Inc.).

Given the difficulty of sequencing homopolymeric regions, we estimated the size of the putative introns from a subset of 22 samples on 1.5% polyacrylamide gels stained with RedSafe (FroggaBio Inc, Concord, ON) using a 100 bp DNA ladder (FroggaBio Inc.) following electrophoresis at 120 V for 45 min.

Amplification of the 5' end of *CO1 sequence*

To obtain the 5' end of the *CO1* gene, we acquired sequences from representative ascidian tunicates across a number of families from GenBank using complete mitochondrial genomes available in 2014. Then we aligned sequences using CLUSTALW v.2.0 (Larkin et al. 2007), to obtain a contig of the region corresponding to the forward Folmer and Geller *CO1* barcoding primers (Table 2.2). This approach allowed us to determine the consensus nucleotides

at the positions that correspond to LCO1490 and jgLCO1490 for each of three orders of ascidians: Stoliodobranchia, Phlebobranchia, and Aplousobranchia. Positions with fourfold degeneracy were replaced with dITP. We synthesized positions with twofold degeneracy with mixed nucleotides to create a primer pool. Relative to the jgLCO1490 primer, we modified two sites to create a new primer, named OvanjgLCO1490. Initially, we used primer OvanjgLCO1490 in combination with OvanCO1EX2R5' GAACACTATGATTACTCTGTG 3' to amplify the 5' end of the barcoding region. However, both the size of the PCR product as visualized by agarose gel electrophoresis and sequence results indicated the presence of a homopolymeric region approximately 150 bp from the OvanjbLCO1490 primer (data not shown). Thus, we designed a new primer, OvanCO1EX1R, to amplify exon 1 with OvanjgLCO1490. Samples that amplified weakly at exon 1 using the OvanjgLCO1490/OvanCO1EX1R primer combination were re-amplified using the primer OvanCO1EX1F with OvanCO1EX1R (Table 2.1).

Validation of putative *CO1* homopolymeric non-coding regions

To confirm the absence of the homopolymeric region 1 and 2 in the mature RNA transcript of *CO1*, we completed one-step RT-PCRs using the primer pairs OvanjgLCO1490 (Table 2.1) and OvanCO1EX3R- 5'- GATGCTGAAAAAGAACTGCGT 3.' Use of the primer pair OvanjgLCO1490 (Table 2.1) and OvanCO1EX2R (Table 2.1) demonstrated the absence of intron 1. Each 50 µL reaction contained 3 µL of RNA, 10 µL of 5x one step RT PCR buffer, 2 µL dNTPs, 2 µL of one step RT-PCR enzyme mix (Qiagen Inc), 3.5 µL of 10 µM reverse primer, 3.5 µL of 10 µM of forward primer and 26 µL of RNase free water (Qiagen Inc). In order t to synthesize cDNA we transferred tubes to a thermocycler pre-heated to 50℃ for 30 min. PCR was initiated at 95℃ for 15 min, followed by 42 cycles of PCR. The PCR cycling profile was 94℃ for 50 sec, 48℃ for 50 sec, and 72℃ for 1 min with a final extension at 72℃ for 10 min. To demonstrate the absence of intron 3 and an additional intron identified outside the barcoding

region (appendix 1), we used the primers OvanCO1EX2bF- 5'-TTCTGGTGAGAAGAGCCAGGTCTAG3' and OvanCO1PR5R- 5'-ATAGCCTTGTAAAAATGGGCCAGTAG 3' as described above, but with an annealing temperature of 52.1℃. Amplified PCR product purification used the QIAquick PCR purification kit (Qiagen Inc.). Genome Quebec sequenced purified products (exons and RT PCR products), using BigDye Terminator chemistry on the ABI Prism 3730 DNA Analyzer (Applied Biosystems Inc., Foster City, USA) using Sequencing Analysis v5.0 software.

Genetic Variation and Genealogic Relationships among Haplotypes

CO1 segments were trimmed and edited using Sequencher v. 4.9 (Gene Codes Corporation, Ann Arbor, MI). For each sample, we joined sequences representing the three *CO1* exons into one sequence spanning the entire length of the barcoding region. Sequencher v 4.9 (Gene Codes Corporation) enabled alignment of the complete *CO1* sequences for all samples into a contig.

We used DNASP v.5 (Rozas et al. 2003) to identify individual haplotypes and calculate haplotype and nucleotide diversity, and then selected the best fit nucleotide substitution model under the Akaike Information Criterion (AIC; Posada and Buckley 2004) using MEGA v. 5.2 (Tamura et al. 2011). MEGA v.4.0 (Tamura et al. 2011) then calculated sequence divergences between haplotypes by the Kimura 2 parameter model. We examined genetic differentiation between populations by computing Φ_{ST} (10,000 permutations) using the Kimura 2parameter substitution model in ARELEQUIN v. 3.1 (Excoffier et al. 2005). By generating a statistical parsimony network at the 95% connection limit and applying the Median-Joining algorithm with TCS v.1.21 (Clement et al. 2000), we examined relationships among the *CO*1 haplotypes.

Neutrality and Historical Demographic Analyses

To assess departure from neutrality (mutation-drift equilibrium) we calculated Fu's Fs (Fu 1997) and Tajima's D statistic (Tajima 1989), where significant negative values may indicate population expansion and/or purifying selection. Neutrality was examined by comparing the rates of non-synonymous substitutions (d_N) and synonymous substitutions (d_S) using Z-tests of selection with the Nei-Gojobori method and Jukes-Cantor distance correction (Nei & Gojobori 1986) for the alternate hypotheses of strict neutrality (Ha: $d_N \neq d_S$) using MEGA. A greater number of synonymous mutations relative to non-synonymous substitutions ($d_N - d_S < 0$) provides evidence of purifying selection.

To assess whether populations exhibited evidence of expansion, we performed a mismatch distribution analysis for each sampling location, and for the data set overall, using both DNASP (Librado & Rozas 2009) and ARELEQUIN. A mismatch distribution is a histogram of the pairwise differences among haplotypes (Slatkin and Hudson 1991; Rogers and Harpending 1992). The distribution is expected to be smooth and unimodal in populations that experienced a single burst of population growth in the past, and ragged and polymodal distribution indicate populations that have been stationary for a long time (Harpending 1994). By plotting the mismatch frequency distributions in DNASP we determined whether the populations exhibited evidence of spatial range expansion or a stationary population history (Tajima 1989). We compared the observed mismatch distribution to a null model of demographic equilibrium using 10,000 bootstrap replicates. The SSD (sum of squared deviations) statistic tested the goodness of fit between the observed and expected mismatch distributions. We then used the raggedness index (r; Harpending 1994) to quantify the smoothness of the observed mismatch distribution. Both the raggedness and SSD statistics were tested against the null hypothesis of population

expansion and evaluated using a parametric bootstrap approach with 10,000 replicates. We computed all these statistics and tests of significance in Arelequin (Excoffier et al. 2005). The R^2 statistic contrasts the number of singletons and the mean number of differences; it approaches low positive values in the case of a recent population growth event (Ramos-Onsins and Rozas 2002). Ramos-Onsins and Rozas (2002) demonstrated that this statistic has the greatest power to detect population expansion for non-recombining regions of the genome under a variety of different circumstances, especially with large population sample sizes (~50, Fu's Fs) or small sample sizes (~10, R^2). We used DnaSP to calculate R^2 and the assessed significance of these tests using 10,000 replicates of coalescent simulations.

Since population expansion is indicated by an excess of haplotypes, as well as an excess of singleton mutations, we calculated Strobeck's *S* statistic (Strobeck 1987) in addition to Fu's Fs using DNASP. This statistic expresses the probability of observing a sample of sequences with a number of haplotypes equal to or lower than *h*. We evaluated the significance of Strobeck's S statistic using 10,000 bootstrap replicates in DNASP.

RESULTS

Structure of *CO1*

In this study, we identified three putative introns and three exons in the typical animal barcoding region of *CO1* in *Oikopleura vanhoeffeni* (Figure 2.1). We obtained sequences only for the three exons because the homopolymeric nature of introns hampers sequencing efforts. The homopolymeric regions are inserted in each case beginning with a string of 6 Ts, and one of the homopolymeric regions is inserted within the reverse Folmer priming sequence (Figure 2.2, Table 2.3).

Size of *CO1* Homopolymeric Regions

We estimated the size of the first three introns from PCR products visualized using gel electrophoresis for 22 individuals representing 14 haplotypes (Table 2.4). In all samples, the first intron from the 5' end of the *CO1* sequence represented the largest intron. Six different sizes of insertion regions occurred at the site of intron 1, ranging from 650-1400 bp. Six different sized introns occurred for intron 2, ranging in size from 250-700 bp. This intron was absent in haplotype 13. At the third intron site, we identified four different sized insertion regions within a relatively small size range (400-650 bp). Given a sample size of one for most haplotypes, we could not determine whether intron size was conserved within haplotypes. However, all nine samples of haplotype 1 displayed identically sized introns at each of the three insertion sites (Table 2.4). For haplotype 12, we determined intron sizes for two individuals and found them to be identical for intron 2 but different for introns 1 and 3. Each haplotype contained a unique combination of insert sizes at the three introns. Finally, where we observed introns of the same size at the same insertion site in multiple individuals, there was no restriction of a particular intron to any of the three populations sampled.

Diversity and Population Structure

We sequenced three exons from the 5' end of the *CO1* gene for each of 31 individuals from three Newfoundland harbours (see Figure 2.1) and combined them after trimming to give a total sequence length of 498 bp. The final alignment contained 29 polymorphic sites (17 parsimony informative), with 26 synonymous substitutions and 3 non-synonymous substitutions, translated using the invertebrate mitochondrial deoxyribonucleic acid (mtDNA) code. The 3 non-synonymous substitutions occurred in haplotypes Ovan10, Ovan 17 and Ovan4. In haplotype Ovan10, the substitution corresponded to a leucine to phenylalanine change. In Ovan17 and Ovan4, the non-synonymous substitution in each case changed from glycine to serine.

Interestingly, translation of haplotypes Ovan17 and Ovan4 using the ascidian mtDNA code eliminates the non-synonymous amino acid change.

Among the 31 sequences spanning the typical animal barcoding region, we identified seventeen haplotypes (Figure 2.3, Table 2.4). Pairwise sequence divergence between the haplotypes ranged from 0.2% to 3.3% (Figure 2.3). Ovan1 was the dominant haplotype recorded in each harbour (50% in Logy Bay, 45% in Admiral's Cove, and 33% in Harbour Main), and the only haplotype shared among all three harbours (Fig. 2.1). Harbour Main and Admiral's Cove, shared one other haplotype, Ovan12. The three NL populations shared 15 private haplotypes. Haplotype diversity was high for all three locations, ranging from 0.79 to 0.91 (Table 2.5). The lowest nucleotide diversity for the Admiral's Cove population (0.011) contrasted the and highest value in Logy Bay (0.014) (Table 2.5). Pairwise Φ_{ST} values indicated no significant spatial structure among locations ($P > 0.5$ in all comparisons, data not shown).

Phylogenetic Analyses

The maximum likelihood phylogenetic analysis revealed that all *Oikopleura vanhoeffeni* haplotypes correspond to two well-supported clades plus a third putative group that was not well supported (Figure 2.4). A 95% parsimony haplotype network identified the two well-supported clades in the neighbour joining tree (Figure 2.5). One clade included the dominant Ovan1 haplotype, shared among all three populations, and haplotypes Ovan2-Ovan5, and Ovan12-Ovan17. The second clade included haplotypes Ovan6-Ovan11. Each group contained one haplotype with a non-synonymous mutation. The minimum number of substitutions between the most closely related haplotypes was one. Group 1 included a maximum of 12 substitutions between the most distantly related haplotypes, whereas a maximum of 7 substitutions separated the most distance haplotypes in group 2.

Demographic Analyses

Tajima's D values were negative but not statistically significant for the entire data set (D = -0.68, $p > 0.5$) and for each population. The results of Fu's F_S test, which is based on the distribution of haplotypes, were not significant ($p > 0.05$). The ratio of non-synonymous to synonymous substitutions over all haplotypes was consistent with purifying selection (Z-test: $d_N - d_S = -4.06$, $p < 0.05$).

The mismatch distribution was multimodal for all three populations sampled and overall (Figure 2.6), indicating populations in demographic equilibrium. The distribution did not differ significantly from one generated under a null model of population expansion, because the raggedness and the sum of squared deviations were both low and non-significant ($p > 0.5$). Ramos-Onsins and Rozas statistic (R^2) was low and non-significant for the entire dataset and each sub-population ($p > 0.5$). Strobeck's S statistic, though not significant for individual populations, was significant for the entire dataset (S=0.98, $p < 0.05$).

DISCUSSION

Genetic Diversity, Selection, and Population Genetics

In this study, we aimed to evaluate the utility of the *CO1* gene for species level identification in the cold ocean tunicate *Oikopleura vanhoeffeni*. The *CO1* gene in appendicularians has been difficult to sequence (Sherlock et al. 2017). We show that the *CO1* region in *Oikopleura vanhoeffeni* is difficult to sequence with a universal primer pair from gDNA due to the presence of non-coding regions throughout the gene. Such findings are unexpected in the mitochondrial gene of an animal. Additionally, our study presents the first report of genetic diversity in *Oikopleura vanhoeffeni* using a portion of the mitochondrial *CO1*

gene. Although we found no genetic structuring among the three populations, we found considerable haplotypic diversity among the 31 individuals. Of 17 haplotypes, 15 were private. The large number of haplotypes, with some pairwise distances greater than 3%, may indicate that *O. vanhoeffeni* in Newfoundland waters may contain one or more cryptic species. Few reports of genetic structuring are available for appendicularians, so not much is known about the presence of cryptic species in this group. This may not be surprising, as the majority of larvacean species have not been studied (Sherlock et al. 2017). Only about 70 species are currently recognized (Deibel and Lowen 2012). One study using a 307bp fragment from the 5' barcoding region of CO1 suggested at least three cryptic species among *Oikopleura Longicauda* (Garic et al. 2019). However, these findings need to be explored further using other mitochondrial genes (Garic et al. 2019).

The multimodal mismatch distributions for each population suggests populations in dynamic equilibrium. The mismatch distributions and neutrality statistics offer no evidence to reject an expansion in this species, given the non-significant values for the raggedness index, the sum of squared deviations, the R-squared values and Strobeck's S statistic for each population. However, the small sample size for each population suggests we lacked sufficient power to reject the null hypothesis. For the entire dataset, Strobeck's S statistic was significant, potentially indicating a deviation from neutrality resulting from selection or a population expansion. We are hesitant to draw conclusions based on the demographic statistics presented here as zooplankton populations are very large in size, and our sample sizes are tiny in comparison. However, changes in population size and/or evidence of selection could be a result of anthropogenic stressors such as climate change.

Our study three non-synonymous mutations among three haplotypes. Two of the non-synonymous mutations corresponded to a glycine to serine change when translated using the invertebrate mitochondrial code. However, when translated using the ascidian mitochondrial code, this change was synonymous, because the product of codons AGA and AGG changed from serine to glycine within the ascidian lineage. Other researchers used the ascidian mitochondrial code to translate *Oikopleura dioica* sequences (Denoeud et al. 2010). In this study, translation of all three *CO1* fragments using the ascidian mitochondrial code did not significantly affect demographic and neutrality results (data not shown). Oikopleurids may also share the ascidian mitochondrial code. If this is the case, this divergent feature can be added to the peculiar features of the *Oikopleura dioica* mitochondrial genome that includes novel gene order, interrupted coding sequences, and the non-standard location of rRNA and tRNA coding genes (Denoeud et al. 2010).

Structure of *CO1*

Our study characterized the typical animal barcoding region of the *CO1* gene of *Oikopleura vanhoeffeni* as three exons and three putative introns. Although mitochondrial introns rarely occur in animals, little evidence exists to support the hypothesis that the three homopolymeric regions of the *CO1* gene in *O. vanhoeffeni* sequenced in this project represent a pseudogene. Previous studies report mitochondrial pseudogenes (numts), which arise from the integration of mtDNA into the nuclear genome, in many eukaryotic taxa (Bensasson et al. 2001). Since pseudogenes may experience more relaxed selection pressures relative to coding sequences, numts may be identified by the presence of indels, non-synonymous changes, or frame stop codons. We identified only three non-synonymous changes in the 498 bp partial sequence of *CO1*, no indels, and no stop codons. Although the location of in-frame stop codons and/or indels may potentially be located downstream of the typical animal barcoding region of

CO1, it is unlikely the case here, because previous work identified the presence of homopolymeric regions in the *CO1* gene of the closely related species *Oikopleura dioica* (Denoeud et al. 2010). Within both *O. dioica* and *O. vanhoeffeni*, a sequence of 6 T's marks the presence of the non-transcribed homopolymeric regions that interrupt the open reading frame (this study, Denoeud et al. 2010). To further support our claim that these homopolymeric regions represent some type of intron, we performed RT-PCR to generate evidence for the absence of the poly-T regions in the mRNA transcript of *CO1*. Previous work indicates that in *O. dioica*, RNA editing, rather than intron-like splicing, removes these homopolmyeric regions, because potential intermediates of a deletion process have been detected in EST's in this species (Denoeud et al. 2010).

The finding of non-coding, mononucleotide tandem repeats in the *CO1* gene of *Oikopleura vanhoeffeni* (this study) and *Oikopleura dioica* (Denoeud et al. 2010) represent an unusual feature of protein coding mitochondrial genes in eukaryotes. Many other features that deviate from those typically observed in animal genomes have been documented in the both the nuclear and mitochondrial genomes of *O. dioica*. Examples of such features include a rapidly evolving, highly compact nuclear genome relative to vertebrates and cephalochordates, peculiar patterns of amino acid substitution, absence of the minor spliceosome, lack of DNA repair pathway genes, less conserved gene order, high intron turnover, and a divergent intron-exon organization (Berna and Alvarez-Valin 2014; Berna et al. 2012; Chavali et al. 2011; Denoeud et al. 2010).

Though the identification of tandem repeats within the *CO1* gene is unusual among chordates, 10 to 20% of eukaryotic genes contain this structure (Gemayel et al. 2010). Despite the small genome size of *Oikopleura dioica*, 16.6% of annotated genes contain tandem repeats

(Chavali et al. 2011). Tandem repeats accelerate evolution by influencing characteristics associated with genetic and epigenetic changes (Gemayel et al. 2010). On a broader level, the repeats link to phenotypic variability (Gatchel and Zoghbi 2005). We identified all three homopolymeric regions in the *CO1* sequence of *Oikopleura vanhoeffeni* in 21 of 22 individuals observed from three populations (Table 2.4). As these tandem repeats occur widely among *O. vanhoeffeni* populations in Newfoundland waters, the repeats may be selected for. The role of tandem repeats in the unusual genomic architecture observed in oikopleurids requires further research.

The sizes of the three tandem repeat regions observed in *O. vanhoeffeni* were observed in unique combinations for each haplotype, with identical estimated sizes of the three homopolymeric regions among all 9 individuals tested for haplotype Ovan1. Although the exact size of these homopolymeric regions is unknown at this time, we do not recommend further development of this feature of the *CO1* gene as a tool for phylogenetic studies prior to application of a whole mitochondrial genome approach to delimit species of appendicularians. Following a whole mitochondrial genome approach to species identification, fully exploring the potential of tandem repeats as a genetic marker for phylogenetic studies requires further study, since given current lack of knowledge of whether appendicularians other than *O. dioica* and *O. vanhoeffeni* exhibit tandem repeats within the *CO1* gene, and whether the size and position of the repeat regions vary within and between species.

Finally, the documentation of other non-coding regions within the *CO1* gene in other animals, though rare, merits comment. Introns have been identified in cnidarians (e.g. Beagley et al. 1996; van Oppen et al. 2002; Fukami et al. 2007), sponges (e.g. Rot et al. 2006; Wang and Lavrov 2008; Szitenberg et al. 2010), placozoans (Burger et al. 2009; Dellaporta et al. 2006) and

an annelid worm (Vallès et al. 2008). Placozoans are even capable of a more sophisticated type of transcript processing, where *CO1* coding regions located on distinct RNA molecules are spliced together (Burger et al. 2009).

Utility of the *CO1* gene for DNA Barcoding in Appendicularians

The utility of DNA barcoding relies on fast, inexpensive, accurate, and efficient amplification and/or sequencing of the marker of choice in the target species. To facilitate amplification, a DNA barcode may even have conserved primer sites in the regions flanking the target gene such that the same primer pairs may be used across a wide spectrum of organisms (so-called "universal" primers). The presence of introns requires multiple primer pairs to amplify the entire *CO1* barcoding fragment of *Oikopleura vanhoeffeni*. Thus, the *CO1* gene is not an ideal candidate gene for use in DNA barcoding in the genus *Oikopleura*.

As sequence divergence among species normally greatly exceeds the distance between congeneric species (Hebert et al. 2003), researchers have used DNA barcoding for both the assignment of specimens to known species and the discovery of new taxa (Witt et al, 2006). There was no *CO1* barcoding fragment available to evaluate interspecific genetic distances between *Oikopleura vanhoeffeni* and other species in the genus *Oikopleura*. Consequently, we were unable to evaluate the presence or absence of a clear gap between the genetic distances calculated at intraspecies and at congeneric levels. However, the large number of distinct haplotypes, and a few intraspecific pairwise genetic distances of over 3% may indicate that *O. vanhoeffeni* in Newfoundland represents a cryptic species complex. If this is the case, the *CO1* gene may be inadequate for DNA barcoding in this genus. Several authors suggested that the invasive tunicate *Botryllus schlosseri* experiences on-going speciation events and consists of at least three cryptic species complexes (Bock et al. 2011, Griggio et al. 2014). In this ascidian,

previous work found the *CO1* barcode fragment inadequate to recognize the subtle, yet biologically significant intraspecies differences identified by whole-mitogenome comparisons (Bruetti et al 2020). Cryptic species have been identified in other ascidians such as *Ciona intestinalis* (Suzuki et al. 2005; Caputi et al. 2007; Zhan et al. 2010), *Clavelina lepadiformis* (Tarjuelo et al. 2001), *Pseudodistoma crucigaster* (Tarjuelo et al. 2004) and *Pycnoclavella communis* (Perez-Portela and Turon 2008). Previous studies have used a whole mitochondrial genome approach for *Ciona intestinalis* to identify the presence of cryptic species (Suzuki et al.2005, Caputi et al. 2007). Future research to address taxonomic questions across short phylogenetic distances for oikopleuids should be based on genetic information from entire mitochondrial genomes. This approach has been shown to be as effective as using a fragment of *CO1* for DNA barcoding in 28 tunicate species (Griggio et al 2014).

CONCLUSIONS

In this study, we identified three mononucleotide, non-coding regions in the *CO1* gene of *Oikopleura vanhoeffeni* and provided the first information on intraspecific genetic diversity in this appendicularian. Given the interrupted *CO1* coding regions, we suggest this gene is unsuitable as a marker for the quick identification of specimens at the species level (DNA barcoding), because it would require several primer pairs to amplify. Further research should utilize a whole mitochondrial genome approach to aid in the diagnosis of cryptic species and identify a suitable gene for DNA barcoding. However, once species limits are known within the genus *Oikopleura*, the tandem repeat regions of the *CO1* gene may be developed as a useful tool for population geneticists.

The presence of tandem repeats in the *CO1* gene, though unusual for animals, has been previously identified in *Oikopleura dioica.* Oikopleurids appear to be unique among the tunicates

with a small genome, and a mutation rate three times as fast as humans and twice as fast as the fast-evolving *Ciona intestinalis* (Berná and Alvarez-Valin 2014). Thus, the high number of haplotypes identified for this species in Newfoundland waters was not surprising and may indicate a cryptic species complex.

Table 2.1: Primer sequences and PCR conditions developed in this study for gene regions of cytochrome oxidase 1 in *Oikopleura vanhoeffeni*. Abbreviation: T_A, annealing temperature.

Primer Pair	Gene Segment	Nucleic Acid Template	Oligonucleotides (5'-3')	T_A (°C)
OvanjgLCO1498	Exon 1	gDNA	T I TC IAC IAAYCAYAARGAYATTRG	51.7
OvanCO1EX1R			GAACACTATGATTACTCTGTG	
OvanCO1EX1F	Exon1	gDNA	TTGTCGACGAATCAYAAGGA	51.7
OvanCO1EX1R			GAACACTATGATTACTCTGTG	
OvanCO1EX2F	Exon 2	gDNA	TTTGATTGGAGGTTTCGGTAATTG	53
OvanCO1EX2R			TACTACTTCTCAGAATACTAGAAAC	
OvanCO1EX3F	Exon3	gDNA	ACTTGGGTGCATTTAGCACC	53
OvanCO1EX3R			GATGCTGAAAAAGAACTGCGT	
OvanCO1IN1F	Intron 1	gDNA	GACACAGAGTAATCATAGTGTTC	54.2
OvanCO1IN1R			ATCAATTACCGAAACCTCCAATC	
OvanCO1IN2F	Intron 2	gDNA	ACTATTTACCCTCCGCTCTCCACTA	53.6
OvanCO1IN2R			TACGGCTGAGAAGACTTGGGTGC	
OvanCO1IN3F	Intron 3	gDNA	TTCGTACTTCGATATGGGTGGAGGA	53.9
OvanCO1IN3R			CATGACTTAAAAGCCCAAAACCCGGT	
OvanjgLCO1498	Exons 1 and 2	RNA	T I TC IAC IAAYCAYAARGAYATTRG	48
OvanCO1EX2R			TACTACTTCTCAGAATACTAGAAAC	
OvanCO1EX2bF	Exons 2, 3 and outside DNA barcoding region to 3' end of CO1	RNA	TTCTGGTGAGAAGAGCCAGGTCTAG	52.1
OvanCO1PR5R			ATAGCCTTGTAAAAATGGGCCAGTAG	
OvanjgLCO1498	Exons 1, 2 and 3 (entire "animal barcoding" region)	RNA	T I TCIACIAAYCAYAARGAYATTRG	48
OvanCO1EX3R			GATGCTGAAAAAGAACTGCGT	

Table 2.2: Consensus sequences from alignments in 2014 of the *cytochrome oxidase subunit I* from three suborders of ascidians, forward primer LCO1490 (Folmer et al. 1994) and forward primer jgLCO1490 (Gellar et al. 2014). We used the alignment of the consensus sequences to create a new forward primer designed to amplify the 5' end of COI in *Oikopleura vanhoeffeni* with species-specific reverse primers (see Table 2.1). Mitochondrial genomes and GenBank accession numbers of sequences used for alignments*:* Stolidobranchia*: Styela plicata* (AM292601), *Botryllus schlosseri* (FM177702*), Botrylloides nigrum* (HF548559), *Botrylloides violaceus* (HF548552), *Styela clava* (HG931920), *Polycarpa mytiligera* (HF548556), *Botrylloides leachii* (HF548553), *Halocynthia roretzi* (AB024528), *Halocynthia spinosa* (HF548558), *Herdmania momus* (FM296153), *Microcosmus sulcatus* (AM292321), *Pyura gangelion* (HF548557). Phlebobranchia: *Ciona intestinalisA* (AJ517314), *Ciona intestinalisB* (AM292218),*Ciona savignyi* (AB07978), *Rhodosoma turcicum* (HF548560), Ascidiella aspersa (HF548561), *Phallusia fumigata* (AM292602), *Phallusia mammillata* (AM292329). Aplousobranchia: Aplidium conicum (FN313538), *Clavelina phlegraea* (AM292604), *Clavelina lepadiformis* (AM292603), *Diplosoma listerianum* (FN313539).

LCO1490	GGT CAA CAA ATC ATA AAG ATA TTG G
Stolidobranchia	GRT C I A C I A AYC AYA ARG AYA TTR G
Phlebobranchia	GKT C I A CWA AYC ATA ARG ATA TTR G
Aplousobranchia	GRT CWA CWA ATC AYA AAG ATA TTR G
jgLCO1490	T I T C I A C I A AYC AYA ARG AYA TTG G
OvanjgLCO1490	G I T C I A C I A AYC AYA A RG AYA TTR G

Table 2.3: Sequences flanking the insertion sites of homopolymeric cytochrome oxidase introns in *Oikopleura vanhoeffeni*. Insertion sites are located at homopolymeric sequences of six T's and indicated by ^. Introns 1 and 3 include two sextet T regions. In this study, we amplified intron 1 and 2 using one set of primers to amplify the entire region containing both homopolymeric T regions. The 3' end of the *CO1* sequence flanking intron 3 contains the Folmer reverse primer, which is shown in blue.

	DNA Sequence
Intron 1	AGAGTAATCATAGTGTTCTTTTTTATCATACC**TTTTTT**^GATTGGAGGTTTCGGTAA
Intron 2	CTAGTATCCTGAGAAGTAGTAA**TTTTTT**^GGTTACAATAGTGCTAAATG
Intron 3	GGGACGCAGTTCTTTTTCAGCATCTTTTTTGG**TTTTTT^**GGTCATCCTGAAGTCTACG

Table 2.4: Approximate Size (# of base pairs, bp) of the first three introns from the 5' end of cytochrome oxidase 1 in *Oikopleura vanhoeffeni*. We estimated intron size from PCR products generated from *Oikopleura vanhoeffeni* DNA using agarose gel electrophoresis (1.5%) with 100 bp ladder (Frogga Bio). H is the haplotype name and n is the number of specimens.

H	*n*	Approximate Intron Size (bp)		
		Intron 1	Intron 2	Intron 3
Ovan1	9	1200	500	500
Ovan3	1	1000	500	500
Ovan4	1	650	400	400
Ovan5	1	1200	500	650
Ovan6	1	650	250	500
Ovan7	1	750	350	500
Ovan8	1	650	300	500
Ovan10	1	650	500	500
Ovan12	1	1400	700	400
Ovan12	1	1200	700	500
Ovan13	1	1400	absent	500
Ovan14	1	950	500	600
Ovan15	1	1400	500	500
Ovan16	1	1400	500	400

Table 2.5: Locations of *Oikopleura vanhoeffeni* populations and genetic diversity indices for cytochrome oxidase 1 with n, sample size; NH, number of haplotypes; h, haplotype diversity; π nucleotide diversity.

.

Location and Geographic Coordinates (Latitude, Longitude)	N	NH	Haplotypes Identified	h	π
Admiral's Cove (47.0958, -52.9109)	11	6	Ovan1, Ovan2, Ovan4, Ovan6, Ovan 9, Ovan12	0.8	0.011
Logy Bay (47.6268, -52.6640)	8	5	Ovan1, Ovan8, Ovan 10, Ovan13, Ovan15	0.786	0.014
Harbour Main (47.4374, -53.1545)	12	9	Ovan1, Ovan3, Ovan 5, Ovan7, Ovan 11, Ovan12, Ovan14, Ovan16, Ovan17	0.909	0.014
Total	31	17	Ovan1 – Ovan17	0.826	0.012

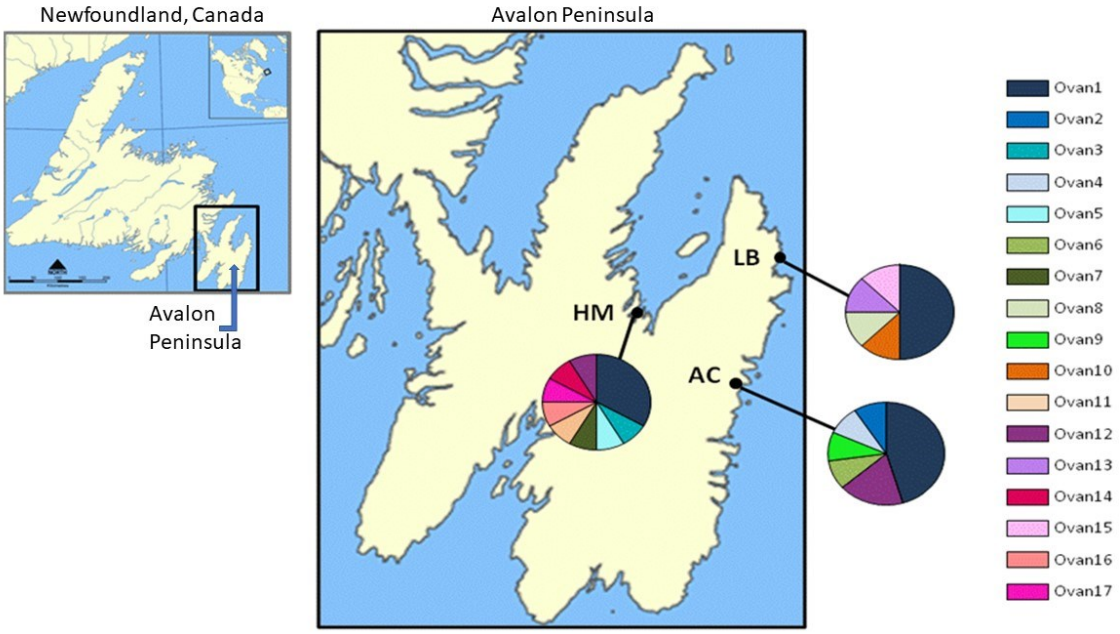

Figure 2.1: Sampling locations and relative frequencies of *Oikopleura vanhoeffeni* mitochondrial cytochrome oxidase 1 haplotypes from three harbours in insular Newfoundland, Canada. HM = Harbour Main, LB = Logy Bay, AC = Admiral's Cove. Map courtesy of T. Wells, Department of Fisheries and Oceans (DFO), Canada.

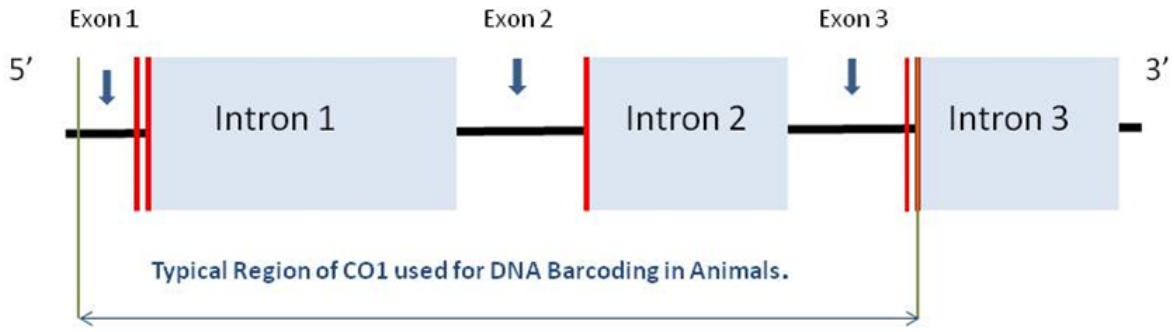

Figure 2.2: Schematic of the typical animal barcoding region of *cytochrome oxidase I* (*CO1*) in *Oikopleura vanhoeffeni*. Red vertical lines indicate a sequence of 6 Ts (TTTTTT), which mark the insertion point for a homopolymeric region found to be absent in *CO1* cDNA transcripts. The vertical green lines represent the positions of the Folmer and Gellar primers for amplification of a fragment of *CO1* in invertebrates and marine invertebrates respectively. (Sequence lengths not to scale.)

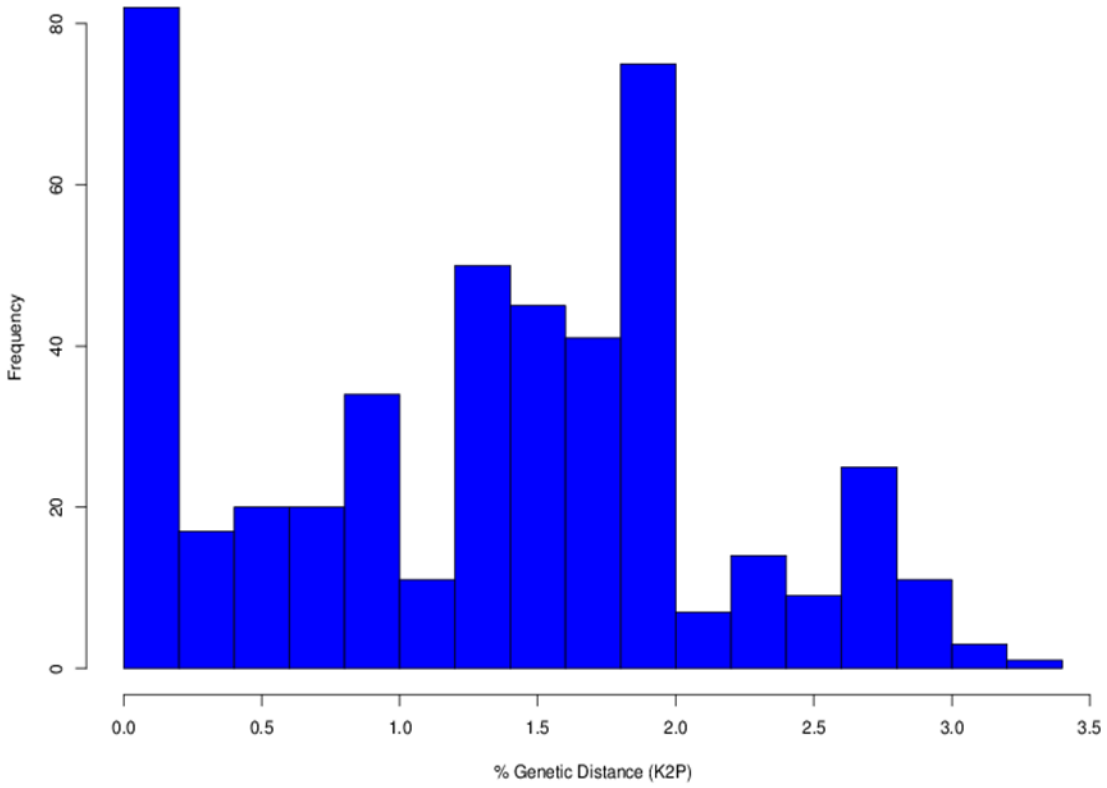

Figure 2.3: Histogram showing the relative distributions of pairwise intraspecific divergence for 29 *Oikopleura vanhoeffeni cytochrome oxidase* 1 sequences, estimated using the Kimura 2 parameter (K2P) model.

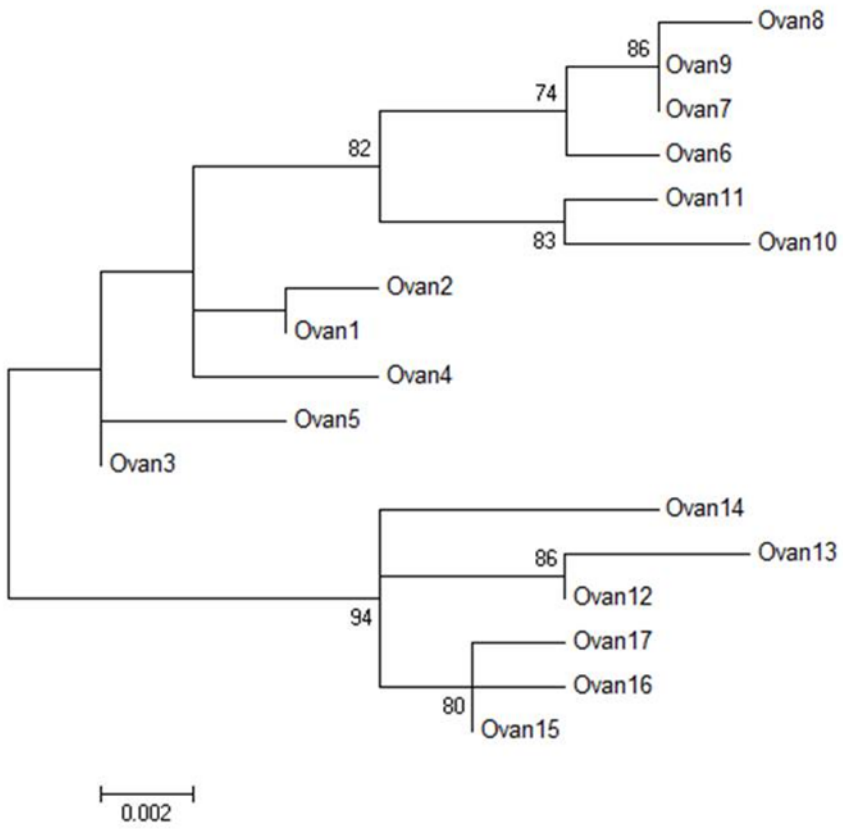

Figure 2.4: Maximum likelihood phylogeny of *Oikopleura vanhoeffeni* in Newfoundland harbours based on a 498 bp region of *cytochrome oxidase 1*. Numbers at nodes indicate bootstrap support (1000 replicates) where values were >70.

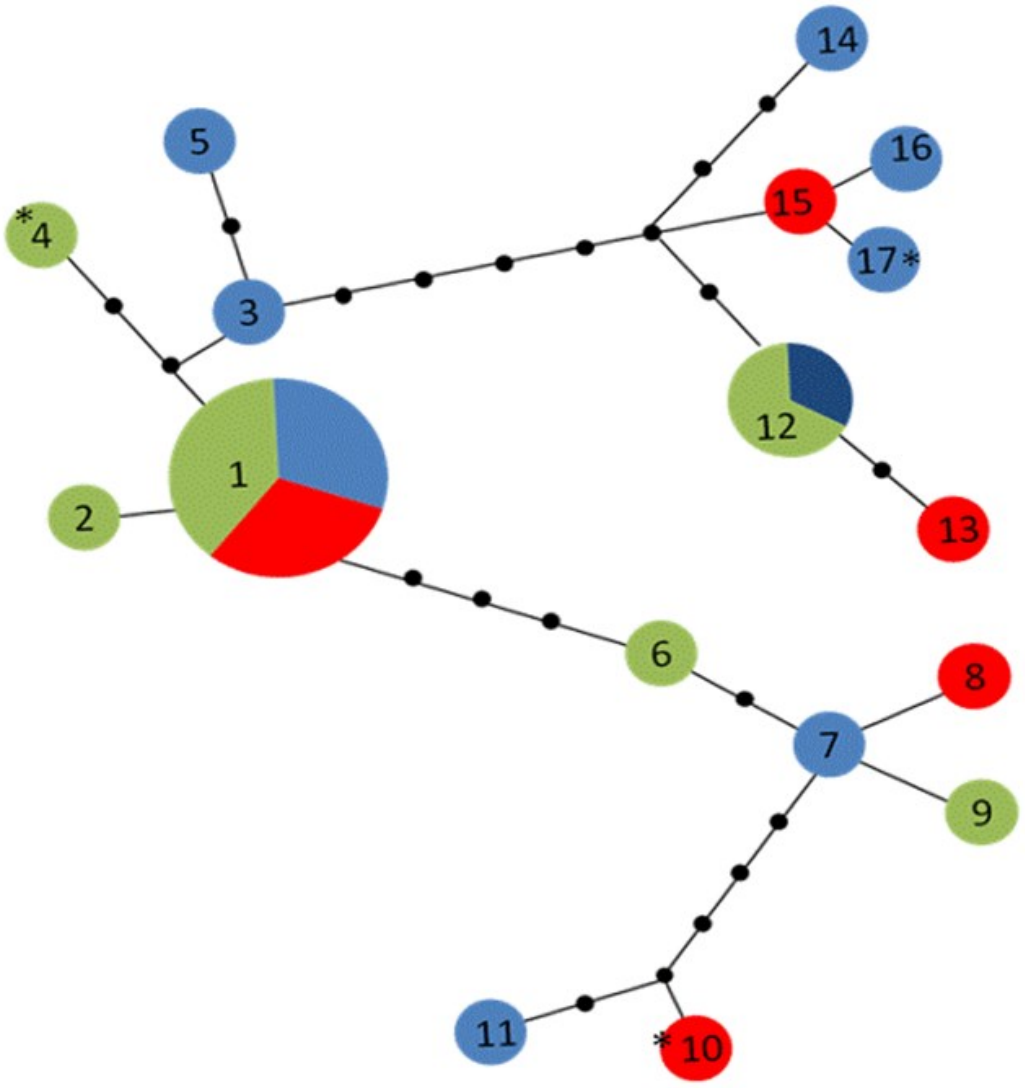

Figure 2.5: Statistical parsimony network of *Oikopleura vanhoeffeni* cytochrome oxidase subunit I (*CO1*) haplotypes Ovan1-Ovan17 (numbered 1-17). Circle size is proportional to haplotype frequency, and small black dots indicate unsampled haplotypes inferred from the data. The non-synonymous substitutions are indicated by an asterisk. Groups 1 and 2 correspond with clades identified by the maximum likelihood phylogenetic analyses. The colours denote sampling locations: green indicates Admirals' Cove, red indicates Logy Bay, and blue indicates Harbour Main.

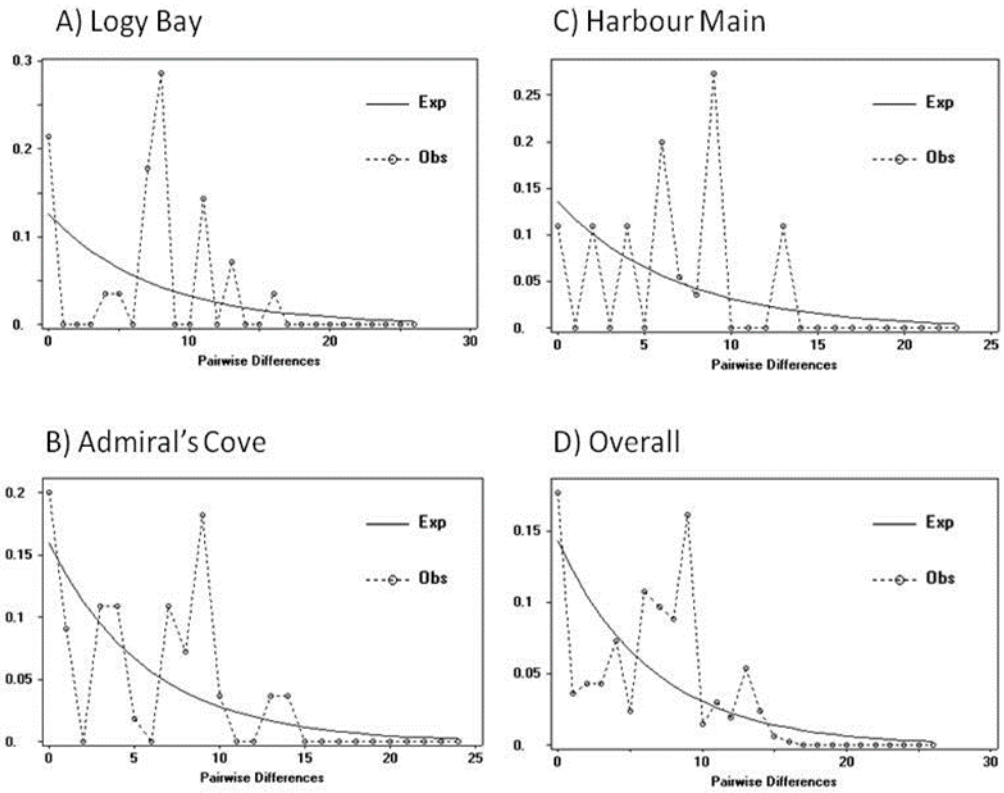

Figure 2.6: Mismatch Distribution obtained of 17 *cytochrome oxidase 1* haplotypes for *Oikopleura vanhoeffeni* in three harbours in Newfoundland. The observed distribution is represented by a dotted line, and the expected distribution based on a model of exponential population growth is represented by a solid line. A) Logy Bay, B) Admiral's Cove, C) Harbour Main and D) Overall.

Chapter 3: Regional phylogeography of the non-native ascidian *Botryllus schlosseri* in Newfoundland harbours

INTRODUCTION

Better management of changing marine ecosystems requires understanding the factors and evolutionary processes that influence the distribution patterns of species in the wild. When species expand their habitat range into geographic areas not previously colonized by the species, either naturally or through accidental or deliberate human intervention, some will fail to become established, some may spread but remain spatially restricted, and some will spread more widely (Roman and Darling 2007). Collectively, differences in the patterns of spatial spread across a variety of climates can be viewed as a series of natural experiments, which can provide insights into the important evolutionary factors that drive adaptation and range expansion (Estoup et al. 2004). Some of the evolutionary processes influencing the patterns of distribution of marine species include secondary contact among lineages (e.g. Jeffery et al. 2017b), adaptation along an environmental gradient (e.g. Sandford and Kelly 2011), or some combination of physical barriers, genetic drift and selection (Harrison and Larson 2016).

The evolutionary processes influencing fine scale population structure are increasingly revealed through genetic data. The relationship between intra-population genetic diversity and the success of new populations is not well established (Roman and Darling 2007). Small populations are thought to face higher risk of extinction as a result of loss of genetic variation (e.g. Frankham and Ralls 1998). Empirical evidence shows that some, but not all non-indigenous species loss of genetic diversity in their non-native range (Dlugosch and Parker 2008, Roman and Darling 2007), which could be due to founder effects and/or bottlenecks. Given the expectation that non-neutral, intra-specific genetic variation has evolved in species on micro-

geographical scales as a result of differences in the selection pressures associated with different environments (Allendorf and Lundquist 2003), suggests lower fitness in species with lower levels of genetic variation in their non-native ranges.

Evidence suggests that some species with reduced levels of genetic diversity in their non-native distributions can become established, and, in some cases, spread rapidly. Multiple, frequent introductions from genetically diverse source populations, or introductions to a single founding population from multiple source populations may facilitate the establishment of species in new environments (Roman and Darling 2007). Independent populations may then hybridize, and spread out, across new environments via human-mediated and natural migration processes (Jeffrey et al, 2017a; 2017b; Lehnert et al, 2018).

Expanding ranges in many non-invasive species may be considered natural experimental models of genetic diversity in the context of range expansion and population success. In these models molecular markers enable study of the number, sources, and patterns of connectivity of established populations across areas of introduction, as well as the micro-evolutionary forces that act on the genetic structure of populations (e.g. genetic drift, types of selection) (e.g. Darling & Folino-Rorem 2009; Rollins *et al* 2009; Jeffery et al. 2017).

Population Genetics of *Botryllus schlosseri*

Botryllus schlosseri, one of the most successful marine invasive species (Bock et al. 2011), has achieved a worldwide distribution outside of the polar continents (Carver et al. 2006, Lambert 2007). Known as the golden star tunicate, this colonial tunicate native to the Mediterranean (Berrill 1950, Carver et al. 2006) was probably introduced to North America in the early 1800s (Carver et al. 2006). In insular Newfoundland, spread of *Botryllus schlosseri* is primarily attributed to fouling of transport vessels, noting its short larval stage and sessile adult

stage (Lacoursière-Roussel et al. 2016). In eastern Canada (excluding NL), this dominant non-indigenous species (Lacoursiere-Roussel et al. 2012) recorded on recreational boats is thought to have increased in abundance in the recent past (Sephton et al. 2011). Empirical evidence suggests that frequent primary introductions of *Botryllus schlosseri* in Nova Scotia occur by ships with subsequent spread by recreational boating (Lacourisiere-Roussel et al. 2012a, b).

Genetic assessments using the *CO1* region have identified 60 unique haplotypes (Bock et al. 2011), and high levels of population subdivision (Ben-Shlomo et al. 2010, Lopez-Lengtil et al. 2006, Lejeusne et al. 2011). Results from one phylogenetic study using *CO1* and the 18S rRNA subunit indicated that this species consists of at least five morphologically cryptic species, with only one geographically widespread clade (Bock et al. 2011). In Nova Scotia, researcher have identified ten *Botryllus schlosseri* haplotypes among 26 studied populations (Lacourisiere-Roussel et al. 2012a).

In Newfoundland harbours, the patchy distribution of *Botryllus schlosse*ri has (Callahan et al. 2010) has not reached "invasive" levels of abundance. Although one of the first records of *B. schlosseri* in NL was recorded in 1975 on the west coast (Hooper 1975), it currently occurs only on the Avalon Peninsula and the southern coast (Callahan et al. 2010). Newfoundland waters may define the northern range limit of this species, resulting in more intense selection pressures resulting from differences in the magnitude and duration of the minimum ocean temperature compared to its habitat in other parts of the world. New evidence suggests that cold temperatures are shaping the life history strategies of *B. schlosseri* so that it can achieve reproductive success in a colder, sub-arctic environment (Ma et al. 2017). Thus, although some studies (e.g. Lejeusne et al. 2011; Bock et al. 2011) have examined gene flow in non-invasive ascidians (including *Botryllus schlosseri*) in other parts of Canada, the factors contributing to the

successful spread of *B. schlosseri* in Newfoundland harbours may be site specific. A better understanding of the regional distribution pattern and species diversity of *B. schlosseri* may facilitate discernment of anthropogenic factors leading to range expansion in marine biota, and enable diagnosis of changes to ocean ecosystems that occur in response to anthropogenic stressors.

OBJECTIVES

The 5' region of the *CO1* gene exhibits sufficient intra-specific variability as a species-specific marker for ascidian tunicates in Newfoundland (Callahan et al. 2010), although no population genetic studies of *Botryllus schlosseri* in NL have yet been completed. The objectives of this study are to: i) characterize genetic variation of *Botryllus schlosseri* in Newfoundland populations, and ii) identify regional patterns of gene flow and population subdivision in Newfoundland populations using the typical animal barcoding region of *CO1*.

MATERIALS AND METHODS

Sample Collection, DNA Extraction, Amplification and Sequencing

We sampled *Botryllus schlosseri* from seven harbours on the coast of insular Newfoundland, Canada (Table 3.1, Figure 3.1) between 2011 and 2012. Colonies from all populations except Foxtrap were collected by SCUBA diving. The sampling of *B. schlosseri* in Foxtrap occurred by excising colonies from floating docks immediately after dock removal from the water in late fall of 2011. To avoid re-sampling of clones we collected all tissue samples with at least 1 m between adjacent specimens. Samples were preserved in 95% ethanol at 4°C for genetic analyses.

To maximize the quality of the DNA extraction, we extracted 15mg of zooids from the tunic using a dissecting microscope and preserved them in 95% ethanol. Total genomic DNA

was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Mississauga, ON) following the manufacturer's protocol. Three washes with nuclease-free water removed residual ethanol prior to DNA extraction.

We amplified a fragment of the *cytochrome oxidase subunit 1* (*CO1*) gene using the LCO1490 and HCO2198 invertebrate primers (Folmer et al. 1994). Amplifications were carried out in a volume of 25 µL, with 100 ng of DNA, 0.25 units of Top Taq DNA polymerase (Qiagen), 2.5 µL of 10x PCR buffer, 0.2 µL of dNTP's and 0.4 µL of each primer. We performed thermal cycling parameters using the profile from Lejeusne et al. (2011), and purified PCR products using the QIAquick Purification Kit (Qiagen, Mississagua, ON). Genome Quebec sequenced purified PCR products, using BigDye Terminator 3.1 chemistry and an ABI automated sequencer (Applied Biosystems, Foster City, CA). Mitochondrial *CO1* sequence alignment and editing used Sequencher v. 4.9 (Gene Codes Corporation, Ann Arbor, MI). To facilitate comparison between *B. schlosseri* populations between Newfoundland and other parts if the world, we added data from Genbank to our dataset for twenty-five populations in Nova Scotia, twelve populations in Eastern Europe, and three populations on the Pacific coast of North America (Appendix 3.1).

Population Genetic Analysis

We used DNASP v.5 (Rozas et al. 2003) to identify individual haplotypes, calculate the haplotype diversity and nucleotide diversity, and test whether sequences evolved neutrally using Tajima's D statistic. MEGA v. 5.2 (Tamura et al. 2011) identified the best fit nucleotide substitution model under the Akaike Information Criterion (AIC; Posada and Buckley 2004). We examined genetic differentiation among populations by computing ΦST (10,000 permutations) using the Tamura and Nei substitution model in ARELEQUIN v. 3.1 (Excoffier et al. 2005), with

significance levels adjusted using sequential Bonferroni corrections (Rice 1989). Results were visualized with non-metric multidimensional scaling (NMDS) plots using SPSS v.19. To assess genetic differentiation among sampling sites, we conducted a hierarchical analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) based on 10,000 random permutations in ARLEQUIN v.3.1 (Excoffier et al. 2005). Molecular variance was partitioned into four levels for the Newfoundland populations: among populations, within populations, among bays, and within bays. For the levels where sampled populations were sorted into bays, Conception Bay included the sampled locations of Long Pond and Foxtrap; Fortune Bay included the sampled population in Hermitage, and Placentia Bay included four sampled populations (Placentia, Fox Harbour, Spencer's Cove, Arnold's Cove). We used AMOVA to explore hypotheses further around interprovincial genetic structure between Newfoundland and Nova Scotia populations. For Nova Scotia, we grouped populations into three groups based on the method of Lacoursière-Roussel et al. (2012). Briefly, data from 26 locations in Nova Scotia with sample sizes ranging from 16-63 colonies were divided into three geographic groups, the Strait of Canso, the Bras d'Or Lake Region, and the Southeast Coast.

We used a Mantel test to assess correlation between genetic distance [ΦST / (1 − ΦST)]) and geographic distance between Newfoundland populations under an isolation by distance model, implemented using the Isolation By Distance Web Service (Jensen et al, 2005). Geographical distances were calculated as the minimum coastline distances between adjacent sampling locations using GOOGLE EARTH v.4.3 (beta).

Phylogeographic Analysis

We examined relationships among the *CO1* haplotypes in Newfoundland waters using a statistical parsimony network generated at the 95% connection limit and applying the Median-Joining algorithm with TCS v. 1.21 (Clement et al. 2000).

RESULTS

Population Diversity and Differentiation of *Botryllus schlosseri* in Insular Newfoundland, Canada

The final *CO1* alignment was 524 bp in length and contained 30 polymorphic sites among 210 individuals (Newfoundland harbours), all of which were parsimony informative. Our analysis identified no non-synonymous substitutions and six haplotypes, all recorded previously (Lejeusne et al. 2011; Bock et al. 2012), among the 7 Newfoundland harbours (Figure 3.1). All six haplotypes identified in Newfoundland were previously recorded in Nova Scotia (Lacoursière-Roussel et al. 2012), and three of these haplotypes (HA, HO and HB) have also been identified in Europe (Lejeusne et al 2011).

We identified two private haplotypes (Bs14 and HB) and found Haplotype Bs14 in four sampled colonies in Long Pond and haplotype HB in two individuals in Foxtrap. The dominant haplotype was HO (50.95%) followed by Bs2 (40.0%), Bs8 (3.33%), HA (2.86%), Bs14 (1.90%) and HB (0.95%). The number of haplotypes ranged from 1 to 5 in each population. Haplotype diversity overall was $h = 0.581$ and ranged from $h = 0$ in Hermitage to $h = 0.706$ in Long Pond. Haplotype diversity generally increased from west to east in insular Newfoundland (Figure 3.1, although an IBD analysis indicated no statistically significant pattern ($p > 0.5$, data not shown). Nucleotide diversity ranged from $\pi = 0$ to $\pi = 0.02$ (Table 3.1). In Newfoundland waters, the spatial distribution of haplotypes was geographically discontinuous, with 57% (12 of 21 comparisons) of population pairwise ΦST values remained significant after sequential Bonferroni corrections (Table 3.2). Of the seven populations in Newfoundland harbours, based on pairwise comparisons, Spencer's Cove and Hermitage differed significantly from Foxtrap, Arnold's Cove, Long Pond, Fox Harbour and Placentia but not from each other (Table 3.2).

Without *a priori* grouping, a hierarchical analysis of molecular variance (AMOVA) suggested significant differentiation of individuals within and among the Newfoundland populations (Table 3.3). When considering only the three Newfoundland bays (i.e. Fortune Bay, Placentia Bay and Conception Bay), most variation was within and among populations, rather than among bays (F_{CT} P=0.360, Table 3.3).

Genetic Structure in Newfoundland Harbours Relative to Nova Scotia and European Populations

Between the Nova Scotia and Newfoundland populations, 45% (76 of 168) of pairwise comparisons remained significant after Bonferroni corrections (data not shown). Among Newfoundland and Europe, 97.8% (82 of 84) of pairwise ΦST values remained significant after Bonferroni corrections (data not shown). The NMDS plots showed five of the seven Newfoundland populations clustered closely with Nova Scotia populations. Spencer's Cove and Hermitage did not cluster with the Nova Scotia populations or the four other Newfoundland populations, nor did they cluster with the European or Pacific populations (Figures 3.2 and 3.3). Generally, the Avalon Peninsula populations from Newfoundland waters clustered closely with the Nova Scotia populations. To further explore hypotheses to explain different population groupings in Nova Scotia and Newfoundland, we grouped Nova Scotia populations into three geographical groups as per Lacoursière-Roussel et al. (2012) and compared to grouped populations from each of the three Newfoundland Bays using AMOVA analyses. With all *a priori* groupings, we found significant genetic differentiation of individuals within populations and among populations within groups (Table 3.4). Except for the grouping of populations in Conception Bay, NL with Bras'd'Or Lake Region in Nova Scotia (F_{CT} =0.427, $p<0.05$), no other groupings showed statistically significant genetic subdivision (Table 3.4).

Phylogenetic Reconstruction

The parsimony haplotype network identified three distinct groups of haplotypes at the 95% confidence level (Figure 3.4). The two private haplotypes identified in Newfoundland waters separated only from one of the most common haplotypes in Newfoundland, Bs2, by one mutational step (Figure 3.4). This common haplotype, identified on the Avalon Peninsula and in Placentia Bay, appears separate from the other dominant Newfoundland haplotype HO by 15 mutational steps (Figure 3.4). A third haplogroup included private haplotype HB and haplotype HA, both previously reported in Europe (Lejeusne et al. 2011). In this third group, haplotype HA is the most distantly related to a common ancestor of the three haplogroups, by an estimated by 13 mutational steps. We inferred HB to relate most closely to a common ancestor of the three groups (Figure 3.4).

DISCUSSION

Levels of Genetic Diversity in Newfoundland Waters Relative to its Native Range

Compared to its native range in Europe*, Botryllus schlosseri* in Newfoundland apparently has slightly lower genetic diversity, and a different haplotype composition. Specifically, a previous study in Southern Europe, identified 16 *CO1* haplotypes 181 specimens (López-Legentil et al. 2006). The 16 European *B. schlosseri* haplotypes in the Mediterranean region represented three strongly supported clades, and perhaps two additional clades (López-Legentil et al. 2006, Lejeusne et al. 2011, Bock et al. 2012). The haplotypes identified in Newfoundland, in contrast, all belong to a single geographically widespread clade in Europe. This result was not surprising given expected lower genetic diversity in the non-native range (Allendorf and Lundquist 2003).

Low haplotype diversity characterized many of the sampled populations in Newfoundland waters (Figure 3.1, Table 3.1). Populations in small harbours and marinas likely experience high stochasticity in the arrival of individuals, which may result in lower genetic diversity (Dupont et al. 2007, Perez-Portela et al, 2012). This reduced diversity may occur in Newfoundland *B. schlosseri* populations sampled in Hermitage and Spencer's Cove, where the combined effects of bottlenecks, genetic drift, and/or adaptation to a subarctic environment have resulted in reduced genetic diversity. The sampling in Hermitage occurred in April, in contrast to sampling of populations in Conception Bay in late summer and early fall. Sampling during the same time window can improve the identification of sources of introduced populations (Goldstein et al. 2013). As the population size (measured as percent cover of colonies on wharf pilings) of this species in Newfoundland fluctuates throughout the year, peaking in September (Ma et al. 2017), we recommend sampling all sites in the fall to describe more accurately the genetic structure of this species in subarctic waters.

Geographic Patterns of Genetic Diversity in Newfoundland and Nova Scotia Waters

Our results suggest that regional boating traffic is a likely vector of introduction for *B. schlosseri* in insular Newfoundland. Comparable levels of genetic diversity of *Botryllus schlosseri* populations in Nova Scotia and Newfoundland waters (Figures 3.2, 3.3; Table 3.4) reflect the shared dominant haplotypes, Bs2 and HO, although the relative frequency of each haplotype in all populations surveyed differed between the two provinces. Haplotype HO accounts for just over 50% of individuals in Newfoundland harbours, whereas the most common haplotype in Nova Scotian waters, Bs2, occurred in 76% of individuals (Lacoursière-Roussel et al. 2012). The highest number of haplotypes recorded in any Nova Scotia port or high traffic marina was four (Lacoursière-Roussel et al. 2012), compared to five in one Newfoundland harbour. Populations in Newfoundland displayed higher overall diversity indices (haplotype

diversity of $h = 0.581$ and nucleotide diversity of pi = 0.01846) relative to ports ($h = 0.5$ and, pi = 0.0123) or marinas ($h = 0.294$, pi = 0.0079) in Nova Scotia (Lacoursière-Roussel et al. 2012). Additionally, private *CO1* haplotypes in Nova Scotia (Bs4, Bs13, Bs15, Bs16) were not identified in Newfoundland (Lacoursière-Roussel 2012). Two haplotypes identified in Newfoundland (Bs8 and Bs14) occurred in low frequencies in Nova Scotia populations. Bs8 occurred primarily in marinas in the Cape Breton region, and Bs14 in Nova Scotia in a high traffic yacht club. In Newfoundland, we recorded Bs14 only recorded in Long Pond (a private haplotype in NL), the location of the yacht club, suggesting regional boating connectivity. To support this interpretation, we found neither Bs14 nor Bs8 in Nova Scotia ports. In Nova Scotia, despite generally higher genetic diversity of *B. schlosseri* populations in ports relative to marinas, evidence suggests that primary introductions occur occasionally through recreational boating (Lacoursière-Roussel et al. 2012).

Probable Source Populations for *Botryllus schlosseri* in Newfoundland Harbours

Taken together, these results strongly suggest a Nova Scotian origin for the Newfoundland populations of *Botryllus schlosseri*. However, the lack of genetic information for *B. schlosseri* populations from Prince Edward Island or New Brunswick means we cannot rule out these locations as a point of origin. Additionally, we cannot exclude the hypothesis of at least one introduction from Europe. Genetic evidence shows that another marine, globally invasive species, European green crab, *Carcinus maenas*; twice invaded Northwest Atlantic waters, and spread to Placentia Bay, NL from Nova Scotia via anthropogenic transport (Blakeslee et al 2010; Jeffrey et al, 2017a; 2017b; Lehnert et al, 2018). Strong evidence suggests the Placentia Bay populations share genetic similarity with populations on the central and western Scotian Shelf (Blakeslee et al 2010), although an alternative hypothesis of admixture from northern and

southern (the Bay of Fundy to New York) populations in the North West Atlantic cannot be ruled out completely (Blakeslee et al 2010).

Although both regional and international shipping traffic may have introduced *B. schlosseri* to Newfoundland, regional shipping traffic likely provides a more significant vector for spread of marine invertebrates. A Canadian study of shipping traffic, biofouling, and invasion success showed the greater probability of ships from neighbouring ecosystems establishing populations of non-indigenous species compared to ships from more distant locations (Lacoursière-Roussel et al. 2016). Shorter trips may limit transit times and exposure to abiotic conditions (e.g. harsh oceanic currents), and thus increase the probability of biofouling species surviving transit (Lacoursière-Roussel et al. 2016). Additionally, greater environmental similarity between source and recipient regions likely increases the likelihood of establishment (MacIsaac et al. 2002; Smith et al. 1999). Previous work showed that shipping traffic correlates positively with genetic diversity of *B. schlosseri* (Lacoursière-Roussel et al. 2016). In particular, in eastern Canada, the presence of non-merchant vessels (defined as commercial fishing vessels, cruise ships, government boats, ferries, tugs, barges, crane ships, drill rigs, dredgers, research vessels and offshore support vessels) predicts genetic diversity of *B. schlosseri* better than the presence of merchant (i.e. commercial goods carriers) ships (Lacoursière-Roussel et al. 2016).

If non-merchant shipping traffic is the best predictor of genetic diversity in insular Newfoundland, then we might predict highest genetic diversity in populations outside Placentia Bay. As home to the only oil refinery in Newfoundland, Placentia Bay is one of the busiest bays in Newfoundland and receives national and international shipping traffic, as well as commercial fishing vessels and some recreational boating activity (Blakeslee et al. 2010). In our study, the highest genetic diversity in Newfoundland populations indeed occurs in Placentia Bay. Thus,

non-merchant boating activity and regional boating connectivity in Newfoundland waters may play a greater role in the spread of *B. schlosseri* in Newfoundland than international and/or other commercial shipping. However, support for this hypothesis requires further studies to examine the roles of the various types of boating, vessel fouling, and boating connectivity.

Selection Pressure, Phenotypic Plasticity, and Evolution

The stochastic demographic processes (founder effects and bottlenecks) associated with colonization suggest that strong genetic drift should reduce genetic diversity of invasive populations (Nei et al. 1975). A range expansion also exposes non-indigenous species to new environmental stressors that could result in strong natural selection (Sakai et al. 2001).

Botryllus schlosseri has become established in insular Newfoundland, despite very low genetic variation in some populations shown in our study. Additional empirical evidence suggests establishment of some invasive species despite low levels of genetic diversity (Dlugosch and Parker 2008, Roman and Darling 2007), and putative adaptive evolution of species during invasion occurs frequently (Whitney and Gabler 2008). Limitations of the approach to measuring genetic variation could explain differences between theory and empirical evidence. Researchers routinely use neutral genetic markers to study the range expansions of invasive species to identify sources of introductions, pathways of introduction, patterns of population connectivity, and demographic history (Tepolt et al. 2009; Bock et al. 2015). However, neutral gene variants do not directly affect population fitness and therefore provide limited information on the adaptive potential of invasive populations (Holderegger et al. 2006; Lal et al. 2016).

Newfoundland waters represent the northern range limit for *Botryllus schlosseri* in eastern North America, extending into subarctic waters. *B schlosseri* in Newfoundland harbours

likely experiences selection pressure given the role of temperature is a predominant seasonal trigger for the onset of recruitment in colonial ascidians (Westerman et al. 2009); previous studies correlate temperature with the number and density of colonies in its native Mediterranean Sea (Rinkevich et al. 1998). In Arnold's Cove, colonies occur year-round, even though the percent cover of the colonies on wharf structures ranges from <1% in May to a maximum of c. 3% in October (Ma et al 2017). These percent cover values are generally much smaller than in other regions within the non-native range of *B. schlosseri* (McNaught and Norden 2011). Additionally, the seasonal window for sexual reproduction in this population, spans about 2.5 months (Ma et al 2017), a much shorter period than reported for this species in the literature (about 5 months, see Table 1 in Ma et al. 2017). Despite these findings, 90% of autumn-recruited colonies survive winter in Arnold's Cove and reproduce sexually the following spring (Deibel et al. 2014). That *B. schlosseri* has shortened its reproductive period in Newfoundland indicates a physiological or behavioral response (Ma et al. 2017), known as activational plasticity (Snell-Rood 2013). Additionally, developmental phenotypic plasticity may allow a single genotype to express different phenotypes depending on the environment it was exposed to during development (West-Eberhard 2005), providing scope for phenotypic changes suited to colder ocean temperatures.

To survive in novel environments, *Botryllus schlosseri* has potentially undergone adaptive evolution, assuming the presence of sufficient genetic variation for traits that confer a fitness advantage in sub-Arctic environments (Fisher 1930). Other authors suggest evolution of *Botryllus schlosseri*, such that three to five cryptic, sibling species comprise the species (Bock et al. 2012; Nydam et al. 2017). Of the sibling species, only one species is geographically widespread, and the *B. schlosseri* haplotypes identified in Newfoundland all belong to the

geographically widespread species. Despite no apparent latitudinal difference in genetic structure of *B. schlosseri* in Newfoundland in our study, survival and adaptation along an environmental gradient can influence fine-scale genetic structure (Endler 1977; Sanford and Kelly 2011; Stanley et al. 2018). Reduced genetic diversity was reported for *B. schlosseri* along the Atlantic Coast in Europe (Ben-Shlomo et al. 2006). Researchers have provided evidence of genetic-environmental associations in European Green crab in Eastern North America (Jeffery et al. 2018, lobster (Benestan et al. 2016), and Atlantic cod (Bradbury et al. 2010). Despite these studies, most evidence for genetic adaptation in marine invasive species comes from quantitative studies that show divergence in a range of morphological or physiological traits, rather than identifying the genetic loci underlying adaptive traits (reviewed in Tepolt 2015). Genomic tools provide a means to assess the role of adaptive genetic variation and plasticity in the success of marine invasions by characterizing adaptive genomic variation, identifying candidate genes underlying adaptive traits, and understanding the role of epigenetics in plastic responses (Tepolt 2015). As *Botryllus schlosseri* occupies the edge of its biogeographic range in Newfoundland, and the early phase of range expansion presumably favours evolution of increased plasticity (Lande 2015), further studies utilizing NGS technologies may facilitate knowledge of the genetics of range expansion in ascidian tunicates.

CONCLUSIONS

Our results identify regional boating traffic as a likely vector of introduction for *B. schlosseri* in Newfoundland harbours, but international shipping and local spread may also play an important role. The use of microsatellite loci, or a next generation sequencing method (e.g. RAD seq) would elucidate finer levels of genetic structure *of B. schlosseri* in Newfoundland waters, potentially allowing the identification of clonal samples and a more precise identification

of multiple introductions, primary introductions, and source population(s). The use of these additional techniques in future studies would also validate this study's results, as well as address the limitations of using a single mitochondrial gene to infer species patterns of migration and gene flow.

Global warming may promote the spread of invasive species into previously uninhabited areas as warming sea temperatures creates more favourable habitats (Stachowicz et al. 2002), and a higher volume of maritime traffic associated with a longer ice-free season facilitates the transport of non-native ascidians (Lambert et al. 2010). As the fouling of boat hauls likely provides the dominant vector of *Botryllus schlosseri* transport, warming sea temperatures may also increase the probability of individuals surviving transport and increase survivorship of introduced individuals. Understanding of the evolution of range expansions in the sea will depend on combining information on marine transport, environmental heterogeneity, neutral genetic diversity, epigenetics, genomic variation, and biogeographic distributions of species.

Table 3.1: Locations of *Botryllus schlosseri* sampling sites in Newfoundland waters and genetic diversity indices for *cytochrome oxidase 1* with *n*, sample size; *Nh*, number of haplotypes; *h*, haplotype diversity; *π*, nucleotide diversity; *Hap*, haplotype

Location and Geographic Coordinates (Latitude and Longitude)	ID	n	Nh	Hap code	h	π
Hermitage (47.5599, -55.9258)	H	30	1	HO	0	0
Spencer's Cove (47.6677, -54.0786)	SC	30	3	HO, HA, Bs2	0.131	0.005
Arnold's Cove (47.7584, -53.9878)	AC	29	3	HO, HA, Bs2	0.569	0.020
Placentia (47.2660, -53.9057)	PL	29	2	Bs2, HO	0.192	0.007
Fox Harbour (47.3206, -53.9123)	FH	29	2	Bs2, HO	0.379	0.013
Foxtrap (47.5135, -52.9982)	FT	29	5	HO, Bs2, HB, Bs8,	0.700	0.017
Long Pond (47.5106, -52.9734)	LP	34	5	Bs2, HA, Bs8, HO, Bs14	0.706	0.019

Table 3.2: Pairwise Φ_{ST} values between populations of *Botryllus schlosseri* in Newfoundland harbours. Asterisks indicate significant values after Bonferroni corrections (α=0.05). Table 3.1 provides locality codes.

	LP	**FT**	**FH**	**SC**	**PL**	**AC**	**H**
LP							
FT	-0.025						
FH	0.016	0.013					
SC	0.458*	0.505*	0.656*				
PL	0.148	0.162	0.031	0.810*			
AC	0.075	0.093	0.230*	0.202*	0.413*		
H	0.556*	0.615*	0.753*	0.018	0.895*	0.316*	

Table 3.3: Analysis of molecular variance (AMOVA) from mtDNA *CO1* sequences from seven *Botryllus schlosseri* populations in Newfoundland waters. d.f. denotes degrees of freedom

Source of Variation	d.f.	Sum of Squares	Variance Components	% Variation	Fixation indices	P value
Among Populations	6	415.983	2.203 Va	40.24		
Within Populations	203	664.156	3.272 Vb	59.76	F_{ST}: 0.411	0
Among Bays	2	160.814	0.251 Va	4.51	F_{CT}: 0.045	0.361
Among Populations within Bays	4	255.169	2.034 Vb	36.60	F_{SC}: 0.383	0
Within Populations	203	664.156	3.272 Vc	58.89	F_{ST}: 0.402	0

Table 3.4: AMOVA table of differentiation indices for *Botryllus schlosseri CO1* sequences for different hypotheses of geographical population groupings. Bold values indicate significant differences at P<0.05.

Comparison	Differentiation indices		
	F_{ST}	F_{SC}	F_{CT}
regional			
Newfoundland vs Nova Scotia	**0.459**	**0.289**	**0.239**
subregional			
Conception Bay, NL (FT/LP) vs south-east Coast of Nova Scotia	**0.129**	**0.174**	-0.054
Conception Bay, NL (FT/LP) vs Strait of Canso Region of Nova Scotia	**0.135**	**0.075**	0.683
Conception Bay, NL (FT/LP) vs Bras d'Or Lake Region of Nova Scotia	**0.488**	**0.107**	**0.427**
Placentia Bay, NL(AC/SC/PL/FH) vs south-east Coast of Nova Scotia	**0.286**	**0.287**	-0.001
Placentia Bay, NL(AC/SC/PL/FH) vs Strait of Canso Region of Nova Scotia	**0.363**	**0.261**	0.138
Placentia Bay, NL(AC/SC/PL/FH) vs Bras d'Or Lake Region of Nova Scotia	**0.677**	**0.398**	0.464
Fortune Bay, NL (H) vs south-east Coast of Nova Scotia	**0.596**	**0.223**	0.479
Fortune Bay, NL (H) vs Strait of Canso Region of Nova Scotia	**0.721**	**0.118**	0.683
Fortune Bay, NL (H) vs Bras d'Or Lake Region of Nova Scotia	**0.947**	**0.276**	0.927

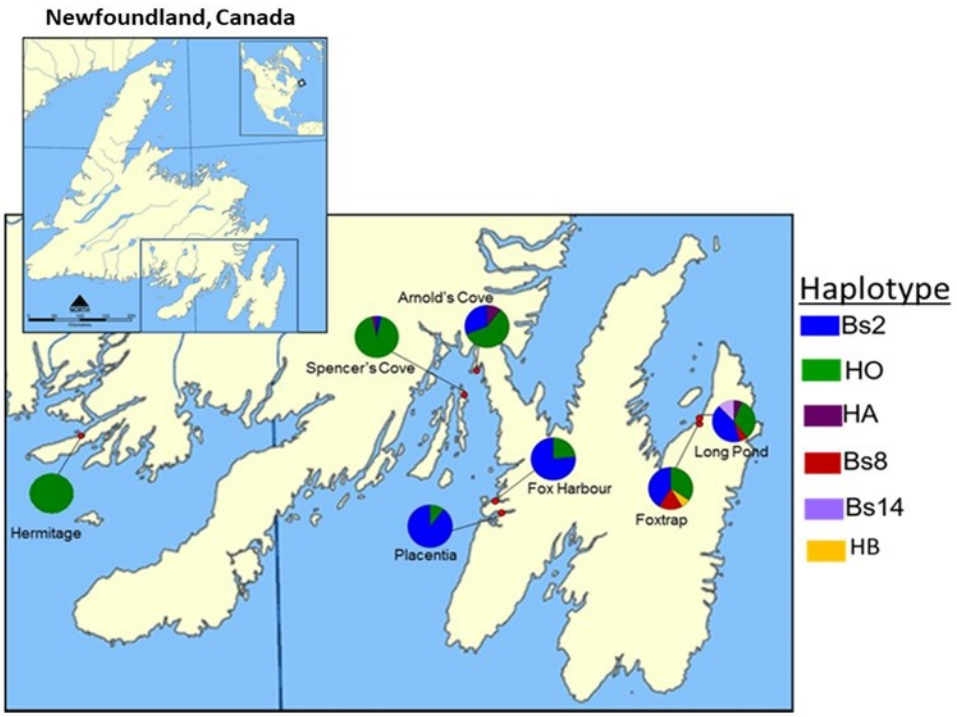

Figure 3.1: Sampling locations and relative frequencies of *Botryllus schlosseri* mitochondrial *cytochrome oxidase 1* (*CO1*) haplotypes from seven harbours in insular Newfoundland, Canada. Map courtesy of T. Wells, Department of Fisheries and Oceans, Canada.

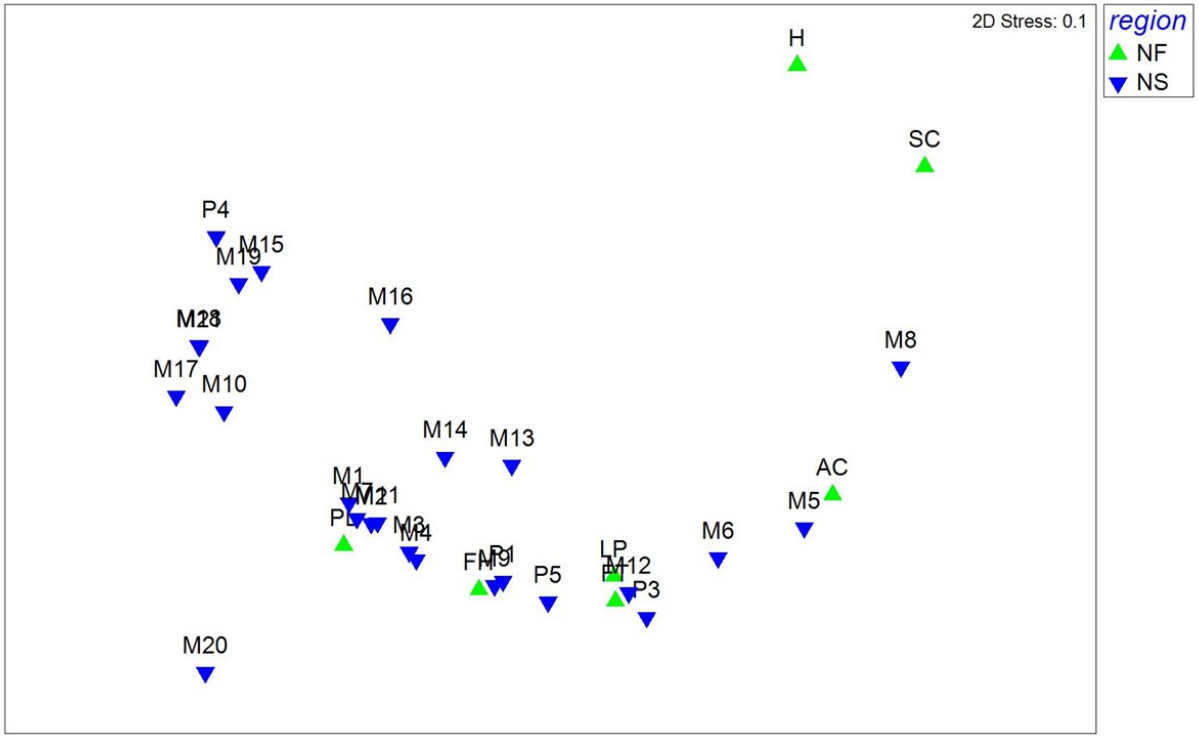

Figure 3.2: A non-metric multidimensional scaling (NMDS) plot of *Botryllus schlosseri* communities across harbours in Newfoundland (NF) and Nova Scotia (NS), Canada based on pairwise ΦST **values estimated from mitochondrial *cytochrome oxidase 1*.** Green triangles denote locations in insular Newfoundland, whereas inverted blue triangles denote locations in Nova Scotia. Location abbreviations for Nova Scotian populations match those listed in Appendix 3.1. For Newfoundland populations AC (Arnold's Cove), SC (Spencer's Cove), H (Hermitage), PL (Placentia), FH (Fox Harbour), FT (Foxtrap), and LP (Long Pond).

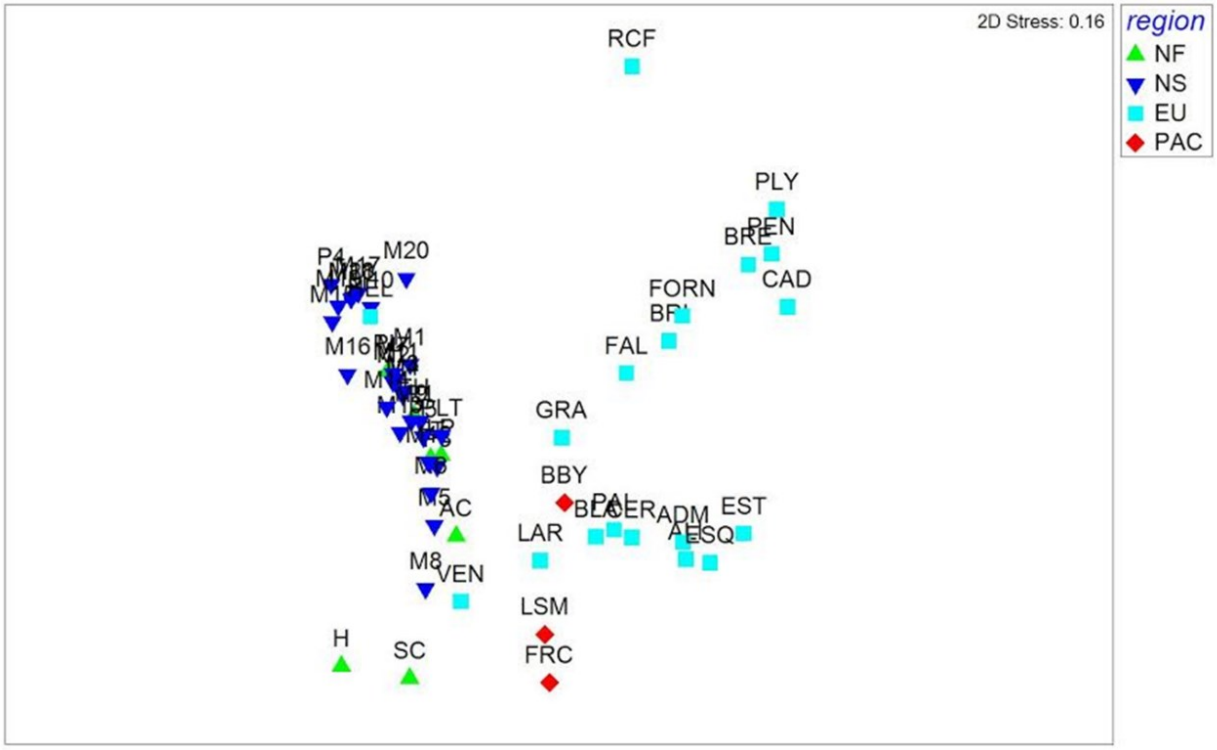

Figure 3.3: A non-metric multidimensional scaling (NMDS) plot of *Botryllus schlosseri* communities across harbours in Canada and eastern Europe, based on pairwise ΦST values estimated from mitochondrial *cytochrome oxidase 1*. Green triangles denote locations in insular Newfoundland, inverted blue triangles denote locations in Nova Scotia, blue squares denote locations in Europe, and red diamonds denote locations on the Pacific coast of North America. Location abbreviations match those listed in Appendix 3.1 For Newfoundland populations AC (Arnold's Cove), SC (Spencer's Cove), H (Hermitage), PL (Placentia), FH (Fox Harbour), FT (Foxtrap), and LP (Long Pond).

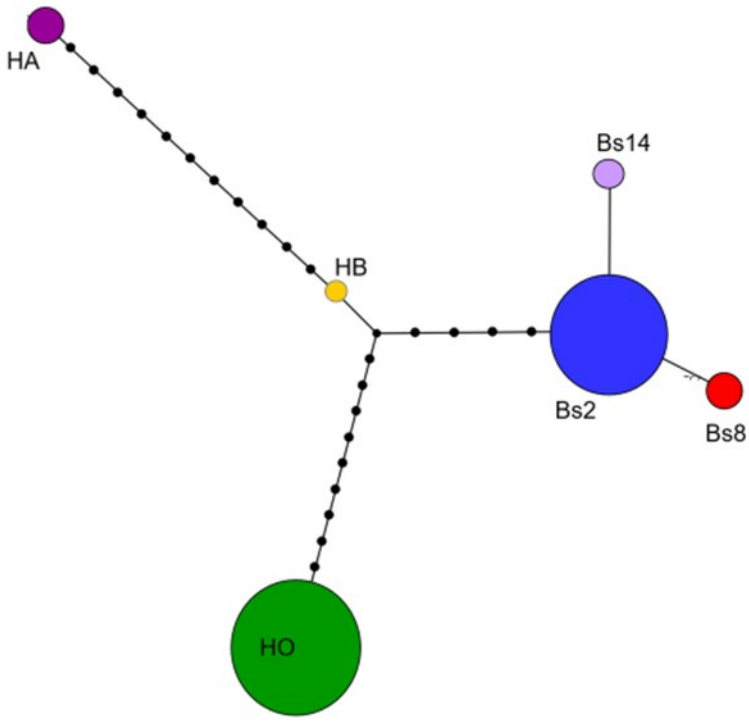

Figure 3.4: Statistical parsimony network of *Botryllus schlosseri* in Newfoundland waters based on *cytochrome oxidase 1* haplotypes. Circle size is proportional to haplotype frequency, and small black dots indicate unsampled haplotypes inferred from the data.

CHAPTER 4: GENERAL CONCLUSION

In the two empirical studies presented above, we sequenced a segment of the cytochrome oxidase 1 gene in two tunicate species, to address questions of genetic diversity and structure of these species in insular Newfoundland. For one of the species, *Oikopleura vanhoeffeni*, we provide the first report on the genetic structure of the *CO1* gene. The paucity of genetic information available for the *CO1* gene in this species likely relates to the finding here that several non-coding, repetitive regions interrupt the gene. Acknowledging extensive reports of genetic structure based on a fragment of the *CO1* gene for *Botryllus schlosseri* from many localities around the globe, we provide the first report of genetic structure for this species in Newfoundland harbours. For both *Oikopleura vanhoeffeni*, and *Botryllus schlosseri*, the use of a fragment of the *CO1* gene has limitations as a DNA barcode for species identification, and as a molecular marker to address species ranges limits and patterns of connectivity. In this thesis, we suggest that *Oikopleura vanhoeffeni* may represent a cryptic species complex, but this hypothesis requires further evaluation, because sequences from congeneric species were unavailable for distance comparisons. Other authors have suggested that *Botryllus schlosseri* represents a cryptic species complex based on whole mitochondrial genome comparisons (Bock et al. 2011; Griggio et al. 2014). Another tunicate, *Ciona intestinalis*, previously thought to represent a single species based on the 5' region of *CO1*, was also found to comprise multiple cryptic species on further evaluation with additional genetic markers (Zhan et al. 2010).

Accurate species identification represents one of the most fundamental aspects of ecological research and monitoring (Baerwald et al. 2020). Without accurate species identification of *Oikopleura vanhoeffeni*, we may never fully understand species-specific differences in life history characteristics and population dynamics of this species, and a

morphologically similar, co-occurring species, *Oikopleura labradoriensis*. Given the success of whole mitochondrial genome approaches in the identification of cryptic and morphologically similar ascidian species, we recommend this approach for appendicularians. Distinguishing appendicularian species at the molecular level using an approach that considers multiple markers throughout the genome may enable the use of a "mini barcode" (Meusnier et al. 2008) with a universal appendicularian primer set based on a portion of *CO1* and other genes. The utility of this tool may facilitate the rapid identification of appendicularians in environmental DNA samples, or where life stage or damage during sampling limits the use morphology of the animals. In addition to its limitations for species level identification in tunicates, the use of *CO1* as the only molecular marker in my work on the non-native species *Botryllus schlosseri*, revealed limited genetic variation in this species in Newfoundland waters. Researchers should consider use of microsatellite markers or a whole genome approach to further elucidate source locations of introduction, and pathways of spread of this non-native ascidian in sub-Arctic waters. In addition, whereas most studies of marine invasive species have focused on population connectivity and neutral genetic variation (Sherman et al. 2016), the lack of recognition of the role of plastic responses and the evolutionary potential of these species represents a key limitation of existing management approaches to invasive species (Whitney & Gabler 2008). Epigenetic changes represent one way that non-native species may achieve plasticity. Past studies provide evidence of epigenetic changes induced from environmental cues in the invasive tunicate ascidian *Didemnum vexillum* (Hawes et al. 2018). Given *B. schlosseri* is at the northern edge of its range in Newfoundland waters, this species could serve as a model organism for examining the epigenetic processes facilitating range expansion in invasive marine species.

Bibliography

Allendorf FW and Lundquist LL (2003) Introduction: population biology, evolution, and control of invasive species. Conserv Biol 17: 24-30

Alldredge AL (2005) The contribution of discarded appendicularian houses to the flux of particulate organic carbon from the ocean surface waters. *In* Gorsky, G., Youngbluth MJ, and Deibel D (eds), Response of Marine Ecosystems to Global Change: Ecological impact of appendicularians. Contemporary Pub. Int. Paris, France, pp.309-326

Baerwald MR, Goodbla A, Nagarajan RP, Gootenberg JS, Abudayyeh O, Zhang F, Schreier AD (2020) Rapid and accurate species identification for ecological studies and monitoring using CRISPR based SHERLOCK. Mol Ecol Res 20: 961-970

Beagley CT, Okada NA, Wolstenholme DR (1996) Two mitochondrial group I introns in a metazoan, the sea anemone Metridium senile: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. Proc Natl Acad Sci USA, 93: 5619-5623

Ben-Shlomo R, Paz G, Rinkevich B (2006) Postglacial-period and recent invasions shape the population genetics of Botryllid ascidians along European Atlantic Coasts. Ecosystems 9: 1118-1127

Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJ et al, (2016) Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). Mol Ecol 25: 5073-5092

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 16: 314-321

Berná L and Alvarez-Valin F (2014) Evolutionary genomics of fast evolving tunicates, Genome Biol Evol 6: 1724–1738

Berná L, D'Onofrio G, Alvarez-Valin F (2012) Peculiar patterns of amino acid substitution and conservation in the fast evolving tunicate *Oikopleura dioica*. Mol Phyl Evol 62: 708-717

Berrill NJ (1928) The identification and validity of certain species of ascidians. J. Mar. Biol. Assoc. U. K., 15: 159-175

Berrill NJ (1950) The Tunicata: With an account of the British species. The Ray Society, London, 354 pp

Blakeslee AMH, McKenzie CH, Darling JA, Byers JE, Pringle JM, Roman J (2010) A hitchhikers guide to the Maritimes: anthropogenic transport facilitates long-distance dispersal of an invasive marine crab to Newfoundland. Diversity and Distribution 16: 879-891

Bock DG, Caseys C, Cousens RD, Ma Hahn, Heredia SM, Hübner S,Turner KG, Whitney KD, Rieseberg LH (2015) What we still don't know about invasion genetics. Mol Ecol 24: 2277–2297

Bock DG, MacIsaac HJ, Cristescu ME (2012) Multilocus genetic analyses differentiate between widespread and spatially restricted cryptic species in a model ascidian. Proc Biol Sci 279: 2377-2385

Bock DG, Zhan A, Lejeusne C, MacIsaac HJ, Cristescu ME (2011) Looking at both side of the invasion: patterns of colonization in the violet tunicate *Botrylloides violaceus*. Mol Ecol 30: 503-516

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ (2006) Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444: 85–88

Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Hardie DC (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. Proc R Soc B: Biol Sci 277: 3725-3734

Brasier MJ, Wiklund H, Neal L, et al (2016) DNA barcoding uncovers cryptic diversity in 50% of deep-sea Antarctic polychaetes. R Soc open sci 3: 160432

Briski E, Cristescu ME, Bailey SA, MacIsaac HJ (2011) Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. Biol Invasions 13: 1325–1340

Brunetti R Griggio F Mastrototaro F, Gasparini F, Gissi C (2020) Toward a resolution of the cosmopolitan *Botryllus schlosseri* species complex (Ascidiacea, Styelidae): mitogenomics and morphology of clade E (Botryllus gaiae). Zool J Linn Soc: zlaa023

Bucklin A, Frost BW, Bradford-Grieve J, Allen LD, Copley NJ (2003) Molecular systematic and phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. Mar Biol 142: 333–43

Bucklin A, Wiebe PH, Smolenack SB, Copley NJ, Beaudet JG, et al (2007) DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). J Plankton Res 29: 483–93

Bucklin A, Hopcroft RR, Jennings RM, Nigro LM, Ortman BD, et al (2010a) DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. Deep-Sea Res II 57: 40-48

Bucklin A, Steink, D, Leocadio BB (2010b) DNA barcoding of Marine Metazoa. Ann Rev Mar Sci 3: 471-508

Bucklin A, Ortman BD, Jennings RM, Nigro L, Sweetman CJ, et al. (2010c) A "Rosetta Stone" for zooplankton: DNA barcode analysis of holozooplankton diversity of the Sargasso Sea (NW Atlantic Ocean) Deep-Sea Res. II 57: 2234-2247

Burger G Yan Y, Javadi P and Lang BF (2009) Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. Trends Genetics 25: 381-386

Callahan AG, Deibel D, McKenzie CH, Hall JR, and Rise ML (2010) Survey of harbours in Newfoundland for indigenous and non-indigenous ascidians and an analysis of their cytochrome c oxidase I gene sequences. Aquat Invasions 5: 31-39

Caputi L, Andreakis N, Mastrotoaro F, Cirino P, Vassillo P, and Sordino P (2007) Cryptic speciation in a model invertebrate chordate. Proc Natl Acad Sci USA 104: 9364-9369

Cardenas P, Menegola C, Rapp HT, Diaz MC (2009) Morphological description and DNA barcodes of shallow-water Tetractinellida (Porifera: Demospongiae) from Bocas del Toro, Panama, with description of a new species. Zootaxa 2009: 1–39

Carver CE, Mallet AL, and Vercaemer B (2006) Biological synopsis of the colonial tunicates, *Botryllus schlosseri* and *Botrylloides violaceus*. Can Man Report Fish and Aquat Sci 2747

Chavali S, Morais DA, Gough J, Babu MM (2011) Evolution of eukaryotic genome architecture: insights from the study of a rapidly evolving metazoan, *Oikopleura dioica*: non-adaptive forces such as elevated mutation rates may influence the evolution of genome architecture. Bioessays 33: 592-601

Choe N, Deibel D (2008) Temporal and vertical distributions of three appendicularian species (Tunicata) in Conception Bay, Newfoundland. J Plankton Res 30: 969-979

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9: 1657-1659

Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh, RT, Hajibabeal M, Hebert PDN (2007) Biological identifications through DNA barcodes: the case of the Crustacea. Can J Fish Aquat Sci 64: 272-295

Dagg MJ and Brown SL (2005) The potential contribution of fecal pellets from the larvacean *Oikopleura diocia* to vertical flux of carbon in a river dominated by coastal margin. In Gorsky G, Youngbluth MJ and Deibel D (eds), Response of Marine Ecosystems to Global Change: Ecological Impact of Appendicularians. Contemporary Publishing International, Paris, France, pp. 293-308.

Darling JA and Folino-Rorem NC (2009) Genetic analysis across different spatial scales reveals multiple dispersal mechanisms for the invasive Hydrozoan *Cordylophora* in the Great Lakes. Mol Ecol 18: 4827-4840

Deibel D and Lowen B (2012) A review of the life cycles and life-history adaptations of pelagic tunicates to environmental conditions. ICES J Mar Sci: J Conseil 69:358–369

Deibel D, McKenzie CH, Rise ML, Thompson RJ and others (2014) Recommendations for the eradication and control of non-indigenous, colonial ascidian tunicates in Newfoundland harbours. Can Manuscr Rep Fish Aquat Sci: 3039

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B (2006) Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. Proc Natl Acad Sci USA 103: 8751-8756

Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439: 965-968

Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H (2008) Additional molecular support for the new chordate phylogeny. Genesis 46: 592-604

Denoeud F, Henriet S, Mungpakdee S, et al (2010) Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330: 1381-1385

Dlugosch KM and Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol Ecol 17: 431-449

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749

Dupont L, Viard F, David P, Bishop JDD (2007) Combined effects of bottlenecks and selfing in populations of *Corella eumyota*, a recently introduced sea squirt in the English Channel. Divers and Distrib 13: 808-817

Endler JA (1977) Geographic Variation, Speciation, and Clines. Princeton, NJ: Princeton Univ. Press. 262 pp.

Erpenbeck D, Aryasari R, Hooper JNA, Worheide G (2015) A mitochondrial intron in a verongid sponge. J Mol Evol 80: 13–17

Estoup A, Beaumont M, Sennsedot F, Moritz C, Cornuet JM (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. Evolution 58: 2021-2036

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479-491

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47-50

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Ann Rev Ecol Evol Sys 40: 481-501

Fisher RA (1930) The genetical theory of natural selection. Clarendon Press

Flood PR, Deibel D, and CC Morris (1992) Filtration of colloidal melanin from sea water by planktonic tunicates. Nature 355:630-632

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotech 3: 294-299

Frankham R and Ralls K (1998) Inbreeding leads to extinction. Nature 392: 441-442

Fu YX (1997) Statistical test of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915-925

Fukami H, Chen CA, Chi-Yung C, Knowlton N (2007) Novel group I introns encoding a putative homing endonuclease in the mitochondrial cox1 Gene of Scleractinian Corals. J Mol Evol 64: 591-600

Galt CP, and Flood PR (1998) Bioluminescence in the Appendicularia. In Bon, Q. (ed.), The Biology of Pelagic Tunicates. Oxford University Press, Oxford, UK pp. 215-229

Garić R, Dénes M, Batistic M (2019) Cryptic Diversity of Appendicularian *Oikopleura longicauda*. Rapp. Comm. Int. Mer Médit 42

Gatchel JR and Zoghbi HY (2005) Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet 6: 743-755

Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all taxa biotic surveys. Mol Ecol Res 13: 851-861

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. Ann Rev Genet 44: 445-477

Goetze E (2010) Species discovery in marine planktonic invertebrates through global molecular screening. Mol Ecol 19: 952-967

Goldstien JS, Inglis GJ Schiel DR, Gemmell NJ (2013) Using temporal sampling to improve attribution of source populations for invasive species PLoS ONE 8: e65656Gorsky G and Fenaux R (1998) The role of Appendicularia in marine food webs. In Bone, Q. (ed.), The Biology of Pelagic Tunicates. Oxford University Press, Oxford, UK, pp. 161–169

Griggio F, Voskoboynik A, Iannelli F, Justy F, Tilak M, Xavier T, Pesole G, Douzery EJP, Mastrototaro F and Gissi C (2014) Ascidian mitogenomics: comparison of evolutionary rates in closely related taxa provides evidence of ongoing speciation events. Genome Biol Evol 6: 591–605

Halliburton R (2004) Introduction to Population Genetics. Pearson Prentice Hill, Upper Saddle River, NJ, 623 pp.Hardy SM, Carr CM, Hardman M. et al. (2011) Biodiversity and phylogeography of Arctic marine fauna: insights from molecular tools. Mar Biodiv 41:195–210

Harpending RC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Human Biol 66: 591-600

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol Ecol 25: 2454–2466

Hawes NA, Temblay LA, Xavier P, Dunphy B, Fidler A, Smith KF (2018) Effects of temperature and salinity stress on DNA methylation in a highly invasive marine invertebrate, the colonial ascidian *Didemnum vexillum*. PEER J 6: e5003

Hawlitschek O, Morinière J, Dunz A, Franzen M, Rödder D, Glaw F, Haszprunar G (2016) Comprehensive DNA barcoding of the herpetofauna of Germany. Mol Ecol Res 16: 242-253

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270: 313-21

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landscape Ecol 21:797–807

Hooper R (1975) Bonne Bay marine resources: an ecological and biological assessment. Parks Canada Contract No. ARO-74-83

Huang D, Meier R, Todd PA, Chou LM (2008) Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J Mol Evol 66: 167–74

Jaffarali HA, Akram S, Arshan MLK (2017) DNA barcoding of a colonial ascidian, *Lissoclinum fragile* (Van Name, 1902). Mitochondrial DNA A 28: 810-813

Jeffery NW, DiBacco C, Van Wyngaarden M, Hamilton LC, Stanley RRE, Bernier R, FitzGerald J, Matheson K, McKenzie CH, Nadukkalam Ravindran P, Beiko R, Bradbury IR (2017a) RAD sequencing reveals genomewide divergence between independent invasions of the European green crab (Carcinus maenas) in the Northwest Atlantic. Ecol and Evol 7:2513-2524

Jeffery NW, DiBacco C, Wringe BF, Stanley RRE, Hamilton LC, Ravindran PN, Bradbury IR (2017b) Genomic evidence of hybridization between two independent invasions of European green crab (Carcinus maenas) in the Northwest Atlantic. Heredity 119: 154-165

Jeffrey NW, Bradbury IR, Stanley RRE, Wringe BF, Wyngaarden MVm Lowen JB, McKenzie CH, Matheson K, Sargent PS, DiBacco C (2018) Genomewide evidence of environmentally mediated secondary contact of European green crab (*Carcinus maenas*) lineages in eastern North America. Evol App 11: 869-882

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genet 6: 13

Jue NK, Batta-Lona PG, Trusiak S, et al (2016) Rapid evolutionary rates and unique genomic signatures discovered in the first reference genome for the Southern Ocean salp, *Salpa thompsoni* (Urochordata, Thaliacea). Genome Biol Evol 8: 3171 - 3186

Kemppainen P, Panova M, Hollander J, Johannesson K (2009) Complete lack of mitochondrial divergence between two species of NE Atlantic marine intertidal gastropods. J Evol Biol 22: 2000–11

Lacoursière-Roussel A, Bock DG, Cristescu ME, Guichard F, McKindsey CW (2016) Effect of shipping traffic on biofouling invasion success at population and community levels. Biol Invasions 18: 3681-3695

Lacoursière-Roussel A, Bock DG, Cristescu M, Guichard F, Girard P, Legendre P, McKindsey CW (2012) Disentangling invasion processes in a dynamic shipping–boating network. Mol Ecol 21: 4227-4241

Lacoursière-Roussel A, Forrest BM, Guichard F, Piola RF, McKindsey CW (2012) Modeling biofouling from boat and source characteristics: a comparative study between Canada and New Zealand. Biol Invasions 14: 2301-2314

Lal MM, Southgate PC, Jerry DR, Zenger KR (2016) Fishing for divergence in a sea of connectivity: the utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. Mar Genomics 25:57–68

Lambert G (2007) Invasive sea squirts: A growing global problem. J Exp Marine Biol Ecol 342: 3-4

Lambert G, Shenkar N, Swalla BJ (2010) First Pacific record of the north Atlantic ascidian *Molgula citrina*-bioinvasion or circumpolar distribution? Aquat Invasions 5: 369-378

Lande R (2015) Evolution of phenotypic plasticity in colonizing species. Mol Ecol 24: 2038–2045

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinform 23: 2947-2948

Lehnert SJ, DiBacco C, Jeffery NW, Blakeslee AMH, Isaksson J, Roman J, Wringe BF, Stanley RRE, Matheson K, McKenzie CH, Hamilton LC, Bradbury IR (2018) Temporal dynamics of genetic clines of invasive European green crab (*Carcinus maenas*) in eastern North America. Evol Appl 11:1656-1670

Lejeusne C, Bock DG, Therriault TW, MacIsaac HJ, Cristescu ME (2011) Comparative phylogeography of two colonial ascidians reveals contrasting invasion histories in North America. Biol Invasions 13: 635-650

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool 10: 34

Librado P and Rozas J (2009) DNAsp v 5. A software for comprehensive analysis of DNA polymorphism data. Bioinform 25: 1451-1452

Locke A, Hanson JM, Ellis KM, Thompson J, Rochette R (2007) Invasion of the southern Gulf of St. Lawrence by the clubbed tunicate (*Styela clava* Herdman): potential mechanisms for invasions of Prince Edward Island estuaries. J Exp Mar Biol Ecol 342: 69-77

López-Legentil S, Turon X, Planes S (2006) Genetic structure of the sea squirt, *Botryllus schlosseri*, introduced in sourthern European harbours. Mol Ecol 15:3957-3969

Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Mol Ecol 7: 1441-1455

Ma CK, Deibel D, Ben Lowen J, McKenzie, CH (2017) Spatio-temporal dynamics of ascidian larval recruitment and colony abundance in a non-indigenous Newfoundland population. Mar Ecol Prog Ser 585: 99-112

MacIsaac JH, Robbins TC, Lewis MA (2002) Modeling ships' ballast water as invasion threats to the Great Lakes. Can J Fish Aquat Sci 59: 1245-1256

McNaught DC, Norden WS (2011) Generalized regional spatial patterns of larval recruitment of invasive ascidians, mussels, and other organisms along the coast of Maine. Aquat Invasions 6: 519-523

Meusnier I, Singer GA, Landry JF et al. (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9

Milan D, Mendes IS, Damasceno JS, Teixeiria DF, Sales NG, Carvalho DC (2020) New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment. Sci Rep 10: 17966

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Bio Evol 3: 418-426

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in Populations. Evol 29: 1–10

Nishikawa T, Oohara I, Saitoh K, et al (2014) Molecular and morphological discrimination between an invasive ascidian, *Ascidiella aspersa*, and its congener *A. scabra* (Urochordata: Ascidiacea). Zool Sci 31(3): 180-185

Nydam ML, Giesbrecht KB, Stephenson EE (2017) Origin and dispersal history of two colonial ascidian clades in the *Botryllus schlosseri* species complex. PLoS ONE 12(1): e016994

Pearman JK, El-Sherbiny MM, Lanzén A, Al-Aidaroos AM, Irigoien X. (2014) Zooplankton diversity across three Red Sea reefs using pyrosequencing. Front Mar Sci 1:27

Perez-Portela R, Turon X, Bishop JDD (2012) Bottlenecks and loss of genetic diversity: spatio-temporal patterns of genetic structure in an ascidian recently introduced in Europe. Mar Ecol Prog Ser 451: 93-105.

Perez-Portela R, Turon X (2008) Phylogenetic relationships of the Clavelinidae and Pycnoclavellidae (Ascidiacea) inferred from mtDNA data. Invertebr Biol 127:108-120

Posada D, Buckley TR. (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Bio 53(5):793-808

Purcell JE, Sturdevant MV, Galt CP (2005) A review of appendicularians as prey of invertebrate and fish predators. In, Response of marine ecosystems to global change: ecological impact of appendicularians, Gorsky G, Youngbluth MJ and Deibel D, eds. pp.359-435

Rach J, Bergmann T , Paknia O, DeSalle R, Schierwater B, Hadrys H (2017) The marker choice: Unexpected resolving power of an unexplored *CO1* region for layered DNA barcoding approaches. PLoS ONE 12: e0174842

Ramos-Onsins, S. E. & J. Rozas (2002) Statistical properties of new neutrality tests against population growth. Mol Biol and Evol 19: 2092-2100

Rinkevich B, Porat R, Goren M (1998) Ecological and life history characteristics of *Botryllus schlosseri* (Tunicata) populations inhabiting undersurface shallow-water stones. Mar Ecol 19: 129-145

Rice WR (1989) Analyzing tables of statistical tests. Evol 43: 223-225

Rogers, A. R. & H. Harpending (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9: 552-569

Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB (2009) Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. Mol Ecol 18: 1560-1573

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. TREE 22: 454 - 464

Rot C, Goldfarb I, Ilan M and Huchon M (2006) Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. BMC Evol Biol 6: 71

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145: 1219-1228

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinform 19: 2496-2497

Sanford E and Kelly MW (2011) Local adaptation in marine invertebrates. Ann Rev Mar Sci 3: 509-535

Sato R, Tanaka Y, Ishimaru T (2003) Species-specific house productivity of appendicularians. Mar Ecol Prog Ser 259: 163-172Sephton D, Vercaemer B, Nicolas JM, Keays J (2011) Monitoring the invasive tunicates in Nova Scotia, Canada (2006-2009). Aquat Invasions 6: 391-403

Shearer T, Van Oppen M, Romano S, Wörheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol. Ecol 11: 2475–87

Sherlock, RE, Walz KR, Schlining KL, Robison BH (2017) Morphology, ecology, and molecular biology of a new species of giant larvacean in the eastern North Pacific*: Bathochordaeus mcnutti sp.* Mar Biol 164: 16

Sherman CDH, Lotterhos KE, Richardson MF, Tepolt CK, Rollins RA, Palumbi SR, Miller AD (2016) What are we missing about marine invasions? Filling in the gaps with evolutionary genomics. Mar Biol 163: 198

Shiga N (1993) Regional and vertical distributions of *Oikopleura vanhoeffeni* in the Northern Bering Sea. Bull. Plank. Soc. Japan 39: 117-126

Slatkin M and Hudson RR (1991) Pairwise comparisons of Mitochondrial DNA sequence in stable and exponentially growing populations. Genetics 129: 555-562

Smith DL, Wonham MJ, McCann LD, Ruiz GM, Hines AH et al. (1999) Invasion pressure to a ballast-flooded estuary and an assessment of inoculants survival. Biol Invasions 1:67-87

Snell-Rood EC (2013) An overview of the evolutionary causes and consequences of behavioural plasticity. Ani Behav 85: 1004-1011

Stach T, Turbeville JM (2002) Phylogeny of Tunicata inferred from molecular and morphological characters. Mol Phyl and Evol 25: 408–428

Stach T (2007) Ontogeny of the appendicularian *Oikopleura dioica* (Tunicata, Chordata) reveals characters similar to ascidian larvae with sessile adults. Zoomorphology 126: 203-214

Stachowicz JJ, Terwin JR, Whitlatch RB, Osman RW (2002) Linking climate change and biological invasions: Ocean warming facilitates nonindigenous species invasions. Proc Natl Acad Sci USA, 99: 15497-15500

Stanley RRE, Dibacco C, Lowen B, Beiko RG, Jeffery NW, Van Wyngaarden M, Bentzen P, Brickman D, Benestan L, Bernatchez L, Johnson C, Snelgrove PVR, Wang Z, Wringe BF, Bradbury IR (2018) A climate-associated multispecies cryptic cline in the northwest Atlantic. Sci Adv 4: EAAQ0929

Steinke D, Zemlak TS, Boutillier JA, Hebert PDN (2009) DNA barcoding Pacific Canada's fishes. Mar Biol 156: 2641–47

Suzuki MM, Nishikawa T, Bird A (2005) Genomic approaches reveal unexpected genetic divergence within *Ciona intestinalis*, J Mol Evol 6: 627-635

Szitenberg A, Rot C, Ilan M, Huchon D (2010) Diversity of sponge mitochondrial introns revealed by *cox 1* sequences of Tetillidae. BMC Evol Biol 10: 288

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123: 585-595

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739

Tarjuelo I, Posada D, Crandall K, Pascual M, Turon X (2001) Cryptic species of Clavelina (Ascidiacea) in two different habitats: harbours and rocky littoral zones in the northwestern Mediterranean. Mar Biol 139: 455-462

Tarjuelo I, Posada D, Crandall K, Pascual M, Turon X (2004) Phylogeography and speciation of colour morphs in the colonial ascidian *Pseudodistoma crucigaster*. Mol Ecol 13: 3125-3136

Tepolt CK (2015) Adaptation in marine invasion: a genetic perspective. Biol Invasions 17:887–903

Tepolt CK, Darling JA, Bagley MJ, Geller JB, Blum MJ, Grosholz ED (2009) European green crabs (*Carcinus maenas*) in the northeastern Pacific: genetic evidence for high population connectivity and current-mediated expansion from a single introduced source population. Divers Distrib 15: 997–1009

Vallès Y, Halanych KM, Boore JL (2008) Group II introns break new boundaries: presence in a bilaterian's genome. PLoS One 1:41488

Van Oppen MJH, Catmull J, McDonald BJ, Hislop NR, Hagerman PJ, Miller DJ (2002) The mitochondrial genome of *Acropora tenuis* (Cnidaria; Scleractinia) contains a large group I intron and a candidate control region. J Mol Evol 55: 1-13

Vella A, Vella N, Schembri S (2017) A molecular approach towards taxonomic identification of elasmobranch species from Maltese fisheries landings. Mar Genomics

Wang X, Lavrov DV (2008) Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Demospongiae. PLoS ONE 3: e2723

West-Eberhard MJ (2005) Developmental plasticity and the origin of species differences. Proc Natl Acad Sci USA 102:6543-6549

Westerman EL, Whitlatch R, Dijkstra JA, Harris LG (2009) Variation in brooding period masks similarities in response to changing temperature. Mar Ecol Prog Ser 391:13-19

Whitney KD, Gabler CA (2008) Rapid evolution in introduced species, "invasive traits" and recipient communities: challenges for predicting invasive potential. Divers and Distrib 14: 569–580

Witt JDS, Threloff DL, Hebert PDN (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Mol Ecol 15: 3073-3082

Zhan A, MacIsaac HJ, Cristescu ME (2010) Invasion genetics of the *Ciona intestinalis* species complex: from regional endemism to global homogeneity. Mol Ecol 19: 4678-4694

Appendix

Appendix 1: Additional sequences and genetic diversity indices for *cytochrome oxidase 1* in *Botryllus schlosseri* used for comparison with populations in insular Newfoundland harbours, with *N*, sample size; *Nh*, number of haplotypes; *h*, haplotype diversity; and π, nucleotide diversity. Genbank accession numbers are also indicated; a, Lacoursière-Roussel et al. (2012); b, Bock et al (2012); c, Leujeuse et al. (2011); d, Lopez-Legentil et al. (2006)

Location	ID	N	Nh	Hap Code	Reference/Genbank Accession no.
NOVA SCOTIA, CANADA					
Yarmouth	P1	29	3	Bs2, HO, HB	**a** JN561069-JN561072
Point Tupper, Port Hawkesbury and Mulgrave	P3	18	4	Bs2, HO, Bs13, Bs15	
Little Narrows	P4	63	1	Bs2	
Sydney (P5)	P5	30	4	Bs2, HO, Bs4, Bs16	
Digby Marina	M1	16	3	Bs2, HO, HA	
Lunenburg Yacht Club	M2	40	3	Bs2, HO, Bs14	
Mahone Bay Classic Boat Marina	M3	34	4	Bs2, HO, Bs8, Bs14	
Oak Island Marina	M4	33	3	Bs2, HO, Bs8	
South Shore Marine	M5	30	3	Bs2, HO, Bs14	
Chester Yacht Club and The Ripe Loft Restaurant	M6	36	3	Bs2, HO, Bs14	
Hubbards Yacht Club	M7	31	2	Bs2, HO	
Shining Waters (M8)	M8	29	3	Bs2, HO, HA	
Ballantyne's Cove	M9	35	2	Bs2, HO	
Cribbon's Point	M10	34	3	Bs2, HO, Bs8	
Guysborough Marina	M11	28	2	Bs2, HO	
Canso Marina	M12	33	2	Bs2, HO	
Petit de Grat Marina (M13)	M13	40	2	Bs2, HO	
Isle Madame Boat Club	M14	35	2	Bs2, HO	
Lennox Passage Yacht Club	M15	37	1	Bs2	
St Peters Marina	M16	37	2	Bs2, HB	
Barra Strait Marina	M17	41	2	Bs2, Bs8	
Baddeck Marine and Bras d'Or Yacht Club	M18	37	2	Bs2, Bs8	
Cape Breton Boat Yard and Inverary resort	M19	39	1	Bs2	

Whycocmagy Harbour	M20	37	2	Bs2, Bs8	
Ross Ferry Marina	M21	37	2	Bs2, Bs8	
Port La Tour	PLT	16	6	Bs2, Bs4, Bs6, Bs7, HA, HO	**c** GQ365697, GQ365699, GQ365701, GQ365702, DQ340205, DQ340216
EUROPE					
Brixham, England	BRI	21	6	HH, HD, Bs18, Bs32, Bs34, Bs35	**b** DQ340209, DQ223766, JN248359, JN248373, JN248375, JN248376
Falmouth, England	FAL	43	5	HA, HB, HD, Bs18, Bs34	**b** DQ340205, DQ223766, DQ223767, JN248359, JN248375
Plymouth, England	PLY	34	6	Bs12, Bs18, Bs32, Bs34, Bs35	**b** GQ365707, JN248359, JN248373, JN248375, JN248376
Port Pendennis, England	PEN	18	3	HB, Bs18, Bs34	**b** DQ223766, JN248359, JN248375
Brest, France	BRE	37	4	HD, HH, Bs12, Bs18, Bs33	**b** GQ365707, JN248359, GQ365707, JN248359, JN248374
Canet en Roussillon, France	CER	36	4	HA, HB, Bs27, Bs29	**b** DQ340209, DQ223766, JN24868, JN248370
Estaque, France	ESQ	31	4	HA, HD*, HB*, HP	**c, d** DQ340205, DQ223766, DQ223767, DQ340217
Estrait, France	EST	16	2	HA, HB	**d** GQ340205, DQ223767
La Rochelle, France	LAR	18	2	HD*, HO	**d** DQ223766, DQ340216
Palvas, France	PAL	19	4	HA, HB, Bs28, Bs30	**b** GQ340205, DQ223767, JN248369, JN248371
Roscoff, France	RCF	35	1	Bs31	**b** JN248372
Helgoland, Germany	HEL	19	1	Bs2	**b** GQ365697
Venice, Italy	VEN	28	8	HA, HB, HO, Bs22, Bs23, Bs24, Bs25, Bs26	**b** DQ340205, DQ223767, DQ340216, JN248363, JN248364, JN248365, JN248366, JN248367
Alicante, Spain	ALI	28	5	HA, HB*, Bs3, Bs5, Bs9	**b, c** GQ365698, GQ365700, DQ365704, DQ340204, DQ223767
Arenys de Mar, Spain	ADM	19	2	HA, HB*	**c** DQ340205, DQ223767
Blanes, Spain	BLA	24	3	Bs3, HA, HB*	**b** GQ365698, DQ340205, DQ223767
Cadaques, Spain	CAD	25	2	HB, HC*	**d** DQ223767, DQ223768
Fornelos, Spain	FORN	18	3	HL, HM, HN	**d** DQ340213- DQ34215

Grana, Spain	GRA	25	4	HC*, HH, HI, HJ	**d** DQ223768, DQ340209, DQ340210, DQ340211
West Coast of North America					
Bodega Bay, CA	BBY	36	4	Bs1, Bs8, Bs10, Bs36	**b** GQ365969, GQ365703, GQ365705, JN248377
French Creek, British Columbia	FRC	28	1	Bs1	**c** GQ365696
Ladysmith, British Columbia	LSM	17	2	Bs1, Bs2	**c** GQ365696, GQ365697

*Corrected haplotypes: Haplotype HD was reported erroneously as HB, haplotype HC was reported erroneously as HD, and haplotype HB was reported erroneously as HC by c, Leujeuse et al. (2011).